طريقة تقدير الدرجات التقليدية وأثرها على الخصائص السيكومترية لاختبار اختيار من متعدد يتطلب بديلين صحيحة

الكلمات المفتاحية : تقدير الدرجات ؛ الخصائص السيكومترية ؛ اختيار من متعدد

د. محمد بن راشد عبدالكريم الزهراني

أستاذ مشارك بقسم التربية وعلم النفس – كلية التربية جامعة الباحة

Mohamm2315@gmail.com

الملخص

هدفت هذه الدراسة إلى التعرف على طريقة تقدير الدرجات التقليدية وأثرها على الخصائص السيكومترية لاختبار اختيار من متعدد يتطلب بديلين صحيحة ، ولتحقيق غرض الدراسة تم إعداد اختبار من نوع اختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات للصف الثالث المتوسط في وحدة (كثيرات الحدود)، وتم تطبيق الدراسة على عينة مكونة من (١٣٢) طالباً من طلبة الصف الثالث المتوسط للعام الدراسي ١٤٤٠/ ١٤٤١هـ بتعليم محافظة القنفذة ، وأظهرت نتائج الدراسة عن وجود فروق دالة إحصائياً عند مستوى (٠.٠١) في معامل ارتباط الفقرة بالدرجة الكلية لاختبار اختيار من متعدد يتطلب بديلين صحيحة لصالح طريقة المعرفة التامة في تقدير الدرجات ، ووجود فروق دالة إحصائياً عند مستوى (٠.٠١) بين متوسطي معاملات الصعوبة والتمييز لاختبار اختيار من متعدد يتطلب بديلين صحيحة لصالح طريقة المعرفة التامة في تقدير الدرجات ، وعدم وجود فروق دالة إحصائياً في معامل الثبات للاختبار تعزى إلى طريقة تقدير الدرجات التقليدية.

المقدمة

يهتم القائمون ببناء الاختبارات والمقاييس بكتابة وانتقاء مفردات عالية الجودة تحقق الشروط الأساسية في تكوين هذه الاختبارات والمقاييس لقياس السمات الإنسانية قياساً دقيقاً، كما يجب مراعاة كثيراً من المفردات وصياغتها والتحقق بالأساليب المنطقية وأحكام الخبراء من صدق محتوى كل مفردة على حده وكذلك صدق محتوى الاختبار ككل، غير أنه مهما بلغت دقة هذه الأساليب والأحكام فإنها لن تغني عن التجريب الميداني للاختبار وتحليل درجات مفرداته باستخدام الأساليب الإحصائية وتحديد العلاقة بين ما تقيسه المفردات وبين استجابات الأفراد لها (علام ، ٢٠١١،٢٦٧).

وتصنف أدوات القياس والتقويم حسب طريقة الإجابة ، فمنها ما يتطلب من الطالب انتقاء إجابة واحدة وهي الاختبارات التي أشكال فقراتها من نوع المطابقة والصواب والخطأ والاختيار من متعدد حيث يقدم للطالب ثلاث بدائل أو أكثر وعليه أن يختار الإجابة الصحيحة أو الأصح ، وفقرات يقوم فيها الطالب بصياغة إجابة وتتضمن ذات الاجابة محددة مثل التكميل والإجابة القصيرة (عودة ،٢٠١٠، ٦٢).

أن اختبارات الاختيار من متعدد هي أكثر انواع الاختبارات الموضوعية انتشارا وأهمية ، نظراً لتعدد إمكانية صياغة مفرداتها بكل دقة ، فضلا عن إمكانيات قياسها لمستويات متعددة من المجال المعرفي وفق مراحله المعروفة حسب تصنيفاته، وهي مستويات يصعب قياسها بالأنواع الأخرى من الاختبارات (الجبة ، ٢٠١٢ ، ١٤٤).

وعند تطبيق اسئلة الاختيار من متعدد على الطلاب ظهرت بعض الأخطاء في القياس ، ومن هذه الأخطاء خطأ التخمين عندما لا يعرف الطالب الإجابة الصحيحة من بين البدائل المعطاه ؛ فيقوم بالتخمين وقد يحصل على درجة جيدة رغم انها لا تمثل قدرته الحقيقية التي يهدف الاختبار إلى قياسها، ومما يترتب عليه أن النتائج قد تكون مضلله في بعض الاحيان ولا تقيس الهدف المراد قياسه بدقة وخاصة عندما يكون الهدف من الاختبار التشخيص أو الاختيار أو التنبؤ (مجيد ، ٢٠١٤).

لذا فإنه يصعب على أسئلة الاختيار من متعدد التي تتضمن إجابة صحيحة واحدة قياس نواتج العملية التعليمية المستهدفة،فقد يكون قياس الهدف ناقصاً إذا تم استخدام إجابة واحدة صحيحة. مما يضطر واضعي الاختبار غلى تكرار السؤال مرة ثانية مع وضع المفردة الصحيحة الأخرى لقياس هدف واحد ، وهذا من شأنه أن يقلل فعالية الاختبار لتكرار نفس الإجابة لنفس السؤال لذا ظهرت الحاجة إلى نوع جديد من اختبارات الاختيار من متعدد يتطلب اختيار أكثر من بديل صحيح مع الحفاظ على الخصائص السيكومترية لمثل هذا النوع من الاختبارات (الجبة ، ٢٠١٢).

ويعد تقدير الدرجات من الأساليب الهامة للقياس والتقويم التربوي في تطوير العملية التعليمية وذلك من خلال التعرف على مستوى أدائها وبما تحققه من أهداف ؛ لذا فإن تقدير الدرجات يحظى باهتمام كبير من قبل الباحثين والمهتمين لما له من أهميه في مجال تطوير نظم وأساليب التقويم والاختبارات.

و تعد طرق تقدير الدرجات من الطرق التي يعتمد عليها في تصحيح مفردات أو فقرات الاختبار،وقد تختلف بناء على صعوبة مفردات الاختبار ، ومن اكثر هذه الطرق استخداما الطريقة التقليدية التي تقوم على إعطاء المصحح درجة واحدة فقط عن كل إجابة صحيحة ودرجة صفر في حالة الإجابة الخاطئة (عبدالوهاب ، ٢٠٠١).

ولذا فإن اختبارات الاختيار من متعدد تقوم في الغالب على فكرة إجابة صحيحة واحدة وباقي البدائل خاطئة مما يوحي في كثير من الأحيان بالإجابة الصحيحة للطالب وبالتالي تكون درجة التخمين عالية في حين أن فكرة اختبار الاختيار من متعدد التي تتطلب بديلين صحيحة تعزز من عملية التعلم وتقلل من نسبة التخمين وتتيح للطالب فرصة أعمق في الإجابة لعدم وجود مشتتات كثيرة وبالتالي فإن الباحث رأى القيام بهذه الدراسة بهدف استخدام الفقرات التي تتطلب بديلين صحيحة في اختبارات الاختيار من متعدد بعد معرفة أثر الطريقة التقليدية لتقدير الدرجات على الخصائص السيكومترية لهذا النوع من الاختبارات.

مشكلة الدراسة :

ينشغل المختصون في القياس والتقويم بالسعي الجاد نحو تحقيق الدقة في القياسات التربوية والنفسية التي تعتمد على أدوات قياس يتحرى الدقة في بنائها وتطبيقها وتجميع بياناتها وتفسير نتائجها سعيا نحو تحقيق شروط الاختبار الجيد من صدق وثبات وموضوعية ومعايير من اجل تحقيق جودة في البيانات والنتائج والتفسير والاستنتاجات التي تُبنى على أساس إجابات الطلاب عن فقرات الاختبار.

ومن أهم هذه الاختبارات اختبار الاختيار من متعدد يتطلب بديلين صحيحة ، إلا أن تلك الاختبارات تعاني من عدم اتفاق على طريقة تصحيحها ، إذ يمكن تصحيحها بعدة طرق ربما يكون لها آثار مختلفة على الخصائص السيكومترية لفقرات الاختبار.

وتشير نتائج الدراسات السابقة إلى ارتباط الخصائص السيكومترية للاختبار الذي يتطلب بديلين صحيحة بطريقة تقدير الدرجات كدراسة Ketterlin-Geller, L. R.; Perry, T., & Platas, P. M. (2019) ، ودراسة اللحياني (٢٠١٢) ، ودراسة الجبة (٢٠١٢)، ودراسة صمصم(٢٠١٢) التي تباينت في نتائجها وفقاً لمعاملات الصعوبة والتمييز والصدق والثبات.

و أن هناك تباين في نتائج الدراسات حول معامل الصدق للاختبار الذي يتطلب بديلين صحيحة حيث كانت بمستويات جيدة كدراسة Abreu-Mendoza , R. A.; Chamorro,

Breuer, S.; Scherndl, T., & Y., & Matute, T. (2019) بينما عكس دراسة Ortner, T. M. (2020) التي أظهرت في نتائجها تدني مستويات الصدق.

نلاحظ من خلال الدراسات السابقة تضارب النتائج حول الخصائص السيكومترية لاختبار الاختيار من متعدد الذي يتطلب بديلين صحيحة بالإضافة إلى التباين في معاملات الصعوبة والتمييز.

و أن الباحث من خلال تجاربه مع الطلاب عند تناول اختبار الاختيار من متعدد الذي يتطلب إجابة صحيحة واحدة لم يلاحظ اهتمام الطلاب بفحص البدائل بشكل جيد بل انهم في كثير من الأحيان يسعون للخلاص من هذا النوع من الاسئلة بأي شكل وذلك بإعطاء أية إجابات وخاصة عندما تكون البدائل متشابه على حد ما ، وهذا بدوره يؤثر على الخصائص السيكومترية للاختبار مما دفع الباحث إلى تناول طريقة تقدير الدرجات التقليدية وأثرها على الخصائص السيكومترية لاختبار اختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات بوحدة (كثيرات الحدود) للصف (الثالث المتوسط) للعام الدراسي ١٤٤٠ / ١٤٤١هـ.

أسئلة الدراسة :

السؤال الأول : هل توجد فروق دالة إحصائياً بين قيم معامل الارتباط المصحح بالدرجة الكلية لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

السؤال الثاني : هل توجد فروق دالة إحصائياً بين متوسطي معاملات الصعوبة لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

السؤال الثالث : هل توجد فروق دالة إحصائياً بين متوسطي معاملات التمييز لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

السؤال الرابع : هل توجد فروق دالة إحصائياً بين معاملي ثبات ألفا كرونباخ لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

أهداف الدراسة :

١. التعرف عن قيم الخصائص السيكومترية لاختبار الاختيار من متعدد يتطلب بديلين صحيحة عند تصحيح الاختبار وفقراته وفق طرق مختلفة في تقدير الدرجات.
٢. الكشف على دلالة الفروق في معاملات (الصدق والصعوبة والتمييز والثبات) لاختبار الاختيار من متعدد يتطلب بديلين صحيحة والتي تعزى إلى طريقة تقدير الدرجات.

أهمية الدراسة :

١. قد تسهم هذه الدراسة في تطوير صياغة نمط اختبار الاختيار من متعدد الذي يتطلب بديلين صحيحة إذا كانت خصائصها السيكومترية أفضل من نمط اختبار الاختيار من متعدد الذي له إجابة صحيحة واحدة.
٢. قد تؤدي هذه الدراسة إلى لفت نظر المسئولين و المتخصصين إلى العناية والاهتمام بإعداد اختبارات الاختيار من متعدد الذي يتطلب بديلين صحيحة وطرق تقدير درجاتها.
٣. مساعدة المعلمين ومعدي الاختبارات على معرفة طرق تقدير الدرجات التي تؤدي إلى خصائص سيكومترية جيدة لاختبار الاختيار من متعدد الذي يتطلب بديلين صحيحة.
٤. قد تؤدي هذه الدراسة إلى إمكانية قيم الصدق والثبات لاختبار الاختيار من متعدد الذي يتطلب بديلين صحيحة.

مصطلحات الدراسة :

طريقة تقدير الدرجات التقليدية : تُعرف طريقة تقدير الدرجات التقليدية بأنها القاعدة التي يتم على أساسها إعطاء قيم عددية لاستجابة الطالب لأسئلة وعناصر الاختبار ، وهي طريقة تحويل المعلومات الخاصة إجابات الطلاب إلى أرقام تعكس مستوى أدائهم وفهمهم لعناصر الاختبار (Stankous, 2016, 69).

وتعرف إجرائياً : بأنها عبارة عن طريقتين المعرفة التامة والمعرفة الجزئية في اختبار الرياضيات من نوع الاختيار من متعدد الذي يتطلب بديلين صحيحة.

الخصائص السيكومترية : تعرف بأنها هي الخصائص المرتبطة بذات الاختبار ، والتي يمكن التعبير عنها بدلالات رقمية ، سواء تلك الخصائص المتعلقة بفقرات الاختبار (الصعوبة ،

التمييز ، التخمين) أو تلك الخصائص المتعلقة بالدرجة الكلية للاختبار (الثبات ، الصدق) (غيث،٢٠٠٧).

اختبار الرياضيات الذي يتطلب بديلين صحيحة : يُعرف هذا النوع من الاختبارات بأنه ذلك الاختبار للاختيار من متعدد الذي يتضمن أسئلة تحتمل بديلين أو خيارين صحيحين تقدم الحلول للمسائل الحسابية ، حيث تتضمن اختيارات الإجابة بديلين صحيحين والعديد من الإجابات المستبعدة التي تُعرف باسم "المشتتات" ، ويتسم هذا النوع من الاختبارات بسرعة التصحيح وحساب الدرجات بدقة وتجنب التوزيع العشوائي للدرجات (Thissen, Wainer, & Wang, 2017, 113) .

ويعرف إجرائياً : اختبار تحصيلي من نوع اختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات في وحدة (كثيرات الحدود) للصف (الثالث المتوسط) للعام الدراسي ١٤٤٠ / ١٤٤١هـ.

حدود الدراسة :

الحدود البشرية : طلبة الصف الثالث متوسط (بنين).

الحدود المكانية : تم تطبيق الدراسة في محافظة القنفذة بمكتب التعليم بالمظيلف.

الحدود الزمانية : تم تطبيق هذه الدراسة في الفصل الدراسي الثاني من العام الدراسي ١٤٤٠/١٤٤١هـ

الاطار النظري :

تتكون اختبارات الرياضيات التي تتطلب بديلين صحيحين من السؤال (الجذر) وبدائل الخيارات التي توجد منها بديلين صحيحين فضلاً عن بدائل أخرى غير صحيحة (مشتتات) وتتمثل معايير جودة هذا النوع من الاختبارات في استعمال اللغة السهلة المباشرة والأسئلة القصيرة وتجنب الإجابات المطلقة مثل "جميع الإجابات صحيحة " أو "لا شيء مما سبق " وتعتبر البدائل جزء أصيل من اختبارات الرياضيات التي تتطلب بديلين صحيحين ، فهي التي تعطي المصداقية للاختبار فضلاً عن معلومات هامة يمكن أن يسترشد بها الطالب في إجابته (Camilli, & Chiu, 2018, 52).

ويصف (Forsyth, & Spratt, 2016, 31) اختبارات الرياضيات التي تتطلب بديلين صحيحين بأنها اختبارات تتألف من عناصر يحصل فيها المستجيب على درجة (١) إذا أجاب

عن كل البدائل بطريقة صحيحة ، وعلى ثلثي الدرجة في حالة الاستجابة بطريقة صحيحة على ثلاث بدائل فقط ، وعلى ثلث درجة إذا استجاب لبديلين فقط بصورة صحيحة ، وعلى درجة صفر في حالة عدم الإجابة.

و يتم الاعتماد على اختبارات الرياضيات التي تتطلب بديلين صحيحين نظراً لسهولة تقدير درجاتها وثبات تلك الطريقة والقدرة على تغطية العديد من الموضوعات في مقررات الرياضيات ونتائجها التي تعكس نتائج الأداء المرتفع فيها من جانب الطلاب.

طرق تقدير الدرجات التقليدية لاختبارات الرياضيات التي تتطلب بديلين صحيحين:

تُعرف طريقة تقدير الدرجات التقليدية بأنها القاعدة التي يتم على أساسها إعطاء قيم عددية لاستجابة المفحوص على اسئلة وعناصر الاختبار ، وهي طريقة تحويل المعلومات الخاصة باستجابات المفحوصين إلى كميات وأرقام تعكس مستوى أدائهم وفهمهم لعناصر الاختبار (Stankous, 2016, 69).

أما بالنسبة لاختبارات الرياضيات التي تتطلب بدائل صحيحة، يتم تقدير الدرجات وفقاً لطريقتين هما : (١) يتم تقدير صحة إجابات المفحوصين على العناصر وفقاً لمجموعة من الأسس لتحديد مدى دقة وصحة البدائل.

(٢) يتم تجميع الإجابات الصحيحة تراكمياً بإتباع مجموعة أخرى من الأسس للوصول إلى درجة إجمالية للاختبار.

ويعبر مجموع الإجابات على عناصر الاختبار كدليل على أداء المفحوصين في الاختبار ، ولتقدير الدرجات ، يتم مقارنة البدائل بالإجابات الصحيحة على سبيل المثال : إذا اختار المفحوص الإجابة بالبدائل الصحيحة من بين مجموعة من الإجابات الخاطئة ، يتم اعتبار الإجابة صحيحة ويتم إعطاء عدد الدرجات الكاملة (Rust, 2019, 27).

وتعد عملية اختيار العناصر وإجراء تقدير الدرجات عملية معقدة ؛ حيث يجب أن تتوازى مع الغرض من الاختبار ودعم جدوى وتفسير النتائج. وتتأثر طريقة تقدير الدرجات بصيغة عناصر الاختبار (الاختيار من بين بدائل متعددة والأسئلة البنائية)،وبالنسبة لاختبارات الرياضيات التي تتطلب بديلين صحيحين ، يتعين على المفحوصين اختيار بديلين من بين مجموعة من الإجابات المحتملة ويتم تقدير الدرجات وفقاً لقاعدة توزيع تحدد القيمة للبدائل الصحيحة فقط يجب أن تكون إجراءات تقدير الدرجات في اختبارات الرياضيات التي تتطلب

بديلين صحيحين بطريقة تقدم معلومات مفيدة في تفسير الدرجات ، ويجب أن تبدأ طريقة تقدير الدرجات بتحديد الاسلوب المناسب للتقدير مثل الدرجات الخام أو الدرجات المبنية على نموذج معين مثل نظرية الاستجابة للمفردة Item Response Theory (IRT) (Briere, 2018, 46).

و أشار (Fehintola, & Akingbade, 2019, 13) إلى أن أفضل طرق تقدير الدرجات في اختبارات الرياضيات التي تتطلب بديلين صحيحين هي طريقة التقدير الجزئي للدرجات Partial Credit Scoring (PCS) والتي تسمح بقياس أفضل لإنجاز الطالب في الرياضيات ؛ حيث تمتاز طريقة التقدير الجزئي للدرجات بالقدرة على رصد مستوى معرفة الطالب بكل اختيار مقدم في عناصر الاختبار،و تتضمن طريقة التقدير الجزئي للدرجات ثلاثة أشكال رئيسية :

(أ) اختبار الاختيار من متعدد الحر الذي يسمح للطالب باختيار أكثر من إجابة على السؤال في حالة عدم التأكد من الإجابة الصحيحة.

(ب) اختبار الإقصاء ET حيث يتعين على الطالب تحديد الاختيارات غير الصحيحة ضمن الأسئلة ذات البديلين الصحيحين ، بمعنى أن الطالب يحذف الاختيارات الخاطئة.

(ج) قياس الثقة CW حيث يحدد الطالب اعتقاده حول البدائل الصحيحة ومدى ثقته في اختياره.

وقد ذكر (Skolnick, 2018, 123) أن استخدام معادلة التصحيح Formula Scoring (FS) يعتبر من بين الطرق الشائعة في تصحيح اختبارات الرياضيات التي تتطلب بديلين صحيحين ، وتعتمد هذه الطريقة على حساب عدد البدائل الصحيحة (العدد الأقصى بديلين) وإجابات البدائل الخاطئة والمحذوفات ، ويهدف هذا الأسلوب في التصحيح إلى خفض معدل التخمين وزيادة التركيز على الإجابة بالبدائل الصحيحة.

الخصائص السيكومترية لاختبارات الرياضيات التي تتطلب بديلين صحيحين:

يعتمد القياس النفسي (السيكومتري) على مبادئ رئيسية يتم من خلالها الحكم على جودة التقييم، ومن أهما : الصدق ، والثبات ، والصعوبة ، والتمييز.

(أ) **الصــدق** : يعبر مفهوم صدق الاختبار validity عن ملائمة وجدوى تطبيق أحد الاختبارات أو درجة قياس الاختبار للهدف الموضوع من أجله ، وبالتالي فإنه يعبر عن مدى

الاستنتاجات والنتائج والقرارات المتخذة على أساس ملائمة وكفاءة درجات الاختبار. أيضاً يوصف صدق الاختبار بأنه إجمالي التقييم التقديري لدرجة مساندة الأدلة التجريبية والأساس النظري لكفاءة وملائمة التفسيرات في ضوء درجات الاختبار أو غيرها من أنماط التقييم ، وتنطبق جميع مبادئ الصدق على جميع أنواع التقييمات التي تقدر الدرجات وفقاً للثبات الملحوظ (Cortina, 2019, 102).

ومن أهم أنواع صدق الاختبار صدق المحتوى Content Validity: الذي يشير إلى الدرجة التي يستطيع عندها الاختبار التقييم الشامل للمجال محل الاختبار (مثل الرياضيات)، وينقسم إلى نوعين هما : الصدق الظاهري والصدق المنطقي. يعبر مفهوم الصدق الظاهري Face Validity عن الدرجة التي يبدو عندها محتوى الاختبار قادراً على التقييم الشامل لجميع المجالات المعنية في المتغير الكامن. أما الصدق المنطقي Logical Validity فإنه عبارة عن صورة أكثر تعقيداً من الصدق الظاهري. لا توجد تقنيات إحصائية معينة لتقييم صدق المحتوى كما هو الحال بالنسبة للصدق المعياري أو البنائي ، بل فقط توجد مجموعة من الأسس التي تساعد في تطوير أداة قياس ذات صدق محتوى قوي ، ومن بين الطرق الأخرى للحصول على صدق محتوى قوي تحديد المجالات المختلفة التي يتألف منها الموضوع الذي يتم تقييمه عن طريق الاختبار (Albanese, & Sabers, 2018, 53).

(ب) الثبات : يشير مفهوم الثبات Reliability إلى استقرار ودقة الدرجات المتحصله من أحد الاختبارات إما خلال حالة تطبيق واحدة أو عدة تطبيقات على فترات زمنية معينة (Bland, Kreiter, &Gordon, 2019, 6).

ومن أهم انواع ثبات الاختبار الثبات الداخلي ، وثبات الإعادة وفيما يلي عرض موجز لهما :

- **الثبات الداخلي** Internal Consistency Reliability: يتم استخدام حساب الثبات الداخلي لتقدير مدى التجانس بين الأسئلة والعناصر داخل الاختبار أو المقياس (بمعنى أنها تقيس سمة واحدة مثل مهارات الرياضيات البسيطة) . بالتالي فإن الثبات الداخلي ملائم لتقييم الاختبارات المتجانسة من الناحية النظرية ، ومن الممكن تقييم الثبات الداخلي في أحد مواقف الاختبار ، لذا فإنه يتجنب الكثير من المشكلات المرتبطة بالاختبار المتكرر الموجودة في تقديرات الثبات الأخرى مثل ثبات الاختبار– إعادة الاختبار، وتعتبر أشهر أشكال الثبات الداخلي حساب ثبات التجزئة النصفية Split Half Reliability ويتم الحساب عادةً باستخدام

معامل ألفا كرونباخ، ويقتضي تقدير الثبات الداخلي باستخدام حساب التجزئة النصفية تقسيم الاختبار إلى نصفين في محاولة لجعلهما متكافئين حيث يجب أن تكون الدرجات المتحصله من كل نصف لها نفس الدرجة الصحيحة ومعامل الخطأ. فإذا كانت أنصاف الاختبارات متكافئة ، يمكن حساب الثبات الداخلي للاختبار كاملاً باستخدام معادلة سبيرمان- براون نظراُ لكفاءة المعادلة في تصحيح الثبات المنخفض في كل نصف من الاختبار. يمثل نصفي الاختبار أشكال بديلة لبعضهما البعض ، فعندما ينتج عن معامل ألفا كرونباخ قيمة مرتفعة ، يكون الثبات الداخلي للاختبار كاملاُ مرتفع (Blair, Lenton, & Hastie, 2016, 594).

- ثبات الاختبار- إعادة الاختبار Test-retest Reliability: يشير إلى عملية تقييم نفس الشخص مرتين باستخدام نفس الاختيار ثم الربط بين النتائج. ويعتبر معامل بيرسون أكثر الأساليب الإحصائية الشائعة في تقييم ثبات الاختبار- إعادة الاختبار ، ومن ثم فإن تقديرات الثبات تتراوح بين -١ إلى +١ فإذا حصل جميع المفحوصين على نفس الدرجات خلال التطبيقين ، يتم الحصول على مستويات ثبات مثالية وعلاقة موجبة بين الدرجات. تكمن أهمية حساب ثبات الاختبار في أن نتائجه تمثل تقديرات لنسبة تباين الدرجة الفعلية بالنسبة لتباين الدرجة الملحوظة أو مقدار التباين المشترك عبر تطبيقي الاختبار (Cicchetti, 2017, 69).

(ج) الصعوبة : معامل الصعوبة للفقرة تعني النسبة المئوية للإجابات عليها ، فإذا كانت النسبة صفر تشير إلى صعوبة الفقرة حيث أجاب جميع الطلاب على الفقرات بشكل خاطئ ، اما إذا كانت النسبة ١٠٠% فهذا يشير إلى سهولة الفقرة حيث جاءت إجابات جميع الطلاب صحيحة ، ويمكن حساب هذه النسبة من المعادلة التالية : (عدد الطلاب الذين اجابوا إجابة صحيحة للفقرة على العدد الكلي لجميع الطلاب الذين حاولوا الإجابة على الفقرة) (الغامدي ، ٢٠٠٨).

وتنحصر الاختبارات المقبولة بين مدى سهولة يتراوح بين (٠.٣٠ - ٠.٧٠) حيث تكون قادرة على التمييز بين الطلاب الحاصلين على درجات عليا والطلاب الحاصلين على درجات دنيا ، إلا أن خروج درجات الفقرات عن هذه النسبة لا تعني أنها غير مناسبة إحصائياً فقد يلجأ واضع الاختبارات إلى فقرات سهلة بهدف تشيع الطلاب (علام ، ٢٠٠٧).

(د) التمييز : ذكر علام (٢٠٠٧ ،٥٤) بأن معامل تمييز المفردة " يشير إلى درجة تمييز المفردة بين مرتفعي التحصيل ومنخفضي التحصيل من الطلاب في الاختبار".ويمكن حساب معامل التمييز للفقرة بطرح نسبة عدد الطلاب الذين أجابوا إجابة صحيحة عن الفقرة في الإرباعي الادنى (صد) من نسبة عدد الطلاب الذين أجابوا إجابة صحيحة عن الفقرة في الإرباعي الأعلى (صع)(فتح الله ، ٢٠٠٦).

العلاقة بين طريقة تقدير الدرجات والخصائص السيكومترية:

توجد علاقة قوية بين طريقة تصحيح الاختبارات وخصائصها السيكومترية حيث تعتبر الخصائص السيكومترية في مجال الرياضيات برهان قوي على الكفاءة والملائمة وقابلية الاستخدام لإجراءات التصحيح المختلفة. ويدرك الكثير من المعلمين والمتخصصين في مجال تقويم تعلم الرياضيات أن دقة الدرجات ودلالتها التفسيرية تعتمد بشكل مباشر على الخصائص الداخلية (السيكومترية) فالخصائص السيكومترية مثل الثبات أو ملائمة الاستخدام هي الأساس للحكم على جودة طريقة تصحيح الدرجات في اختبارات الرياضيات تحت الظروف المختلفة (Brombacher, 2015, 3).

تعد الخصائص السيكومترية من العناصر الجوهرية المحددة لجودة طريقة التصحيح في اختبارات الرياضيات. فهناك بعض الخصائص السيكومترية التي تقدم الدليل على جودة طريقة تصحيح الاختبار كاملاً، في حين توجد بعض الخصائص السيكومترية الأخرى التي تثبت كفاءة التصحيح لأجزاء معينة فقط أو حتى نوعية أسئلة بعينها. على سبيل المثال، يمكن أن تبرهن الخصائص السيكومترية على دقة حساب درجات عناصر أو أبعاد أو الاختبار كاملاً في مجال تقويم الرياضيات

(Brombacher, Bulat, King, Kochetkova, & Nordstrum, 2015, 27).

كذلك من بين الجوانب الأخرى الهامة للخصائص السيكومترية ذات التأثير على طريقة حساب الدرجات في اختبارات الرياضيات إمكانية حساب تلك الخصائص بطريقة كمية في صورة معاملات أو كميات رقمية تعبر عن مدى الصدق والثبات. على سبيل المثال، يمكن أن يكون حساب "الثبات" على شكل قيمة عددية وكذلك الحال بالنسبة للخصائص السيكومترية الأخرى (Dai, Wang, & Svetina, 2016, 29).

و يمكن تفسير أثر طريقة تقدير الدرجات على الخصائص السيكومترية لاختبارات الرياضيات في ضوء النظرية الكلاسيكية للقياس التي تمثل مفهوم نشط لتفسير الإجراءات الإحصائية لصدق وثبات الاختبارات. تشير النظرية الكلاسيكية إلى أن الدرجة الحقيقية التي يحصل عليها المفحوص في اختبار معين (اختبار للرياضيات) تعكس معرفة وأداء المفحوص في تلك المادة أو مجال الاختبار، ويجب أن تتشابه هذه الدرجة مع ما يحصل عليه المفحوص في الأداء المستقبلي على نفس الاختبار (بمعنى أن الاختبار يكون ذو ثبات اختبار/إعادة الاختبار) ؛ حيث ترتبط أهمية حساب الصدق والثبات (الخصائص السيكومترية) بتعميم درجات الاختبار بمعنى قابلية الأداء للتكرار خلال الاختبارات المستقبلية، وهو ما يرتبط بصورة مباشرة بنوعية إجراءات تقدير الدرجات (Haberman, 2018, 205).

الدراسات السابقة :

دراسة Anderson, H. R. (2020) التي هدفت إلى التعرف على الخصائص السيكومترية لاختبار في الرياضيات قائم على أكثر من بديل صحيح من خلال التعرف على صدق المحتوى والصدق البنائي ومعامل الثبات الداخلي وفحص خصائص عناصر الاختبار والتعرف على الصدق المعياري المحتمل للاختبار واستخدمت الدراسة المنهج الوصفي الكمي ، وتكونت عينة الدراسة من (٦٠٧) طالب وطالبة في الصف العاشر من التعليم الأساسي بأحد المدارس المتكاملة في ولاية إلينوي الأمريكية ، كما تم تطوير اختبار في الرياضيات قائم على الخيارات البديلة الصحيحة عبارة عن (٣٠) سؤال مدته (٤٥) دقيقة ،و يتكون من ثلاث أقسام : العمليات الحسابية والجبر والهندسة،وقد أظهرت نتائج الدراسة أن عناصر الاختبار المتعلقة بالجبر والهندسة كانت سهلة بصفة عامة مع أفراد العينة ، في حين كانت أسئلة العناصر الحسابية أكثر صعوبة ، كما أسفرت النتائج عن بقاء جميع تقديرات الثبات الداخلي للاختبار قوية وفي المدى المتوسط ما بين (٠.٧٣٦ إلى ٠.٩٥٠).

دراسة Breuer, S.; Scherndl, T., & Ortner, T. M. (2020) التي هدفت إلى فحص تأثير نمط الإجابة (الاختبارات التي تتطلب بديلين صحيحين مقابل اختيار محدد) على الخصائص السيكومترية لاختبار الرياضيات في ضوء عوامل الجنس وقلق الاختبار، وقد استخدمت الدراسة التصميمين الكمي والنوعي ، وتكونت عينة الدراسة من (٤٣٣) طالب وطالبة متوسط أعمارهم ١٤ عام في المانيا ، تم توزيعهم عشوائياً على مجموعتين (مجموعة

الطلاب الذين حصلوا على اختبار في الرياضيات يتكون من (٢٥) فقرة يتطلب الإجابة ببديلين صحيحين مقابل المجموعة الأخرى التي حصلت على نفس الاختبار ولكن بنظام الإجابة الصحيحة الواحدة ، وقد توصلت الدراسة إلى النتائج التالية :أظهر تحليل البيانات أن نسخة الاختبار التي تتطلب بديلين صحيحين للإجابة أظهر مستويات أعلى في الثبات لكن مستويات أدنى في الصدق المعياري بالمقارنة مع نسخة اختبار الرياضيات ذات الإجابة الصحيحة الواحدة المحددة ، كما لم توجد علاقة ذات دلالة إحصائية بين نمط الإجابة على اختبار الرياضيات بنوعيه وكل من الجنس أو قلق الاختبار ، كما أظهرت نتائج الدراسة إلى وجود علاقة ارتباطيه بين نمط الإجابة (الاختبار الرياضيات الذي يتطلب بديلين صحيحين) وبين مستويات الصدق والثبات للاختبار.

دراسة Abreu-Mendoza , R. A.; Chamorro, Y., & Matute, T. (2019) التي هدفت إلى توضيح الدرجات المعيارية وفحص الخصائص السيكومترية لاختبار في الرياضيات يتم تطبيقه على الطلاب في المكسيك ، وقد استخدمت الدراسة المنهج الوصفي الكمي للتعرف على الخصائص السيكومترية لاختبار الرياضيات الذي يتطلب الإجابة ببديلين صحيحين ، وكانت عينة الدراسة مكونة من (١٣١٨) طالب وطالبة بالسنة التاسعة من التعليم الأساسي ، وتمثلت الأدوات المستخدمة في جمع البيانات من : اختبار الرياضيات الذي يتطلب بديلين صحيحين (تم تقنينه على عينة من الطلاب في الولايات المتحدة الأمريكية يقدر عددهم بـ ٣٠٢١) ، والاختبار الحسابي الفرعي ضمن مقياس ويكسلر للذكاء (WISC) ، وقد أسفرت نتائج الدراسة عما يلي :

- ظهور مستويات جيدة من الصدق الداخلي لاختبار الرياضيات ومستويات خصائص سيكومترية مناسبة ، وهو ما يبرهن على فاعليته في تقييم الأداء في تعلم الرياضيات بين الطلاب.
- أسفر التحليل العاملي التفسيري عن وجود ثلاثة عوامل للاختبار وهي : الحساب والكسور والعمليات الجبرية الأساسية.
- ظهور فروق نوعية بين الذكور والإناث في عدد الإجابات الصحيحة بينما كان حجم التأثير صغيراً.

دراسة Ketterlin-Geller, L. R.; Perry, T., & Platas, P. M. (2019) التي هدفت إلى فحص مدى فاعلية طريقة تقدير الدرجات في اختبارات الرياضيات على كفاءة استخدام الاختبار في ضوء الخصائص السيكومترية من خلال المقارنة بين ثلاث طرق لتقدير الدرجات باستخدام اختبار للرياضيات يتطلب الإجابة ببديلين صحيحين ، وتم استخدام المنهج الوصفي للدراسة ، وقد تكونت عينة الدراسة من (٢٨٦٠) طالب وطالبة متوسط أعمارهم (٨.٩) سنوات وقع الاختيار عليهم بعد تجاوز اختبار تمهيدي يغطي المهارات والمعارف الأساسية في الرياضيات. تم تطبيق اختبار تقييم الرياضيات EGMA الذي يتكون من (٨) مجموعات من الأسئلة التي تتطلب بديلين صحيحين للإجابة مع أفراد العينة مع تطبيق ثلاث طرق لتقييم الدرجات (تقييم الدرجة الكلية وتقييم الدرجات الفرعية وطريقة الدرجة المجمعة) من جانب (٥٦) مصحح للتعرف على فاعليتها في استخدام الاختبار مع الطلاب ، وقد أظهرت نتائج الدراسة ارتباط الخصائص السيكومترية للاختبار (حساب الصدق والثبات وتفرد النتائج) بفاعلية الإجراء المستخدم في تقدير الدرجات على مستوى الدرجات الكلية والفرعية والدرجة المجمعة ، كما توصلت الدراسة إلى وجود علاقة ذات دلالة إحصائية بين طريقة تقييم الدرجة المجمعة لاختبار الرياضيات الذي يتطلب الإجابة ببديلين وبين الكفاءة السيكومترية وهو ما انعكس على كفاءة استخدام الاختبار.

دراسة الثمالي (٢٠١٦) التي هدفت إلى التعرف على أثر اختلاف عدد البدائل على الخصائص السيكومترية لاختبار اختيار من متعدد التي يتضمن إجابات مركبة تحتوي على أكثر من إجابة صحيحة، والكشف عن العدد الأمثل لعدد البدائل في فقرات الاختيار من متعدد التي تتضمن إجابات مركبة تحتوي على أكثر من إجابة صحيحة ، وتكونت عينة الدراسة من (٢٥٠) طالباً من طلاب المرحلة الثانوية بالطائف ، ويتألف الاختبار من (٣٠) فقرة من نوع الاختيار من متعدد ذو استجابات مركبة يتضمن اجابتين صحيحة ، مكون من ثلاث صور تتفق في جذر السؤال وتختلف في عدد البدائل فكانت على ثلاث صور ، صورة ذات خمسة بدائل ويحذف بديل لتصبح أربعة بدائل ويحذف بديل لتصبح ثلاثة بدائل ، وتم تطبيق الاختبار على نفس العينة بحيث يجيب كل طالب على النماذج الثلاثة ولكن بفاصل زمني مدته أسبوعين بين كل نموذج والآخر ، وقد أشارت نتائج الدراسة إلى عدم وجود فروق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة والتمييز بين صور الاختبار ذات

الثلاثة البدائل والأربع بدائل والخمس بدائل ، بينما توجد فروق ذات دلالة إحصائية بين متوسطات صدق وثبات الفقرات بين صور الاختبار الثلاث لصالح الاختبار ذو الخمسة بدائل.

دراسة الجبة (٢٠١٢) التي هدفت إلى تحديد أوجه الاختلاف والاتفاق التي تنطوي عليها النظرية التقليدية في تناول الخصائص السيكومترية لمفردات اختبار اختيار من متعدد تتضمن أكثر من إجابة صحيحة ، وتكونت عينة البحث من (٥٣) طالباً من طلاب التربية الخاصة بكلية التربية بجامعة الملك عبدالعزيز ، وعلى ضوء أداء الطلاب في اختبار تحصيلي موضوعي في مقرر القياس والتقويم ، أظهرت النتائج أن المفردات التي تتضمن أكثر من إجابة صحيحة أعلى في الصعوبة والتمييز والثبات من المفردات التي تتضمن إجابة صحيحة واحدة ، وفقاً للمعايير المحددة في البحث لقبول المفردة ، إلا أن متوسطات معاملات الصعوبة والتمييز والثبات مجتمعة تقع في النطاق المسموح به لقبول المفردة ، كما توصل البحث إلى أن الزيادات الطفيفة في صعوبة وتمييز وثبات المفردات التي تتضمن أكثر من إجابة صحيحة ليس كافياً لتبرير تجنب استخدامها في الاختبارات الموضوعية.

أما دراسة صمصم (٢٠١٢) التي هدفت إلى مقارنة أثر ثلاث طرق لتصحيح اختبارات الاختيار من متعدد لضبط التخمين على مؤشرات ثبات وصدق الفقرات وهذه الطرق هي الطريقة التقليدية ومعادلة التصحيح لأثر التخمين وطريقة الاختبار الجزئي ، وتكونت عينة الدراسة من (٩٠) طالبة من طالبات الصف الخامس الأساسي ، وظهرت النتائج أن هناك فروق ذات دلالة إحصائية بين متوسطات مؤشرات ثبات ومعاملات تمييز الفقرة عند مستوى دلالة (٠.٠٥) لصالح طريقة التصحيح المعتمدة على الاختبار الجزئي يليه طريقة التصحيح لأثر التخمين ثم الطريقة التقليدية بينما معاملات الصعوبة كانت لصالح الطريقة التقليدية للتصحيح يليه طريقة الاختبار الجزئي ثم طريقة التصحيح لأثر التخمين.

دراسة اللحياني (٢٠١٢) التي كانت بعنوان " أثر بعض تقدير الدرجات للمفردات على ثبات وصدق درجات اختبار تحصيلي في الرياضيات ذي الاختيار من متعدد لدى طالبات الصف الأول الثانوي بمكة المكرمة " ، وهدفت إلى مقارنة أثر أربع طرائق لتقدير الدرجات في صدق اختبار تحصيلي من نوع الاختيار من متعدد في مادة الرياضيات وهي (التقليدية ، والمكافأة ، والاحتمالية ، والتجريبية) ، ويتألف الاختبار من (٢٠) مفردة اختباريه لكل منها أربعة بدائل

وكذلك أربع نماذج مختلفة لتعليمات الاختبار حسب طريقة تقدير الدرجات المتبعة ، وطبقت الدراسة على عينة مكونة من (٤٠٠) طالبة ، وقد توصلت الدراسة إلى نتائج أهمها أن الطريقتين التقليدية والاحتمالية تحققان ثباتاً أعلى للاختبار ، وأوصت بالابتعاد عن طريقة المكافأة.

التعقيب على الدراسات السابقة:

أن الدراسات السابقة التي تناولت الخصائص السيكومترية لاختبار الاختيار من متعدد الذي يتطلب بديلين صحيحة كان هناك تباين فيما توصلت إليه من نتائج وفقاً لما يلي:

-دراسات أظهرت في نتائجها أن الاختبار الذي يتطلب بديلين صحيحة بشكل عام كانت مستويات الثبات فيها عالية مثل دراسة Breuer, S.; Scherndl, T., & Ortner, T. M. (2020) ،ودراسة Anderson, H. R. (2020) ، ودراسة الجبة (٢٠١٢).

-دراسات أظهرت في نتائجها أن معامل الصدق للاختبار الذي يتطلب بديلين صحيحة كانت بمستويات جيدة كدراسة Abreu-Mendoza , R. A.; Chamorro, Y., & Matute, T. (2019) بعكس دراسة Breuer, S.; Scherndl, T., & Ortner, T. M. (2020) التي اظهرت في نتائجها تدني مستويات الصدق .

-دراسات أظهرت في نتائجها ارتباط الخصائص السيكومترية للاختبار الذي يتطلب بديلين صحيحة بطريقة تقدير الدرجات كدراسة Ketterlin-Geller, L. R.; Perry, T., & Platas, P. M. (2019) ، ودراسة اللحياني (٢٠١٢) ، ودراسة الجبة (٢٠١٢)، ودراسة صمصم(٢٠١٢) التي تباينت في نتائجها وفقاً لمعاملات الصعوبة والتمييز والصدق والثبات.

منهج وإجراءات الدراسة :

منهج الدراسة :

استخدم الباحث المنهج شبة التجريبي لملائمة للإجابة عن اسئلة الدراسة.

مجتمع الدراسة :

تكون مجتمع الدراسة من جميع طلاب الصف الثالث المتوسط (بنين) بتعليم محافظة القنفذة خلال الفصل الدراسي الثاني من العام الدراسي ١٤٤٠ / ١٤٤١هـ والبالغ عددهم (٢٤١٣) طالباً يتوزعون على (٦) مكاتب تعليمية كما هو موضح بالجدول التالي :

جدول (١) توزيع مجتمع الدراسة على المكاتب التعليمية بمحافظة القنفذة

مكتب التعليم	عدد طلاب الصف الثالث المتوسط
القنفذة	٦٥٤
حلي	٢٨٢
العرضية الشمالية	٤٠٦
العرضية الجنوبية	٢٢٧
المظيلف	٤٣٤
القوز	٤١٠
المجموع	٢٤١٣

عينة الدراسة :

تكونت عينة الدراسة من (١٣٢) طالباً من طلبة الصف الثالث المتوسط والذي تم اختيارهم بطريقة عشوائية طبقية من مدارس مكتب التعليم بالمظيلف.

أداة الدراسة :

لتحقيق هدف الدراسة تم إعداد اختبار من نوع اختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات في وحدة (كثيرات الحدود) للصف (الثالث المتوسط) للعام الدراسي ١٤٤٠ / ١٤٤١هـ وفق الخطوات التالية:

١. تحديد المحتوى والمتمثل في وحدة كثيرات الحدود للصف الثالث المتوسط.
٢. بناء جدول المواصفات للاختبار وفق الأهمية النسبية للمواضيع مع الأهمية النسبية للأهداف لكل موضوع كما هو موضح بالجدول التالي:

جدول (٢) جدول مواصفات اختبار اختيار من متعدد يتطلب بديلين صحيحين في مادة الرياضيات بوحدة كثيرات الحدود للصف الثالث المتوسط للفصل الدراسي الثاني من العام الدراسي ١٤٤٠ / ١٤٤١هـ

الموضوعات	الاهداف التربوية				مجموع الاهداف	مجموع الحصص	الوزن النسبي للموضوعات
	تذكر	فهم	تطبيق	تحليل			
ضرب وحيدات الحد	–	١	١	–	٢	٢	%١٦.٦٦
قسمة وحيدات الحد	–	١	٢	–	٣	٢	%١٦.٦٦
كثيرات الحدود	١	١	١	–	٣	١	%٨.٣٣
جمع كثيرات الحدود وطرحها	–		٢	–	٢	٢	%١٦.٦٦
ضرب وحيدة حد في كثيرة حدود	–	١	٢	١	٤	١	%٨.٣٣
ضرب كثيرات حدود باستعمال خاصية التوزيع	–	–	٣	–	٣	٢	%١٦.٦٦
حالات خاصة من ضرب كثيرات الحدود	–	–	٢	١	٣	٢	%١٦.٦٦
الوزن النسبي للأهداف	%٥	%٢٠	%٦٥	%١٠	٢٠	١٢	%١٠٠

٣- تحكيم الاختبار تم عرض فقرات الاختبار على مجموعة من المحكمين من ذوي الاختصاص والبالغ عددهم (١١) وكانت نسبة اتفاق المحكمين على فقرات الاختبار ٩٠ % على جميع فقرات الاختبار بصورته النهائية بعدد (٢٠) فقرة.

٤ – التجربة الأولية للاختبار على عينة استطلاعية مكونة من (٢٤) طالباً من طلاب الصف الثالث المتوسط تم اختيارهم عشوائياً من خارج عينة الدراسة بهدف التأكد من وضوح التعليمات والصياغة اللغوية للفقرات والزمن المناسب للإجابة على اسئلة الاختبار وقد تم تحديد الزمن اللازم للإجابة على اسئلة الاختبار وقدر بـ (٣٠) دقيقة .

٥- الصورة النهائية للاختبار بعد التأكد من وضوح التعليمات وصياغة المفردات وتحديد زمن الاختبار وقد تم وضع الاختبار على الرابط التالي :

https://cutt.us/1HiPV

٦- تصحيح الاختبار تم استخدام طريقتين لتصحيح الاختبار:

- طريقة المعرفة التامة : يعطى الطالب (١) درجة إذا اختار البديلين الصحيحين وإذا اختار بعض البدائل يعطى (٠).
- طريقة المعرفة الجزئية : يعطى الطالب (١) عن كل بديل صحيح حيث تتراوح درجة إجابة كل سؤال من (٠ – ٢ درجتان).

الأساليب الإحصائية :

للإجابة على اسئلة الدراسة ، فقد تم استخدام الاساليب الإحصائية التالية :

- معاملات الصعوبة والتمييز.
- معامل ثبات الفا كرونباخ
- معامل ارتباط بيرسون لحساب صدق الاتساق الداخلي لفقرات الاختبار تبعاً لطريقة التصحيح.
- اختبار(ت) لعينتين مرتبطتين لمعرفة دلالة الفروق بين متوسطي معاملات الصعوبة والتمييز وصدق الاتساق الداخلي للاختبار تعزى لطريقة تقدير الدرجات.
- الاختبار الإحصائي (Feldt) لمعرفة دلالة الفروق بين معاملي ثبات الفا كرونباخ للاختبار تعزى لطريقة تقدير الدرجات.

نتائج الدراسة وتفسيرها:

السؤال الأول : هل توجد فروق دالة إحصائياً بين قيم معامل الارتباط المصحح بالدرجة الكلية لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

للإجابة على هذا السؤال تم حساب معاملات ارتباط درجة كل فقرة بالدرجة الكلية باستخدام معامل ارتباط بيرسون كما هو موضح بالجدول التالي:

جدول (٣) معامل الارتباط المصحح بالدرجة الكلية لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية

طريقة المعرفة التامة لتقدير الدرجات				طريقة المعرفة الجزئية لتقدير الدرجات			
رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط
١	٠.٦٨**	١١	٠.٦٨**	١	٠.٣٢*	١١	٠.٣٢*
٢	٠.٦٥**	١٢	٠.٨٦**	٢	٠.٤٤**	١٢	٠.٣٠*
٣	٠.٦٥**	١٣	٠.٨٠**	٣	٠.٣٥**	١٣	٠.٢٧*
٤	٠.٨٩**	١٤	٠.٦٨**	٤	٠.٢٣*	١٤	٠.٣٢*
٥	٠.٦٨**	١٥	٠.٧٦**	٥	٠.٣٢*	١٥	٠.٣٠*
٦	٠.٨٠**	١٦	٠.٥٩**	٦	٠.٢٨*	١٦	٠.٣٢*
٧	٠.٥٨**	١٧	٠.٥٩**	٧	٠.٣٣**	١٧	٠.٤٤**
٨	٠.٦٨**	١٨	٠.٥٦**	٨	٠.٣٢*	١٨	٠.٤٤**
٩	٠.٦٧**	١٩	٠.٦٨**	٩	٠.٣٣**	١٩	٠.٣٣**
١٠	٠.٥٩**	٢٠	٠.٦٨**	١٠	٠.٤٤**	٢٠	٠.٣١*

بالنظر إلى الجدول السابق نجد أن قيم معاملات الارتباط بين درجة كل فقرة والدرجة الكلية للاختبار وفق طريقة المعرفة التامة لتقدير الدرجات تراوحت من (٠.٥٦ – ٠.٨٩) قد كانت موجبة و دالة إحصائياً ، أما بالنسبة لقيم معاملات الارتباط بين درجة كل فقرة والدرجة الكلية للاختبار وفق طريقة المعرفة الجزئية لتقدير الدرجات تراوحت من (٠.٢٣ – ٠.٤٤) وقد كانت موجبة و دالة إحصائياً أيضاً ولمعرفة الفروق في قيم معاملات الارتباط وفق طريقتي تقدير الدرجات تم استخدام اختبار(ت) لعينتين مترابطتين كما هو موضح بالجدول التالي :

جدول (٤) اختبار (ت) لعينتين مترابطتين لمتوسطات قيم معاملات الارتباط وفق طريقة تقدير الدرجات التقليدية

طريقة تقدير الدرجات التقليدية	عدد فقرات الاختبار	المتوسط	الانحراف المعياري	درجات الحرية	قيمة (ت)	الدلالة
طريقة المعرفة التامة لتقدير الدرجات	٢٠	٠.٦٩	٠.٠٩٢	١٩	١١.٠٥	٠.٠٠٠
طريقة المعرفة الجزئية لتقدير الدرجات	٢٠	٠.٣٤	٠.٠٥٩			

يتضح من الجدول السابق أن قيمة (ت) تساوي ١١.٠٥ وهي دالة إحصائية عند مستوى (٠.٠١) مما يعني أنه يوجد فروق بين متوسطات قيم معاملات الارتباط وفق طريقتي تقدير الدرجات وكانت الفروق لصالح طريقة المعرفة التامة لتقدير الدرجات ، وبالتالي هناك تأثير لطريقة تقدير الدرجات على صدق الاختبار.

وتتفق نتيجة هذه الدراسة مع دراسة Abreu-Mendoza , R. A.; Chamorro, Y., & Matute, T. (2019) التي اشارت إلى ظهور مستويات جيدة من الصدق الداخلي لاختبار الرياضيات الذي يتطلب الاجابة ببديلين صحيحين ومستويات خصائص سيكومترية مناسبة ، بينما اختلفت مع دراسة Breuer, S.; Scherndl, T., & Ortner, T. M. (2020) التي أشارت نتائجها أن نسخة الاختبار التي تتطلب بديلين صحيحين للإجابة أظهر مستويات أعلى في الثبات لكن مستويات أدنى في الصدق المعياري بالمقارنة مع نسخة اختبار الرياضيات ذات الإجابة الصحيحة الواحدة المحددة.

وربما يعود السبب في ذلك إلى انه تبعاً لطريقة المعرفة التامة لتقدير الدرجات فإن الطالب يحصل على درجة واحدة إذا اختار جميع البدائل الصحيحة ، وهذا مما يزيد من ثقة الفقرات للاختبار وبالتالي يزيد معامل صدقه ، أما بالنسبة لطريقة المعرفة الجزئية لتقدير الدرجات فغن الطالب يكون لديه ثقة أكبر بالحصول على درجة إذا اختار إحدى البدائل الصحيحة مما يساعد على زيادة التخمين وبالتالي يقلل من صدق الاختبار في هذه الحالة.

السؤال الثاني : هل توجد فروق دالة إحصائياً بين متوسطي معاملات الصعوبة لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

للإجابة على هذا السؤال تم حساب قيم معاملات الصعوبة لكل فقرة من فقرات الاختبار كما هو موضح بالجدول التالي:

جدول (٥) معاملات الصعوبة وفق طريقة تقدير الدرجات التقليدية

طريقة المعرفة التامة لتقدير الدرجات				طريقة المعرفة الجزئية لتقدير الدرجات			
رقم الفقرة	معامل الصعوبة	رقم الفقرة	معامل الصعوبة	رقم الفقرة	معامل الصعوبة	رقم الفقرة	معامل الصعوبة
١	٠.٦٣	١١	٠.٩١	١	٠.٢٩	١١	٠.٢٤
٢	٠.٧٢	١٢	٠.٩١	٢	٠.٥٦	١٢	٠.٤٤
٣	٠.٤٤	١٣	٠.٩٢	٣	٠.٢٦	١٣	٠.٣١
٤	٠.٥٥	١٤	٠.٦٣	٤	٠.٤١	١٤	٠.٥٠
٥	٠.٥٦	١٥	٠.٨٢	٥	٠.٤٢	١٥	٠.٣٨
٦	٠.٩١	١٦	٠.٨١	٦	٠.٢٧	١٦	٠.٢٥
٧	٠.٦٣	١٧	٠.٧٥	٧	٠.٥٩	١٧	٠.٤٣
٨	٠.٩٢	١٨	٠.٧٢	٨	٠.٣٦	١٨	٠.٢٦
٩	٠.٤٢	١٩	٠.٧٠	٩	٠.٣٢	١٩	٠.٣٨
١٠	٠.٧٥	٢٠	٠.٦٤	١٠	٠.٤٠	٢٠	٠.٢٧

بالنظر إلى الجدول السابق نجد أن قيم معاملات الصعوبة للاختبار وفق طريقة المعرفة التامة لتقدير الدرجات تراوحت من (٠.٤٢ – ٠.٩٢) ، أما بالنسبة لقيم معاملات الصعوبة للاختبار وفق طريقة المعرفة الجزئية لتقدير الدرجات تراوحت من (٠.٢٤ – ٠.٥٩) ولمعرفة الفروق في قيم معاملات الصعوبة وفق طريقتي تقدير الدرجات تم استخدام اختبار(ت) لعينتين مترابطتين كما هو موضح بالجدول التالي :

جدول (٦) اختبار (ت) لعينتين مترابطتين لمتوسطات قيم معاملات الصعوبة وفق طريقة تقدير الدرجات التقليدية

طريقة تقدير الدرجات التقليدية	عدد فقرات الاختبار	المتوسط	الانحراف المعياري	درجات الحرية	قيمة (ت)	الدلالة
طريقة المعرفة التامة لتقدير الدرجات	٢٠	٠.٧٢	٠.١٦	١٩	٧.٩٤	٠.٠٠٠
طريقة المعرفة الجزئية لتقدير الدرجات	٢٠	٠.٣٧	٠.١٠			

يتضح من الجدول السابق أن قيمة (ت) تساوي ٧.٩٤ وهي دالة إحصائية عند مستوى (٠.٠١) مما يعني أنه يوجد فروق بين متوسطات قيم معاملات الصعوبة وفق طريقتي تقدير الدرجات وكانت الفروق لصالح طريقة المعرفة التامة لتقدير الدرجات ، وبالتالي هناك تأثير لطريقة تقدير الدرجات على معاملات الصعوبة.

وتتفق نتيجة هذه الدراسة مع دراسة الجبة (٢٠١٢) التي أشارت إلى أن المفردات التي تتضمن أكثر من إجابة صحيحة أعلى في الصعوبة من المفردات التي تتضمن إجابة صحيحة واحدة ، كما تتفق مع دراسة صمصم (٢٠١٢) التي أشارت إلى أن هناك فروق ذات دلالة إحصائية بين متوسطات مؤشرات صعوبة الفقرة عند مستوى دلالة (٠.٠٥) لصالح الطريقة التقليدية للتصحيح يليه طريقة الاختبار الجزئي ثم طريقة التصحيح لأثر التخمين.

ويرى الباحث ان السبب في ذلك ربما يعود إلى أن طريقة المعرفة الجزئية تتيح للطالب فرصة تخمين الإجابة مما يؤدي إلى أن تكون معاملات صعوبة الفقرات أقل بخلاف طريقة المعرفة التامة التي تؤدي إلى تحفظ الطالب وعدم الاستعجال في الإجابة على الفقرات مما يعطي تقديراً اعلى لمعاملات الصعوبة في هذه الطريقة.

السؤال الثالث : هل توجد فروق دالة إحصائياً بين متوسطي معاملات التمييز لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

للإجابة على هذا السؤال تم حساب قيم معاملات التمييز لكل فقرة من فقرات الاختبار كما هو موضح بالجدول التالي:

جدول (٧) معاملات التمييز وفق طريقة تقدير الدرجات التقليدية

طريقة المعرفة التامة لتقدير الدرجات				طريقة المعرفة الجزئية لتقدير الدرجات			
رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز	رقم الفقرة	معامل التمييز
١	٠.٧٠	١١	٠.٦٨	١	٠.١٨	١١	٠.٢٤
٢	٠.٤٦	١٢	٠.٥١	٢	٠.٢٧	١٢	٠.٢٨
٣	٠.٣٧	١٣	٠.٢٤	٣	٠.١٦	١٣	٠.٣٤
٤	٠.٤١	١٤	٠.٨٢	٤	٠.٣٣	١٤	٠.١٠
٥	٠.٦١	١٥	٠.١٧	٥	٠.٣٧	١٥	٠.٣٣
٦	٠.٤٩	١٦	٠.٣٩	٦	٠.٣٠	١٦	٠.٠٩
٧	٠.٦١	١٧	٠.٤٦	٧	٠.١٢	١٧	٠.١٢
٨	٠.٥٩	١٨	٠.٢٨	٨	٠.٢٠	١٨	٠.١٥
٩	٠.٢١	١٩	٠.٦٥	٩	٠.١٧	١٩	٠.٧٢
١٠	٠.٥٦	٢٠	٠.٧١	١٠	٠.١١	٢٠	٠.٣٧

بالنظر إلى الجدول السابق نجد أن قيم معاملات التمييز للاختبار وفق طريقة المعرفة التامة لتقدير الدرجات تراوحت من (٠.١٧ – ٠.٨٢) ، أما بالنسبة لقيم معاملات التمييز للاختبار وفق طريقة المعرفة الجزئية لتقدير الدرجات تراوحت من (٠.٠٩ – ٠.٧٢) ولمعرفة الفروق في قيم معاملات التمييز وفق طريقتي تقدير الدرجات تم استخدام اختبار(ت) لعينتين مترابطتين كما هو موضح بالجدول التالي :

جدول (٨) اختبار (ت) لعينتين مترابطتين لمتوسطات قيم معاملات التمييز وفق طريقة تقدير الدرجات التقليدية

طريقة تقدير الدرجات التقليدية	عدد فقرات الاختبار	المتوسط	الانحراف المعياري	درجات الحرية	قيمة (ت)	الدلالة
طريقة المعرفة التامة لتقدير الدرجات	٢٠	٠.٥٠	٠.١٨	١٩	٥.٣٠	٠.٠٠٠
طريقة المعرفة الجزئية لتقدير الدرجات	٢٠	٠.٢٢	٠.٠٩			

يتضح من الجدول السابق أن قيمة (ت) تساوي ٥.٣٠ وهي دالة إحصائية عند مستوى (٠.٠١) مما يعني أنه يوجد فروق بين متوسطات قيم معاملات التمييز وفق طريقتي تقدير الدرجات وكانت الفروق لصالح طريقة المعرفة التامة لتقدير الدرجات ، وبالتالي هناك تأثير لطريقة تقدير الدرجات على معاملات تمييز الفقرات.

وتتفق نتيجة هذه الدراسة مع دراسة الجبة (٢٠١٢) التي أشارت إلى أن المفردات التي تتضمن أكثر من إجابة صحيحة أعلى في التمييز من المفردات التي تتضمن إجابة صحيحة واحدة ، وتختلف مع نتيجة دراسة صمصم (٢٠١٢) التي أشارت إلى أن هناك فروق ذات دلالة إحصائية بين متوسطات معاملات تمييز الفقرة عند مستوى دلالة (٠.٠٥) لصالح طريقة التصحيح المعتمدة على الاختبار الجزئي يليه طريقة التصحيح لأثر التخمين ثم الطريقة التقليدية.

وربما يعود السبب في ذلك إلى أن طريقة المعرفة الجزئية تتيح للطالب فرصة تخمين الإجابة مما يؤدي إلى أن تكون معاملات تمييز الفقرات بين الفئة العليا والدنيا أقل مما هي عليه ، وعلى عكس طريقة المعرفة التامة التي تؤدي إلى تحفظ الطالب وعدم الاستعجال في الإجابة على الفقرات مما يعطي تقديراً أعلى لمعاملات التمييز بين الفئة العليا والدنيا لفقرات الاختبار.

السؤال الرابع : هل توجد فروق دالة إحصائياً بين معاملي ثبات ألفا كرونباخ لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية؟

للإجابة على هذا السؤال تم حساب معاملات الثبات باستخدام معامل ألفا كرونباخ للاختبار تبعاً لطريقتي تقدير الدرجات ، كما تم اختبار الفروق بين معاملي ثبات ألفا كرونباخ باستخدام الاختبار الإحصائي Feldt حسب المعادلة التالية:

$$W=\frac{1-\alpha_2}{1-\alpha_1}$$

(Feldt & Kim , 2006)

جدول (٩) اختبار Feldt (W) لمعرفة الفروق بين معاملي ثبات ألفا كرونباخ لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات التقليدية

طريقة تقدير الدرجات التقليدية	قيمة معامل ألفا كرونباخ	Feldt (W)	df	قيمة t عند مستوى (0.05)
طريقة المعرفة التامة لتقدير الدرجات	٠.٨٤٣	١.٠٥	١٩	١.٦٩
طريقة المعرفة الجزئية لتقدير الدرجات	٠.٨٠٢			

يتضح من الجدول السابق أن قيمة معامل ألفا كرونباخ للاختبار وفق طريقة المعرفة التامة لتقدير الدرجات تساوي (٠.٨٤٣) بينما كانت قيمة معامل ألفا كرونباخ للاختبار وفق طريقة المعرفة الجزئية لتقدير الدرجات تساوي (٠.٨٠٢) ، كما نجد أن قيمة (t) الجدولية تساوي (١.٦٩) عند مستوى (٠.٠٥) وهي أكبر من قيمة (W) المحسوبة بطريقة Feldt والتي تساوي (١.٠٥) مما يعني أنه لا توجد فروق بين معاملي ثبات ألفا كرونباخ لاختبار الاختيار من متعدد يتطلب بديلين صحيحة في مادة الرياضيات تعزى إلى طريقة تقدير الدرجات ، وبالتالي فإن ليس هناك تأثير لطريقة تقدير الدرجات على ثبات الاختبار.

تختلف نتيجة هذه الدراسة مع دراسة Anderson, H. R. (2020) التي أسفرت نتائجها عن بقاء جميع تقديرات الثبات الداخلي لاختبار الرياضيات القائم على أكثر من بديل صحيح ، وتختلف مع دراسة الجبة (٢٠١٢) التي أشارت إلى أن المفردات التي تتضمن أكثر من إجابة صحيحة أعلى في الثبات من المفردات التي تتضمن إجابة صحيحة واحدة ، وتختلف أيضاً مع نتيجة دراسة صمصم (٢٠١٢) التي أشارت إلى أن هناك فروق ذات دلالة إحصائية بين متوسطات مؤشرات ثبات الفقرة عند مستوى دلالة (٠.٠٥) لصالح طريقة التصحيح المعتمدة على الاختبار الجزئي يليه طريقة التصحيح لأثر التخمين ثم الطريقة التقليدية.

وربما يعود السبب في ذلك إلى أن معامل ثبات ألفا كرونباخ يتعامل مع نوع واحد من مصادر الخطأ وهو مضمون البنود ، والأفضل أن يتم حساب معامل الثبات بطريقة إعادة الاختبار وهذا ما أكده (Haberman, 2018, 205) الذي يرى أن الدرجة الحقيقية التي يحصل

عليها المفحوص في اختبار معين (اختبار للرياضيات) تعكس معرفة وأداء المفحوص في تلك المادة أو مجال الاختبار، ويجب أن تتشابه هذه الدرجة مع ما يحصل عليه المفحوص في الأداء المستقبلي على نفس الاختبار (بمعنى أن الاختبار يكون ذو ثبات اختبار/إعادة الاختبار) ؛ حيث ترتبط أهمية حساب الصدق والثبات (الخصائص السيكومترية) بتعميم درجات الاختبار بمعنى قابلية الأداء للتكرار خلال الاختبارات المستقبلية ، وهو ما يرتبط بصورة مباشرة بنوعية إجراءات تقدير الدرجات.

التوصيات والمقترحات:

في ضوء ما توصلت إليه الدراسة من نتائج فيمكن التوصية بما يلي :

- استخدام الفقرات التي تتطلب بديلين صحيحة في اختبارات الاختيار من متعدد نظراً لكفاءة خصائصها السيكومترية لضعف التخمين في هذا النوع مما أنعكس على كفاءة استخدامه.
- استخدام نمط الفقرات التي تتطلب بديلين صحيحة بدلاً من نمط الفقرات التي لها إجابة واحدة صحيحة ، وبخاصة عندما يتعلق القياس بأهداف معينة كحل المسائل الرياضية التي تحوي على أكثر من طريقة للحل.
- إدخال مواضيع بناء وتصميم الفقرات التي تتطلب بديلين صحيحة في اختبارات الاختيار من متعدد وطرق تقدير درجاتها ضمن مقررات القياس والتقويم التربوي بكليات التربية.
- إقامة دورات تدريبية للمعلمين بالتعاون مع إدارات التعليم في بناء الاختبارات وبالذات بنوع الفقرات التي تتطلب بديلين صحيحة في اختبارات الاختيار من متعدد لمختلف المراحل الدراسية.
- إجراء المزيد من الدراسات حول طريقة تقدير الدرجات التقليدية وأثرها على الخصائص السيكومترية لاختبار يتطلب بديلين صحيحة في مقررات دراسية مختلفة وبمختلف المراحل الدراسية.
- إجراء دراسات حول استخدام طرق أخرى لتقدير الدرجات وأثرها على الخصائص السيكومترية لاختبار يتطلب بديلين صحيحة .

Abstract

Grades Assessment Way and Its Effect on The Psychometrics Properties of Multiple Choices Test which Required Two Correct Alternatives.

Key Words: Grades Assessment, The Psychometrics Properties, Multiple choices test

Dr. Mohammad Rashid Abdulkarim ALzahrani. Associate professor in Education and Psychology Department. College of Education at ALBaha University.

The study aimed to identify the grades assessment way and its effect on the psychometrics properties of multiple choices test which required two correct alternatives. The multiple choices test required two correct alternatives was made in the math for the third level in middle school for the (Unit of Polynomial). The data of study was collected from (123) male students of the third level of middle school in 1440/1441year from AL-Qunfudhah Governorate education administration. The results of the study showed there were significant differences (P=0.01) for the correlation of the item and the total scores of The multiple choices test required two correct alternatives, also there were significant differences (P=0.01) between means of difficulty and discrimination of coefficients for multiple choices test which required two correct alternatives for the full knowledge of the grades assessment, and there were not significant differences for the test reliability caused by the grades assessment .

المصادر

- الثمالي، عبدالله ابن عبدالهادي، (٢٠١٦).أثر اختلاف عدد البدائل على الخصائص السيكومترية لفقرات اختبار اختيار من متعدد يتطلب أكثر من إجابة صحيحة، رسالة ماجستير غير منشورة ، كلية التربية ، جامعة أم القرى.
- الجبة، عصام الدسوقي إسماعيل،(٢٠١٢).الخصائص السيكومترية لمفردات اختيار من متعدد تتضمن اكثر من إجابة صحيحة،مجلة كلية التربية بالمنصورة،ع٨١، ١٤١-١٧٣.
- الغامدي ، عبدالله أحمد،(٢٠٠٨).أثر عدد البدائل في الخصائص السيكومترية للاختبار التحصيلي في الرياضيات ، رسالة ماجستير غير منشورة ، كلية التربية جامعة أم القرى.
- اللحياني، عفاف بنت راضي مشخص،(٢٠١٢).أثر بعض طرق تقدير الدرجات على ثبات وصدق درجات اختبار تحصيلي في الرياضيات ذي الاختيار من متعدد لدى طالبات الصف الأول الثانوي بمكة المكرمة،رابطة التربويون العرب،ع٢٢(٢)،٤٨٧-٥١٦.

- صمصم، ميساء محمود،(٢٠١٢).أثر طرق التصحيح لضبط التخمين في اختبارات الاختيار من متعدد على كل من مؤشر ثبات الفقرة ومؤشر صدقها، رسالة ماجستير غير منشورة ، جامعة مؤتة ،الاردن.
- عبدالوهاب ، صلاح شريف ، (٢٠٠١).أثر بعض الطرق الوزنية لتقدير الدرجات على صدق الاختبارات مرجعية المحك ذات الاختيار من متعدد ، مجلة كلية التربية ببنها ، ٩(٣)،٣١٩-٣٣٣.
- علام ، صلاح الدين محمود ،(٢٠٠٧).القياس والتقويم التربوي ، ط١، عمان : دار المسيرة للنشر.
- علام ، صلاح الدين محمود ،(٢٠١١).القياس والتقويم التربوي والنفسي ، القاهرة : دار الفكر العربي.
- عودة ، أحمد ، (٢٠١٠).القياس والتقويم في العملية التدريسية ، الأردن : دار الأمل.
- غيث ، محمد عبدالله، (٢٠٠٧). الخصائص السيكومترية لاختبارات عينة من المواد في مشروع الاختبارات المدرسية للصف الأول متوسط بمنطقة المدينة المنورة ، رسالة ماجستير غير منشورة، كلية التربية ، جامعة أم القرى.
- فتح الله ، مندور عبدالسلام،(٢٠٠٦).التقويم التربوي ، ط٢ ، الرياض: دار النشر الدولي.
- مجيد ، سوسن شاكر ، (٢٠١٤). أسس بناء الاختبارات والمقاييس النفسية والتربوية ، عمان : مركز ديبوتو لتعليم التفكير.
- Abreu-Mendoza , R. A.; Chamorro, Y., & Matute, T. (2019). Psychometric Properties Of Math Computation Test Which Requires Two Correct Alternatives With Mexican Students, *Journal of Psychoeducational Assessment; 1 (8).*
- Albanese, M. A., & Sabers, D. L. (2018). Multiple True-False Items: A Study Of Interitem Correlations, Scoring Alternatives, And Reliability Estimation, Journal Of Educational Measurement; 25 (2).
- Anderson, H. R. (2020). A Psychometric Properties Of A Mathematics Test Based On Correct Alternatine Choices, *PhD Thesis, Kent State University: USA.*
- Blair, I. V., Lenton, A. P., & Hastie, R. (2016). The Reliability Of The DRM Paradigm As A Measure Of Individual Differences In False Memories. Psychological Bulletin & Review, 9 (3), 590 – 596.
- Bland, A. C.; Kreiter, C. D., &Gordon, J. A. (2019). The Psychometric

Properties Of Five Scoring Methods Applied To Math Test, Academics; 80 (4).

- Breuer, S.; Scherndl, T., & Ortner, T. M. (2020). Effects of Response Format on Psychometric Properties of a Mathematics Test: Two Correct Choices vs. Free Response Test, *Frontiers in Education; 5 (15).*
- Briere, J. L. (2018). Determining The Psychometric Properties Of The Retrieval-Induced Forgetting Procedure, Phd Thesis, University Of Saskatchewan.
- Brombacher, A. (2015). National Intervention Research Activity For Early Grade Mathematics In Jordan. In X. Sun, B, Kaur, & J. Novotná (Eds.) Conference Proceedings Of The Twenty-Third ICMI Study: Primary Mathematics Study On Whole Numbers.
- Brombacher, A., Bulat, J., King, S., Kochetkova, K., And Nordstrum, L. (2015). National Assessment Survey Of Learning Achievement At Grade 2: Results For Early Grade Reading And Mathematics In Zambia. Research Triangle Park, NC: RTI International.
- Camilli, G. &. Chiu, T. (2018). Two New 1RT-Based Corrections For Guessing. University Of Colorado, Boulder.
- Cicchetti, D. V. (2017). The Precision Of Reliability And Validity Estimates Re-Visited: Distingquishing Between Clinical And Statistical Significance Of Sample Size Requirements, Journal Of Clinical And Experimental Neuropsychology, 23 (5).
- Cortina, J. M. (2019). What Is Coefficient Alpha? An Examination Of Theory And Applications. Journal Of Applied Psychology, 78, 98 – 104.
- Dai, S., Wang, X., & Svetina, D. (2016). Subscore: Computing Subscores In Classical Test Theory And Item Response Theory. R Package. Bloomington, Indiana: Indiana University.
- Fehintola, J.O., & Akingbade, O. (2019). Comparison Of Psychometric Properties Of Multiple-Choice Test Using Confidence And Number Right Scoring Among Senior Secondary School Students In Ibadan Metropolis, Al-Hikmah Journal Of Education; 6 (1).
- Feldt, L. S. & Kim, S.(2006). Testing the difference between tow alpha coefficients with small samples of subjects and raters. Educational and Psychological Measurement, 66(4), 589-600.
- Forsyth, R. A., & Spratt, K. F. (2016). Measuring Problem Solving Ability In Mathematics With Multiple-Choice Items: The Effect Of Item Format On Selected Item And Test Characteristics, Journal Of Educational Measurement; 17 (1).
- Haberman, S. J. (2018). When Can Subscores Have Value? Journal Of Educational And Behavioral Statistics, 33, 204–229.
- Ketterlin-Geller, L. R.; Perry, T., & Platas, P. M. (2019). Effects of Test Scoring Procedures in Mathematic Tests with Multiple Right Alternatives on Test Uses in Light of Psychometric Properties, *Global Education*

Review; 5 (3).

- Rust, J. (2019). The Psychometric Principles Of Assessment, Research Matters; 3 (22).
- Skolnick, A. (2018). A Structure And Scoring Method For Judging Alternatives, IEEE Transactions ON Engineerinlj MANAGEMENT; 16 (2).
- Stankous, N. V. (2016). Constructive Response Vs. Multiple-Choice Tests In Math: American Experience And Discussion (Review), European Scientific Journal; 1 (19).
- Thissen, D., Wainer, H., & Wang, X. (2017). Are Tests Comprising Both Multiple-Choice And Free-Response Items Necessarily Less Unidimensional Than Multiple-Choice Tests An Analysis Of Two Tests. Journal Of Educational Measurement, 31(2).

طريقة تقدير الدرجات وأثرها على الخصائص السيكومترية لاختبار اختيار من متعدد يتطلب بديلين صحيحة

(تعليمات الاختبار)

عزيزي الطالب :

بين يديك اختبار من نوع اختيار من متعدد يتطلب أكثر من بديل صحيح في وحدة كثيرات الحدود بمادة الرياضيات للصف الثالث متوسط ويتكون الاختبار من (٢٠) سؤالاً كل سؤال له أربع بدائل ويوجد له إجابتين صحيحة.

الرجاء التفضل بقراءة كل سؤال بعناية ثم اختيار الإجابة الصحيحة بوضع علامة (✓) ولا تترك أي سؤال دون الإجابة عليه.

مثال :

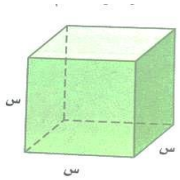
ناتج ما يلي : (أ -٢)@=			
أ	أ –٢أ ب +٤	ب✓	أ@–٤أ +٤
ج✓	أ@ –٤أ +٢@	د	أ –٢أ@ ب +٤

علماً بأن هذا الاختبار يستخدم للبحث العلمي فقط.

ونشكر لك حسن تعاونك،،،

	أوجد ناتج ما يلي:٣س@ ×٤س% =

١	أ	١٢س&	ب	٧س &
	ج	٦×٢س&	د	١٢ س (!
٢	ك (٥ك)@=			
	أ	٥٠ك#	ب	٥٠ك@
	ج	٢٥ك#	د	٢٥ك×ك×ك
٣	١٤ص@ = ٧ ص			
	أ	٢ص	ب	٢
	ج	ص	د	ص+ص
٤	١٥أ@ب% جـ = ٣أب@جـ			
	أ	٥أب#جـ(	ب	١٢أ#ب
	ج	١٢أب#	د	٥أب#
٥	١٦س@ص% ع = ٤أس@ص #			
	أ	٤ص ع*	ب	٢(٢ص* ع)
	ج	٨ص ع%ـ٣	د	٤ص *ع !
٦	كثيرة الحدود س # +٥س@ + ٤ من الدرجة :			
	أ	١	ب	١+٢
	ج	٣	د	٠
٧	٣س+س@ +٥س# - ٤ إذا رتبت على الصورة القياسية فإن:			
	أ	معاملها الرئيس ٥	ب	معاملها الرئيس ٣
	ج	درجتها الثالثة ومعاملها الرئيس ٥	د	درجتها الرابعة
٨	(٧س@ -٤س)+ (٣س@ -٣س)=			
	أ	(١٠س@ -٧س)	ب	(٤س@ -٧س)
	ج	(٢١س@-١٢س)	د	-٧س +١٠س@
٩	(س@ -٢س)- (س@ -٣س)=			
	أ	- (- س)	ب	س
	ج	س@-٥س	د	-٥س
١٠	٢س (س@ -٢س+ ٥)=			
	أ	٢س# -٤س@+ ١٠س	ب	(س@ -٢س+ ٥)
	ج	(س@ -٢س+ ١٠س)	د	-٤س@+ ١٠س+٢س#
١١	عند حل المعادلة : -٦(١١-٢جـ) =٧(-٢-٢جـ) فإن جـ =			
	أ	٢	ب	-٢
	ج	٢÷٥٢	د	٥٢÷٢٦

١٢	(س -٣) (س -٢)=			
	أ	س@ -٥ س-٦	ب	س@ -٥س+ ٦
	ج	س@ -٢س-٣س+ ٦	د	س@ +٥س+ ٦
١٣	(٦س+٥)(٢س@ -٣ س-٥)=			
	أ	١٢ س #-٨س@ -٤٥ س-٢٥	ب	- ٨س@ + ١٢ س # -٢٥ س-٤٥
	ج	-٨س@ -٤٥ س-٢٥+ ١٢ س #	د	١٢ س #-٨س@ -٢٥
١٤	(ب-٦)(ب+٦)=			
	أ	ب@-٣٦	ب	ب@+٣٦
	ج	-٣٦+٢ب	د	-٣٦+ب@
١٥	(ص+٣)@=			
	أ	ص@+٣@	ب	ص@+٦ص+٣@
	ج	ص@+٦ص+٩	د	٢ص+٩
١٦	(٣ أ@ب #)(=			
	أ	١	ب	٠
	ج	-١(-١)	د	١-١
١٧	حجم المكعب = 			
	أ	س#	ب	٣س
	ج	٢س(س)	د	س@(س)
١٨	حدد وحيدة الحد فيما يأتي			
	أ	٥س	ب	٣س(
	ج	٤ / س	د	٥ س¯ #
١٩	محيط الشكل المجاور ٢س ٩س			
	أ	٢٢ س	ب	١٨ س
	ج	٢(٢س+٩س)	د	١٨ س@
٢٠	بسط العبارة ٣ ¯@×٧(×٩			
	أ	٩	ب	١
	ج	٣٦	د	٩(

الاجابة النموذجية للاختبار

م	الخيارات الصحيحة	م	الخيارات الصحيحة
١	أ ، ج	١١	أ ، د
٢	ج ، د	١٢	ب ، ج
٣	أ ، د	١٣	أ ، ج
٤	أ ، د	١٤	أ ، د
٥	ب ، د	١٥	ب ، ج
٦	ب ، ج	١٦	أ ، ج
٧	أ ، ج	١٧	أ ، د
٨	أ ، د	١٨	أ ، ب
٩	أ ، ب	١٩	أ ، ج
١٠	أ ، د	٢٠	ب ، د