المجلة الدولية للدراسات التربوية والنفسية
e-ISSN 2520-4149 , p-ISSN 2520-4130
المجلد٦- العدد١- ٢٠١٩، ص١-١٧
https://doi.org/DOI:10.31559/EPS2019.6.1.1

رفاد للدراسات والأبحاث
www.refaad.com

أثر شكل فقرة الاختيار من متعدد والصواب والخطأ على الخصائص السيكومترية وفق نظرية استجابة الفقرة لاختبار مادة الحاسب الآلي للصف الأول ثانوي في مدينة تبوك

صبري محمد إسماعيل عبد العال
أستاذ مساعد- قسم التربية وعلم النفس- جامعة تبوك- المملكة العربية السعودية
sabdelaal@ut.edu.sa

محمد عويض رجاء العنزي
معلم- المملكة العربية السعودية
abyzm22@gmail.com

الملخص:

هدف هذا البحث إلى التعرف على أثر شكل فقرات الاختيار من متعدد والصواب والخطأ على معالم الفقرة والخصائص السيكومترية للاختبار وفق نموذج ثنائي المعلم لنظرية استجابة الفقرة، ولتحقيق هذا الهدف تم بناء شكلين لاختبار في مادة الحاسب الآلي: الاختيار من متعدد والصواب والخطأ، تكون كل شكل من شكلي الاختبار في صورته النهائية من (٢٤) فقرة، وتم تطبيق الاختبار على (٤٢١) طالباً.

أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية في صعوبة الفقرة بين شكلي الاختبار، إلا إن النتائج كشفت وجود فروق ذات دلالة إحصائية في تمييز الفقرة لصالح اختبار الصواب والخطأ، من ناحية أخرى أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية في معامل الثبات التجريبي بين شكلي الاختبار.

وتوصي نتائج البحث بتوعية المعلمين إلى التنويع بأشكال فقرات الاختبارات التي يستخدمونها بحيث تحتوي على الشكلين الاختيار من متعدد والصواب والخطأ، واستعمال فقرات الصواب والخطأ على نطاق أوسع مما هو عليه الآن كونها أقوى تمييزًا.

الكلمات المفتاحية: الخصائص السيكومترية، شكل فقرة الاختبار، فقرة الاختيار من متعدد، فقرة الصواب والخطأ، نظرية استجابة الفقرة.

المقدمة:

يعتبر الاختبار التحصيلي الذي يعده المعلم ونطلق عليه الاختبار التحصيلي الصفي تمييزاً له عن الاختبار التحصيلي المقنن وهو من أهم أساليب القياس التي يستخدمها المعلم في جمع المعلومات اللازمة لعملية التقويم التربوي، حيث يبقى الاختبار الأنسب والأفضل لأغراض التقويم الصفي وتوفير معلومات للمعلم عن سير العملية التعليمية التعلمية، والمعلومات التي توفرها الاختبارات متنوعة وضرورية جداً وذلك من أجل اتخاذ قرارات مختلفة تتعلق بدرجة إتقان التلاميذ لوحدة دراسية معينة أو بمدى تقدم التلاميذ في مقرر معين بعد فترة زمنية محددة أو ترفيع التلاميذ من صف لأخر. (أبو ناهية، ١٩٩٤: ١٩٧).

وأشار علام (٢٠١٤: ١٢٢) أنه يستخدم مفهوم التحصيل للإشارة إلى درجة أو مستوى النجاح الذي يحرزه الطالب في مجال دراسي عام أو متخصص، فهو يمثل اكتساب المعارف والمهارات والقدرة على استخدامها في مواقف حالية أو مستقبلية، ويعد التحصيل هو الناتج النهائي للتعلم.

وعلى الرغم من أهمية الاختبارات في العملية التربوية، إلا إنّ هناك وجهات نظر مختلفة حول طبيعتها وأشكالها وطرق إجرائها، وكذلك صلتها بالأهداف التربوية وطرق تدريسها، ولهذا فلا غرابة أن تولي المؤسسات التربوية عناية خاصة بالاختبارات التحصيلية، حتى أصبحت الاختبارات تحتل المكانة الأولى في المدارس والجامعات.

يهتم القائمون ببناء الاختبارات والمقاييس بكتابة أو انتقاء مفردات عالية الجودة لقياس السمات الإنسانية قياسًا دقيقًا، لذلك يراعون كثيرًا من الشروط في تكوين هذه المفردات وصياغتها والتحقق بالأساليب المنطقية وأحكام الخبراء من صدق محتوى كل مفردة على حدة وكذلك صدق محتوى الاختبار ككل، غير أنه مهما بلغت دقة هذه الأساليب والأحكام فإنها لا تغني عن التجريب الميداني للاختبار وتحليل درجات مفرداته باستخدام الأساليب الإحصائية وتحديد العلاقة بين ما تقيسه المفردات وبين استجابات الأفراد لها (علام، ٢٠١١: ٢٦٧).

أن التعرف على الأسئلة الموضوعية المختلفة يتيح للطالب والمعلم وواضعي الأسئلة ببنوك الأسئلة فرصة الألفة بتلك الأنواع والتدريب على إعداد نماذج أخرى على شاكلتها والتعرف على المواصفات الفنية المطلوبة في كل نوع على حدة' حتى يتمكن من عمل مخزن أسئلة جيد الصياغة (سليمان،٢٠١٠:٢٤٥).

وذكر النبهان (٢٠١٣:٨٥) أن الخطأ والصواب ما هي إلا حالة خاصة من فقرات الاختيار من متعدد.

واكتسبت الاختبارات الموضوعية شهرة وانتشارًا كبيرين لما تتمتع به من مستوى مرتفع من الموثوقية والبعد عن التحيز في إعطاء الدرجات، ورغم ذلك نجدها تواجه بالنقد في بعض الحالات، كما تقابل بالتأييد والاستحسان في حالات أخرى (عمر، وفخرو، والسبيعي، وتركي،٢٠١٠:٣٧٩)

وأشار عودة (٢٠١٠:١٦٤) أيضًا أن لكل نوع من أنواع الفقرات جوانب ضعف أو عيوب تحد أو لا تشجع من استعمال هذه الفقرات في مواقف معينة وجوانب قوة ومزايا ربما تشجع على استعمالها في مواقف أخرى، ولذلك فإن القرار بشأن نوع من أنواع الفقرات في فئة معينة وتفضيله على نوع آخر من نفس الفئة أو فئة أخرى يتطلب من المعلم إجراء موازنة بين جوانب الضعف والقوة في ضوء الغرض من التقويم والزمن المتوفر للإجابة.

وتُعد نظرية استجابة المفردة من النظريات السيكومترية المعاصرة التي اتضحت فائدتها في التغلب على الكثير من مشكلات القياس التي عجزت النظرية الكلاسيكية عن مواجهتها، فهذه النظرية تحاول نمذجة العلاقة القائمة بين مستوى سمة معينة لدى الفرد التي يقيسها اختبار معين واستجابته لمفردة من مفردات الاختبار (علام،٢٠٠٥:٥٣).

فقد أشار علاونة (٢٠١٦) أن هناك فروقاً ذات دلالة إحصائية في معاملات ثبات الاختبار التحصيلي تعزى لمتغير شكل الفقرة لصالح اختبار الاختيار من متعدد، كما تبين على عدم وجود فروق ذات دلالة إحصائية في معاملات الصعوبة ومعاملات التمييز لفقرات الاختبار تعزى لمتغيرات شكل الفقرة، وتوصي هذه الدراسة إلى إجراء دراسة مماثلة لها باستخدام نظرية القياس الحديثة.

وأشارت الذنيبات (٢٠١٣) أنه لا يوجد فروق ذات دلالة إحصائية بين اختبار الاختيار من متعدد واختبار الصواب والخطأ يعزى للثبات والصدق والتمييز بينما يوجد فروقا ذات دلالة إحصائية بين الاختيار من متعدد والصواب والخطأ يعزى لمعامل الصعوبة وذلك لصالح اختبار الصواب والخطأ، وقد أوصت هذه الدراسة إجراء دراسات تقارن بين الخصائص السيكومترية للاختبارات الموضوعية وفق النموذج الثنائي لنظرية استجابة الفقرة.

وقد أظهرت نتائج دراسة قام بها أبو جراد (٢٠١٢) أن معاملات ثبات تقديرات الصعوبة للفقرات والأفراد في اختبار التكميل أعلى منها لاختبار الاختيار من متعدد.

وقد أجرى موباليغ وباراتي (Mobalegh&Barati,2012) دراسة تبين التحليلات الكمية لنتائج الاختبارات أن الاختبارات بأشكاله المختلفة لم تختلف اختلافا كبيراً في معامل الثبات.

وقد أظهرت نتائج دراسة تاسديمير(Tasdemir,2010) أن اختبارات الصواب والخطأ ليست أسهل من الاختيار من المتعدد وأن اختبارات الاختيار من متعدد لا تتفوق على اختبارات الصواب والخطأ في تحصيل الطلاب ولا يوجد فروق في معامل الصعوبة ومعامل التمييز بين الاختبارين.

كما أظهرت دراسة طعامنة (٢٠٠٧) عدم وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقديرات معالم الصعوبة للفقرات وكذلك بين معاملات ثبات الأفراد تعزى لعدد البدائل.

وأظهرت دراسة الكحلوت (٢٠٠٢) أن معاملي ثبات اختبار التكميل المحسوبين بطريقة الإعادة وباستخدام معادلة كرونباخ ألفا أكبر من معاملي ثبات اختبار الاختيار من متعدد كما بينت النتائج أن معاملات الصعوبة لاختبار الاختيار من متعدد أكبر من متوسط معاملات الصعوبة لاختبار التكميل، وأن متوسط معاملات التمييز لاختبار التكميل أكبر من متوسط معاملات التمييز لاختبار الاختيار من متعدد.

وأشار جاد الله (Gadalla,1999) أظهرت نتائجها وجود فروق ذات دلالة إحصائية بين متوسط علامات الاختبار في الاختيار من متعدد وبين علامات الاختبار ذي الفقرات القصيرة والمحددة، حيث كان في علامات اختبار الاختيار من متعدد أعلى منه في علامات اختبار التكميل ذي الإجابة القصيرة والمحددة وذلك للصفين الثاني والثالث، كما أشارت نتائج الدراسة إلى عدم وجود فروق ذات دلالة إحصائية لصيغة الفقرات على أداء الطلبة في الصفوف من الرابع إلى السادس.

لقد شاع استخدام شكلين من الأسئلة، الاختيار من متعدد والصواب والخطأ، ولكل من الشكلين مزايا وعيوب نتج عنها وجود جدل حول استخدامات هذين الشكلين وأيهما أفضل، حيث إنه من المهم عند اختيار شكل معين من الأسئلة في اختبار ما الأخذ بعين الاعتبار بعدة عوامل أهمها زمن وطريقة تصحيح الاختبار، ومدى ملاءمة شكل الأسئلة لأهداف الاختبار والنتائج التعليمية التي سيقيسها الاختبار، بالإضافة إلى مدى تمتع أي من الشكلين بخصائص سيكومترية أفضل، وقد أشارت نتائج الدراسات التي أجريت للمقارنة بين أشكال الفقرة من حيث خصائصهما السيكومترية إلى اختلاف في نتائجها وبين أشكال الفقرة التي قورنت بينها بالإضافة إلى استخدم نظرية القياس المستخدمة في الدراسة، وقد يحدد هذا البحث الشكل

الأكثر فاعلية والأفضل والمناسب بين الشكلين من خلال الخصائص السيكومترية للفقرة وللاختبار، وتأتي أهمية هذا البحث في المقارنة بين الخصائص السيكومترية لفقرات الاختيار من متعدد وفقرات الصواب والخطأ وفقًا لنموذج ثنائي المعلم حيث يوفر هذا النموذج وحدة قياس واحدة لكل من الأفراد والفقرات مما يعطي دقة أعلى في إجراء المقارنات.

مشكلة البحث وأسئلته:

يدور جدل كبير حول الكيفية التي تطور بواسطتها عملية بناء الاختبارات حول الاستخدام الأفضل لشكل الفقرة، أما شكل الاختيار من متعدد ووجود مموهات قد تؤثر في إجابات الطالب وتشتته عن بقية الفقرات أو شكل الصواب والخطأ، ونظراً لشيوع استخدام كل من فقرات الاختيار من متعدد وفقرات الصواب والخطأ وانتشارها في الميدان التربوي في قياس تحصيل الطلاب أنهما يقدمان عمليات عقلية مختلفة.

وقد أوصت الدراسات السابقة إلى إجراء دراسة للمقارنة بين أشكال الفقرة من حيث خصائصهم السيكومترية، والتي أشارت إلى اختلاف في نتائجها وبين أشكال الفقرة التي قورنت بينها بالإضافة إلى اختلاف نظرية القياس المستخدمة في الدراسة، وافتقرت دراسة هذين الشكلين وفق نظرية استجابة الفقرة بنماذجها الثلاثة، ويختلف هذا البحث عن الدراسات السابقة بأنه يقارن بين شكلين من أشكال الفقرة الاختيار من متعدد والصواب والخطأ، بناء على افتراضات ونماذج وإحصائيات النظرية الحديثة في القياس والتي لم تتناولها الدراسات السابقة مع هذين الشكلين ومن هنا جاء هذا البحث للمقارنة بين هذين الشكلين ومدى ملاءمة كل منهما للغرض المنشود من استخدامهما والمساعدة في تقليص حجم الاختلاف والتناقض حولهما.

وقد تناول هذه البحث مقارنة الخصائص السيكومترية من الثبات التجريبي للاختبار وتحليل فقرات الشكلين على النموذج اللوجستي ثنائي المعلم لشكلين من فقرات الاختبار الموضوعية الاختيار من متعدد وفقرات الصواب والخطأ، وقد أجاب هذا البحث عن الأسئلة التالية:

١. هل يوجد فرق دال إحصائيًا بين المتوسطات الحسابية لمعلمة صعوبة الفقرة تعزى لاختلاف شكل الفقرة (الاختيار من متعدد – الصواب والخطأ)؟
٢. هل يوجد فرق دال إحصائيًا بين المتوسطات الحسابية لمعلمة تمييز الفقرة تعزى لاختلاف شكل الفقرة (الاختيار من متعدد – الصواب والخطأ)؟
٣. هل يوجد فرق دال إحصائيًا بين معاملات الثبات التجريبي يعزى لاختلاف شكل الفقرة (الاختيار من متعدد – الصواب والخطأ)؟

أهداف البحث:

يهدف هذا البحث عن الكشف عن أثر شكل فقرات الاختبار في الخصائص السيكومترية وفقاً لنظرية استجابة الفقرة، لفقرات الأكثر شيوعًا واستخدامًا لدى الكثير من المعلمين والمعلمات في المدارس: فقرات الاختيار من متعدد مع فقرات اختبار الصواب والخطأ التي تناظرها في المحتوى، وذلك باستخدام أشهر نموذج ثنائي المعلم، والكشف عن شكل الاختبار الأفضل من حيث ثباته التجريبي وتمييز وصعوبة فقراته.

أهمية البحث:

١. تأتي أهمية هذا البحث في المقارنة بين الخصائص السيكومترية لفقرات الاختيار من متعدد وفقرات الصواب والخطأ وفقًا لنموذج ثنائي المعلم.
٢. الكشف عن إمكانية المفاضلة بين شكلين من الاختبارات يعدان من أكثر الاختبارات استعمالاً واختيار الأفضل منهما.
٣. الإسهام في تحسين العملية التعليمية، فنتائج هذا البحث ستمكن الباحث من أن يوصي باستخدام الشكل المناسب من الفقرات في الاختبارات.
٤. اختلاف نتائج الدراسات السابقة من حيث إمكانية تفوق أي شكل من أشكال الفقرة، وكذلك توصياتها على إجراء المقارنة بين أشكال الفقرة وفقًا لنظرية استجابة الفقرة فكان ذلك باعثًا على إجراء هذا البحث.
٥. قد تشكل نتائج هذا البحث منطلقًا لإجراء المزيد من الدراسات للمقارنة بين أشكال أخرى من الاختبارات وملاحظة أثرها على الخصائص السيكومترية للاختبار.

مصطلحات البحث:

الخصائص السيكومترية: 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)(الزهراني،٢٠١٠:٢٠٥) ، ويقصد بالخصائص السيكومترية حسب نظرية الاستجابة الفقرة: معلم الصعوبة، ومعلم التمييز، ومعلم التخمين للفقرة.

ويعرف إجرائيًا: عرض معاملات تمييز وصعوبة الفقرات وفق نظرية استجابة الفقرة.

الاختبارات Tests: وسيلة منظمة لتقويم قدرات الطلاب ولتحديد مستوى تحصيل المعلومات والمهارات عندهم في مادة دراسية تعلموها مسبقًا، وذلك من خلال إجابتهم على مجموعة من الفقرات التي تمثل محتوى مادة دراسية. (دعمس،٢٠٠٨:٦٥)

ويعرف إجرائيًا: بأنه مجموعة من الفقرات الموضوعة لقياس التحصيل في مادة الحاسب وتقنية المعلومات١ للصف الأول ثانوي، والذي يتكون من (٢٤) فقرة في صورته النهائية.

الاختيار من متعدد: يتألف سؤال الاختيار من متعدد من متن السؤال يحدد فيه المعلم المشكلة التي قد تصاغ في هيئة سؤال أو عبارة استفهامية أو شكل عبارة ناقصة، ويلي هذه المقدمة قائمة من الحلول المحتملة تسمى بدائل وأحد هذه البدائل هو الإجابة الصحيحة أو الأكثر صحة ثم يطلب من الطالب أن يقرأ السؤال وقائمة البدائل قراءة متعمقة ثم ينتقي البديل الصحيح لكل سؤال. (مراد،وسليمان٢٠٠٥:١٨٦)

ويعرف إجرائيًا: عدد من الفقرات المعدة لقياس التحصيل في الحاسب وتقنية المعلومات١ للصف الأول ثانوي، ويتطلب من الطالب اختيار البديل الصحيح من بين البدائل الخاطئة.

الصواب والخطأ: هي عبارة عن جمل قابلة لأن تكون صحيحة أو غير صحيحة وعلى الطالب أن يحدد فيما إذا كانت صحيحة أو غير صحيحة. (الطريري،١٩٩٧:٣١١)

ويعرف إجرائيًا: عدد من الجمل الخبرية المعدة لقياس التحصيل في الحاسب وتقنية المعلومات١ للصف الأول ثانوي قابلة لأن تكون صحيحة أو غير صحيحة وعلى الطالب أن يحدد فيما إذا كانت صحيحة أو غير صحيحة

نظرية استجابة الفقرة: وهي مجموعة الطرق الإحصائية التي تستخدم في حساب معالم الفقرات والأفراد الخاصة بالاختبار من صعوبة وتمييز وتخمين ودالة معلومات، والمكونة من نماذج أحادية البعد هي: (نموذج راش أحادي البارامتر، نموذج لورد ثنائي البارامتر، نموذج بيرنيوم ثلاثي البارامتر). (زكري،٢٠٠٨:١٤) (زكري،١٤٢٩:١٤)

ويعرف إجرائيًا: هو استخدام أسس نظرية استجابة الفقرة في حساب معاملات الفقرة (الصعوبة والتمييز) وفق النموذج اللوجستي ثنائي المعلم والثبات التجريبي للاختبار.

النموذج اللوجستي ثنائي المعلم Two Parameters Logistic Model: ويسمى بنموذج" لورد" وفيه يتم إضافة معلم تمييز الفقرة (ai) إضافة إلى معلم صعوبة الفقرة (bi) في نموذج أحادي المعلم لكل فقرة، كما يفترض هذا النموذج انعدام التخمين (ci). (زكري،٢٠٠٨:٥٠).

ويعرف إجرائيًا: هو حساب معالم الفقرة (الصعوبة والتمييز) وفق النموذج اللوجستي ثنائي المعلم أحد نماذج نظرية استجابة الفقرة.

حدود البحث:

الحدود المكانية: تم تطبيق أداة البحث في المدارس الثانوية في مدينة تبوك.

الحدود الزمانية: تم تطبيق أداة البحث في الفصل الدراسي الأول للعام الدراسي ١٤٣٨-١٤٣٩هـ

الحدود الديموغرافية:

تم تطبيق أداة البحث على طلاب الصف الأول ثانوي، في مادة الحاسب وتقنية المعلومات١ الوحدة الثانية (برامج إدارة المواقع على الإنترنت) "الجزء النظري".

منهجية البحث وإجراءاته:

منهجية البحث:

استخدم الباحث المنهج شبه التجريبي، وطبقت أدوات البحث وفقاً لتصميم القياسات المتكررة ويعد هذا التصميم أقوى من تصميم المجموعات العشوائية الذي اعتمدته معظم الدراسات السابقة. (أبو جراد،٢٠١٢:٣٩٠)

مجتمع البحث:

طلاب الصف الأول ثانوي بالمدارس الحكومية بمدينة تبوك للعام الدراسي ١٤٣٨-١٤٣٩هـ، الفصل الدراسي الأول، وبعد الرجوع للسجلات والتقارير الرسمية لإدارة التخطيط والتطوير بالإدارة العامة لتعليم بمنطقة تبوك تبين لنا أن عدد طلاب أول ثانوي بالمدارس الحكومية بمدينة تبوك (٣٢٦٨) طالبًا بتاريخ ١٤٣٩/٤/٢هـ والذي يمثل هذا العدد حجم مجتمع البحث.

عينة البحث:

تم أخذ عينة البحث من مجتمع البحث باستخدام العينة العشوائية العنقودية من المدارس التي داخل نطاق مدينة تبوك، وبلغ عدد أفراد العينة (٤٢١) طالبًا.

حيث اشتملت عينة البحث ما هو موضح في جدول رقم (١):

جدول (١): توزيع أفراد عينة البحث من حيث المدارس وعدد الشعب والطلاب

م	اسم المدرسة	عدد الشعب	عدد الطلاب	أرقام الطلاب	النسبة من العينة الكلية
١	ثانوية تبوك	٣	٧٧	١٠٠ - ١٧٦	%١٨,٢٩
٢	ثانوية ابن حزم	٢	٦٥	١٧٧ – ٢٤١	%١٥,٤٤
٣	ثانوية أسامة بن زيد	٥	١٥٩	٢٤٢ – ٤٠٠	%٣٧,٧٧
٤	ثانوية مجمع الأمير فهد	٢	٦١	٤٠١ – ٤٦١	%١٤,٤٩
٥	ثانوية يوسف القاضي	٢	٥٩	٤٦٢ – ٥٢٠	%١٤,٠١
	المجموع	١٤	٤٢١	١٠٠ - ٥٢٠	%١٠٠,٠٠

أداة البحث:

تم بناء اختبار تحصيلي في مادة (الحاسب وتقنية المعلومات للصف الأول ثانوي) في الوحدة الثانية (تصميم وإدارة المواقع والشبكات الاجتماعية) من المقرر، وتم تحديد هذه المادة نظرًا لتدريس الباحث هذه المادة خلال السنوات الماضية وتمكنه من المادة ومحتوى مفرداتها، وتكون الاختبار من شكلين: الأول شكل اختيار من متعدد من أربعة بدائل والثاني من نوع الصواب والخطأ، حيث تساوى الشكلان بعدد الفقرات والمتن، وقد تم بناء هذا الاختبار بناءً على الخطوات المتبعة في بناء مثل هذه الاختبارات وهي:

تحديد الغرض من الاختبار:

تم بناء الاختبار على أساس أنه اختبار تحصيلي في صورتين (الاختيار من متعدد) و(الصواب والخطأ) لتقييم طلبة الصف الأول ثانوي في منهج الحاسب وتقنية المعلومات كاختبار فترة أولى.

تحليل محتوى المادة:

تم تحليل محتوى مادة الحاسب وتقنية المعلومات للصف الأول الثانوي في الوحدة الثانية (تصميم وإدارة المواقع والشبكات الاجتماعية)، وتم صياغة الأهداف السلوكية لكل موضوع في الوحدة حسب عناصر المحتوى ومستويات نواتج التعلم المعرفية، حيث تم صياغة (٢٤) هدفًا تعليميًا سلوكيًا والتي يفترض تحققها عند انتهاء الطالب من دراسة الوحدة.

بناء جدول المواصفات:

تم بناء جدول المواصفات بتوزيع الأهمية النسبية لكل موضوع في الوحدة التي تم فيها الاختبار اعتمادًا على عدد الأهداف التي يلزم المعلم شرحها لكل موضوع.

بناء الفقرات:

تم صياغة بناء الفقرات على شكل اختبار اختيار من متعدد (أربع بدائل) وبناءً على هذا الشكل تم بناء شكل اختبار الصواب والخطأ وذلك بحذف البدائل ووضع بديل صحيح أو خاطئ وإعادة صياغة الفقرة على شكل جملة خبرية سليمة من الأخطاء، وبهذا يكون مساويًا للشكل الأول من حيث عدد الفقرات، وقد بلغ عدد الفقرات في صورتها الأولية (٣٤) فقرة لقياس أهداف المحتوى، وقد رُعِيت الإرشادات والأسس المتبعة في بناء الفقرات في كلا الشكلين.

تحكيم الاختبار والتحقق من صدق المحتوى:

ولغرض تحكيم الأداة والتحقق من صدق المحتوى، تم تزويد كل محكم بفقرات الاختبارين تتضمن أهداف المادة وأمام كل هدف الفقرة المرتبطة بالهدف على شكلين (الاختيار من متعدد والصواب والخطأ) حيث طُلب منهم تحكيم الفقرات من خلال استمارة التحكيم وتحديد الفقرات غير المناسبة وإجراء التعديلات اللازمة.

وقد تم عرض الفقرات على (١٢) محكم من ذوي الخبرة: (٣) متخصصين في القياس والتقويم و(٥) من مشرفي المادة و(٤) من معلمي المادة من ذوي الخبرة وطُلِب منهم إبداء ملاحظاتهم في الفقرات من حيث سلامتها اللغوية ودقتها في قياس الأهداف التي وضعت من أجلها وفاعلية المموهات والتأكد من صدق المحتوى للاختبار.

ومن خلال ملاحظات المحكمين تم تعديل الفقرات وتحسينها وإخراجها بصورة واضحة وسليمة من الأخطاء، واعتماد الشكل النهائي لفقرات الاختبارين، وتجهيزها لتطبيقها على العينتين الاستطلاعية والأساسية.

التطبيق الأولي للاختبار:

بعد حصول الباحث على خطاب تسهيل المهمة تم تطبيق الاختبار المكون من (٣٤) فقرة على شكلين من أشكال الفقرة: الاختيار من متعدد، والصواب والخطأ على عينة استطلاعية مكونة من (٥٠) طالبًا وذلك للتأكد من وضوح الفقرات وتقدير الزمن اللازم للإجابة (حيث أنهى جميع الطلبة الإجابة عن فقرات الاختبار بشكليه بعد إعطائهم الوقت الكافي للإجابة بمقدار خمس وعشرين دقيقة لشكل الاختيار من متعدد، وخمسة عشر دقيقة

لشكل الصواب والخطأ) وحساب معاملات الصعوبة والتمييز لفقرات شكلي الاختبار وكذلك لحساب معامل الثبات وذلك حسب النظرية الكلاسيكية، وتم توزيع الاختبار في الشعبة الواحدة بشكليه.

تصحيح الأداة المطبقة على العينة الاستطلاعية:

تم إعطاء لكل طالب رقم موحد بين الشكلين وتم إدخال إجابات الطلاب إلى برنامج الإكسل لكل من الشكلين، وتم تصحيح إجابات الطلاب من خلال البحث والاستبدال من نفس برنامج (استبدال الإجابة الصحيحة برقم (١) والإجابة الخاطئة بالرقم (٠).

تحليل فقرات الأداة المطبقة على العينة الاستطلاعية:

بعد تطبيق الاختبار وتصحيحه تم حساب معاملات الصعوبة والتمييز لفقرات شكلي الاختبار من خلال برنامج البايلوج وثبات الاختبار معامل ألفا كرونباخ من خلال برنامج SPSS وذلك حسب النظرية الكلاسيكية، ويوضح جدول رقم (٢) معاملات الصعوبة والتمييز التي حسبت لفقرات الاختبار من متعدد والصواب والخطأ:

جدول (٢): معاملات الصعوبة والتمييز لفقرات الاختبار من متعدد والصواب والخطأ

الفقرة	شكل الاختيار من متعدد		شكل الصواب والخطأ	
	معامل الصعوبة	معامل التمييز	معامل الصعوبة	معامل التمييز
١	٠,٤٢	٠,٤٩٩	٠,٦٢	٠,٣٧٧
٢	٠,٦٤	٠,٦٢٧	٠,٨٠	٠,٠٨٠
٣	٠,٧٠	٠,٥١٣	٠,٨٠	٠,٤٦٩
٤	٠,٥٤	٠,٥٣٩	٠,٤٨	٠,٥٣٨
٥	٠,٦٦	٠,٥٠٥	٠,٦٦	٠,٤٣٠
٦	٠,٤٤	٠,٦٠٨	٠,٢٦	٠,٠٢٨
٧	٠,٦٦	٠,٥١٣	٠,٤٤	٠,٣٥٥
٨	٠,٦٤	٠,٣١٨	٠,٨٨	٠,٠٠٤
٩	٠,٦٤	٠,٤٩٦	٠,٦٢	٠,٥٦٤
١٠	٠,٧٠	٠,٥٢٢	٠,٨٢	٠,٣٢٢
١١	٠,٩٢	٠,٣٧٤	٠,٨٨	٠,٢٨٠
١٢	٠,٦٨	٠,٣٨١	٠,٦٢	٠,٥٧٥
١٣	٠,٦٤	٠,٦٦٠	٠,٧٤	٠,٤٠٠
١٤	٠,٩٢	٠,٥١٢	٠,٨٦	٠,٢٣٩
١٥	٠,٨٤	٠,٣٨٣	٠,٩٢	٠,٢٨١
١٦	٠,٧٦	٠,٥٤٥	٠,٨٤	٠,٠٣١
١٧	٠,٤٦	٠,٥٩٢	٠,٥٦	٠,٥٠٩
١٨	٠,٨٤	٠,٥٩٢	٠,٩٤	٠,٢٤٣
١٩	٠,٦٠	٠,٦٦٧	٠,٨٢	٠,٥٢٤
٢٠	٠,٦٤	٠,٧٠٢	٠,٨٤	-٠,٠٦٥
٢١	٠,٧٠	٠,٢٨٢	٠,٦٢	٠,٣٧٧
٢٢	٠,٦٦	٠,٤٨٠	٠,٧٤	٠,٢٩٩
٢٣	٠,٧٢	٠,٢٣٥	٠,٧٨	٠,٢٩٩
٢٤	٠,٧٠	٠,٥١٣	٠,٧٨	٠,٥٧٦
٢٥	٠,٥٢	٠,٤٤٠	٠,٤٨	٠,٥٢٧
٢٦	٠,٥٠	٠,٣٢١	٠,٥٤	٠,٤٧٤
٢٧	٠,٦٢	٠,٤٠١	٠,٤٦	٠,٥٥٢
٢٨	٠,٤٢	٠,٦٠٢	٠,٧٢	٠,٥١٦
٢٩	٠,٣٦	٠,٦٧٨	٠,٥٨	٠,٤٦٠
٣٠	٠,٥٠	٠,٥٦٦	٠,٧٢	٠,٤٣٠
٣١	٠,٣٨	٠,٦٧٧	٠,٧٢	٠,٢٤٨
٣٢	٠,٤٢	٠,٤٨٤	٠,٦٦	٠,٠٧٦
٣٣	٠,٤٠	٠,٦٦٢	٠,٤٦	٠,٥٤١
٣٤	٠,٤٥	٠,٢٩٧	٠,٤٠	٠,٢٩٣
المتوسط	٠,٦١	٠,٥٠	٠,٦٨	٠,٣٥

يتضح من الجدول رقم (٢) أن معاملات الصعوبة في شكل الاختيار من متعدد تراوحت بين (٠,٣٨) و (٠,٩٢) بمتوسط حسابي قدره (٠,٦١) بينما تراوحت قيم معاملات التمييز في نفس الشكل بين (٠,٣٠) و (٠,٧٠) بمتوسط حسابي قدره (٠,٥٠).

كم يوضح الجدول رقم (٢) أن معاملات الصعوبة في شكل الصواب والخطأ تراوحت بين (٠,٢٦) و(٠,٩٤) بمتوسط حسابي قدره (٠,٦٨) بينما تراوحت قيم معاملات التمييز في نفس الشكل بين(-٠,٠٦) و(٠,٥٨) بمتوسط حسابي قدره (٠,٣٥).

ويذكر عوده (٢٨٧:٢٠١٠) أن أي فقرة ضمن توزيع لمعاملات الصعوبة يتراوح بين (٠,٢٠) و(٠,٨٠) بمتوسط مقداره (٠,٥٠) يمكن أن تكون مقبولة، وينصح بالاحتفاظ بها في ملف أو بنك الفقرات.

وتتراوح قيمة معامل التمييز بين (-١) و(+١)، فإذا كانت قيمته (٠,٤٠) أو أكبر فإن هذا يكون دليلاً على أن المفردة تميز بين المجموعتين بدرجة جيدة، وإذا كانت بين (٠,٢٠–٠,٤٠) فإن التمييز يكون لا بأس به وإذا قلت عن (٠,٢٠) يكون التمييز ضعيفاً، وإذا كانت قيمة التمييز صفراً فإن هذا يعني أن المفردة في غاية السهولة أو الصعوبة وينبغي حذف هذه المفردة من الاختبار. (علام،٢٨٩:٢٠١١)

ومن جدولي (٢) فقد تم حذف (١٠) فقرات للأسباب التالية:

- فقرة (٦، ٣٢، ٢) من شكل الصواب والخطأ لانخفاض معامل التمييز لشكل، وبالتالي تم حذفها من شكل الاختيار من متعدد.
- فقرة (٨، ١٦، ٢٠) لشكل الصواب والخطأ وذلك لارتفاع معامل صعوبتها وانخفاض تمييزها، كما تم حذفها من شكل الاختيار من متعدد.
- فقرة (١١، ١٤، ١٥، ١٨) لارتفاع معامل صعوبتها في كلا الشكلين.

كما تم حساب معامل ثبات ألفا كرونباخ لأداة البحث ويوضح جدول رقم (٣) قيم معاملات الثبات:

جدول(٣): قيم معاملات الثبات لشكلي الاختبار

شكل الاختبار	قيمة معامل ألفا كرونباخ
الاختيار من متعدد	٠,٨٦
الصواب والخطأ	٠,٧٨

ويتضح من جدول رقم (٣) أن معاملات الثبات لشكلي الاختبار باستخدام معامل ألفا كرونباخ بلغت (٠,٨٦) لشكل الاختيار من متعدد ، و(٠,٧٨) لشكل الصواب والخطأ.

إعداد الصورة النهائية لفقرات الاختبار لكلا الشكلين:

بعد تحليل فقرات اختبار العينة الاستطلاعية تم إعداد الصورة النهائية لفقرات الاختبار الاختيار من متعدد واختبار الصواب والخطأ بعد تحديد الفقرات التي تم حذفها ووضع تعليمات الاختبار وإخراجها بالشكل النهائي.

تطبيق الاختبار بصورته النهائية:

تم تطبيق الاختبار على عينة البحث وبفقرات الاختبار بشكليه بصورته النهائية والبالغ عددها (٢٤) فقرة حيث تم الآتي:

١. الحصول على موافقة إدارة التعليم لإجراء البحث وتطبيق الاختبار.
٢. الحصول على أعداد طلاب الصف الأول ثانوي في مدينة تبوك.
٣. اختيار عينة البحث.
٤. تمت زيارة المدارس التي تم تطبيق الاختبار فيها لتحديد وقت الاختبار والجلوس مع معلمين الصف الأول ثانوي لمادة الحاسب وتقنية المعلومات عن آلية تطبيق الاختبار وإعطاء الأمر للطلاب بالاستعداد للاختبار كمنتصف فصلي والجدية في الاختبار واعتماد نتائجه كتقويم للطلبة في المادة.
٥. تم توضيح طريقة الاختبار للمعلمين الذين شاركوا في تطبيق الاختبار والحرص على تنظيم الطلاب في كل لجنة اختبار.
٦. تم توزيع الاختبار في الشعبة الواحدة بشكليه حيث بدأ صف بأسئلة الاختبار من متعدد ويبدأ الصف الأخر بأسئلة الصواب والخطأ وبعد الانتهاء من إجابة هذه الأسئلة أعطى الطلبة الذين بدأوا بأسئلة الاختيار من متعدد أسئلة الصواب والخطأ، وأعطى الطلبة الذين بدأوا بأسئلة الصواب والخطأ أسئلة الاختيار من متعدد، مع الحرص الشديد على عدم الجمع بين الشكلين عند الطالب وعدم السماح له بالغش.
٧. تم جمع أوراق الاختبار وإعطاء لكل طالب رقم موحد بين الشكلين وتم إدخال إجابات الطلاب إلى برنامج الإكسل، وتم تصحيح إجابات الطلاب من خلال البحث والاستبدال من نفس برنامج (استبدال الإجابة الصحيحة برقم (١) والإجابة الخاطئة بالرقم (٠) وتصديرها إلى البرامج الإحصائية الأخرى (SPSS.19)،(BILOGMG 3.0) وذلك لعمل التحاليل الإحصائية المطلوبة.

متغيرات البحث:

المتغير المستقل: شكل الفقرة في الاختبار اختيار من متعدد (ذو أربع بدائل) - اختبار الصواب والخطأ.

المتغيرات التابعة: خصائص الاختبار السيكومترية والمتمثلة بمعالم الفقرة بـ: معلمة الصعوبة ومعلمة التمييز، ومعامل الثبات التجريبي للاختبار.

الأساليب الإحصائية:

بعد إجراء تطبيق الاختبارين وإدخال إجابات المفحوصين إلى البرامج الإحصائية ومن أجل الإجابة عن أسئلة البحث ، فقد تم استخدام المعالجات الإحصائية التالية:

١. معامل ثبات ألفا لكرونباخ.
٢. نسبة الذين أجابوا إجابة صحيحة لكل فقرة (معامل الصعوبة للنظرية التقليدية).
٣. معامل بايسيريال (معامل التمييز للنظرية التقليدية).
٤. اختبار التحليل العاملي للتحقق من افتراض أحادية البعد والاستقلال الموضعي.
٥. مربع كاي لمطابقة الفقرات.
٦. اختبار(ت) (T) Sample test paired لعينتين مرتبطتين لدراسة الفروق في الأوساط الحسابية لمتوسطات الخصائص السيكومترية للفقرة.
٧. معامل الصعوبة ومعامل التمييز للفقرات وثبات الاختبار التجريبي وفق نظرية استجابة الفقرة .
٨. اختبار فلدت (Feldt & Alswalmeh test) لفحص فرضية تساوي معاملي ثبات مرتبطين (Alswalmeh&Feldt,1994:187)

عرض النتائج ومناقشتها وتوصياتها:

التحقق من افتراضات نظرية استجابة الفقرة:

أولاً: أحادية البعد:

تم التحقق من افتراض أحادية البعد باستخدام التحليل العاملي للمكونات الرئيسية وباستخدام الرزمة الإحصائية (SPSS.19) لشكلي الاختبار:

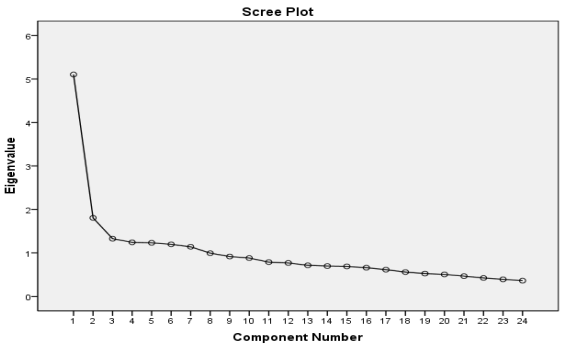
١. **اختبار الاختيار من متعدد: يبين الجدول رقم (٤) قيم الجذر الكامن ونسبة التباين المفسر للعوامل المستخلصة.**

جدول (٤): قيم الجذر الكامن ونسبة التباين المفسر للعوامل المستخلصة ونسبة التباين التراكمية لكل عامل من العوامل في العينة الرئيسية في نموذج الاختيار من متعدد

العامل	الجذر الكامن	التباين المفسر %	النسبة التراكمية للتباين %
١	٥,١٠	٢١,٢٦	٢١,٢٦
٢	١,٨١	٧,٥٢	٢٨,٧٨
٣	١,٣٣	٥,٥٣	٣٤,٣١
٤	١,٢٤	٥,١٨	٣٩,٤٩
٥	١,٢٣	٥,١٣	٤٤,٦٢
٦	١,٢٠	٤,٩٩	٤٩,٦١
٧	١,١٤	٤,٧٥	٥٤,٣٦

يلاحظ من جدول رقم (٤) أن قيمة الجذر الكامن للعامل الأول (٥,١) وأنه فسر ما نسبته (٢١,٢٦%) من التباين الكلي للاختبار، بينما كانت قيمة الجذر الكامن للعامل الثاني (١,٨١) وأنه فسر ما نسبته (٧,٥٢%) من التباين الكلي للاختبار، كما نلاحظ أن قيمة التباين المفسر للعامل الأول أكبر من (٢٠) مما يعني أن الاختبار يقيس بعدًا واحدًا. (Georgiev,2008:10)

ويلاحظ من خلال استخدام مخطط سكري (Scree Plot) للعوامل مع الجذور الكامنة كما يظهر في الشكل رقم (١) أن التمثيل البياني يظهر فيه العامل الأول ينفرد بقيمة جذر كامن مرتفعة نسبيًا مقارنة مع العوامل الأخرى التي تظهر بجذور كامنة صغيرة نسبيًا ومتقاربة وهذا أيضًا يعتبر مؤشرًا على أحادية البعد.



شكل(١): مخطط سكري لقيم الجذور الكامنة للعوامل المستخلصة من التحليل العاملي لاختبار الاختيار من متعدد

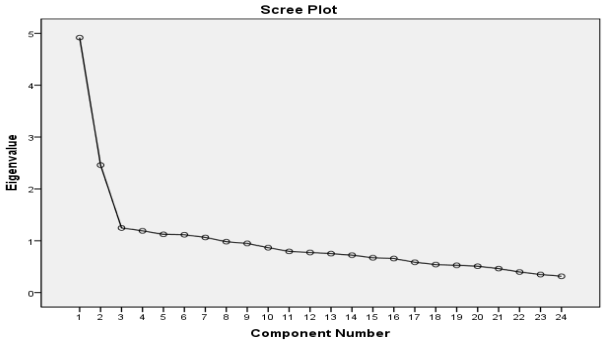
٢. **اختبار الصواب والخطأ: يبين الجدول رقم (٥) قيم الجذر الكامن ونسبة التباين المفسر للعوامل المستخلصة.**

جدول(٥): قيم الجذر الكامن ونسبة التباين المفسر للعوامل المستخلصة ونسبة التباين التراكمية لكل عامل من العوامل في العينة الرئيسية في نموذج الصواب والخطأ

العامل	الجذر الكامن	التباين المفسر %	النسبة التراكمية للتباين %
١	٤,٩٢	٢٠,٤٩	٢٠,٤٩
٢	٢,٤٦	١٠,٢٥	٣٠,٧٣
٣	١,٢٥	٥,٢٠	٣٥,٩٣
٤	١,١٩	٤,٩٧	٤٠,٩٠
٥	١,١٣	٤,٦٩	٤٥,٥٩
٦	١,١٢	٤,٦٦	٥٠,٢٥
٧	١,٠٧	٤,٤٤	٥٤,٦٨

يلاحظ من جدول رقم(٥) أن قيمة الجذر الكامن للعامل الأول(٤,٩٢) وأنه فسر ما نسبته (٢٠,٤٩%) من التباين الكلي للاختبار، بينما كانت قيمة الجذر الكامن للعامل الثاني (٢,٤٦) وأنه فسر ما نسبته (١٠,٢٥%) من التباين الكلي للاختبار كما نلاحظ أن قيمة التباين المفسر للعامل الأول أكبر من (٢٠) مما يعني أن الاختبار يقيس بعداً واحداً.

ويلاحظ من خلال استخدام مخطط سكري للعوامل مع الجذور الكامنة كما يظهر في الشكل (٢) أن التمثيل البياني يظهر فيه العامل الأول ينفرد بقيمة جذر كامن مرتفعة نسبيًا مقارنة مع العوامل الأخرى التي تظهر بجذور كامنة صغيرة نسبيًا ومتقاربة وهذا أيضًا يعتبر مؤشرًا على أحادية البعد.



شكل(٢): مخطط سكري لقيم الجذور الكامنة للعوامل المستخلصة من التحليل العاملي لاختبار الصواب والخطأ

ثانياً: الاستقلال الموضعي:

يتحقق هذا الافتراض إذا كانت جميع المفردات تقيس سمة كامنة أو قدرة واحدة، فأحادية البعد تتطلب إذن وجود سمة كامنة واحدة بحيث يكون جميع أزواج المفردات تحقق افتراض الاستقلال، أي أن الاستقلال الموضعي يعد مؤشراً لأحادية البعد، ونظراً لتداخل هذا الافتراض مع افتراض أحادية البعد يمكن التحقق من ملائمة افتراض الاستقلال الموضعي بتحقق افتراض أحادية البعد. (علام،٢٠٠٥:٦٣)

ثالثاً: التحرر من السرعة في الأداء:

تم التحقق من افتراض السرعة في الأداء، حيث أنهى جميع الطلبة الإجابة عن فقرات الاختبار بشكليه بعد إعطائهم الوقت الكافي للإجابة بمقدار خمس وعشرين دقيقة لشكل الاختيار من متعدد، وخمسة عشر دقيقة لشكل الصواب والخطأ، بالإضافة إلى أنه لم يشتكِ أي طالب أثناء أو بعد تطبيق الاختبار بضيق الوقت، وأن إخفاق الأفراد في الإجابة يعود إلى انخفاض قدراتهم وليس إلى تأثير عامل السرعة في الإجابة.

التحقق من مطابقة الفقرات:

تم استخدام برنامج (BILOG- MG) الذي تم اعتماده في البحث في تحليل البيانات وفي تقدير معالم الفقرات والأفراد وفق طريقة الأرجحية العظمى، بحسب النموذج ثنائي المعلمة من نماذج نظرية استجابة الفقرة.

تم فحص مطابقة البيانات من خلال استخدام الإحصائي مربع كاي حيث أظهرت النتائج إلى أن هناك خمس فقرات من الاختبار الاختيار من متعدد كانت ضعيفة المطابقة عند مستوى الدلالة (٠,٠٥) وهي (١٩،٩،٨،٦،٤)، كما أن هناك خمس فقرات من الاختبار الصواب والخطأ كانت ضعيفة المطابقة عند مستوى الدلالة (٠,٠٥) وهي (١٩،٩،٨،٦،٤) ويوضح جدول رقم (٦) قيم اختبار مربع كأي للفقرات ومستوى دلالتها:

جدول (٦): قيم اختبار مربع كأي للفقرات ومستوى دلالتها

رقم الفقرة	الاختيار من متعدد		الصواب والخطأ	
	مربع كأي	مستوى دلالتها	مربع كأي	مستوى دلالتها
١	٨,٠٠	٠,٢٤	٨,٤	٠,٠٨
٢	٥,٤	٠,٢٤	٠,٨٠	٠,٩٤
٣	٧,٨٠	٠,٥٥	٧,٠٠	٠,٤٣
٤	١٥,٢٠	٠,٠٢	٩,٩٠	٠,٠٢
٥	١٢,٥	٠,١٩	١٠,١٠	٠,١١
٦	١٤,٢٠	٠,٠١	١٣,٦٠	٠,٠٠
٧	٠,٦٠	٠,٩٩	٢,٨٠	٠,٧٣
٨	١٩,٩٠	٠,٠٠	٩,٣٠	٠,٠٣
٩	١٨,٢٠	٠,٠١	١٨,٩٠	٠,٠٠
١٠	٥,٨٠	٠,٦٧	٠,٧٠	٠,٩٧
١١	٤,٣٠	٠,٠٩	٧,٥٠	٠,٢٨
١٢	١٠,٢٠	٠,٢٥	٧,١٠	٠,٣١
١٣	٣,٤٠	٠,٩٤	٣,٦٠	٠,٨٣
١٤	٥,٩٠	٠,٥٦	١٠,٤٠	٠,١١
١٥	٦,٦٠	٠,٣٦	٦,٤٠	٠,١٧
١٦	١,٥٠	٠,٩٩	٢,٩٠	٠,٧٢
١٧	٤,١٠	٠,٩١	٧,٧٠	٠,٣٦
١٨	٤,٦٠	٠,٦٠	١,٩٠	٠,٧٦
١٩	٢٠,٨٠	٠,٠١	٢٠,٧٠	٠,٠١
٢٠	١٣,٤٠	٠,٠٦	٩,٩٠	٠,٢٧
٢١	٣,٤	٠,٧٦	٨,٠٠	٠,٣٣
٢٢	٤,٠٠	٠,٨٥	١٤,٠٠	٠,٠٨
٢٣	٩,٩٠	٠,٣٦	١٠,٢٠	٠,١٨
٢٤	٣,٨٠	٠,٩٣	٨,٦٠	٠,٣٨

وبعد حذف الفقرات غير المطابقة والإبقاء على الفقرات المطابقة للنموذج أصبح عدد الفقرات لكل شكل من شكلي الاختبار (١٩) فقرة خضعت للمعالجات الإحصائية المناسبة.

مطابقة المفحوصين للنموذج اللوجستي ثنائي المعلم:

تم التحقق من مطابقة المفحوصين للنموذج اللوجستي ثنائي المعلم من خلال برنامج (BILOG-MG) ولم يتم استبعاد أي مفحوص لشكلي الاختبار.

الإجابة عن أسئلة الدراسة:

السؤال الأول: هل يوجد فرق دال إحصائياً بين المتوسطات الحسابية لمعلمة صعوبة الفقرة تعزى لاختلاف شكل الفقرة (الاختيار من متعدد– الصواب والخطأ)؟

وللإجابة عن هذا السؤال تم تقدير معلمة الصعوبة للفقرات المطابقة للنموذج من خلال إجابات المفحوصين المطابقين للنموذج حيث تم اعتماد نموذج ثنائي المعلم لفقرات اختبار الاختيار من متعدد، والصواب والخطأ، ويوضح جدول رقم (٧) قيم معلمة الصعوبة والخطأ المعياري لفقرات اختبار الاختيار من متعدد والصواب والخطأ.

جدول (٧): قيم معلمة الصعوبة لفقرات اختبار الاختيار من متعدد والصواب والخطأ

م	رقم الفقرة	الاختيار من متعدد		الصواب والخطأ	
		معلمة الصعوبة	الخطأ المعياري	معلمة الصعوبة	الخطأ المعياري
١	١	٠,٥٨-	٠,١١	٠,٢٧-	٠,٠٩
٢	٢	١,١٤-	٠,١٣	١,٥٣-	٠,٢٨
٣	٣	٠,٣١-	٠,١٦	٠,٢٩-	٠,١٣
٤	٥	٢,٧٣-	٠,٧٤	٠,٨٩-	٠,٢٧
٥	٧	١,٨٦-	٠,٣٨	١,٦٧-	٠,٣٠
٦	١٠	٠,٣٢-	٠,١٠	٠,٣٠-	٠,٠٩
٧	١١	٠,٥٥-	٠,١٠	١,١٠-	٠,١٥
٨	١٢	١,٠٨-	٠,٢٦	٠,٧٣-	٠,١٨
٩	١٣	٠,٥١-	٠,٢١	١,٦٣-	٠,٤١
١٠	١٤	١,٨٠-	٠,٤٦	٢,٣٤-	٠,٥٤
١١	١٥	١,٢٣-	٠,١٤	١,٤٤-	٠,١٩
١٢	١٦	٠,٢٨-	٠,١٥	٠,٣٤-	٠,١٣
١٣	١٧	٠,٥١-	٠,١٥	٠,٩٧-	٠,١٨
١٤	١٨	٠,٨٤-	٠,١٣	٠,٣٦-	٠,١٢
١٥	٢٠	٠,١٧	٠,١٤	٠,١٥-	٠,١١
١٦	٢١	٠,٥٧-	٠,١٠	٠,٩٧-	٠,١٥
١٧	٢٢	٠,٠٧-	٠,١٠	٠,٥٩-	٠,٠٨
١٨	٢٣	٠,١٢	٠,١٤	٠,٠٦	٠,١٣
١٩	٢٤	١,١٤	٠,٢٣	٠,٨٤	٠,٢٢

نلاحظ من خلال جدول رقم (٧) تراوحت قيم معلمة الصعوبة كالتالي:

اختبار الاختيار من متعدد: تراوحت قيم معلمة الصعوبة بين ١,١٤ و -٢,٧٣

اختبار الصواب والخطأ: تراوحت قيم معلمة الصعوبة بين ٠,٨٤ و -٢,٣٤

ويرى بيكر (Baker,2001:5) أن الصعوبة تتراوح درجاته من (+∞) إلى (-∞) ، إلا أنَّه ولاعتبارات علمية عادة ما نحدد مدى القدرة (من –٣ إلى + ٣).

ثم حسبت المتوسطات الحسابية لمعلمة صعوبة الفقرة لفقرات شكلي الاختبار الاختيار من متعدد والصواب والخطأ واستخدام اختبار (Sample paired T – Test) لعينيتين مرتبطتين للكشف عن الفروق بين متوسطات معلمة الصعوبة ويوضح جدول (٨) نتائج هذا الاختبار:

جدول (٨): اختبار (T) لعينتين مرتبطتين للكشف عن الدلالة الإحصائية بين متوسطات معلمة الصعوبة لشكلي الاختبار

شكل الاختبار	العدد	المتوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة (T)	مستوى الدلالة Sig.
الاختيار من متعدد	١٩	٠,٦٨-	٠,٨٥	١٨	٠,٥٣٧	٠,٦٠
الصواب والخطأ	١٩	٠,٧٦-	٠,٧٦			

ويتضح من جدول رقم (٨) عدم وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq$ ٠,٠٥) بين متوسطات الصعوبة لشكلي الاختبار.

السؤال الثاني: هل يوجد فرق دال إحصائياً بين المتوسطات الحسابية لمعلم تمييز الفقرة تعزى لاختلاف شكل الفقرة (الاختيار من متعدد– الصواب والخطأ)؟

وللإجابة عن هذا السؤال تم تقدير معلمة التمييز للفقرات المطابقة للنموذج من خلال إجابات المفحوصين المطابقين للنموذج حيث تم اعتماد نموذج ثنائي المعلم لفقرات اختبار الاختيار من متعدد، والصواب والخطأ، ويوضح جدول رقم (٩) قيم معلمة التمييز والخطأ المعياري لفقرات اختبار الاختيار من متعدد والصواب والخطأ.

جدول(٩): قيم معلمة التمييز لفقرات اختبار الاختيار من متعدد والصواب والخطأ

م	رقم الفقرة	الاختيار من متعدد		الصواب والخطأ	
		معلمة التمييز	الخطأ المعياري	معلمة التمييز	الخطأ المعياري
١	١	٠,٨٢	٠,١٣	٠,٩٦	٠,١٤
٢	٢	١,٢٩	٠,٣٥	١,٢٨	٠,٤٤
٣	٣	٠,٤٨	٠,٠٨	٠,٥٧	٠,١٠
٤	٥	٠,٢٤	٠,٠٦	٠,٣٣	٠,٠٧
٥	٧	٠,٨٣	٠,٢٦	٠,٩٣	٠,٢٨
٦	١٠	٠,٧٨	٠,١٠	٠,٩٣	٠,١٣
٧	١١	٠,٩٠	٠,٠١٣	٠,٩٣	٠,١٧
٨	١٢	٠,٣٧	٠,٠٧	٠,٤٦	٠,٠٨
٩	١٣	٠,٣٧	٠,٠٨	٠,٣٤	٠,٠٨
١٠	١٤	٠,٣٣	٠,٠٨	٠,٣٥	٠,٠٩
١١	١٥	٠,٩٣	٠,١٧	١,٠٣	٠,٢١
١٢	١٦	٠,٥٠	٠,١٠	٠,٥٩	٠,١٠
١٣	١٧	٠,٥٢	٠,٠٩	٠,٥٦	٠,١٠
١٤	١٨	٠,٧٢	٠,١٢	٠,٦٦	٠,١١
١٥	٢٠	٠,٦١	٠,٠٩	٠,٨٢	٠,١١
١٦	٢١	٠,٩٣	٠,١٤	٠,٧٨	٠,١٣
١٧	٢٢	١,٢٦	٠,١٨	١,٢١	٠,٢٦
١٨	٢٣	٠,٥٠	٠,٠٨	٠,٥٥	٠,٠٨
١٩	٢٤	٠,٥٠	٠,٠٩	٠,٤٢	٠,٠٧

نلاحظ من خلال جدول السابق تراوحت قيم معلمة التمييز كالتالي:

اختبار الاختيار من متعدد: تراوحت قيم معلمة التمييز بين ٠,٢٤ و ١,٢٩

اختبار الصواب والخطأ: تراوحت قيم معلمة التمييز بين ٠,٣٣ و ١,٢٨

ويذكر بيكر(Baker,2001:34) مسميات قيم معلم التمييز للفقرة كما موضح في جدول رقم (١٠) التالي :

جدول(١٠): مسميات قيم معلم التمييز للفقرة

مدى القيم	المسمى اللفظي
صفر	غير مميز
٠,٠١ – ٠,٣٤	منخفض جداً
٠,٣٥ – ٠,٦٤	منخفض
٠,٦٥ – ١,٣٤	متوسط
١,٣٥ – ١,٦٩	مرتفع
١,٧ فأكثر	مرتفع جداً
+ ما لا نهاية	تام

ثم حسبت المتوسطات الحسابية لمعلم تمييز الفقرة لفقرات شكلي الاختبار الاختيار من متعدد والصواب والخطأ واستخدام اختبار (Sample paired T Test) لعينيتين مرتبطتين للكشف عن الفروق بين معاملات التمييز ويوضح جدول (١١) نتائج هذا الاختبار:

جدول (١١): اختبار(T) لعينتين مرتبطتين للكشف عن الدلالة الإحصائية بين متوسطات معلم التمييز لشكلي الاختبار

شكل الاختبار	العدد	المتوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة (T)	مستوى الدلالة .Sig
الاختيار من متعدد	١٩	٠,٦٨	٠,٣٠	١٨	٢,١٣-	٠,٠٤
الصواب والخطأ	١٩	٠,٧٢	٠,٢٩			

ويتضح من جدول رقم (١١) وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (α≤٠,٠٥) بين متوسطات معلمة التمييز لشكلي الاختبار وذلك لصالح اختبار الصواب والخطأ.

السؤال الثالث: هل يوجد فرق دال إحصائياً بين معامل الثبات التجريبي يعزى لاختلاف شكل الفقرة (الاختيار من متعدد – الصواب والخطأ)؟

تم حساب معاملات الثبات التجريبي لكل من الاختبار الاختيار من متعدد واختبار الصواب والخطأ والتي بلغت (٠,٨٦)، (٠,٨٧) على التوالي ولمعرفة الدلالة الإحصائية للفروق بين معاملات الثبات التجريبي لكل الشكلين فقد تم استخدام الصورة المعدلة لاختبار فلدت (Feldt test) لفحص فرضية تساوي معاملي ثبات مرتبطين (Alswalmeh&Feldt,1994:187) ويستعمل في هذه الصورة الإحصائي المعطى في القانون التالي:

$$T = \frac{1-\hat{\rho}_1}{1-\hat{\rho}_2}$$

حيث أن : $\hat{\rho}_1$ = معامل ثبات الأقل. $\hat{\rho}_2$ = معامل ثبات الأكبر.

وتقوم هذه الصورة على كون التوزيع العيني للاختبار الإحصائي فلدت يقترب من التوزيع الفائي (F) بدرجات حرية خاصة تقدر بحجم العينة.

$$T = \frac{1-0.86}{1-0.87} = 1.08$$

نلاحظ أن قيمة الاختبار الإحصائي فلدت القيمة المحسوبة تساوي (١,٠٨) وبدرجة حرية عدد العينة (٤٢١) فإن القيمة الحرجة من جدول (F) تساوي (١,٢٢) (الكيلاني والشريفين،٢٠١٤:٤٩٨) ونلاحظ أن القيمة المحسوبة أقل من القيمة الحرجة عند مستوى الدلالة (α≤٠,٠٥) وبالتالي لا توجد فروق ذات دلالة إحصائية بين معاملات الثبات التجريبي لشكلي الاختبار

مناقشة النتائج وتفسيرها:

مناقشة النتائج المتعلقة بالسؤال الأول:

نص السؤال الأول: **هل هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعلمة صعوبة الفقرة تعزى لاختلاف شكل الفقرة (الاختيار من متعدد–الصواب والخطأ)؟**

أظهرت النتائج المتعلقة بالسؤال الأول أن الفروق في الوسط الحسابي لمعلمة الصعوبة كانت غير دالة إحصائياً وعلى مستوى دلالة (α≤٠,٠٥) أي أن معاملات الصعوبة لا تتأثر بشكل الفقرة، واتفقت هذه النتائج مع دراسة (علاونة،٢٠١٦) في عدم وجود فروق ذات دلالة إحصائية بين متوسطات الصعوبة لشكلي الاختبار.

مناقشة النتائج المتعلقة بالسؤال الثاني:

نص السؤال الثاني: **هل هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعلمة التمييز الفقرة تعزى لاختلاف شكل (الاختيار من متعدد–الصواب والخطأ)؟**

أظهرت النتائج المتعلقة بالسؤال الثاني أن الفروق في الوسط الحسابي لمعلمة التمييز كانت دالة إحصائياً وعلى مستوى دلالة (α≤٠,٠٥) وكانت لصالح الصواب والخطأ، أي أن معاملات التمييز تتأثر بشكل الفقرة، واختلفت هذه النتائج مع دراسة (علاونة،٢٠١٦) والذي قد يعود هذا الاختلاف إلى استخدامه عينات مستقلة في دراسته.

مناقشة النتائج المتعلقة بالسؤال الثالث:

نص السؤال الثالث: **هل هناك فروق ذات دلالة إحصائية بين معاملات الثبات التجريبي يعزى لاختلاف (الاختيار من متعدد–الصواب والخطأ)؟**

أظهرت النتائج المتعلقة بالسؤال الثالث أن الفروق بين معاملات الثبات التجريبي لشكل الفقرة كانت غير دالة إحصائياً وعلى مستوى دلالة (α≤٠,٠٥) أي أن معاملات الثبات التجريبي لا تتأثر بشكل الفقرة، واختلفت هذه النتائج مع دراسة (علاونة،٢٠١٦) والذي قد يعود هذا الاختلاف إلى استخدمه عينات مستقلة في دراسته.

التوصيات:

في ضوء ما توصل إليه البحث من نتائج فيمكن التوصية بما يلي:

١. توعية المعلمين إلى التنويع بأنواع الاختبارات التي يستخدمونها بحيث تحتوي على النوعين الاختيار من متعدد والصواب والخطأ.
٢. استعمال فقرات الصواب والخطأ على نطاق أوسع مما هو عليه الآن كونها أقوى تمييزاً.

المقترحات:

١. إجراء المزيد من الدراسات حول أثر شكل الفقرة في بناء الاختبارات في مباحث أخرى غير الحاسب وعلى مستويات مختلفة من الصفوف والتحقق من نتائج هذه الدراسة.
٢. إجراء دراسات تقارن الخصائص السيكومترية بين شكل فقرة الاختيار من متعدد والصواب والخطأ وفق النموذج اللوجستي ثلاثي المعلم.
٣. إجراء دراسات على أشكال مختلفة للاختبارات الموضوعية.

المراجع:

أولاً: المراجع العربية:

١. أبو جراد، حمدي يونس، (٢٠١٢) الخصائص السيكومترية لاختبارات الاختيار من متعدد والتكميل :دراسة مقارنة باستخدام نظرية الاستجابة للمفردة. مجلة العلوم التربوية والنفسية– البحرين.مج١٣،ع٣، ٣٧٥-٤٠٤
٢. أبو ناهية، صلاح الدين محمد،(١٩٩٤) القياس التربوي (ط١)، القاهرة: مكتبة الأنجلو المصرية
٣. دعمس، مصطفى نمر،(٢٠٠٨) استراتيجيات التقويم التربوي الحديث وأدواته، عمان: دار غيداء للنشر والتوزيع.
٤. الذنيبات، سجى أحمد،(٢٠١٣) أثر اختلاف شكل الاختبار على الخصائص السيكومترية للاختبار، رسالة ماجستير منشورة، جامعة مؤتة، مؤتة، الأردن
٥. زكري، علي محمد، (٢٠٠٨) الخصائص السيكومترية لاختبار (أوتيس-لينون) للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية، رسالة دكتوراه، جامعة أم القرى
٦. الزهراني، محمد راشد، (٢٠١٠) الخصائص السيكومترية لاختبار (أوتيس-لينون) للقدرة العقلية لدى طلبة المرحلة المتوسطة بمحافظة القنفذة، مجلة كلية التربية بالزقازيق، مصر، ع٦٨
٧. الطريري، عبدالرحمن سليمان، (١٩٩٧) القياس النفسي والتربوي نظريته أسسه تطبيقاته (ط١)، الرياض: مكتبة الرشد
٨. طعامنة، إيمان صالح صلاح، (٢٠٠٧) أثر عدد البدائل في اختبار الاختيار من متعدد على تقديرات القدرة للأفراد وتقديرات الصعوبة للفقرات ودالة المعلومات للفقرات والاختبار ، جامعة اليرموك، الأردن
٩. علام، صلاح الدين محمود،(٢٠١٤) الاختبارات والمقاييس التربوية والنفسية (ط٤)، عمان: دار الفكر
١٠. علام، صلاح الدين محمود،(٢٠١١) القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة (ط٥)، القاهرة: دار الفكر العربي
١١. علام، صلاح الدين محمود،(٢٠٠٥) نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي (ط١)، القاهرة: دار الفكر العربي
١٢. علاونة، معزوز جابر،(٢٠١٦) أثر حجم العينة و نمط الفقرة و جنس الطلبة على الخصائص السيكومترية لاختبار تحصيلي في الرياضيات لدى طلبة الصف الحادي عشر في المدارس الحكومية الفلسطينية، مجلة مجمع جامعة المدينة العالمية ماليزيا. ع١٧ ، ٤٢٦-٤٧٣
١٣. عمر، محمود & فخرو، حصة & السبيعي، تركي & تركي، آمنة،(٢٠١٠) القياس النفسي والتربوي (ط١)، عمان: دار المسيرة للنشر والتوزيع والطباعة
١٤. عودة، أحمد، (٢٠١٠م) القياس والتقويم في العملية التدريسية.(الطبعة الرابعة). عمان: دار الأمل.
١٥. الكحلوت، أحمد إسماعيل، (٢٠٠٢) مقارنة بين الخصائص السيكومترية لكل من اختبارات الاختيار من متعدد واختبارات التكميل، مجلة مركز البحوث التربوية، جامعة قطر، (٢٢)، ١٢٧ – ١٥٣
١٦. الكيلاني، عبدالله زيد، & الشريفين، نظال كمال، (٢٠١٤) مدخل إلى البحث في العلوم التربوية والاجتماعية(ط٤)، عمان: دار المسيرة للنشر والتوزيع والطباعة
١٧. مراد، صلاح أحمد & سليمان، أمين علي، (٢٠٠٥) الاختبارات والمقاييس في العلوم النفسية والتربوية (ط٢)، القاهرة: دار الكتاب الحديث.
١٨. النبهان، موسى،(٢٠١٣) أساسيات القياس في العلوم السلوكية (ط٢)، عمان: دار الشروق للنشر والتوزيع

ثانياً: المراجع الأجنبية:

[1] Alswalmeh. Y.& Feldt. L., Testing the equality of two related intraclass reliability coefficients, Applied Psychological Measurement, 18(2)(1994),183- 190, https://doi.org/10.1177/014662169401800207

[2] Baker. F, The Basics of Item Response Theory. Eric Clearing House on Assessment and Evaluation, University of Mary Land, College Park, MD, (2001)

[3] Gadalla. M., Multiple choice versus constructed response tests، in the assessment of mathematics computation skills. Paper presented at the Annual Meeting of the American Educational Research Association، Canada. (ERIC Document Reproduction Service .ED431813), (1999)

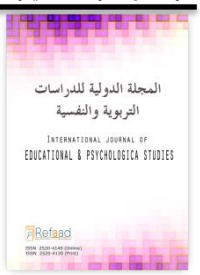
[4] Georgiev. N., Item Analysis of C, D and E Series from Raven′s Standard Progressive Matrices With Item Response Theory Two Parameter Logistic Model, Europe′s Journal of Psychology,4(3)(2008),1-17, https://doi.org/10.5964/ejop.v4i3.431

[5] Mobalegh. A & Barati. H, Multiple True-false (MTF) and Multiple-choice (MC) Test Formats: A Comparison between Two Versions of the Same Test Paper of Iranian NUEE, Journal of Language Teaching and Research, 3(5)(2012), 1027-1037, https://doi.org/10.4304/jltr.3.5.1027-1037

[6] Tasdemir. M, A Comparison of Multiple-Choice Tests and True-False Tests Used in Evaluating Student Progress, Journal of Instructional Psychology ،37(3)(2010)

Refaad
Research, Enlightment, Findings Accelerated Applications Development
رفاد للدراسات والأبحاث
www.refaad.com

المجلة الدولية للدراسات التربوية والنفسية
e-ISSN 2520-4149 , p-ISSN 2520-4130
المجلد٦- العدد١- ٢٠١٩، ص١-١٧
https://doi.org/DOI:10.31559/EPS2019.6.1.1



The Effect of Item Format The Multiple choice And True- false On Psychometrics Properties According To Item Response Theory (IRT) of The Computer Test for The First Secondary Grade in The City of Tabuk

Sabri Mohamed Ismail Abdel- Aal
Assistant Professor- Department of Education and Psychology, Tabuk University, KSA
sabdelaal@ut.edu.sa

Mohammad Oweid Raja Al Enezi
Teacher- KSA
abyzm22@gmail.com

Abstract: This study aimed on recognizable the effect items format of multiple choice and true-false on psychometrics properties according two-parameter logistic model of the (IRT), two format of computer test have been created: multiple choice and True-false, In its final format consisted of (24) items. The test has been applied on (421) students. The results showed no statistically significant differences in the item difficulty between the test formats, however he results revealed statistically significant differences in the item discrimination were in favor of True- false test. On the other hand, the results showed no statistically significant differences in the empirical reliability between test formats. The results of the research recommend educating teachers to the types of tests they use so that the two types contain multiple choice, right and wrong, and use the right and wrong paragraphs on a wider scale than it is now being more discrimination.

Keywords: *Psychometrics Properties, Format Of Test item, Multiple Choice Item, True-False Item, Item Response Theory (IRT)*

References:

[1] ʿlam. Ṣlaḥ Ạldyn Mḥmwd, Ạlakẖtbaṛat Wạlmqayys Ạltrbwyh Wạlnfsyh (Ṭ4), ʿman: Dạr Ạlfkr, (2014)

[2] ʿlam. Ṣlaḥ Ạldyn Mḥmwd, Ạlqyas Wạltqwym Ạltrbwy Wạlnfsy Ạ̉sạsyath Wtṭbyqath Wtwjhath Ạlmʿạṣrh (Ṭ5), Ạlqahrh: Dạr Ạlfkr Ạlʿrby, (2011)

[3] ʿlam. Ṣlaḥ Ạldyn Mḥmwd, Nmạdẖj Ạlastjabh Llmfrdh Ạlakẖtbaryh Ạ̉hạdyẗ Ạlbʿd Wmtʿddẗ Ạlạ̉bʿạd Wtṭbyqathạ Fy Ạlqyas Ạlnfsy Wạltrbwy (Ṭ1), Ạlqahrh: Dạr Ạlfkr Ạlʿrby, (2005)

[4] ʿlawnh. Mʿzwz Jạbr, Ạ̉thr Ḥjm Ạlʿynh W Nmṭ Ạlfqrh W Jns Ạltlbh ʿla Ạlkẖṣạ̉ṣ Ạlsykwmtryh Lạkẖtbạr Thṣyly Fy Ạlryạdyạt Lda Ṭlbẗ Ạlṣf Ạlhạdy ʿshr Fy Ạlmdạrs Ạlḥkwmyh Ạlflsṭynyh, Mjlẗ Mjmʿ Jạmʿẗ Ạlmdynh Ạlʿạlmyh Mạlyzyạ. (17)(2016), 426-473

[5] ʿmr. Mḥmwd , Fkẖrw. Ḥṣh , Ạlsbyʿy. Trky & Trky. Ả̂mnh, Ạlqyas Ạlnfsy Wạltrbwy (Ṭ1), ʿman: Dạr Ạlmsyrh Llnsẖr Wạltwzyʿ Wạlṭbạʿh, (2010)

[6] ʿwdh. Ạ̉hmd, Ạlqyạs Wạltqwym Fy Ạlʿmlyh Ạltdrysyh.(Ạlṭbʿh Ạlrạbʿh). ʿman: Dạr Ạlạ̉ml, (2010)

[7] Ẩbw Jrad. Ḥmdy Ywns, Ạlkẖṣạỷṣ Ạlsykwmtryh Lạkẖtbạrạt Ạlạkẖtyạr Mn Mtʿdd Wạltkmyl: Drạsh Mqạrnh Bạstkẖdạm Nẓryẗ Ạlạstjạbh Llmfrdh, Mjlẗ Ạlʿlwm Ạltrbwyh Wạlnfsyh –Ạlbḥryn, 13(3)(2012), 375-404

[8] Ẩbw Nạhyh. Ṣlạḥ Ạldyn Mḥmd, Ạlqyạs Ạltrbwy (Ṭ1), Ạlqạhrh: Mktbẗ Ạlẩnjlw Ạlmṣryh, (1994)

[9] Ạldẖnybạt. Sja Ẩḥmd, Ẩthr Ạkẖtlạf Shkl Ạlạkẖtbạr ʿla Ạlkẖṣạỷṣ Ạlsykwmtryh Llạkẖtbạr, Rsạlẗ Mạjstyr Mnshwrh, Jạmʿẗ Mẃth, Mẃth, Ạlẩrdn, (2013)

[10] Ạlkẖlwt. Ẩḥmd ASmạʿyl, Mqạrnh Byn Ạlkẖṣạỷṣ Ạlsykwmtryh Lkl Mn Ạkẖtbạrạt Ạlạkẖtyạr Mn Mtʿdd Wạkẖtbạrạt Ạltkmẏl, Mjlẗ Mrkz Ạlbḥwtẖ Ạltrbwyh, Jạmʿẗ Qṭr, (22)(2002), 127–153

[11] Ạlkylạny. ʿbdạllh Zyd, & Ạlshryfyn. Nẓạl Kmạl, Mdkẖl Ạla Ạlbḥtẖ Fy Ạlʿlwm Ạltrbwyh Wạlạjtmạʿyh (Ṭ4), ʿmạn: Dạr Ạlmsyrh Llnshr Wạltwzyʿ Wạlṭbạʿh, (2014)

[12] Ạlnbhạn. Mwsa, Ẩsạsyạt Ạlqyạs Fy Ạlʿlwm Ạlslwkyh (Ṭ2), ʿmạn: Dạr Ạlshrwq Llnshr Wạltwzyʿ, (2013)

[13] Alswalmeh. Y.& Feldt. L., Testing the equality of two related intraclass reliability coefficients, Applied Psychological Measurement, 18(2)(1994),183- 190, https://doi.org/10.1177/014662169401800207

[14] Ạlṭryry. ʿbdạlrḥmn Slymạn, Ạlqyạs Ạlnfsy Wạltrbwy Nẓryth Ẩssh Tṭbyqạth (Ṭ1), Ạlryạḍ: Mktbẗ Ạlrshd, (1997)

[15] Ạlzhrạny. Mḥmd Rạshd, Ạlkẖṣạỷṣ Ạlsykwmtryh Lạkẖtbạr (Ẩwtys-Lynwn) Llqdrh Ạlʿqlyh Lda Ṭlbẗ Ạlmrḥlh Ạlmtwsṭh Bmḥạfẓẗ Ạlqnfdẖh, Mjlẗ Klyẗ Ạltrbyh Bạlzqạzyq, Mṣr, (68)(2010)

[16] Baker. F, The Basics of Item Response Theory. Eric Clearing House on Assessment and Evaluation, University of Mary Land, College Park, MD, (2001)

[17] Dʿms. Mṣṭfy̲ Nmr, Ạstrạtyjyạt Ạltqwym Ạltrbwy Ạlḥdytẖ Wẩdwạth, ʿmạn: Dạr Gẖydạʾ Llnshr Wạltwzyʿ, (2008)

[18] Gadalla. M., Multiple choice versus constructed response tests، in the assessment of mathematics computation skills. Paper presented at the Annual Meeting of the American Educational Research Association، Canada. (ERIC Document Reproduction Service .ED431813), (1999)

[19] Georgiev. N., Item Analysis of C, D and E Series from Raven′s Standard Progressive Matrices With Item Response Theory Two Parameter Logistic Model, Europe′s Journal of Psychology,4(3)(2008),1-17, https://doi.org/10.5964/ejop.v4i3.431

[20] Mobalegh. A & Barati. H, Multiple True-false (MTF) and Multiple-choice (MC) Test Formats: A Comparison between Two Versions of the Same Test Paper of Iranian NUEE, Journal of Language Teaching and Research, 3(5)(2012), 1027-1037, https://doi.org/10.4304/jltr.3.5.1027-1037

[21] Mrạd. Ṣlạḥ Ẩḥmd & Slymạn. Ẩmyn ʿly, Ạlạkẖtbạrạt Wạlmqạyys Fy Ạlʿlwm Ạlnfsyh Wạltrbwyh (Ṭ2), Ạlqạhrh: Dạr Ạlktạb Ạlḥdytẖ, (2005)

[22] Ṭʿạmnh. Ạymạn Ṣạlḥ Ṣlạḥ, Ẩthr ʿdd Ạlbdạỷl Fy Ạkẖtbạr Ạlạkẖtyạr Mn Mtʿdd ʿla Tqdyrạt Ạlqdrh Llẩfrạd Wtqdyrạt Ạlṣʿwbh Llfqrạt Wdạlẗ Ạlmʿlwmạt Llfqrạt Wạlạkẖtbạr, Jạmʿẗ Ạlyrmwk, Ạlẩrdn, (2007)

[23] Tasdemir. M, A Comparison of Multiple-Choice Tests and True-False Tests Used in Evaluating Student Progress, Journal of Instructional Psychology ،37(3)(2010)

[24] Zkry. ʿly Mḥmd, Ạlkẖṣạỷṣ Ạlsykwmtryh Lạkẖtbạr (Ẩwtys-Lynwn) Llqdrh Ạlʿqlyh Mqdrh Wfq Ạlqyạs Ạlklạsyky Wnmwdẖj Rạsẖ Lda Ṭlbẗ Ạlmrḥlh Ạlmtwsṭh Bmḥạfẓẗ Ṣbyạ Ạltʿlymyh, Rsạlẗ Dktwrạh, Jạmʿẗ Ạm Ạlqra, (2008)