

تقويم الاختبارات التحصيلية بكلية التربية والعلوم التطبيقية بأرحب جامعة صنعاء في ضوء معايير الجودة

د. عبد الحق عبد الرحيم البازلي

الأستاذ المساعد كلية التربية والعلوم التطبيقية/أرحب/ جامعة صنعاء

د. أحلام محمد محمد شرف الدين

الأستاذ المساعد قسم الدراسات العربية _ جامعة صنعاء

خلاصة البحث:

هدفت الدراسة الحالية إلى تحديد معايير جودة الاختبار، ومن ثم محاولة الوقوف على مدى التزام أساتذة التعليم الجامعي بمعايير الجودة في أوراق الاختبارات التحصيلية بكلية التربية والعلوم التطبيقية بأرحب جامعة صنعاء، وتحقيقاً لذلك قام الباحثان ببناء وتصميم مقياس تقدير جودة الورقة الاختبارية؛ شمل على (٦٤) معياراً، موزعة على ثلاثة محاور رئيسة (معايير الإطار العام للاختبار من حيث الشكل والمضمون، وجودة صياغة أنواع الأسئلة المقالية والموضوعية، وأخيراً معايير الطباعة وإخراج)، واعتمدت على المنهج الوصفي القائم على أسلوب التحليل، وتم تطبيق مقياس تقدير جودة الاختبار (أداة الدراسة)، في تحليل عينة عشوائية بلغت (٤٧) اختباراً من أوراق اختبارات نهاية الفصل الدراسي الثاني للعام الجامعي (٢٠٢١-٢٠٢٢)، وبعد جمع وتفرغ البيانات وتحليلها باستخدام برنامج الرزم الإحصائية (SPSS)، توصلت الدراسة إلى مجموعة من النتائج ويتمثل أبرزها في:

ضعف وتدني مستوى جودة الاختبارات التحصيلية المعدة من أعضاء الهيئة التدريسية بكلية التربية أرحب، إذ لم تتجاوز نسب التزامهم بمعايير وموصفات الجودة فيما يعدونه من اختبارات (٤٦%)، وبدرجة ضعيفة على مستوى المحاور الرئيسة جميعها المكونة للمقياس ككل، وقد تراوحت نسب التزامهم بمعايير محور جودة الصياغة لأنواع الأسئلة المقالية والموضوعية بأبعاده الفرعية الخمسة بين (٤٤% - ٦٢%)، وأستحق المرتبة الأولى من بين بقية المحاور الرئيسة المكونة للمقياس، في حين جاء مستوى التزامهم بمعايير الطباعة والإخراج لورقة الاختبار في المرتبة الثانية، بنسبة (٥٠%)، وبدرجة ضعيفة، بنما تراوحت نسب مستوى التزامهم بمعايير الإطار العام للورقة الاختبارية ببعديه الفرعيين من حيث الشكل والمضمون ما بين (٣٤% - ٤٦%) على الترتيب، وأستحق المرتبة الثالثة والأخيرة، وقدمت الدراسة مجموعة من التوصيات من أبرزها ضرورة تنفيذ دورات تنمية مهنية للأساتذة في مجال بناء وتصميم الاختبارات، وصياغة فقراتها، وضرورة اخضاع عينة من أوراق الاختبارات للمراجعة والتقييم بصورة مستمرة.

Evaluating achievement tests at the College of Education and Applied Sciences at Arhab Sana'a University, , in light of quality standards

Dr. Abdulhak Abdulrheem Albazeli

Assistant Professor of Curriculum and Teaching Methods at the
College of Education and Applied Sciences in Arhab - Sana'a University

Dr. Ahlam mohammed mohammed shrf al din

Assistant Professor at the Department of Arabic Language Studies,
Sana'a University

Abstract:

The current study aimed to determine test quality standards, and then attempt to determine the extent to which university education professors adhere to quality standards in achievement test papers at the College of Education and Applied Sciences at Arhab Sana'a University. To achieve this, the researchers built and designed a scale to evaluate the quality of the test paper; It included (64) standards, distributed over three main axes (the standards for the general framework of the test in terms of form and content, the quality of wording of the types of essay and objective questions, and finally the standards for printing and production). It relied on the descriptive approach based on the analysis method, and the test quality assessment scale (study tool) was applied in analyzing a random sample of (47) tests from the examination papers at the end of the second semester of the university year (2021-2022), and after collecting and transcribing the data and analyzing it using... Statistical Packages Program (SPSS), the study reached a set of results, the most prominent of which are:

The weakness and low level of quality of achievement tests prepared by faculty members at the College of Education, Arhab, as their percentages of commitment to quality standards and specifications in the tests they prepare did not exceed (46%), and to a weak degree on the level of all the main axes that make up the scale as a whole, the rates of their commitment to the standards of the quality of formulation axis for the types of essay and objective questions with its five sub-dimensions ranged between (44%-62%), and deserved first place among the rest of the main axes that make up the scale, while their level of commitment to the standards of printing and production of the test paper came in second place with percentage (50%), and to a weak degree, while the percentages of their level of commitment to the standards of the general framework of the test paper with its two sub-dimensions in terms of form and content ranged between (34% - 46%), respectively, and deserved the third and last rank. The study presented a set of recommendations, the most prominent of which is the necessity Implementing professional development courses for professors in the field of constructing and designing tests, drafting their paragraphs, and the necessity of subjecting a sample of test papers to continuous review and evaluation.

مقدمة الدراسة:

تعد عملية القياس والتقويم من العمليات المهمة والقديمة قدم تاريخ الإنسان نفسه منذ أن وجد على وجه الكرة الأرضية؛ فقد استخدم الإنسان القديم الطرائق البائبة الساذجة البسيطة في عملية القياس والتقويم، فندما كان يحاول اصطياد حيوان ما فإنه كان عليه أن يقدر المسافات وتحديد حجم الصخرة التي يقذفه بها لاصطياده ، وعندما كان ينظر إلى جماعته من حوله فإنه يعرف أن زميله الضخم أقوى من في الجماعة ، أن الآخر اسرعه في العدو ، كل هذه الأمور الحياتية وغيرها تحتوي على عمليات القياس والتقويم، (مجيد، سوسن: ٢١٠٤، ص ١١)

وفي الوقت الحاضر يعني النظام التربوي والتعليمي بعملية القياس والتقويم ؛ إذ أن نتائج التقويم توجه المربين نحو اختيار أهداف معينة وتحقيقها فهي عملية ضرورية للمعلم والمتعلم.

وللتقويم دور كبير في تطوير العملية التعليمية حيث يعد الأساس الذي يبنى عليه أي تطوير، ومما يضاعف من أهمية التقويم أيضا هو تغير وتطور دور التعليم الجامعي وظائفه؛ حيث يؤكد الخبراء أن وظيفة الجامعة قد تغيرت وتطورت ومازالت تتطور عما كانت عليه في بداية ومنتصف القرن العشرين، وحتى نهايته، فقد أصبح هناك توجه نحو اقتصاد المعرفة ومجالات إنتاجها ووسائلها... إلخ.

وللتقويم أدوات عدة أهمها: الاختبارات ؛ حيث تحتل موقع الصدارة بين وسائل التقويم المستخدمة في الجامعة ، بل قد تكون الوسيلة الوحيدة أحيانا، وهذا يعني أن الأحكام التي تصدر في الجامعة وعلى الأصعدة جميعها يتم في ضوء نتائج تلك الاختبارات بالدرجة الأولى.

وتحتل الاختبارات النهائية في الجامعة أهمية أكبر؛ حيث تستحوذ على نسبة ٧٠% من الدرجة الكلية في أغلب المقررات الدراسية في بعض الكليات ، أي أنها أهم أداة للحكم على نجاح أو رسوب الطالب الجامعي، (الصلاحي، ١٩٩٨م، ص ١٥).

ومما يضاعف من أهمية الاختبارات في الجامعة وجود العديد من التحديات المعاصرة التي تواجه الأمة والتي فرضت معايير عالمية للجودة والنوعية في إعداد وتأهيل الطالب وفي مخرجات التعليم الجامعي ؛ حيث أن مخرجات الجامعة قد يخضعون للتقويم من جديد وفقا لمعايير علمية للحصول على الرخصة الدولية في التخصص، وهذا يفرض على التقويم المتبع مراعاة المعايير العلمية، وإلا فإن ذلك قد يؤدي إلى أضرار عدة تمس هذه المخرجات وتمس سمعة النظام التربوي في المجتمع ككل.

إن هذه الأهمية التي تحتلها الاختبارات الجامعية يفرض عليها مسؤولية كبيرة، ولذلك يجب أن تتصف بالمعايير العلمية ومعايير الجودة التي تتناسب وأهميتها، وإلا أصبحت أداة مضللة تقدم بيانات غير دقيقة قد تقود إلى أحكام غير صادقة، وهنا تكون العواقب وخيمة على عملية التقويم الجامعي والمجتمع، ونظراً لذلك ولكي تؤدي الاختبارات دورها ينبغي أن توضع وفقاً لأسس ومعايير علمية ذات جودة، وأيضاً تقويمها بشكل مستمر وبالتالي تطويرها، (هاشم: ٢٠١٤م، ص ١٧).

لذا، هدفت الدراسة الحالية إلى تحديد معايير جودة الاختبار، ومن ثم محاولة التعرف على مدى توافر هذه المعايير في أوراق الاختبارات التحصيلية المعدة من أعضاء الهيئة التدريسية بكلية التربية والعلوم التطبيقية بأرحب جامعة صنعاء، وتحقيقاً لذلك سيتم إعداد وتصميم مقياس تقدير جودة الورقة الاختبارية؛ سيشمل على (٦٤) معياراً، موزعاً على ثلاثة محاور رئيسة (معايير الإطار العام للاختبار من حيث الشكل والمضمون، وجودة صياغة أنواع الأسئلة المقالية والموضوعية، وأخيراً معايير الطباعة وإخراج)، معتمدين على المنهج الوصفي القائم على أسلوب التحليلي، ومن ثم سيقوم الباحثان بتحليل عينة من الاختبارات النهائية للفصل الدراسي الثاني للعام الجامعي (٢٠٢١-٢٠٢٢) المقدمة للطلبة الكلية في للأقسام العلمية والأدبية.

الفصل الأول.

مشكلة الدراسة:

لمس الباحثان بحكم عملهما الأكاديمي في كلية التربية والعلوم التطبيقية بأرحب - جامعة صنعاء - أن ما يطبق من الاختبارات التحصيلية بالكلية يشوبها الكثير من أوجه القصور والنقص، وأنها تفتقد إلى الكثير من الأسس والمعايير العلمية والشروط والمواصفات الفنية التي ينبغي أن يلتزم بها الأساتذة، فضلاً عن مخالفة نماذج أسئلة الاختبار لقواعد وأسس الصياغة العلمية للسؤال الجيد، مما يعطي مؤشراً أن تلك الاختبارات يتم إعدادها بطريقة عفوية وارتجالية، ولا تخضع في بنائها لأسس وشروط وقواعد البناء للاختبار، كما تكشف عن تدني مهارات بناء وتصميم وإخراج الاختبار وصياغة فقراته لدى معديها، الأمر الذي نتج عنه وجود اختبارات تفتقر إلى الدقة والصحة، بل وتفتقد إلى الوثوق بمحتواها علمياً أو الاطمئنان إلى القرارات التي تتخذ في ضوءها.

وفي السياق ذاته تؤكد نتائج معظم الدراسات عن وجود حالة من القصور والضعف في نسب استيفاء معايير ومواصفات الجودة في أوراق الاختبارات، كدراسة ساعد وعامر (٢٠١٧) التي خلصت إلى أن توافر معايير كتابة تعليمات الاختبار وإعداد فقراته كانت بدرجة ضعيفة، وما كشفته دراسة السفياني (٢٠١٩) من ضعف نسب توافر معايير المضمون والشكلية للورقة.

وعلى الرغم من تناقض الدراسات السابقة؛ إلا أنها وبصفة عامة قد أكدت في توصياتها على أهمية تقويم وتحليل أسئلة الاختبارات للوقوف على جوانب القوة والضعف فيها، وضرورة تنمية مهارات بناء وتصميم الاختبارات وصياغة فقراتها لدى معديها، من هنا جاءت هذه الدراسة للتعرف على مدى التزام أساتذة التعليم الجامعي بمعايير جودة الاختبارات، والكشف عن مستوى جود هذه الاختبارات في كلية التربية بأرحب جامعة صنعاء، ويمكن تحديد مشكلة الدراسة إجرائياً في السؤال الرئيس التالي:

ما مدى جودة الاختبارات التحصيلية في كلية التربية والعلوم التطبيقية بأرحب- جامعة صنعاء في ضوء معايير الجودة الشاملة؟

ويتفرع من هذا السؤال الأسئلة الفرعية التالية:

- ١- ما مدى التزام معدي الاختبارات بمواصفات الورقة الاختبارية من حيث (الشكل والمضمون)؟
- ٢- ما مدى التزام معدي الاختبارات بمعايير كتابة أنواع الأسئلة الاختبارية (المقالية، والموضوعية)؟
- ٣- ما مدى التزام معدي الاختبارات بمعايير الطباعة والإخراج للورقة الاختبارية ؟

أهمية الدراسة:

تتجلى أهمية القيام بهذا الدراسة في أن نتائجها قد:

- ١- تساعد في الكشف عن مدى كفاءة أساتذة التعليم الجامعي بكلية التربية أرحب - جامعة صنعاء في بناء وتصميم الاختبارات التحصيلية، ومدى التزامهم بمعايير الجودة للورقة الاختبارية، ونسب ما تم استيفاءه أو إغفاله منها، ومن ثم بيان وتحديد جوانب القوة، وأبرز مواطن القصور فيما يعدونه من اختبارات، والمجالات التي يجب زيادة الاهتمام بها، وترتيبها حسب أهميتها، وبالتالي تجنب ما قد يترتب عليه من آثار سلبية على العملية التربوية والتعليمية، أي أن نتائجها قد تفيد الباحثين.
- ٢- لفت أنباه أساتذة التعليم الجامعي إلى أهمية التزامهم بمعايير ومواصفات الجودة في بنائهم للاختبارات التحصيلية، وأهمية تحكيمهم لها وتطويرها.
- ٣- تقديم تغذية راجعة لأساتذة الكلية للوقوف على جوانب القوة والضعف في ما يعدونه من اختبارات، وإعادة النظر في كيفية بنائها وصياغتها، والعمل على تحسين وتطوير ممارساتهم في هذا المجال، وبما يسهم في تجويد الاختبارات مستقبلاً.
- ٤- توفر معلومات وبيانات لأصحاب القرار والمعينين بالجودة والقياس والتقويم للوقوف على طبيعة الاختبارات داخل الجامعة والكليات التابعة لها، والتعرف على جوانب القوة ومواطن الضعف فيها،

وتعديل بعض الممارسات أو تقنياتها، وتوجيه السياسات ووضع الخطط التطويرية لأنظمة التكوين بالجامعة، واعداد برامج التدريب الملائمة لتنمية مهارات الأساتذة في مجال بناء الاختبارات.

٥- توفر أداة قياس معيارية محكمة لتقييم أوراق الاختبارات تحصيلية في ضوء معايير الجودة، الأمير الذي يشكل مدخلاً لتحسين وتطوير الاختبارات التحصيلية داخل الجامعة وكلياتها المختلفة.

٦- كما ترجع أهميتها إلى كونها من أوائل الدراسات تناولاً لجودة الاختبارات في كلية التربية أرحب، وإسهاماً لإثراء المكتبة اليمنية بهذا النوع من الدراسات، وفتح آفاق جديدة للقيام بدراسات مماثلة.

أهداف الدراسة:

تسعى الدراسة الحالية إلى تحقيق الأهداف الآتية:

التعرف على مدى التزام أستاذ التعليم الجامعي بمعايير جودة الإطار العام للورقة الاختبارية منحيث (الشكل والمضمون)، اللازم مراعاتها في الاختبارات التحصيلية الجيدة.

الكشف عن مدى التزام أساتذة التعليم الجامعي بمعايير صياغة أنواع فقرات الأسئلة الاختبارية (المقالية، والموضوعية) فيما يقدمونه من الاختبارات التحصيلية النهائية للطلبة.

الوقوف على مدى التزام أساتذة التعليم الجامعي بمعايير الطباعة والإخراج للورقة الاختبارية في الاختبارات التحصيلية النهائية المقدمة للطلبة.

حدود الدراسة:

التزمت الدراسة بالحدود الآتية:

حدود زمانية: تم إجراء الدراسة في نهاية الفصل الثاني من العام الجامعي ٢٠٢١-٢٠٢٢.

حدود مكانية: أقتصر تطبيق مقياس جودة الاختبار على الاختبارات التحصيلية النهائية التي أعدها أعضاء الهيئة التدريسية بكلية التربية والعلوم التطبيقية بأرحب - جامعة صنعاء، لطلبة الأقسام العلمية (كيمياء، فيزياء، رياضيات)، والأقسام الأدبية (إسلامية، قرآن، عربي، تاريخ، جغرافيا).

حدود موضوعية : سيتم دراسة جودة الاختبارات في ضوء ثلاثة محاور رئيسة هي: معايير الإطار العام للورقة الاختبارية: وله بعدين فرعيين، من حيث (الشكل، والمضمون)، ومعايير صياغة أنواع فقرات الأسئلة الاختبارية، وله خمسة أبعاد فرعية هي: الأسئلة (المقالية، الصح والخطأ، الإكمال، اختيار من متعدد، المطابقة)، وأخيراً محور معايير الطباعة والإخراج.

مصطلحات الدراسة:

١- الاختبار التحصيلي:

- يعرفه علام بأنه: "تلك الأسئلة التي يهدف من خلالها إلى تقويم تحصيل المتعلمين الدارسين في نهاية الفترات أو النقل أو الشهادات العامة، وتعد من أهم وسائل تقويم التحصيل وتحديد مستوى الطالب التحصيلي".
- ويقصد بها في هذه الدراسة: الاختبارات التحصيلية التي أعدها أعضاء هيئة تدريس للمقررات النظرية لقياس مستوى تحصيل طلبة الأقسام المختلفة بكلية التربية أرحب - جامعة صنعاء، وذلك خلال نهاية الفصل الدراسي الثاني من العام الجامعي ٢٠٢١.

٢- معايير جودة الاختبار:

عرف المتخصصون المعايير بعدد من التعريفات، فقد عرفت بأنها:

- "المحك أو الإطار المرجعي الذي يحدد على أساسه الأداء الواقعي في مدى الابتعاد أو الاقتراب من المراجع". (السنبل: ٢٠٠١، ص ٥٢).
- مستوى مرجعي للقياس يستدل عليه من خلال مجموعة من المؤشرات تستخدم كمواصفات للأداء النموذجي. (Arter&mctighe: ٢٠٠١، ٦٤)
- "مجموعة من الشروط والأحكام المضبوطة علمياً التي توظف كأساس للمقارنة والحكم على القيمة النوعية أو الكمية لتشخيص مواطن القوة لتعزيزها، وجوانب الضعف لعلاجها". (حلس: ٢٠٠٤، ص ١١)

وتعرف معايير جودة الاختبار بأنها: مجموعة من الأسس والمعايير العلمية والشروط والمواصفات الفنية التي ينبغي توافرها في أوراق الاختبارات التحصيلية، وأن يلتزم بها الأساتذة بنسبة ٨٠% في عينة الأوراق الاختبارية، وذلك من خلال قياس متوسط ونسب توافرها في ضوء مقياس تقدير جودة الورقة الاختبارية المعتمدة كأداة في هذه الدراسة.

٣- كلية التربية والعلوم التطبيقية بأرحب:

هي إحدى فروع كليات التربية التابعة لجامعة صنعاء، وتقع في مديرية أرحب، تأسست عام (١٩٧٠)، وتضم عدد من الإقسام التربوية لإعداد وتأهيل المعلمين، وتنمخ درجة البكالوريوس، في الأقسام الأدبية (علوم القرآن، الدراسات الإسلامية، اللغة العربية، لغة انجليزية، التاريخ، الجغرافيا)، والأقسام العلمية (رياضيات، فيزياء، كيمياء وعلوم الحياة، الحاسوب).

الفصل الثاني.

أولاً: الخلفية النظرية.

أولاً: مفهوم الاختبارات التحصيلية:

طريقة منظمة لتحديد مستوى تحصيل الطالب للمعارف والمهارات والاتجاهات في مادة دراسية، سبق له تعلمها بصفة رسمية، من خلال إجاباته على عينة من الأسئلة أو الفقرات التي تمثل محتوى المادة الدراسية (أبو مغلي، وسلامة: ٢٠١٠، ص ٢٧)، أي أنه: إجراء منظم لقياس عينة ممثلة من المهمات أو المخرجات التعليمية. (عبد المقصود: ٢٠٠٩، ص ٢٩١)

مما سبق يتضح أن الاختبار التحصيلي عبارة عن عينة مختارة من الأسئلة (المقالية والموضوعية) التي يقررها واضع الاختبار، مراعيًا مجموعة من القواعد والأسس العلمية، لقياس مدى تحقيق الأهداف التعليمية المحددة مسبقاً لدى المتعلمين، ومن ثم الحكم على الدرجة التي حدث بها التعلم بصورة كمية، وتقع الأسئلة التحريرية (الكتابية) على عدة أشكال لكل منها إجابياتها وسلبياتها.

ثانياً: أهداف الاختبارات التحصيلية:

تستخدم اختبارات التحصيل في كل أنواع التقويم (التمهيدي، والبنائي، والختامي) ولأغراض مختلفة، ويؤكد (العجيلي، ٢٠١١، ص ٥٥) أن الاختبارات تستخدم في التقويم الختامي لأغراض بالغة الأهمية، واتخاذ قرارات حاسمة تتعلق بالنجاح والرسوب، ونقل الطلبة إلى مستويات عليا، أو تخرجهم ومنحهم شهادات تخرج لمرحلة معينة، فضلاً عن تحديد جوانب الضعف والإخفاق في المناهج والمقررات الدراسية، وطرائق وأساليب التدريس المتبعة.

ويستنتج من خلال ما سبق أن أهداف الاختبارات التحصيلية تتمثل في الآتي:

١. وصف المستوى العام لتحصيل الطلبة العلمي، ومستوى تحقيقهم لأهداف المقرر.
٢. قياس الفارق بين الأهداف المتوقعة والأهداف المحققة.
٣. الكشف عن الفروق الفردية بين الطلبة، وتحديد نقاط القوة والضعف لديهم.
٤. تحديد ورصد علامات الطلبة في سجلات خاصة، وإصدار أحكام تتعلق بهم كالنجاح والإكمال والرسوب.
٥. اتخاذ قرارات مهمة بنقل الطلبة من مستوى لآخر، أو منح شهادة تخرج في التخصصات والكليات المختلفة.
٦. إجراء مقارنة بين نتائج الطلبة في الأقسام المختلفة، ونتائجهم لمقرر في البرنامج نفسه أو مع برامج نظير في مؤسسة أخرى.
٧. الحكم على جهود الأساتذة، وفاعلية أساليب وطرائق تدريسهم، والمقررات وأساليب التقويمية المتبعة.

ثالثاً: أسس بناء الاختبار التحصيلي:

- يورد مجموعة من الباحثين منهم (Nuryulia, I. R : ٢٠٠٩, ٢٥ عدد من الأسس التي يجب أن يلتزم بها الأساتذة أثناء بناء الاختبارات التحصيلية، لعل أهمها تتمثل في الآتي:
- ١- أن يتم التخطيط لها بشكل جيد، ويتم بنائها وفق اسس علمية دقيقة.
 - ٢- أن تحدد أهداف ومخرجات التعلم التي سيقاسها الاختبار بشكل دقيق.
 - ٣- أن تتنوع الأسئلة المستخدمة في الاختبار لقياس نواتج ومخرجات التعلم المطلوب قياسها.
 - ٤- أن تكون أسئلة الاختبار مما درسه الطالب في المحتوى الدراسي.
 - ٥- أن يسعى الاختبار إلى توظيف الحقائق والمبادئ المختلفة التي درسها الطلبة.
 - ٦- أن يتناسب الاختبار مع مستوى نمو الطلبة المراد تطبيق الاختبار عليهم.
 - ٧- يجب أن يحتوي الاختبار على أسئلة موضوعية وأخرى مقالية، يمكن من خلالها قياس قدرات الطلبة على التنظيم والتحليل والتركيب وتطبيق القواعد على مواقف جديدة.

رابعاً: شروط الاختبار الجيد:

- (١) الارتباط بالأهداف: لما كانت غاية التقويم هي التعرف على مدى تحقيق الأهداف، لذا يجب أن يرتبط الاختبار بمخرجات التعلم وأهداف توصيف المقرر.
- (٢) الشمول: فيجب أن تكون أسئلة الاختبار شاملة لمخرجات التعلم وأهداف التدريس المحددة في توصيف المقرر، في جوانب التعلم المعرفية والمهارية والوجدانية، وشاملة للمحتوى.
- (٣) الموضوعية: ويقصد به عدم تأثر درجات الاختبار بذاتية المصحح وانطباعاته، لتضمنها مؤشرات محددة مسبقاً تمنع الذاتية، ومن ثم لا تختلف تقدير درجة الطالب في الاختبار باختلاف المصححين، ومما يساعد على تحقيق الموضوعية أن يفهم الطلاب تعليمات الاختبار بدقة، وأن يكون هناك تفسير واحداً للأسئلة والإجابات المطلوبة منه، والعدل في إعطاء الدرجات.
- (٤) الصدق: ليكون الاختبار صادقاً يجب أن يقيس ما وضع لقياسه، وينقسم الصدق نفسه على أنواع منها:

- الصدق الظاهري: يكون الاختبار صادقاً ظاهرياً إذا كانت فقراته تقيس السلوك أو السمة المراد قياسها، ويهتم هذا الصدق بنوعية الفقرات وطريقة صياغتها، وبوضوح التعليمات، ويسمى صدق البناء، ويتحقق بالعرض على المحكمين.
- صدق المحتوى: ويتمثل في درجة تمثيل الفقرات للسّمات المراد قياسها، ويهتم بشكل كمي بعدد الفقرات ومدى ملائمتها، ومدى تغطيتها للمحتوى والسّمات المراد قياسها، أي بتوزيع الأسئلة على

صفحات وفصول وأهداف الكتاب، وأن يتضمن عدد كافي من الأسئلة لتمثيل المادة، وأن لا يتضمن أسئلة اختيارية؛ لأن ذلك يؤثر على درجة الصدق، ويتحقق بآراء المحكمين، وبجدول المواصفات.

٥) الثبات: أي تعطينا الاختبار نتائج ثابتة في حالة إعادة تطبيقه أكثر من مرة على الطلاب تحت ظروف نفسها، أي الدرجة نفسها أو قريباً منها حالة تكرار عملية القياس، وهو مؤشر لمقدار خلو الدرجات من الأخطاء غير المنتظمة في القياس، ومما يؤثر على الثبات التعليمات الغامضة، فإذا كانت تعليمات الاختبار غير واضحة، أو قابلة للفهم والتفسير على عدة أوجه؛ فإنها تساهم في ضعف ثبات الاختبار، ويجعل الطالب يكسب بعض العلامات بالصدفة ويخسرهما بالصدفة أيضاً.

٦) التمييز: وهي قدرة الاختبار على إظهار ما بين الطلبة من فروق فردية بشكل واضح.

خامسا: أنواع الاختبارات التحصيلية:

تصنف الاختبارات إلى أنواع كثيرة تبعاً لمجموعة من المعايير، (العبادي: ٢٠١٨، ص ١٦١٩-١٦٣١)، وعلى النحو الآتي:

- ١) تصنف الاختبارات تبعاً لطريقة التفسير إلى: اختبارات محكية المرجع، والمحك هو مستوى أداء معين محدد مسبقاً، قد يكون كمياً كتحديد الجامعات درجة النجاح بـ (٥٠%)، أو نوعياً كمقارنة أداء الفرد بمستوى أداء محدد في الأهداف، واختبارات معيارية المرجع تهتم بمقارنة أداء الفرد بالجماعة.
- ٢) تصنف الاختبارات تبعاً لزمان الإجابة إلى: اختبارات السرعة، وتهتم بقياس كم الإنجاز خلال زمن معين، واختبار الكفاءة أو القوة، ولا يرتبط بزمن محدد.
- ٣) تصنف الاختبارات تبعاً للقائمين على إعدادها إلى: اختبارات غير مقننة (غير رسمية)، تهتم بقياس أهداف محددة أو محتوى معين، واختبارات مقننة رسمية، وهي التي تقيس أجزاء واسعة من المعلومات والمهارات، وتخضع للتجربة والمرجعة والتحليل.
- ٤) تصنف الاختبارات تبعاً لطريقة تقديمها أو أداء الطالب إلى: اختبارات شفوية تتطلب إجابات لفظية، كاختبار القدرة اللغوية، واختبارات أدائية: وتهتم بقياس بعض الجوانب المهارية والفنية، واختبارات تحريرية (كتابية)، والتي تسمح للطلاب بإنشاء وتنظيم الإجابة، وتكون مفردات الاختبار على شكل أسئلة مقالية أو موضوعية، أو رسوم وأشكال ...

سادسا: مواصفات ومعايير جودة الورقة الاختبارية:

وجهت مختلف نظم التعليم في العالم الاهتمام بالمواصفات والمعايير التي تضمن جودة المخرجات بداية من أدوات القياس ومنها الاختبارات لتحسين نوعية التعليم تماشياً مع الاتجاهات المعاصرة.

ومن هنا فإن المفهوم الحديث للقياس والتقويم يركز توجيه الاهتمام إلى معيارية بناء أداة القياس والتقويم، وإعدادها ضمن معايير ومواصفات وشروط الإعداد والبناء، ويقوم بها معلمون ذو كفاءات مهنية ومهارية عالية، بحيث تستطيع أداة القياس قياس المحتوى المستهدف بدقة للحصول على نتائج حقيقية يترتب عليها إصدار الاحكام التقويمية الصحيحة، لاتخاذ قرارات تربوية وتعليمية فاعلة تسهم في تحسين العملية التعليمية. (Stufflebeam, D. 1997, 11))

أي أن الاختبار لا يكون فاعلاً إلا إذا كان معيارياً في بنائه، ومستوفياً لمواصفاته وشروطه (الحيلة: ٢٠٠٢، ٢٥)، لأن أي قصور في عملية إعدادها سيؤدي حتماً إلى خلل في عملية التقويم، والتأثير على العملية التعليمية برمتها، لذلك يحدد العلماء مجموعة من معايير ومواصفات الجودة للورقة الاختبارية، والتي يجب على القائم ببنائها الالتزام بها أثناء إعداد وتصميم الاختبارات التحصيلية، سواء المعايير والمواصفات المتعلقة بالإطار العام للورقة الاختبارية من حيث الشكل ومن حيث المضمون، أم المتعلقة بجودة الصياغة لأنواع الفقرات والأسئلة الاختبارية، أم المتعلقة بطباعة وإخراج الورقة الاختبارية، وعلى النحو الآتي:

أولاً: المعايير الشكلية والفنية للورقة الاختبارية:

وتتعلق هذه المعايير بمواصفات الورقة الاختبارية، يمكن تصنيفها ضمن نوعين رئيسيين هما معايير الورقة الاختبارية من حيث الشكل، ومعايير الورقة من حيث المضمون، وعلى النحو الآتي:

١- المعايير الشكلية للورقة الاختبارية:

تهتم الموصفات الشكلية بكليشة الورقة، وتعليمات الاختبار (العامة، والخاصة)، ومن تلك المواصفات (جيوسي: ٢٠١٦، ص ٢٧) ما يأتي:

١- مقدمة ورقة الاختبار (كليشة الورقة الاختبارية): وتسمى ترويسة أو كليشة الورقة الاختبارية، والتي

ينبغي أن تحتوي على المعلومات العامة والبيانات الأساسية الآتية:

- في الجانب الأيمن (اسم الجامعة، والكلية، والقسم، الشعبة) ضمن مربع مخفي.
- في الجانب الأيسر (المادة، التخصص، المستوى، يوم وتاريخ الاختبار، الزمن الكلي للاختبار).
- وسط الجانبين السابقين يوضع شعار الجامعة.
- بيانات نوع الاختبار المقرر وزمنه (فصلي، سنوي)، ويكتب بينط عريض بارز وتوسيط.
- أن تتضمن ورقة الاختبار بيانات الطالب (الاسم، التخصص، الرقم الجامعي..).

- ٢- التعليمات العامة للاختبار: وهذه التعليمات تمثل إرشادات مهمة توجه الطلبة وترشدتهم إلى كيفية الاستجابة لفقرات الاختبار، وتكتب بعد الإكليشة مباشرة، ويجب أن توضح التعليمات العامة:
- الدرجة الكلية للاختبار، العدد الكلي للأسئلة، عدد الصفحات.
 - مكان الإجابة على الأسئلة، وطريقة ونوع الإجابة المطلوبة بدقة ووضوح.
 - أن يتم الفصل بينها وبين الأسئلة بفواصل (خط)، ويرى البعض أن تكون التعليمات في صفحة مستقلة وخاصة إن كانت كثيرة.
 - أن توضح البيانات المطلوبة باستخدام الغلاف الموحد لورقة الاختبار.

ونظراً لأهمية هذه التعليمات يؤكد (كاظم: ٢٠١٠، ص ٩٣) أنه يجب:

- أن تكون التعليمات مصاغة في عبارات موجزة مفهومة ومبسطة وقصيرة.
- أن تكون متسلسلة على شكل نقاط واضحة.

٣- التعليمات الخاصة بالأسئلة والفقرات الاختبارية: ويقصد بالتعليمات الخاصة التعليمات التي تأتي بداية كل سؤال من أسئلة الاختبار، لترشد وتوجه الطلبة، كما تقلل من كثرة استفسارات الطلبة أثناء أدائهم للاختبار، وتؤكد (الريماوي: ٢٠١٧، ص ٧٤) أنه ينبغي:

- أن تكون تعليمات كل سؤال واضحة ومحددة ومكتوبة بخط بارز لتقليل احتمالية عدم انتباه الطلبة لها.
- أن توضح الدرجة المرصودة لكل سؤال وفقراته الفرعية، وتوزيعها حسب أهمية كل منها.
- أن توضح مكان الإجابة عن فقرات السؤال، وطريقة الإجابة، ونوعية وكيفية الإجابة.

٢- معايير المضمون للورقة الاختبارية:

يتفق عدد من الباحثين منهم (العبادي: 2006، ص ١٤)، و (مخائيل: ٢٠١٦، ص ١٢٨-١٣٠)، وقطامي، يوسف وقطامي، نايفه: ٢٠٠١، ص ٥٦٧)، إلى أن هناك مجموعة من المعايير التربوية التي يجب الالتزام بها لإعداد اختبار تحصيلي جيد، من أبرزها:

- ١- شمول الأسئلة الاختبارية لموضوعات ومفردات ومحتوى المقرر.
- ٢- قياس أسئلة الاختبار مخرجات التعلم في جوانب التعلم الثلاثة للمقرر.
- ٣- أن يصاغ كل سؤال بطريقة واضحة ودقيقة، ويحدد المهمة المطلوبة بشكل مفهوم لكل طالب.
- ٤- تجنب وضع أسئلة تتسم بالعمومية والغموض والتعقيد اللفظي، أو القابلة للتأويل.
- ٥- أن يقيس كل سؤال في الاختبار فكرة أو ناتجاً تعليمياً واحداً، فلا يتضمن أي أسئلة مركبة.
- ٦- كل سؤال في الاختبار مستقل بذاته، ولا يعد الإجابة عنه شرطاً للإجابة عن سؤال يليه.
- ٧- تقيس أسئلة الاختبار مستويات التفكير العقلية العليا والدنيا المختلفة.
- ٨- تنوع أسئلة في الاختبار ما بين الأسئلة المقالية، والموضوعية.

- ٩- ترتيب أسئلة الاختبار منطقياً بحسب (مستوى الصعوبة، موضوعات المقرر، نوع الفقرة)، لأن ذلك يشكل ويوفر حافزاً ودافعاً للمتعلم لاستكمال الإجابة على الاختبار.
- ١٠- أن يتناسب عدد الأسئلة مع الزمن المحدد للاختبار.

ثانياً: الدراسات السابقة:

تعددت الدراسات التي تناولت موضوع الاختبارات التحصيلية تبعاً لتعدد جوانب الموضوع والمتغيرات التي تناولتها كل دراسة، وفيما يأتي عرضاً لأهم الدراسات التي أتيحت للباحثين:

دراسة هلال السفيني (٢٠١٩): هدفت إلى التعرف على مدى توافر مواصفات إعداد ورقة الامتحان الشكلية والضمنية، وكذا معايير صياغة فقراتها، وتحقيقاً لذلك قام بإعداد استبانة تكونت من قسمين (مواصفات الامتحان الجيد، ومعايير صياغة فقرات الامتحان)، واستخدم المنهج الوصفي التحليلي، وتكونت عينة الدراسة من (٨) اختبارات نهائية من اختبارات المقررات التربوية بكلية التربية المهرة، توصلت إلى مجموعة من النتائج من أبرزها: أن نسبة توافر المواصفات الجيدة للورقة الاختبارية التي يعدها أعضاء الهيئة التدريسية بكلية التربية في المهرة كانت بدرجة عالية، ونسبة (٧١.٨٩%)، بينما جاءت معايير صياغة فقرات الأسئلة الموضوعية المرتبة الأولى بنسبة (٨٦.٢%)، يليها المقالية، في حين توافرت المواصفات الشكلية للورقة بنسبة (٦٩.٤)، بينما توافرت المعايير الضمنية لورقة الامتحان بنسبة (٥٩%)، وبدرجة متوسطة، وحصدت المرتبة الأخيرة

كما أجرت أماني مكاي (٢٠١٩): دراسة هدفت إلى التعرف على جودة الاختبارات التحصيلية في التعليم الجامعي، من خلال الوقوف على مستوى معارف أعضاء هيئة التدريس بمعايير ومواصفات الاختبار الجيد، ومستوى التزامهم بها في ورقة الاختبار، واستخدمت الاستبانة كأداة لجمع البيانات، وحددت المجتمع الأصلي في جميع الأساتذة بكلية التربية بعفيف، وشملت العينة (٤٦) أستاذاً، ومن أبرز النتائج التي توصلت إليها: أن معرفة أعضاء هيئة التدريس بكلية التربية بعفيف بمعايير الاختبار كان بدرجة متوسطة، وأن مستوى التزامهم بها كان بدرجة جيد.

بينما هدفت دراسة كلا من ساعد وعامر (٢٠١٧): إلى التعرف على مدى توافر معايير الاختبار التحصيلي الجيد في الاختبارات المعدة من قبل أساتذة التعليم الجامعي، وتحقيقاً لذلك قام الباحثان بإعداد قائمة بمعايير الاختبار الجيد، واستخدما المنهج الوصفي التحليلي، وطبقا الأداة على عينة عشوائية بلغ حجمها (٣٦) اختباراً تحصيلياً من اختبارات جامعة محمد خيضر - الجزائر، للعام الجامعي (٢٠١٢-٢٠١٣)، وخلصت الدراسة إلى عدد من النتائج من أبرزها أن مستوى توافر معايير كتابة تعليمات الاختبار التحصيلي وإعداد فقراته كانت بدرجة ضعيفة، أما معايير الطباعة والإخراج .

فكانت بنسبة متوسطة، وأن أكثر الأسئلة استخداماً لدى الأساتذة الأسئلة المقالية.

وأجرى كل من الحسينان، وآخرون (٢٠١٤): دراسة هدفت إلى تقويم اختبارات المقررات التربوية في كلية التربية بجامعة المجمعة في ضوء معايير الجودة، للوقوف على مدى توافر تلك المعايير في أوراق الاختبار، وتحقيقاً لذلك أعدى قائمة بمعايير جودة الورقة الاختبارية، ضمن خمسة محاور هي (البيانات الأساسية - التعليمات - لغة الاختبار - ارتباط الاختبار بالتوصيف - مضمون الاختبار)، وقاما بتحليل عينة من اختبار المقررات التربوية للأعوام من (1431 وحتى 1435 هـ) في ضوء تلك القائمة، ومن أهم النتائج التي توصلت إليها أن استيفاء المعايير (البيانات الأساسية وتعليمات الاختبار - ولغة الاختبار) بنسبة ٩٠%، وأن أعلى نسب استيفاء المعايير معيار لغة الاختبار بنسبة (٩٦%)، وأدنى نسبة مئوية للاستيفاء كان لمعيار مضمون الاختبار، بنسبة ٦٨%.

بينما هدفت دراسة القرشي (٢٠١٠): إلى تقويم الورقة الاختبارية لطالبات البكالوريوس بكلية التربية جامعة الطائف وفق معايير الجودة؛ وتحقيقاً لذلك قامت الباحثة ببناء قائمة بمعايير الجودة، واعتمدت على المنهج الوصفي التحليلي، واستخدمت الأداة لتحليل عينة من الأوراق الاختبارية لأعضاء هيئة التدريس بالكلية، بلغت (٣٠) ورقة من أوراق اختبارات لفصل الدراسي الأول للعام الجامعي (١٤٣١)، ومن أبرز النتائج التي توصلت إليها: أن الاختبارات تفتقر إلى الصدق والشمول والموضوعية، كما يركز معظمها على قياس التذكر والحفظ مع إهمال المستويات العليا، وأن هناك معوقات تحول دون تصميم الورقة الاختبارية وفق معايير الجودة.

كما أجرى عبدالمنعم (٢٠٠٩): دراسة هدفت إلى تحليل أسئلة الاختبارات التحصيلية بكلية المعلمين جامعة الملك فيصل في ضوء معايير الجودة، للتعرف على واقع أسئلة الاختبارات الفصلية، واستخدمت الدراسة المنهج الوصفي التحليلي، واعتمدت الاستبانة كأداة لجمع البيانات، وتمثل المجتمع في اختبارات الفصل الدراسي الثاني ٢٠٠٧ / ٢٠٠٨، للطلاب الخريجين من البكالوريوس، وطبقت الأداة على عينة بلغت (٣٠) ورقة اختبارية في التخصصات الأدبية والعلمية بالكلية، وقد خلصت الدراسة إلى أن الاختبارات اقتصرت على قياس التحصيل في الجانب المعرفي، وفي أدنى مستوياته التذكر، بينما أهملت قياس مستويات التفكير العليا، وجوانب التعلم الأخرى، وافتقارها للتنوع والشمول.

التعليق على الدراسات السابقة:

يتضح من خلال الدراسات السابقة أن هذه الدراسات قد ركزت واهتمت بجوانب متعددة؛ منها ما يتصل بتحليل الأسئلة وتقويمها، ومنها ما أهتم بمدى إلمام الأساتذة بالجوانب المعرفية للجودة، ومدى التزامهم بها، كما أقتصرت بعضها على اختبارات المقررات التربوية، وبعضها الآخر على السنة الرابعة فقط، وقد اختلفت وتباينت تلك الدراسات في النتائج التي توصلت إليها، فبينما أكد بعضها أن توافر معايير الجودة في أوراق الاختبارات كان بدرجة عالية، توصلت أخرى إلى أن توافرها كان بدرجة ضعيفة، أو متوسطة.

وتتفق الدراسة الحالية مع الدراسات السابقة في تناولها موضوع جودة أوراق الاختبارات التحصيلية على المستوى الجامعي، وفي استخدامها للمنهج الوصفي التحليلي، لكنها تختلف عنها في مكان وزمان إجراءاتها، وشمول عينتها لاختبارات الأقسام العلمية والأدبية، ولم تقتصر على اختبارات المقررات التربوية، أو على مستوى معين من المستويات الدراسية، كما اختلفت عنها في تصنيفها لمعايير الجودة، حيث تتميز الدراسة الحالية عنها في كونها تتناول معايير ورقة الاختبار كاملة، وتصنيفها ضمن ثلاثة محاور رئيسة هي (التحليل العام للورقة الاختبارية، وله بعدين فرعيين (الشكل، والمضمون)، ومحور معايير صياغة فقرات أنواع الأسئلة الاختبارية المقالية والموضوعية، وله خمسة أبعاد فرعة، إضافة إلى محور الطباعة والإخراج، وذلك للحصول على إجابات حول مستوى استيفاء معايير كل محور من هذه المحاور، وعلى مستوى كل بعد من الأبعاد الفرعية لها، وأثره في نجاح أو فشل العملية الاختبارية والتي تتم داخل الجامعة.

الفصل الثالث.

منهجية الدراسة وإجراءاتها:

أولاً: منهج الدراسة:

اعتمدت هذه الدراسة على المنهج الوصفي، القائم على الأسلوب التحليلي، لملائمته لأهداف الدراسة وتساؤلاتها، فهذا المنهج يهتم بجمع المعلومات والبيانات ذات الصلة بالمشكلة كما هي في الواقع، والتعبير عنها كمياً وكيفياً، بهدف وصفها وصفاً دقيقاً (ملحم: ٢٠١٠، ص ٣٠٧)، و (الحاج: ٢٠١٣، ص ٧٥)، حيث سيتم بواسطته تحليل عينة من الأوراق الاختبارية في ضوء معايير جودة الورقة الاختبارية للوصول إلى وصف الظاهرة المدروسة من حيث طبيعتها، ودرجة وحجم وجودها، والوصول إلى استنتاجات تسهم في تحسين الواقع وتطويره.

ثانياً: مجتمع الدراسة وعينتها:

تحدد مجتمع الدراسة في جميع أوراق الاختبارات النهائية التي أعدها أعضاء الهيئة التدريسية بكلية التربية والعلوم التطبيقية بأرحب - جامعة صنعاء، لطلبة الكلية بالأقسام العلمية (كيمياء، فيزياء، رياضيات)، والأقسام الأدبية (إسلامية، قرآن، لغة عربية، تاريخ، جغرافيا)، خلال الفصل الدراسي الثاني من العام الجامعي (٢٠٢١ / ٢٠٢٢)، والبالغ عددها (٧٥) اختباراً تحصيلياً من اختبارات المقررات النظرية في مختلف الأقسام، والتي أتاحت للباحثين وتمكنا من الحصول عليها.

وقد تم تطبيق هذه الدراسة على عينة من أوراق الاختبارات، وعددها (٤٧) ورقة اختبارية من أوراق اختبارات نهاية الفصل الدراسي الثاني لمختلف الأقسام، وبذلك مثلت عينة الدراسة نسبة (٦٢%) من أفراد المجتمع الأصلي (الاختبارات)، تم اختيارها بطريقة عشوائية بسيطة، روعي عند اختيارها تمثيل المواد التربوية وغير التربوية، والمستويات الدراسية لكل تخصص، ويبين الجدول توزيع العينة بحسب المتغيرات.

جدول (١)

يبين توزيع أفراد عينة الدراسة (الاختبارات) حسب النوع والتخصص والجنس

نوع المواد			التخصص										الجنس	
تربوية	غير تربوية	المجموع	علوم قرآن	إسلامية	لغة عربية	تاريخ	جغرافيا	رياضيات	فيزياء	كيمياء	علوم حياة	المجموع	نساء	ذكور
20	27	47	6	6	6	6	6	6	6	5	47	19	28	47

أداة الدراسة:

اعتمدت الدراسة الحالية في جمع البيانات والمعلومات مقياس تقدير جودة الاختبار التحصيلي، نظراً لمناسبته لمنهجية وأهداف هذا الدراسة وأسئلتها، والذي تم بنائه بعد الاطلاع على الأدبيات التربوية والدراسات السابقة ذات الصلة بمشكلة الدراسة، وفي ضوء أهداف ومعطيات الدراسة الحالية وتساؤلاتها، وللوصول بمقياس تقدير جودة الاختبار التحصيلي لشكله النهائي، والتأكد من عامل الصدق والثبات تم اتباع الخطوات الإجرائية الآتية:

١- صدق المحكمين:

قام الباحثان بإعداد وتصميم مقياس تقدير جودة الاختبار التحصيلي، اشتمل في صورته الأولية على ثلاثة محاور هي: معايير الإطار العام لورقة الاختبار من حيث (الشكل والمضمون)، ومعايير صياغة الأسئلة المقالية والموضوعية- معايير الطباعة والإخراج)، وسبعة أبعاد فرعية، أدرج منها بعدين فرعيين تحت تحليل الإطار العام للورقة الاختبارية هما من حيث (الشكل - المضمون)، وخمسة أبعاد فرعية تتدرج تحت محور معايير صياغة أنواع فقرات الأسئلة الاختبارية (المقالية، صح وخطأ، التكميل، المزاوجة، اختيار من متعدد)، وقد شمل المقياس في صورته الأولية على (٨٥) معياراً، ومن ثم تم عرضه على مجموعة من الأساتذة المتخصصين في المناهج والقياس والتقويم، لإبداء الرأي والحكم على مناسبة الفقرات للمحاور وكذلك انتمائها لها، وإبداء الرأي حول مناسبتها لما وضع لقياسه، وفي ضوء آراء المحكمين تم حذف ٦ فقرات، ودمج ١٠ فقرات داخل المحاور الثلاثة، وتعديل صياغة عدد منها، ونتيجة لذلك أصبح المقياس مشتملاً على (٦٤) معياراً، ومن ثم عرض المقياس على المحكمين مرة أخرى، أظهرت استجابتهم الاتفاق على مناسبة الأداة للقياس، وبذلك أطمئن الباحثان من توافر وتحقيق الصدق الظاهري لمحتوى المقياس، واستقر المقياس في صورته النهائية على ٦٤ معياراً، موزعة على المحاور الرئيسية والفرعية للمقياس، ويقابل كل معيار منها مقياس رباعي متدرج للـ (يكرت)، متوافر بدرجة (عالية، متوسطة، ضعيفة، منعدمة)، وتكون المقياس في صورته النهائية مما يأتي:

القسم الأول: يحتوي على البيانات الأولية، سواء الخاصة بمعد الاختبار (الاسم، الجنس، المؤهل، الدرجة، نوع وأنماط الأسئلة)، أم المتعلقة بالاختبار (القسم، المادة، المستوى الدراسي...)

القسم الثاني: ويحتوي على المحاور الرئيسية والفرعية المكونة للمقياس، والذي شمل في صورته النهائية (٦٤) معياراً، موزعة على تلك المحاور الثلاثة والأبعاد الفرعية وفقاً للجدول التالي:

الجدول (٢)

يبين المحاور الرئيسية والفرعية المكونة للمقياس

العدد	الأبعاد الفرعية	المحاور الرئيسية	
8	من حيث الشكل	معايير الإطار العام للورقة الاختبارية	المحور الأول
9	من حيث المضمون		
4	أسئلة المقال	معايير الصياغة لأنواع الأسئلة الاختبارية	المحور الثاني
6	أسئلة الصح والخطأ		
6	أسئلة المطابقة		
4	أسئلة الإكمال		
11	أسئلة الاختيار من متعدد		
16	_____	معايير الطباعة والإخراج	المحور الثالث
64	المجموع الكلي للمعايير		

١- ثبات المقياس:

للتأكد من ثبات التحليل باستخدام أداة مقياس تقدير جودة الاختبار، تم حساب نوعين من الثبات هما الثبات عبر الزمن والأشخاص، حيث تم تحليل (٥) أوراق اختبارية من خارج عينة الدراسة حيث قام الباحثان كل منهما على حدة بتحليل نفس عينة الاختبار نفسها ، كما قام أحد الباحثين بإعادة التحليل مرة أخرى للعينة ذاتها بعد مدة زمنية مناسبة، ولحساب معامل الثبات، تم حساب معامل الاتفاق بين تحليل الباحثان، وكذا التحليل الأول والثاني للباحث عبر الزمن وفقاً للمعادلة، $2N + 1 / 2N + 1 =$

جدول (٣)

يبين معامل ثبات (ألفا كرونباخ)

الاتفاق بين تحليل	عدد مرات الاتفاق	معامل ألفا كرونباخ
الباحث الأول والثاني	56	0.88
التحليل الأول والثاني للباحث الثاني	60	0.94
المتوسط المرجح	0.91	0.84

فكان معامل الثبات ٠.٩١%، وهو معدل ثابت وعالٍ ويسمح بإجراء الدراسة.

ثالثاً: تطبيق أداة الدراسة:

بعد التأكد من صدق وثبات الأداة وصلاحيه استخدامها؛ قام الباحثان بتطبيق مقياس تقدير جودة الورقة الاختبارية على عينة الاختبارات التحصيلية للدراسة، ومن ثم جمع وتفريغ البيانات على الحاسوب، وإجراء العمليات الإحصائية المناسبة؛ باستخدام برنامج (SPSS)، وفقاً للإجراءات الآتية:

- تم تفريغ البيانات وتحويلها إلى درجات، حيث تم إعطاء ٣ درجات عند توفر المعيار بدرجة عالية، و ٢ درجة إذا توافر بدرجة متوسطة، ودرجة واحدة إذا توافر بدرجة ضعيفة، و (٠) إذا لم يتوفر، كما تم حذف درجة أي محور إذا ارتبط بنوع معين من الأسئلة ولم يتوفر في ورقة الاختبار، فإذا اعتمد الاختبار على الأسئلة الموضوعية دون الأسئلة المقالية تحذف درجة المعيار الخاص بالأسئلة المقالية وهكذا...

- تم حساب المتوسطات الحسابية، والانحرافات المعيارية، والنسب المئوية لكل فقرة من فقرات مقياس جودة الاختبارات، البالغ عددها (٦٤) فقرة، وكذلك الدرجة الكلية على مستوى كل محور من المحاور الرئيسة الثلاثة والأبعاد الفرعية المكونة لها، وعلى مستوى محاور المقياس ككل، وتم تحديد الرتبة والتقدير اللفظي لكل فقرة من فقرات كل محور أو بعد اعتماداً على المتوسطات الحسابية، وعلى الانحرافات المعيارية عند تساوي المتوسطات، ومن ثم تم محاكمة وتفسير النتائج في ضوء المحك المعتمد لهذه الدراسة، وبين الجدول (٤) تقديرات معايير تفسير لنتائج.

متوسط الفقرة	3 – 2.34	2.33-1.67	1.66- 1	أقل من ١
الحكم اللفظي	عالية	متوسطة	ضعيفة	منعدم

رابعاً: الأساليب الإحصائية:

استخدمت الدراسة الأساليب الإحصائية الآتية:

- ١- التكرارات والنسب المئوية.
- ٢- المتوسطات الحسابية والانحرافات المعيارية.
- ٣- معادلة (ألفا كرنباخ) لحساب معامل الثبات.

خامساً: عرض نتائج الدراسة ومناقشتها:

للإجابة عن السؤال الأول:

ونصه "ما درجة التزام معدي الاختبارات بمواصفات الورقة الاختبارية من حيث (الشكل والمضمون)؟" تم حساب المتوسطات الحسابية، والانحرافات المعيارية، والنسب المئوية لفقرات محور الإطار العام للورقة الاختبارية، وعددها (١٧) فقرة، والمصنفة ضمن بعدين فرعيين، وكذلك الدرجة الكلية لكل بعد منها، وللمحور ككل، وكما تبينه النتائج المعروضة في الجدول الآتي:

جدول (٥)

يبين المتوسطات الحسابية والانحرافات المعيارية للالتزام بمعايير الإطار العام للورقة الاختبارية ببعد

الشكل والمضمون

المحاور الأول	الأبعاد الفرعية	المتوسط الحسابي	الانحراف المعياري	%	النتيجة
معايير جودة الإطار العام للورقة الاختبارية	من حيث الشكل	47	1.02	474.	34% ضعيف
	من حيث المضمون	47	1.37	263.	46% ضعيف
المتوسط الكلي لمحور الاطار العام ككل		47	1.20	335.	40% ضعيف

يتبين من الجدول (٥) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة الإطار العام للورقة الاختبارية على مستوى المحور ككل بلغ (١.٢٠)، وانحراف معياري (335)، وبدرجة تقدير (ضعيف)، وقد حصل بعد التزامهم بمعايير جودة ورقة الاختبار من حيث المضمون على الترتيب الأول، بمتوسط حسابي (١.٣٧)، وانحراف معياري (٠.٢٦٣)، بينما حصل بعد التزامهم بالمعايير الشكلية على الترتيب الثاني، بمتوسط حسابي (١.٠٢)، وانحراف معياري (٠.٤٧٤)، وبشكل عام تؤكد هذه النتائج أن التزامهم بالمعايير الشكلية لورقة الاختبار كان أكثر تدنياً وضعفاً من معايير المضمون، إذ لم تتجاوز استيفاء المعايير الشكلية نسبة (٣٤%).

وفيما يلي عرضاً لنتائج كل فقرة من الفقرات على مستوى كل بعد من الأبعاد لهذا المحور، وعلى النحو الآتي:

١-المعايير الشكلية للورقة الاختبارية:

جدول (٦)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة للمعايير الفنية للورقة الاختبارية من حيث الشكل مرتبة

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
1	(الشعار، الجامعة، الكلية، القسم) في الجانب الأيمن	ت	8	20	15	4	1.68	0.86	56%
		%	17.0	42.6	31.9	8.5			
2	بيانات (المادة، التخصص، المستوى الدراسي) في الجانب الأيسر	ت	11	10	23	3	1.62	0.92	54%
		%	23.4	21.3	48.9	6.4			
3	بيانات الزمن: (يوم وتاريخ الاختبار، الفصل، العام الدراسي)	ت	3	7	34	3	1.21	0.66	40%
		%	6.4	14.9	72.3	6.4			
4	بيانات الاختبار: العدد الكلي لـ (الأسئلة، الدرجة، الصفحات).	ت	0	6	14	27	0.55	0.72	18%
		%	0.0	12.8	29.8	57.4			
5	بيانات الطالب: (الاسم، التخصص، الرقم الجامعي..).	ت	4	12	10	21	0.98	1.03	33%
		%	8.5	25.5	21.3	44.7			
6	تحدد التعليمات الدرجة أمام كل سؤال (رئيسي/ فرعي) حسب أهميته.	ت	3	4	11	29	0.60	0.90	20%
		%	6.4	8.5	23.4	61.7			
7	توضح التعليمات (مكان- وطريقة) الإجابة لكل سؤال بدقة.	ت	0	8	20	19	0.77	0.73	26%
		%	0.0	17.0	42.6	40.4			
8	تم استخدام الغلاف الموحد لورقة الاختبار لتوحيد البيانات.	ت	3	5	17	22	0.77	0.89	26%
		%	6.4	10.6	36.2	46.8			
	المتوسط العام للبعد الشكلي					1.02	0.47	34%	ضعيف

تبين نتائج الجدول (٦) أن المتوسط العام لدرجة التزام معدي الاختبارات بالمواصفات الشكلية للورقة الاختبارية على مستوى البعد ككل بلغ (1.02)، وبانحراف معياري (0.47)، وهذا يعني أن الالتزام بالمواصفات الشكلية للورقة الاختبارية من قبل معديها جاء بتقدير (ضعيف)؛ وفقاً لمعيار التصحيح، وقد تراوحت قيم المتوسطات على جميع الفقرات هذا البعد بين (1.68 في حدها الأعلى، وكانت لفقرة تضمن ورقة الاختبار لـ (الشعار، الجامعة، الكلية، القسم) في الجانب الأيمن"، وبين (٠.٥٥) في حدها الأدنى، وكانت لفقرة " بيانات الاختبار (عدد الأسئلة، الدرجة الكلية، عدد الصفحات) للاختبار"، مما يدل على أن التزام الأساتذة بالمواصفات الشكلية قد انحصر بين التقدير (المتوسط، والضعيف، والمنعدم)، ولم تظهر النتائج حصول أي عبارة على درجة تقدير (عالية) على مستوى المحور ككل.

أما ترتيب فقرات هذا البعد فجاءت حسب متوسطها الحسابي؛ ووفقاً لمعيار التصحيح كما يأتي:

- تحقق معيار واحد فقط من المعايير الشكلية بدرجة (متوسطة)، مثلته الفقرة (١)، بمتوسط (1.68)، واستحقت الرتبة الأولى، لكن قيمة هذا المتوسط في أدنى درجاته، لأنه يقترب كثيراً من الفئة الضعيفة في حدها الأعلى، والمحددة بـ (1.66)، وقد يرجع سبب التدني في توافره لى عدم قيام الكلية بإعداد نموذج غلاف موحد للاختبار، وغياب المتابعة للالتزام معديها بالنموذج الموحد.
- تحقق ٢ من المعايير بدرجة (ضعيفة)، مثلتهما الفقرتان (٢، ٣)، وجاءا في المرتبتين الثانية والثالثة، ومثلت المعايير المتعلقة بـ (المادة والمستوى والتخصص، ويوم وزمن وتاريخ الاختبار).
- تحققت ٥ معايير بدرجة (منعدمة)، ومثلتها الفقرات ذوات الأرقام (٤، ٥، ٦، ٧، ٨)، وقد تراوح المتوسط الحسابي لها في حده الأعلى والأدنى بين (٠.٩٨ - ٠.٥٥)، وهي قيمة تقع ضمن الفئة المنعدمة؛ لأنها أقل من الواحد الصحيح وفقاً لمعيار التصحيح، ونالت الخمس الرتب الأخيرة من الرابعة وحتى الثامنة من ترتيب الفقرات لهذا المحور.

يتضح من خلال ما أظهرته النتائج المعروضة سابقاً لهذا البعد من أبعاد المحور الأول؛ أن درجة التزام معدي الاختبارات بالمواصفات الشكلية للورقة قد انحصرت بين التقدير (المتوسط والضعيف والمنعدم)، إذ لم تُظهر النتائج أي درجة توافر (عالية) على أي فقرة من الفقرات المكونة لهذا البعد، حيث أن قيمة المتوسط في الحد الأدنى للمستوى المرتفع وفقاً للمعيار المعتمد في الدراسة هو (٢.٣٤)، فأكثر، في حين أن أعلى متوسط أظهرته النتائج لهذه الفقرات كان (١.٦٨) وعلى فقرة واحدة فقط، وهي قيمة يقابلها وفقاً لمعيار التفسير مستوى توافر بدرجة (متوسطة)، أما أدنى متوسط لهذه الفقرات فكان (٠.٥٥)، وهي درجة توافر منعدمة، وتعني هذه النتيجة إجمالاً بأن هناك ضعفاً وتدنياً واضحاً في مستوى التزام معدي الاختبارات بالمعايير الشكلية اللازم مراعاتها عند تصميم الاختبارات، أو حتى العمل على توافر الحد الأدنى من تلك

- المعايير على أقل تقدير؛ بل أن هناك انعدام شبه كلي لبعض الجوانب الشكلية التي على درجة بالغة من الأهمية، منها بيان (المستوى الدراسي والتخصص) الموجه لهم الاختبار، الأمر الذي يحتاج معه الأمر إلى لفظة جادة، وخطوات عملية من القائمين؛ ويعزو الباحثان سبب ذلك إلى واحد أو أكثر من الأسباب الآتية.
- غياب وضعف المتابعة المستمرة من الكلية، أو إلزام الهيئة التدريسية باستخدام نموذج موحد ومعتمد سبق إعداده، لضمان توحيد البيانات اللازم توافرها بحددها الأدنى.
 - غياب الخطط التقويم المزمدة المحددة مسبقاً، وإعلان مواعيدها بصورة مفاجئة؛ مما أنعكس سلباً على مستوى اهتمام معدي تلك الاختبارات بالمواصفات الشكلية لضيق الوقت أمامهم.
 - كثرة أعباء ومهام ومسؤوليات أعضاء الهيئة التدريسية.
- وتتفق هذه النتيجة مع نتائج دراسة كل من ساعد، ووسيلة (٢٠١٣) التي توصلت إلى أن معايير كتابة تعليمات الاختبار جاءت بنسب ضعيفة، بينما اختلفت مع نتائج دراسة كل من الحسينان، وآخرون (٢٠١٤)، والسفياني (٢٠١٩)، فقد توصلت الأولى في نتائجها عن تجاوز نسب استيفاء المعايير البيانات الأساسية وتعليمات الاختبار ٩٠%، وما كشفته الأخرى من توافر المواصفات الشكلية بدرجة عالية.
- ب- معايير جودة الورقة الاختبارية من حيث المضمون:

جدول (٧)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير لورقة الاختبارية من حيث المضمون

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
9	يغطي الاختبار موضوعات مختلفة للمقرر.	0	8	39	0	1.17	0.38	39%	منعدمة
		0.0	17.0	83.0	0.0				
10	تقيس الأسئلة جوانب التعلم الثلاثة قدر الإمكان.	0	4	43	0	1.09	0.28	36%	ضعيف
		0.0	8.5	91.5	0.0				
11	تقيس الأسئلة مستويات تفكير عقلية (عليا ودنيا).	0	4	43	0	0.83	0.56	28%	منعدم
		0.0	8.5	91.5	0.0				
12	يقيس كل سؤال ناتج تعليمي واحد (غير مركب).	7	19	21	0	1.70	0.72	57%	متوسط
		14.9	40.4	44.7	0.0				
13	يحدد السؤال المهمة المطلوبة بدقة، بعيد عن التعقيد اللفظي.	10	16	21	0	1.77	0.79	59%	متوسط
		21.3	34.0	44.7	0.0				
14	تتنوع أسئلة الاختبار ما بين (المقالية، الموضوعية).	0	4	31	12	0.83	0.56	28%	منعدم
		0.0	8.5	66.0	25.5				
15	ترتيب الأسئلة حسب (مستوى الصعوبة، نوع السؤال).	3	15	29	0	1.45	0.62	48%	ضعيف
		6.4	31.9	61.7	0.0				

16	يتناسب عدد الأسئلة مع زمن الاختبار.	ت	8	33	6	0	2.04	0.55	68%	عالي
		%	17.0	70.2	12.8	0.0				
17	صحة مجموع الدرجات لورقة الاختبار.	ت	13	8	2	24	1.21	1.33	40%	ضعيف
		%	27.7	17.0	4.3	51.1				
المتوسط العام للمحور										
							1.37	0.26	46%	ضعيف

يتبين من نتائج الجدول (٧) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة الورقة الاختبارية من حيث المضمون بلغ (1.37)، وبانحراف معياري (0.26)، وهذا يعني أن درجة التزامهم بتلك المعايير جاء بدرجة (ضعيفة) بشكل عام، وقد تراوحت قيم المتوسطات على جميع فقرات هذا البعد بين ((2.04 في حدها الأعلى، وكانت لفقرة " يتناسب عدد الأسئلة مع زمن الاختبار"، وبين (0.83) في حدها الأدنى، وكانت لفقرة " تقيس الأسئلة مستويات تفكير عقلية (عليا ودنيا)"، وفقرة "تتنوع أسئلة الاختبار ما بين (المقالية، الموضوعية) وفق طبيعة المادة"، وهذا يدل على أن التزام معدي الاختبارات بمعايير مضمون ورقة الاختبار قد انحصر بين التقدير (العالي، والمنعدم) على جميع الفقرات المندرجة تحت هذا البعد ككل،

أما ترتيب فقرات هذا البعد فقد جاءت بحسب متوسطها الحسابي؛ ووفقاً لمعيار التصحيح كما يأتي:

- تحقق معيار واحدة فقط من معايير بعد المضمون بدرجة (عالية)، مثلته الفقرة (١٦)، بمتوسط حسابي (٢.٠٤)، واستحقت الرتبة الأولى، تمثلت في النص " يناسب عدد الأسئلة مع زمن الاختبار"
- تحقق ٢ معايير بدرجة (متوسطة)، مثلتهما الفقرتان (١٣، ١٢)، وبمتوسط حسابي (١.٧٧، ١.٧٠) على الترتيب، واستحقا الرتبتين الأولى والثانية، تعلق نصهما بمعياريين من المعايير العامة للأسئلة من حيث (عدم التركيب، وضوح الصياغة بعيداً عن التعقيد اللفظي).
- تحقق ٤ معايير بدرجة (ضعيفة)، مثلتها الفقرات (١٥، ١٧، ٩، ١٠)، وأنحصر متوسطاتها الحسابية بين (١.٤٥ - ١.٠٩)، ونالت الرتب من الرابعة وحتى السابعة، ونصها (ترتيب الأسئلة حسب مستوى الصعوبة، صحة مجموع الدرجات، شمول الأسئلة لموضوعات المقرر وجوانب التعلم).
- جاء ٢ من المعايير بدرجة (منعدمة)، مثلتهما الفقرتان (١١، ١٤)، بمتوسط حسابي (٠.٨٣) لكلا الفقرتين، إضافة إلى تساويهما في الانحراف المعياري؛ لذلك استحققتا المرتبة الثامنة والأخيرة، ونصتا على (تتنوع الأسئلة بين المقالية الموضوعية، وتقيس الأسئلة مستويات تفكير عليا ودنيا)، ما يعني اقتصرهم على أنماط محددة من الأسئلة وبشكل خاص على الأسئلة المقالية دون غيرها، أو نمط معين من الأسئلة الموضوعية، ويعزى السبب في ذلك إلى سهولة إعدادها، مقارنة مع ما تطلبه بقية الأنماط من وقت وجهد، ومما يؤكد هذا التفسير تركيز الأسئلة على قياس المستويات الدنيا من التفكير، وبالمقابل غياب الأسئلة التي تتطلب القيام بعمليات التفكير العليا (التحليل، التركيب، التقويم)، وهذا الأمر يؤكد

انصراف اهتمام معدي الاختبارات على قياس الحفظ والاستظهار، على الرغم من أن نمط الأسئلة المقالية هو أكثر أنواع الأسئلة استخداماً من قبل الأساتذة، كان ينبغي أن يقتصر استخدامها على قياس المستويات العليا، إلا أنها استخدمت لقياس مستويات الدنيا من التفكير.

وتشير ما بينته النتائج المعروضة سابقاً إجمالاً من نسب ومتوسطات لدرجة التزام معدي الاختبارات بمعايير المضمون على درجة كبيرة من الضعف والقصور، وأن درجة التزامهم بتلك المعايير لا يرقى إلى المستوى المطلوب، إذ أقتصرت التحقق والتوافر بدرجة عالية على معيار واحد فقط، بينما كان التزامهم على معيارين بدرجة متوسط، في حين كان التزامهم بتحقيق معظم معايير جودة الورقة الاختبارية في بعد المضمون وعددها ٦ معايير ضمن الفئتين (الضعيفة والمنعدمة)، كما يتبين أن تدني وضعف النسب كان للمعايير المتضمنة مضامينها مهام يغلب عليها الطابع العملي أو التطبيقي، منها على سبيل التمثيل المعايير المتعلقة بشمول الأسئلة لجوانب التعلم الثلاثة، وشمولها لمستويات التفكير العليا والدنيا لكل مجال، إضافة إلى قصور شمولها لتغطي موضوعات ومفردات المقرر المختلفة.

ويرجع الباحثان السبب في ذلك إلى أن إعداد وبناء الاختبارات عملية خاضعة للاجتهاد الشخصي، وأن واضعوها لا يلتزمون بالخطوات العلمية في بنائها ووفقاً لجدول المواصفات، وهذا الأمر جعل تلك الاختبارات تفتقر إلى أهم معايير الاختبار الجيد، من أبرزها معيار شمول الأسئلة لجوانب المحتوى، ولمجالات الأهداف ومستوياتها المختلفة، وتبعاً للأهمية النسبية والجهد والوقت المبذول، ومن ثم افتقار تلك الاختبارات لسمة الصدق والموضوعية نتيجة لذلك، لإهمالها أجزاء كبيرة ومهمة من المحتوى، واقتصارها على مجال ومستوى دون آخر، ولا يتأتى أو يتحقق كل ذلك إلا باتباع الخطوات العلمية لوضع الاختبار، وبناءها وفق جدول المواصفات.

وتتفق هذه النتيجة مع ما خلصت إليه نتائج دراسة الحسينان، وآخرون (٢٠١٤) من تدني نسب استيفاء معيار مضمون الاختبار، إذ لم يتجاوز نسبة (٦٨%)، وما كشفتته دراسة السفيناني (٢٠١٩) من تدني نسب المعايير الضمنية أخيراً بنسبة (٥٩%)، ودراسة القرشي (٢٠١٠) من أن الاختبارات تفتقر إلى الصدق والشمول والموضوعية، وتركيزها على قياس التذكر مع إهمال المستويات العليا.

للإجابة عن السؤال الثاني: ونصه "ما مدى التزام معدي الاختبارات بمعايير كتابة أنواع الأسئلة الاختبارية (المقالية، والموضوعية)؟"، فقد تم حساب المتوسطات الحسابية، والانحرافات المعيارية، والنسب المئوية لكل فقرة من فقرات قياس معايير جودة صياغة كل نوع من أنواع الأسئلة (المقالية، والصواب والخطأ، والإكمال، والمطابقة، والاختيار من متعدد)؛ المصنفة ضمن ٥ أبعاد فرعية مكونة لهذا المحور، والبالغ عددها (٣١) معياراً، وكذلك الدرجة الكلية على مستوى كل بعد من الأبعاد الفرعية وعلى مستوى المحور ككل كما هو مبين في الجداول التالي.

جدول (٨)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة للالتزام بمعايير الصياغة لأنواع الأسئلة المقالية والموضوعية

المحور الثاني	الأبعاد الفرعية	العدد	المتوسط الحسابي	الانحراف المعياري	%	الحكم
معايير صياغة فقرات الأسئلة الاختبارية (مقالي، موضوعي)	الأسئلة المقالية	42	1.49	.544	50%	ضعيف
	أسئلة الصح والخطأ	25	1.59	.367	53%	ضعيف
	أسئلة المطابقة	3	1.33	.00000 ^a	44%	ضعيف
	أسئلة الإكمال	21	1.86	.322	62%	متوسط
	الاختبار من متعدد	12	1.46	.197	49%	ضعيف
اجمالي المتوسط للمحور		47	1.52	.386	51%	ضعيف

يتبين من الجدول (٨) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة الصياغة لأنواع الأسئلة الاختبارية على مستوى هذا المحور بأبعاده الفرعية الخمسة ككل بلغ (١.٥٢)، وانحراف معياري (٠.٣٨٦)، وبدرجة تقدير (ضعيفة)، وقد جاء التزامهم بمعايير جودة صياغة أسئلة الإكمال في المرتبة الأولى، بمتوسط حسابي (١.٨٦)، وانحراف معياري (٣٢٢.٠)، وبدرجة تقدير (متوسطة)، يليه بعد معايير صياغة الصواب والخطأ في المرتبة الثانية، بمتوسط حسابي (١.٥٩)، وانحراف معياري (٠.٣٦٧) وبدرجة (ضعيفة)، وحصل بعد معايير جودة صياغة أسئلة المقال على الرتبة الثالثة، بمتوسط حسابي (١.٤٩) وانحراف معياري (٥٤٤.٠) وبدرجة (ضعيفة)، بينما جاء بعد معايير جودة صياغة أسئلة الاختيار من متعدد في المرتبة الرابعة، بمتوسط حسابي (١.٤٦)، وانحراف معياري (٠.١٩٧)، وبدرجة (ضعيفة)، في حين حصل بعد معايير صياغة أسئلة المطابقة المرتبة الخامسة والاختيرة، بمتوسط (١.٣٣) وانحراف معياري (٠.٠٠٠)، وبدرجة (ضعيفة).

كما يتضح من خلال ما أظهرته النتائج المعروضة سابقاً لهذا المحور بأبعاده الفرعية ككل، أن أكثر أنواع الأسئلة الشائعة استخداماً من قبل الأساتذة هي الأسئلة المقالية، يليها أسئلة الصح والخطأ، ثم الإكمال، ثم الاختيار من متعدد، وأخير المطابقة، وفيما يلي عرضاً لنتائج كل فقرة من الفقرات على مستوى كل بعد من الأبعاد الخمسة السابقة المندرجة والمكونة لهذا المحور وعلى النحو الآتي:

جدول (٩)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير جودة صياغة الأسئلة المقالية

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
18	يطرح السؤال مشكلة واضحة ومحددة تعطي معناً واحد لدى كل الطلبة.	ت	9	17	18	1.80	0.76	60%	متوسط
		%	20.5	38.6	40.9				
19	يقتصر استخدامها على قياس مستويات التفكير العليا (تحليل، تركيب، تقويم).	ت	0	5	20	0.68	0.67	23%	ضعيف
		%	0.0	11.4	45.5				
20	تستخدم أسئلة قصيرة يدور كل منها حول فكرة .	ت	12	22	8	2.00	0.81	67%	متوسط
		%	27.3	50.0	18.2				
21	تحديد كم الإجابة المطلوبة في حالة الأسئلة المقالية (الإنشائية).	ت	6	13	22	1.50	0.82	50%	ضعيف
		%	13.6	29.5	50.0				
المتوسط الكلي للبعد									
						1.48	0.54	49%	ضعيف

تبين النتائج المعروضة في الجدول (٩) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة صياغة الأسئلة المقالية بلغ (١.٤٨)، وبانحراف معياري (٠.٥٤)، وهي قيمة متوسط يقابلها درجة تقدير (ضعيف)، وقد تراوحت قيم المتوسطات على جميع فقرات هذا البعد بين (٢.٠٠) في حدها الأعلى، وكانت لفقرة تضمن ورقة الاختبار لـ " أسئلة قصيرة يدور كل منها حول فكرة، وبين (٠.٦٨) في حدها الأدنى، وكانت لفقرة " يقتصر استخدامها على قياس المستويات العليا من التفكير (تحليل، تركيب، تقويم)، ما يدل على أن درجة التزام معدي الاختبارات بمعايير صياغة الأسئلة المقالية قد انحصر بين التقدير (المتوسط، والمنعدم) على مستوى هذا البعد ككل.

أما ترتيب الفقرات فقد جاءت حسب المتوسط الحسابي، ووفقاً لمعيار التصحيح على النحو الآتي:

- تحقق ٢ من معايير جودة صياغة أسئلة المقال بدرجة (متوسطة)، مثلثتها الفقرتان (٢٠، ١٨)، كان المتوسط الحسابي لهما (٢.٠٠، ١.٨٠) على الترتيب، واستحقا المرتبتين الأولى والثانية، ونصتا على أن الاختبارات تستخدم (أسئلة مقالية قصيرة يدور كل منها حول فكرة واحدة، وأن الأسئلة تطرح مشكلة واضحة ومحددة تعطي معناً واحداً لدى كل الطلبة)، مما يدل على حرص معدي الاختبارات في الابتعاد عن الأسئلة الإنشائية المفتوحة لتقليل تدخل أثر عامل الذاتية وتحقيق قدر من الموضوعية.

- جاء معيار واحد فقط بدرجة (ضعيفة)، مثلته الفقرة (٢١)، بمتوسط (١.٥٠)، جاء في المرتبة الثالثة، مما يشير إلى أن معظم الأساتذة لا يحددون كم الإجابة المطلوبة حالة استخدامهم الأسئلة المقالية.
- جاء معيار واحد بدرجة (منعدمة)، مثلته الفقرة (١٩)، بمتوسط حسابي (٠.٦٨)، لذلك استحققت المرتبة الأخيرة، تمثل بالنص "يقتصر استخدامها على قياس المستويات العليا من التفكير (تحليل، تركيب، تقويم)، مما يدل على أن معدي الاختبارات يستخدمون الأسئلة المقالية في قياس المستويات الدنيا من التفكير كالحفظ والاستظهار أو الفهم على أحسن تقدير.

جدول (١٠)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير جودة صياغة الأسئلة الصح والخطأ

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
22	أن تكون العبارة قصيرة نسبياً تقيس فكرة أساسية وجوهرية	ت	8	17	2	0	2.22	0.58	74%
		%	29.6	63.0	7.4	0.0			
23	صياغة العبارة بإحكام بحيث تكون صحيحة تماماً أو خطأ.	ت	2	15	10	0	1.70	0.61	57%
		%	7.4	55.6	37.0	0.0			
24	توزيع العبارات الصحيحة والخطأ بشكل عشوائي.	ت	0	1	26	0	1.04	0.19	35%
		%	0.0	3.7	96.3	0.0			
25	تساوي الفقرات الصحيحة والخطأ في الطول قدر الإمكان.	ت	0	6	21	0	1.22	0.42	41%
		%	0.0	22.2	77.8	0.0			
26	تجنب عبارات التعميم والتخصيص الموحية بالإجابة (دائماً، غالباً، فقط..).	ت	11	6	10	0	2.04	0.90	68%
		%	40.7	22.2	37.0	0.0			
27	عدد العبارات الصحيحة غير مساوٍ للعبارات الخطأ	ت	0	9	17	1	1.30	0.54	43%
		%	0.0	33.3	63.0	3.7			
المتوسط الكلي									
						1.59	0.37	53%	ضعيف

تبين نتائج المعروضة في الجدول (١٠) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة صياغة أسئلة الصواب والخطأ بلغ (١.٥٩)، وانحراف معياري (٠.٣٧)، وهي قيمة متوسط يقابلها درجة تقدير (ضعيفة)؛ ويقترّب كثيراً من تقدير الفئة المتوسطة، وقد تراوحت قيم متوسطات جميع فقرات هذا البعد بين (٢.٢٢) في حدها الأعلى، وكانت لفقرة " العبارة قصيرة نسبياً تقيس فكرة أساسية وجوهرية "، وبين

(١٠٤) في حدها الأدنى، وكانت لفقرة " توزيع العبارات الصحيحة والخطأ بشكل عشوائي"، ما يدل على أن درجة التزامهم بمعايير صياغة هذا الأسئلة قد انحصر بين التقدير (المتوسط، والضعيف).

أما ترتيب فقرات هذا البعد فقد جاءت حسب المتوسط الحسابي، ووفقاً لمعيار التصحيح كما يأتي:

- تحقق ٣ من معايير صياغة أسئلة الصواب والخطأ بدرجة (متوسطة)، تمثلت في الفقرات ذوات الأرقام (٢٢، ٢٣، ٢٦)، نالت المتوسطات الحسابية (٢.٢٢، ١.٠٤، ١.٧٠) على الترتيب، وحصلت على الثلاث المراتب الأولى، تمثلت بالنص على أن صياغة هذا النوع من الفقرات (قصير تقيس فكرة جوهرية، وتتجنب عبارات " دائماً، غالباً، فقط،.." الموحية بالإجابة، ومصاغة بإحكام بحيث تكون صحيحة تماماً أو خطأ تماماً)، وقد يرجع الارتفاع النسبي لدرجة التزامهم بمعايير صياغة هذه الفقرات إلى الخبرة التي أكتسبها معديها طوال فترة عملهم الأكاديمي، وقد يعود إلى سهولة صياغة هذا النوع من الفقرات، وبساطة الالتزام بتلك المعايير لأنها لا تطلب الكثير من الجهد.
- تحقق ٣ من معايير الصياغة لهذا النوع من الأسئلة بدرجة (ضعيفة)، تمثلت في الفقرات ذوات الأرقام (٢٧، ٢٥، ٢٤)، كان متوسطها الحسابي (١.٣٠، ١.٢٢، ١.٠٤) على الترتيب، وحصلت على الثلاث الرتب الأخيرة من الرابعة وحتى السادسة، تمثلت بالنص على معايير (عدم تساوي عبارات الصح والخطأ في العدد، وتساوي أطوال الفقرات قدر الإمكان، وتوزيع الفقرات بشكل عشوائي)، ويرجع انخفاض نسب درجة الالتزام على هذه المعايير إلى ما يتطلبه مراعاتها من معدي الامتحان من جهد ووقت أكثر في الصياغة والفرز والتنسيق، الأمر الذي يؤكد ما تم الذهاب إليه من تفسير سابق لأسباب ارتفاع التزامهم بمعايير الفقرات الثلاث الأولى؛ لأنها لا تتطلب ذلك الجهد.

جدول (١١)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير جودة صياغة أسئلة المطابقة

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
28	تجانس محتوى جميع العبارات في كلا القائمتين.	0	3	0	0	2.00	0.0	67%	متوسط
29	عبارات إحدى القائمتين أكثر من الأخرى، لإبعاد عنصر التخمين.	0	0	0	3	0.00	0.0	0%	منعدم
30	تقع جميع فقرات القائمتين (أ) و (ب) في صفحة واحدة ..	3	0	0	0	3.00	0.0	100%	عالي
31	يتراوح عدد البنود في كلا القائمتين بين (٥-٨).	0	0	3	0	1.00	0.0	33%	ضعيف
32	ترتب المقدمات والاستجابات في القائمتين ترتيباً منطقياً.	0	0	3	0	1.00	0.0	33%	ضعيف
33	المقدمات إلى اليمين تأخذ أرقاماً، والاستجابات في اليسار مرقمة حروفاً.	0	0	3	0	1.00	0.0	33%	ضعيف
المتوسط الكلي						1.33	0.0	44%	ضعيف

العدد ٤ - المجلد ٤٨ - كانون الأول لسنة ٢٠٢٣

مجلة أبحاث البصرة للعلوم الإنسانية

تبين النتائج المعروضة في الجدول (١١) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة صياغة أسئلة المطابقة ككل (١.٣٣)، وانحراف معياري (٠.٠٠)، وبدرجة تقدير (ضعيفة)، وقد تراوحت قيم المتوسطات على جميع فقرات هذا البعد بين (٣.٠٠) في حدها الأعلى، وكانت لفقرة " تقع جميع فقرات القائمتين (أ) و (ب) في صفحة واحدة "، وبين (٠.٠٠) في حدها الأدنى، وكانت لفقرة " عبارات إحدى القائمتين أكثر من الأخرى لإبعاد عنصر التخمين"، مما يدل على أن درجة التزام معدي الاختبارات بمعايير الصياغة لأسئلة المطابقة قد انحصر بين التقدير (العالي، والضعيف) على مستوى البعد ككل.

أما ترتيب فقرات هذا البعد فقد جاءت حسب المتوسط الحسابي، ووفقاً لمعيار التصحيح كما يأتي:

- كان التزامهم بمعيار واحد من معايير صياغة أسئلة المطابقة بدرجة (عالية)، مثلته الفقرة رقم (٣٠)، بمتوسط (٣.٠٠)، ونالت المرتبة الأولى، ونصها " تقع جميع فقرات القائمتين في صفحة واحدة.
- كان التزامهم بمعيار واحد بدرجة (متوسطة)، مثلته الفقرة رقم (٢٨)، بمتوسط حسابي (٢.٠٠)، وحصلت على الرتبة الثانية، تمثلت بالنص " تجانس محتوى جميع العبارات في كلا القائمتين".

- كان التزامهم ب ٣ من المعايير بدرجة (ضعيفة)، تمثلت في الفقرات ذوات الأرقام (٣٢، ٣١، ٣٣)، كان المتوسط الحسابي لها (٠.٠٠)، واستحقت جميعها المرتبة الثالثة، لتساويها في الانحراف المعياري، الأمر الذي يدل على تجانس البيانات وعدم اختلافها لدى جميع أفراد العينة في مستوى الالتزام بالمعايير التي نصها " يتراوح عدد البنود في كلا القائمتين بين (٥-٨)، وترتب المقدمات والاستجابات في القائمتين ترتيباً منطقياً، وتكتب المقدمات إلى اليمين وتأخذ أرقاماً، والاستجابات في اليسار مرقمة حروفاً".

- كان التزامهم بمعيار واحد بدرجة (منعدمة)، مثلته الفقرة رقم (٢٩)، بمتوسط حسابي (٠.٠٠)، واستحقت الرتبة الرابعة والأخيرة، و نصت على أن تكون "عبارات إحدى القائمتين أكثر من الأخرى لإبعاد عنصر التخمين"، ما يعني عدم التزام معدي نمط هذه الأسئلة بهذا المعيار بشكل كلي.

وتكشف النتائج المعروضة سابقاً عن قصور وتدني شديد في درجة الالتزام بمعايير الصياغة لهذا النوع من أنواع الأسئلة، إضافة إلى تدني عدد الاختبارات المستخدمة لهذا النمط من الأسئلة في الاختبارات، والذي قد يرجع سبب ذلك إلى قناعة الأساتذة بعدم مناسبة هذا النوع من الأسئلة لمستويات الطلبة المتقدم في المرحلة الجامعية، وأن استخدامها يتناسب مع صغار السن في التعليم العام فقط.

جدول (١٢)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير جودة صياغة أسئلة الإكمال

م	العبارة		عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
34	أن تكون الكلمة الناقصة في أسئلة الإكمال جوهريّة.	ت	9	12	1	0	2.36	0.58	79%	عالي
		%	40.9	54.5	4.5	0.0				
35	تصاغ الفقرات بشكل محدد لها إجابة واحدة لا يحتمل غيرها.	ت	3	2	15	2	1.27	0.83	42%	ضعيف
		%	13.6	9.1	68.2	9.1				
36	لا تتضمن العبارة أكثر من فراغين كحد أقصى.	ت	9	2	9	2	1.82	1.10	61%	متوسط
		%	40.9	9.1	40.9	9.1				
37	يقع الفراغ نهاية العبارة أو قريب منها وليس في أولها.	ت	7	7	7	1	1.91	0.92	64%	متوسط
		%	31.8	31.8	31.8	4.5				
المتوسط الكلي										
							1.86	0.32	62%	متوسط

يتضح من خال النتائج المعروضة في الجدول (١٢) أن المتوسط العام لهذا البعد المتعلق بمعايير صياغة أسئلة المطابقة بلغ (١.٨٦)، وبانحراف معياري (٠.٣٢)، وهي قيمة متوسط يقابلها درجة تقدير (متوسطة)، وقد تراوحت قيم المتوسطات على جميع الفقرات لهذا البعد بين (٢.٣٦) في حدها الأعلى،

وكانت لفقرة " أن تكون الكلمة الناقصة في أسئلة الإكمال جوهريّة"، وبين (١.٢٧) في حدها الأدنى، وكانت لفقرة " تصاغ الفقرات بشكل محدد لها إجابة واحدة لا يحتمل غيرها"، ويعني هذا أن درجة التزام بمعايير الصياغة عند كتابة نمط أسئلة الإكمال لدى معدي الاختبارات قد انحصر بين التقدير (العالي، والضعيف) على مستوى البعد ككل.

أما ترتيب فقرات هذا البعد فقد جاءت حسب المتوسط الحسابي، ووفقاً لمعيار التصحيح كالاتي:

-تحقق معيار واحد فقط من معايير أسئلة الإكمال بدرجة (عالية)، مثلته الفقرة (٣١)، بمتوسط (٢.٣٦)، ونالت المرتبة الأولى، وتمثلت بالنص " تصاغ الفقرات بشكل محدد لها إجابة واحدة لا يحتمل غيرها".

-تحقق ٢ من المعايير بدرجة (متوسطة)، مثلتهما الفقرتان (٣٦، ٣٧)، متوسطهما الحسابي على الترتيب

-(١.٩١، ١.٨٢)، ونالا الرتبة الثانية والثالثة، تمثلت بالنص " يقع الفراغ نهاية العبارة أو قريب منها

-وليس أولها، ولا تتضمن العبارة أكثر من فراغين كحد أقصى".

-جاء معيار واحد فقط بدرجة (ضعيفة)، مثلته الفقرة رقم (٣٥)، بمتوسط حسابي (١.٢٧)، وحصلت على

الرتبة الرابعة الأخيرة، ونصها "تصاغ الفقرات بشكل محدد لها إجابة واحدة لا يحتمل غيرها".

ويرجع الباحثان سبب الارتفاع النسبي لنسبة الالتزام بمعايير صياغة أسئلة الإكمال في هذا البعد مقارنة بغيره من أنواع الأسئلة في الأبعاد الخمسة إلى سهولة وبساطة كتابة هذا النوع من الفقرات، حيث لا تتطلب من المعدين امتلاك الكثير من المهارات عند بنائها مقارنة ببقية الأنماط الأخرى.

جدول (١٣)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير جودة صياغة أسئلة الاختيار من متعدد

م	العبارة	ن	م	متوسط	انحراف معياري	رتبة	الحكم
38	يعرض متن السؤال مشكلة واضحة ومحددة ودقيقة وقصيرة .	0.67	1.58	0	6	5	ضعيف
				0.0	50.0	41.7	
39	يتضمن متن السؤال كل الكلمات التي تتكرر في جميع البدائل.	1.00	1.92	1	3	4	متوسط
				8.3	25.0	33.3	
40	يتراوح عدد البدائل بين (٤-٥) لتقليل نسبة التخمين.	0.97	1.75	0	7	1	متوسط
				0.0	58.3	8.3	
41	تجانس جميع البدائل في محتواها	0.49	1.33	0	8	4	ت

م	العبارة		عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
	وترتبط بمجال السؤال.		0.0	33.3	66.7	0.0				ضعيف
42	ترتب بدائل الاجابة أبجديا للكلمات، وتنازلياً أو تصاعدياً للأرقام.		0	1	11	0	1.08	0.29	36%	ضعيف
	0.0	8.3	91.7	0.0						
43	تتساوى بدائل الإجابة في الأطوال والعدد والتنسيق غالباً.		0	6	6	0	1.50	0.52	50%	ضعيف
	0.0	50.0	50.0	0.0						
44	تجنب استخدام كلمات التعميم والتخصيص (غالباً، دائماً....		0	7	5	0	1.58	0.51	53%	ضعيف
	0.0	58.3	41.7	0.0						
45	تجنب صياغة بدائل كل / جميع ما تقدم" و " لاشيء مما ذكر" ما أمكن.		1	6	5	0	1.67	0.65	56%	متوسط
	8.3	50.0	41.7	0.0						
46	تجنب النفي والنفي المزدوج في المتن والبدايل، وتمييز النفي للتنبيه.		0	6	6	0	1.50	0.52	50%	ضعيف
	0.0	50.0	50.0	0.0						
47	تجنب التطابق اللغوي بين متن السؤال والبدايل (تذكيراً وتأنيثاً وجمعا وتنثية) الموحية بالإجابة.		0	0	12	0	1.00	0.00	33%	ضعيف
	0.0	0.0	100.0	0.0						
48	مراعاة توزيع بدائل الإجابات الصحيحة عشوائياً.		0	2	10	0	1.17	0.39	39%	ضعيف
	0.0	16.7	83.3	0.0						
	المتوسط الكلي لهذا البعد									
							1.46	1.97	49%	ضيف

تبين نتائج المعروضة في الجدول (١٣) أن المتوسط العام لهذا البعد المتعلق بدرجة الالتزام بمعايير صياغة أسئلة الاختيار من متعدد بلغ (١.٤٦)، وبانحراف معياري (١.٩٧)، وهي قيمة متوسط يقابلها درجة تقدير (ضعيفة)، وقد تراوحت قيم المتوسطات على جميع فقرات هذا البعد بين (١.٩٢) في حدها الأعلى، وكانت لفقرة " يتضمن متن السؤال جميع الكلمات التي تتكرر في جميع البدائل المقترحة"، وبين (١.٠٠) في حدها الأدنى، وكانت لفقرة " تجنب الارتباط والتطابق اللغوي بين متن السؤال والبدايل (تذكيراً وتأنيثاً وجمعا وتنثية) الموحية بالإجابة"، مما يدل على أن التزام الاساتذة بمعايير صياغة أسئلة الاختيار من متعدد قد انحصر بين التقدير (المتوسط، والضعيف)، ولم تظهر النتائج أي تقدير بدرجة عالية على أي فقرة على مستوى هذا البعد ككل.

أما ترتيب فقرات هذا البعد فجاءت حسب المتوسط الحسابي، ووفقاً لمعيار التصحيح كما يأتي :

- تحققت ٣ معايير من معايير صياغة أسئلة الاختيار من متعدد بدرجة (متوسطة)، تمثلت في الفقرات ذوات الأرقام (٣٩، ٤٠، ٤٥)، نالت المتوسطات الحسابية (١.٩٢، ١.٧٥، ١.٦٧) على الترتيب، واستحقت الثلاث المراتب الأولى، تمثلت بالنص " يتضمن متن السؤال جميع الكلمات التي تتكرر في جميع البدائل المقترحة، يتراوح عدد البدائل بين (٤-٥) لتقليل نسبة التخمين، تجنب صياغة بدائل "كل / جميع ما تقدم" و " لا شيء مما ذكر" ما أمكن، ما يعني أن ٥٦% - ٦٤% من عينة الاختبارات قد التزم معديها بهذه المعايير.

- تحقق ٧ من معايير صياغة هذا النوع من الأسئلة بدرجة (ضعيفة)، مثلتها الفقرات ذوات الأرقام (٤٤، ٣٨، ٤٣، ٤١، ٤٨، ٤٧، ٤٢) على الترتيب، تراوحت متوسطها الحسابي في حده الأعلى والأدنى بين (١.٥٨، ١.٠٠)، وحصلت على الرتبة الرابعة وحتى العاشرة، ما يدل على أن مستوى التزام معدي الاختبارات كان متدنياً وضعيفاً في المعايير الآتية (تجنب استخدام كلمات التعميم والتخصيص...، وعرض المتن لمشكلة واضحة وقصيرة ، وتساوى البدائل في الطوال والعدد والتنسيق، ومراعاة تجانس البدائل، وتوزيع بدائل الإجابة عشوائياً، وكذا ترتيب البدائل ترتيباً أبجدياً أو منطقياً، وتجنب التطابق اللغوي بين المتن والبدائل " تذكيراً وتأنياً وجمعاً وتثني"، وأي ارتباط موحى بالإجابة).

ومما ينبغي التأكيد عليه هنا أن أشد هذه المعايير تدنياً حسب ما أظهرته النتائج تمثلت في عدم ترتيب البدائل (أبجدياً للكلمات، وتنازلياً أو تصاعدياً للأرقام والتواريخ)، وكذا في عدم تجنب التطابق والارتباط اللغوي أو التراكبي بين متن السؤال والبدائل؛ إذ أن المتوسطات الحسابي لهما قد اقترب كثيراً من قيم الفئة المنعدمة، كما تظهر النتائج في الجدول أن ٣٣% فقط من الاختبارات ألتزم معديها بهذه المعايير؛ في حين أن ٦٧% من الاختبارات لم تلتزم بتلك المعايير، مما أفقد تلك البدائل أو (الموهات/المشتتات) التي يضعونها سمة الفعالية أو ما يعرف بالقدرة التمييزية؛ لأنها تصبح سهلة الاستبعاد.

ويرجع الباحثان ما أظهرته النتائج المعروضة سابقاً إجمالاً من ضعف وتدني التزام معدي الاختبارات بمعايير بصياغة نمط أسئلة الاختيار من متعدد في هذا البعد إلى صعوبة إعداد وصياغة هذا النوع من فقرات الأسئلة، وما يتطلبه وضعها وصياغة فقراتها من مهارة وخبرة عالية لدى معديها، وما يفرضه ذلك بالضرورة إقامة دوار تدريبية لتنمية مهارات وكفايات الأساتذة بصورة مستمرة وتوفير فرص الخبرة والممارسة، وهو الأمر الذي يغيب غياباً شبه كلي إن لم يكن منعدماً.

تكشف النتائج المعروضة سابقاً إجمالاً أن ما أظهرته النتائج من نسب ومتوسطات لدرجة الالتزام بمعايير الصياغة الجيدة لأنواع الأسئلة على مستوى كل بعد من الأبعاد الفرعية الخمسة المكونة لهذا المحور، وعلى مستوى المحور ككل غير كافية ولا ترقى إلى المستوى المطلوب أو المتوقع، إذ أقتصرت التحقق

والتوافر على بعد واحد فقط بدرجة متوسطة، هو بعد معايير صياغة أسئلة الإكمال، وبنسبة (٦٢%)، في حين كان جميع الأبعاد الفرعية المكونة لهذا المحور ضمن الفئة الضعيفة، إذ تراوح المتوسط الحسابي لها بين ((1.59 في حده الأعلى، وهو بعد معايير صياغة أسئلة الصح والخطأ، وبمتوسط (١.٣٣) في حده الأدنى، والمتعلق ببعد أسئلة المطابقة، ونتيجة لذلك فقد جاء متوسط درجة التزامهم بمعايير صياغة أنواع الأسئلة الاختبارية بدرجة ضعيفة على مستوى هذا المحور ككل بجميع أبعاده الفرعية.

وتتفق هذه النتائج مع ما توصلت إليه نتائج دراسة كل من ساعد، ووسيلة (٢٠١٣) من أن معايير كتابة فقرات الاختيار جاءت بنسب ضعيفة.

للإجابة على السؤال الثالث:

ونصه " ما مدى التزام معدي الاختبارات بمعايير الطباعة والإخراج في الورقة الاختبارية؟، تم حساب المتوسطات الحسابية، والانحرافات المعيارية، والنسب المئوية لكل فقرة من فقرات المحور، والبالغ عددها (١٦) معياراً، وكذلك الدرجة الكلية على مستوى المحور ككل.

جدول (١٤)

يبين المتوسطات الحسابية والانحرافات المعيارية والرتبة لمعايير جودة الطباعة والإخراج للورقة الاختبارية

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
49	طباعة الاختبار آلياً وليس بخط اليد.	46	1	0	0	2.98	0.15	99%	عالي
		97.9	2.1	0.0	0.0				
50	طباعة الاختبار بحجم خط مناسب واضح ومقروء.	21	18	8	0	2.28	0.74	76%	متوسط
		44.7	38.3	17.0	0.0				
51	رأس السؤال مكتوب بخط ١٦ عريض لتمييزه، مثل (السؤال الأول).	12	7	14	14	1.36	1.17	45%	ضعيف
		25.5	14.9	29.8	29.8				
52	وجود مسافة تفصل بين كل سؤال والذي يليه بمقدار ١,٥.	5	15	22	5	1.43	0.83	48%	ضعيف
		10.6	31.9	46.8	10.6				
53	خلو ورقة الاختبار من الأخطاء اللغوية والإملائية والمطبعية ونحوها.	8	12	27	0	1.60	0.77	53%	ضعيف
		17.0	25.5	57.4	0.0				
54	مراعاة علامات الترقيم في الكتابة.	5	6	36	0	1.34	0.67	45%	ضعيف
		10.6	12.8	76.6	0.0				
55	وجود جدول خاص لدرجات الأسئلة والمجموع	4	0	1	27	0.41	1.01	14%	منعدم

تقويم الاختبارات التحصيلية بكلية التربية والعلوم التطبيقية

بأرحب جامعة صنعاء في ضوء معايير الجودة : —

م	العبارة	عالي	متوسط	ضعيف	غير متوفر	M	N	%	الحكم
	النهائي على غلاف الاختبار.	12.5	0.0	3.1	84.4				
56	يوجد مستطيل لدرجة أعمال السنة على غلاف الورقة.	4	0	0	28	0.38	1.01	13%	منعدم
		12.5	0.0	0.0	87.5				
57	توجد نماذج لتفريغ الإجابة على ورقة غطاء الاختبار (لأسئلة الموضوعية).	0	0	3	29	0.09	0.30	3%	منعدم
		0.0	0.0	9.4	90.6				
58	تعطى أسئلة الاختبار وفقراتها الفرعية أرقاماً منظمة ومتسلسلة.	13	20	13	1	1.96	0.81	65%	متوسط
		27.7	42.6	27.7	2.1				
59	تنسيق ورقة الاختبار بشكل جذاب (تنسيق النص، المسافة بين الفقرات)	2	9	33	3	1.21	0.62	40%	ضعيف
		4.3	19.1	70.2	6.4				
60	عدم ضغط الأسئلة في الورقة بحيث تبدوا مكدسة ومربكة للطلبة عند قراءتها .	3	17	27	0	1.49	0.62	50%	ضعيف
		6.4	36.2	57.4	0.0				
61	الفراغات المخصصة للإجابة كافية (إذا كانت الإجابة في ذات الورقة).	1	8	20	3	1.22	0.66	41%	ضعيف
		3.1	25.0	62.5	9.4				
62	ترقيم صفحات الورقة الاختبارية، والتنويه بكلمة (يتبع) أسفل الورقة .	8	5	9	25	0.91	1.16	30%	منعدم
		17.0	10.6	19.1	53.2				
63	تنزيل الورقة — (انتهت الأسئلة، وبدعاء للطلبة)	4	9	30	4	1.28	0.74	43%	ضعيف
		8.5	19.1	63.8	8.5				
64	كتابة أسم أستاذ المقرر نهاية ورقة الاختبار.	33	2	2	10	2.23	1.25	74%	متوسط
		70.2	4.3	4.3	21.3				
	المتوسط الكلي للمحور					1.49	0.44	50%	ضعيف

تبين النتائج المعروضة في الجدول (١٤) أن المتوسط العام للمحور الخاص بمعايير الطباعة والإخراج للورقة الاختبارية بلغ (1.49)، وبانحراف معياري (0.44)، وهذا يعني أن درجة الالتزام بمعايير الطباعة والإخراج للورقة الاختبارية من قبل معدي الاختبارات كان بدرجة (ضعيفة) وفقاً لمعيار التصحيح، وقد تراوحت قيم المتوسطات على جميع فقرات هذا المحور بين (٢.٩٨) في حدها الأعلى، وكانت لفقرة "طباعة الاختبار آلياً وليس بخط اليد"، وبين (٠.٠٩) في حدها الأدنى، وكانت لفقرة "توجد نماذج لتفريغ الإجابة على ورقة غطاء الاختبار حال كانت أسئلة موضوعية"، مما يدل على أن مستوى التزام معدي الاختبارات بمعايير الطباعة والإخراج قد انحصر بين التقدير (العالي، والمنعدم) على مستوى المحور ككل.

أما ترتيب الفقرات لهذا المحور فجاءت حسب المتوسط الحسابي؛ ووفقاً لمعيار التصحيح كما يأتي:

- تحقق معيار واحدة فقط من معايير الطباعة والإخراج بدرجة (عالية)، مثلته الفقرة رقم (٤٩)، بمتوسط حسابي (٢.٩٨)، واستحقت الرتبة الأولى، تمثلت بالنص "طباعة الاختبار آلياً وليس يدوياً"، وهي نتيجة متوقعة فجميع الاختبارات يتم طباعتها، إلا أن نموذج اختباري واحد تضمن بعض الرسوم والكتابة يدوياً للتفاصيل، مما أدى إلى ظهور تكرار واحد بدرجة متوسطة.
- تحقق ٣ معايير بدرجة (متوسطة)، تمثلت في الفقرات ذوات الأرقام (٥٠، ٦٤، ٥٨)، كانت متوسطاتها الحسابية (٢.٢٨، ٢.٢٣، ١.٩٦) على الترتيب، وحصلت على الرتبة الثانية وحتى الرابعة، نصت على معايير (طباعة الاختبار بحجم خط مناسب ومقروء، وكتابة أسم أستاذ المقرر نهاية الورقة، وإعطاء الأسئلة وفقراتها الفرعية أرقاماً منظمة ومتسلسلة)، وتظهر هذه النتيجة قصوراً غير مبرر في التزام معدي الاختبار بهذه المعايير؛ على الرغم من بساطة تحقيقها، الأمر الذي يعكس ضعف درجة الاهتمام والحرص وغياب المتابعة.
- تحققت ٨ معايير بدرجة (ضعيفة)، تمثلت في الفقرات ذوات الأرقام (٥٣، ٦٠، ٥٢، ٥١، ٥٤، ٦٣، ٦١، ٥٩)، حيث تراوح متوسطها الحسابي في حده الأعلى والأدنى بين (١.٦٠ - ١.٢١)، وحصلت على الرتب من الخامسة وحتى الثانية عشرة، والتي نصت بـ (ترتيب الأسئلة حسب مستوى الصعوبة، صحة مجموع الدرجات، شمول الأسئلة لموضوعات المقرر، ولجوانب التعلم).
- جاء ٤ من المعايير بدرجة (منعدمة)، مثلتها الفقرات ذوات الأرقام (٦٢، ٥٥، ٥٦، ٥٧)، حيث كانت المتوسطات الحسابية لها (٠.٩١، ٠.٤١، ٠.٣٨، ٠.٠٩) على الترتيب، وتقع قيم هذا المتوسط ضمن الفئة المنعدمة؛ لأنها أقل من ١ الصحيح؛ لذلك استحقت الأربع الرتب الأخيرة، وتدل هذه النتيجة على عدم مراعاة معدي الاختبارات لعدد من المعايير هي عدم ترقيم الصفحات والتتويه بكلمة (يتبع) أسفل الورقة، وعدم وجود مستطيل لدرجة أعمال السنة، وأخيراً وأكثر واشد المعايير انعداماً عدم وجود نماذج تفريغ للإجابة على ورقة غطاء الاختبار (الأسئلة الموضوعية).

الإجابة عن السؤال الرئيس للدراسة:

جدول (١٥)

يبين المتوسطات الحسابية والانحرافات المعيارية والترتيب لمعايير الجودة للمحاور الرئيسة الثلاثة لورقة الاختبار

م	المحاور الرئيسة	العدد	المتوسط الحسابي	الانحراف المعياري	%	الحكم اللفظي
1	معايير الاطار العام للورقة الاختبارية ببدي (الشكل والمضمون).	17	1.20	.335	40%	ضعيف
2	معايير جودة الصياغة لأنواع الأسئلة الاختبارية	31	1.52	.386	51%	ضعيف
3	معايير الطباعة والإخراج.	16	1.49	.437	50%	ضعيف
	متوسط العام لمحاور المقياس ككل	64	1.37	.289	46%	ضعيف

تبين النتائج في الجدول (١٥) أن المتوسط الكلي لدرجة التزام معدي الاختبارات بمعايير جودة

الاختبارات على مستوى المحاور الثلاثة الرئيسة المكونة لمقياس تقدير جودة الاختبارات ككل بلغ (١.٣٧)، وبانحراف معياري (289)، وهذا يعني أن التزامهم بمعايير جودة الاختبار كان بدرجة (ضعيفة) على مستوى جميع المحاور المكونة للمقياس بشكل عام، وقد حصل المحور الثاني الخاص بمعايير الصياغة لأنواع الأسئلة المقالية والموضوعية بأبعاده الفرعية الخمسة على المرتبة الأولى، بمتوسط حسابي (1.52)، وانحراف معياري (386)، يليه محور معايير الطباعة والإخراج لورقة الاختبار في المرتبة الثانية، بمتوسط حسابي (1.37)، وانحراف معياري (289)، بينما جاء المحور الأول الخاص بمعايير جودة الإطار العام للورقة الاختبارية ببديه الفرعيين من حيث (الشكل والمضمون) في المرتبة الثالثة والأخيرة، بمتوسط حسابي (1.20) وانحراف معياري (335)،

وتشير النتائج المعروضة إجمالاً على مستوى المحاور الثلاثة الرئيسية بأن هناك ضعفاً وتدنياً واضحاً في مستوى التزام معدي الاختبارات بالمعايير اللازم مراعاتها عند إعداد وتصميم وإخراج ورقة الاختبارات التحصيلية وصياغة فقراتها، إذ لم تظهر النتائج درجة توافر واستيفاء بدرجة (عالية)، أو على الأقل بدرجة (متوسطة) على أي محور من المحاور الرئيسة بأبعادها الفرعية المكونة لمقياس تقدير جودة الورقة الاختبارية، عدى بعد واحد من الأبعاد الفرعية الخمسة للمحور الثاني، كان بدرجة متوسطة، والمتعلق بمعايير صياغة أسئلة الإكمال فقط، بينما كانت قيمة متوسطات بقية المحاور والأبعاد الرئيسة والفرعية داخل المقياس ككل تقع ضمن فئة درجة التقدير الضعيفة، مما يعطي مؤشراً، ويدل بجلاء على واقع التدهور الشديد

في مستوى جودة الاختبارات التحصيلية بالكلية، وضعف درجة التزام الأساتذة بتلك المعايير والمواصفات، ودرجة استيفائهم لها والعمل على توافرها فيما يعدونه من تلك الاختبارات.

لأن ما أظهرته النتائج من نسب ومتوسطات حسابية لكل محور من المحاور، وعلى مستوى المحاور الثلاثة الرئيسية ككل، وما بينته من نسب للعبارة المتضمنة للمواصفات والمعايير المختلفة داخل تلك المحاور والأبعاد الفرعية لها غير كافية، ولا ترقى إلى المستوى المطلوب والمتوقع تربوياً.

ويعود السبب في ذلك غالباً إلى ضعف مستوى التكوين لدى الأساتذة سواء في المسار الدراسي أو المهني، وافتقار معظم الكوادر إلى مهارات وكفايات بناء الاختبارات، وبخاصة في ظل غياب التنمية الذاتية وبرامج التدريب المتسمة المستهدفة تنمية مهارات وكفايات بناء الاختبارات التحصيلية وصياغة فقراتها وفق جدول مواصفات، وقد يرجع ذلك إلى أن وضع الأساتذة للاختبارات يتم بطريقة عشوائية وارتجالية، بعيداً عن الخطوات العلمية لإعداد الاختبارات ووفقاً لجدول المواصفات، نتيجة واستسهالهم للأساليب التقليدية التي اعتادوا عليها، وخاصة مع غياب وضع آليات الرقابة والمتابعة المستمرة من الكلية، مما انعكس سلباً على تدني جودة ونوعية ما يقدمونه وما ينجزونه من أعمال ومهام، وما رافق ذلك من انعدام تام لوسائل التجويد والحوافز المادية أو الاهتمام والتشجيع للإبداع والمتميزين داخل الكلية والجامعة.

وتتفق نتائج الدراسة الحالية إجمالاً مع ما خلصت إليه نتائج دراسة كلاً من ساعد، ووسيلة (٢٠١٣) من أن معايير كتابة تعليمات الاختبار التحصيلي الجيد واعداد فقراته جاءت بنسب ضعيفة، وأن أكثر الأسئلة استخداماً لدى الأساتذة الأسئلة المقالية، إلا أنها اختلفت معها جزئياً فيما كشفتته من توافر معايير الطباعة والإخراج بدرجة متوسطة، كما تتشابه نتائج الدراسة الحالية جزئياً مع نتائج دراسة السفياي (٢٠١٩)، والحسينان، وآخرون (٢٠١٤)، ومكاوي (٢٠١٩) والتي خلصت إلى تدني نسب معايير المضمون، كما تتفق مع نتائج دراستي القرشي (٢٠١٠)، وعبدالمع (٢٠٠٩) اللتان خلصتا إلى افتقار معظم نماذج الاختبارات إلى الصدق والشمول والموضوعية، وأن تلك الاختبارات اقتصر على قياس التحصيل في الجانب المعرفي، وفي أدنى مستوياته الحفظ والتذكر، وأهملت قياس المستويات العليا، وجوانب التعلم المهارية والوجدانية، وافتقارها للتنوع والشمول.

الفصل الرابع.

نتائج الدراسة وتوصياته ومقترحاتها:.

أولاً- النتائج:

(١) ضعف وتدني مستوى جودة الاختبارات التحصيلية المعدة من قبل أساتذة التعليم الجامعي بكلية التربية أرحب - جامع صنعاء، إذ لم تتجاوز نسب توافر معايير ومواصفات الجودة في أوراق الاختبارات التحصيلية (٤٦%)، وبدرجة تقدير (ضعيفة) لجميع المحاور الرئيسية الثلاثة بأبعادها الفرعية المكونة للمقياس ككل، وفقاً لمعيار تفسير النتائج المعتمد في هذه الدراسة بشكل عام.

(٢) كانت درجة التزام معدي الاختبارات بمعايير ومواصفات الإطار العام للورقة الاختبارية بدرجة تقدير ضعيفة، إذ لم تتجاوز نسبة استيفائهم لمعايير هذا المحور ككل (٤٠%)، وقد جاء بعد معايير المضمون أولاً، بنسبة (٤٦%)، يليه بعد المعايير الشكلية ثانياً، بنسبة (٣٦%)، وهذه النسب تعد أدنى نسب المعايير استيفاء على مستوى المحاور الرئيسة الثلاثة المكونة للمقياس، لذلك جاء هذا المحور ببعديه في المرتبة الأخيرة من بين المحاور الثلاثة المكونة للمقياس.

(٣) إن نسبة التزام معدي الاختبارات بمعايير جود الصياغة لأنواع الأسئلة الاختبارية المقالية والموضوعية بنسبة (٥١%)، وبدرجة ضعيفة، بينما تراوحت نسب التزامهم بمعايير أنواع الأسئلة في الأبعاد الفرعية الخمسة تحت هذا المحور بين (٤٤% - ٦٢%)، وبدرجة ضعيفة لجميع الأبعاد، ما عدى بعد معايير صياغة أسئلة الإكمال كان بدرجة متوسطة، وقد أحتل هذا المحور بأبعاده الفرعية الخمسة المرتبة الأولى من بين المحاور الرئيسية الثلاثة المكونة للمقياس.

(٤) إن نسبة التزام معدي الاختبارات بمعايير الطباعة والإخراج للورقة الاختبارية بدرجة ضعيفة، إذ لم تتجاوز نسبة استيفائهم لمعايير المحور ككل نسبة (٥٠%)، وحصل على المرتبة الثانية من بين المحاور الرئيسة للمقياس.

ثانياً- التوصيات :

في ضوء ما كشفته النتائج والاستنتاجات، فإن الدراسة تقترح التوصيات العلمية والعملية الآتية:

- (١) ضرورة قيام جامعة صنعاء وكلية التربية بشكل خاص بعقد دورات تدريبية وورش عمل حول بناء الاختبارات وفقاً لجدول المواصفات، وجودة صياغة أنواع فقرات الأسئلة الاختبارية بشكل خاص.
- (٢) حث أساتذة الكلية على تصميم أوراق الاختبار بحسب المعايير الفنية والتربوية المتعارف عليها، والالتزام بنموذج موحد لتحقيق الاختبارات الغرض منها بشكل أفضل.

- (٣) تفعيل دور قسم الجودة في مساعدة الأساتذة ذوي الخبرات الأقل في بناء الاختبارات وفق المواصفات العلمية، وإخضاع عينة من أوراق الاختبارات للمراجعة والتحليل والتقويم المستمر.
- (٤) ضرورة إعداد ونشر دليل المعلم الإرشادي بأدوات وأساليب التقويم الحديثة، والخطوات العلمية في بناء الاختبارات التحصيلية وفقاً لجدول المواصفات ومخرجات التعلم.
- (٥) ضرورة التحول نحو فلسفة التقويم التكويني المستمر في مختلف الأقسام بالكلية، والأخذ بنظام ملفات تقويم أداء الطالب وبما يحقق الشمولية والتنوع في القياس والتقويم.
- المقترحات.

المقترحات:

- اجراء دراسة عن مدى وعي أعضاء هيئة التدريس بمعايير بناء الاختبارات وفق جدول مواصفات ومخرجات التعلم.
- ٢_ اجراء دراسات مماثلة مقارنة بين أكثر من كلية في الجامعات اليمنية.
- ٣_ اجراء دراسة حول فاعلية برنامج تدريبي مقترح في تنمية مهارات بناء الاختبارات لدى الاستاذ الجامعي.

المراجع:

أولاً- المراجع العربية:

- ١- أبو علام، رجاء محمود (١٩٨٧): قياس وتقويم التحصيل الدراسي، دار العلم ، الكويت.
- ٢- أبو مغلي، سمير وسلامة، عبد الحافظ (٢٠١٠): القياس والتقويم في التربية الخاصة، دار اليازوري العلمية للنشر، الأردن.
- ٣- جيوسي، مهدي (٢٠١٦): تحليل أسئلة الامتحانات النهائية في جامعة فلسطين التقنية للأهداف التعليمية تبعاً لهرم بلوم في ضوء معايير الورقة الامتحانية الجيدة، المجلة الدولية التربوية، ج (٥)، ع (٨)، ص ٣٧-١٥.
- ٤- الحاج، أحمد علي، وماجد إسماعيل (٢٠١٣): مناهج البحث العلمي، المتفوق للطباعة، صنعاء.
- ٥- الحسينان، إبراهيم وآخرون (٢٠١٤): تقويم اختبارات المقررات التربوية في كلية التربية جامعة المجمعة في ضوء معايير الجودة، مجلة كلية التربية، ج٢٤، ع٤- جامعة الإسكندرية، ص٢٣-٩٠.
- ٦- حلس، داور درويش (٢٠٠٤): دراسة تقويمية للأخطاء الكتابية الشائعة لدى تلاميذ الصف السادس الأساسي في محافظة غزة"، أطروحة دكتوراه (غير منشورة)، جامعة الخرطوم.
- ٧- حمادة، شريف (٢٠١٣): جودة أسئلة الامتحانات النصفية لبرنامج التربية بجامعة القدس المفتوحة في ضوء معايير الجودة، ورقة مقدمة إلى مؤتمر جودة الامتحانات الجامعية" المنعقد في ٢٠١١، غزة.

- ٨- الحيلة، محمد (٢٠٠٢): مهارات التدريس الصفي، دار المسيرة للنشر والتوزيع، عمان، الأردن.
- ٩- دعمس، مصطفى (٢٠١٠): استراتيجيات التقويم التربوي الحديث وأداته، دار غيدا للنشر، عمان.
- ١٠- رشيد خضر (٢٠٠٤): التقويم لتربوي، دار القلم للنشر، دبي، الإمارات العربية المتحدة.
- ١١- الريماوي، عمر (٢٠١٧): بناء وتصميم الاختبارات والمقاييس النفسية والتربوية، دار أمجد، عمان.
- ١٢- ساعد، صباح وعامر، وسيلة (٢٠١٧): تقييم كفاية بناء الاختبارات التحصيلية لدى أساتذة التعليم الجامعي وفق معايير الاختبار الجيد، مجلة العلوم الإنسانية والاجتماعية، ع (٢٨).
- ١٣- السفياني، هلال محمد (٢٠١٩): تقييم أوراق امتحانات المقررات التربوية بكلية التربية بالمهرة وفقاً لمواصفات الامتحان ومعايير صياغة فقراته، مجلة مركز جزيرة العرب للبحوث النفسية والتربوية، ج (١)، ع (٥)، ص ص ٣١-١.
- ١٤- السنبل، عبد العزيز (٢٠٠١): مبادئ وإجراءات ضبط الجودة النوعية، مجلة تعليم الجماهير، ع (٤٨)، المنظمة العربية للتربية والتعليم والعلوم.
- ١٥- الصلاحي، عبدالسلام محمد (١٩٩٨): بناء معايير للاختبارات الوزارية في الثانوية العامة في مادة التاريخ، وتطورها في اليمن، رسالة دكتوراه غير منشورة، كلية التربية، ابن رشد، العراق.
- ١٦- العبادي، رائد خليل (٢٠٠٦): الاختبارات المدرسية، مكتبة المجتمع العربي، الأردن.
- ١٧- العبادي، عضيد عبد (٢٠١٨): الاختبارات في التربية الإسلامية (أنواعها، سماتها، عيوبها، كيفية صياغتها)، مجلة كلية التربية الأساسية للعلوم التربوية والإنسانية، ع (٣٨)، جامعة بابل.
- ١٨- عبد الرحمن، أحمد (٢٠١٠): تصميم الاختبارات أسس نظرية وتطبيقات عملية، دار أسامة للنشر، الأردن.
- ١٩- عبد المقصود، محمد (٢٠٠٩): استراتيجيات تدريس الاجتماعيات، دار المعرفة، الإسكندرية.
- ٢٠- عبد المنعم، محمد (٢٠٠٩): واقع أسئلة الاختبارات الفصلية للطلاب الخريجين بكلية المعلمين جامعة الملك فيصل في ضوء معايير جودة التقويم، مجلة البحوث النفسية والتربوية، ج (٤)، ع (٣)، كلية التربية - جامعة المنوفية.
- ٢١- العجيلي: صباح حسين (٢٠١١): القياس والتقويم التربوي، ط٤، مركز التربية للنشر، صنعاء.
- ٢٢- عمر، محمود وآخرون (٢٠١٠): القياس النفسي والتربوي، دار المسيرة للنشر، عمان.
- ٢٣- القرشي، خديجة ضيف الله (٢٠١٠): تقويم الورقة الاختبارية لطالبات البكالوريوس بكلية التربية جامعة الطائف وفق معايير الجودة، بحث مقدم للمؤتمر الإقليمي الثاني لعلم النفس، المنعقد بالقاهرة، خلال الفترة من ٢٨ نوفمبر - ١ ديسمبر، رابطة الاخصائيين النفسيين المصرية، مصر.
- ٢٤- قطامي، يوسف وقطامي، نايفه (٢٠٠١): سيكولوجية التدريس، دار الكندي للنشر، الأردن.
- ٢٥- كاظم، علي مهدي (٢٠٠١): القياس والتقويم في التعليم والتعلم، دار الكندي للنشر، الأردن.

- ٢٦- مجيد، سوسن شاكر (٢٠١٤): أسس بناء الاختبارات والمقاييس النفسية والتربوية، مركز دبيونو لتعليم التفكير، ط: ٣.
- ٢٧- مخايل، امطانيوس (٢٠١٦): بناء الاختبارات والمقاييس النفسية والتربوية، دار الإعصار العلمي للنشر، عمان.
- ٢٨- مكاوي، أماني (٢٠١٩): مدى جودة الاختبارات التحصيلية في التعليم الجامعي في ضوء معايير الجودة الشاملة بكلية التربية بعفيف جامعة شقراء، مجلة إدارة الجودة الشاملة، ج (٢٠)، ع (١)، جامعة السودان للعلوم والتكنولوجيا، السودان، ص ص ٩٦ - ١٠٧.
- ٢٩- ملحم، سامي محمد (٢٠١٠): مناهج البحث في التربية وعلم النفس، دار المسيرة، الأردن.
- ٣٠- هاشم، علي محمد هادي (٢٠١٤): إعداد نماذج للاختبارات التحصيلية في مادة القرآن الكريم في ضوء معايير الاختبار الجيد، رسالة ماجستير غير منشورة.

ثانيا- المراجع الأجنبية:

- ٣١- Arter&mctighe (2001):scoring rubrics in the classroom. Using performance criteria for assessing and improving student performance thousand oaks.corwing press, inc.
- ٣٢- Nuryulia, I. R. (2009): ITEM ANALYSIS OF TEST IN FINAL TEST FOR THE SEVENTH GRADE STUDENTS OF SMP NI MOGA PEMALANG in the academic year of 2008/2009 (Doctoral dissertation, UniversitasNegeri Semarang).
- ٣٣- Stufflbeam, D (1997):Professional Standard for Educational Evaluation International Journal of Education Research, 11, (1)