

الخصائص السيكومترية لاختبار تحصيلي محكي المرجع في مهارات

القياس والتقويم لدى طلبة كلية التربية- جامعة إب.

Psychometric Characteristics of Criterion-Referenced Achievement Test in Measurement and Evaluation Skills among Graduates of Faculty of Education – University of Ibb

إعداد

حليمة علي صالح الحدأ

Halima Ali Saleh Al Hada

مدرس مساعد- قسم العلوم النفسية والتربوية كلية التربية جامعة إب- اليمن

أ.د/ مأمون علي ناجي البناء

Prof. Mamoun Ali Naji Al-Binaa

أستاذ القياس والتقويم والإحصاء التربوي المشارك - كلية التربية جامعة إب- اليمن

Doi: 10.21608/jasep.2021.181939

قبول النشر: ٢٠٢١/٣/٦

استلام البحث: ٢٠٢١/٢/٢٢

الحدأ ، حليمة علي صالح - البناء ، مأمون علي ناجي (٢٠٢١). الخصائص السيكومترية لاختبار تحصيلي محكي المرجع في مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب. مج ٥، ٢١٤، المجلة العربية للعلوم التربوية والنفسية، المؤسسة العربية للتربية والعلوم والآداب ، ص ص ٢١٥ – ٢٤٤.

الخصائص السيكومترية لاختبار تحصيلي محكي المرجع في مهارات القياس والتقويم لدى
طلبة كلية التربية- جامعة إب

المستخلص:

هدفت الدراسة الحالية إلى بناء اختبار تحصيلي محكي المرجع في مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- الجمهورية اليمنية، و التعرف على مدى تمتع درجاته بخاصيتي (الصدق الوصفي، والصدق الوظيفي)، كذلك معرفة مدى تمتع درجاته بخاصية الثبات وفقاً لمعامل ليفنجستون، بالإضافة إلى تحديد درجة القطع المناسبة التي من خلالها يتم تصنيف الطلبة إلى متقنين وغير متقنين لمهارات القياس والتقويم التي ينبغي تمكنهم منها، وتمثل الهدف الأخير للدراسة بمعرفة مدى تمكن طلبة كلية التربية- جامعة إب من مهارات القياس والتقويم التي تضمنها الاختبار الذي تم بناؤه. منهج بحوث التقويم هو المنهج المستخدم في الدراسة وتكون مجتمع الدراسة من جميع طلبة المستوى الرابع الذين تم تدريسهم مقرر القياس والتقويم من كلية التربية- جامعة إب للعام الجامعي (٢٠١٨-٢٠١٩م) والذي بلغ عددهم (٤٢٥) طالباً وطالبة، أما عينة الدراسة فقد كانت نوعان هما عينة للتجريب الأول والتي تكونت من (٢٩٠) طالباً وطالبة وعينة التطبيق النهائي والتي تكونت من (١١٨) طالباً وطالبة، ولتحقيق أهداف الدراسة المتمثلة ببناء اختبار تحصيلي محكي المرجع في مهارات القياس والتقويم لدى طلبة كلية التربية فقد تم اتباع أربع مراحل في بناء الاختبار، هي: (مرحلة التحليل، مرحلة البناء، مرحلة التجريب، ومرحلة الإخراج والتطبيق النهائي للاختبار). كان أهم النتائج التي توصلت إليها الدراسة هو تمتع درجات الاختبار الذي تم بناؤه في مهارات القياس والتقويم بصدق (وصفي، وظيفي جيد)، كذلك تمتعها بخاصة ثبات عالٍ وفقاً لمعامل ليفنجستون؛ إذ بلغ (٠,٩٢)، ودرجة قطع وفقاً لطريقة أنجوف بلغت (٠,٧٥) تبين من خلالها تدني مستوى طلبة كلية التربية- جامعة إب في مهارات القياس والتقويم، وتكون الاختبار بصورته النهائية من (٨٢) مفردة اختبارية من نوع الاختيار من متعدد ذي أربع بدائل.

الكلمات المفتاحية: الاختبارات التحصيلية، الاختبار محكي المرجع، درجة القطع، القياس والتقويم.

Abstract:

This study aimed at constructing criterion-referenced achievement test in measurement and evaluation skills among graduates of faculty of education, University of Ibb, Republic of Yemen. Moreover, it aimed at identifying range to which degrees of criterion-referenced achievement test constructed and used to measure skills of measurement and evaluation of the property of descriptive validity and functional validity, and identifying range to which the degrees of achievement test have reliability feature in

accordance with coefficient of Livingstone. In addition, determining appropriate Cut-off score according to Ansoff's method through which students are classified into proficient and non-proficient for measurement and evaluation skills to be available to them. The final aim is identifying range to which, the students of Faculty of Education- University of Ibb, were able to measurement and evaluation skills included in the test which was constructed. The evaluation research methodology is the method was used in the study. Research population consisted of (425) students at the fourth level, who were taught the course of measurement and evaluation at Faculty of Education- University of Ibb for the academic year (2018-2019). Population of research was divided into two samples according to its procedures: the first experiment sample consisted of (290) students, and the second application sample consisted of (118) students. In order to achieve the aims of study represented in constructing a reference-based achievement test in the measurement and evaluation skills of the students of the faculty of Education, four phases were followed in the construction of the test, namely: (the analysis phase, the construction phase, the experimentation phase, the output phase and the final application of the test. The most important results of the study were that the degrees of the criterion-referenced achievement test which was constructed for measuring measurement and evaluation skills had appropriate (descriptive validity and functional validity). Moreover, it has appropriate reliability (0.92) according to Livingstone coefficient. The Cut-off score according to Angoff's method was (0.75). So, there was a weakness in measurement and evaluation skills among students of Faculty of Education - University of Ibb. And the final test consisted of (82) test items of the multiple-choice type with four alternatives.

Key words: achievement tests, criterion-referenced test, degree of cutting, measurement and evaluation.

المقدمة Introduction:

يشكل التعليم الجامعي قمة الهرم في منظومة التعليم المؤسسي العام، وتعتمد جودة التعليم الجامعي على مدى قدرته على تخرج أفراد يحملون من المعارف والمهارات المهنية التي تتلاءم مع متطلبات سوق العمل وتحقيق الأهداف العملية والتعليمية في شتى مجالات المعرفة؛ حيث يعد تقويم أداء طلبة التعليم الجامعي هو الركن الأهم في العملية التعليمية، كون أهمية التقويم تبرز في مدى قدرته على تحديد وانتقاء الطلبة الذين حققوا أهداف التعليم المرجوة (نور الدين، ٢٠١٧، ٤٨).

لذا فإن الجامعات تسعى إلى تدريب وتأهيل الشباب وتزويدهم بالمهارات والمعارف والمعلومات في شتى مجالات المعرفة، ومن أهم تلك المهارات (المهارات القياسية) التي تعد من أهم الأهداف الرئيسية للتربية المعاصرة (الترتوري، ٢٠١٨: ٧٦). كون قياس وتقييم تحصيل مخرجات التعليم خطوة في غاية الأهمية؛ لما يقدمان من معلومات مفيدة في التعرف على مقدار التقدم في تحصيل الطلبة ومدى تحقيق الأهداف التعليمية المرجوة (علي، ٢٠١٢: ٢).

فيلزم إعداد أدوات قياس تتميز بالموضوعية والمصدقية، يمكن من خلالها التحقق من مستوى تمكن أو إتقان الطلبة للمهارات التعليمية وتحقيق الأهداف التعليمية المرجوة، مما يستدعي تطبيق بعض الإجراءات والأساليب التي يمكن باستخدامها تحقيق بعض متطلبات الدقة والموضوعية المنشودة كالاختبارات التحصيلية.

فالاختبارات التحصيلية ذو أهمية بالغة في اتخاذ القرارات المتعلقة بالمتعلمين أو بمستوياتهم التعليمية أو غيره، ومن تلك الاختبارات (الاختبارات محكية المرجع) فهي أكثر أنواع الاختبارات ملاءمة لقياس تحصيل الطلبة وتقويم أدائهم؛ كونها تركز على اكتساب الطلبة مهارات ومعارف معينة، ومدى تحقيقهم لأهداف محددة، واتخاذ قرارات تعليمية معينة تتعلق بمدى إتقان الطلبة للمحتوى التعليمي المحدد، ناهيك أنها لا تعتمد على مقارنة أداء الفرد بأداء أقرانه كالاختبارات معيارية المرجع (علام، ٢٠٠١: ٢٥٤)، بل تعتمد في جوهرها على تقدير مستوى أداء الفرد بالنسبة لمحك أو مستوى أداء أو درجة قطع محددة تفصل بين المتقنين وغير المتقنين للمحتوى التعليمي المقاس (الشريفين، ٢٠٠٦: ٨٢-٨٣).

وبناءً على ما سبق تبدو الحاجة ملحة لتوفر أداة قياس موضوعية، يمكن من خلالها قياس مستوى تحصيل طلبة البكالوريوس في الجامعات اليمنية في مهارات القياس والتقويم.

مشكلة الدراسة Study Problem:

تحدد مشكلة الدراسة بالإجابة عن الأسئلة الآتية:

- ما مهارات القياس والتقويم التي ينبغي امتلاكها لدى طلبة كلية التربية- جامعة إب- كما يراها أعضاء هيئة التدريس المختصين في القياس والتقويم وكما هي في توصيف محتوى وأهداف مقرر القياس والتقويم المقدمة لهم؟

- هل تتمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الصدق الوصفي؟

- هل تتمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الصدق الوظيفي؟

- هل تتمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الثبات وفقاً لمعامل (ليفنجستون)؟

- ما درجة القطع المناسبة التي من خلالها يتم تصنيف الطلبة إلى متقنين وغير متقنين لمهارات القياس والتقويم التي ينبغي توفرها لديهم؟

- ما مستوى تمكن طلبة كلية التربية- جامعة إب- من مهارات القياس والتقويم التي تضمنها الاختبار الذي تم بناؤه؟

أهداف الدراسة : Study Aims :

تهدف الدراسة إلى التعرف على:

- مهارات القياس والتقويم التي ينبغي ان يمتلكها طلبة كلية التربية- جامعة إب.

- مدى تمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الصدق الوصفي.

- مدى تمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الصدق الوظيفي.

- مدى تمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الثبات وفقاً لمعامل (ليفنجستون).

- تحديد درجة القطع المناسبة التي من خلالها يتم تصنيف الطلبة إلى متقنين وغير متقنين لمهارات القياس والتقويم التي ينبغي توفرها لديهم.

- مستوى تمكن طلبة كلية التربية- جامعة إب- من مهارات القياس والتقويم التي تضمنها الاختبار الذي تم بناؤه.

أهمية الدراسة : Study Importance:

تبرز أهمية الدراسة الحالية في الآتي:

- تزويد المختصين بمجال القياس والتقويم من أعضاء هيئة التدريس في الجامعات اليمنية بالمهارات القياسية الرئيسية التي ينبغي توفرها لدى طلبة كليات التربية، ومراعاة هذه المهارات عند توصيف مقرر القياس والتقويم وتدريبهم له.
- استفادة أعضاء هيئة التدريس من الاختبار التحصيلي الذي تم بناؤه المتميز بمواصفات الاختبار الجيد، ما يمكنهم من استخدامه مؤشراً على مستوى إتقان الطلبة لمهارات القياس والتقويم.
- يعد الاختبار أداة مناسبة لتشخيص ومعرفة الأهداف التي تحققت، والأهداف التي لم تتحقق في تدريس مقرر القياس والتقويم في كليات التربية بالجامعات اليمنية، واقتراح الأساليب التعليمية والتدريبية في ضوء نتائج عملية التشخيص.
- تفيد الدراسة الحالية طلبة الدراسات العليا في اتباع الإجراءات المستخدمة لجمع دلالات (الصدق، الثبات) للاختبارات التحصيلية محكية المرجع سواء كانت تلك الاختبارات في مجال التعليم الجامعي أو التعليم العام.
- تسهم نتائج الدراسة في فتح المجال أمام الباحثين للقيام بدراسات وبحوث مماثلة وعلى نطاق أوسع في مقررات تعليمية أخرى ولمختلف مراحل التعليم الجامعي.
- تمثل الاختبارات محكية المرجع أبرز التوجهات المعاصرة للقياس النفسي والتربوي، لذلك تعد هذه الدراسة إسهام علمي متواضع في التراكم العلمي المعرفي القائم في هذا الجانب وفق الاتجاه المعاصر.

مصطلحات الدراسة: Definition of Terms

١. الخصائص السيكومترية: The Psychometric Characteristics

عرف (مراد وسليمان، ٢٠٠٢: ٣٥٠) الخصائص السيكومترية بأنها: تلك الخصائص الضرورية والمتعلقة بالصدق والثبات والمعايير التي يتم حسابها بعد تجريب الاختبار على عينة ممثلة للمجتمع. ومن أهم تلك الخصائص التي يعتمد عليها في إصدار القرارات حول دقة نتائج الاختبارات وصلاحيتها، هي (الصدق، الثبات)؛ حيث إن الثبات يهتم باتساق ما يقيس، في حين أن الصدق ينظر إلى ما إذا كان الاختبار يقيس ما نريد قياسه أم لا (أبو علام، ٢٠٠٥: ٣٧٣-٣٧٨).

ويعرفها الباحثان إجرائياً:

بأنها تلك الخصائص المرتبطة بذات الاختبار، التي يمكن التعبير عنها بدلالات رقمية، كالخصائص المتعلقة بالدرجة الكلية للاختبار (الصدق، الثبات).

٢. الصدق: Validity

عرفه (علام، ٢٠٠٦: ٣٩) بأنه: مدى استفادتنا من الدرجات التي يحصل عليها الطلبة في اتخاذ قرارات مناسبة تتعلق بمواقف معينة بقدر كبير من الثقة. أما (العجيلي، ٢٠١١: ١١٧)

عرفه بأنه: "مدى قدرة الاختبار وصلاحيته في قياس الظاهرة (الخاصية، أو السمة، أو القدرة) التي صمم لقياسها. ويعرفه الباحثان إجرائياً:

بأنه مدى قدرة الاختبار التحصيلي محكي المرجع المستخدم في قياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب على قياس ما وضع لقياسه- وذلك من خلال استخدام عدة طرائق للصدق منها (الصدق الوصفي، الصدق الوظيفي).

٣. الثبات: Reliability

عرفه (علام، ٢٠٠٠: ١٣١) بأنه: مدى خلو الدرجات من الأخطاء غير المنتظمة التي تشوب القياس، أي مدى قياس الاختبار للمقدار الحقيقي للسمة التي تهدف لقياسها. ويعرف بأنه مدى اتساق البيانات التي تجمع عن طريق إعادة المقاييس نفسها على الأفراد أو الظواهر نفسها وتحت الظروف نفسها أو تحت ظروف مشابهة إلى أكبر قدر ممكن (ملحم، ٢٠٠٢: ٤٥).

ويعرفه الباحثان إجرائياً:

بأنه مدى تمتع درجات الاختبار التحصيلي محكي المرجع المستخدم في قياس مهارات القياس والتقويم بثبات جيد وفقاً لمعامل (ليفنجستون).

٤. الاختبارات التحصيلية Achievement Tests

يرى (أبو حطب، ١٩٨٧: ٣٨٧) أن الاختبارات التحصيلية هي: "الأداة التي تُستخدم في قياس المعرفة والفهم والمهارات في مادة دراسية أو تدريبية معينة أو مجموعة من المواد. وعرفت بأنها إجراءات منظمة تهدف إلى قياس كمية من المعلومات التي يتقنها الطلبة أو لتحديد مقدار ما تعلمه الطالب (ملحم، ٢٠٠٩: ٢٠٦).

٥. الاختبار محكي المرجع: Criterion Referenced Test

عرفه بابام (Popham, 1978: 93-94) بأنه: ذلك الاختبار الذي يستخدم للتأكد من وضع الفرد أو حالته، بالنسبة لنطاق سلوكي محدد تحديداً دقيقاً. في حين عرفه كل من كوبزن وبوريتش (Kubiszyn & Borich, 2000: 33) بأنه: الاختبار الذي يخبرنا عن مستوى أداء الطالب في بعض المهارات المحددة، وذلك بمقارنة أدائه بمستوى إتقان محدد مسبقاً يسمى المحك.

ويعرفه الباحثان إجرائياً:

بأنه الاختبار الذي يستخدم في تقدير أداء طلبة كلية التربية - جامعة إب- لمهارات القياس والتقويم والذي تم تحديد درجة القطع له مسبقاً، التي ينبغي على الطلبة الحصول عليها لتحقيق النجاح.

٦. درجة القطع Cut-off-Score:

يرى هامبلتون (Hambleton, 1978: 279) أنها: نقطة على متصل درجات الاختبار، تستخدم لتصنيف الطلبة إلى فئتين تعكس كل منهما مستويات الأداء المختلفة بالنسبة لهدف معين يراد قياسه في الاختبار. في حين عرفها (علام، ٢٠٠١: ٢٥٣) بأنها: الدرجة التي تفصل بين المتمكنين وغير المتمكنين، أو بين مستويات مختلفة من التمكن على متصل درجات النطاق السلوكي الشامل للاختبار ويعرفها الباحثان إجرائياً:

بأنها الدرجة المعتمدة محكاً للإتقان التي تفصل بين طلبة كلية التربية المتقنين وغير المتقنين لمهارات القياس والتقويم، والتي سوف تحدد وفقاً لطريقة أنجوف.

٦. القياس والتقويم Evaluation and Measurement:

- القياس: Measurement:

عرفه (عودة، ٢٠٠٥: ٣٤) بأنه: "العملية التي يتم بواسطتها التعبير عن الخصائص والسمات بالأرقام". وعرفه (مخائيل، ٢٠١٥: ١٨) بأنه: التقدير الكمي باستخدام لغة العدد في التعبير عن الظواهر والسمات المقاسة. ويعرفه الباحثان إجرائياً:

بأنه العملية التي سيتم بناءً عليها إعطاء الأفراد أو خصائصهم قيمة رقمية وفق قاعدة معينة، وذلك لبيان درجة تواجد الخاصية المقاسة لديهم.

- التقويم: Evaluation:

عرفه (العجيلي، ٢٠١١: ١١) بأنه: عبارة عن إصدار حكم قيمي على خصائص الأشياء المقدره تقديراً كمياً أو كيفياً في ضوء معيار أو محك، واتخاذ القرارات بشأنها. بينما (مجيد، ٢٠١٤: ٢٩) عرفه بأنه: الحكم على الشيء ويتجاوز هذا الحكم إلى التحسين والتطوير. ويعرفه الباحثان إجرائياً:

بأنه عملية تهدف إلى جمع البيانات للطلبة حول الاختبار التحصيلي محكي المرجع في مهارات القياس والتقويم وتحليلها من أجل إصدار حكم على مدى توفر تلك المهارات وفقاً لدرجة القطع المحددة مسبقاً لغرض التحسين والتطوير لتلك المهارات.

حدود الدراسة Limitations: Study

تحدد بدراسة الخصائص السيكومترية (الصدق، الثبات) لاختبار تحصيلي محكي المرجع في مهارات القياس والتقويم لدى الطلبة الذين تم تدريسهم مقرر القياس والتقويم بكلية التربية- جامعة إب للعام الجامعي (٢٠١٨/٢٠١٩م).

منهج الدراسة Method of Study:

استخدم الباحثان منهج بحوث التقويم؛ كون الدراسة تهدف إلى بناء اختبار تحصيلي محكي المرجع لمهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب، لأنه المنهج المناسب لتحقيق أهداف دراستهم؛ فالغرض من بحوث التقويم عادة تطوير العملية التعليمية، وتقويم

الطالب والمعلم، كذلك تقويم العملية التعليمية بوجه عام، بغرض النهوض بها، والتأكد من تحقيق الأهداف (أبو علام، ٢٠١١: ٣٥٢).

مجتمع الدراسة: Study Population:

تكون مجتمع الدراسة من جميع طلبة البكالوريوس في كلية التربية جامعة إب الذين درسوا مقرر القياس والتقويم في المستوى الرابع بالفصل الدراسي الأول للعام الجامعي (٢٠١٨/٢٠١٩م) كونه مقرر متطلب في جميع التخصصات بكلية التربية، وبلغ عددهم (٤٢٥) طالبًا وطالبة موزعين على (١٢) تخصص، والجدول (١) يوضح مجتمع الدراسة.

جدول (١) مجتمع الدراسة

العدد			الجنس التخصص	العدد			الجنس التخصص
المجموع	إناث	ذكور		المجموع	إناث	ذكور	
١١	١٠	١	تربية خاصة	١٢٥	٩٤	٣١	تكنولوجيا تعليم
١٦	١١	٥	تربية فنية	٤١	٢٩	١٢	رياضيات
٦٧	٦٥	٢	رياض أطفال	٢٤	١٨	٦	فيزياء
١٤	١٢	٢	إرشاد تربوي	٢٩	٢٣	٦	كيمياء
٦	٦	-	تعليم كبار	٣٢	٢٨	٤	لغة عربية
١٥	٢	١٣	إدارة وتخطيط	٤٥	٢٦	١٩	علوم قرآن
١٢٩	١٠٦	٢٣	المجموع	٢٩٦	٢١٨	٧٨	المجموع
٤٢٥							المجموع الكلي

بما أن الدراسة تستهدف الدارسين لمقرر القياس والتقويم فقط وهم طلبة المستوى الرابع والموضح تفصيلهم بالجدول (١) في مجتمع الدراسة، والاختبار الذي تم بناؤه يحتاج لعينتين وفقاً لخطوات بناء الاختبار، فقد تم اعتماد المجتمع كاملاً عينةً للدراسة وقد قسم المجتمع إلى عينتين، هما: عينة التطبيق الأول وعينة التطبيق النهائي، والغرض هو تجنب انتقال أثر التعليم، وفيما يلي جدولان يوضحان خصائص العينتين:

جدول (٢) عينة التطبيق الأول حسب التخصص والجنس

العدد			الجنس التخصص	العدد			الجنس التخصص
المجموع	إناث	ذكور		المجموع	إناث	ذكور	
٧	٦	١	تربية خاصة	١٢٥	٩٤	٣١	تكنولوجيا تعليم
١٦	١١	٥	تربية فنية	٣٨	٢٨	١٠	رياضيات
٦٧	٦٥	٢	رياض أطفال	٢٤	١٨	٦	فيزياء
٩٠	٨٢	٨	المجموع	١٣	١١	٢	ارشاد تربوي
				٢٠٠	١٥١	٤٩	المجموع
٢٩٠							المجموع الكلي

يتضح من الجدول (٢) أن عينة التطبيق الأول تمثلت بطلبة المستوى الرابع لـ (٧) أقسام بلغ عددهم (٢٩٠) طالبًا وطالبة؛ حيث تم تطبيق الاختبار بصورته الأولية عليهم للتحقق من صدق درجات الاختبار وثباتها، وتحليل المفردات الاختبارية (الصعوبة، التمييز)، ومعرفة وضوح التعليمات، ووضوح مفردات الاختبار لجميع الطلبة، وتقدير الوقت اللازم لإجراء الاختبار، وإتمام خطوات بناء الاختبار.

جدول (٣) عينة التطبيق النهائي حسب التخصص والجنس

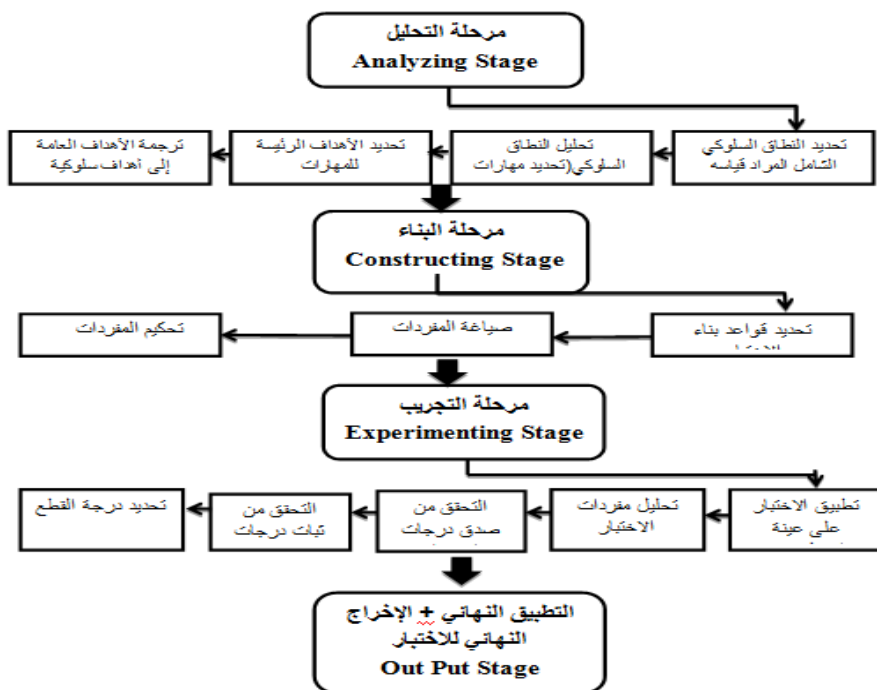
العدد			الجنس التخصص
المجموع	إناث	ذكور	
٣٢	٢٨	٤	لغة عربية
٤٥	٢٦	١٩	علوم قرآن
٢٧	٢١	٦	كيمياء
٥	٥	-	تعليم كبار
٩	٢	٧	إدارة وتخطيط
١١٨	٨٢	٣٦	المجموع
١١٨			المجموع الكلي

الجدول (٣)، يوضح توزيع عينة التطبيق النهائي المتمثلة بطلبة المستوى الرابع من (٥) أقسام هي قسم (اللغة العربية، علوم قرآن، الكيمياء، تعليم كبار، الإدارة والتخطيط التربوي) البالغ عددهم (١١٨)؛ حيث تم تطبيق الاختبار بصورته النهائية عليهم للتحقق من خصائصه السيكومترية مرة أخرى وذلك لتحقيق الأهداف المرجوة من الدراسة.

خطوات بناء الاختبار : Steps of Constructing Test

تم بناء اختبار تحصيلي محكي المرجع لقياس مدى تمكن طلبة كلية التربية من مهارات القياس والتقويم، وذلك في ضوء خطوات معتمدة عند بناء هذا النوع من الاختبارات كما أشارت مراجع ودراسات الاختبارات المحكية، ومنها (علام، ٢٠٠١؛ الشرفين، ٢٠٠٦؛ البناء، ٢٠٢٠؛ الحموري والكحلوت، ٢٠٠٩؛ ناجي، ٢٠١٠؛ البناء، ٢٠١١؛ علي، ٢٠١٢؛ السامراني والخفاجي، ٢٠١٢؛ سليمان والصالح، ٢٠١٦؛ سرققوة، ٢٠١٦؛ العبيدي، ٢٠١٧؛ عليلي، ٢٠١٧؛ الزبيدي، ٢٠١٨). والشكل (١) يوضح خطوات بناء الاختبار:

شكل (١) خطوات بناء الاختبار



نتائج الدراسة Results of Study:

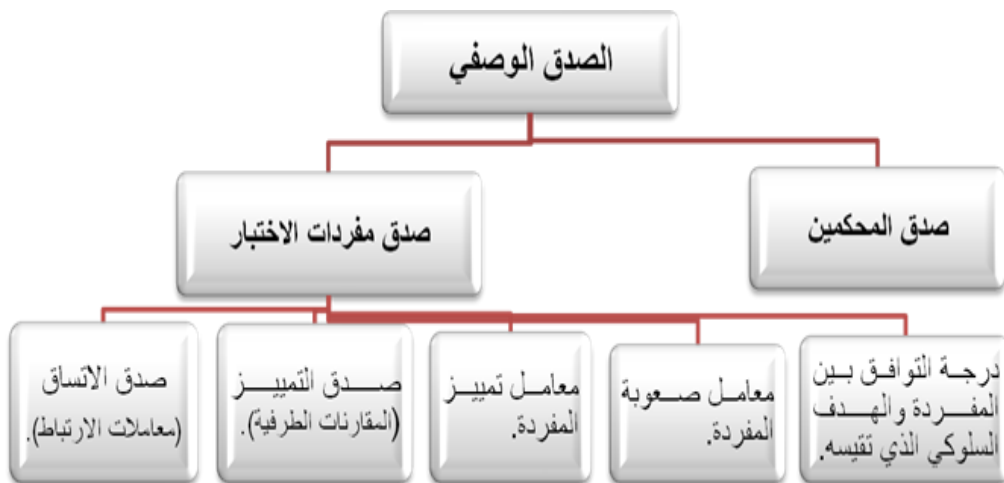
السؤال الأول: ما مهارات القياس والتقويم التي ينبغي امتلاكها لدى طلبة كلية التربية- جامعة إب- كما يراها أعضاء هيئة التدريس المختصين في القياس والتقويم، وكما هي في توصيف محتوى وأهداف مقرر القياس والتقويم المقدمة لهم؟

وللإجابة عن هذا السؤال: تم تحليل التوصيف والأهداف الخاصة بمقرر القياس والتقويم المقرر لطلبة البكالوريوس بكليات التربية في الجامعات اليمنية، ومنها جامعة (إب، ذمار، صنعاء، تعز، وعدن)، كذلك تحليل الكتاب المقرر في القياس والتقويم على طلبة كلية التربية بجامعة- إب، بالإضافة إلى الاطلاع على عدد من مراجع وأدبيات القياس والتقويم، ومن ثم تم بناء استبانة تضمنت مجموعة من مفردات مهارات القياس والتقويم وعرضها على (١٠) محكمين من الأساتذة والخبراء في التخصص وممن لديهم خبرة ودرسوا مقرر القياس والتقويم في الجامعات اليمنية وجامعات عربية لتحكيم مفرداتها من حيث الأهمية، حيث حددت خمس مهارات في القياس والتقويم هي: (التعريف بالمفاهيم والمصطلحات الأساسية بالقياس والتقويم، الإلمام بمفهوم الصدق وطرائق التحقق منه والعوامل المؤثرة فيه، الإلمام بمفهوم الثبات وطرائق التحقق منه والعوامل المؤثرة فيه، بناء الاختبارات التحصيلية بأنواعها المختلفة، إعداد المقاييس النفسية والتربوية وبنائها).

ومن ثم تم اتباع خطوات بناء الاختبارات محكية المرجع لقياس تلك المهارات؛ حيث تم تحديد الأهداف السلوكية وبناء المفردات الاختبارية؛ حيث أن كل مفردة تقيس هدفًا سلوكيًا مختلفًا وعرضها على (٢٥) محكمًا وخبيرًا، والأخذ بملاحظاتهم وآرائهم وإجراء التعديلات اللازمة، حتى تم إخراج الاختبار بصورته النهائية مكونًا من (٨٢) مفردة اختبارية من نوع الاختيار من متعدد بأربعة بدائل.

السؤال الثاني: هل تتمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الصدق الوصفي؟ وللإجابة عن هذا السؤال: تم اتباع إجراءات معينة لكي يتسنى للباحثان الحكم على الصدق الوصفي لدرجات الاختبار، واعتبرت الإجراءات التي تمت في بناء أداة الدراسة دليلاً على تحقيقه، وفي ما يلي شرح مفصل لما تم القيام به للتحقق من كل نوع من أنواع الصدق الوصفي المستخدمة في تحديد مدى صدق درجات مفردات الاختبار الذي تم بناؤه والموضحة في الشكل (٢):

شكل (٢) أنواع الصدق الوصفي



- الوصفي لدرجات الاختبار بدلالة صدق المحكمين:

تم الاستعانة بأراء مجموعة من الخبراء والمختصين بمجال (القياس والتقويم) من خلال الأخذ برأي (١٠) من المحكمين والخبراء لتحديد أهمية المهارات التي ينبغي امتلاكها لدى طلبة كلية التربية- جامعة إب. ومن ثم أخذ رأي (١٥) محكمًا وخبيرًا للحكم على مدى سلامة ووضوح صياغة الأهداف السلوكية (الإجرائية) ومدى انتماءها وارتباطها بالأهداف الرئيسية للمجالات المراد قياسها. وكذلك الحكم من قبل (٢٠) محكمًا على مدى التناسق بين مفردات الاختبار وأهدافها السلوكية والمجالات المنتمية لها التي صممت لقياسها. ومن ثم تم تحديد نسب الاتفاق بين المحكمين حيث أظهرت نتائج التحكيم نسب اتفاق عالية بين المحكمين، وانتهاءً بتحكيم (١١) محكمًا وخبيرًا لتحديد درجة قطع باستخدام طريقة أنجوف.

- الوصفي لدرجات الاختبار بدلالة صدق مفردات الاختبار:

وذلك للتعرف على مدى فاعلية مفردات الاختبار، ومدى إسهامها في الحصول على خصائص سيكومترية تتفق مع خصائص الاختبار الجيد، ولتحقيق هذا النوع من الصدق تم اتباع عدة خطوات هي (إيجاد درجة التوافق بين المفردة والهدف السلوكي الذي تقيسه، وإيجاد درجة صعوبة كل مفردة من المفردات وقدرتها التمييزية، واستخدام الاختبار التائي للمقارنة بين درجات المجموعتين العليا والدنيا، وحساب معامل ارتباط بيرسون بين درجات

المجالات والدرجة الكلية للاختبار، ومعامل ارتباط بايسيريال بين درجات المفردات ودرجة المجالات المنتمية إليها).

درجة التوافق بين المفردة والهدف السلوكي الذي تقيسه:

بعد صياغة مفردات الاختبار وطباعتها، وللتحقق من درجة توافقها مع الهدف السلوكي الذي تقيسه، تم عرضها على مجموعة محكمين من ذوي الاختصاص والخبرة بلغ عددهم (٢٠) محكمًا، ومن ثم القيام بحساب نسبة الاتفاق والاختلاف بين المحكمين والخبراء حول مدى التوافق والارتباط بين تلك المفردات والأهداف السلوكية (الإجرائية) التي وضعت لقياسها، والأخذ بنسب الاتفاق والاختلاف بين المحكمين لتكون معيارًا للحكم على مدى قياس المفردة للهدف السلوكي بالمهارات الرئيسة واعتمدت المفردات التي حصلت على نسبة قبول (٨٠%) فأكثر من نسب الاتفاق بين المحكمين، وبناء على آراء المحكمين تبين أن الاختبار ممثل للمحتوى (النطاق السلوكي) المراد قياسه.

معامل صعوبة المفردة Items Difficulty:

لتحديد معامل الصعوبة لكل مفردة اختبارية تم تحديد مجموعتين (عليا، دنيا) باستخدام الإرباعيات، حيث بلغ عدد أفراد كل مجموعة (٣٠) طالبًا، ثم تم حساب معامل الصعوبة لكل سؤال وذلك بقسمة حاصل جمع الذين اجابوا اجابة صحيحة في المجموعة العليا والدنيا على مجموع عدد المجموعتين وهو ٦٠ وكانت معاملات الصعوبة كما هو موضح بالجدول التالي:

جدول (٤) معاملات صعوبة مفردات الاختبار.

رقم المفردة	قيمة معامل صعوبتها	رقم المفردة	قيمة معامل صعوبتها	رقم المفردة	قيمة معامل صعوبتها	رقم المفردة	قيمة معامل صعوبتها	رقم المفردة	قيمة معامل صعوبتها
١	0.92	١٨	0.57	٣٥	0.52	٥٢	0.90	٦٩	0.35
٢	0.82	١٩	0.60	٣٦	0.70	٥٣	0.88	٧٠	0.38
٣	0.78	٢٠	0.72	٣٧	0.52	٥٤	0.93	٧١	0.27
٤	0.62	٢١	0.85	٣٨	0.68	٥٥	0.72	٧٢	0.48
٥	0.92	٢٢	0.85	٣٩	0.43	٥٦	0.82	٧٣	0.80
٦	0.88	٢٣	0.48	٤٠	0.45	٥٧	0.55	٧٤	0.70
٧	0.75	٢٤	0.33	٤١	0.50	٥٨	0.83	٧٥	0.63
٨	0.77	٢٥	0.67	٤٢	0.55	٥٩	0.77	٧٦	0.65
٩	0.92	٢٦	0.70	٤٣	0.48	٦٠	0.78	٧٧	0.60
١٠	0.97	٢٧	0.52	٤٤	0.33	٦١	0.40	٧٨	0.48

0.55	٧٩	0.38	٦٢	0.75	٤٥	0.37	٢٨	0.68	١١
0.48	٨٠	0.42	٦٣	0.58	٤٦	0.42	٢٩	0.48	١٢
0.57	٨١	0.47	٦٤	0.73	٤٧	0.73	٣٠	0.73	١٣
0.73	٨٢	0.73	٦٥	0.48	٤٨	0.67	٣١	0.63	١٤
		0.35	٦٦	0.78	٤٩	0.52	٣٢	0.78	١٥
		0.65	٦٧	0.92	٥٠	0.67	٣٣	٠,٢٤	١٦
		0.30	٦٨	0.75	٥١	0.70	٣٤	0.37	١٧

نلاحظ من الجدول (٤) أن هناك تنوعاً بمعاملات الصعوبة بين (٠,٢٤ - ٠,٩٧) كما يلاحظ أن اغلب المفردات كانت متوسطة في معامل صعوبتها حيث تراوح معامل الصعوبة لأغلب الفقرات بين ٠,٣٠ و ٠,٧٠ .

معامل تمييز المفردة Items Discrimination :

تم الاعتماد على طريقة المقارنة الطرفية لحساب معامل التمييز بين درجات الفقرات للمجموعتين (العليا والدنيا) باستخدام الإرباعيات؛ حيث بلغ عدد أفراد كل مجموعة (٣٠) طالباً، وتم حساب معامل التمييز بقسمة حاصل الفرق بين عدد الذين اجابوا صح من المجموعة العليا والدنيا على نصف عدد المجموعتين وهو ٣٠ وكانت معاملات التمييز كما هو موضح بالجدول التالي :

جدول (٥) معاملات تمييز مفردات الاختبار.

رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها
١	٠,٥١	١٨	٠,٤٣	٣٥	٠,٧٠	٥٢	٠,٥٧	٦٩	٠,٣٠
٢	٠,٤٣	١٩	٠,٦٧	٣٦	٠,٦٧	٥٣	٠,٦٧	٧٠	٠,٦٣
٣	٠,٥٠	٢٠	٠,٤٧	٣٧	٠,٤٧	٥٤	٠,٥٠	٧١	٠,٤٣
٤	٠,٤١	٢١	٠,٤٠	٣٨	٠,٦٣	٥٥	٠,٦٧	٧٢	٠,٤٣
٥	٠,٦٧	٢٢	٠,٥٩	٣٩	٠,٥٠	٥٦	٠,٣٩	٧٣	٠,٤٠
٦	٠,٥٧	٢٣	٠,٧٧	٤٠	٠,٤٧	٥٧	٠,٤٠	٧٤	٠,٥٣
٧	٠,٤٩	٢٤	٠,٥٣	٤١	٠,٨٧	٥٨	٠,٦٧	٧٥	٠,٦٠
٨	٠,٥٣	٢٥	٠,٤٠	٤٢	٠,٣٨	٥٩	٠,٤٣	٧٦	٠,٤٠
٩	٠,٤٧	٢٦	٠,٤٣	٤٣	٠,٥٧	٦٠	٠,٢٩	٧٧	٠,٦٠
١٠	٠,٧٠	٢٧	٠,٦٣	٤٤	٠,٧٧	٦١	٠,٤٢	٧٨	٠,٧٧

رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها	رقم المفردة	قيمة معامل تمييزها
١١	٠,٥٠	٢٨	٠,٦٠	٤٥	٠,٤٠	٦٢	٠,٥٣	٧٩	٠,٧٠
١٢	٠,٥٧	٢٩	٠,٥٧	٤٦	٠,٣٠	٦٣	٠,٦٣	٨٠	٠,٩٠
١٣	٠,٤٧	٣٠	٠,٤٧	٤٧	٠,٣٩	٦٤	٠,٥٧	٨١	٠,٧٣
١٤	٠,٤٧	٣١	٠,٥٣	٤٨	٠,٤٩	٦٥	٠,٤٥	٨٢	٠,٤٨
١٥	٠,٦٣	٣٢	٠,٧٠	٤٩	٠,٤٠	٦٦	٠,٥٣		
١٦	٠,٥٧	٣٣	٠,٤٧	٥٠	٠,٦٧	٦٧	٠,٤٧		
١٧	٠,٦٠	٣٤	٠,٥١	٥١	٠,٣٩	٦٨	٠,٥٦		

يتضح من الجدول (٥) أن معظم مفردات الاختبار المحكي ذات تمييز أكثر من (٠,٤٠)، وهي معاملات تمييز يمكن القول إنها جيدة جداً أو ممتازة، وهنا يمكننا القول إن درجات مفردات الاختبار تتمتع بمؤشرات تمييزية جيدة.

الصدق الوصفي لدرجات الاختبار بدلالة الصدق التمييزي (صدق المقارنات الطرفية): Discriminant Validity

وللتحقق من الصدق الوصفي لدرجات الاختبار بدلالة المقارنة الطرفية تم استخدام الاختبار (t) لعينتين مستقلتين، وذلك بعد تحديد مجموعتين (عليا- دنيا) باستخدام الإرباعيات؛ حيث كانت قيمة الإرباعي الأدنى (٤١ درجة وأدنى)، وقيمة الإرباعي الأعلى (٥٩ درجة وأعلى)، كما بلغ عدد أفراد كل مجموعة (٣٠) طالباً، والجدول الآتي يوضح نتائج المقارنة بين المجموعتين (الصدق التمييزي).

جدول (٦) الصدق التمييزي بطريقة المقارنة الطرفية بين المجموعتين (العليا- الدنيا) للمجالات والدرجة الكلية للاختبار.

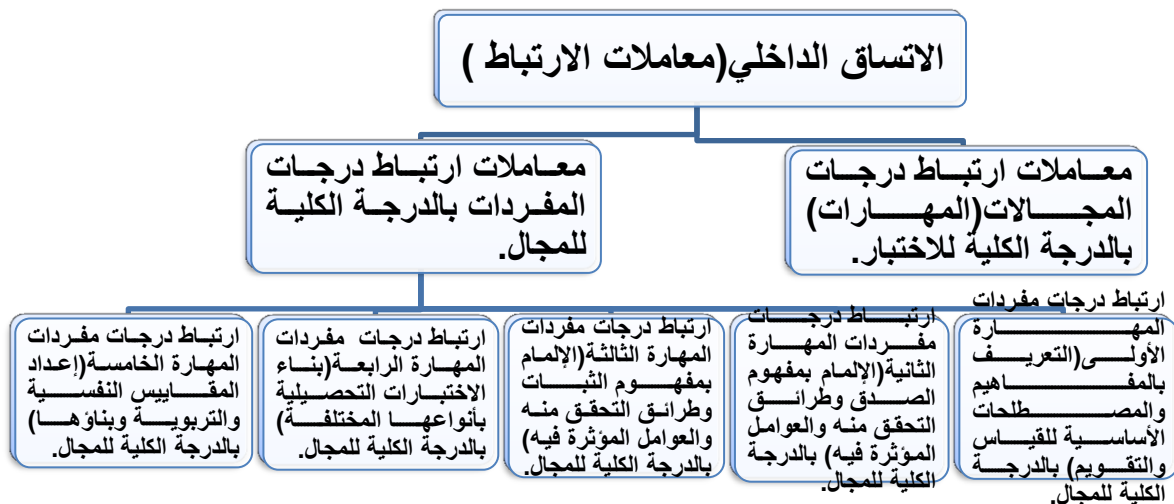
المجالات (المهارات)	المجموعة	العينة	المتوسط	الانحراف المعياري	درجة الحرية	قيمة (t)	مستوى الدلالة	القرار
المهارة الأولى	العليا	٣٠	١٧,٦٦	١,٢١	58	١٢,72	٠,٠٠	دالة
	الدنيا	٣٠	١٢,٠٠	٢,١١				
المهارة الثانية	العليا	٣٠	٩,٧٣	1.59	58	١٣,٩١	٠,٠٠	دالة
	الدنيا	٣٠	٤,١٠	1.53				
المهارة الثالثة	العليا	٣٠	٩,٨٠	2.02	58	١٠,٢٩	٠,٠٠	دالة
	الدنيا	٣٠	٤,٦٠	1.88				
المهارة الرابعة	العليا	٣٠	١٩,٠٣	2.49	58	١٠,٠٨	٠,٠٠	دالة

المجالات (المهارات)	المجموعة	العينة	المتوسط	الانحراف المعياري	درجة الحرية	قيمة (t)	مستوى الدلالة	القرار
	الدنيا	٣٠	١٢,٠٦	2.83				
المهارة الخامسة	العليا	٣٠	٩,٧٠	1.20	58	١٦,١٨	٠,٠٠٠	دالة
	الدنيا	٣٠	٣,٦٦	1.64				
الدرجة الكلية	العليا	٣٠	٦٥,٩٣	5.04	58	٢٤,٧٣	٠,٠٠٠	دالة
	الدنيا	٣٠	٣٦,٤٣	4.14				

بالنظر إلى الجدول (٦) يتضح أن قيم (t) للمقارنة بين درجات المجموعتين العليا والدنيا سواء في المجالات الخمسة (المهارات) أو الدرجة الكلية للاختبار، كانت دالة إحصائياً عند مستوى الدلالة (٠,٠٠٠)، وهي أقل من مستوى الدلالة (٠,٠٥)؛ وبذلك يمكن القول إنه يوجد فروق ذات دلالة إحصائية بين المجموعتين العليا والدنيا في الاختبار التحصيلي محكي المرجع لمهارات القياس والتقويم لصالح المتوسطات الأكبر، وهي متوسطات المجموعات العليا، وهذا مؤشر جيد على أن الاختبار يتمتع بصدق وصفي جيد (من الناحية التمييزية) بدليل المقارنات الطرفية بين المجموعتين العليا والدنيا.

- الصدق الوصفي لدرجات الاختبار بدلالة الاتساق الداخلي (معاملات الارتباط):
وللتحقق من الصدق الوصفي لدرجات الاختبار بدلالة معاملات الارتباط (الاتساق الداخلي) تم إيجاد معامل ارتباط درجة كل مجال مع الدرجة الكلية للاختبار باستخدام معامل ارتباط بيرسون، بالإضافة إلى إيجاد معامل ارتباط درجة المفردة بالدرجة الكلية للمجال الذي تنتمي إليه باستخدام معامل الارتباط الثنائي الأصيل (بوينت بايسيريال) كما هو موضح في الشكل رقم (٣)، وكانت النتائج على النحو التالي:

شكل (٣) أنواع معاملات الارتباط
أولاً- ارتباط درجات المجالات بالدرجة الكلية للاختبار:



جدول (٧) معاملات ارتباط بيرسون بين درجات المجالات والدرجة الكلية للاختبار.

القرار	مستوى الدلالة	معامل الارتباط	اسم المجال (المهارة)
دالة	0.000	0.86**	التعريف بالمفاهيم والمصطلحات الأساسية بالقياس والتقويم.
دالة	0.000	0.88**	الإلمام بمفهوم الصدق وطرائق التحقق منه والعوامل المؤثرة فيه.
دالة	0.000	0.87**	الإلمام بمفهوم الثبات وطرائق التحقق منه والعوامل المؤثرة فيه.
دالة	0.000	0.78**	بناء الاختبارات التحصيلية بأنواعها المختلفة.
دالة	0.000	0.79**	إعداد المقاييس النفسية والتربوية وبنائها.

يتضح من الجدول (٧) أن معاملات ارتباط بيرسون بين درجات المجالات الخمسة و الدرجة الكلية للاختبار تراوحت بين (٠,٧٨ – ٠,٨٨)، وهي معاملات ارتباط دالة إحصائياً عند مستوى دلالة (٠,٠٠٠)؛ وعليه يمكن القول إن درجات الاختبار المحكي المرجع تتمتع بصدق وصفي جيد وفقاً لمؤشر الاتساق الداخلي.
ثانياً- ارتباط درجات المفردات بالدرجة الكلية للمجالات كالاتي:

١- ارتباط درجة مفردات المجال الأول (التعريف بالمفاهيم والمصطلحات الأساسية بالقياس والتقييم) بالدرجة الكلية للمجال.
جدول (٨) معاملات ارتباط بوينت بايسيربال بين درجات المفردات والدرجة الكلية للمجال الأول.

المفردة	معامل الارتباط	المفردة	معامل الارتباط	المفردة	معامل الارتباط
١	0.60	٨	0.51	١٥	0.79
٢	0.65	٩	0.67	١٦	0.52
٣	0.70	١٠	0.55	١٧	0.56
٤	0.54	١١	0.74	١٨	0.68
٥	0.56	١٢	0.56	١٩	0.55
٦	0.69	١٣	0.80	٢٠	0.78
٧	0.52	١٤	0.69	٢١	0.67

من خلال الجدول (٨) تبين أن جميع معاملات ارتباط بوينت بايسيربال بين درجات مفردات المجال الأول و الدرجة الكلية للمجال تراوحت بين (٠,٥١ - ٠,٨٠)، وهذا دليل على صدق وصفي جيد لدرجات مفردات المجال الأول.

٢- ارتباط درجات مفردات المجال الثاني (الإلمام بمفهوم الصدق وطرائق التحقق منه والعوامل المؤثرة فيه) بالدرجة الكلية للمجال.
جدول (٩) معاملات ارتباط بوينت بايسيربال بين درجات لدرجات المفردات والدرجة الكلية للمجال الثاني.

المفردة	معامل الارتباط	المفردة	معامل الارتباط
١	0.48	٧	0.58
٢	0.54	٨	0.53
٣	0.56	٩	0.51
٤	0.57	١٠	0.50
٥	0.56	١١	0.59
٦	0.65	١٢	0.68

يتضح من قيم الجدول (٩) أن جميع معاملات ارتباط بوينت بايسيريال بين درجات مفردات المجال الثاني و الدرجة الكلية للمجال تراوحت بين (٠,٤٨ - ٠,٦٨)، وهي معاملات ارتباط جيدة، وهذا دليل على صدق وصفي جيد لدرجات مفردات هذا المجال.

٣- ارتباط درجات مفردات المجال الثالث (الإلمام بمفهوم الثبات وطرائق التحقق منه والعوامل المؤثرة فيه) بالدرجة الكلية للمجال.

جدول (١٠) معاملات ارتباط بوينت بايسيربال بين درجات المفردات والدرجة الكلية للمجال الثالث.

المفردة	معامل الارتباط	المفردة	معامل الارتباط
١	0.46	٨	0.59
٢	0.43	٩	0.60
٣	0.53	١٠	0.52
٤	0.54	١١	0.45
٥	0.42	١٢	0.56
٦	0.57	١٣	0.59
٧	0.51		

من خلال الجدول (١٠) تبين أن جميع معاملات ارتباط بوينت بايسيريال بين درجات مفردات المجال الثالث و الدرجة الكلية للمجال تفاوتت قيمها بين (٠,٤٢ - ٠,٦٠)، وهي معاملات ارتباط جيدة، وعليه يمكن القول إن جميع درجات مفردات المجال الثالث تتمتع بصدق وصفي جيد وفقاً لمؤشر صدق الاتساق الداخلي.

٤- ارتباط درجات مفردات المجال الرابع (بناء الاختبارات التحصيلية بأنواعها المختلفة) بالدرجة الكلية للمجال.

جدول (١١) معاملات ارتباط بوينت بايسيربال بين درجات المفردات والدرجة الكلية للمجال الرابع.

معامل الارتباط	المفردة	معامل الارتباط	المفردة	معامل الارتباط	المفردة	معامل الارتباط	المفردة	معامل الارتباط	المفردة
٠,٧٤	٢١	٠,٦٦	١٦	٠,٦٤	١١	٠,٨٦	٦	٠,٥٩	١
٠,٨٢	٢٢	٠,٥٧	١٧	٠,٧٣	١٢	٠,٦٧	٧	٠,٦٨	٢
٠,٥٩	٢٣	٠,٥٨	١٨	٠,٧٨	١٣	٠,٧٩	٨	٠,٧٥	٣
٠,٧٤	٢٤	٠,٦٩	١٩	٠,٧٥	١٤	٠,٥٢	٩	٠,٨٩	٤
٠,٨٩	٢٥	٠,٨٣	٢٠	٠,٦٥	١٥	٠,٥٨	١٠	٠,٧٧	٥

بالنظر في قيم الجدول (١١) يتضح أن جميع معاملات ارتباط بوينت بايسيربال لدرجات مفردات المهارة الرابعة مع الدرجة الكلية للمجال تراوحت قيمها بين (٠,٥٢ - ٠,٨٩)، وهي معاملات ارتباط جيدة، وهو مؤشر ودليل على صدق وصفي جيد لدرجات مفردات المجال الرابع (بناء الاختبارات التحصيلية بأنواعها المختلفة).

٥- ارتباط درجات مفردات المجال الخامس (إعداد المقاييس النفسية والتربوية وبنائها) بالدرجة الكلية للمجال.

جدول (١٢) معاملات ارتباط بوينت بايسيربال بين درجات المفردات والدرجة الكلية للمجال الخامس.

معامل الارتباط	المفردة	معامل الارتباط	المفردة
٠,٦٣	٧	٠,٨٨	١
٠,٦٧	٨	٠,٨٦	٢
٠,٦٠	٩	٠,٧٥	٣
٠,٧٠	١٠	٠,٥٥	٤
٠,٦٦	١١	٠,٥٩	٥
		٠,٦٠	٦

بالنظر في قيم الجدول (١٢) يتبين أن جميع معاملات ارتباط بوينت بايسيربال بين درجات مفردات المجال الخامس و الدرجة الكلية للمجال تفاوتت قيمها بين (٠,٥٥ - ٠,٨٨)، وهي

معاملات ارتباط جيدة، وعليه يمكن القول إن جميع درجات مفردات المجال الخامس تتمتع بصدق وصفي جيد وفقاً لمؤشر صدق الاتساق الداخلي.

السؤال الثالث: هل تتمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الصدق الوظيفي؟

لأن محور اهتمام هذا النوع من الصدق يتركز على مدى صلاحية الاختبار من الناحية الوظيفية أو العملية، فإجراءاته تركز على مدى تقديم الاختبار أو تحقيقه لغرض معين، لذلك يعتمد بالدرجة الأولى على معامل الارتباط بوصفه مؤشراً إحصائياً لتقدير درجة صدق مفردات الاختبار، أي أن الارتباط بين نتائج الاختبار الذي نريد معرفة مدى صدق درجات مفرداته ونتائج محك آخر (عودة، ٢٠٠٥: ٤٦٥). فقد تم التحقق من الصدق الوظيفي لدرجات الاختبار، بإيجاد معامل ارتباط بيرسون بين درجات الطلبة في الاختبار التحصيلي محكي المرجع الذي تم بناؤه ودرجاتهم في اختبار نهاية الفصل الدراسي الأول لمقرر القياس والتقويم للعام (٢٠١٨- ٢٠١٩م). والجدول الآتي يوضح نتائج الصدق الوظيفي لدرجات الاختبار:

جدول (١٣) نتائج الارتباط بين درجات الاختبار ودرجات المحك

معامل الارتباط بيرسون	مستوى الدلالة	معامل التحديد
r	Sig	R ²
0.87**	0.000	0.76

يتضح من الجدول (١٣) أن قيمة معامل ارتباط بيرسون (r) بلغت (٠,٨٧)، وهي قيمة دالة إحصائياً عند مستوى دلالة (٠,٠٥)؛ ما يدل على وجود علاقة ارتباط موجبة بين درجات الطلبة في الاختبار محكي المرجع الذي تم بناؤه ودرجات المحك، وعليه يمكن القول إن درجات مفردات الاختبار تتمتع بصدق وظيفي جيد بدلالة المحك.

السؤال الرابع: هل تتمتع درجات الاختبار التحصيلي محكي المرجع- الذي تم بناؤه- لقياس مهارات القياس والتقويم لدى طلبة كلية التربية- جامعة إب- بخاصية الثبات وفقاً لمعامل (ليفنجستون)؟

تم حساب قيمة معامل ثبات ليفنجستون من خلال حساب معامل ألفا كرونباك ومن ثم تم حساب قيم معامل ليفنجستون للاختبار الكلي ومجالاته الفرعية، كما هو موضح في الجدول الآتي.

جدول (١٤) معامل الثبات لكل مجال (مهارة) وللإختبار بصورة كلية

معامل	معامل	المهارة
ليفنجستون	ألفا كرونباك	
٠,٨٣	٠,٦٧	المهارة الأولى: التعريف بالمفاهيم والمصطلحات الأساسية بالقياس والتقويم.
٠,٨٨	٠,٧١	المهارة الثانية: الإلمام بمفهوم الصدق وطرائق التحقق منه والعوامل المؤثرة فيه.
٠,٨١	٠,٦٩	المهارة الثالثة: الإلمام بمفهوم الثبات وطرائق التحقق منه والعوامل المؤثرة فيه.
٠,٨٥	٠,٦٨	المهارة الرابعة: بناء الإختبارات التحصيلية بأنواعها المختلفة.
٠,٨٩	٠,٧٤	المهارة الخامسة: إعداد المقاييس النفسية والتربوية وبنائها.
٠,٩٢	٠,٧٢	الإختبار الكلي

يتضح من القيم الموضحة بالجدول (١٤) أن درجات الإختبار تمتاز بمؤشرات ثبات عالية سوى كان ذلك في مجالات الإختبار او في الدرجة الكلية للإختبار ؛ حيث بلغ معامل ثبات الدرجة الكلية للإختبار (٠,٧٢) بمعامل ألفا كرونباخ وبلغت (٠,٩٢) بمعامل ليفنجستون، وهو دليل على ثبات جيد لدرجات الطلبة في النطاق السلوكي الذي يقسه الإختبار، أما مجالات الإختبار فتراوحت معاملات ثباتها وفقاً لمعامل ليفنجستون بين (٠,٨١ - ٠,٨٩) وهي معاملات ثبات جيدة.

السؤال الخامس: ما درجة القطع المناسبة التي من خلالها يتم تصنيف الطلبة إلى متقنين وغير متقنين لمهارات القياس والتقويم التي ينبغي توفرها لديهم ؟
حددت درجة القطع للإختبار وفقاً لطريقة أنجوف المعتمدة على المحكمين وتم تحديد عدد الأسئلة المطلوب الإجابة عنها في كل مجال وفي الإختبار الكلي ودرجة النجاح لكل مجال وللدرجة الكلية كما هي في الجدول الآتي:

جدول (١٥) درجة القطع لكل مجال (مهارة) وللاختبار الكلي

درجة التمكن وفقاً لدرجة القطع	عدد الأسئلة المطلوب إجابتها	درجة القطع	المهارة
١٦ من ٢١	١٦	٠,٧٥	المهارة الأولى: التعريف بالمفاهيم والمصطلحات الأساسية بالقياس والتقويم.
٩ من ١٢	٩	٠,٧٥	المهارة الثانية: الإلمام بمفهوم الصدق وطرائق التحقق منه والعوامل المؤثرة فيه.
١٠ من ١٣	١٠	٠,٧٥	المهارة الثالثة: الإلمام بمفهوم الثبات وطرائق التحقق منه والعوامل المؤثرة فيه.
١٩ من ٢٥	١٩	٠,٧٦	المهارة الرابعة: بناء الاختبارات التحصيلية بأنواعها المختلفة.
٨ من ١١	٨	٠,٧٢	المهارة الخامسة: إعداد المقاييس النفسية والتربوية وبنائها.
٦٢ من ٨٢	٦٢	٠,٧٥	الاختبار الكلي

يتضح من الجدول (١٥) أن درجة القطع للاختبار الكلي كما يرى المحكمون هي (٠,٧٥)، أي أن على الطالب أن يجيب عن (٦٢) سؤالاً من أصل (٨٢) سؤالاً اختبارياً، أي أن يحصل على (٦٢) درجة من أصل (٨٢) درجة؛ إذ لكل سؤال درجة واحدة حتى يتم تصنيفه ضمن فئة المتمكنين من مهارات القياس والتقويم التي يقسها الاختبار، وكانت درجة القطع في المهارة الأولى (٠,٧٥)، أي أن على الطالب أن يجيب عن (١٦) سؤالاً من أصل (٢١) سؤالاً اختبارياً حتى يتم تصنيفه ضمن فئة المتمكنين من المهارة الأولى، وهكذا في بقية المهارات.

السؤال السادس: ما مستوى تمكن طلبة كلية التربية- جامعة إب- من مهارات القياس والتقويم التي تضمنها الاختبار الذي تم بناؤه؟

تم استخدام الاختبار (t) لعينة واحدة one sample t test لكل للتعرف على مستوى تمكن طلبة كلية التربية من مهارات القياس والتقويم وكانت النتائج كما هو في الجدول التالي:

جدول (١٦) الاختبار (t) لعينة واحدة لكل مهارة وللإختبار الكلي

المهارة	عدد العينة	المتوسط الفرضي (درجة القطع)	المتوسط الواقعي	الانحراف المعياري	درجة الحرية	قيمة (t)	مستوى الدلالة	القرار
المهارة الأولى	١١٨	١٦	١٤,٥٩	٢,٨٢	١١٧	٤,٤٤	٠,٠٠٠	منخفض
المهارة الثانية	١١٨	٩	٦,٦٩	٢,٦٥	١١٧	٩,٤٤	٠,٠٠٠	منخفض
المهارة الثالثة	١١٨	١٠	٧,٢٤	٢,٦٦	١١٧	١٠,١٩	٠,٠٠٠	منخفض
المهارة الرابعة	١١٨	١٩	١٥,٢٢	٣,٥٨	١١٧	١٠,٧٠	٠,٠٠٠	منخفض
المهارة الخامسة	١١٨	٨	٦,٥٥	٢,٨٣	١١٧	٧,٥٥	٠,٠٠٠	منخفض
الدرجة الكلية	١١٨	٦٢	٥٠,٣١	١١,٥٤	١١٧	١٠,٥٣	٠,٠٠٠	منخفض

من الجدول (١٦) يتضح أن كل قيم (t) للمهارات والاختبار الكلي دالة إحصائياً؛ لأن مستوى الدلالة المقترن بها أقل من (٠,٠٥)؛ ما يدل على أن هناك فروقاً بين المتوسطات الفرضية والواقعية لصالح المتوسطات الفرضية لجميع المهارات والاختبار الكلي، وبذلك يمكن القول إن هناك تدنياً واضحاً في إتقان الطلبة لمهارات القياس والتقويم، كذلك تدني درجاتهم للاختبار بشكل عام، وهذه النتيجة تتفق مع ما توصلت إليه عدد من الدراسات السابقة كدراسة (الثبيتي، ٢٠١٤)؛ (الحسن، ٢٠١٦)؛ (سرفيوه، ٢٠١٦)، التي أشارت إلى وجود تدني واضح في مستوى إتقان عيناتها لمهارات القياس والتقويم بشكل عام ومهارات بناء الاختبارات التحصيلية بشكل خاص.

الاستنتاجات Conclusions:

- من خلال النتائج التي توصلت إليها الدراسة يمكن تقديم الاستنتاجات الآتية:
- تكون الاختبار التحصيلي محكي المرجع في مهارات القياس والتقويم بصورته النهائية من (٨٢) مفردة من نوع الاختيار من متعدد بأربعة بدائل.
- تمتع الاختبار التحصيلي محكي المرجع في مهارات القياس والتقويم بدلائل متعددة من الصدق الوصفي منها:
- تمتع الاختبار بصدق ظاهري جيد وفقاً لآراء الخبراء والمحكمين.
- تمتع الاختبار بصدق وصفي جيد وفقاً لدلالة الاتساق الداخلي بدليل أن ارتباط درجات المجالات (المهارات) مع الدرجة الكلية للاختبار، وارتباط درجات المفردات مع الدرجة الكلية للمجالات كانت دالة إحصائياً.

- تمتع الاختبار بصدق وصفي جيد وفقاً لدلالة المقارنة الطرفية (الصدق التمييزي) بدليل أن هناك فروقاً ذا دلالة إحصائية بين المجموعتين العليا والدنيا.
- تمتع الاختبار التحصيلي محكي المرجع في مهارات القياس والتقويم بصدق وظيفي جيد وفقاً لدلالة معامل ارتباط بيرسون بين درجات الاختبار المبني ودرجات المحك.
- تمتع مفردات الاختبار التحصيلي محكي المرجع في مهارات القياس والتقويم بدرجة عالية من الفاعلية، دلت عليها التقديرات الخاصة بمؤشرات معاملات (الصعوبة، التمييز) التي كانت ضمن المعايير المقبولة التي أوردتها كتب القياس والتقويم.
- تمتع الاختبار التحصيلي محكي المرجع في مهارات القياس والتقويم بدرجة عالية من الثبات، دلت عليها معاملات الثبات المستخرجة بمعامل ثبات ألفا كرونباخ وليفنجستون.
- بلغت درجة القطع (٠,٧٥) وفقاً لطريقة أنجوف.
- يوجد تدني في مستوى تمكن طلبة كلية التربية- جامعة إب من مهارات القياس والتقويم.
- يوجد احتياج في الجامعات اليمنية لكوادر متخصصة في القياس والتقويم كونها تعاني من نقص شديد في الكوادر المؤهلة والمتخصصة في القياس والتقويم .
- يبدو أن هناك عدم اتساق بين خطط الأقسام النفسية في الجامعات اليمنية لتأهيل كوادرها وبين ما تحتاجه من تلك الكوادر والتخصصات.

التوصيات Recommendations :

توصي الدراسة الحالية بالآتي:

- استخدام الاختبار الحالي في الجامعات اليمنية بوصفه أداةً تشخيصية يتم من خلاله تحديد مستوى تمكن طلبة كليات التربية من مهارات القياس والتقويم، نظراً ل تمتع الاختبار المبني بدلالات صدق وثبات جيدة.
- ينبغي على الجامعات اليمنية تبني إقامة برامج تدريبية لطلبة كليات التربية بمختلف المراحل التعليمية (بكالوريوس، ماجستير، دكتوراه) لتنمية مهاراتهم بالقياس والتقويم.
- ضرورة إنشاء مراكز ووحدات قياس في الجامعات اليمنية يكون من ضمن مهامها وأهدافها تزويد المعلمين والمتعلمين بمعلومات عن أهمية القياس والتقويم في جميع مجالات العملية التعليمية والتربوية، وذلك من خلال إقامة محاضرات وورش عمل وندوات ومؤتمرات علمية تدعم ذلك.
- يفضل تعاون كل من الجامعات ووزارة التربية والتعليم ومراكز البحوث والتطوير بإقامة دورات تدريبية لتأهيل المعلمين والمعلمات في أثناء الخدمة في مجال القياس والتقويم، والتركيز على أبعاد بناء الاختبارات التحصيلية .

المقترحات Suggestions:

استكمالاً وتطويراً لأهداف الدراسة تم اقتراح إجراء مجموعة من الدراسات الآتية:

- بناء اختبار محكي المرجع لمهارات القياس والتقويم وكفايات الإحصاء التربوي لطلبة كلية التربية في الجامعات اليمنية باستخدام النظرية الحديثة في القياس (نظرية الاستجابة للمفردة).
- إجراء بحوث مماثلة لبناء اختبارات محكية المرجع لقياس مهارات المقررات الجامعية الأخرى لدى طلبة الجامعة، وتصميم بنوك أسئلة للمقررات المختلفة ذات مواصفات جيدة.
- إجراء دراسة للتعرف على اتجاهات طلبة كلية التربية نحو مقررات القياس والتقويم.

المراجع

- أبو حطب، فؤاد (١٩٨٧): التقويم النفسي، ط٣، مكتبة الأنجلو المصرية، القاهرة- مصر.
- أبو علام، رجاء محمود (٢٠٠٥): تقويم التعليم، ط١، دار المسيرة للنشر والتوزيع، عمان- الأردن.
- _____، (٢٠١١): مناهج البحث في العلوم النفسية والتربوية، ط٦، دار النشر للجامعات، القاهرة- مصر.
- البناء، مأمون، (٢٠١٦): المهارات الإحصائية للباحث التربوي، ط١، دار وائل للنشر، عمان.
- _____، (٢٠٢٠): أساسيات القياس والتقويم في التربية وعلم النفس، ط٢، مركز دمشق، إب، اليمن.
- الترتوري، محمد عوض (٢٠١٨): تحليل احتياج طلبة السنة الأولى المشتركة للمهارات الجامعية من وجهة نظر القيادات الأكاديمية بجامعة الملك سعود، المجلة الدولية التربوية المتخصصة، المجلد (٧)، العدد (٧)، ٧٦.
- الثبيني، سالم عبدالله (٢٠١٤): بناء اختبار تشخيصي مرجعي المحك لقياس مهارات المعلمين في بناء الاختبارات التحصيلية لمقرر الحاسب الآلي، رسالة ماجستير، قسم علم النفس، كلية التربية، جامعة أم القرى، مكة المكرمة- السعودية.
- الحسن، أسماء عدنان (٢٠١٦): بناء اختبار محكي المرجع لقياس درجة إتقان المعلمين لقواعد بناء الاختبارات التحصيلية، رسالة ماجستير، قسم القياس والتقويم التربوي والنفسي، كلية التربية، جامعة دمشق، دمشق، سوريا.
- سرفقوة، عبد السلام عوض (٢٠١٦): بناء اختبار تشخيصي محكي المرجع في موضوعات من مقرر القياس والتقويم في جامعة عمر المختار، كلية التربية، البيضاء، المجلة الليبية العالمية، العدد (٩).
- الشرفين، نضال كمال (٢٠٠٦): الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي، مجلة العلوم التربوية والنفسية، المجلد (٧)، العدد (٤)، ٨٠- ١٠٩.
- الشعري، أماني (٢٠١٩): اتجاه طلبة كلية التربية جامعة إب نحو مقرر القياس والتقويم التربوي، بحث بكالوريوس، قسم العلوم النفسية والتربوية، كلية التربية، جامعة إب.
- العجيلي، صباح حسين (٢٠١١): مدخل إلى القياس والتقويم التربوي والنفسي، ط٥، مركز التربية للطباعة والنشر، كلية التربية، صنعاء- اليمن.
- علي، نداء بهاء الدين (٢٠١٢): فاعلية استخدام نموذج راش في بناء اختبار تحصيلي محكي المرجع لمقرر القياس والتقويم في التربية، رسالة دكتوراه، كلية التربية، جامعة دمشق- سوريا.

- علام، صلاح الدين (٢٠٠٠): تحليل البيانات في البحوث النفسية والتربوية والاجتماعية، ط١، القاهرة، دار الفكر العربي.
- _____، (٢٠٠١): الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية، ط٢، دار الفكر العربي، القاهرة- مصر.
- _____، (٢٠٠٦): الاختبارات والمقاييس التربوية والنفسية، ط١، المطابع المركزية، عمان.
- عودة، أحمد (٢٠٠٥): القياس والتقويم في العملية التدريسية، ط٣، دار الأمل للنشر، إربد- الأردن.
- مخائيل، امطانيوس نايف (٢٠١٥): القياس والتقويم النفسي والتربوي للأسوياء وذوي الحاجات الخاصة، ط١، دار الاصدار العلمي للنشر والتوزيع، عمان- الأردن.
- مراد، صلاح أحمد؛ سليمان، أمين علي (٢٠٠٢): الاختبارات والمقاييس في العلوم النفسية والتربوية، دار الكتاب الحديث، القاهرة- مصر.
- ملحم، سامي محمد (٢٠٠٢): مناهج البحث في التربية وعلم النفس، ط٢، دار المسيرة، عمان- الأردن.
- _____، (٢٠٠٩): القياس والتقويم في التربية وعلم النفس، ط٤، دار المسيرة، عمان- الأردن.
- نور الدين، أمين صبري (٢٠١٧): فعالية اختبار تحصيلي توافقي محوسب في تقويم طلاب جامعة الجوف، ندوة: التقويم في التعليم الجامعي " مرتكزات وتطلعات"، كلية التربية، جامعة الجوف، المملكة العربية السعودية.
- Hambelton, R.K. et al., (1978): Referenced Testing and Measurement, A Review of Technical Issue and Development Review of Educational Research, 48, 4.467-510 .
- Kubiszyn, T. & Borich, G., (2000): Educational Testing and Measurement, 6th ed., N.Y.: John Wiley & Sons, Inc.
- Popham, W. James,(1978): Criterion-Referenced Measurement, Englewood Cliffs, N.Y.: Prentice-Hall, Inc.
- Qasem, M., Almoshigah, T., & Gupta, S. (2014). The effect of number of alternatives (Two/Three/ Five) on Alpha Cronbach reliability coefficient and also on the validity in Likert scale. International journal of innovative research & studies,13 (6), 324 – 333 .

- Qasem, M., Altrairy, A., & Govil, P. (2012). Constructing Criterion Referenced Test to measure the statistical competencies of the Post-graduate Students in Education Colleges in Yemeni Universities. Retrieved on 19/9/ 2013 from <http://eric.ed.com> .
- Rajasekar, S., Philominathanet, P., & Chinnathambi, V. (2013). Research methodology. Retrieved on 11/11/ 2013 from www.physics.qc.edu .
- Ritter, N. (2010). Understanding a widely misunderstood statistic: Cronbach's α . Paper presented at the annual meeting of the Southwest, Educational Research Association, New Orleans, 2(8), 77-81.
- Thanasegaran, G. (2005). Reliability and validity issues in research. Malaysia: Integration & dissemination .
- Tighezza, A. (2011). The logical structure of Cronbach coefficient Alpha and its implications. KSA: Published by King Saud University .
- Zaza, H. (2011). The degree of agreement validity and reliability indices of standardized tests used in the theses presented in the faculties of education in Jordanian Universities, with indices of the original versions. Journal of Educational Sciences, 38 (2), 2400-2416.