العـــــدد (21)، الجــزء الثاني، اكتوبر 2023، ص ص 212 - 246

أثر طول اختبار الاختيار من متعدد وطريقة ترتيب فقراته على دقة تقدير بارامترات فقراته وقدرات المختبرين وفق النموذج اللوجستي ثلاثي البارامتر

اعداد

د. أحمد بن رزق الله الزهراني

وزارة التعليم – الإدارة العامة للتعليم بمحافظة جدة

أثر طول اختبار الاختيار من متعدد وطريقة ترتيب فقراته على دقة تقدير بارامترات فقراته وقدرات المختبرين وفق النموذج اللوجستي ثلاثي البارامتر

د. أحمد بن رزق الله الزهراني

الملخـــص

هدف البحث إلى استقصاء أثر طول اختبار الاختيار من متعدد في مادة الرياضيات للصف السادس الابتدائي على دقة تقدير بارامترات الفقرات وقدرات المختبرين باستخدام النموذج اللوجستي ثلاثي البارامتر، واتبعت الدراسة المنهج شبه التجريبي، وتمثلت أداة الدراسة في (4) صور اختبارية في مادة الرياضيات وفقاً للأهداف التعليمية للوحدة، تختلف فيما بينها في عدد الفقرات (30، 60) و / أو طريقة ترتيب فقرات الاختبار (حسب المحتوى / عشوائيًا)، وتم تطبيقهما على عينة البحث المكونة من (2400) تلميذ بمدارس مدينة جدة الابتدائية، وبعد التحقق من الخصائص السيكومترية للصور الاختبارية، تم تحليل نتائجهما وفقاً للنموذج اللوجستي ثلاثي البارامتر وباستخدام البرامج الإحصائية (SPSS, Bilog-mg3)، وأظهرت النتائج وجود فروق بين متوسطي الأخطاء المعيارية في تقدير بارامترات الصعوبة والتمييز وقدرات المختبرين تعزى لاختلاف عدد فقرات الصورتين الاختباريتين، حيث كلما زادت عدد فقرات الاختبار زادت دقة تقديرات بارامترات صعوبة وتمييز الفقرة وقدرات المختبرين، بينما أظهرت النتائج عدم وجود فروق بين متوسطي الأخطاء المعيارية في تقدير بارامتر التخمين بحسب عدد الفقرات، كما توصل البحث إلى وجود فروق ذات دلالة إحصائية بين متوسطي الأخطاء المعيارية في تقدير بارامتر صعوبة الفقرة وقدرات الأفراد بحسب ترتيب فقرات الاختبار؛ حيث إذا رُتبت فقرات الاختبار حسب المحتوى زادت دقة تقديرات هذه البارامترات، بينما لم توجد فروق ذات دلالة إحصائية بين متوسطي الأخطاء المعيارية في تقدير بارامتري التمييز والتخمين بحسب ترتيب الفقرات للاختبار.

الكلمات المفتاحية: عدد فقرات الاختبار – طريقة ترتيب فقرات الاختبار – بارامترات الفقرات – قدرة المختبرين – نظرية الاستجابة للمفردة – النموذج اللوجستي ثلاثي البارامتر.

المقـدمـــة:

يعد القياس والتقويم وأدواتهما جانبًا من الجوانب المهمة والحيوية وركنًا أساسيًا للمسؤولين وأصحاب القرار في التربية أو الإدارة أو غيرها من المجالات ذات الصلة، فمن خلالهما يتم تقدير الظواهر والخصائص بدقة وموضوعية، وكذلك الوصول إلى براهين تساعد في اتخاذ القرارات التربوية، حيث أن القياس والتقويم أمر ضروري في العلوم الإنسانية وحتى الطبيعية، وتأتي الأهمية التربوية لهما من خلال دراسة المعرفة والحرص على ما تم إنجازه من أهداف المجتمع التربوية أو الإدارية أو غيرهما (الهويدي، 2004).

وقد استخدمت نظرية القياس الكلاسيكية بشكل كبير ومنتشر في تصميم وبناء الاختبارات النفسية والتربوية في القرن العشرين، ولكنها لم تكن بتلك الفاعلية المنشودة للتغلب على كثير من القضايا المنهجية المهمة والمشكلات السيكومترية المعاصرة، على سبيل المثال / تصميم بنوك الأسئلة وكذلك الواجبات الأدائية، وإنشاء نظم اختبارات موائمة حاسوبياً (علام، 2005).

وقد واجهت هذه النظرية معضلة تتمثل في أن المؤشرات الكلاسيكية للمفردات أو الفقرات غير ثابتة؛ والسبب يعود إلى أن الإجراءات المتبعة في هذه النظرية لانتقاء الفقرات تعتمد على المجموعة المستخدمة في تحديد بارامترات الفقرة ودرجة تمثيل المجموعة للمجتمع (Bond & Fox,2007).

وللتغلب على عيوب وانتقادات أساليب القياس وفق النظرية الكلاسيكية للقياس Classical Test Theory (CTT)، قام علماء القياس بمحاولات مستمرة وجهود بحثية لتطوير نظرية سيكومترية، أُطلق عليها نظرية الاستجابة للمفردة Item Response Theory (IRT)، ومع تطور هذه النظرية منذ القرن الماضي وحتى الوقت الحاضر، انبثق عنها نماذج متعددة مستندة إلى التقنيات المتقدمة وتوفر برمجيات الحاسب.

وقد حظيت هذه النظرية باهتمام واسع من قبل المختصين بالقياس في الدول المتقدمة والمتطورة، حيث تم اعتمادها في تطوير الاختبارات التحصيلية المقننة، واختبارات الاستعدادات، وكذلك مقاييس الذكاء، والاتجاهات، والشخصية، وغيرها من جوانب الوجدان الأخرى، وكذلك العديد من التطبيقات الأخرى.

وتتمثل الفكرة الأساسية لنماذج الاستجابة للمفردة (IRT) في محاولة اشتقاق وتقدير بارامترات الفقرات والأفراد في السمات أو القدرات موضع القياس، واستخدام هذه القيم المشتقة والمقدرة للبارامترات في تفسير أداء كل فرد في كل فقرة من فقرات الاختبار (علام، 2005).

تغلبت نظرية الاستجابة للفقرة على كثير من مشكلات القياس التي لم تستطيع النظرية الكلاسيكية للقياس مواجهتها والتغلب عليها، حيث أتت هذه النظرية للتغلب على كثير من جوانب القصور والضعف التي انتابت مبادئ ومفاهيم النظرية الكلاسيكية للقياس (علام، 2005)، ولعل أهم مواطن الضعف في نظرية القياس الكلاسيكية هو تأثر خصائص الاختبار والفقرة وفقاً للنظرية الكلاسيكية بخصائص العينة وموقف الاختبار، بينما في نظرية الاستجابة للفقرة تكون قدرة المفحوص ثابتة ولا تتأثر بخصائص العينة (عودة، 2010)، ويكون تقدير بارامترات الفقرات مستقلاً عن خصائص الأفراد المستخدمة في تقدير هذه البارامترات، وكذلك تقدير قدرة الفرد مستقلة عن عينة الفقرات المطبقة عليه (Hambelton, 1994)، وهذا ما يعنى باستقلالية القياس التي تعد النقطة المفصلية بين النظرية الكلاسيكية ونظرية الاستجابة للمفردة، وهو ما ذكره Lord (1980) بخاصية اللاتغاير (Invariance)، واذا ما تحققت خاصية اللاتغاير يمكن القيام ببعض التطبيقات: بنوك الأسئلة، ومعادلة الاختبارات، والكشف عن تحيز الفقرات، والقياس التكيفي (Rup & Zumbo,2006)، وتقدم نظرية الاستجابة للفقرة وما يرتبط بها من نماذج عدة طرق للوصول إلى ما يعرف بالاختبار المثالي، كما أنها أيضاً تحقق أكبر قدر من الموضوعية لمعالجة قصور أساليب القياس في النظرية الكلاسيكية (Tay & Smith,2015).

ويتطلب المجال التربوي دقة عالية في إعداد الاختبارات التحصيلية والمقاييس، والاختبارات النفسية، والقدرات، والذكاء، واختبارات الاستعداد، وغيرها؛ وذلك لاتخاذ القرارات بدقة وموضوعية، وتمثل الاختبارات التحصيلية أكثر أدوات القياس استخداما في قياس التحصيل المعرفي للطلاب، فهي مجموعة من المفردات تكون موضوعية أو مقالية أو رسوماً أو أشكالاً تعطى للطالب للإجابة عنها تحريرياً أو شفوياً (العبادي،2006)، لقياس التحصيل المعرفي ومعرفة ما تحقق عنده من نتاج التعلم المنشود، كما يرى التربويون أن الاختبارات تمثل عاملاً أساسياً ومهماً في تقويم عملية التعلم، بل يعتبرونها الأداة الوحيدة المحددة ما لدى الطالب من مهارة ومعرفة، فالاختبارات مؤشرات جيدة للعديد من التغيرات المطلوبة في سلوك الطالب، ومقاييس لكثير من مخرجات العملية التعليمية (علام، 2011)، ويعد الهدف من تطبيق هذه الاختبارات هو إعطاء درجات للشخص المفحوص، تعكس مستوى مهاراته أو مقدار ما يمتلك من معارف وخبراته في المقرر الذي يقوم بدراسته (Hambleton ,2004).

وتعد اختبارات الاختيار من متعدد أكثر أنواع الاختبارات شيوعاً واستخداماً في النظام التربوي، حيث يمكن استخدامها في فقياس كافة الأهداف التربوية المعرفية بمستوياتها المختلفة،

كما تتميز بكفاءتها وتعدد استعمالها في المباحث والمراحل التعليمية المختلفة، كما أنها تسمح بمعاينة أكبر قدر من المجال السلوكي للمقياس (ظاظا، 2012)، وهي اختبارات حديثة نسبياً بالنسبة للاختبارات المقالية حيث كان أول ظهور لها في عام 1915م في أمريكا، وتتمتع باستقلالية تامة بين علامة الطالب وشخصية المصحح وأهوائه، وبذلك تكون الاختبارات موضوعية وشمولية أيضًا لتمثيلها الواسع للمادة التعليمية.

وتتكون فقرات الاختيار من متعدد من متن متبوعاً بثلاث بدائل أو أكثر، وتكون أحد هذه البدائل هي الإجابة الصحيحة وما تبقى عبارة عن مموهات أو مشتتات، ودائماً ما يهدف المتن إلى اظهار المشكلة حتى يتسنى للطالب اختيار الإجابة الصحيحة لتلك المشكلة، ويراعى عند كتابة المتن البساطة والوضوح (Gronlund & Linn, 1990) وتكون المموهات بدائل تشبه الإجابة الصحيحة، لكنها مشتتة وخاطئة، وهي غالباً تشير إلى إجابات ذات علاقة بالسؤال المطروح للتأثير على استجابة الفرد وعلى فقرة الاختبار، وذلك للحد من نسبة التخمين لدى الأفراد (Nitko, 2001)، وتشتيت التفكير لدى المفحوصين الذين لا يعرفون الإجابة الصحيحة، وتبين مدى حاجة المفحوص للمعلومات اللازمة للوصول للإجابة الصحيحة، وكذلك معرفة جوانب الضعف والقصور لدى الأفراد نحو توجههم عند بديل مشتت غير صحيح، حيث يجب أن تكون المموهات جاذبة وقريبة جداً للإجابة الصحيحة (McMillan, 2004).

وتتمتع فقرات الاختيار من متعدد بكثير من المزايا إلا أنها تتطلب عدد كاف ومقبول من المموهات وخاصة في مراحل التعليم الأساسية بسبب قلة المفردات والمعلومات التي يمتلكونها في محتوى الاختبار، وكذلك كتابتها تستغرق وقتاً أطول مقارنة بأنواع أخرى من الفقرات الموضوعية، خاصة كتابة المموهات الجيدة (McMillan,2004).

ونظرًا لتأثير الخطأ في القياس على عملية اتخاذ القرار؛ فقد تم اعتماد إجراءات وقواعد دقيقة في بناء المقاييس، للتقليل والقضاء على أخطاء عملية القياس، متزامناً في الوقت نفسه مع توافر الخصائص السيكومترية للمقاييس أو الاختبار وفقراته بدرجة دقيقة وعالية، وكذلك لخبرة مصمم وباني المقاييس دور هام في صياغة الفقرات بصورة واضحة وموثوق بنتائجها، ومن خلال دقة هذه الإجراءات التي تتم في بناء المقاييس أو الاختبارات يتم الحكم إلى أي درجة يمكن الوثوق بدقة النتائج ودقة اتخاذ القرار في ضوئها (عودة،2010)، ومن ثم فإن دراسة العوامل التي قد تؤثر في دقة تقدير قدرات الأفراد أو الخصائص السيكومترية لتلك الاختبارات في ضوء نظرية الاستجابة للفقرة يعد أمرًا مهما لمزيد من الدقة والموثوقية في اتخاذ القرارات التربوية المختلفة.

مشكلة البحث:

نظراً للعيوب جراء استخدام النظرية الكلاسيكية في القياس والتقويم جاءت نظرية الاستجابة للفقرة والتي ركزت على التقدير الاحصائي لبارامترات الفقرات وقدرة الأفراد، ويعد النموذج ثلاثي البارامتر أحد نماذج نظرية الاستجابة للفقرة الذي يمثل امتداداً للنموذج اللوجيستي ثنائي البارامتر، حيث أضاف بيرنبوم بارامترًا ثالثاً أطلق عليه بارامتر الخط التقاربي، أو بارامتر التخمين (ci) ويشير إلى احتمال توصل الأفراد ذوي القدرة المتدنية جدًا إلى الإجابة الصحيحة بطريقة التخمين(Hambleton & Swaminathan, 1985)، وهذا يتناسب مع طبيعة أسئلة الاختيار من متعدد التي يمكن إجابتها عن طريق التخمين.

ويعد الكشف عن دقة تقدير بارامترات الفقرات والأفراد أمرًا ضروريًا للاطمئنان لنتائج التحليل باستخدام نماذج الاستجابة للفقرة، ويشير اجريستي وفاينالي (Agresti & Finnlay, 2009) إلى أن الخطأ المعياري يعد مؤشراً احصائياً يعتمد عليه الباحثون لإصدار حكم على دقة تقدير العينة للمجتمع. والباحثون الذين يستخدمون نظرية الاستجابة للفقرة يأخذون الخطأ المعياري لكل من تقدير بارامتر القدرة وبارامترات الفقرات، حيث أن للدلالة على الدقة في تقدير بارامترات النموذج المستخدم، حيث أنه إذا كان هناك مقدار قليل للخطأ المعياري فهذا يدل على دقة القياس، علماً بأن للخطأ المعياري دوراً هاماً في دالة معلومات الاختبار، ودالة معلومات الاختبار، ويعد مؤشرًا جيدًا يستدل به على ثبات الاختبار في نظرية الاستجابة للفقرة (Embretson & Reiase, 2000).

وقد سعت بعض الدراسات إلى دراسة تأثير عدد فقرات الاختبار على دقة تقديرات بارمترات الفقرات وقدرات المختبرين، مثل دراسات (بني عطا، 2014؛ الحواري، 2015؛ الدرابيع ، 2001؛ Barnes & Wise, 1991; Glass, 2005; Fitzpatric, 2009; Fitzpatric & Ann, 2001; Hambleton & Cook, 1980; Pommerich, 2007)، وقد توصلت كل هذه الدراسات إلى أفضلية تقدير بارامترات الفقرات وقدرات الأفراد مع زيادة عدد الفقرات، ما عدا دراسة الحواري (2015) التي توصلت إلى أفضليتها في حالة النموذج الأقصر.

كما سعت الدراسات السابقة إلى دراسة تأثير طريقة ترتيب الفقرات على الخصائص السيكومترية للاختبارات والمقاييس النفسية، أو الأداء على الاختبار، وقد تضاربت نتائجها، كما تبين وجود نقص كبير في الدراسات التي سعت إلى دراسة تأثير اختلاف طريقة ترتيب الفقرات

على هذه الخصائص السيكومترية وفق نظرية الاستجابة للفقرة بشكل عام، ووفق النموذج ثلاثي البارامتر على وجه الخصوص.

وبالبحث في مصادر وقواعد المعلومات المختلفة لم يتم الحصول على دراسات تناولت تأثير عدد فقرات الاختبار وطريقة ترتيب الفقرات معًا على دقة تقديرات بارامترات الفقرات وقدرات المختبرين وفق النموذج ثلاثي البارامتر، وهو ما يسعى إليه هذا البحث، وبالتالي تتمثل مشكلة البحث في التساؤلات الآتية:

1- ما أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وفقاً للنموذج اللوجستي ثلاثي البارامتر؟

2- ما أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير قدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر؟

3- ما أثر طريقة ترتيب فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وفقاً للنموذج اللوجستي ثلاثي البارامتر؟

4- ما أثر طريقة ترتيب فقرات اختبار الاختيار من متعدد على دقة تقدير قدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر؟

أهــداف البحــث:

هدف البحث إلى استقصاء أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وقدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر في اختبار مادة الرياضيات للصف السادس الابتدائي، وبذلك تكون أهداف هذه الدراسة كما يلي:

- الكشف عن أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وفقاً للنموذج اللوجستي ثلاثي البارامتر.
- الكشف عن أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير قدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر.
- الكشف عن أثر طريقة ترتيب فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وفقاً للنموذج اللوجستي ثلاثي البارامتر.
- الكشف عن أثر طريقة ترتيب فقرات اختبار الاختيار من متعدد على دقة تقدير قدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر.

أهـــمية البحـــث:

تكمن أهمية الدراسة فيما يلي:

1- إثراء الدراسات والبحوث المتعلقة بتحديد عدد فقرات اختبار الاختيار من متعدد المناسبة في تقدير ما يمتلكه الطلاب من مهارات أو قدرات وفق النموذج اللوجستي ثلاثي البارامتر.

2- دراسة الجودة المتعلقة ببناء الاختبار وفق النظرية الحديثة وفي ضوء النموذج اللوجستي ثلاثي البارامتر.

3- تقديم تغذية راجعة للباحثين ومعدي الاختبارات والمهتمين بوضع الفقرات الاختبارية التي تتميز بارتفاع في التقديرات لبارامترات الفقرات وقدرات الأفراد.

4- قد يستفيد من نتائج الدراسة القائمون على العمل الأكاديمي والتعليمي ودعوتهم إلى ضرورة الاهتمام بجودة بناء الاختبارات.

مصطلحات البحث:

- اختبار الاختيار من متعدد: يشير العزاوي (2007) إلى أن اختبار الاختيار من متعدد يتكون من جزأين، حيث الجزء الأول هو الأساس وخاص بصياغة المسألة التي يُسأل عنها الطالب، أما الجزء الآخر يتكون من عدد من البدائل، وعلى الطالب اختيار الإجابة الصحيحة من بينها، ولابد أن تُطرح المسألة أو المشكلة في صيغة استفهامية أو على شكل عبارة منقوصة في النهاية.

والتعريف الاجرائي: هو ما تقيسه الصورتين الاختباريتين بحسب (عدد الفقرات)، حيث تقدم للطالب مجموعة من فقرات الاختيار من متعدد المكون من أسئلة ممثلة لكتاب الرياضيات للصف السادس الابتدائي، والذي تم تطبيقه على عينة ممثلة خلال الفصل الدراسي الثاني للعام الدراسي 1442هـ.

- طول الاختبار: هو عدد فقرات الاختبار التي يتضمنها الاختبار، وتم استخدام طولين مختلفين من الفقرات (30، 60) فقرة.
- نظرية الاستجابة للمفردة: مجموعة من الطرق الإحصائية تُستخدم لحساب بارامترات الفقرات، من حيث الصعوبة، والتمييز، والتخمين، ودالة المعلومات، وكذلك تُستخدم هذه النظرية لحساب قدرات الأفراد (علام، 2005).

- دقة التقدير: يذكر بني عطا (2014) أنه مصطلح يُشير إلى الجودة في التقدير، حيث تتميز تلك الجودة باحتمالية أن التقدير قريب من البارامترات الحقيقية بدرجة كبيرة.

والتعريف الاجرائي: هو الجذر التربيعي لمتوسطات مربعات الانحرافات للفروق بين البارامترات المقدرة والبارامترات الحقيقية ومعاملات الارتباط بينهما.

- قدرات المختبرين (θ): هي مستوى القدرة الكاملة الوحيدة التي تفسر أداء الفرد في الاختبار، عندما تكون جميع فقرات الاختبار متجانسة، أي أنها أحادية البعد (Sinharay & Holland, 2006)، عند وتقدير قدرة الفرد يتم وفقاً للقدرة الحقيقية التي تكمن خلف استجابته عن فقرات الاختبار، وكذلك لتصنيف الأفراد.

والتعريف الاجرائي: هي مستوى القدرة أو السمة المقاسة لدى المختبرين للإجابة على فقرات الاختبار المعد لقياس هذه السمة.

- الصعوبة (b): وصف سلوك وتحديد موقع المفردة على متصل القدرة، فتكون ازاحتها لذوي القدرة العالية إذا كانت أكثر صعوبة والعكس صحيح، وبيانياً تمثل الصعوبة القدرة المقابلة لنقطة انقلاب المنحنى (Hambleton & Swaminathan, 1985).
- التمييز (a): وصف مدى قدرة المفردة للتمييز بين الأفراد ذوي القدرات العالية من القدرة، والأفراد ذوي القدرات المتدنية من القدرة، ورياضياً تمثل قيمة المنحنى، أي أنه كلما زاد ميل المنحنى أصبح التمييز عالي، واذا كانت القيمة تساوي الصفر فهذا يدل أن احتمال الإجابة الصحيحة للأفراد ذوي القدرة العالية والأفراد ذوي القدرة المتدنية نفسه (Hambleton & Swaminathan, 1985).
- التخمين (c): أن يحصل الطالب على الإجابة الصحيحة باختياره للإجابة بشكل عشوائي، أي من خلال التخمين وليس القدرة، وبيانياً هو تقاطع المحنى مع الاحداثي الصادي (Hambleton & Swaminathan, 1985).

حدود البحث:

- الحدود الموضوعية: أثر عدد فقرات اختبار الاختيار من متعدد (30، 60) على دقة تقدير بارامترات الفقرات وقدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر.
- الحدود البشرية: اقتصرت الدراسة على تلاميذ الصف السادس الابتدائي بمدينة جدة.
- الحدود الزمانية: تم تطبيق الدراسة في الفصل الدراسي الثاني للعام 1442هـ / 2020م.

- الحدود المكانية: تم تطبيق الدراسة بمدارس مدينة جدة بمنطقة مكة المكرمة بالمملكة العربية السعودية.
- اختبار تحصيلي له (4) صور اختبارية تختلف فيما بينها في طريقة ترتيب فقرات الاختبار وعدد فقرات الاختبار.

الدراسات السابقة التي تناولت تأثير عدد فقرات الاختبار:

هدفت دراسة **هامبلتون وكوك (Hambleton & Cook, 1980)** إلى تعرف أثر حجم العينة وعدد فقرات الاختبار في تقدير قدرات الأفراد وفقاً للنموذج الثلاثي البارامتر، حيث تمت دراسة ثلاثة مستويات لحجم العينة (50، 200، 1000)، وكذلك ثلاثة مستويات لعدد فقرات الاختبار (10،20،80)، وبينت نتائج الدراسة أن القيمة (2.19) كانت أكبر قيمة للخطأ المعياري للتنبؤ إذا كان عدد فقرات الاختبار (10) فقرات، وعند حجم عينة (50) فرداً، وتراوحت قيمة الخطأ المعياري لعدد فقرات الاختبار (20) فقرة بين 0.88 إلى 1.50، عند حجم عينة (200) فرد، وكذلك أن قيمة الخطأ المعياري للتنبؤ تقل عند اختبار (80) فقرة، بحجم عينة (1000) فرداً، وبالتالي فإن نتائج الدراسة أثبتت بشكل عام أن دقة التقدير تزداد بزيادة حجم العينة وعدد فقرات الاختبار.

وقد سعت **دراسة بارنس وويس (Barnes & Wise, 1991)** إلى تعرف أثر حجم العينة وعدد فقرات الاختبار في تقدير قدرات الأفراد وفقاً لنموذج راش، حيث كانت العينة على ثلاثة مستويات (200، 100، 50) فرداً، وهناك مستويين لعدد فقرات الاختبار (25 – 50) فقرة، وأثبتت نتائج الدراسة أن عدد الفقرات لها تأثير على قدرة المفحوص، وليس هناك تأثير لحجم العينة على قدرة المفحوص، وأن عدد الفقرات للاختبار لا يؤثر على بارامتر الصعوبة، وكان هناك تغيراً بسيطاً عندما تغير حجم العينة، بسبب الفروق بين حجم العينات كان صغيراً.

كما قام **الدرابيع (2001)** بدراسة هدفت إلى التعرف على أثر حجم العينة وعدد فقرات الاختبار في تقدير قدرات الأفراد وبارامتر صعوبة الفقرة وفقاً لنموذج راش، حيث كانت العينة على ثلاثة مستويات (50، 100، 500) وهناك ثلاثة مستويات لعدد فقرات الاختبار (25، 50، 300)، وأظهرت النتائج بشكل واضح وجوهري وجود فروق لتفاعل كل من متغير حجم العينة وعدد الفقرات في دقة تقدير قدرة الفرد، بينمالم تتضح هذه الفروق في دقة تقدير بارامتر الصعوبة.

بينما أجريت **دراسة فيتزباتريك وآن (Fitzpatric & Ann, 2001)** بهدف تعرف أثر حجم العينة وعدد فقرات الاختبار في ثبات الاختبار ومعادلة الاختبار مستخدمة نموذج التقدير الجزئي، حيث كانت العينة على ثلاثة مستويات (200، 500، 1000) فرداً، وهناك خمس مستويات لعدد فقرات الاختبار (2، 4، 8، 12، 20)، وتم استخدام المحاكاة بغرض توليد استجابات لاختبار في الرياضيات، وتم استخدام معيار جذر معدل مربعات الفروق لتقييم دقة تقديرات البارامترات، وقد أثبتت نتائج الدراسة أنه كلما زاد حجم العينة كانت هناك دقة أكثر في تقييم البارامترات، وكذلك توصلت إلى أن الاختبارات التي تتكون من (8) فقرات على الأقل تكون قابلة للاستخدام، وأنه كلما زاد عدد فقرات الاختبار كانت دقة تقييم البرامترات أفضل.

وقد سعت **دراسة جلاس (Glass, 2005)** إلى تعرف أثر حجم العينة وعدد فقرات الاختبار في تقدير قدرات الأفراد وفق طريقة بييز، حيث كانت العينة على ثلاثة مستويات (500، 1000، 2000)، وهناك مستويين لعدد فقرات الاختبار (200، 440)، وولدت بيانات ثنائية التدريج، وأظهرت النتائج أنه بزيادة عدد الفقرات عند حجم عينة (500، 2000) تقل الأخطاء المعيارية، حيث تزيد دقة التقدير.

ما سعت **دراسة بوميرتش (Pommerich, 2007)** إلى اختبار البرامج التي تستخدم الكمبيوتر والاختبار الورقي، حيث اهتمت بدراسة استقرار تقدير بارامترات الفقرات، من خلال اختبارات متفاوتة في عدد الفقرات، وأظهرت النتائج أنه بزيادة عدد الفقرات، يزداد الاستقرار في بارامترات الفقرات.

كما هدفت **دراسة فيتزباتريك (Fitzpatric, 2009)** إلى تعرف أثر عدد فقرات الاختبار في تقدير بارامترات الفقرات وفقاً لنموذج راش، وكانت عدد الفقرات للاختبار (15 – 10 – 5)، وأظهرت النتائج والتوصيات أنه يجب أن يزيد عدد فقرات الاختبار عن (15) فقرة لزيادة الاستقرار في تقدير بارامتراتها.

وقد قام **بني عطا (2014)** بدراسة لتعرف أثر حجم العينة وعدد فقرات الاختبار في تقدير قدرات الأفراد وبارامترات الفقرة وفقاً للنموذج الثلاثي البارامتر، وتم توليد بيانات ثنائية الاستجابة خمسين مرة، حيث كانت حجم العينة على ستة مستويات (100 ،250، 500، 1000، 2000،4000)، وهناك أيضاً ستة مستويات لعدد فقرات الاختبار (10، 25، 50، 75، 100، 300)، وأظهرت النتائج وجود أثر لعدد فقرات الاختبار وحجم العينة في دقة تقدير بارامترات

الفقرة وقدرات الأفراد، حيث كلما زادت عدد فقرات الاختبار عن (50) فقرة وزاد حجم العينة عن (2000) كلما زادت دقة تقديرات بارامترات الفقرة وقدرات الأفراد.

وقد هدفت **دراسة الحواري (2015)** إلى تعرف أثر عدد فقرات الاختبار في تقدير قدرات الأفراد وبارامتر صعوبة الفقرات وفقاً لنموذج راش، وتم توليد بيانات لحجم عينة (1000) فرداً، وهناك مستويين لعدد فقرات الاختبار (30، 60)، وأظهرت النتائج وجود فروق دالة إحصائيا في دقة تقدير بارامترات قدرات الأفراد وبارامترات الفقرات لصالح الاختبار ذي (30 فقرة) إذ زود بمعلومات أكثر.

ويتضح من خلال عرض نتائج الدراسات السابقة أن أغلب الدراسات توصلت إلى أفضلية تقدير بارامترات الفقرات وقدرات الأفراد مع زيادة عدد فقرات الاختبار، ما عدا دراسة الحواري (2015).

الدراسات السابقة التي تناولت تأثير طريقة ترتيب فقرات الاختبار:

هدفت دراسة **دويرنر وكالهون (Doerner & Calhoun, 2009)** إلى تعرف أثر ترتيب الفقرات على تحصيل طلاب مقرر الاقتصاد، وكُونت ثلاثة نماذج من الاختبار النهائي، حيث كل نموذج عدد فقراته (75) فقرة من اختيار من متعدد، حيث كان عدد الطلاب (1048) من (1296) طالب مسجل بمقرر الاقتصاد، وكان النموذج الأول مرتب تصاعدياً أي من السهل للأصعب، والنموذج الثاني مرتب تنازلياً أي من الصعب للأسهل، بينما النموذج الثالث رُتب بشكل عشوائي، وأظهرت نتائج الدراسة أن أعلى الدرجات للذين استجابوا للنموذج الأول، أي الذين تناولوا النموذج المرتب تصاعدياً أكثر من أقرانهم.

كما سعت **دراسة سيو (Sue, 2009)** إلى تقصي أثر ترتيب فقرات الاختبار على تحصيل طلاب مقرر مبادئ الاقتصاد، حيث كانت عينة الدراسة (81) طالباً، وكان هناك نموذجين للاختبار، النموذج الأول مرتب بحسب المحتوى، والنموذج الآخر مرتب بشكل عشوائي، وأظهرت نتائج الدراسة أنه لا توجد فروق دالة احصائياً في الدرجات تُعزى إلى ترتيب الفقرات.

بينما أجريت **دراسة رمضان (2010)** بهدف تعرف أثر ترتيب فقرات الاختبار بحسب درجة الصعوبة، حيث تكونت عينة الدراسة من (200) طالب من المرحلة الثانوية في محافظة المنوفية بمصر، وكان عدد فقرات الاختبار (23) فقرة اختيار من متعدد، كما أنه تم حساب الصدق والثبات للاختبار ومعاملات الصعوبة لكل فقرة، وكانت هناك ثلاثة نماذج للاختبار،

النموذج الأول مرتب ترتيب تصاعدي أي أنه من السهل للأصعب، والنموذج الثاني مرتب بشكل تنازلي أي أنه من الصعب للأسهل، والنموذج الثالث مرتب عشوائياً، وأظهرت النتائج أنه أعلى المتوسطات الحسابية للنماذج الثلاثة كان الترتيب التصاعدي أي الترتيب الذي يبدأ من الأسهل للأصعب.

كما قام **اللحياني (2010)** بدراسة هدفت إلى تعرف الطرق المهمة لترتيب فقرات الاختبار حسب المحتوى الدراسي، وبحسب السهولة والصعوبة وفق النظرية التقليدية للقياس، وكذلك للكشف عن النمط الأفضل لترتيب فقرات الاختبار من خلال الخصائص السيكومترية، وقد تم بناء اختبار تحصيلي ذو عدة بدائل وكان هناك خمسة نماذج للاختبار حيث عدد فقرات الاختبار الواحد (25) فقرة مرتبة حسب المحتوى الدراسي، والثاني من السهل للأصعب، والثالث من الصعب للأسهل، والرابع مرتب بشكل دائري (ترتيب تبادلي)، والخامس مرتب بشكل عشوائي، وكانت العينة مكونة من (600) طالب، وأظهرت النتائج أنه توجد فروق ذات دلالة إحصائية بين قيم معاملات الثبات تُعزى لاختلاف ترتيب فقرات الاختبار لصالح النموذج الأول (ترتيب الفقرات حسب المحتوى الدراسي).

وقد هدفت **دراسة العمري (2011)** إلى استكشاف أثر ترتيب الفقرات في المقاييس النفسية، من خلال خصائص المقياس وفقراته، وتم استخدام مقياس الثقة بالنفس، حيث تم إعادة ترتيب فقراته بأربعة طرق، في أربعة نماذج للمقياس، النموذج الأول تم ترتيب فقراته حسب مؤشر تمييز الفقرة، والنموذج الثاني تم ترتيب فقراته حسب مؤشر صدق الفقرة، والنموذج الثالث تم ترتيب فقراته حسب ثبات الفقرة، والنموذج الأخير تم ترتيب فقراته بشكل عشوائي، حيث كانت عينة الدراسة (264) طالباً، حيث تم تقسيمهم إلى أربع مجموعات، حيث تلقى كل مفحوص جميع الأنماط الأربعة للقياس، وأظهرت نتائج الدراسة أن ترتيب الفقرات معتمداً على مؤشر الثبات قد تقدم على الطرق الأخرى في تحسين متوسطات الخصائص السيكومترية.

وقد سعت **دراسة شين (Chen, 2012)** إلى تعرف أثر ترتيب الفقرات للاختبار بحسب الصعوبة على القلق وأداء الطالب، وفقاً للنظرية التقليدية للقياس، حيث كانت عينة الدراسة (250) طالباً، وكان هناك نموذجين للاختبار، النموذج الأول اختبار تحصيلي في المعرفة الثقافية، والنموذج الآخر مقياس للقلق، وتم تقسيم الطلاب إلى ثلاث مجموعات بحسب مستوى القلق لديهم، كما أنه تم ترتيب فقرات الاختبار تنازلياً أي أنه من الصعب للأسهل، وأظهرت نتائج

الدراسة ومن خلال اختبار (ت) إلى وجود أثر دال احصائياً لترتيب الفقرات على الأداء الأكاديمي لذوي القلق المرتفع والمتوسط.

كما هدفت دراسة **نايبي (Naibi, 2013)** إلى تعرف أثر ترتيب الفقرات على أداء الطلاب، حيث كانت العينة (480) طالباً وطالبة من المرحلة الثانوية، وكان هناك نموذجين للاختبار التحصيلي، ففي النموذج الأول رُتبت فقرات الاختبار تصاعدياً أي من الأسهل للأصعب، والنموذج الآخر رُتبت فقراته بشكل عشوائي، وأظهرت النتائج وجود أثر لترتيب الفقرات على أداء الطلاب في الاختبار، لصالح النموذج المرتب بشكل عشوائي.

بينما أجريت **دراسة الحاج (2015)** بهدف تعرف أثر ترتيب فقرات الاختبار ومستوى القلق لدى الطالبات على الأداء في الاختبار، حيث كانت عينة الدراسة (454) طالبة من جامعة أم القرى، وكان هناك ستة نماذج مختلفة الترتيب، حيث تم ترتيبهم بحسب درجات الصعوبة، فتم إيجاد ثلاثة نماذج مرتبة (تصاعدي، وتنازلي، وعشوائي)، وكذلك تم ترتيب الفقرات للنماذج الأخرى بحسب المحتوى المعرفي (تسلسلي، وعكسي، وعشوائي)، وأظهرت نتائج الدراسة أنه لا توجد فروق ذات دلالة إحصائية بين النماذج المرتبة بحسب درجة الصعوبة في متوسط الأداء للطالبات، بينما توجد فروق ذات دلالة إحصائية في متوسط أداء الطالبات من خلال النماذج المرتبة حسب المحتوى المعرفي في متوسط الأداء للطالبات، حيث كان أفضل أداء للطالبات على النموذجين التسلسلي والعكسي.

كما هدفت **دراسة سليمان (2015)** إلى تقصي أثر ترتيب الفقرات لاختبار تحصيلي على معاملات الصعوبة والتمييز للفقرة، وكذلك صدق وثبات الاختبار، حيث كانت العينة (300) طالباً، وكان هناك ثلاثة نماذج من الاختبارات وعدد فقرات كل نموذج (30) فقرة، وكان النموذج الأول مرتب بحسب مستوى الهدف الذي تقيسه الفقرة وفق تصنيف بلوم، والنموذج الثاني مرتب بحسب تسلسل المحتوى الدراسي، والنموذج الثالث رُتبت فقراته تصاعدياً بحسب معاملات صعوبة الفقرات، وأظهرت نتائج الدراسة أنه لا توجد فروق ذات دلالة إحصائية في المتوسطات الحسابية لمعاملات صعوبة الفقرات تُعزى لاختلاف ترتيب الفقرات، بينما هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعاملات التمييز تُعزى للنموذج الثاني الذي رُتبت فقراته تصاعدياً بحسب الصعوبة، وأنه لا توجد فروق ذات دلالة إحصائية بين معاملات ثبات وصدق الاختبار تُعزى لترتيب الفقرات.

وقام كل من **كاقوندو وروس (Kagundu & Ross, 2015)** بدراسة هدفت إلى التعرف على أثر ترتيب الفقرات لاختبار اختيار من متعدد على أداء الطلاب في دورة الاقتصاد، حيث كانت عينة الدراسة (515) طالباً وطالبة، وكان هناك أربعة نماذج للاختبار مختلفة وفق طريقة ترتيب الفقرات، النموذج الأول مرتب بحسب تسلسل المحتوى والثلاثة النماذج الأخرى مرتبة بشكل عشوائي، وأظهرت نتائج الدراسة أن اختلاف ترتيب الفقرات لا يؤثر على أداء الطلاب في تلك الدورة.

وقد سعى **عبابنة (2017)** إلى التعرف على أثر ترتيب الفقرات لاختبار اختيار من متعدد، على خصائص الاختبار السيكومترية ومفرداته، وكذلك أداء الطلاب عليه، وفقاً للنظرية التقليدية للقياس، حيث تكونت عينة من (1798) طالباً وطالبة، وتم بناء اختبار تحصيلي مكون من (64) فقرة من نوع اختيار من متعدد، وكان هناك ثلاثة نماذج للاختبار، النموذج الأول رُتب بشكل تصاعدي أي أنه من السهل للأصعب، والنموذج الثاني رُتبت فقراته تنازلياً أي أنه من الصعب للأسهل، والنموذج الثالث رُتبت فقراته بشكل عشوائي، وأظهرت نتائج الدراسة أنه توجد فروق ذات دلالة إحصائية في معاملات الصعوبة والتمييز للفقرة في النماذج الثلاثة، بينما لا توجد فروق ذات دلالة إحصائية في معاملات الصدق والثبات بين تلك النماذج الثلاثة، وكما أنه لا توجد فروق ذات دلالة إحصائية في متوسط أداء الطلاب.

كما هدفت دراسة **سالم وعليوة (2018)** إلى تعرف أثر اختلاف توزيع فقرات اختبار مرجعي المحك وتم اعداده في اطار المحتوى المعرفي، ورُتب بطريقتين (تسلسل معرفي – عشوائي)، وكانت عينة الدراسة (858) طالباً، حيث تكون الاختبار بعدد فقرات (100) فقرة، وكان هناك نموذجين للاختبار، النموذج الأول رُتبت فقراته بحسب تسلسل المحتوى، والنموذج الآخر رُتبت فقراته عشوائياً، وأظهرت نتائج الدراسة أنه لا توجد فروق ذات دلالة إحصائية في معاملات الصعوبة والتمييز تعود لترتيب الفقرات بحسب المحتوى متسلسلاً أو عشوائياً، بينما توجد فروق ذات دلالة إحصائية بين النموذجين في أداء الطلاب التحصيلي لصالح النموذج المرتب بتسلسل معرفي.

وقد أجرى **ساتي ورفاقه (Satti, et al., 2019)** دراسة بهدف كشف أثر ترتيب الفقرات للاختبار على أداء الطلاب ومستوى الصعوبة لطلاب كلية الطب، حيث تكونت عينة الدراسة من (98) طالباً وطالبة، وكان هناك أربعة نماذج للاختبار، حيث كان النموذج الأول مرتب بحسب عرض المعلومات أثناء التدريس داخل القاعة الدراسية، بينما النموذجين الثاني

والثالث رُتبت فقراتهما بطريقة عشوائية، أما النموذج الرابع رُتبت فقراته عكس ترتيب فقرات النموذج الأول، وأظهرت نتائج الدراسة أنه لا وجود لأثر ترتيب فقرات الاختبار على أداء الطلاب ومستوى صعوبة فقرات الاختبار.

وبالتالي يتضح من خلال العرض السابق للدراسات السابقة تضارب النتائج حول تأثير اختلاف طريقة ترتيب الفقرات على الخصائص السيكومترية للاختبار وفقراته أو الأداء عليه، كما يتضح وجود نقص كبير في الدراسات التي سعت إلى دراسة تأثير اختلاف طريقة ترتيب الفقرات على هذه الخصائص السيكومترية وفق نظرية الاستجابة للفقرة بشكل عام، ووفق النموذج ثلاثي البارامتر على وجه الخصوص.

مجتمع البحث:

تكون مجتمع البحث من جميع طلاب الصف السادس الابتدائي في مدارس التعليم العام الحكومي وتحفيظ القرآن الكريم في مدينة جدة التعليمية للعام الدراسي 1441هـ – 1442هـ والبالغ عددهم (22167) طالباً موزعين على (248) تبعاً للدليل الاحصائي للإدارة العامة للتعليم بمحافظة جدة.

عينة البحث:

تمثلت عينة البحث في (2400) تلميذ تم اختيارهم بطريقة عشوائية عنقودية من مجتمع البحث، وتكونت، حيث تم اختيار (20) مدرسة عشوائية، ثم اختيار عدد من الطلاب في كل مدرسة.

أداة البحث:

من خلال الإطلاع على البحوث والدراسات المرتبطة بموضوع دراسته تم اعداد اختبار تحصيلي معرفي في مادة الرياضيات للصف السادس الابتدائي اشتمل على (62) فقرة وذلك وفقاً للأهداف التعليمية للوحدة، شاملة المستويات الثلاثة من المجال المعرفي، واعتمد أثناء تصميم الاختبار على قياس ما اكتسبه الطالب من معارف ومهارات وقوانين في المحتوى الدراسي، وذلك من خلال فقرات الاختيار من متعدد، حيث تتكون المفردة الاختبارية من جزءين هما (المتن – البدائل)، مع مراعاة الدقة والوضوح في صياغة المتن والبدائل الأربعة، وفقا لما أورده توماس ومايكل (Thomas & Michael, 2002) من أسس عامة عند بناء فقرات اختبار الاختيار من متعدد، وقد مرت عملية بناء الاختبار التحصيلي بالخطوات الآتية:

- تحديد محتوى الاختبار: ينحصر مقياس التحصيل في كتاب الرياضيات للصف السادس الابتدائي، للفصل الدراسي الثاني من العام الدراسي 1442هـ/2020م.
- صياغة الأهداف السلوكية: تمت صياغة الأهداف السلوكية لكي تشمل كامل المقرر لمادة الرياضيات، بحيث يمكن قياسها من خلال فقرات الاختيار من متعدد، وذلك وفق متطلبات الدراسة.
- بناء فقرات الاختبار التحصيلي: حيث لكل هدف سلوكي تم صياغة فقرة اختبارية من نوع الاختيار من متعدد، مع مراعاة شروط صياغة ذلك النوع من الفقرات، من خلال جدول المواصفات المعد في البحث الحالي، علماً أن بعض الأهداف تتطلب أكثر من فقرة لقياسها..
- تحكيم المحتوى والأهداف السلوكية والبنود: قُدمت تلك الفقرات والأهداف إلى نخبة من أهل الخبرة والاختصاص من أساتذة الجامعات وقد بلغ عددهم (10) محكمين، من أجل الاستئناس والأخذ بوجهة نظرهم المعتبرة، واقتراح ما يرونه مناسباً، وتم حذف فقرتين بعد الاطلاع على رأي السادة المحكمين أهل الاختصاص، حيث بلغ عدد الفقرات بصورتها النهائية (60) فقرة، لقياس الأهداف المراد قياسها.
- تم التحقق من الخصائص السيكومترية لأداة البحث وفق الخطوات الآتية:

أولاً: صدق الأداة:

صدق المحتوى: تم التأكد من صدق المحتوى للاختبار، بإعداد لائحة مواصفات لبناء الاختبار، للتأكد من أن الفقرات تحقق درجة تمثيل الموضوعات التي يتكون منها الاختبار، حيث تكون الاختبار من (62) فقرة، ومن ثم عرض الاختبار التحصيلي على (10) محكمين من أعضاء هيئة التدريس بجامعة أم القرى، وجامعة الملك عبد العزيز، ومشرفي ومعلمي الرياضيات، وقد تم حذف فقرتين لعدم ملاءمتها للاختبار، وجاءت نسبة الاتفاق بين المحكمين على باقي العبارات 90%.

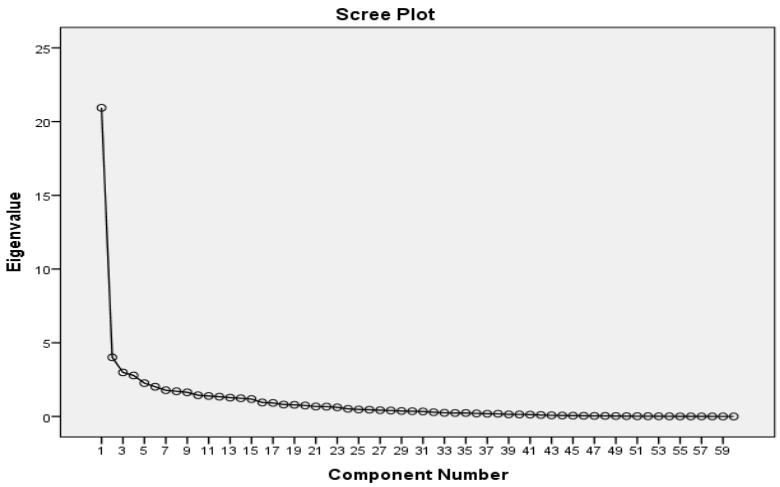
الصدق العاملي: استخدم الباحث الصدق العاملي أيضاً كمؤشر لصدق الاختبار، حيث تم إجراء التحليل العاملي باستخدام طريقة المكونات الرئيسة، وباستخدام التدوير المتعامد لاستخلاص العوامل المسؤولة عن الأداء في الاختبار، حيث تم استخراج قيمة الجذر الكامن، ونسبة التباين المفسر لكل عامل من العوامل.

كما أنه بينت نتائج التحليل العاملي باستخدام التدوير المتعامد، تشبع الفقرات على (15) عاملاً كانت قيمة جذورها الكامنة أعلى من الواحد، حيث فسرت (79.96%) من تباين الأداء على الاختبار، كما يُظهر جدول (1) التالي نسبة التباين المفسر، وقيمة الجذر الكامن، والنسبة التراكمية للتباين المفسر.

جدول (1): **قيم الجذور الكامنة ونسبة التباين المفسر ونسبة التباين التراكمية لكل عامل من العوامل للاختبار**

العامل	الجذر الكامن	نسبة التباين المفسر	التباين المفسر التراكمي	العامل	الجذر الكامن	نسبة التباين المفسر	التباين المفسر التراكمي
1	20.936	34.894	34.894	9	1.637	2.728	66.860
2	4.011	6.685	41.579	10	1.442	2.404	69.264
3	2.987	4.978	46.557	11	1.389	2.315	71.579
4	2.781	4.635	51.192	12	1.341	2.235	73.813
5	2.261	3.768	54.959	13	1.279	2.132	75.945
6	2.017	3.361	58.320	14	1.232	2.053	77.998
7	1.779	2.965	61.286	15	1.179	1.965	79.964
8	1.708	2.847	64.132				

يلاحظ من جدول (1) أن قيمة الجذر الكامن للعامل الأول (20.936) وأنه فسر ما نسبته (34.8%) من التباين الكلي للاختبار، حيث كانت قيمة الجذر الثاني (4.0)، وفسر ما نسبته (6.6%) من التباين الكلي للاختبار، ويُشير ريكاس (Reckas, 1985) أنه اذا كانت نسبة الجذر الكامن للعامل الأول كبيرة بالنسبة للجذر الكامن الثاني، وهذا يُعد مؤشراً على أحادية البعد.



شكل (1): التمثيل البياني لقيم الجذور الكامنة لكل عامل من العوامل للاختبار.

ويتضح من خلال الرسم أن الجذر الكامن للعامل الأول مرتفع جدًا مقارنة بباقي العوامل، وأن هناك انحدارًا شديدًا بين الجذرين الكامنين للعاملين الأول والثاني، مما يدل على أحادية البعد.

ثانياً: التحليل الأولي لفقرات الاختبار:

طُبق الاختبار المكون بصورته النهائية من (60) فقرة على (75) طالباً من طلاب الصف السادس الابتدائي من مدرسة ياسر بن عامر الابتدائية، حيث تم اختيارها عشوائياً من خارج عينة البحث؛ وذلك بهدف معرفة مدى وضوح الفقرات وتعليماته، مع مراعاة ملاحظات الطلاب عن الاختبار، حيث مكن الطلاب من ابداء آرائهم واستفساراتهم عن الفقرات، وكانت التعليمات واضحة لدى مجمل الطلاب، واستغرق أول طالب للانتهاء من الاختبار (55) دقيقة، وأغلب الطلاب انهوا اختبارهم خلال (70) دقيقة، كما أن آخر طالب أنهى اختباره خلال (78) دقيقة، وقد راعى الباحث عند تطبيق الصور الاختبارية أن يزيد الزمن المحدد للاختبار عن (78) دقيقة؛ حتى لا يكون عامل السرعة مؤثرًا في أداء الطلاب على الاختبار، ولم توجد أية ملاحظات تُلزم إجراء بعض التعديلات على فقرات الاختبار.

معاملات الصعوبة والتمييز وفق مبادئ النظرية الكلاسيكية:

تم حساب معامل الصعوبة بحساب النسبة المئوية من الطلاب الذين أجابوا عن الفقرة إجابة صحيحة. كما تم حساب معامل التمييز (اعتماداً على طريقة المقارنة الطرفية)، وكانت المعاملات كما في جدول (2).

جدول (2): **معاملات الصعوبة والتمييز لفقرات الاختبار**.

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
1	0.26	0.35	16	0.67	0.6	31	0.20	0.6	46	0.40	0.25
2	0.26	0.4	17	0.21	0.55	32	0.29	0.8	47	0.52	0.8
3	0.30	0.55	18	0.24	0.75	33	0.21	0.7	48	0.45	0.2
4	0.26	0.5	19	0.27	0.7	34	0.24	0.7	49	0.48	0.2
5	0.28	0.8	20	0.20	0.7	35	0.29	0.75	50	0.51	0.35
6	0.27	0.65	21	0.20	0.5	36	0.23	0.75	51	0.28	0.2
7	0.23	0.65	22	0.22	0.45	37	0.26	0.5	52	0.24	0.25
8	0.29	0.5	23	0.23	0.7	38	0.29	0.55	53	0.47	0.35
9	0.24	0.75	24	0.35	0.8	39	0.29	0.6	54	0.52	0.25
10	0.66	0.65	25	0.20	0.65	40	0.27	0.45	55	0.52	0.2
11	0.36	0.8	26	0.25	0.8	41	0.20	0.5	56	0.25	0.8
12	0.29	0.55	27	0.24	0.7	42	0.29	0.55	57	0.49	0.35
13	0.25	0.7	28	0.29	0.7	43	0.29	0.45	58	0.43	0.2
14	0.23	0.75	29	0.21	0.5	44	0.24	0.5	59	0.49	0.25
15	0.29	0.6	30	0.22	0.75	45	0.37	0.2	60	0.31	0.2

يتضح من جدول (2) أن معاملات الصعوبة للفقرات تراوحت بين (0.2 – 0.67) وهيي معاملات مقبولة وفق ما أشار إليه هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985)، وبمعاملات تمييز تراوحت بين (0.2 – 0.8)، وهي معاملات تمييز مقبولة وجيدة وفق ما أشار إليه ايبل وفريسبي (Ebel & Frisbie, 1991).

ثالثاً: ثبات الأداة:

تم حساب معامل الثبات باستخدام طريقة كيودر – ريتشاردسون 20 (KR–20)، حيث بلغت قيمة معامل ثبات الاختبار (0.9)، وتمثل درجة مقبولة من الثبات، ومما تقدم حول اختبار التحصيل لمادة الرياضيات لطلاب الصف السادس المعد لأغراض البحث الحالي، يتضح أنه يتمتع بالقدر الكافي والمناسب من الصدق والثبات والتمييز والصعوبة.

إجراءات البحث:

أعد الباحث أربعة نماذج للاختبار، حيث كان النموذج الأول عبارة عن (60) فقرة، أما النموذج الثاني فقد كان (30) فقرة، أما النموذج الثالث فقد كان (60) فقرة مرتبة وفق تسلسل المحتوى، والنموذج الرابع (60) فقرة مرتبة بشكل عشوائي، وقد طُبقت نماذج الاختبار بشكل عشوائياً على أفراد العينة والبالغ عددهم (2400) وبعد استلام إجابات الطلاب، تم فرز كل نموذج منفرداً عن الآخر، وتم تصحيح الاختبار، وإدخال البيانات في الحاسب الآلي، حيث استخدم الباحث البرنامج الاحصائي (SPSS) وبرنامج البايلوج (Bilog-Mg3) لتحليل بيانات البحث.

كما تم التحقق من افتراضات نماذج نظرية الاستجابة للفقرة أحادية البعد كما يلي:

أولاً: التحقق من افتراض أحادية البعد:

استخدم الباحث التحليل العاملي للمكونات الرئيسة ومن خلال برنامج (SPSS) لكل اختبار من الاختبارات المنبثقة من الاختبار التحصيلي الرئيس، وتم أولاً التحقق من افتراضاته كما يلي:

تم حساب القيمة المطلقة لمحدد مصفوفة الارتباطات، وتم التأكد من أنها أكبر من (0.00001) في جميع الصورة الاختبارية، حيث عند حساب هذه القيمة للاختبارات الأربعة اتضح أن الناتج أكبر من ذلك فقد كانت (0.0004)، (0.0003)، (0.0005)، (0.0004) على التوالي.

ومن خلال قيمة مربع كاي لاختبار برتليت، تحقق الباحث من تجانس العينة واختبار فرضية عدم تطابق مصفوفة الارتباط الأصلية، فقد أشارت النتيجة إلى أن مربع كاي قيمته في الاختبارات الأربع كانت (14316.711) و (6197.356) و (7490.089) و (13280.990) بدرجات حرية (1770) و (435) و (1770) و (1711)، وجميعها دالة عند مستوى دلالة (0.01)، مما يشير إلى أن البيانات مناسبة لإجراء التحليل العاملي.

ومن خلال اختبار كايزر – أولكين (KMO)، تم التحقق من مدى كفاية حجم العينة لإجراء التحليل العاملي، حيث يجب أن لا يقل عن (0.5) بناء على محك كايزر، حيث كانت قيمة (KMO) في الاختبارات الأربع أكبر من محك كايزر الذي قيمته (0.5)، وهذا يشير إلى كفاية حجم العينة للتحليل العاملي.

وتم التحقق من افتراض أحادية البعد باستخدام التحليل العاملي الاستكشافي، مستخدماً طريقة المكونات الرئيسة، مستعملاً التدوير المتعامد لإيجاد قيم الجذور الكامنة للعاملين الأول والثاني ونسبة التباين المفسر في الصور الاختبارية الأربعة، كما يبينه جدول (3) التالي:

جدول (3) **قيم الجذور الكامنة للعاملين الأول والثاني ونسبة التباين المفسر في الصور الاختبارية الأربعة**

الصورة الاختبارية	العامل	الجذر الكامن	نسبة التباين المفسر
الأولى	1	17.662	29.436
	2	2.46	3.409
الثانية	1	9.711	32.369
	2	1.499	4.997
الثالثة	1	20.313	33.855
	2	2.391	3.985
الرابعة	1	17.401	29.493
	2	2.573	4.361

يلاحظ من جدول (3) السابق أن قيمة الجذر الكامن للعامل الأول أكبر بكثير من قيمة الجذر الكامن للعامل الثاني، والنسبة بينهما تفوق القيمة (2) التي أشار إليها ريكاس (Reckas, 1985) للدلالة على تحقق أحادية البعد.

ثانياً: التحقق من افتراض الاستقلال الموضعي: عندما يتحقق افتراض أحادية البعد يعني تحقق افتراض الاستقلال الموضعي، حيث أنهما مفهومان متكافئان (Hambleton & Rogers, 1999).

ثالثاً: التحقق من التحرر من السرعة في الأداء: تم التحقق من أن الاختبار ليس اختبار سرعة من خلال حصر المفحوصين الذين لم يستطيعوا اكمال الاختبار محل التطبيق، حيث كان ما نسبته 96% من الطلاب قد انهوا الاختبار قبل الوقت المحدد لتسليم الإجابات.

نتائج البحث:

قبل الإجابة عن تساؤلات البحث، تم تحليل البيانات وفقاً للنموذج الأكثر ملاءمة ودقة للفقرات، وهو النموذج اللوغاريتمي ثلاثي المعلمة، وقد تم استخدام برنامج Bilog_mg3، والتأكد من مدى

ملاءمة الفقرات لافتراضات النموذج، وتم تحويل الصعوبة المقدرة بوحدة اللوجيت إلى صعوبة مقدرة بوحدة المنف، وكذلك القدرة المقدرة بوحدة اللوجيت إلى قدرة مقدرة بوحدة المنف؛ وذلك لتجنب القيم السالبة، من خلال المعادلتين:

الصعوبة بالمنف = (الصعوبة باللوجيت 5x) +50 (أبو جراد، ٢٠٠٨).

القدرة بالمنف = (القدرة باللوجيت 5x) + 50 (أبو جراد، ٢٠٠٨).

ثم تم حساب المتوسط الحسابي والانحراف المعياري لكل بارامتر من بارامترات الفقرات، وكذلك للأخطاء المعيارية في تقدير بارامترات الفقرات، وجاءت النتائج كما في جدول (4).

جدول(4): قيم المتوسط الحسابي والانحراف المعياري لبارامترات الفقرات والخطأ المعياري في تقديرها في الصور المختلفة.

الصورة	البارامتر	المتوسط	الانحراف المعياري	الصورة	البارامتر	المتوسط	الانحراف المعياري
الأولى	الصعوبة بالمنف	48.35	6.93	الثالثة	الصعوبة	44.42	7.94
	الخطأ المعياري للصعوبة	50.88	0.51		الخطأ المعياري للصعوبة	51.71	1.52
	التمييز	1.22	0.65		التمييز	1.10	0.61
	الخطأ المعياري للتمييز	0.19	0.14		الخطأ المعياري للتمييز	0.22	0.21
	التخمين	0.01	0.03		التخمين	0.003	0.01
	الخطأ المعياري للتخمين	0.014	0.013		الخطأ المعياري للتخمين	0.009	0.005
	قدرات الأفراد بالمنف	49.99	54.98		قدرات الأفراد	50.01	54.75
	الخطأ المعياري للقدرة	0.236	0.005		الخطأ المعياري للقدرة	0.321	0.103
الثانية	الصعوبة بالمنف	45.65	5.86	الرابعة	الصعوبة	39.53	17.63
	الخطأ المعياري للصعوبة	51.59	0.74		الخطأ المعياري للصعوبة	55.40	5.2
	التمييز	0.55	0.25		التمييز	0.95	0.60
	الخطأ المعياري للتمييز	0.10	0.03		الخطأ المعياري للتمييز	0.27	0.19
	التخمين	0.001	0.0001		التخمين	0.005	0.012

الخطأ المعياري للتخمين	0.008	0.001	الخطأ المعياري للتخمين	0.008	0.13
قدرات الأفراد	45.65	54.33	قدرات الأفراد	43.98	53.88
الخطأ المعياري للقدرة	0.579	0.13	الخطأ المعياري للقدرة	0.631	2.767

الإجابة عن التساؤل الأول:

ينص على: "ما أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وفقاً للنموذج اللوجستي ثلاثي البارامتر؟"، وللإجابة عن هذا التساؤل، تم استخدام اختبار ت للمقارنة بين متوسطي بارامترات الصعوبة والتمييز والتخمين وكذلك متوسطي الأخطاء المعيارية في تقدير كل منهم، وذلك في الصورتين الأولى والثانية المختلفتين فقط في عدد فقراتهما، وجاءت النتائج كما في جدول (5).

جدول(5): اختبار (ت) لدراسة دلالة الفروق بين متوسطي بارامترات الفقرات، وكذلك متوسطي الأخطاء المعيارية في دقة تقديرها في الصورتين الأولى والثانية

دلالة فروق متوسطي بارامتر صعوبة الفقرات بالمنف					
	العدد	**المتوسط**	**الانحراف المعياري**	**ت**	**الدلالة**
فقرات الاختبار الأول	**60**	**48.35**	**6.93**	**2.89**	**دالة عند**
فقرات الاختبار الثاني	**30**	**45.65**	**5.86**		**0.01**
دلالة فروق متوسطي بارامتر تمييز الفقرات					
	العدد	**المتوسط**	**الانحراف المعياري**	**ت**	**الدلالة**
فقرات الاختبار الأول	**60**	**1.22**	**0.65**	**5.29**	**دالة عند**
فقرات الاختبار الثاني	**30**	**0.55**	**0.25**		**0.01**
دلالة فروق متوسطي بارامتر تخمين الفقرات					
	العدد	**المتوسط**	**الانحراف المعياري**	**ت**	**الدلالة**
فقرات الاختبار الأول	**60**	**0.01**	**0.03**	**1.52**	**غير دالة**
فقرات الاختبار الثاني	**30**	**0.001**	**0.0001**		
دلالة فروق متوسطي الأخطاء المعيارية في دقة تقدير بارامتر الصعوبة					
	العدد	**المتوسط**	**الانحراف المعياري**	**ت**	**الدلالة**
فقرات الاختبار الأول	**60**	**50.88**	**0.51**	**5.27-**	**دالة عند**
فقرات الاختبار الثاني	**30**	**51.59**	**0.74**		**0.01**

دلالة فروق متوسطي الأخطاء المعيارية في دقة تقدير بارامتر التمييز

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
فقرات الاختبار الأول	60	0.19	0.14	3.478	دالة عند 0.01
فقرات الاختبار الثاني	30	0.10	0.03		

دلالة فروق متوسطي الأخطاء المعيارية في دقة تقدير بارامتر التخمين

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
فقرات الاختبار الأول	60	0.014	0.013	1.82	غير دالة
فقرات الاختبار الثاني	30	0.008	0.001		

ويتضح من جدول (5) وجود فروق بين متوسطي صعوبة الفقرات للاختبار الأول (60) ومتوسط صعوبة فقرات الاختبار الثاني (30) فقرة، حيث أن متوسط معلمة الصعوبة لفقرات الاختبار الأول (60) فقرة كان أعلى من متوسط معلمة صعوبة الاختبار الثاني (30) فقرة، كما يتضح أيضًا وجود فروق ذات دلالة إحصائية بين متوسطي الأخطاء المعيارية في دقة تقدير بارامتر صعوبة الفقرات، وتتفق هذه النتيجة مع دراسة (بني عطا، 2014؛ الدرابيع، 2014؛ Barnes & Wise, 1991 Fitzpatric& Ann, 2001; Glass, 2005; Hambleton & Cook,1980) التي أظهرت وجود أثر لعدد فقرات الاختبار في دقة تقدير بارامترات الفقرات حيث كلما زادت عدد فقرات الاختبار، وتختلف هذه النتيجة مع دراسة الحواري (2015)، التي أظهرت نتائجها وجود أثر لصالح الاختبار ذو (30 فقرة) أي أنه الأكثر تقديراً لبارامتر الصعوبة إذ زُود بمعلومات أكثر.

كما يتضح وجود فروق بين متوسطي تمييز الفقرات للاختبار الأول (60) ومتوسط تمييز فقرات الاختبار الثاني (30) فقرة، حيث أن متوسط معلمة التمييز لفقرات الاختبار الأول (60) فقرة كان أعلى من متوسط معلمة التمييز للاختبار الثاني (30) فقرة، كما يتضح أيضًا وجود فروق ذات دلالة إحصائية بين متوسطي الأخطاء المعيارية في دقة تقدير بارامتر تمييز الفقرات، وتتفق هذه النتيجة مع دراسة (بني عطا، 2014؛ الدرابيع، 2014؛ Barnes & Wise, 1991 Fitzpatric& Ann, 2001; Glass, 2005; Hambleton & Cook,1980) التي أظهرت وجود أثر لعدد فقرات الاختبار في دقة تقدير بارامترات الفقرات حيث كلما زادت عدد فقرات

الاختبار، وتختلف هذه النتيجة مع دراسة الحواري (2015)، التي أظهرت نتائجها وجود أثر لصالح الاختبار ذو (30 فقرة) أي أنه الأكثر تقديراً لبارامتر التمييز إذ زُود بمعلومات أكثر.

كما يتضح أيضًا عدم وجود فروق ذات دلالة إحصائية بين متوسطي بارامترات التخمين، ومتوسطي الخطأ المعياري في تقديرها تبعاً لنموذجي الاختبار (60، 30) فقرة.

الإجابة عن التساؤل الثاني:

ينص التساؤل الثاني على: "ما أثر عدد فقرات اختبار الاختيار من متعدد على دقة تقدير قدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البارامتر؟

وللإجابة عن هذا التساؤل، تم استخدام اختبار ت للمقارنة بين متوسطي بارامترات قدرات المختبرين وكذلك متوسطي الأخطاء المعيارية في تقديرها، وذلك في الصورتين الأولى والثانية المختلفتين فقط في عدد فقراتهما، وجاءت النتائج كما في جدول (6).

جدول(6): اختبار (ت) لدراسة دلالة الفروق بين متوسطي بارامترات قدرات المختبرين، وكذلك متوسطي الأخطاء المعيارية في دقة تقديرها في الصورتين الأولى والثانية

دلالة فروق متوسطي قدرة المختبرين بوحدة المنف					
العينة	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
قدرات الأفراد في الاختبار الأول	600	49.99	4.98	3.12	دالة عند 0.01
قدرات الأفراد في الاختبار الثاني	600	45.65	2.88		
دلالة فروق متوسطي الأخطاء المعيارية لتقدير قدرة المختبرين					
العينة	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
قدرات الأفراد في الاختبار الأول	600	0.236	0.005	-62.01	دالة عند 0.01
قدرات الأفراد في الاختبار الثاني	600	0.579	0.13		

ويتضح من جدول (6) وجود فروق ذات دلالة إحصائية بين متوسطي قدرة الأفراد، ومتوسطي الأخطاء المعيارية في تقدير القدرة، وفقاً لنموذجي الاختبار (60، 30) فقرة، حيث

يلاحظ أنه فيما يخص متوسط القدرة بالمنف، أن متوسط الاختبار (60) فقرة كان أعلى من متوسط الاختبار (30)، مما يدل على أن عدد فقرات الاختبار الأول أثرت على قدرة الأفراد، وتتفق هذه النتيجة مع دراسة (بني عطا، 2014؛ الدرابيع، 2014؛ Barnes & Wise, 1991; Fitzpatric& Ann, 2001; Glass, 2005; Hambleton & Cook,1980)، وتختلف هذه النتائج مع دراسة (الحواري، 2015)، التي أظهرت نتائجها وجود أثر لعدد فقرات الاختبار في دقة تقدير قدرات الأفراد لصالح الاختبار ذو (30 فقرة) أي أنه الأقل فقرات إذ زُود بمعلومات أكثر، ومن ثم فإنه للحصول على تقديرات جيدة للقدرة، لابد من زيادة عدد فقرات الاختبار.

الإجابة عن التساؤل الثالث:

ينص التساؤل على: "ما أثر طريقة ترتيب فقرات اختبار الاختيار من متعدد على دقة تقدير بارامترات الفقرات وفقاً للنموذج اللوجستي ثلاثي البرامتر؟"

تم استخدام اختبار ت للمقارنة بين متوسطي بارامترات الصعوبة والتمييز والتخمين وكذلك متوسطي الأخطاء المعيارية في تقدير كل منهم، وذلك في الصورتين الثالثة والرابعة المختلفتين فقط في طريقة ترتيب فقراتهما (حسب المحتوى – عشوائيًا) على الترتيب، وجاءت النتائج كما في جدول (7).

جدول(7): اختبار (ت) لدراسة دلالة الفروق بين متوسطي بارامترات الفقرات، وكذلك متوسطي الأخطاء المعيارية في دقة تقديرها في الصورتين الثالثة والرابعة

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
دلالة فروق متوسطي بارامتر صعوبة الفقرات بالمنف					
فقرات الاختبار الثالث	60	44.42	7.94	2.18	دالة عند
فقرات الاختبار الرابع	60	39.53	17.63		0.05
دلالة فروق متوسطي بارامتر تمييز الفقرات					
فقرات الاختبار الثالث	60	1.10	0.61	1.32	غير دالة
فقرات الاختبار الرابع	60	0.95	0.6		
دلالة فروق متوسطي بارامتر تخمين الفقرات					

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
فقرات الاختبار الثالث	60	0.003	0.009	-1.071	غير دالة
فقرات الاختبار الرابع	60	0.005	0.197		

دلالة فروق متوسطي الأخطاء المعيارية في دقة تقدير بارامتر الصعوبة

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
فقرات الاختبار الثالث	60	51.71	1.52	-5.27	دالة عند 0.01
فقرات الاختبار الرابع	60	55.40	5.2		

دلالة فروق متوسطي الأخطاء المعيارية في دقة تقدير بارامتر التمييز

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
فقرات الاختبار الثالث	60	0.22	0.21	-1.24	غير دالة
فقرات الاختبار الرابع	60	0.27	0.19		

دلالة فروق متوسطي الأخطاء المعيارية في دقة تقدير بارامتر التخمين

	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
فقرات الاختبار الثالث	60	0.009	0.005	0.29	غير دالة
فقرات الاختبار الرابع	60	0.008	0.13		

يتضح من جدول (7) وجود فروق بين متوسطي صعوبة الفقرات للاختبار الثالث (مرتب حسب المحتوى) ومتوسط بارامترات صعوبة فقرات الاختبار الرابع (بشكل عشوائي)، حيث أن متوسط بارامترات الصعوبة لفقرات الاختبار الثالث (مرتب حسب المحتوى) كان أعلى من متوسط بارامترات صعوبة الاختبار الرابع (بشكل عشوائي)، كما يتضح وجود فروق بين متوسطي الأخطاء المعيارية في دقة تقدير بارامترات الصعوبة وفقاً لنموذجي الاختبار (مرتب حسب المحتوى، بشكل عشوائي)، مما يؤكد وجود فروق ذات دلالة إحصائية بين متوسطي الأخطاء المعيارية في دقة تقدير بارامترات صعوبة الفقرات.

كما يتضح عدم وجود فروق بين متوسطي تمييز الفقرات للاختبار الثالث (مرتب حسب المحتوى)، ومتوسط تمييز فقرات الاختبار الرابع (بشكل عشوائي)، وكذلك عدم وجود فروق ذات دلالة إحصائية بين متوسطي الأخطاء المعيارية في دقة تقدير بارامتر تمييز الفقرات.

وكذلك يتبين عدم وجود فروق ذات دلالة إحصائية بين متوسطي بارامترات التخمين ومتوسط الخطأ المعياري في تقديرها تبعاً لنموذجي الاختبار (مرتب حسب المحتوى، بشكل عشوائي).

أي أن ترتيب فقرات اختبار الاختيار من متعدد حسب المحتوى يزيد من دقة تقديرات بارامترات الصعوبة، ولكن ليس لترتيب الفقرات أثر على دقة تقديرات بارامتري التمييز والتخمين.

الإجابة عن التساؤل الرابع:

ينص التساؤل على: "ما أثر طريقة ترتيب فقرات اختبار الاختيار من متعدد على دقة تقدير قدرات المختبرين وفقاً للنموذج اللوجستي ثلاثي البرامتر؟"

تم استخدام اختبار ت للمقارنة بين متوسطي بارامترات قدرات المختبرين وكذلك متوسطي الأخطاء المعيارية في تقديرها، وذلك في الصورتين الثالثة والرابعة المختلفتين فقط في طريقة ترتيب فقراتهما (حسب المحتوى – عشوائيًا) على الترتيب، وجاءت النتائج كما في جدول (8).

جدول(8): اختبار (ت) لدراسة دلالة الفروق بين متوسطي بارامترات قدرات المختبرين، وكذلك متوسطي الأخطاء المعيارية في دقة تقديرها في الصورتين الثالثة والرابعة

دلالة فروق متوسطي قدرة المختبرين بوحدة المنف				
العينة	العدد	المتوسط	الانحراف المعياري	ت	الدلالة
قدرات الأفراد في الاختبار الثالث	600	50.006	4.75	2.96	دالة عند 0.01
قدرات الأفراد في الاختبار الرابع	600	43.982	2.32		
دلالة فروق متوسطي الأخطاء المعيارية لتقدير قدرة المختبرين					
العينة	العدد	المتوسط	الانحراف المعياري	ت	الدلالة

قدرات الأفراد في الاختبار الثالث	600	0.321	0.103	2.739-	دالة عند 0.01
قدرات الأفراد في الاختبار الرابع	600	0.631	2.767		

ويتضح من جدول (8) وجود فروق ذات دلالة إحصائية بين متوسطي قدرة الأفراد، ومتوسطي الأخطاء المعيارية في تقدير القدرة، وفقاً لنموذجي الاختبار (مرتب حسب المحتوى، بشكل عشوائي)، حيث يلاحظ أنه فيما يخص متوسط القدرة بالمنف، أن متوسط الاختبار (المرتب حسب المحتوى) كان أعلى من متوسط الاختبار (بشكل عشوائي)، مما يدل على أن ترتيب فقرات الاختبار حسب المحتوى أثرت على قدرة الأفراد، وتختلف هذه النتائج مع نتائج دراسات (الليحاني، 2010؛ الحاج، 2015؛ سليمان، 2015؛ سالم وعليوة، 2018؛ Sue, 2009; Kagundu & Ross, 2015; Satti, et al, 2019)، التي أثبتت أنه لا توجد فروق ذات دلالة إحصائية مهما رُتبت فقرات الاختبار، سواءاً رُتبت حسب المحتوى أو بشكل عشوائي، كما تختلف مع نتائج دراسة (Naibi, 2013) التي أظهرت النتائج وجود أثر لترتيب الفقرات لصالح النموذج المرتب بشكل عشوائي.

توصيات البحث:

1. اعتماد عدد فقرات الاختبار (60) فقرة في الاختبارات التحصيلية، لأنها تعطي تقديرات أكثر دقة لمعالم القدرة للمفحوصين ومعالم الفقرة.
2. اعتماد ترتيب الفقرات بحسب تسلسل المحتوى في الاختبارات التحصيلية، حيث أنها تعطي تقديرات أكثر دقة لمعالم القدرة للمفحوصين وكذلك معالم الفقرة.
3. اعداد وتهيئة نخبة من أعضاء هيئة التدريس المختصين في بناء وتصميم الاختبارات، وإيجاد صور متكافئة أيضاً للاختبارات.
4. العمل على دورات تدريبية لأعضاء هيئة التدريس، لبرنامج (Bilog-mg3) والحث عليه، لتحليل خصائص فقرات الاختبار، وتقدير قدرات الأفراد.
5. توجيه القائمين على بناء الاختبارات في جميع المراحل التعليمية بأهمية الاستفادة من نظرية الاستجابة للفقرة، لما تحتويه هذه النظرية من كم هائل ومميزات تساهم في رفع كفاءة الاختبارات.

الدراسات والبحوث المقترحة:

1. إجراء دراسات أخرى تتناول الكشف عن أثر عدد فقرات الاختبار بطرق أخرى مختلفة وفق دراسة متغيرات أخرى مثل (حجم العينة، شكل الفقرات، عدد البدائل ... الخ).
2. إجراء دراسات أخرى تتناول الكشف عن أثر ترتيب فقرات الاختبار بطرق أخرى مختلفة وفق دراسة متغيرات أخرى مثل (حجم العينة، شكل الفقرات، عدد البدائل ... الخ)
3. إجراء دراسة أخرى تكشف أثر عدد الفقرات وترتيبها في مواد دراسية أخرى.
4. إجراء دراسة أخرى للكشف عن أثر عدد الفقرات وترتيبها في المقاييس النفسية والتربوية على معالم الأفراد والفقرات ومقارنتها بالاختبارات التحصيلية.
5. إجراء دراسة أخرى، للكشف عن أثر عدد الفقرات وترتيبها باستخدام نموذج راش، والنموذج اللوجيستي ثنائي المعلمة.
6. إجراء دراسة أخرى تبحث أثر عدد فقرات الاختبار وترتيبها وفقاً للنظرية التقليدية للقياس ومقارنتها مع النظرية الحديثة لاستجابة الفقرة.

المراجع:

أبو جراد، حمدي يونس (٢٠٠٨). استخدام نموذج (راش) في تطوير اختبار كاتل الثالث للذكاء الصورة (أ). *مجلة الجامعة الإسلامية*. شؤون البحث العلمي والدراسات العليا بالجامعة الإسلامية. غزة، فلسطين، *16* (٢). 555-583.

بني عطا، زايد صالح (2014). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة. *مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية*، 11 (2). 1-37.

الحاج، فريال محمود (2015). أثر كل من ترتيب فقرات الاختبار ومستوى القلق على أداء الطلاب في اختبار من نوع الاختيار من متعدد. *مكتب التربية العربي لدول الخليج*. *36* (136). 15-30.

الحواري، أروى (2015). أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وفق نموذج راش في نظرية استجابة الفقرة. *مجلة جامعة النجاح للأبحاث*، العلوم الإنسانية، *29* (8). 1463-1488.

الدرابيع، ماهر (2001). فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة ((نموذج راش)) في دقة تقدير قدرة الفرد ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار. *مجلة دراسات العلوم الإنسانية*، 11 (6). 208- 197.

رمضان، رمضان (2010). أثر ترتيب فقرات الاختبار حسب درجة صعوبتها على تحصيل طلاب المرحلة الثانوية لمادة الرياضيات، *مجلة الأبحاث التربوية*. كلية التربية. جامعة الأزهر (8)، 157-176.

سالم، هانم أحمد و عليوة، محمد مصطفى (2018). تأثير اختلاف ترتيب مفردات الاختبار التحصيلي مرجعي المحك على معاملات الصعوبة والتمييز لمفردات الاختبار والأداء الأكاديمي وفق نظرية الاستجابة للمفردة لدى طلاب كلية التربية. *مجلة كلية التربية*. جامعة بنها. 116، 58-115.

سليمان، شاهر خالد (2015). أثر ثلاث نماذج لترتيب فقرات اختبار تحصيلي في التقويم التربوي على الخصائص السيكومترية للاختبار وفقراته: *مجلة كلية التربية*. جامعة الأزهر، *165*، 15-30.

ظاظا، حيدر، (2012). الكشف عن مدى انتهاك قواعد صياغة فقرة الاختيار من متعدد في أسئلة امتحان شهادة الثانوية العامة في الأردن، *المجلة الأردنية للعلوم التربوية*، *8* (4). 81-91.

عبابنة، سناء حسين (2017). *أثر ترتيب فقرات اختبار اختيار من متعدد تبعاً لعدد خطوات حلها على الخصائص السيكومترية للاختبار وفقراته وأداء الطلبة عليه*. رسالة ماجستير غير منشورة، جامعة اليرموك.

العبادي، رائد خليل، (2006). *الاختبارات المدرسية*. عمان: مكتبة المجتمع العربي.

علام، صلاح الدين، (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. القاهرة: دار الفكر العربي.

علام، صلاح الدين، (2011). *القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة: دار الفكر العربي.

العمري، حسان (2011). أثر ترتيب الفقرات في المقاييس النفسية على التجانس الداخلي للمقياس وخصائص فقراته: *مجلة جامعة دمشق*.*43* (27)، 111- 129.

عودة، أحمد، (2010). *القياس والتقويم في العملية التدريسية*. اربد: دار الأمل.

الليحاني، عدنان (2010). *أثر ترتيب فقرات الاختبار على صدق وثبات اختبار اختيار من متعدد البدائل في مادة رياضيات المرحلة الثانوية بمكة المكرمة*. رسالة ماجستير غير منشورة. جامعة أم القرى بمكة المكرمة.

الهويدي، زيد، (2004). *أساسيات القياس والتقويم التربوي*. العين: دار الكتاب الجامعي.

Agresti, A., & Finnlay, B. (2009). *Statistical Methods for the Social Science* (4th Ed). Upper Saddle River, NJ: Person- Prentice Hall.

Barnes, L.& Wise (1991). The Utility of AModified One - Parameter IRT Model with Small Samples, *Journal of Applied Measurement in Education, 4* (2),143– 153.

Bond TG, Fox CM. (2007). *Applying the Rasch Model – Fundamental Measurement in the Human Sciences*. Mahwah, NY: Lawrence Erlbaum Associates.

Chen, H. (2012). The Moderating Effects of Item order Arranged by Difficulty on the Relationship between Test Anxiety and Test Performance. *Creative Education, 3* (3), 328-333.

Doerner, W. M. & Calhoun, J. P. (2009). *The Impact of the Order of Test Questions in Introductory Economics*, (April 2, 2009). Available at SSRN: https://ssrn.com/ abstract=1321906 or http://d x.doi.org/10.2139/ ssrn.1321906.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of Educational Measurement* (5th ed.). New Jersey, NJ: Prentice-Hall, Englewood Cliffs.

Embretson, S. E. & Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey, Mahwah Lawrence Erlbaum Associates Publishers.

Fitzpatric, A. R. (2009). The Impact of Anchor Test Configuration on Student Proficiency Rates. Educational Measurement: *Issues and Practice*. *27* (4). 34-40 Win 2008.

Fitzpatrick,& Ann. R. (2001). The Effects of Test Length and Sample Size on the Reliability and Equating of Tests Composed of Constructed – Response Item. *Applied Measurement in Education, 14* (1): 412-425.

Glass, G. (2005). The Impact of Item Parameter Estimation of Computerized Adaptive Testing with Item Cloning. Law School Admission Council Computerized Testing Report 02-06 November.

Gronlund, N. & Linn, R. (1990). *Measurement and Evaluation in Teaching.* New York: Macmillan publishing Co., Inc.

Gronlund, N. (1997). *Measurement and Evaluation in Teaching*. London: Macmillan Co., LTD.

Hambleton, R. (24-25/nov/2004).Traditional and Moderns Approaches to Outcomes Measurement, *Paper presented on Conference on Improving Health Outcomes Assessment based on Modern Measurement Theory and Computerized Adaptive Testing.* Retrieved from www.cancer.gov/irt/ hambleton.pdf Date of entry 12/11/2015 at 10:00pm

Hambleton, R., Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff.

Hambleton, R.K. Cook, L. (1980). *The Robustness of Latent Trait Models and Effect of Test Length and Sample Size on the Precision of Ability Estimate*. Boston: Kluwer Nijhoff Publishers Group.

Kagundu, P., & Ross, G. (2015). The Impact of Question Order on Multiple Choice Exams on Student Performance in an Unconventional Introductory Economics Course. *Journal for Economic Educators,* 15(1), 19-36.

Lord, F. (1980). *Application of Item Response Theory to Practical Testing Problems.* New Jersey: Lawrence Erlbaum Associates, Publishers.

McMillan, W. (2004). Classroom Assessment: Principles and Effective Instruction. *Journal Educational Measurement, 22* (3), 4-54.

Naibi, L. (2013). Effect of Item Arrangement on Test Reliability Coefficients: Implications for Testing. *Journal of Research in Education and Society, 4* (3), 54- 62.

Nitko, A. (2001). *Educational Assessment of Students* (3rd ed.). New Jersey: Prentice-Hall, Inc

Pommerich, M. (2007). The Effect of Using Item Parameters Calibrated from Paper Administrations in Computer Adaptive Test Administrations". *Journal of Technology. Learning. and Assessment .5*(7). 231-135.

Reckase, M. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implicions. *Journal Educational statislics,*4,207-230.

Rup, A. & Zumbo, B. (2006). Understanding Parameter Invariance in Unidimensional IRT Models. *Educational and Psychological Measurement, 66* (2), 63-84.

Satti, I., Hassan, B., Alamri, A., Khan, M. A. & Patel, A. (2019). The Effect of Scrambling Test Item on Students' Performance and Difficulty Level of MCQS Test in a College Medicine, KKU. *Creative Education, 10*, 1813- 1818.

Sinharay, S, & Holland, P, (2006).*The Correlation between the Scores of a Test,* Research Report, Ets, March.

Sue, D. L. (2009). The Effect of Scrambling Test Questions on Student Performance in a Small Class Setting. *Journal for Economic Educators,* 9 (1), 32-41.

Tay, L., Meade, M., & Smith, M. (2015). An Overview and Practical Guide to IRT Measurement Equivalence Analysis. *Organizational Research Methods*,18, 3-46.