International Journal for Research in Education

Volume 43 | Issue 3 Article 6

2019

Information Function for Item and Test and Reliability when Using Three Forms of Multi-Choice Test under Item Response Theory

mo'en slman al-Nasraween Dr
assistent teacher, mueen1234@hotmail.com

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/ijre

Part of the Art Education Commons, Bilingual, Multilingual, and Multicultural Education Commons, Curriculum and Instruction Commons, Disability and Equity in Education Commons, Educational Administration and Supervision Commons, Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, and the Gifted Education Commons

Recommended Citation

al-Nasraween, mo'en slman Dr (2019) "Information Function for Item and Test and Reliability when Using Three Forms of Multi-Choice Test under Item Response Theory," *International Journal for Research in Education*: Vol. 43 : Iss. 3 , Article 6.
Available at: https://scholarworks.uaeu.ac.ae/ijre/vol43/iss3/6

This Article is brought to you for free and open access by Scholarworks@UAEU. It has been accepted for inclusion in International Journal for Research in Education by an authorized editor of Scholarworks@UAEU. For more information, please contact fadl.musa@uaeu.ac.ae.

DOI : http://doi.org/10.36771/ijre.43.3.19-pp159-181

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

Information Function for Item and Test and Reliability When Using Three Forms of Multi-Choice Test Under Item Response Theory

Mo'en Salman Alnasraween
Assistant Professor –Educational and Phsychological College Amman Arab Uinversity
mueen@aau.edu.jo & mueen1234@hotmail.com

Abstract

This study aimed at revealing the effect of the number of alternatives in the multiple choice tests on the information function of the item and the test according to the three-parameter model under the item response theory. To achieve the objective of the study, a multi-choice achievement test was constructed in the second part of mathematics subject for the 10th grade students in the public schools in the capital city of Amman. The final test consists of 38 paragraphs and three models are prepared, which differ only in the number of item alternatives. The sample of the study consisted of 1530 students. The results of the study showed statistically significant differences in reliability in favor of the five- alternative form, as well as the four-alternative form, the results also showed no statistically significant differences between the arithmetic means of the information function due to the variable number of item alternatives.

Keywords: number of alternatives; multiplechoice test; three parameter model.

DOI : http://doi.org/10.36771/ijre.43.3.19-pp159-181

دالة معلومات الفقرة والاختبار والثبات عند استخدام ثلاثة نماذج من اختبار الاختيار من متعدد في إطار نظرية الاستجابة للفقرة

معين سلمان سليم النصراوين
أستاذ مساعد في القياس والتقويم- كلية العلوم التربوية والنفسية جامعة عمان العربية- الأردن
mueen@aau.edu.jo , mueen1234@hotmail.com

مستخلص البحث:
هدفت هذه الدراسة إلى الكشف عن أثر عدد البدائل في اختبار الاختيار من متعدّد على دالة معلومات الفقرة والاختبار وفق النموذج الثلاثي المعلمة في إطار نظرية الاستجابة للفقرة؛ ولتحقيق هدف الدراسة تمَّ بناء اختبار تحصيلي من نوع الاختيار من متعدّد في الجزء الثاني من مبحث الرياضيات لطلبة الصف العاشر الأساسي في المدارس الحكومية في محافظة العاصمة عمّان، تكوّن الاختبار في صورته النهائية من (38) فقرة وتم إعداد ثلاثة نماذج منه تختلف فقط في عدد بدائل فقراتها، تكوّنت عينة الدراسة من (1530) طالبًا من طلبة الصف العاشر، أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية في الثبات لصالح النموذج ذي الخمسة بدائل، والنموذج ذي الاربعة بدائل، كما أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لدالة المعلومات يُعزى لمتغير عدد البدائل للفقرة.

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

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

المقدمة

انتشر استخدام الاختبارات انتشاراً واسعاً في العديد من المجالات، وتستخدم للإشارة إلى درجة أو مستوى النجاح الذي يحرزه المفحوص في مجال دراسي عام أو متخصص، وتكتسب عملية تقويم الأفراد أهمية كبيرة بقدر أهمية القرارات التي تبنى عليها، وبقدر خطورة القرارات الخاطئة التي يمكن أن تترتب على ذلك في المواقف والمجالات المتعددة، وبالتالي يجب أن توفر الاختبارات معلومات تتسم بالصدق والدقة، وهذا يمكن أن يتحقق من خلال التخطيط والإعداد الجيد لهذه الاختبارات.

أدّى انتشار الاختبارات الموضوعية إلى التعدّد في الأشكال التي تتخذها هذه الاختبارات، فهنالك اختبارات المزاوجة، واختبارات الصواب والخطأ، واختبارات التكميل، واختبارات الاختيار من متعدد Multiple Choice Question التي تُعدّ أكثر أنواع الأسئلة الموضوعية استخدامًا في المؤسسات التعليمة؛ وأكثرها استخداماً وانتشارًا في الاختبارات التحصيلية وغيرها من الاختبارات، ويتكّون سؤال الاختيار من متعدّد من قسمين رئيسين هما: متن السؤال الذي يحتوي على المشكلة، ومجموعة من البدائل أو المموهات بالإضافة للإجابة الصحيحة، وعمومًا يُعدّ هذا النمط من الاختبارات الموضوعية الأكثر استخدامًا عند بناء بنوك الأسئلة، أو إجراء الاختبارات التكيفية؛ وذلك لما يتمتع به هذا النمط من مرونة، وقدرة على اختبار عددٍ كبيرٍ من المفحوصين، وسهولة التصحيح والموضوعية وثبات الدرجات، وتوفر الاختبارات الموضوعية تغطية جيدة للمادة الدراسية، وتقيس نواتج التعلم في المستويات العقلية العليا من المجال المعرفي، ويُعد اختيار عدد بدائل الفقرة المناسب من القضايا التي جرى الاهتمام بها، لمعرفة الأثر الناتج منها على دالة المعلومات وعلى الخطأ المعياري في التقدير في نظرية الاستجابة للفقرة (Sahin & Anil, 2017).

ونظرية الاستجابة للفقرة تستند على افتراض أنَّ احتمال إجابة المفحوص إجابة صحيحة عن الفقرة يتحدّد بقدرة الفرد وبصعوبة الفقرة، كما توضّح هذه النظرية العلاقة بين قدرة المفحوص واحتمال إجابته على الفقرة إجابة صحيحة؛ وذلك من خلال نماذج رياضية تصف هذه العلاقة بدلالة قدرة الفرد ومعالم الفقرة، وتفترض هذه النظرية إمكانية التنبؤ بأداء الأفراد، أو تفسير أدائهم في اختبار نفسي أو تربوي معين في ضوء خاصية أو خصائص مميزة لهذا الأداء، وقد أنبثق عن هذه النظرية مجموعة من النماذج تُعرف بنماذج السمات الكامنة تصف العلاقة بين أداء الفرد على الفقرة والقدرة الكامنة وراء هذا الأداء (Kim & Lee, 2017).

وتقوم نظرية الاستجابة للفقرة على افتراضات أساسية هي: (Demars, 2010; Onn, 2013)

(أ) أحادية البعد Uni- Dimensionality **:** ويتحقق هذا الافتراض حينما تقيس الفقرات المكونة للاختبار بعدًا واحدًا فقط؛ ممّا يعني وجودَ عاملٍ مؤثرٍ واحدٍ فقط على الأداء، ويُسمى هذا العامل بالقدرة المقاسة بالاختبار، والذي يمكن أن تُعزى إجابة المفحوص لفقرات الاختبار لهذه القدرة.

(ب) الاستقلال الموضعي Local Independence: ويعني أنَّ احتمال إجابة المفحوص عن الفقرة إجابة صحيحة يكون مستقلاً إحصائيًا عند مستوى قدرة معين عن إجابته على أية فقرة أخرى في الاختبار.

(ج) منحنى خصائص الفقرة Item Characteristic Curve (ICC): أي أنه يمكن التعبير بيانياً باقتران متصاعد يُسمى منُحنى خصائص الفقرة عن العلاقة بين أداء الفرد والسمة التي يقيسها الاختبار.

(د) انعدام تأثير السرعة في الإجابة Speededness: أي أن عدم إجابة الفرد إجابة صحيحة عن فقرات الاختبار لا يعود لضيق الوقت، بل لافتقاره للقدرة اللازمة للإجابة عن السؤال إجابة صحيحة.

النموذج ثلاثي المعلمة

يفترض هذا النموذج أن الفقرات تتباين من حيث درجة الصعوبة والتمييز، كما أنَّ مستويات التخمين لدى المفحوصين تختلف من فقرة لأخرى، واحتمال أن يجيب المفحوص عن الفقرة إجابة صحيحة من خلال التخمين العشوائي تساوي (1) مقسومًا على عدد بدائل الفقرة وتأخذ معادلة هذا النموذج الشكل الآتي(De Ayala, 2010 : , p.149)

$$p_i(\theta) = Ci + (1 - Ci)\frac{e^{Dai(\theta - bi)}}{1 + e^{Dai(\theta - bi)}}$$

حيث إن:

$P_i(\theta)$: تمثل احتمال إجابةِ الفردِ ذي القدرة (θ) على الفقرة (i) إجابة صحيحة.

D: قيمة ثابتة وتساوي (1.7).

b_i: معلمة الصعوبة للفقرة (i).

e. الأساس اللوغاريتمي الطبيعي ويساوي (2.718).

دالة معلومات الفقرة والاختبار

وتُعدّ دالة المعلومات من المفاهيم الأساسية في نظرية الاستجابة للفقرة؛ ومن خلال دالة المعلومات يمكن تحديد النقطة على متصل القدرة حيث يمكن الحصول على أقصى معلومات عند مستوى قدرة (θ) محدّد، وهي أداة لوصف كل من الفقرات والاختبار، كما أنها وظيفة للعلاقة بين قدرة الفرد والمعلومات التي تقدمّها الفقرة، إذ تعبّر هذه الدالة عن كمية المعلومات التي يمكن أن توفرها فقرات الاختبار عند مستوى قدرة محدّد، ولعل من أهم التطبيقات التي تزود بها دالة المعلومات هي (Ha, 2017):

أ- المساهمة في بناء الاختبارات بأنواعها المختلفة، واختيار فقرات الاختبار بناءً على كمية المعلومات التي يمكن أن تقدّمها كل فقرة وبما يتوافق مع هدف الاختبار والغرض الذي صمم لأجله.

ب- المقارنة بين الاختبارت المختلفة، وتقييم مدى دقة أدوات القياس، والمقارنة بين طرق التصحيح المستخدمة في الاختبارات المختلفة، ولعل أهم ما يُميز دالة المعلومات أنها يمكن أن تُعَرف عند كل مستوى من مستويات القدرة (θ) على المتصل، ولا تتأثر دالة المعلومات لأية فقرة بالمعلومات التي تقدمها الفقرات الأخرى بنفس الاختبار، والمعادلة الآتية تبين دالة معلومات الفقرة (Yen & Fitzpatrick, 2006 , p.126):

$$I_i(\theta) = \frac{P'(\theta)^2}{Pi(\theta)Qi(\theta)}$$

حيث إن:

$I_i(\theta)$: دالة المعلومات للفقرة (i).

$P'(\theta)$: المشتقة الأولى لاحتمال الإجابة الصحيحة عن الفقرة (i).

$Pi(\theta)$: احتمال الإجابة الصحيحة عن الفقرة (i).

$Qi(\theta)$: احتمال الإجابة الخطأ عن الفقرة (i).

علاقة معالم الفقرات بدالة المعلومات

يمكن وصف علاقة كل من معلمة الصعوبة والتمييز والتخمين للفقرة، وتحديد الأثر الذي تتركه الفقرات على دالة المعلومات من خلال المعادلة الآتية (Linden & Hambleton,1997):

$$I(\theta,Ui)max = \frac{D^2 ai^2}{8(1-ci^2)}[1 - 20ci - 8ci^2 + (1+8ci)^{3/2}]$$

حيث إن:

$I(\theta,Ui)max$: قيمة دالة المعلومات، c_i : معلمة التخمين.، a_i : معلمة التمييز، D: ثابت ويساوي (1.77).

وتبيّن المعادلة أعلاه أن هناك أثراً لكل من معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين على دالة المعلومات على النحو الآتي:

أ- كلما كانت قيمة معلمة الصعوبة (b) قريبة من مستوى القدرة (θ) للمفحوص أدى ذلك لزيادة كمية المعلومات التي تقدمها الفقرة.

ب- تزداد قيمة دالة المعلومات كلما قلت معلمة التخمين (c) واقتربت من الصفر، وأعلى قيمة لدالة المعلومات يمكن الوصول لها حينما تصبح قيمة معلمة التخمين مساوية للصفر، وعندها فإن قيمة القيمة القصوى لدالة المعلومات ($_{max}\theta$) تساوي معلمة الصعوبة (b)، وعندما تكون قيمة معلمة التخمين أكبر من صفر (c >0) فإن قيمة ($_{max}\theta$) تصبح إلى يمين معلمة الصعوبة (b).

ج- بزيادة قيمة معلمة التمييز (a) يقل التباين، ويرتفع منحنى المعلومات وبالتالي تزداد قيمة المعلومات.

دالة معلومات الاختبار

إن دالة المعلومات للاختبار عند مستوى قدرة (θ) هي عبارة عن مجموع دوال المعلومات للفقرات المكونة لهذا الاختبار عند ذلك المستوى للقدرة (θ)، والمعادلة الآتية تبين كيفية إيجاد دالة المعلومات للاختبار :

$$I(\theta) = \sum_{i=1}^{n} I_i \ (\theta)$$

حيث إن:

I(θ) : دالة معلومات الاختبار، دالة معلومات الفقرة. $I_i(\theta)$

ويرتبط مقدار المعلومات التي يزوّد بها الاختبار عند مستوى قدرة (θ) عكسيًا مع مستوى الدقة كما يتّضح من المعادلة الآتية (Partchev ,2004)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

حيث إن:

SE(θ) : الخطأ المعياري في التقدير (Standard Error of Estimate)، I(θ) : دالة المعلومات.

إن معدل الخطأ المعياري في تقدير معلمة القدرة للمفحوص يرتبط مع الثبات وفقًا للنظرية الكلاسيكية في القياس المعروف بالعلاقة $r_{xx}=1-(SEE)^2$؛ فالاختبار ذو الثبات المرتفع تكون قيمة الخطأ المعياري في التقدير له متدنية، وكمية المعلومات التي يقدمها تكون كبيرة؛ فالخطأ المعياري

للتقدير كما تبيّن المعادلة أعلاه يرتبط عكسيًا مع الجذر التربيعي لدالة المعلومات للاختبار (Emberston & Reise, 2000).

الدراسات السابقة

أجرى واغنر وهارفي (Wagner & Harvey, 2003) دراسة هدفت إلى تطوير أداة جديدة لأوجه التفكير الناقد لاختبار واتسون-جليسر باستخدام النموذج الثلاثي المعلمة، وقام الباحثان بزيادة عدد البدائل (المموهات) لتكوين اختبار جديد Wagner Assessment Test(WAT) للحد من درجة التخمين، حيث تكونت عينة الدراسة من (407) طالبٍ وطالبة في المستوى الجامعي، وقد أشارت النتائج إلى وجود علاقة طردية بين زيادة عدد البدائل ونسبة التخمين في الاختبار المطور (WAT)، وكذلك تحسّن في دالة معلومات الفقرة للاختبار المطور، وكانت أخطاء القياس في اختبار واتسون- جليسر (WGCT) أكبر من أخطاء القياس للاختبار المطور (WAT) بمقدار 50% على الأقل.

أجرى عيلبوني (2007) دراسة هدفت إلى معرفة أثر عدد البدائل وجاذبيتها في فقرات الاختيار من متعدد على التوافق مع النموذج ثلاثي المعلمة، ولتحقيق هدف الدراسة قام الباحث ببناء اختبار تحصيلي في مادة الرياضيات للصف التاسع الأساسي، تكونت أداة الدراسة من ثلاثة أشكال اختبارية لكل منها (50) فقرة، في الشكل الأول خمس بدائل للفقرة، والشكل الثاني له ثلاثة بدائل بحذف بديلين عشوائيًا من الشكل الأول، والشكل الثالث له ثلاثة بدائل أيضاً بحذف البديلين الأقل تمييزًا من الشكل الأول، تكونت عينة الدراسة من (1656) طالبًا وطالبة من طلبة الصف التاسع في مدارس مديرية التربية والتعليم إربد الثانية، تم تطبيق الاختبارات بأشكالها الثلاثة على أفراد الدراسة، ومن ثم حُللت الاستجابات، وأظهرت النتائج توافق الفقرات مع النموذج الثلاثي المعلمة لدى الاختبار بأشكاله الثلاثة، كما أظهرت النتائج أيضاً تطابق الفقرات للأشكال الثلاثة للاختبار مع النموذج ثلاثي المعلمة بعد حذف (3) فقرات من الشكل الأول للاختبار والذي يحتوي على (50) فقرة. وأظهرت النتائج ومن خلال مقارنة خصائص الفقرات والاختبار للأشكال الثلاثة من حيث معاملات الصعوبة والتمييز والتخمين ودالة المعلومات وثبات الاختبار إلى أنَّ الشكلَ الأولَ ذا الخمسةِ بدائل هو الأفضل مقارنة مع الشكلين الثاني والثالث.

وأجرى الغامدي (2008) دراسة هدفت لمعرفة أثر عدد البدائل في الخصائص السيكومترية لاختبار تحصيلي في الرياضيات للصف التاسع الأساسي في المملكة العربية السعودية، تكونت عينة الدراسة من (1400) طالبٍ، وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية عند (α=0.05) بين متوسطات قيم معاملات الصعوبة للنموذجين ذي الخمسة بدائل وذي الثلاثة بدائل وكانت هذه الفروق لصالح النموذج ذي الثلاثة بدائل، كما أظهرت النتائج فروقاً ذات دلالة إحصائية بين متوسطات قيم معاملات الصعوبة للنموذجين ذي الأربعة بدائل وذي الثلاثة بدائل وكانت الفروق لصالح النموذج ذي الثلاثة بدائل، ولم تظهر النتائج أية فروق دالة إحصائياً تعزى للنماذج فيما يتعلق بمعلمة التخمين.

أما دراسة الشريفين وطعامنة (2009) 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 (40) فقرة. وتمّ تشكيل ثلاث صور من الاختبار تختلف عن بعضها بعضًا 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 (600) طالبٍ وطالبة بواقع (200) طالبٍ وطالبة لكل صورة، أظهرت النتائج عدم وجود

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

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، بينما كانت معاملات الثبات للفقرات متساوية، كما أظهرت النتائجُ وجودَ فروقٍ ذاتِ دلالةٍ إحصائيَّةٍ بين متوسطات الأخطاء المعياريَّة في تقديرات معالم القدرة للأفراد؛ حيث كانت التَّقديرات أكثر دقة لقدرات أفراد الصورة الأولى للاختبار من قدرات أفراد الصورة الثالثة له، وكانت التقديرات لقدرات أفراد الصورة الثانية أكثر دقة منها للصورة الثالثة، بينما لم تختلف دقة تقدير معالم القدرة للأفراد لكل من الصورتين الأولى والثانية، إضافة إلى ذلك أعطت الصورة الأولى كمية معلومات أكبر من الصورتين الثانية والثالثة عند مستويات القدرة المنخفضة، بينما أعطت الصورة الثالثة للاختبار كمية أكبر من المعلومات عند مستويات القدرة المتوسطة والمرتفعة من الصورتين الأخريين. كما بينت النتائج وجود فروق دالة إحصائيًا بين معاملات صدق المحك ولصالح الصورة الثانية للاختبار.

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، تختلفان 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 الممّوه الضعيف في عملية الإجابة ليتم حذفه وتشكيل الصورة الثانية ذات الثلاثة بدائل من الاختبار، تكونت صورة الاختبار النهائية من (41) فقرة في كل نموذج (ثلاثة بدائل، أربعة بدائل)، وأشارت نتائج الدراسة إلى أنَّ الصورة الثانية (ثلاثة بدائل) كانت أكثر فاعلية على الرغم من قلة المموهات؛ وذلك بسبب قوة هذه المموهات.

كما أجرى ملياني (2010) 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. ولتحقيق الغرض من الدراسة تم بناء اختبار تحصيلي من نوع الاختيار من متعدد في الرياضيات لطلبة الصف الثاني الثانوي مكون من (25) فقرة بصورته النهائية من نوع الاختيار من متعدد في ثلاثة نماذج (ثلاثي البدائل، رباعي البدائل، خماسي البدائل)، تم تحليل استجابات المفحوصين البالغ عددهم (582) طالبًا لجميع نماذج الاختبار الثلاثة وفق النظرية الكلاسيكية والنموذج اللوجستي أحادي المعلمة، أظهرت النتائجُ وجودَ أثرٍ ذي دلالة إحصائيَّة للنموذج على قدرات الأفراد المقدرة وفق النظرية الكلاسيكية ولصالح النموذج ثلاثي البدائل، في حين لم يتبين وجودَ أثرٍ ذي دلالة إحصائيَّة للنموذج على قدرات الأفراد المقدرة وفق نموذج راش، كما كشفت نتائج الدراسة عن وجود علاقة ارتباطيَّة ذات دلالة إحصائيَّة بين قيم معاملات الصُّعوبة المقدرة وفق النَّظريَّة الكلاسيكيَّة 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 النموذج أحادي المعلمة ولصالح النموذج ثلاثي البدائل.

أما دراسة باغاي وأمراهي (Baghaei & Amrahi, 2011) فهدفت إلى الكشف عن أثر اختلاف عدد البدائل في أسئلة الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته، كما هدفت الدراسة إلى تحديد العدد الأمثل من البدائل المستخدمة في أسئلة الاختيار من متعدد والتي يكون صدقها وثباتها عاليًا، تكونت عينة الدراسة من(180) طالبًا في إيران، 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، تم تطبيقها على أفراد الدراسة، أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين

أثر النماذج الثلاثة للاختبار(ثلاثة، وأربعة، وخمسة) بدائل على صدق وثبات الاختبار حيث كانت متوسطات كل من الثبات والصدق 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، ولتحقيق هدف الدراسة تم بناء (9) اختبارات من نوع الاختيار من متعدد ذات أعداد مختلفة من البدائل (ثلاثة بدائل، أربعة بدائل، خمسة بدائل) طبقت على أفراد عينة الدراسة البالغ عددهم (41) طالبًا وطالبة جامعيين 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.

وأجرى نوادنجول ونابل (Nwadinigwe1 & Naibi1 , 2013) دراسة هدفت التعرف إلى أثر عدد البدائل في اختبار الاختيار من متعدد على خصائص الاختبار، تكونت عينة الدراسة من (159) طالباً وطالبة من الصف السادس الأساسي في نيجيريا، وتكونت أداة الدراسة من صورتين لاختبار اختيار من متعدد في الرياضيات الصورة الأولى احتوت خمسة بدائل للفقرة، والصورة الثانية ثلاثة بدائل للفقرة، وأظهرت نتائج الاختبار أنَّ عدد البدائل يؤثر بشكل دال إحصائياً على المتوسط الحسابي للدرجات وعلى مؤشرات الصعوبة والتمييز للفقرات، وبشكل عام دعمت النتائج استخدام اخبارات بثلاث بدائل للفقرة.

وأجرى بني عطا والرباعي (2013) دراسة هدفت الكشف عن أثر عدد بدائل الفقرة وموقع المموه القوي في فقرات اختبار اختيار من متعدد على الخصائص السيكومترية للاختبار، ومعالم الفقرات، تم بناء اختبار تحصيلي في الرياضيات لطلبة الصف العاشر مكوّن من (41) فقرة اشتمل الاختبار على أربعة نماذج حسب عدد البدائل وموقع المموه القوي، تكونت عينة الدراسة من (2111) طالبًا وطالبة لجميع نماذج الاختبار الأربعة، وفقًا للنموذج ثلاثي المعلمة، أظهرت نتائج الدراسة بعدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تُعزى لعدد بدائل الفقرة وموقع المموه القوي، وأظهرت نتائج الدراسة بأن دالة المعلومات للاختبار قد تباينت بتباين نماذج الاختبار، وأنَّ هنالك فروقاً دالة إحصائيًا بين قيم معاملات الثبات الإمبريقي وجاءت لصالح النموذج الذي يحوي خمسة بدائل.

وأجرت أبو مسلم (2016) دراسة هدفت إلى معرفة أثر كل من عدد البدائل وترتيبها لمفردات اختبار تحصيل دراسي من نوع الاختيار من متعدد على الخصائص السيكومترية للاختبار ومفرداته، تكونت عينة الدراسة من (113) طالباً وطالبة من معهد البحوث بجامعة القاهرة، تكونت أداة الدراسة من (50) فقرة من نوع الاختيار من متعدد وله ثلاث صور تختلف في عدد البدائل لكل فقرة، أظهرت نتائج الدراسة أن صدق وثبات الاختبار من نوع الاختيار من متعدد يزداد عندما يكون الاختبار ثلاثي البدائل عما لو كان أكثر من ذلك (رباعي أو خماسي البدائل)، ويزداد ثبات الاختبار عندما يتوزع البديل الصحيح عشوائياً بين البدائل الأخرى، وتكون فقرات الاختبار أكثر صعوبة كلما ازداد عدد بدائل الفقرة، كما تزداد القدرة التمييزية للفقرة كلما ازداد عدد البدائل للفقرة، وكلما كان توزيع البديل الصحيح عشوائياً بين البدائل الأخرى.

وأجرى زواهرة وكوفحي (2017) دراسة هدفت إلى استقصاء أثر تصفية المموهات في فقرات الاختبار من متعدد على معادلة الاختبار، وتم إعداد اختبار في الرياضيات من نوع الاختيار من متعدد بستة بدائل، وكونت أربعة نماذج من لكل منها أربعة بدائل للفقرة بعد حذف مموهين من بدائل فقرات الصورة النهائية للاختبار، وجرى تطبيق النماذج الاربعة على عينة تكونت من (880) طالباً من طلبة الصف السابع الأساسي، أظهرت النتائج أن فاعلية المعادلة وفق نموذج راش تقل عندما

لا تكون البيانات متطابقة مع النموذج، كما أظهرت النتائج التأكيد على فعالية معادلة الدرجات باستخدام نماذج تصفية المموهات.

تعقيب على الدراسات السابقة

اختلفت الدراسات السابقة فيما بينها من حيث المرحلة التي تم استهدافها، فبعضها استهدفت طلبة المرحلة الأساسية كدراسة بني عطا والرباعي (2013)؛ ودراسة الشريفين وطعامنة (2009)، في حين استهدف بعضها الآخر طلبة المرحلة الثانوية كدراسة ملياني (2010)؛ وهكذا يتضح أنَّ الدراسات لم تقتصر على مرحلة دراسية دون أخرى، كما اختلفت الدراسات السابقة في نتائجها وفي البرامج الحاسوبية والنماذج المستخدمة في تحليل فقرات الاختبار وتدريجها؛ إذ إنَّ معظم هذه الدراسات استخدمت النموذج أحادي المعلمة وثُنائي المعلمة كدراسة كل من: ملياني (2010)؛ وفرانسيسكو، جوليو، ودراسة أبو مسلم (2016)؛ تارنت وويز (Tarrant & Ware, 2010)؛ باغاي و أمراهي (Baghaei & Amrahi, 2011)؛ دراسة يامان (Yaman, 2011)، باستثناء دراسة كل من: واغنر وهارفي (Wagner & Harvey, 2003)؛ عيلبوني (2007)؛ دراسة بني عطا والرباعي (2013)، وتشابهت الدراسة الحالية مع دراسة نوادنجول ونابل (2013 , Nwadinigwe1 & Naibi1) من حيث استخدام نفس عدد البدائل للاختبارات إلا أنَّ الأخيرة استندت في تحليل البيانات على نظرية القياس التقليدية، بينما تستند الدراسة الحالية على نظرية الاستجابة للفقرة.

يُلاحظ عدم وجود اتفاق بين نتائج الدراسات والأبحاث التي أجريت في مجال الكشف عن أثر عدد بدائل الفقرة على هذه المتغيرات، وبالتالي جاءت هذه الدراسة لمعرفة أثر عدد بدائل الفقرة (خمسة بدائل؛ أربعة بدائل؛ ثلاثة بدائل) في اختبار الاختيار من متعدد على دالة معلومات الفقرة والاختبار وعلى الخطأ المعياري في التقدير؛ وذلك باستخدام أحد نماذج نظرية الاستجابة للفقرة وهو النموذج ثُلاثي المعلمة، وبالتالي تميّزت الدراسة الحالية عن غيرها من الدراسات السابقة من حيث مجتمعها وعينتها والتي تكونت من طلبة الصف العاشر الأساسي.

مشكلة الدراسة

تُعد الاختبارات التحصيلية أداة بالغة الأهمية في قياس وتقويم تحصيل الطلبة، ويُعد امتلاك المعلم لمبادئ بناء الاختبارات التحصيلية هدفاً بالغ الأهمية، فالمعلم مسؤول عن متابعة طلبته والإشراف عليهم وتقويم مسارهم، وإعداد الاختبارات وتطبيقها وتصحيحها في جميع مراحل التعليم المختلفة، وتتطلب اختبارات الاختيار من متعدد مهارات في إعدادها واستخدامها، 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 ذي درجة مرتفعة من الجاذبية والتجانس مع بقية البدائل، ويلاحظ أن هنالك ضعفاً في طريقة إعداد الاختبارات من قبل المعلمين بالإضافة لعدم قدرتهم على تحليل نتائج هذه الاختبارات وهذا جزء ذو أهمية كبيرة؛ فتحليل نتائج الاختبار يُمكّن المعلم من الوقوف على جوانب الضعف في العمليَّة التعليميَّة برمتها، والكثير من المعلمين يكثرون من استخدام الأسئلة ذات الإجابة المنتقاة في قياسهم لتحصيل الطلبة، الأمر الذي يستدعي إجراء المزيد من البحث في الأثر الذي يتركه عدد بدائل السؤال الموضوعي في خصائص الاختبار، وفي دالة المعلومات، من هنا جاءت هذه الدراسة بغرض الكشف عن أثر عدد البدائل على دالة المعلومات وعلى الثبات للاختبار وفقًا للنموذج ثُلاثي المعلمة.

أسئلة الدراسة

السؤال الأول: هل توجد فروق ذات دلالة إحصائية بين قيم معاملات الثبات المقدّرة باستخدام النموذج الثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدد تُعزى لاختلاف عدد بدائل الفقرة (خمسة بدائل؛ أربعة بدائل؛ ثلاثة بدائل)؟

السؤال الثاني: هل يوجد فروق ذات دلالة إحصائية بين متوسطات قيم دالة المعلومات للفقرة والاختبار المُقدَّرة باستخدام النموذج ثُلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (خمسة بدائل؛ أربعة بدائل؛ ثلاثة بدائل)؟

أهداف الدراسة

هدفت الدراسة إلى الكشف عن أثر اختلاف عدد البدائل في أسئلة الاختيار من متعدّد على دالة المعلومات للفقرة والاختبار والثبات؛ باستخدام النموذج الثلاثي المعلمة.

أهمية الدراسة

لهذه الدراسة أهمية نظرية وتطبيقية كما يلي:

الأهمية النظرية.

تتمثل الأهمية النظرية في توفير أدب نظري يتضمن بناء كيفية بناء اختبارات لها خصائص سيكومترية جيدة استنادًا إلى النموذج ثُلاثي المعلمة كواحد من نماذج نظرية الاستجابة للفقرة، كما قد تسهم هذه الدراسة في دعم الأطر النظرية للبحوث المتعلقة بنظرية الاستجابة للفقرة في من حيث انتقائها للفقرات المناسبة لبناء الاختبارات، وعدد بدائل الفقرة الملائمة، كما قد تُسهم نتائج هذه الدراسة في توجه الباحثين للقيام بالمزيد من الدراسات المشابهة.

الأهمية العملية.

تتمثل الأهميَّة العملية في هذه الدراسة بإعداد اختبار تحصيلي ذي خصائص جيدة يمكن أن يستخدم كأداة لقياس تحصيل الطلبة في مبحث الرياضيات للكشف عن جوانب القوة والضعف لديهم؛ كما قد تُسهم نتائج هذه الدراسة في مساعدة متخذي القرار وراسمي السياسات التربوية على تطوير مهارات تدريس وتقويم مبحث الرياضيات، وتحسين العملية التعليمية التعلّمية من حيث أهدافها ووسائلها وتقنياتها من خلال عقد الدورات المتخصصة في مجال بناء الاختبارات التحصيلية، وكيفية تحليل وتفسير النتائج، الأمر الذي سيقود إلى تطوير النظام التربوي والارتقاء بنوعية مخرجاته.

مصطلحات الدراسة

اختبار الاختيار من متعدّد.

اختبار يتكون من عدد من الفقرات، حيث تتكون الفقرة من جزأين، الجزء الأول يسمى المتن وقد يكون على شكل سؤال أو عبارة تتضمن قضية معينة تحتاج إلى تكميل إجابة، ويسمى الجزء الثاني بالبدائل والتي هي بمثابة حلول أو إجابات محتملة للسؤال أو القضية الواردة في المتن، وتتطلب إجابة صحيحة أو أفضل الإجابات، أو إجابة مركبة، أو تفسيرًا لبيانات ومعلومات معطاة" (Thorndike, 2010). ويعرف إجرائياً على أنه الاختبار الذي أعدّه الباحث في محتوى الفصل الدراسي الأول من كتاب الرياضيات المقرّر تدريسه لطلبة الصف العاشر في المدارس الحكومية في محافظة العاصمة عمان لقياس درجة اكتسابهم للمفاهيم الرياضية، والذي يتألف من ثلاثة نماذج دون تغيير في متون فقراتها.

الثبات.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

الحصـــول على نفس النتائج تقريبًا عندما يتم إعادة الاختبار نفســـه على المجموعة نفســـها وتحت نفس الظروف (Pathak, Patro, Pathak, & Valecha, 2013)، أما التعريف الإجرائي للثبات في ســياق الدراســة الحالية فهو قيمة معامل الثبات الإمبريقي المحســوب وفقًا للنموذج ثُلاثي المعلمة.

دالة المعلومات للفقرة.

هي اقتران رياضـــي يعبّر عن كمية المعلومات المتمثلة في تمييز الفقرة بين مســـتويات القدرة، ويتم التوصل إليها من بيانات التطبيق وفي هذه الدراسة تُعرّف بأنها القيمة التي يتم الحصول عليها من .تطبيق المعادلة الآتية: (Kim & Lee, 2017) Ii (θ) = ai2 Pi (θ) Qi (θ).

دالة المعلومات للاختبار.

هي اقتران رياضـــي يعبّر عن مجموع دوال المعلومات لفقرات الاختبار (Hambleton, Swamithan & Roger, 1991). وفي هذه الدراســة تُعرّف دالة معلومات الاختبار بأنها القيمة التي يتم الحصـــول عليها من تطبيق النماذج الثلاثة لاختبار الرياضـــيات الذي أعدّه الباحث لقياس تحصيل طلبة الصف العاشر الأساسي.

النموذج ثُلاثي المعلمة .

أحد نماذج نظرية الاســتجابة للفقرة، ويفترض أن الفقرات تختلف في صــعوبتها وتمييزها وميل المفحوصــين ذوي مســتويات القدرة المنخفضــة للتخمين لدى محاولتهم الإجابة عن الفقرة من نوع الاختيار من متعدّد، ويُرمز لمعامل التمييز بالرمز (a)، ومعامل الصـــعوبة بالرمز (b)، ومعامل التخمين بالرمز (c) في المعادلة الرياضـية التي تربط بين احتمال إجابة المفحوص ذي قدرة معيّنة إجابة صـــحيحة عن فقرة ما من جهة والمعالم الثلاثة للفقرة من الجهة الثانية (Embreston & Reise, 2000).

حدود الدراسة ومحدداتها

يتحدّد تعميم نتائج هذه الدراسـة بخصـائص مجتمع الدراسـة والمجتمعات المماثلة له، إذ تمّ جمع بيانات الدراسـة من خلال تطبيق اختبار تحصـيلي في مبحث الرياضـيات، على عيّنة عشـوائية عنقودية تمّ اختيارها من مجتمع طلبة الصـف العاشـر الأسـاسـي المسـجّلين في المدارس الحكومية التابعة لمديرية التربية والتعليم لواء الجامعة في محافظة العاصـمة عمان، خلال الفصـل الدراسـي الثاني من العام الدراسي(2018/2017).

كما يتحدّد تعميم نتائج هذه الدراسة بالخصائص السيكومترية لاختبار التحصيل المُستخدم في هذه الدراسة، وبالنموذج المُستخدم في تحليل البيانات وهو النموذج الثلاثي المعلمة، كما يتحدّد تعميم نتائج الدراسة بمدى جديّة ومصداقية استجابة الطلبة على أداة الدراسة.

الطريقة والإجراءات

مجتمع الدراسة

تكوّن مجتمع الدراسة من جميع طلبة الصف العاشر المسجّلين في المدارس الحكومية في مديرية التربية والتعليم لواء الجامعة خلال الفصل الثاني من العام الدراسي (2018/2017)، والبالغ عددهم (3840) طالبا وطالبة، وذلك حسب إحصائيات مديرية التربية والتعليم للواء الجامعة.

عينة الدراسة

تمّ اختيار عينة الدراسة بالطريقة العنقودية العشوائية، حيث تم تقسيم مجتمع الدراسة إلى أربع مناطق إدارية وتم اختيار المدارس في تلك المناطق بصورة عشوائية وصولاً إلى شعب الصف العاشر التي تم اختيارها كذلك بصورة عشوائية، وبلغ حجم عينة الدراسة (1530) طالبًا وطالبة يمثّلون ما نسبته (39.8%) من حجم مجتمع الدراسة.

أداة الدراسة

تم بناء ثلاثة نماذج لاختبار تحصيلي في الرياضيات من محتوى الجزء الثاني للصف العاشر الأساسي من نوع الاختيار من متعدّد تختلف في عدد البدائل فقط، حيث كان لفقرات النموذج الأول خمسة بدائل للإجابة، ولفقرات النموذج الثاني أربعة بدائل للإجابة، ولفقرات النموذج الثالث ثلاثة بدائل للإجابة. وتمت مراعاة الخطوات الآتية خلال بناء الاختبار التحصيلي(Popham, 2003):

أولاً: تحديد الهدف من الاختبار وهو قياس تحصيل طلبة الصف العاشر في مقرّر الرضيات في الفصل الدراسي الثاني (2018/2017) في المدارس الحكومية في مديرية التربية والتعليم للواء الجامعة، وتحليل محتوى الجزء الثاني من الكتاب، إذ أمكن اشتقاق (40) هدفًا تعليمياً، وتصميم جدول مواصفات الاختبار، وكتابة فقرات الاختبار لتغطي المحتوى التعليمي، وقد بلغ عدد فقرات الاختبار بالصورة الأولية (50) فقرة.

ثانياً: للتحقق من صدق المحتوى تم عرض الاختبار بصورته الأولية على مجموعة من المحكمين في الجامعات الأردنية، وتمّ إجراء التعديلات على صياغة بعض الفقرات في ضوء الاقتراحات التي قدمها المحكّمون، فقد تم تعديل الفقرة (4) التي نصت على "ميل المستقيم الموازي للمستقيم الذي يمر بالنقطتين (2 ،5) ، (4، 9) ، لتصبح بعد التعديل " ميل المستقيم ل$_1$ الموازي للمستقيم ل$_2$ الذي يمر بالنقطتين ..... " ومن الفقرات الأخرى التي تم تعديلها الفقرة(18) والتي نصت"إذا كانت هـ زاوية وكان جتاهـ=0.6 فإن ج(180) يساوي" لتصبح بعد التعديل " إذا كانت هـ زاوية حادة وكانت جتا هـ =(0.6) فإن جا(180+ هـ) يساوي"، كما تم حذف (12) فقرات ليصبح عدد فقرات الاختبار بصورته النهائية (38) فقرة .

ثالثاً: تطبيق الصورة النهائية من الاختبار على عينة استطلاعيَّة تكوَّنت من (30) طالبًا من خارج عينة الدراسة، وتحليل الفقرات من خلال حساب معاملات الصعوبة والتمييز لكل فقرة من فقرات الاختبار، وحساب فاعلية المموهات لفقرات الاختبار، وفي ضوء نتائج تحليل الفقرات تمّت المفاضلة بين الفقرات التي تشترك بقياس الهدف نفسه، وتم اختيار الفقرات التي تتمتع بأفضل معاملي صعوبة وتمييز، وبهذا تمَّ الاحتفاظ بـ (34) فقرة في الصورة النهائية من الاختبار، إذ بلغ متوسط صعوبة فقرات الاختبار بصورته النهائية (0.50)، ومتوسط معاملات تمييزها (0.43).

رابعاً: حساب فاعلية المموهات لكل فقرة من فقرات الاختبار، وذلك لاستخدامها كمعيار لحذف المموه الأضعف.

خامساً: تطبيق الصورة النهائية من الاختبار بنماذجه الثلاثة على أفراد عيِّنة الدراسة موزّعين على النماذج الثلاثة للاختبار.

سادساً: إعداد مفتاح الإجابة لكل نموذج من نماذج الاختبار.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

سابعاً: استبعاد إجابات الطلبة الذين أخفقوا في الإجابة عن جميع فقرات الاختبار؛ لأنَّ مستوى صعوبة الاختبار أعلى من مستوى قدرتهم، وكذلك استبعاد استجابات الطلبة الذين أجابوا إجابة صحيحة عن جميع الفقرات؛ لأن مستوى صعوبة الاختبار يُعدّ أدنى من مستوى قدراتهم، وبلغ عدد الطلبة الذين تمّ استبعادهم (30) طالباً.

المعالجات الإحصائية

التحقق من افتراض أحادية البعد من خلال:

أولاً: إجراءات التحليل العاملي باستخدام تحليل المكوّنات الأساسية، واستخرجت قيم الجذور الكامنة ونسبة التباين المُفسر لكل عامل من العوامل المستخلصة من التحليل، وتم استخدام تدوير المحاور باستخدام طريقة التدوير المتعامد للعوامل المستخلصة من تحليل المكوّنات الأساسية التي زادت قيم جذورها الكامنة عن الواحد الصحيح.

ثانياً: معامل كرونباخ الفا، والذي تُعدّ القيمة المرتفعة له مؤشرًا على تحقّق أحادية البعد.

ثالثاً: الثبات وتم استخراجه باستخدام برنامج (Bilog-MG).

رابعاً: تمّ استخدام تحليل التباين الأحادي للكشف عن دلالة الفروق بين المتوسطات الحسابية لقيم دالة المعلومات، كما تم التحقق من الافتراضات التي يتطلبها النموذج الثلاثي المّعلمة في نظرية الاستجابة للفقرة، وفيما يلي عرض لهذه الإجراءات.

أولاً- النتائج المتعلقة بالتحقق من افتراض أحادية البعد.

جرى التحقق من توافر افتراض أحادية البعد 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 (1).

جدول 1

معاملات ثبات الاتساق الداخلي لنماذج الاختبار باستخدام معادلة كرونباخ ألفا

النموذج	عدد الفقرات	عدد البدائل	معامل الثبات
الأول	38	خمسة بدائل	0.85
الثاني	38	أربعة بدائل بدائل	0.83
الثالث	38	ثلاثة بدائل	0.87

يُلاحظ من الجدول (1) أنّ قيم معاملات ثبات الاتساق الداخلي باستخدام معادلة كرونباخ ألفا تراوحت بين (0.83 و 0.87)، وهي معاملات مقبولة لأغراض التحقّق من افتراض أحادية البُعد (Dennick & Tavakol, 2011) .

التحقّق من أحادية البعد من خلال إجراء التحليل العاملي: 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، وتبين الجداول (2-5) نتائج التحليل العاملي لكل نموذج من النماذج المستخدمة.

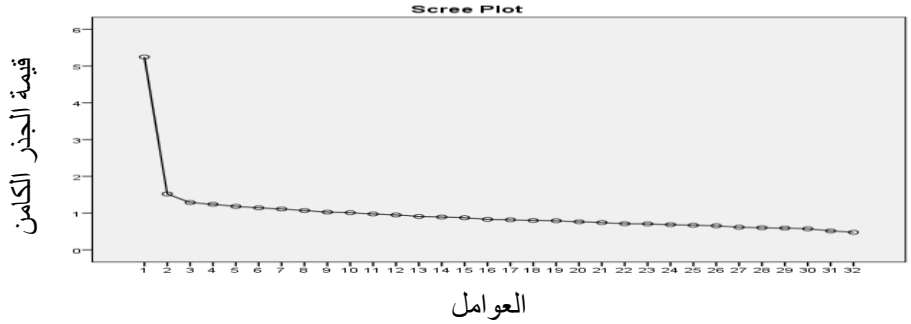
المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

جدول2

نتائج التحليل العاملي لاستجابات أفراد عينة الدراسة على فقرات النموذج ذي الخمسة بدائل وتدوير المحاور تدويرًا متعامدًا

العامل	الجذر الكامن	نسبة التباين المفسّر%	نسبة التباين التراكمي المفسّر %
1	3.30	13.20	13.20
2	1.30	5.20	18.40
3	1.20	4.90	23.30
4	1.20	4.80	28.10
5	1.10	4.50	32.60
6	1.10	4.30	36.90

يُلاحظ من الجدول (2) وجود ستة عوامل تزيد جذورها الكامنة عن الواحد الصحيح؛ إذ إنَّ العامل الأول فسّر ما نسبته (13.20 %) من تباين الأداء على الاختبار، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني زادات عن (2)، الأمر الذي يُعّد مؤشرًا على تحقق سمة أحادية البعد، والتمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لهذا النموذج من الاختبار يؤكد ذلك، كما هو موضّح في الشكل (1).



شكل 1. التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لنموذج الاختبار ذي الخمسة بدائل.

يُلاحظ من الشكل (1) أن الفرق بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالفرق بين العامل الثاني والعامل الثالث، ممّا يُشير إلى تحقّق سمة أحادية البعد في هذا الاختبار.

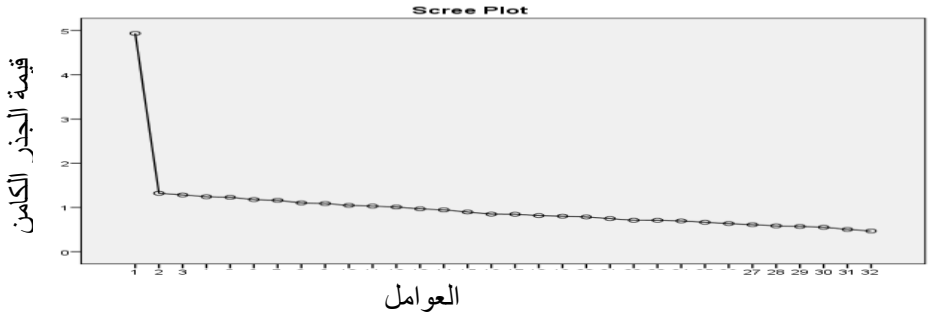
جدول3

نتائج التحليل العاملي لاستجابات أفراد عينة الدراسة على فقرات النموذج ذي الأربعة بدائل

العامل	الجذر الكامن	نسبة التباين المفسّر%	نسبة التباين التراكمي المفسّر%
1	4.94	15.43	15.43
2	1.32	4.13	19.55
3	1.28	4.01	23.56
4	1.24	3.89	27.45
5	1.23	3.84	31.29
6	1.18	3.68	34.98

7	1.16	3.63	38.61
8	1.10	3.45	42.06

يُلاحظ من الجدول (3) وجود ثمانية عوامل تزيد قيم جذورها الكامنة عن الواحد الصحيح، وأن العامل الأول فسر ما نسبته (15.43%) من تباين الأداء على نموذج الاختبار ذي البدائل الأربعة، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (3.736)، وهي تزيد عن (2)، وهذا يُعُد مؤشرًا على تحقق افتراض أحادية البعد للبيانات في هذا الاختبار، والتمثيل البياني لقيم الجذورالكامنة للعوامل الناتجة عن التحليل العاملي لهذا النموذج من الاختبار يؤكد ذلك، كما هو موضّح في الشكل (2).



شكل 2. التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لنموذج الاختبار ذي الأربعة بدائل

يُلاحظ من الشكل (2) أن مقدار الفرق بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالفرق بين العامل الثاني والعامل الثالث، وبقاء الفروق بعد ذلك متقاربة جدًا بين العوامل المتبقية، ممّا يُشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على الاختبار.

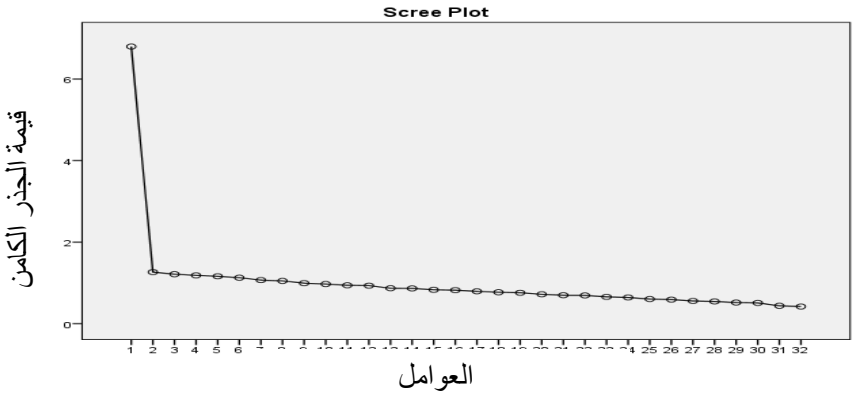
جدول 4

نتائج التحليل العاملي لاستجابات أفراد عينة الدراسة على فقرات النموذج ذي الثلاثة بدائل وتدوير المحاور تدويرًا متعامدًا

العامل	الجذر الكامن	نسبة التباين المفسّر%	نسبة التباين التراكمي المفسّر%
1	6.80	21.24	21.24
2	1.27	3.95	25.19
3	1.21	3.79	28.98
4	1.19	3.70	32.69
5	1.16	3.63	36.32
6	1.13	3.52	39.84
7	1.07	3.34	43.45
8	1.05	3.28	46.45

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

يُلاحظ من الجدول (4) وجود ثمانية عوامل تزيد قيم جذورها الكامنة عن الواحد الصحيح، وأن العامل الأول فسّر ما نسبته (21.24%) من تباين الأداء على هذا النموذج من الاختبار، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (5.37) وهذا يُعُد 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 (3).



شكل 3. التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي لنموذج الاختبار ذي الثلاثة بدائل

يُلاحظ من الشكل (3) أن مقدار الفرق بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالفرق بين العامل الثاني والعامل الثالث، ممَّا يشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على نموذج الاختبار ذي الثلاثة بدائل.

أما بالنسبة لافتراض الاستقلال الموضعي فإنَّ تحقق افتراض أحادية البعد في البيانات، يعني تحقق افتراض الاستقلال الموضعي كما أشار لذلك هامبلتون وسوامنثان (Hamblton & Swaminathan, 1985).

ثانيًا- النتائج المتعلقة بمطابقة الأفراد والفقرات للنموذج ثلاثي المعلمة.

تم استخدام برنامج (BILOG-MG3) ومن خلال مؤشر مربع كاي ($\chi 2$) عند مستوى الدلالة (α= 0.01) حيث أظهرت النتائج مطابقة الأفراد في كل نموذج، أما بالنسبة لمطابقة الفقرات في كل اختبار للنموذج الثلاثي المعلمة بناءً على مؤشر مربع كاي ($\chi 2$) عند مستوى الدلالة (α= 0.01) حيث تُعد الفقرة غير مطابقة للنموذج الثلاثي المعلمة إذا كانت الاحتمالية لهذه الفقرة أقل من (0.01)، والجدول (5) يُبين ذلك:

جدول 5

نماذج الاختبار والفقرات غير المطابقة للنموذج الثلاثي المعلمة المستخدم في تحليل البيانات

العدد الأصلي للفقرات	نموذج الاختبار	الفقرات غير المطابقة	عدد فقرات الاختبار بعد حذف الفقرات غير المطابقة
38	خمسة بدائل	-	38
38	أربعة بدائل	5	33
38	ثلاثة بدائل	6	32

نتائج الدراسة

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

للإجابة عن هذا السؤال تم إيجاد معاملات الثبات باستخدام برنامج (Bilog-MG) لكل نموذج، والجدول (6) يبين ذلك.

جدول 6

معاملات الثبات لكل نموذج من نماذج الاختبار المستخدمة

نموذج الاختبار	معامل الثبات
خمسة بدائل	0.85
أربعة بدائل	0.75
ثلاثة بدائل	0.60

يُلاحظ من الجدول(6) وجود فروق ظاهرية بين معاملات الثبات المستخرجة لنماذج الاختبار المختلفة، حيث تبيّن أن نموذج الاختبار ذو الخمسة بدائل هو الأكثر ثباتًا، يليه نموذج الاختبار ذي الأربعة بدائل، 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 (7).

جدول 7

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

نموذج الاختبار	معامل الثبات	المقارنة	قيمة "ز"	مستوى الدلالة
خمسة بدائل	0.85	4*5	12.16	0.004
أربعة بدائل	0.75	3*5	7.30	0.040
ثلاثة بدائل	0.60	3*4	-4.85	0.023

يُلاحظ من الجدول (7) أن الفروق بين معاملات ثبات كل زوج من أزواج النماذج الثلاثة للاختبار هي فروق ذات دلالة إحصائية بمستوى يقل عن (α=0.05)، ويُلاحظ أيضاً أن نموذج الاختبار ذي الخمسة بدائل هو الأكثر ثباتًا مقارنةً بالنموذجين الأخريين، يليه النموذج ذو الأربعة بدائل.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019

نتائج الإجابة عن السؤال الثاني: هل توجد فروق ذات دلالة إحصائية ($\alpha = 0.05$) بين متوسطات قيم دالة المعلومات للفقرة والاختبار باستخدام النموذج اللوجستس ثلاثي المعلمة في ثلاثة نماذج لاختبار تحصيلي في مادة الرياضيات من نوع الاختيار من متعدّد تُعزى لاختلاف عدد بدائل الفقرة (خمسة بدائل؛ أربعة بدائل؛ ثلاثة بدائل)؟

للإجابة عن هذا السؤال تم استخراج دالة المعلومات لكل فقرة من فقرات كل نموذج من نماذج الاختبار، وتمّ حساب متوسط دوال معلومات الفقرات لكل نموذج لتعبّر عن دالة معلومات الاختبار كما يوضّحها الجدول (8).

جدول 8

قيم دالة المعلومات لكل فقرة من فقرات كل نموذج من نماذج الاختبار

نموذج الخمسة بدائل				نموذج الأربعة بدائل				نموذج الثلاثة بدائل			
الفقرة	دالة المعلومات	الفقرة	دالة المعلومات	الفقرة	دالة المعلومات	الفقرة	دالة المعلومات	الفقرة	دالة المعلومات	الفقرة	دالة المعلومات
1	0.59	21	0.57	1	0.62	21	0.54	1	0.40	21	0.51
2	0.68	22	0.66	2	.451	22	0.98	2	0.40	*22	محذوفة
3	.581	23	0.50	*3	محذوفة	*23	محذوفة	3	0.40	23	0.45
4	0.46	24	0.53	4	0.34	24	0.54	*4	محذوفة	24	1.87
5	0.51	25	0.57	5	1.45	25	0.98	5	1.45	25	1.01
6	0.79	26	0.44	6	0.45	*26	محذوفة	6	0.40	26	0.41
7	0.64	27	0.73	7	0.62	27	0.70	7	0.41	27	0.49
8	0.49	28	0.52	8	0.43	28	0.69	8	0.40	28	0.42
9	0.56	29	0.73	9	0.62	29	0.54	9	0.40	29	0.44
10	0.50	30	0.54	10	0.23	30	0.62	10	0.64	*30	محذوفة
11	0.60	31	0.70	11	1.34	31	0.44	11	0.60	31	1.45
12	0.65	32	0.58	12	0.62	32	0.34	*12	محذوفة	32	0.43
13	0.49	33	0.51	*13	محذوفة	33	0.54	13	0.43	33	محذوفة
14	0.69	34	0.79	14	0.43	34	0.98	14	0.99	34	محذوفة
15	0.49	35	0.50	15	0.23	35	0.50	15	1.49	*35	محذوفة
16	0.61	36	0.60	*16	محذوفة	*36	محذوفة	*16	محذوفة	36	0.51
17	0.58	37	0.65	17	0.54	37	0.60	17	0.49	37	0.45
18	0.46	38	0.51	18	0.62	38	0.52	18	0.48	38	0.39
19	0.58			19	0.43			19	0.45		
20	0.59			20	0.62			20	0.42		
المتوسط الحسابي			24.17	المتوسط الحسابي			20.55	المتوسط الحسابي			19.08

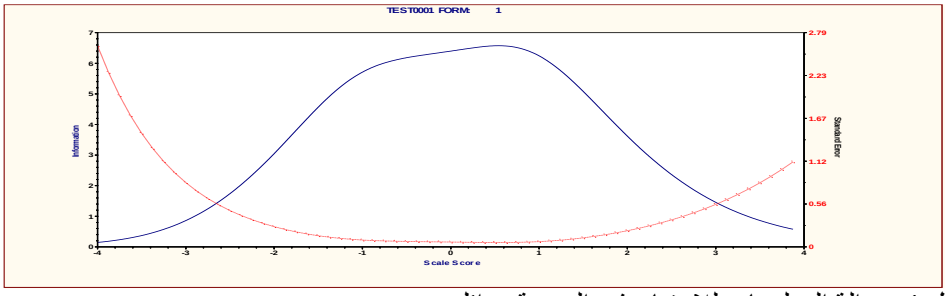
يُلاحظ من الجدول (8) وجود فروق ظاهرية بين المتوسطات الحسابية لقيم دالة معلومات الاختبار باختلاف نموذج الاختبار المستخدم، حيث تبيّن أن نموذج الاختبار ذي الخمسة بدائل يُعطي أكبر معلومات، يليه نموذج الاختبار ذي الأربعة بدائل، ثم الاختبار ذي الثلاثة بدائل، وللكشف عن دلالة الفروق بين هذه المتوسطات تبعًا لنموذج الاختبار، تمّ استخدام أسلوب تحليل التباين الأحادي كما يوضّحها الجدول (9).

جدول 9

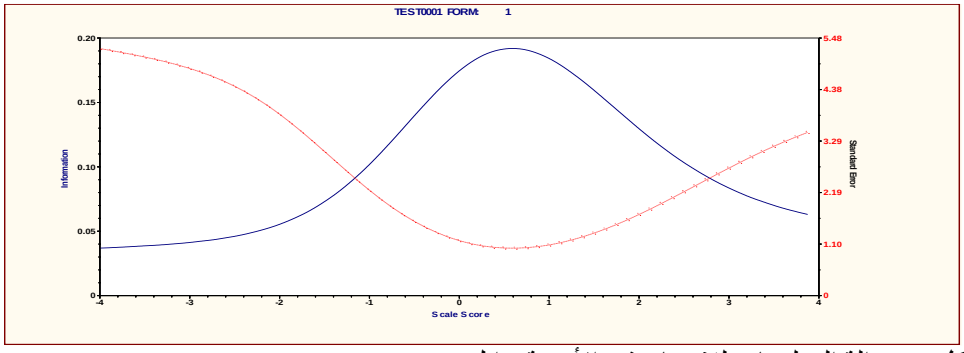
نتائج تحليل التباين الأحادي لفحص دلالة الفروق بين المتوسطات الحسابية لتقدير دالة معلومات الاختبار تبعًا لنموذج الاختبار المستخدم (خمسة بدائل، أربعة بدائل، ثلاثة بدائل)

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "ف"	مستوى الدلالة
نموذج الاختبار	0.01	2	0.01	0.05	0.95
الخطأ	10.04	100	0.10		
الكلي	10.05	102			

يُلاحظ من الجدول (9) عدم وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (α = 0.05) بين المتوسطات الحسابية لتقدير دالة المعلومات يُعزى لمتغير نموذج الاختبار (خمسة بدائل؛ وأربعة بدائل؛ وثلاثة بدائل) حيث بلغت قيمة "ف" المحسوبة (0.05)، وبمستوى دلالة مقداره (0.95)، وهذا يعني عدم اختلاف المتوسطات الحسابية لدالة معلومات الاختبار باختلاف عدد بدائل الفقرة، وتوضّح الأشكال (4، 5، 6) التمثيلات البيانية التي تعبّر عن دالة معلومات الاختبار لكل نموذج.

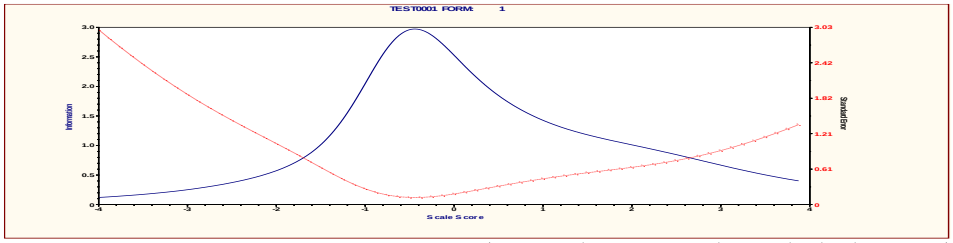


شكل 4. دالة المعلومات للاختبار ذي الخمسة بدائل



شكل 5. دالة المعلومات للاختبار ذي الأربعة بدائل

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد (43) العدد (3) أكتوبر 2019



شكل 6. دالة المعلومات للاختبار ذي الثلاثة بدائل

مناقشة النتائج والتوصيات

أولا-مناقشة النتائج المتعلقة بالإجابة عن سؤال الدراسة الأول

أظهرت نتائج الإجابة عن هذا السؤال أنَّ النموذج الأكثر ثباتًا هو النموذج ذو الخمسة بدائل، يليه النموذج ذو الثلاثة بدائل، وقد تُعزى هذه النتيجة إلى أنَّ عدد بدائل الفقرة يؤثر إيجابًا على ثبات الاختبار؛ إذ تقل فرصة تخمين الإجابة الصحيحية بزيادة عدد البدائل، فالتخمين يؤثر سلباً على ثبات الاختبار، فالتخمين العشوائي يعُدّ من المُهددات لكل من صدق وثبات اختبارات الاختيار من متعدد حيث يرى ميسك (Messick, 1995) أن لجوء الطلبة لعملية التخمين خلال إجابتهم عن فقرات اختبار الاختيار من متعدد يُضيف تبايناً لا يرتبط بالتباين في القدرة الحقيقية، واختلفت نتيجة الإجابة عن هذا السؤال عن نتائج دراسة الشريفين وطعامنة (2009) التي أشارت إلى عدم وجود فروق بين معاملات الثبات تُعزى لعدد البدائل، كما اختلفت النتائج عن دراسة باغاي و أمراهي (Baghaei & Amrahi, 2011) والتي أشارت إلى عدم وجود أثر يُعزى لنموذج الاختبار المستخدم 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 (2013)، والتي أشارت إلى وجود فروق دالة إحصائيًا بين قيم معاملات الثبات في النماذج المختلفة حيث جاءت لصالح النموذجي الخمسة بدائل، كما اتفقت نتائج الإجابة عن هذا السؤال مع نتائج دراسة نوادنجول ونابل (Nwadinigwe1 & Naibil, 2013) والتي أشارت إلى أفضلية لنموذج الاختبار ذي الخمسة بدائل.

ثانيا- مناقشة النتائج المتعلقة بالإجابة عن سؤال الدراسة الثاني

أظهرت نتائج الإجابة عن هذا السؤال عدم وجود فرق ذي دلالة إحصائية عند مستوى الدلالة (α= 0.05) بين المتوسطات الحسابية لدالة المعلومات يُعزى لمتغير عدد بدائل الفقرة، إذ بلغت قيمة "ف" المحسوبة (0.05) وبمستوى دلالة مقداره (0.95)، وهذه القيمة أكبر من مستوى الدلالة (0.05). وقد يُعزى ذلك إلى جديّة الطلبة والتزامهم بالإجابة عن فقرات الاختبار بدقة وعدم لجوئهم للتخمين العشوائي، كما أن دالة المعلومات تعُد مؤشراً على ثبات الاختبار كونها تتناسب عكسياً مع خطأ التقدير الذي يزداد ثبات الاختبار بنقصانه، واختلفت نتائج الإجابة عن هذا السؤال مع نتائج دراسة بني عطا والرباعي (2013) والتي أشارت إلى وجود فروق ذات دلالة إحصائية بين قيم متوسطات دالة المعلومات لنماذج الاختبار ولصالح نموذج الاختبار ذي الخمسة بدائل للفقرة، كما اختلفت نسبياً مع نتائج دراسة نوادنجول ونابل (Nwadinigwe1 & Naibi1 , 2013) والتي أشارت إلى أفضلية

للنموذج ذي الخمسة بدائل للفقرة، كما اختلفت مع نتائج دراسة أبو مسلم (2016) والتي أشارت إلى أن نموذج الاختبار ذي الثلاثة بدائل كان أفضل من النموذج ذي الخمسة أو الأربعة بدائل.

التوصيات

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

أولاً: التوصيات التربوية

1. 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 ذوي مستوى القدرة المرتفع فيفضل استخدام الاختبار ذي الخمسة بدائل كونه يعطي معلومات أكثر عنهم.
2. الاهتمام ببناء الاختبارات التربوية وفقاً لخطوات سليمة من تحليل محتوى وبناء جدول مواصفات، وإعداد مموهات ذات خصائص مناسبة.

ثانياً: التوصيات البحثية

1. إجراء المزيد من الدراسات التي تتناول أشكالاً أخرى من الاختبارات الموضوعية.
2. إجراء دراسات لتحري أثر شكل وطبيعة متن فقرة الاختبار من متعدد على دقة تقديرات معالالفقرات (الصعوبة، والتمييز، والتخمين).
3. إجراء دراسات مماثلة، بعد تقسيم أفراد عينة الدراسة لثلاثة مستويات للقدرة لكل نموذج (مرتفعة، ومتوسطة، ومنخفضة) ودراسة الخصائص السيكومترية التي بحثتها الدراسة الحالية عند كل مستوى.

المصادر والمراجع

اولاً: المراجع العربية

أبو مسلم، مايسة.(2016). أثر كل من عدد البدائل وترتيبها لمفردات اختبار تحصيل دراسي من نوع الاختيار من متعدد على الخصائص السيكومترية للاختبار ومفرداته، **مجلة اتحاد الجامعات العربية للتربية وعلم النفس**، 14(2)، 161-200.

بني عطا، زايد والرباعي، إبراهيم. (2013). أثر عدد البدائل وتغيير موقع المموه القوي في فقرات اختبار الاختيار من متعدد على معالم الفقرات وقدرة الفرد ودالة المعلومات. **المجلة في العلوم التربوية، 9**(3)،319-333.

زواهرة، ريما، كوفحي، موسى.(2017). لمراجع الأجنبية و كوفحي، موسى.(2017). أثر طريقة تصفية المموهات في فقرات الاختيار من متعدد على معادلة الاختبار. **دراسات العلوم التربوية**، 44(2)، 347-362.

الشريفين، نضال وطعامنة، إيمان. (2009). أثر عدد البدائل في اختبار الاختيار من متعدد في تقديرات القدرة للأفراد والخصائص السيكومترية للفقرات والاختبار وفق نموذج راش في نظرية الاستجابة. **المجلة في العلوم التربوية، 5**(4)، 309-335..

عيلبوني، سمير. (2007). **أثر عدد البدائل وتمييز المموهات في فقرات الاختيار من متعدد على التوافق مع نموذج ثلاثي المعلم**. أطروحة دكتوراه (غير منشورة)، جامعة اليرموك، إربد.

الغامدي، عبدالله بن أحمد. (2008). **أثر عدد البدائل في الخصائص السيكومترية للاختبار التحصيلي في الرياضيات.** رسالة ماجستير (غير منشورة)، جامعة أم القرى، المملكة .

مليايي، زياد. (2010). **مقارنة أثر اختلاف عدد البدائل في اختبار الاختيار من متعدد على صعوبة الفقرة وقدرة الفرد في ضوء كلٍ من النظرية الكلاسيكية ونموذج راش**. أطروحة دكتوراه (غير منشورة)، جامعة أم القرى، مكة المكرمة.

ثانياً: المراجع الأجنبية

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.

De Ayala, R. J. (2010). *The theory and practice of item response theory*. Guilford Press, New York, USA.

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.

Embreston, Susan, and Reise, Steven (2000). *Item reesponse theory for psychologis***t**. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ha, D. (2017). Applying multidimensional three-parameter logistic model (m3pl) in validating a multiple-choice test. *International Journal of Scientific and Research Publications*, *7* (2), 2250-3153.

Hambleton R.K. Swaminthan.H, and Rogers.H.J.(1991). Fundamentals of item response theory**.** *International Educational and professional*. Sage Publication, Inc Newbury Park, USA.

Kim, K. Y., & Lee, W. C. (2017). The impact of three factors on the recovery of item parameters for the three-parameter logistic model. *Applied Measurement in Education*, *30* (3), 228-242.

Linden, W. J., and Hambleton, R. K. (1997). *Handbook of modern item response theory***.** Springer-Verlage New York Inc., USA.

Messick, S. (1995). Validation of psychological assessment: validation of inferences from person's response and performances as scientific inquiry into scoring meanings. *American Psychologist, 50*,741-749.

Nwadinigwel, P. & Naibi1, L. (2013)**.** The number of options in a multiple-choice test item and the psychometric characteristics**.** *Journal of Education and Practice*, *4*(28), 22-34.

Onn, D. (2013). Classical test theory versus item response theory: An evaluation of the comparability of item analysis results. *Joint Admissions and Matriculation Board*, 1-23.

Partchev. I. (2004). *A visual guide to item response theory*. Friedrich-Schiller-Universitat¨ Jena.

Pathak, A., Patro, K., Pathak, M., & Valecha, M. (2013). Item response theory. *International Journal of Computer Science and Mobile Computing*, *2* (11), 7.

Popham, James (2003). *Educational assessment***.** Pearson Education Inc., USA.

Sahin, A and Anil, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory and Practice, 17* (1),321-335

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.

Thorndike, R. M., (2010). *Measurement and evaluation in psychology and education* (8th Ed). Upper Saddle River, NJ: Pearson/Merril Prentice Hall.

Wagner, T., & Harvey, R. (2003). Developing a new critical thinking test using item response theory. *Paper presented at the 2003 annual conference of the society for industrial and organizational psychology*. Retrieved June 11, 2016, from http://harvey.psyc.vt.edu/.

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.