المجلد 19 العدد 2 يـونـيـو 2018

الخصائص السيكومترية ومعالم الفقرات والقدرة لاختبار الاختيار من متعدد باختلاف عدد البدائل عند استخدام النموذج ثلاثي المعلمة

د. خالد محمد حياصات
قسم القياس والتقويم
كلية العلوم التربوية – جامعة القصيم
khaledhyasat75@yahoo.com

الخصائص السيكومترية ومعالم الفقرات والقدرة لاختبار الاختيار من متعدد باختلاف عدد البدائل عند استخدام النموذج ثلاثي المعلمة

د. خالد محمد حياصات
قسم القياس والتقويم
كلية العلوم التربوية – جامعة القصيم

الملخص

سعت هذه الدراسة للكشف عن أثر عدد البدائل في اختبار الاختيار من متعدّد على الخصائص السيكومترية للاختبار، والقدرة، ومعالم الفقرات، وفق النموذج الثلاثي المعلمة. ولتحقيق هدف الدراسة قام الباحث ببناء اختبار تحصيلي من نوع الاختيار من متعدّد في الجزء الأول من مبحث الرياضيات المقرّر تدريسه لطلاب الصف الأول الثانوي في المدارس الحكومية في المملكة الاردنية الهاشمية، تكوّن من (٣٢) فقرة بصورته النهائية، وله ثلاثة نماذج تختلف فقط في عدد بدائل فقراتها؛ أربعة بدائل للنموذج الأول، وثلاثة بدائل للنموذج الثاني، وبديلين للنموذج الثالث.

تكوّنت عينة الدراسة من (١٥٠٠) طالبًا من طلبة الصف الأول ثانوي للعام الدراسي (٢٠١٥/٢٠١٦) في المملكة الأردنية الهاشمية. تمّ تحليل البيانات باستخدام الرزمة الإحصائية للعلوم الاجتماعية (SPSS) وبرمجية BILOG-MG في إطار نظرية الاستجابة للفقرة، حيث تم الحصول على تقديرات لقدرات الأفراد وتقديرات لمعالم الفقرات، وقيم معاملات الثبات لكل اختبار.

أظهرت النتائج وجود فروق ذات دلالة إحصائية (α=٠,٠٥) بين المتوسطات الحسابية لتقدير معلمة التمييز تُعزى لعدد بدائل الاختبار، ولصالح النموذج ذي الأربعة بدائل للفقرة. كما أظهرت النتائج وجود فروق دالة إحصائيًا بين المتوسطات الحسابية لقيم معاملات الثبات ولصالح النموذج ذي الأربعة بدائل يليه النموذج ذو الثلاثة بدائل. وأظهرت النتائج المتعلقة بدقة تقديرات معلمة القدرة للمفحوصين وجود فروق دالة احصائيًا، ولصالح نموذج الاختبار ذيالأربعة بدائل، يليه نموذج الاختبار ذو البديلين.

الكلمات المفتاحية: عدد البدائل، اختبار الاختيار من متعدّد، النموذج ثلاثي المعلمة.

* تاريخ تسلم البحث: ٢٠١٧/٢/٤م * تاريخ قبوله للنشر: ٢٠١٧/٦/٢٨م

Psychometric Properties, Item Difficulty and Ability Parameters of a Multiple Choice Test Estimated According to the Number of Options Using the 3-Parameter Model

Dr. Khaled M. A. Al-hyasaat
Colleage of Education Science
Alqasiem University

Abstract

This study aimed at investigating the effect of using a number of alternatives in multiple choice tests on the students' ability, and the estimation of the item difficulty. To achieve the aim of the study, three forms of a multiple choice test consisting of 32 items were constructed to measure students' achievement in mathematics. The forms were different in the number of options. Each item was constructed to have four, three or two alternatives.

The sample of the study consisted of 1500 male students in the 8th grade in schools in Jordan during the 2016/2017 academic year. Each of the three test forms was administered to 500 students.

The three test forms were administered and scored objectively. The statistical Package for Social Sciences (SPSS) and the BILOG-MG3 programs were used to analyze the data, provide estimates of the individual's abilities, item parameters, standard error of measurement, and test reliability for each test form.

The findings of the study indicated statistically significant differences (α= 0.05) between the discrimination parameter mean scores due to the test form in favor of the four-option alternative. Also, the findings showed statistically significant differences between the values of the reliability coefficents in favor of the four- and three-option test forms. As for the results related to the accuracy of the ability parameter estimate, the findings of the study indicated statistically significant differences in favor of the four-, followed by the two- and three option forms of the test.

Keywords: multiple choice test, number of options, the three-parameter logistic model.

المجلد 19 العدد 2 يونيو 2018

الخصائص السيكومترية ومعالم الفقرات والقدرة لاختبار الاختيار من متعدد باختلاف عدد البدائل عند استخدام النموذج ثلاثي المعلمة

د. خالد محمد حياصات
قسم القياس والتقويم
كلية العلوم التربوية – جامعة القصيم

مقدمة

تُعدّ الاختبارات واحدة من وسائل التقويم المتنوعة التي يُستند إليها في اتخاذ القرارات المهمة التي تخصّ الفرد والمجتمع. وهي من أهم الأدوات التي تزودنا بالمعلومات والتي يستند إليها في اتخاذ العديد من القرارات الهامة التي تخص الفرد والمجتمع، حيث تصمم الاختبارات لأهداف متنوعة منها: اختيار شخص لوظيفة ما، أو لأغراض التصنيف كتحديد مسار المتعلمين بما يتناسب مع قدراتهم ومهاراتهم، وفي تقويم تحصيل المتعلمين من خلال الدرجات التي يحصلون عليها في الاختبارات الصفية، وعملية تقويم الأفراد تكتسب أهمية كبيرة بقدر أهمية القرارات المبنية عليها، وبقدر خطورة القرارات الخاطئة التي يمكن أن تترتب على ذلك في المواقف والمجالات المتعددة على مستوى الفرد والمجتمع إلى حد قد يصعب معالجته، أو قد تحتاج إلى وقت طويل ممّا يُعيق عملية التنمية ومواكبة تطور المجتمعات الأخرى، وبذلك يمكن العمل على تحسين وتطوير العملية التربوية والتعليمية والسير بها إلى الأفضل عن طريق تطوير هذه الاختبارات اللفظية أو الأدائية، وتحسين قدرتها لقياس نواتج التعلم (Haladyna, 2004).

تصنف الاختبارات تبعًا لمعايير مختلفة، كأن تُصنّف حسب شكل الفقرة أو نوعها Item-Form، حيث يمكن تصنيفها إلى: ذات الإجابة المنتقاة، وتأتي على شكل فقرات المطابقة أو المزاوجة Matching؛ وفقرات الاختيار من بديلين مثل (صواب، خطأ)؛ وفقرات الاختيار من متعدد Multiple Choice، والفقرات ذات الإجابة المصوغة، وفقرات التكميل Completion؛ وفقرات الإجابة القصيرة Short Answer؛ والفقرات الإنشائية أو المقالية وتُعدّ فقرات الاختيار من متعدد من أفضل أنواع الفقرات الموضوعية، وأكثرها شيوعًا وانتشارًا في الاختبارات التحصيلية؛ فهي تتميّز بسهولة تصحيحها وموضوعيتها، وتوفر تغطية جيدة للمادة الدراسية، وتمتاز علامة الطالب عليها بدرجة عالية من الثبات، إضافة إلى أنها تحدّد

نتائج التعلم المقصودة بدرجة عالية، على الرغم من أنّ إعدادها يتطلب وقتًا طويلاً، وجهدًا كبيرًا، ومهارة فائقة من قبل واضعيها، كما أنّ فقرات الاختيار من متعدد قادرة على قياس نواتج التعلم في المستويات العقلية العليا من المجال المعرفي بدرجة تفوق فقرات المطابقة، وفقرات الصواب والخطأ، وفقرات التكميل، والإجابة القصيرة (Baghaei & Amrahi, 2011).

تُعدّ نظرية الاستجابة للفقرة تطورًا حديثًا ومهمًا في القياس النفسي والتربوي، وهي كما يشير كل من هامبلتون وجونز (Hambleton & Jones, 1993) نظرية إحصائية عامة تتمحور حول الفرد والأداء على فقرات الاختبار، وقد سعت منذ ظهورها إلى تقديم الحلول والتطبيقات المقنعة للكثير من القضايا التي عجزت النظرية الكلاسيكية في القياس عن إيجاد حلول لها مثل القياس التكيفي، وبنوك الأسئلة، ومعادلة الاختبارات (Emberston & Reise, 2000).

حظيت نظرية الاستجابة للفقرة IRT باهتمام علماء القياس المعاصر والمؤسسات المعنية ببناء الاختبارات والمقاييس؛ وذلك لأنها أسهمت في تقديم الكثير من الحلول للمشكلات الناجمة عن استخدام النظرية الكلاسيكية في تصميم الاختبارات. وجوهر هذه النظرية يقوم على وجود متصل للسمة، بحيث يمكن تقدير احتمال إجابة الفرد عن اي فقرة من فقرات الاختبار إجابة صحيحة إذا عُلم موقعه على هذا المتصل. وتفترض هذه النظرية وجود علاقة بين قدرة المفحوص، واحتمال الإجابة الصحيحة عن فقرة معينة، حيث يمكن تفسير أداء المفحوص في هذا الاختبار، في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات(علام، ٢٠٠١).

وبما أن هذه السمات من الصعب ملاحظتها بصورة مباشرة، لذلك ينبغي تقديرها أو الاستدلال عليها من خلال أداء الأفراد على مجموعة من فقرات الاختبار؛ ولهذا السبب سُميت هذه النظرية بنظرية السمات الكامنة. ويمكن وصف العلاقة بين أداء الفرد في الاختبار ومقدار هذه السمة من خلال دالة رياضية متزايدة تسمى منحنى خصائص الفقرة Item Charecterstic Curve (ICC). ويعتمد هذا المنحنى في طريقة رسمه على ثلاثة أنواع من المعالم؛ هي معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين للفقرة. واختلاف هذه المعالم يعود إلى اختلاف منظور علماء القياس. لذلك تعدّدت النماذج المتعلقة بهذه النظرية، وتباينت في أشكال المنحنيات المميزة تبعًا لاختلاف عدد معالم الفقرات. ولهذا فإنه يندرج تحت نظرية السمات الكامنة مجموعة من النماذج تعرف باسم نماذج السمات الكامنة، والتي تهدف إلى بيان العلاقة بين أداء الفرد في الاختبار، والسمة أو السمات التي تكمن وراء هذا الأداء وتفسره (Baker, 2001; Demars, 2010).

أكّدت الدراسات العربية والأجنبية التي أجريت في مجال الكشف عن أثر عدد البدائل على الخصائص السيكومترية على أهمية هذا الموضوع، حيث إن هذه الدراسات اختلفت عن بعضها من حيث الهدف والمنهجية المتبعة والنظرية المستخدمة، وفيما يلي عرضٌ للدراسات العربية والأجنبية التي تناولت أثر عدد البدائل على الخصائص السيكومترية للاختبار وفقراته، مرتبة بحسب تسلسلها الزمني، ووفق مضامينها أو مشكلاتها:

في دراسة أجراها فرانسيسكو و جوليا و فيسنتا (Francisco, julio, & Vicente, 2001) هدفت إلى تحديد العدد الأمثل من البدائل (ثلاثة بدائل، أربعة بدائل، خمسة بدائل) في اختبار الاختيار من متعدد وفق نظرية الاستجابة للفقرة. ولتحقيق غرض الدراسة قام الباحث ببناء اختبار في اللغة الانجليزية مكوّن من (٢٢١) فقرة تم تطبيقه على أفراد عينة الدراسة المكونة من (٤٥٢) طالبًا من طلاب جامعة العاصمة الإسبانية مدريد، تم اختيارهم بالطريقة العشوائية. حُللت استجابات أفراد عينة الدراسة للكشف عن معالم الفقرات (الصعوبة، والتمييز) ودالة المعلومات للاختبار وتقديرات القدرة للأفراد. أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية في أثر عدد البدائل على معالم الفقرات ودالة المعلومات للاختبار وتقديرات القدرة ولصالح الاختبار ذي الثلاثة بدائل.

تُعدّ قضية عدد بدائل الفقرة من القضايا السيكومترية التي اهتمت بها البحوث السيكومترية لمعرفة الأثر الناتج منها على الخصائص السيكومترية للفقرات وللاختبار. فالمتتبع للدراسات والبحوث التي أجريت للكشف عن أثر عدد البدائل على بعض الخصائص السيكومترية للاختبار يجد أن بعضها ركز على النظرية الكلاسيكية في القياس كدراسة فقوسة (١٩٩٥) التي كشفت نتائجها وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تُعزى لعدد بدائل الفقرة، ولصالح النموذج الذي يحتوي على بديلين لكل من المبحثين (الرياضيات، والتاريخ)، أما فيما يتعلق بمعامل تمييز الفقرة، فأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات معاملات تمييز الفقرة للنماذج الأربعة لاختبار الرياضيات (بديلين؛ ثلاثة بدائل؛ أربعة بدائل؛ خمسة بدائل)، ولصالح النموذج ذي الخمسة بدائل. في حين أظهرت دراسة الغامدي (٢٠٠٨) عدم وجود فروق ذات دلالة إحصائية بين نماذج الاختبارات الثلاثة (ثلاثة بدائل، أربعة بدائل، خمسة بدائل) حيث كانت المتوسطات الحسابية لمؤشرات الصدق والثبات متقاربة.

أما دراسة الشريفين وطعامنة (٢٠٠٩)، فقد هدفت إلى الكشف عن أثر عدد البدائل في اختبار الاختيار المتعدد في تقديرات القدرة للأفراد، ومعالم صعوبة الفقرات في ضوء نظرية الاستجابة للفقرة. وقد أشارت نتائجها إلى وجود فروق ذات دلالة إحصائية بين معاملات

صدق المحك، في حين لم يثبت وجود فروق ذات دلالة إحصائية بين معاملات الثبات تُعزى لمتغير عدد البدائل.

وهدفت دراسة يامان (Yaman, 2011) إلى تحديد العدد الأمثل للبدائل في اختبارات الاختيار من متعدد، وذلك عن طريق مقارنة الخصائص السيكومترية لاختبارات الاختيار من متعدد. وتوصلت نتائج الدراسة إلى أن مستوى ثبات فقرات الاختبار المكوّن من ثلاثة وخمسة بدائل أعلى منها في الفقرات المكونة من أربعة بدائل. وأشارت النتائج أيضاً إلى عدم وجود فروق ذات دلالة إحصائية بين الاختبارات الثلاثة في الخصائص السيكومترية للفقرات الخاصة بمعلمة الصعوبة والتمييز.

يُلاحظ مما سبق، عدم وجود اتفاق بين نتائج الدراسات والأبحاث التي أجريت في مجال الكشف عن أثر عدد البدائل على الخصائص السيكومترية للاختبار وفقراته. وتجدر الإشارة كذلك إلى أنَّ بعض هذه الدراسات اعتمدت على النظرية الكلاسيكية في القياس كدراسة فقوسة (١٩٩٤)؛ ودراسة الغامدي (٢٠٠٨)، واعتمد البعض الآخر كدراسة الشريفين وطعامنة (٢٠٠٩)، ودراسة يامان (Yaman, 2010) على نظرية الاستجابة للفقرة ، والتي تُعدّ ثورة في القياس النفسي والتربوي، حيث ساعدت على تقديم الكثير من الحلول للمشاكل المتعلقة في بناء الاختبارات وتطويرها. هذا ما دفع الباحث إلى إجراء هذه الدراسة لمعرفة أثر عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلين) في اختبار الاختيار من متعدد في الخصائص السيكومترية لاختبار تحصيلي في مادة الرياضيات باستخدام أحد نماذج نظرية الاستجابة للفقرة وهو النموذج ثُلاثي المعلمة.

أكّدت الدراسات العربية والأجنبية التي أجريت في مجال الكشف عن أثر عدد البدائل على الخصائص السيكومتريةعلى أهمية هذا الموضوع، حيث إن هذه الدراسات اختلفت عن بعضها من حيث الهدف والمنهجية المتبعة والنظرية المستخدمة، وفيما يأتي عرض لبعض هذه الدراسات مرتبة بحسب تسلسلها الزمني، ووفق مضامينها أو مشكلاتها:

وأجرى الغامدي (٢٠٠٨) دراسة هدفت لمعرفة أثر عدد البدائل في الخصائص السيكومترية لاختبار تحصيلي في الرياضيات للصف التاسع الاساسي في المملكة الاردنية الهاشمية، تكونت عينة الدراسة من (١٤٠٠) طالب، أشارت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية عند (α=٠,٠٥) بين متوسطات قيم معاملات الصعوبة للنموذجين ذي الخمسة بدائل وذي الثلاثة بدائل ولصالح النموذج ذي الثلاثة بدائل وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات قيم معاملات الصعوبة للنموذجين ذي الأربعة بدائل وذي الثلاثة بدائل و لصالح النموذج ذي الثلاثة بدائل.

أما دراسة الشريفين وطعامنة (٢٠٠٩) فهدفت إلى الكشف عن أثر عدد البدائل في اختبار الاختيار من متعدد في تقديرات معلمة القدرة للمفحوصين، وتقديرات معلمة الصعوبة للفقرات، واقتران المعلومات للفقرات والاختبار. ولتحقيق أهداف الدراسة تم بناء اختبار تحصيلي من نوع الاختيار من متعدد في الرياضيات لطلبة الصف العاشر الأساسي تكّون من (٤٠) فقرة. وبنيت ثلاث صور من الاختبار يختلف بعضها عن بعض في عدد بدائل فقراتها فقط (ثلاثة بدائل، أربعة بدائل، خمسة بدائل) لكل صورة من الصور الثلاث على الترتيب. طبقت صور الاختبار على عينة مكوّنة من (٦٠٠) طالب وطالبة بواقع (٢٠٠) طالب وطالبة لكل صورة، حللّت الاستجابات على كل صورة من صور الاختبار بشكل مستقل، أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقديرات معالم الصعوبة للفقرات، وعدم وجود فروق ذات دلالة إحصائية بين معاملات ثبات الأفراد تُعزى لعدد البدائل، كما أشارت النتائج إلى أن التقديرات أكثر دقة لقدرات أفراد الصورة الأولى للاختبار من قدرات أفراد الصورة الثالثة له، وكانت التقديرات لقدرات أفراد الصورة الثانية أكثر دقة منها للصورة الثالثة.

وأجرى تارنت و وير (Tarrant & Ware, 2010) دراسة هدفت إلى مقارنة الخصائص السيكومترية لصورتين من اختبارات الاختيار من متعدّد، المستخدمة في عملية تقييم طلبة تخصص التمريض، تختلف فقراتها في عدد البدائل (ثلاثة بدائل، أربعة بدائل). ولتحقيق أهداف الدراسة قام الباحثان بتطبيق الصورة ذات الأربعة بدائل على عينة استطلاعية، وذلك لفحص ومقارنة الخصائص السيكومترية لفقرات الاختبار، والكشف عن المموّه الضعيف في عملية الإجابة ليتم حذفه وتشكيل الصورة الثانية ذات الثلاثة بدائل من الاختبار. وتكونت الصورة النهائية من الاختبار من(٤١) فقرة في كل نموذج (ثلاثة بدائل، أربعة بدائل). أشارت نتائج الدراسة إلى أن الصورة الثانية (ثلاثة بدائل) كانت أكثر فاعلية على الرغم من قلة المموهات؛ وذلك بسبب قوة هذه المموهات.

كما أجرى ملياني (٢٠١٠) دراسة هدفت إلى مقارنة أثر اختلاف عدد البدائل في اختبار الاختيار من متعدد على صعوبة الفقرة وقدرة الفرد في ضوء كلٍ من النظرية الكلاسيكية ونموذج راش. تم بناء اختبار تحصيلي من نوع الاختيار من متعدد في الرياضيات لطلبة الصف الثاني الثانوي مكون من(٢٥) فقرة بصورته النهائية من ثلاثة نماذج (ثلاثي البدائل، رباعي البدائل، خماسي البدائل). تكونت عينة الدراسة من (٥٨٢) طالبًا لجميع نماذج الاختبار الثلاثة وفق النظرية الكلاسيكية والنموذج اللوجستي أحادي المعلمة (نموذج

راش). أظهرت النتائج وجود أثر ذي دلالة إحصائية للنموذج على قدرات الأفراد المقدرة وفق النظرية الكلاسيكية ولصالح النموذج ثلاثي البدائل، في حين لم يتبين وجود أثر ذي دلالة إحصائية للنموذج على قدرات الأفراد المقدرة وفق نموذج راش. كما كشفت نتائج الدراسة عن وجود علاقة ارتباطية ذات دلالة إحصائية بين قيم معاملات الصعوبة المقدرة وفق النظرية الكلاسيكية ونموذج راش للنماذج الثلاثة، وأظهرت النتائج عدم وجود أثر ذي دلالة إحصائية للنموذج على قيم معاملات الصعوبة المقدرة وفق النظرية الكلاسيكية، في حين تبين وجود أثر ذي دلالة إحصائية للنماذج على قيم معاملات الصعوبة المقدرة وفق نموذج راش ولصالح النموذج ثلاثي البدائل.

أما دراسة باغاي وأمراهي (Baghaei & Amrahi, 2011) فهدفت إلى الكشف عن أثر اختلاف عددالبدائل في أسئلة الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته، وتحديد العدد الأمثل من البدائل المستخدمة في أسئلة الاختيار من متعدد والتي يكون صدقها وثباتها عاليًا. تكونت عينة الدراسة من(١٨٠) طالبًا في إيران. ولتحقيق أهداف الدراسة قام الباحثان ببناء اختبار من نوع الاختيار من متعدد، تألف من ثلاثة نماذج؛ النموذج الأول لكل فقرة من فقراته خمسة بدائل، والنموذج الثاني لكل فقرة من فقراته أربعة بدائل، أما النموذج الثالث فلكل فقرة من فقراته ثلاثة بدائل، طُبِّقت على أفراد الدراسة بعد التحقق من صدقه وثباته. أظهرت نتائج الدراسة عدم وجود فروق ذات دلالة إحصائية بين أثر النماذج الثلاثة للاختبار(ثلاثة، أربعة، خمسة) بدائل على صعوبة الفقرة وكانت متوسطات معاملات التمييز للنماذج الثلاثة متقاربة، أما فيما يتعلق بتمييز الفقرة، فأظهرت النتائج وجود فروق ذات دلالة إحصائية بين النماذج الثلاثة للاختبار(ثلاثة، وأربعة، وخمسة) بدائل على تمييز الفقرة. وأظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين أثر النماذج الثلاثة للاختبار(ثلاثة، وأربعة، وخمسة) بدائل على ثبات الاختبار حيث كانت متوسطات الثبات للنماذج الثلاثة متقاربة.

وهدفت دراسة يامان (Yaman, 2011) إلى تحديد العدد الأمثل للبدائل في اختبارات الاختيار من متعدد، وذلك من خلال مقارنة الخصائص السيكومترية لاختبارات الاختيار من متعدد في تعليم العلوم والتكنولوجيا. ولتحقيق هدف الدراسة تم بناء (٩) اختبارات من نوع الاختيار من متعدد ذات أعداد مختلفة من البدائل (ثلاثة بدائل، أربعة بدائل، خمسة بدائل) طبقت على أفراد عينة الدراسة البالغ عددهم (٤١) طالبًا وطالبة في جامعة كارالماس (Karaelmas) في تركيا. أشارت نتائج الدراسة إلى أن مستوى ثبات فقرات الاختبار المكوّن من ثلاثة أو خمسة بدائل أعلى منها بالنسبة للاختبار المكوّن من أربعة بدائل. كما أشارت

النتائج أيضاً إلى عدم وجود فروق ذات دلالة إحصائية بين الاختبارات الثلاثة في الخصائص السيكومترية للفقرات الخاصة بمعلمة الصعوبة والتمييز لكل منها.

وأجرى بني عطا والرباعي (٢٠١٣) دراسة هدفت إلى الكشف عن أثر عدد بدائل الفقرة وموقع المموه القوي في فقرات اختبار الاختيار من متعدد على الخصائص السيكومترية للاختبار ومعالم الفقرات وقدرة الأفراد. ولتحقيق الغرض من الدراسة تم بناء اختبار تحصيلي من نوع الاختيار من متعدد في الرياضيات لطلبة الصف العاشر مكوّن من (٤١) فقرة بصورته النهائية. اشتمل الاختبار على أربعة نماذج حسب عدد البدائل وموقع المموه القوي. تكونت عينة الدراسة من (٢١١١) طالبًا وطالبة لجميع نماذج الاختبار الأربعة وفقًا للنموذج اللوجستي ثلاثي المعلمة. كشفت نتائج تحليل التباين الثنائي بأنه لا توجد فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تُعزى لعدد بدائل الفقرة وموقع المموّه القوي، وأظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم التخمين للفقرات تُعزى لمتغيري الدراسة.

في ضوء ما تم عرضناه من دراسات يُلاحَظ أنها تناولت أوجهًا مختلفة حيث استخلص الباحث من خلال نتائجها ما يأتي:

اختلاف الدراسات السابقة في نتائجها وفي البرامج الحاسوبية والنماذج المستخدمة في تحليل فقرات الاختبار وتدريجها؛ إذ إن معظم هذه الدراسات استخدمت النموذج أحادي المعلمة وثُنائي المعلمة كدراسة كل من: ملياني (٢٠١٠)؛ وفرانسيسكو، جوليو، و فيسنتا (Francisco, Julio & Vicente, 2001)؛ تارنت و وير (Tarrant & Ware, 2010)؛ باغاي و أمراهي (Baghaei & Amrahi, 2011)؛ دراسة يامان (Yaman, 2011)، باستثناء دراسة كل من: واغنر و هارفي (Wagner & Harvey, 2003)؛ دراسة بني عطا والرباعي (٢٠١٣).

أما بالنسبة لأوجه الاستفادة من الدراسات السابقة: فتُعدُّ هذه الدراسات مصدراً أساسياً لكثير من المعلومات المهمة التي وجهت الباحث في دراسته الحالية من حيث اختيارها وتحديد مشكلتها ومنهجيتها والإجراءات الملائمة لتحقيق أهدافها، وتكوين تصور شامل عن الأطر النظرية التي ينبغي أن تشملها الدراسة الحالية، ومن المؤمل أن تضيف إلى نتائج الدراسات السابقة لأنها تناولت موضوع عدد البدائل في اختبارات الاختيار من متعدد وأثرها على الخصائص السيكومترية للاختبار وفقراته. تميّزت الدراسة الحالية باستخدامها للنموذج ثُلاثي المعلمة كممثل للنظرية الحديثة، في حين ركزت معظم الدراسات السابقة على استخدام النموذج أحادي المعلمة والنموذج ثنائي المعلمة، وتميّزت هذه الدراسة أيضاً بكشفها عن

الخصائص السيكومترية للاختبار وفقراته (الصعوبة، التمييز، التخمين، الصدق، والثبات، وتقديرات القدرة للأفراد، ودالة المعلومات للفقرة وللاختبار)، ومن هذا المنطلق توجب على الباحث تطوير اختبار يقيس تحصيل طلاب الصف الاول ثانوي في محتوى الفصل الدراسي الأول من مبحث الرياضيات في المملكة الأردنية الهاشمية، تكوّن من ثلاثة نماذج دون تغيير متن الفقرة (أربعة بدائل، ثلاثة بدائل، بديلين) للكشف عن أثر عدد البدائل على الخصائص السيكومترية للاختبار وفقراته في ضوء النموذج ثُلاثي المعلمة.

مشكلة الدراسة

نظرًا لأهمية الاختبارات التحصيلية في قياس تحصيل المتعلمين وتقويمه، وأهمية امتلاك المعلم لمبادئ بناء الاختبارات التحصيلية، كون المعلم هو المسؤول عن متابعة طلبته والإشراف عليهم وتقويم مسارهم، ومشاركته لهم مسؤولية نجاح أو فشل العملية التعليمية التعلّمية من خلال مشاركته في إعداد الاختبارات وتطبيقها وتصحيحها في جميع مراحل التعليم الابتدائية والمتوسطة والثانوية، فقد لاحظ الباحث أن هناك قصورًا في طريقة إعداد الاختبارات من قبل المعلمين بجانب قصورهم في تحليل نتائج هذه الاختبارات والتي هي جزء من مسؤوليتهم. ولوحظ أيضاً أن الكثير من المعلمين يكثرون من استخدام الأسئلة ذات الإجابة المنتقاة في قياسهم لتحصيل الطلبة، الأمر الذي يستدعي التوقف عند هذه المسألة، وإجراء المزيد من البحث في الأثر الذي يتركه عدد بدائل الفقرة في الخصائص السيكومترية للاختبار، ومعالم الفقرات، والقدرة.

وجدير بالذكر أن الأسئلة الموضوعية تتطلب مهارات معينة، وخصوصًا اختبارات الاختيار من متعدد، حيث يستخدم بعض المعلمين فقرات تحتوي على خمسة بدائل للحدّ من أثر التخمين، في حين يميل البعض منهم إلى استخدام فقرات تحتوي على عدد أقل من خمسة بدائل بسبب صعوبة إيجاد بديل خامس ذو درجة مرتفعة من الجاذبية والتجانس مع بقية البدائل. ويتوقف اختيار العدد المناسب من البدائل على الخصائص السيكومترية التي تجعل فقرات الاختبار أفضل ما يمكن من حيث الصعوبة والتمييز، وتجعل من الاختبار أفضل ما يمكن من حيث الصدق والثبات الأمر الذي يترتب عليه ظهور المستوى الحقيقي للطالب، من هنا جاءت هذه الدراسة بغرض الكشف عن أثر عدد البدائل على الخصائص السيكومترية للاختبار، وفقراته، والقدرة، وفقًا للنموذج ثُلاثي المعلمة؛ أحد نماذج نظرية الاستجابة للفقرة IRT. وبشكل أكثر تحديدًا حاولت الدراسة الإجابة عن الأسئلة الآتية:

أسئلة الدراسة

- هل يوجد فروق ذات دلالة إحصائية بين متوسطات قيم معلمة صعوبة الفقرة المُقدّرة باستخدام النموذج اللوجستي ثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلين)؟
- هل يوجد فروق ذات دلالة إحصائية بين متوسطات قيم معلمة التمييز المُقدّرة باستخدام النموذج اللوجستي ثُلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلين)؟
- هل يوجد فروق ذات دلالة إحصائية بين متوسطات قيم معاملات الثبات المُقدّرة باستخدام النموذج اللوجستي ثُلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلين)؟
- هل يوجد فروق ذات دلالة إحصائية بين متوسطات قيم الخطأ المعياري في تقدير معلمة القدرة المُقدّرة باستخدام النموذج اللوجستي ثُلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلين)؟

أهداف الدراسة

هدفت الدراسة الحالية إلى الكشف عن أثر اختلاف عدد البدائل في أسئلة الاختيار من متعدّد على الخصائص السيكومترية للاختبار وفقراته في اختبار تحصيلي في مادة الرياضيات وفقًا للنموذج ثُلاثي المعلمة؛ أحد نماذج نظرية الاستجابة للفقرة. وهدفت الدراسة أيضاً إلى تحديد العدد الأمثل من البدائل المستخدمة في أسئلة الاختيار من متعدد والتي تُحقّق درجة مرتفعة من مؤشرات الصدق والثبات لاختبارات التحصيل.

أهمية الدراسة

تكتسب هذه الدراسة أهميتها من أهمية الموضوع الذي تناولته، حيث تناولت الدراسة موضوع عدد البدائل في اختبار الاختيار من متعدّد وأثرها على الخصائص السيكومترية للاختبار وفقراته وفقًا للنموذج ثُلاثي المعلمة في نظرية الاستجابة للفقرة؛ من هنا يمكن

تلخيص أهمية هذه الدراسة في الجوانب التالية:

- التوصل إلى بناء اختبار يتمتع بخصائص سيكومترية جيدة استنادًا إلى النموذج ثُلاثي المعلمة كواحد من نماذج النظرية الحديثة في القياس IRT في تحليل وتفسير نتائج الاختبارات والمقاييس. كما قد تسهم هذه الدراسة في دعم القاعدة النظرية للبحوث المتعلقة بنظرية الاستجابة للفقرة في انتقائها للفقرات المناسبة وعدد بدائل الفقرة. وربما يُستخدم الاختبار الوارد في هذه الدراسة كاختبار لقياس تحصيل الطلبة في مبحث الرياضيات للكشف عن جوانب القوة والضعف في تحصيلهم؛ ليتم تعزيز جوانب القوة، والعمل على تلافي نقاط الضعف لديهم والعمل على تحسين تحصيلهم.
- مساعدة متخذي القرار وراسمي السياسات التربوية في المملكة الأردنية الهاشمية على تطوير مهارات تدريس وتقويم مبحث الرياضيات، ومقارنة المستوى الوطني بالمستوى الدولي، وتحسين العملية التعليمية التعلّمية من حيث أهدافها ووسائلها وتقنياتها من خلال عقد الدورات المتخصصة في مجال بناء الاختبارات التحصيلية، وتحليل وتفسير نتائجها، وتطوير النظام التربوي والارتقاء بنوعية مخرجاته.
- قد تسهم نتائج هذه الدراسة في تشجيع الباحثين على القيام بالمزيد من الدراسات المشابهة وعلى نطاق أوسع وفي مواد دراسية أخرى.

مصطلحات الدراسة

اختبار الاختيار من متعدّد: هو "اختبار يتكون من عدد من الفقرات، الفقرة منه تتكون من جزأين، الجزء الأول يسمى المتن وقد يكون على شكل سؤال أو عبارة تتضمن قضية معينة تحتاج إلى تكميل إجابة، ويسمى الجزء الثاني بالبدائل ووهي عهبارة عن حلول أو إجابات محتملة للسؤال أو القضية الواردة في المتن، وتتطلب إجابة صحيحة أو أفضل الإجابات، أو إجابة مركبة، أو تفسيرًا لبيانات ومعلومات معطاة" (المعتوق، ٢٠٠٦: ١٠٣). أما التعريف الإجرائي لاختبار الاختيار من متعدّد في هذه الدراسة، فهو الاختبار الذي أعدّه الباحث في محتوى الفصل الدراسي الأول من كتاب الرياضيات المقرّر تدريسه لطلاب الصف الاول ثانوي في المدارس الحكومية في المملكة الأردنية الهاشمية لقياس درجة اكتسابهم للمفاهيم الرياضية، والذي يتألف من ثلاثة نماذج دون تغيير في متون فقراتها؛ النموذج الأول لكل فقرة من فقراته أربعة بدائل، والنموذج الثاني لكل فقرة من فقراته ثلاثة بدائل، أما النموذج الثالث فلكل فقرة من فقراته بديلين اثنين.

الخصائص السيكومترية: ويقصد بها مؤشرات الصعوبة والتمييز للفقرات، ومعاملات الثبات والصدق للاختبار؛ وفيما يلي تعريف كل منها:

معلمة الصعوبة: "مؤشر إحصائي يُقَّدر بنسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن فقرة معينة من فقرات الاختبار، ويرمز له بالرمز Pi" (علام، ٢٠٠٢: ٦٧٩).

معلمة التمييز: "مؤشر إحصائي يبين الزيادة في نسبة الطلبة الذين يجيبون إجابة صحيحة عن الفقرة بعد عملية التدريس" (عودة، ٢٠١٠: ٢٩٢).

الثبات: "الحصول على نفس النتائج تقريبًا عندما يتم إعادة الاختبار نفسه على المجموعة نفسها وتحت نفس الظروف" (عبابنة، ٢٠٠٩ : ١١١). أما التعريف الإجرائي للثبات في سياق الدراسة الحالية فهو قيمة معامل الثبات الإمبريقي المحسوب وفقًا للنموذج ثُلاثي المعلمة.

معلمة القدرة: وهي الخصائص المتعلقة بالتوزيع التكراري للقدرة من حيث شكل التوزيع ومتوسطه الحسابي وانحرافه المعياري والتوائه وتفلطحه (Hambleton & Jones, 1994). ويُقصد بمعلمة القدرة في سياق الدراسة الحالية بأنها الخصائص المتعلقة بالتوزيع التكراري لقدرة طلاب الصف الأول الثانوي في مدارس المملكة الأردنية الهاشمية من حيث شكل التوزيع ومتوسطه الحسابي وانحرافه المعياري والتوائه وتفلطحه.

النموذج ثُلاثي المعلمة: هو أحد نماذج النظرية الحديثة في القياس الذي يفترض أن الفقرات تختلف في صعوبتها وتمييزها وميل المفحوصين للتخمين في المستويات المختلفة للقدرة لدى محاولتهم الإجابة عن فقرة من نوع الاختيار من متعدّد. ويُرمز لمعامل التمييز بالرمز (a)، ومعامل الصعوبة بالرمز (b)، ومعامل التخمين بالرمز (c) في المعادلة الرياضية التي تربط بين احتمال إجابة المفحوص بمستوى معيّن من القدرة إجابة صحيحة عن فقرة ما من جهة والمعالم الثلاثة للفقرة من جهة أخرى.

حدود الدراسة ومحدداتها

يتحدّد تعميم نتائج الدراسة الحالية بخصائص مجتمع الدراسة والمجتمعات المماثلة، حيث تمّ جمع بيانات الدراسة من خلال تطبيق اختبار تحصيل في مبحث الرياضيات من إعداد الباحث على عيّنة عشوائية عنقودية تمّ اختيارها من مجتمع طلاب الصف الاول ثانوي المسجّلين في المدارس التابعة لمديرية التعليم المملكة الأردنية الهاشمية خلال الفصل الدراسي الأول من العام الدراسي(٢٠١٥/٢٠١٦)

كما يتحدّد تعميم نتائج هذه الدراسة بالخصائص السيكومترية لاختبار التحصيل المُستخدم في هذه الدراسة، والذي يتألف من ثلاثة نماذج من نوع الاختيار من متعدّد (أربعة

بدائل؛ ثلاثة بدائل؛ بديلين) مع المحافظة على متون الفقرات في النماذج الثلاثة. كما يتحدّد تعميم نتائج الدراسة بالنموذج ثلاثي المعلمة المُستخدم في تحليل البيانات واستخراج مؤشرات حول معالم الفقرات والاختبار وتقدير قدرة الأفراد.

الطريقة والإجراءات

مجتمع الدراسة

تكوّن مجتمع الدراسة المُستهدف من جميع طلبة الصف الأول ثانوي المسجّلين في مدارس الذكور الحكومية في المملكة الاردنية الهاشمية خلال الفصل الأول من العام الدراسي (٢٠١٥/٢٠١٦). أما مجتمع الدراسة الممكن، فقد تكوّن من جميع طلبة الصف الأول ثانوي المسجّلين في مدارس الذكور الحكومية التابعة لمديرية التربية والتعليم لواء الجامعة في الأردنية الهاشمية خلال الفصل الدراسي الأول من العام الدراسي (٢٠١٥/٢٠١٦) البالغ عددهم (١٢٠٠٠) طالب، وذلك حسب إحصائيات وزارة التربية والتعليم لعام (٢٠١٥/٢٠١٦).

عينة الدراسة

استُخدم أسلوب المعاينة العشوائية العنقودية لاختيار أفراد عيّنة الدراسة، حيث تمّ اختيار (٢٥) مدرسة للذكور من مدارس مجتمع الدراسة المتوفر بطريقة عشوائية، ثمّ تمّ اختيار (٦١) شعبة من شعب الصف الأول الثانوي من تلك المدارس بطريقة عشوائية أيضاً، وتمّ اعتبار الطلبة في الشعب التي تمّ اختيارها هم أفراد العيّنة. وقد بلغ حجم العيّنة (١٥٣٠) طالبًا يمثّلون ما نسبته (١٢,٧٥٪) من حجم مجتمع الدراسة.

أداة الدراسة

بُنيت ثلاثة نماذج لاختبار تحصيلي في محتوى الجزء الأول من مبحث الرياضيات للصف الأول الثانوي من نوع الاختيار من متعدّد تختلف في عدد البدائل فقط، حيث كان لفقرات النموذج الأول أربعة بدائل للإجابة، ولفقرات النموذج الثاني ثلاثة بدائل للإجابة، ولفقرات النموذج الثالث بديلان اثنان للإجابة. وفق خطوات بناء اختبارات التحصيل التي أوردها غرونلند ولين (Gronlnud & Linn, 1990) إذ مرّت عملية بناء الاختبار بالخطوات الآتية:

١- تحديد الهدف من الاختبار وهو قياس تحصيل طلبة الصف الأول ثانوي في مبحث الرياضيات للفصل للعام الدراسي (٢٠١٥/٢٠١٦) في المدارس الحكومية في المملكة الاردنية الهاشمية.

٢- تحليل محتوى الجزء الأول من كتاب الرياضيات للصف الأول ثانوي في المملكة الاردنية الهاشمية، وأمكن اشتقاق (٣٢) هدفًا تعليمياً (ملحق ٢)، وبعد ذلك تمّ تصميم جدول مواصفات الاختبار، ثم كتابة فقرات الاختبار لتغطي محتوى الجزء الأول من الكتاب المقرّر. وبلغ عدد فقرات الاختبار بصورته الأولية (٧٠) فقرة إذ تم كتابة فقرتين اثنتين أو ثلاث فقرات لقياس كل هدف من الأهداف.

٣- عرض الاختبار بصورته الأولية على مجموعة من المحكمين من أصحاب الاختصاص والخبرة في الجامعات الأردنية للتأكد من صدق المحتوى للاختبار. وفي ضوء الاقتراحات التي قدمها المحكّمون أجريت التعديلات على صياغة بعض الفقرات، وحذفت ست فقرات ليصبح عدد فقرات الاختبار بصورته المعدّلة (٦٤) فقرة بحيث تمّ الاحتفاظ بفقرتين اثنتين لقياس كل هدف من الأهداف.

٤- تطبيق الصورة المعدّلة من الاختبار على عينة استطلاعية تكوّنت من (٨٠) طالبًا من خارج عينة الدراسة، وجرى تحديد موعد الاختبار والمادة التي يشملها، وبعد الانتهاء من تطبيق الصورة الأولية تمّ تحليل الفقرات من خلال حساب معاملات الصعوبة والتمييز (بوينت بايسيريال) لكل فقرة من فقرات الاختبار، وحساب فاعلية المموهات لكل فقرة. وفي ضوء نتائج تحليل الفقرات تمّت المفاضلة بين الفقرات التي تشترك بقياس الهدف نفسه، وتم اختيار الفقرة التي تتمتع بأفضل معاملي صعوبة وتمييز، وبهذا تمّ الاحتفاظ بـ (٣٢) فقرة في الاختبار لتشكّل الصورة النهائية من الاختبار إذ بلغ متوسط صعوبة فقرات الاختبار بصورته النهائية (٠,٥١)، ومتوسط معاملات تمييزها (٠,٤٩).

شكلت ثلاثة نماذج من الاختبار (أربعة بدائل؛ ثلاثة بدائل؛ بديلين) حيث تمّ حذف المموه الأضعف لتشكيل النموذج ذو الثلاثة بدائل، وتمّ حذف أضعف مموهين لتشكيل النموذج ذو البديلين.

٧- تمّ تطبيق الصورة النهائية من الاختبار بنماذجه الثلاثة على أفراد عيّنة الدراسة موزّعين على النماذج الثلاثة للاختبار بواقع (٥٠٠) طالب لكل نموذج.

٨- إعداد مفتاح الإجابة لكل نموذج من النماذج الثلاثة للاختبار.

٩- تصحيح كراسات الاختبار يدويًا باستخدام مفتاح التصحيح المعدّ مسبقًا لكل نموذج.

١٠- مراجعة استجابات الطلبة يدويًا بهدف استبعاد الطلبة الذين أخفقوا في الإجابة عن جميع فقرات الاختبار لأن مستوى صعوبة الاختبار يُعدّ أعلى من مستوى قدرتهم، واستبعاد الطلبة الذين أجابوا إجابة صحيحة عن جميع الفقرات لأن مستوى صعوبة الاختبار يُعدّ أدنى من مستوى قدراتهم، وبلغ عدد الطلبة الذين تمّ استبعادهم (٣٠) طالبًا. وتمت

مراجعة الفقرات للكشف عن الفقرات التي أجاب عنها جميع الطلبة إجابة صحيحة، أو الفقرات التي أجاب عنها جميع الطلبة إجابة خاطئة ولم يتبيّن وجود أي من هذه الفقرات.

١١- تحليل البيانات باستخدام البرامج الإحصائية الملائمة.

نموذج نظرية الاستجابة للفقرة المستخدم

استخدام النموذج اللوجستي ثلاثي المعلمة، لكونه أكثر ملاءمة لفقرات الاختيار من متعدد، وتم تحليل البيانات واستخراج معالم الفقرات من خلال برنامج (Bilog-MG).

المعالجات الإحصائية

تم التحقق من افتراض أحادية البعد للبيانات من خلال:

- إجراءات التحليل العاملي Factor Analysis حيث تمّ إخضاع استجابات أفراد الدراسة عن فقرات كل نموذج من نماذج الاختبار الثلاثة لأسلوب تحليل المكوّنات الأساسية Principle Component Analysis. وتم استخراج قيم الجذر الكامن ونسبة التباين المفسر لكل عامل من العوامل المستخلصة من التحليل، ثم تمّ تدوير المحاور باستخدام طريقة التدوير المتعامد للعوامل المستخلصة من تحليل المكوّنات الأساسية والتي زادت قيم جذورها الكامنة عن الواحد الصحيح.
- معامل كرونباخ الفا، والذي تُعدّ القيمة المرتفعة له مؤشرًا على تحقّق أحادية البعد.
- الخطأ المعياري للتقدير باستخدام برنامج (Bilog-MG).
- تمّ استخدام تحليل التباين الأحادي للكشف عن دلالة الفروق بين متوسطات أداء الأفراد على النماذج الثلاثة للاختبار باختلاف النموذج المستخدم.

كما تم التحقق من الافتراضات التي يتطلبها النموذج اللوجستي الثلاثي المّعلمة في نظرية الاستجابة للفقرة، وكذلك التحقق من مطابقة الأفراد والفقرات لهذا النموذج، وفيما يلي عرض لهذه الإجراءات.

التحقق من افتراض أحادية البعد

تمّ التحقق من توافر افتراض أحادية البعد كأحد افتراضات نظرية الاستجابة للفقرة باستخدام إجرائين هما: معامل الاتساق الداخلي بدلالة إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا، وإجراء التحليل العاملي باستخدام طريقة المكونات الأساسية، وفيما يلي عرضٌ للنتائج التي تم التوصل إليها:

معامل ثبات الاتساق الداخلي بدلالة إحصائيات الفقرة:جرى استخراج معامل الثبات الاتساق الداخلي بدلالة إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا لكل نموذج من النماذج الثلاثة للاختبار كما هو موضّح في الجدول (٢).

جدول (٢)

معاملات ثبات الاتساق الداخلي لنماذج الاختبار باستخدام معادلة كرونباخ ألفا

النموذج	عدد الفقرات	عدد البدائل	معامل الثبات
الأول	٣٢	أربعة بدائل	٠,٨٩
الثاني	٣٢	ثلاثة بدائل	٠,٨٦
الثالث	٣٢	بديلين	٠,٨٣

يُلاحظ من الجدول (٢) أنّ قيم معاملات الثبات الاتساق الداخلي بدلالة إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا تراوحت بين (٠,٨٢ و ٠,٨٩)، وهي معاملات مقبولة لأغراض التحقّق من افتراض أحادية البُعد استنادًا إلى إشارة دينيك وتفكول (Dennick & Tavakol, 2011) التي تنص على أن قيم معامل الثبات التي تزيد عن (٠,٧٠) تُعدّ مقبولة لأغراض تحقق افتراض أحادية البعد في الأداء على الاختبار، وأن قيم معامل الثبات التي تقل عن ذلك تُشير إلى ارتباطات ضعيفة بين فقرات الاختبارممّا يدل على عدم تجانسها.

التحقّق من أحادية البعد من خلال إجراء التحليل العاملي: للتحقق من توافر أحادية البعد في الأداء على نماذج الاختبار التي تم تطويرها، تمّ استخدام الرزمة الإحصائية للعلوم الاجتماعية (SPSS)؛ حيث تمّ إخضاع استجابات أفراد عيّنة الدراسة لأسلوب تحليل المكوّنات الأساسية، إذ تم استخراج قيم الجذور الكامنة للعوامل المتحققة عليها، إضافة إلى نسب التباين المُفسّر لكل عامل من العوامل، ثمّ جرى تدوير المحاور باستخدام طريقة التدوير المتعامد للعوامل التي تمّ استخلاصها نتيجة تحليل المكوّنات الأساسية، والتي زادت قيم جذورها الكامنة عن الواحد الصحيح.وتبين الجداول (٣-٦) نتائج التحليل العاملي لكل نموذج من النماذج الثلاثة للاختبار.

جدول (٣)

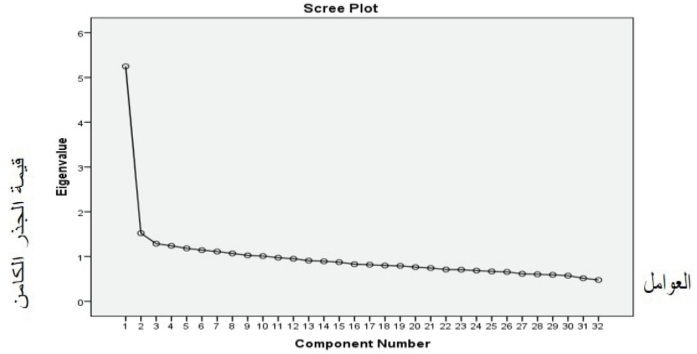
نتائج التحليل العاملي لاستجابات أفراد عينة الدراسة لفقرات النموذج ذي الأربعة بدائل وتدوير المحاور تدويرًا متعامدًا

العامل	الجذر الكامن	نسبة التباين المفسّر٪	نسبة التباين التراكمي المفسّر٪
١	٥,٢٥١	١٥,٠٠٠	١٥,٠٠٠
٢	١,٥٣١	٤,٣٧٧	١٩,٣٧٧

تابع جدول (٢)

العامل	الجذر الكامن	نسبة التباين المفسّر٪	نسبة التباين التراكمي المفسّر٪
٣	١,٣٣٢	٣,٨٠٥	٢٣,١٨٢
٤	١,٢٥٣	٣,٥٨١	٢٦,٧٦٣
٥	١,٢٣٩	٣,٥٤٠	٣٠,٣٠٣
٦	١,١٨٠	٣,٣٧٢	٣٣,٦٧٥

يُلاحظ من الجدول (٣) وجود ستة عوامل تزيد جذورها الكامنة عن الواحد الصحيح؛ إذ إن العامل الأول فسّر ما نسبته (٠٠٠١٥.٪) من تباين الأداء على نموذج الاختبار ذي البدائل الأربعة، وهذه النسبة تشكّل ما مقداره (٤٤,٥٤٣٪) من التباين الكلي الذي تفسّره العوامل الستة مجتمعة. كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن الثاني للعامل الثاني بلغت (٣,٤٢٧) ضعفًا، وهي نسبة كبيرة تزيد عن الضعفين الأمر الذي يُعّد مؤشرًا على تحقق سمة أحادية البعد في الأداء على هذا الاختبار. ويتأكد هذا من خلال التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لهذا النموذج من الاختبار كما هو موضّح في الشكل (١).



شكل (١)

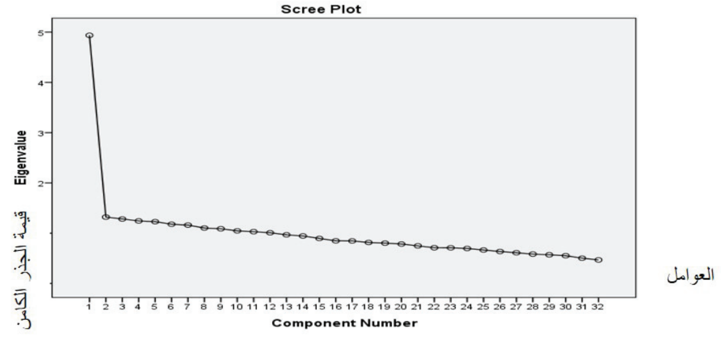
التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لنموذج الاختبار ذو الأربعة بدائل

يُلاحظ من الشكل (١) أن الفرق بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالفرق بين العامل الثاني والعامل الثالث، ثم بقاء الفروق بعد ذلك متقاربة بين العوامل الباقية، ممّا يُشير إلى تحقّق سمة أحادية البعد في تفسير أداء الطلبة على هذا الاختبار، وأنه يقيس سمة واحدة فقط.

جدول (٤)
نتائج التحليل العاملي لاستجابات أفراد عينة الدراسة
لفقرات النموذج ذي الثلاثة بدائل وتدوير المحاور تدويرًا متعامدًا

العامل	الجذر الكامن	نسبة التباين المفسّر٪	نسبة التباين التراكمي المفسّر٪
١	٤,٩٣٦	١٥,٤٢٥	١٥,٤٢٥
٢	١,٣٢١	٤,١٢٩	١٩,٥٥٤
٣	١,٢٨٣	٤,٠١٠	٢٣,٥٦٤
٤	١,٢٤٤	٣,٨٨٨	٢٧,٤٥٢
٥	١,٢٣٠	٣,٨٤٣	٣١,٢٩٤
٦	١,١٧٩	٣,٦٨٤	٣٤,٩٧٩
٧	١,١٦١	٣,٦٣٠	٣٨,٦٠٨
٨	١,١٠٣	٣,٤٤٨	٤٢,٠٥٦

يُلاحظ من الجدول (٤) وجود ثمانية عوامل تزيد قيم جذورها الكامنة عن الواحد الصحيح، وأن العامل الأول فسر ما نسبته (١٥,٤٢٥٪) من تباين الأداء على نموذج الاختبار ذي البدائل الثلاثة، وهذه النسبة تُشكل ما مقداره (٣٦,٦٨٪) من التباين الكلي الذي تفسّره جميع العوامل الثمانية مجتمعة. كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (٣,٧٣٦) ضعفًا، وهي تزيد عن الضعفين، وهذا يُعتبر مؤشرًا على تحقق افتراض أحادية البعد للبيانات في هذا الاختبار. ويتأكد هذا من خلال التمثيل البياني لقيم الجذورالكامنة للعوامل الناتجة عن التحليل العاملي لهذا النموذج من الاختبار كما هو موضّح في الشكل (٢).



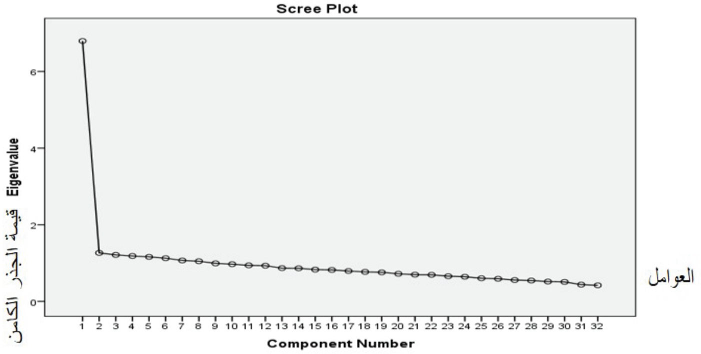
شكل (٢)
التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لنموذج الاختبار ذو الثلاثة بدائل

يُلاحظ من الشكل (٢) أن مقدار الفرق بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالفرق بين العاملالثاني والعامل الثالث، ثم بقاء الفروق بعد ذلك متقاربة جدًا بين العوامل المتبقية، الأمر الذي يُشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على نموذج الاختبار ذي البدائل الثلاثة وأنه يقيس سمة واحدة فقط.

جدول (٥)
نتائج التحليل العاملي لاستجابات أفراد عينة الدراسة لفقرات النموذج ذي البديلين وتدوير المحاور تدويرًا متعامدًا

العامل	الجذر الكامن	نسبة التباين المفسّر ٪	نسبة التباين التراكمي المفسّر ٪
١	٦,٧٩٧	٢١,٢٤٠	٢١,٢٤٠
٢	١,٢٦٥	٣,٩٥٢	٢٥,١٩٢
٣	١,٢١٣	٣,٧٩١	٢٨,٩٨٣
٤	١,١٨٥	٣,٧٠٢	٣٢,٦٨٥
٥	١,١٦٢	٣,٦٣١	٣٦,٣١٦
٦	١,١٢٧	٣,٥٢٢	٣٩,٨٣٨
٧	١,٠٦٨	٣,٣٣٩	٤٣,٤٥٣
٨	١,٠٤٨	٣,٢٧٦	٤٦,٤٥٣

يُلاحظ من الجدول (٥) وجود ثمانية عوامل تزيد قيم جذورها الكامنة عن الواحد الصحيح، وأن العامل الأول فسّر ما نسبته (٢١,٢٤٠٪) من تباين الأداء على هذا النموذج من الاختبار، وهذه النسبة تشكّل ما مقداره (٤٥,٧٢٣٪) من التباين الكلّي الذي تفسّره العوامل الثمانية مجتمعة، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (٥,٣٧٣) ضِعفًا، وهي تزيد عن ضعفين، وهذا يُعتبر مؤشرًا على تحقق أحادية البعد للبيانات في هذا النموذج من الاختبار، ويتأكد هذا من خلال التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لهذا النموذج من الاختبار كما هو موضّح في الشكل (٣).



شكل (٣)
التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لنموذج الاختبار ذو البديلين

يُلاحظ من الشكل (٣) أن مقدار الفرق بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالفرق بين العامل الثاني والعامل الثالث، ثم بقاء الفروق بعد ذلك متقاربة بين العوامل الباقية، ممّا يشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على نموذج الاختبار ذي البديلين.

ونظرًا لتحقق افتراض أحادية البعد في البيانات المتحققة عن نماذج الاختبار الثلاثة، فهذا يعني تحقق افتراض الاستقلال الموضعي Local Independence الذي يُعدّ مؤشرًا على أحادية البعد إذا كان النموذج المستخدم يحدد قيمة تقديرية لقدرة الفرد على بعد أحادي (Hamblton & Swaminathan, 1985).

مطابقة الأفراد والفقرات للنموذج اللوجستي ثلاثي المعلمة:

تم استخدام برنامج (BILOG-MG3) ومن خلال مؤشر مربع كاي ($\chi 2$) كما يبين ملحق (٨) عند مستوى الدلالة (α = ٠,٠١) حيث أظهرت النتائج مطابقة الأفراد في كل نموذج، أما بالنسبة لمطابقة الفقرات في كل اختبار للنموذج اللوجستي الثلاثي المعلمة بناءً على مؤشر مربع كاي ($\chi 2$) عند مستوى الدلالة (α= ٠,٠١) حيث تُعد الفقرة غير مطابقة للنموذج اللوجستي الثلاثي المعلمة إذا كانت الاحتمالية لهذه الفقرة أقل من (٠,٠١)، والجدول (٦) يُبين ذلك:

الجدول (٦)
نماذج الاختبار والفقرات غير المطابقة للنموذج اللوجستي الثلاثي المعلمة

العدد الأصلي للفقرات	نموذج الاختبار	الفقرات غير المطابقة	عدد فقرات الاختبار بعد حذف الفقرات غير المطابقة
٣٢	أربعة بدائل	–	٣٢
٣٢	ثلاثة بدائل	٦	٢٦
٣٢	بديلين	٤	٢٨

انتفاء أثر السرعة Speedness:

تمّ التحقق من أن الاختبار ليس اختبار سرعة من خلال فحص نسبة الأفراد الذين أكملوا الاختبار، وكذلك فحص الفقرات التي لم يجب عنها أي طالب. وفي هذا السياق يُشير كل من هامبلتون و رجورز (Hambleton & Rogers, 1991) أنه إذا كان (٧٥٪) من الأفراد قد أكملوا الإجابة عن فقرات الاختبار، وإذا كان (٨٠٪) من فقرات الاختبار قد تمت الإجابة عنها من قبل الأفراد، فإن السرعة لا تعتبر في هذه الحالة عاملاً مهمًا في الأداء على الاختبار. وحيث إن نسبة الأفراد الذين أكملوا هذا الاختبار هي (١٠٠٪)، وإن جميع الأفراد أكملوا الإجابة عن جميع الفقرات، فإن هذا يعني انتفاء أثر السرعة في هذا الاختبار.

نتائج الدراسة

نتائج الإجابة عن سؤال الدراسة الأول: "هل توجد فروق ذات دلالة إحصائية بين متوسطات قيم معلمة صعوبة الفقرة المقدّرة باستخدام النموذج اللوجستي ثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلين)"؟

للإجابة عن هذا السؤال تم استخراج المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة الصعوبة بحسب نموذج الاختبار المستخدم كما هو موضّح في الجدول (٧).

جدول (٧)
المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة صعوبة الفقرة موزّعة تبعًا لنموذج الاختبار

نموذج الاختبار	المتوسط الحسابي	الخطأ المعياري
أربعة بدائل	-٠,١٣٢	١,٣١٠
ثلاثة بدائل	-١,٧٣٠	٤,٧٣٠
بديلان	٠,٣٣٠	٢,١٠٠

يُلاحظ من الجدول (٧) وجود فروق ظاهرية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة باختلاف نموذج الاختبار، إذ بلغ المتوسط الحسابي لنموذج الاختبار ذي الأربعة بدائل (٠,١٣٢-)، وللنموذج ذو الثلاثة بدائل (١,٧٣٠-)، وفي حالة النموذج ذي البديلين بلغ المتوسط الحسابي (٠,٣٣٠). وللكشف عن دلالة الفروق بين هذه المتوسطات تمّ استخدام أسلوب تحليل التباين الأحادي كما يتّضح من الجدول (٨).

جدول (٨)
نتائج تحليل التباين الأحادي لدلالة الفروق بين المتوسطات الحسابية لتقديرات معلمة صعوبة الفقرة تبعًا لنموذج الاختبار

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة
نموذج الاختبار	٦٢,٧٠١	٢	٣١,٣٥٠	٣,٥٦٤	٠,٠٤١
الخطأ	٧٣٠,١٦١	٨٣	٨,٧٩٦		

يُلاحظ من الجدول (٨) وجود فروق ذات دلالة إحصائية عند مستوى دلالة يقل عن (٠,٠٥) بين المتوسطات الحسابية لتقديرات معلمة صعوبة الفقرة يُعزى لمتغير نموذج الاختبار (أربعة بدائل؛ ثلاثة بدائل؛ بديلين). وللوقوف على مصدر الفروق بين المتوسطات الحسابية الثلاثة تمّ استخدام اختبار "شيفي" للمقارنات البعدية كما هو موضّح في الجدول (٩).

جدول (٩)
نتائج اختبار "شيفيه" للمقارنات البعدية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة تبعًا لنموذج الاختبار

نموذج الاختبار	المتوسط الحسابي	أربعة بدائل	ثلاثة بدائل	بديلين
أربعة بدائل	٠,١٣٣-	-	١,٥٩٥	٠,٤٦٢-
ثلاثة بدائل	١,٧٢٩-		-	٢,٠٥٧-*
بديلان	٠,٣٢٩			-

* $\alpha \leq$ ٠,٠٥

يُلاحظ من الجدول (٩) أن الفرق بين المتوسط الحسابي لتقدير معلمة الصعوبة في نموذج الاختبار ذي الثلاثة بدائل يختلف اختلافًا جوهريًا عنه بالنسبة لنموذج الاختبار ذي البديلين، وهو الذي أحدث الفرق الإجمالي. وهذا يعني ان الاختبار ذا الثلاثة بدائل هو الأكثر صعوبة.

نتائج الإجابة عن سؤال الدراسة الثاني: "هل توجد فروق ذات دلالة إحصائية بين متوسطات قيم معلمة تمييز الفقرة المقدّرة باستخدام النموذج اللوجستي ثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل، ثلاثة بدائل، بديلان)؟

للإجابة عن هذا السؤال تمّ استخراج المتوسطات الحسابية والانحرافات المعيارية لقيم تقديرات معلمة تمييز الفقرة حسب نموذج الاختبار كما هو موضّح في الجدول (١٠).

جدول (١٠)
المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة تمييز الفقرة تبعًا لنموذج الاختبار

نموذج الاختبار	المتوسط الحسابي	الانحراف المعياري
أربعة بدائل	١,٤٢٠	٠,٣٤٨
ثلاثة بدائل	٠,١٤١	٠,١٨٠
بديلان	٠,٥٦٦	٠,٥٩٣

يُلاحظ من الجدول (١٠) وجود فروق ظاهرية بين المتوسطات الحسابية لتقديرات معلمة تمييز الفقرة تبعًا لنموذج الاختبار (أربعة بدائل؛ ثلاثة بدائل؛ بديلان)؛ ففي حالة النموذج ذو الأربعة بدائل بلغ المتوسط الحسابي لمعلمة التمييز (١,٤٢٠) يليه النموذج ذو البديلين بمتوسط حسابي مقداره (٠,٥٦٦)، ثم النموذج ذو الثلاثة بدائل بمتوسط حسابي مقداره (٠,١٤١). ولمعرفة دلالة الفروق بين هذه المتوسطات، تمّ استخدام تحليل التباين الأحادي كما هو موضّح في الجدول (١١).

جدول (١١)
نتائج تحليل التباين الأحادي لدلالة الفروق بين المتوسطات الحسابية لتقديرات معلمة تمييز الفقرة تبعًا لنموذج الاختبار

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "F"	مستوى الدلالة
نموذج الاختبار	٢٤,٩٨٥	٢	١٢,٤٩٢	٧٣,٧٤٠	٠,٠٠١
الخطأ	١٤,٠٦١	٨٣	٠,١٦٩		

يُلاحظ من الجدول (١١) وجود فروق ذات دلالة إحصائية عند مستوى دلالة يقل عن (٠,٠٥) بين المتوسطات الحسابية لتقديرات معلمة تمييز الفقرة يُعزى لمتغير نموذج الاختبار (أربعة بدائل؛ ثلاثة بدائل؛ بديلان)، إذ بلغت قيمة "ف" المحسوبة (٧٣,٧٤٠) وبمستوى دلالة مقداره (٠,٠٠١). وللوقوف على مصدر الفروق بين المتوسطات الحسابية تم استخدام اختبار "شيفي" للمقارنات البعدية والجدول (١٢) يبين ذلك.

جدول (١٢)
نتائج اختبار "شيفيه" للمقارنات البعدية بين المتوسطات الحسابية لتقديرات معلمة تمييز الفقرة تبعًا لنموذج الاختبار

نموذج الاختبار	المتوسط الحسابي	أربعة بدائل	ثلاثة بدائل	بديلين
أربعة بدائل	-٠,١٣٣	-	١,٢٧٩	٠,٨٥٤
ثلاثة بدائل	-١,٧٢٩		-	-٠,٤٢٥*
بديلان	٠,٣٢٩			-

* $\alpha \leq$ ٠,٠٥

يُلاحظ من الجدول (١٢) أن الفروق الإجمالية بين المتوسطات الحسابية لتقديرات معلمة التمييز تُعزى إلى اختلاف متوسط التقديرات للنموذج ذي الثلاثة بدائل مقارنة بمتوسط تقديرات معلمة التمييز للنموذج ذي البديلين. وهذا يعني أن النموذج ذي الثلاثة بدائل هو الأكثر تمييزًا يليه النموذج ذو الأربعة بدائل، ثم النموذج ذو البديلين.

نتائج الإجابة عن سؤال الدراسة الثالث: "هل توجد فروق ذات دلالة إحصائية بين قيم معاملات الثبات المقدّرة باستخدام النموذج اللوجستس ثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلان)؟"

للإجابة عن هذا السؤال تم إيجاد معاملات الثبات باستخدام برنامج (Bilog-MG) لكل نموذج، والجدول (١٣) يبين ذلك.

الجدول (١٣)
معاملات الثبات لكل نموذج من نماذج الاختبار

نموذج الاختبار	معامل الثبات
أربعة بدائل	٠,٨٥
ثلاثة بدائل	٠,٤٥
بديلان	٠,٦٦

يُلاحظ من الجدول(١٣) وجود فروق ظاهرية بين معاملات الثبات المستخرجة لنماذج الاختبار المختلفة، حيث تبيّن أن نموذج الاختبار ذي الأربعة بدائل هو الأكثر ثباتًا، يليه نموذج الاختبار ذي البديلين، ثم نموذج الاختبار ذي الثلاثة بدائل. ولاختبار دلالة الفروق بين معاملات الثبات تم استخدام معادلة "فيشر" لتحويل معاملات الثبات إلى قيم زائيّة وفحص دلالتها باستخدام معادلة "فيشر" (Fisher r-to-z transformation) كما هو موضّح في الجدول (١٤).

جدول (١٤)

معاملات الثبات لكل نموذج من نماذج الاختبار وقيم "ز" المناظرة لها ومستوى دلالتها

نموذج الاختبار	معامل الثبات	المقارنة	قيمة "ز"	مستوى الدلالة
أربعة بدائل	٠,٨٦	٣×٤	١٢,١٦١	٠,٠٠٤
ثلاثة بدائل	٠,٤٧	٢×٤	٧,٣٠٤	٠,٠٣٨
بديلان	٠,٦٨	٣×٢	-٤,٨٥٧	٠,٠٢٣

يُلاحظ من الجدول (١٤) أن الفروق بين معاملات ثبات كل زوج من أزواج النماذج الثلاثة للاختبار (أربعة بدائل؛ ثلاثة بدائل؛ بديلين) هي فروق ذات دلالة إحصائية بمستوى يقل عن (٠,٠٥). ويُلاحظ أيضاً أن نموذج الاختبار ذي الأربعة بدائل هو الأكثر ثباتًا مقارنةً بالنموذجين الآخرين، يليه النموذج ذو البديلان، وأخيرًا النموذج ذو الثلاثة بدائل الأقل ثباتًا.

نتائج الإجابة عن سؤال الدراسة الرابع: "هل تختلف دقة تقديرات معلمة القدرة للفرد المقدّرة باستخدام النموذج اللوجستي ثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (أربعة بدائل؛ ثلاثة بدائل؛ بديلان) ؟

للإجابة عن هذا السؤال تم استخراج المتوسطات الحسابية للخطأ المعياري (Standard Error of Estimate) في تقدير معلمة القدرة في كل نموذج من النماذج الثلاثة للاختبار كما هو موضّح في الجدول (١٧).

جدول (١٧)

المتوسطات الحسابية للخطأ المعياري في تقدير معلمة القدرة والانحراف المعياري لنماذج الاختبار

نموذج الاختبار	المتوسط الحسابي لمعلمة القدرة	المتوسط الحسابي للخطأ المعياري في تقدير معلمة القدرة	الانحراف المعياري
أربعة بدائل	٠,٧٧٠	٠,٣٨٩	٠,٠٦٧
ثلاثة بدائل	٠,٣٣٦	٠,٩٥٧	٠,٠٣٦
بديلان	٠,٦٩٦	٠,٦٠١	٠,٠٤١

يُلاحظ من الجدول (١٧) وجود فروق ظاهرية بين متوسطات قيم الخطأ المعياري لتقدير معلمة القدرة للمفحوصين باختلاف نموذج الاختبار. وتبيّن أن نموذج الاختبار ذي الأربعة بدائل هو الأقل خطأً في تقدير معلمة القدرة، يليه النموذج ذو البديلين، ثم النموذج ذو الثلاثة بدائل. ولاختبار دلالة الفروق بين هذه المتوسطات تمّ استخدام أسلوب تحليل التباين الأحادي كما يوضّحها الجدول (١٨).

الجدول (١٨)
نتائج تحليل التباين الأحادي للكشف عن دلالة الفروق بين المتوسطات الحسابية لقيم الخطأ المعياري في تقدير معلمة القدرة للمفحوصين تبعًا لنموذج الاختبار

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "F"	مستوى الدلالة
نموذج الاختبار	٨٢,٢٢٥	٢	٤١,١١٢	٢٠,٥٥٦	٠,٠٠١
الخطأ	٣,٧٢٨	١٤٩٧	٠,٠٠٢		
الكلي	٨٥,٩٥٢	١٤٩٩			

يُلاحظ من الجدول (١٨) وجود فرق ذي دلالة إحصائية عند مستوى الدلالة (α = ٠,٠٥) بين المتوسطات الحسابية للخطأ المعياري في تقدير معلمة القدرة للمفحوصين يُعزى لمتغير نموذج الاختبار (أربعة بدائل؛ ثلاثة بدائل؛ بديلان)، إذ بلغت قيمة "ف" المحسوبة (٢٠,٥٥٦) وبمستوى دلالة مقداره (٠,٠٠٠). وللكشف عن مصدر الفروق تم استخدام اختبار "شيفي" للمقارنات البعدية كما هو موضّح في الجدول (١٩).

الجدول (١٩)
نتائج اختبار "شيفيه" للمقارنات البعدية بين المتوسطاتالحسابية لقيم الخطأ المعياري في تقدير معلمة القدرة تبعًا لنموذج الاختبار

نموذج الاختبار	المتوسط الحسابي	أربعة بدائل	ثلاثة بدائل	بديلين
أربعة بدائل	٠,٣٨٩	-	-٠,٥٦٧*	٠,٨٥٤
ثلاثة بدائل	٠,٩٥٧		-	٠,٣٥٥*
بديلان	٠,٦٠١			-

* $\alpha \leq$ ٠,٠٥

يُلاحظ من الجدول (١٩) أن الفروق بين المتوسطات الحسابية للخطأ المعياري في تقدير معلمة القدرة للمفحوصين تُعزى للفرق بين متوسط الخطأ المعياري في تقدير القدرة في النموذج ذو الأربعة بدائل مقارنة مع نموذج الاختبار ذي الثلاثة بدائل، والفرق بين متوسط الخطأ المعياري في تقدير القدرة في النموذج ذي الثلاثة بدائل مقارنة مع النموذج ذي البديلين، بينما لم يكن الفرق بين متوسط الخطأ المعياري في تقدير القدرة في النموذج ذي البديلين مقارنة بالنموذج ذي الأربعة بدائل ذا دلالة إحصائية. وهذا يعني أن النموذج ذا الأربعة بدائل هو الأكثر دقة في تقدير معلمة القدرة، يليه النموذج ذو البديلين، ثم النموذج ذو الثلاثة بدائل.

مناقشة النتائج والتوصيات

مناقشة نتائج السؤال الأول

أظهرت نتائج الإجابة عن هذا السؤال أن الفروق بين المتوسطات الحسابية في تقدير معلمة الصعوبة تُعزى للنموذج ذي البديلين، أي أن هذا النموذج أعطى قيمًا مرتفعة لمعلمة صعوبة الفقرة عند مقارنته مع النماذج الأخرى. وقد تحتاج هذه النتيجة إلى المزيد من الدراسة، واختلفت نتائج الإجابة عن هذا السؤال نسبياً عن نتائج دراسة تارنت و وير(& Tarrant Ware, 2010) والتي أشارت إلى تفوق النموذج ذي الثلاثة بدائل في تقدير معلمة الصعوبة، كما اختلفت عن نتائج دراسة ملياني (٢٠١٠) والتي أشارت إلى وجود فروق ذات دلالة إحصائية بين قيم معاملات الصعوبة المقدرة في النماذج الثلاثة للاختبار (ثلاثة بدائل؛ أربعة بدائل؛ خمسة بدائل)، ولصالح النموذج ذي الثلاثة بدائل.

مناقشة نتائج السؤال الثاني

أظهرت نتائج الإجابة عن هذا السؤال أن الفروق بين المتوسطات الحسابية في تقدير معلمة التمييز تُعزى للنموذج ذي الأربعة بدائل عند مقارنته مع كل من النماذج الأخرى، وكذلك نموذج الاختبار ذي البديلين للفقرة، وقد تُعزى هذه النتيجة إلى ان زيادة عدد البدائل تقلل من حصول المفحوص على إجابة صائبة من خلال التخمين العشوائي والذي بدوره يقلّل من قدرة الفقرة على التمييز، واختلفت النتائج المتعلقة بالإجابة عن هذا السؤال مع نتائج دراسة يامان (Yaman, 2011) والتي أشارت إلى عدم وجود فروق ذات دلالة إحصائية بين الاختبارات الثلاثة (ثلاثة بدائل؛ أربعة بدائل؛ خمسة بدائل) في الخصائص السيكومترية المتعلقة بمعلمة التمييز.

مناقشة نتائج السؤال الثالث

أظهرت نتائج الإجابة عن هذا السؤال أن النموذج الأكثر ثباتًا هو النموذج ذو الأربعة بدائل، يليه النموذج ذو البديلين. وقد تُعزى هذه النتيجة إلى أن عدد بدائل الفقرة يؤثر إيجابًا على ثبات الاختبار؛ إذ تقل فرصة تخمين الإجابة الصحيحية بزيادة عدد البدائل. واختلفت نتيجة الإجابة عن هذا السؤال عن نتائج دراسة الشريفين وطعامنة (٢٠٠٩) والتي أشارت إلى عدم وجود فروق بين معاملات الثبات تُعزى لعدد البدائل. كما اختلفت عن نتائج دراسة باغاي و أمراهي (Baghaei & Amrahi, 2011) والتي أشارت إلى عدم وجود أثر يُعزى لنموذج الاختبار المستخدم (ثلاثة بدائل؛ أربعة بدائل؛ خمسة بدائل) على صدق وثبات الاختبار

المستخدم حيث كانت قيم الثبات متقاربة لهذه النماذج. واتفقت نتائج الدراسة الحاليّة المتعلقة بالإجابة عن هذا السؤال نسبياً مع نتائج دراسة يامان (Yaman, 2011) والتي أشارت إلى أن مستوى ثبات الاختبار ذي الثلاثة بدائل أو خمسة بدائل أعلى منها في حالة النموذج ذي الأربعة بدائل. كما اتفقت نتائج الإجابة عن هذا السؤال نسبياً مع نتائج دراسة بني عطا والرباعي (٢٠١٣) والتي أشارت إلى وجود فروق دالة إحصائيًا بين قيم معاملات الثبات في النماذج المختلفة حيث جاءت لصالح النموذج ذو الخمسة بدائل.

مناقشة نتائج السؤال الرابع

فيما يتعلق بالإجابة عن هذا السؤال الخاص بدقة تقدير معلمة القدرة للفرد باستخدام النموذج اللوجستي ثلاثي المعلمة لنماذج الاختبار الثلاثة، تمّ تقدير القيم المتحققة لمعلمة القدرة للأفراد والأخطاء المعيارية في تقديرها لكل نموذج من النماذج الثلاثة للاختبار. وقد أظهرت نتائج الإجابة عن هذا السؤال أن النموذج ذا الأربعة بدائل هو الأكثر دقة مقارنة بالنموذجين الأخريين حيث كان متوسط قيم الأخطاء المعيارية لتقدير معلمة القدرة للأفراد في النموذج ذي الأربعة بدائل هو الأقل مقارنةً بالنموذجين الآخرين. وقد تُعزى هذه النتيجة إلى أن النموذج ذا الأربعة بدائل يحتاج من الطالب القراءة المتمعنة لبدائل الفقرة وقدرة عالية من الربط والدقة للوصول إلى الإجابة الصحيحة. وتبيّن أيضاً أن دقة تقدير معلمة القدرة في النموذج ذي البديلين أفضل منها مقارنة بالنموذج ذي الثلاثة بدائل، واتفقت نتائج الإجابة عن هذا السؤال نسبياً مع نتائج دراسة الشريفين وطعامنة (٢٠٠٩) والتي أشارت إلى وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقديرات معلمة القدرة للمفحوصين تُعزى للنموذج ذي عدد البدائل الأقل إذ كانت الفروق لصالح النموذج ذي الثلاثة بدائل ثم النموذج ذي الأربعة بدائل ثم النموذج ذي البديلين. واختلفت نتائج الإجابة عن هذا السؤال مع نتائج دراسة ملياني (٢٠١٠) والتي أشارت إلى عدم وجود فروق ذات دلالة إحصائية بين متوسطات قيم الأخطاء المعيارية في تقديرات معلمة القدرة في نماذج الاختبار الثلاثة المستخدمة وهي (ثلاثة بدائل؛ أربعة بدائل؛ خمسة بدائل). كما اختلفت مع نتائج دراسة عطا والرباعي (٢٠١٣) والتي أشارت نتائجها إلى عدم وجود فروق ذات دلالة إحصائية في دقة تقدير معلمة القدرة باختلاف عدد البدائل في نماذج الاختبار المستخدمة (خمسة بدائل؛ أربعة بدائل؛ ثلاثة بدائل).

التوصيات

في ضوء النتائج التي خلصت إليها الدراسة الحالية فإن الباحث يوصي بما يأتي:

- اختيار عدد بدائل للفقرة تبعًا لمستوى قدرة المفحوصين، إذ إن الاختبار ذا الأربعة بدائل يتناسب مع الطلبة ذوي القدرة المتوسطة؛ فهو يُعطي معلومات أكثر، أما الطلبة ذوو مستوى القدرة المرتفع فيفضل استخدام الاختبار ذي الأربعة بدائل لأنه يعطي معلومات أكثر عنهم.
- إجراء دراسات لتحري أثر شكل وطبيعة متن فقرة الاختبار من متعدد على دقة تقديرات معالم الفقرات (الصعوبة، والتمييز، والتخمين)، وعلى دالة المعلومات لكل من الفقرة والاختبار.
- إجراء دراسة مشابهة، مع تقسيم طلبة عينة الدراسة بحسب كل نموذج إلى ثلاثة مستويات للقدرة (مرتفع، متوسط، منخفض)، وتحري الخصائص السيكومترية لكل اختبار.

المراجع

بني عطا، زايد والرباعي، إبراهيم. (٢٠١٣). أثر عدد البدائل وتغيير موقع المموه القوي في فقرات اختبار الاختيار من متعدد على معالم الفقرات وقدرة الفرد ودالة المعلومات. المجلة الأردنية في العلوم التربوية، ٩(٣)، ٣١٩-٣٣٣.

بني عطا، زايد والشريفين، نضال. (٢٠١١). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. المجلة الأردنية في العلوم التربوية، ٨(٢)، ١٥١-١٦٦.

حسن، السيد محمد. (٢٠٠٦). دراسة مقارنة بين الطريقة التقليدية ونموذج راش في اختيار فقرات مقياس مداخل الدراسة لدى طلبة الجامعة. رسالة ماجستير غير منشورة. جامعة الزقازيق، القاهرة، مصر.

ربيع، محمد شحاته. (٢٠٠٨). قياس الشخصية. دار المسيرة للنشر والتوزيع والطباعة، عمان: الأردن.

الشريفين، نضال وطعامنة، إيمان. (٢٠٠٩). أثر عدد البدائل في اختبار الاختيار من متعدد في تقديرات القدرة للأفراد والخصائص السيكومترية للفقرات والاختبار وفق نموذج راش في نظرية الاستجابة. المجلة الأردنية في العلوم التربوية، ٥(٤)، ٣٠٩-٣٣٥.

عبابنة، عماد. (٢٠٠٩). الاختبارات محكية المرجع. عمان: دار المسيرة للنشر والتوزيع والطباعة.

علام، صلاح الدين محمود. (٢٠٠١). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية. القاهرة: دار الفكر العربي للنشر والتوزيع.

علام، صلاح الدين. (٢٠٠٢). القياس والتقويم التربوي النفسي أساسياته وتطبيقاته وتوجيهاته المعاصرة. القاهرة: دارالفكر العربي للنشر والتوزيع.

عودة، أحمد. (٢٠١٠). القياس والتقويم في العملية التدريسية (ط٤). إربد: دار الأمل للنشر والتوزيع.

الغامدي، عبدالله بن أحمد. (٢٠٠٨). أثر عدد البدائل في الخصائص السيكومترية للاختبار التحصيلي في الرياضيات. رسالة ماجستير غير منشورة، جامعة أم القرى، المملكة المملكة الاردنية الهاشمية.

فقوسة، محمد. (١٩٩٥). أثر عدد البدائل في فقرات اختبار الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته. رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان.

المعتوق، بدر. (٢٠٠٦). التقويم التربوي، الاختبارات التحصيلية(ط٢). الكويت: وزارة الأوقاف والشؤون الإسلامية.

ملياني، زياد. (٢٠١٠). مقارنة أثر اختلاف عدد البدائل في اختبار الاختيار من متعدد على صعوبة الفقرة وقدرة الفرد في ضوء كل من النظرية الكلاسيكية ونموذج راش. أطروحة دكتوراه غير منشورة، جامعة أم القرِّى، مكة المكرمة.

Baghaei, P., and Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple choice items. *Psychological Test and Assessment Modeling, 53*(2), 192-211.

Baker, F. B. (2004). *The basic of item response theory. ERIC Clearinghouse on Assessment and Evaluation*, University of Maryland, College Park, MD.

Demars, C. (2010). Item response theory. Oxford University Press, Inc. USA.

Dennick, R., and Tavakol, M. (2011). Making sense of Cronback's Alpha. *International Journal of Medical Education, 2*, 53-55.

Ebel, Robert, and David, A. Frisbi (1991). *Essentials of educational measurement* (3rd ed.). USA, New Jersey: Prentice-Hall.

Embreston, S, and Reise, S (2000). I*tem reesponse theory for psychologist*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Francisco, A, Olea, J., and P, V. (2001). Analysis of the optimum number of alternatives from the Item Response *Theory. Psicothema, 13*(1), 152-158

Haladyna, T, M. (2004). *Developing and validating multiple-choice test items*. Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, USA. 25.

Hambelton, R. K. (1993). Item response theory: A broad psychometric framework for measurement advances. *Psychometrika, 6*(3), 353-556.

Hambleton, R. K., and Jones, R. (1993). Item Paramter estimation errors and their influence on test information function. *Applied Measurment in Education, 7*(3), 171-186.

Hambleton, R. K., and Jones, R. (1994). Comparison of Classical Test Theory and Item Response Theory and their Applications to test Development. *Educational Measurement: Issue and Practice, 8*(3), 67-88.

Hambleton, R. K., Swaminthan, H., and Rogers, H. J. (1991). Fundamentals of item response Theory. I*nternational Educational and Professional*. Sage Publication, Inc., Newbury Park, USA.

Lia, Wen-Wei (2012). The four-parameter logistic item response theory model asrobust method of estimating ability despite aberrant responses. *Social Behavior and Personality, 40,* 1679-1694.

Nunnally, J. (1978). *Psychometric Theory*. New York, Mccrow-Hill, USA.

Popham, J. (2003). *Educational Assessment*. Pearson Education Inc., USA.

Razali, N., Razali, W, and Yap, B. (2011). Power comparison of shapiro-wilk, kolomogorov-smirnov, Lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

Shneider, M. C., Eagn, K. L. and Julian, M. W. (2013). Classroom assessment in the context of high-stakes testing. in mcmillan. James H.(Ed), *Research on Classroom Assessment*. Sage Publication.

Tarrant, M., and Ware, J. (2010). A comparison of the psychometric of three and four- option multiple choice questions in nursing assessment. *Nurse Education Today, 30*(6), 539-543.

Wang, S., and Wang, T. (2002). *Relative precision of ability estimation in polytomous cat: a comparison under the generalized partial credit model and graded response model*. Paper Presented at the annual meeting of the American Educational Research Association, New Orleans.

Yaman, S. (2011). The optimal number of choices in multiple–choice tests: Some evidence for science and technology education. *The New Educational Review, 7*(3), 227-241.

Yen, W., and Fitzpatrick, A. R. (2006). Education measurement. In Brennan, Robert L., *Educational Measurement*. Praeger Publishers. USA.