

أثر استراتيجية اختيار المموهات لفقرات اختبارات الاختيار من متعدد على دالة معلومات الاختبار

The Effect of Distracters' Selection Method of Multiple Choice Tests Items on Information Function of the Test

نايل الكعابنه

Nayel Alkaabnh

قسم علم النفس التربوي، كلية التربية، جامعة شقراء، السعودية

بريد الكتروني: nayel_odeh1978@yahoo.com

تاريخ التسليم: (2015/5/3)، تاريخ القبول: (2016/5/12)

ملخص

هدفت هذه الدراسة إلى معرفة أثر استراتيجية اختيار المموهات لفقرات اختبارات الاختيار من متعدد على دالة معلومات الاختبار، ولأغراض هذه الدراسة تم إعداد اختبار من نوع الاختيار من متعدد لمقرر الرياضيات لطلبة الصف التاسع الأساسي مكون من (41) فقرة، وموزع على ثلاثة نماذج (A, B, C)، حيث تم بناء تلك النماذج بثلاثة استراتيجيات مختلفة (النموذج A تم بناء مموهات فقراته من خلال المعلمين، والنموذج B تم بناء مموهات فقراته من خلال التحليل الإحصائي، والنموذج C تم بناء مموهات فقراته من خلال اجابات الطلبة)، وتم تطبيق نماذج الاختبار على عينة مكونة من (1473) طالباً، واستخدم برنامج (SPSS) وبرنامج (BILOG– MG) لتحليل بيانات الدراسة. وأظهرت النتائج أن مقدار دالة معلومات الاختبار لنماذج الاختبار (A, B, C) كانت مختلفة وكان أفضلها في النموذج (B)، يليه النموذج (C) ثم النموذج (A)، وكان هناك فرق قليل بين قيمة دالة المعلومات للنموذجين (B,C)، ولكن عند مستويات مختلفة من القدرة، وكذلك أظهرت النتائج أن أكبر قيمة لدالة معلومات الاختبار كانت عندما تكون قدرات المفحوصين متوسطة.

الكلمات المفتاحية: دالة معلومات الاختبار، استراتيجية اختيار المموهات، اختبارات الاختيار من متعدد.

Abstract

The objective of the study was to investigate the effect of distracters' selection method of multiple choice test items on Information Function.

To achieve the objective of the study, the researcher prepared an achievement test for 9th graders consisting in the final format of 41 items. Three different forms of items distracters (A, B, C) were used. The forms were constructed according to three different strategies: form (A) was constructed based on the information taken from teachers, while the items distracters of the second form (B) were constructed by using the statistical analysis and the distracters of form (C) were constructed according to students' answers. The various forms of the test were applied on sample of 1473 students. The results of the study were analyzed by using SPSS, and BILOG- MG programs. The findings showed that the information function of the test in all the forms were different, and the best information function was in the form (B), then form (C), then form (A), and there was a small difference between the information function for the (A, C) forms, but at different levels of the ability. Moreover, the results revealed that the highest value of the information function was with average students' abilities.

Keywords: Information Function, Multiple choice tests, Distractors' selection method

المقدمة

تُعد الاختبارات من أهم الأدوات التي تزودنا بالبيانات التي يُستند إليها في اتخاذ القرارات في جميع المؤسسات التربوية وغير التربوية، وانتشر استخدام الاختبارات وخاصةً اختبارات الاختيار من متعدد في الآونة الأخيرة بشكل كبير، وزاد من ذلك التطور الكبير في الاستراتيجيات المستخدمة في بنائها، وتحليل الفقرات ونتائج الاختبار وفق النظريتين (التقليدية، والحديثة) في القياس (عوده، 2004).

إلا أن هذه الأدبيات المتوفرة في القياس لم تشر إلى استراتيجية مثلى لاختيار مموهات فقرات الاختيار من متعدد، فقد يتم الاختيار وفق أسس منطقية أو وفق أسس إحصائية.

ويعتمد الكثير من مطوري الاختبارات على الأسس المنطقية لخبرتهم في المادة التعليمية أو خبرتهم في المفحوصين، ولم يكن هناك دراسات أشارت إلى اختيار مموهات فقرات اختبارات الاختيار من متعدد من خلال إجابات المفحوصين أنفسهم، وفحص هذا الاختيار ومقارنته بطرق أخرى لاختيار هذه المموهات من خلال خصائص الفقرات والاختبار في بحوث تجريبية ومواقف اختبارية واقعية.

فعدت تطوير اختبار من نوع الاختيار من متعدد يجب الأخذ بعين الاعتبار عدة أمور تؤثر في الخصائص السيكومترية للاختبار، وفي الخصائص السيكومترية للفقرة (الصعوبة، والتمييز).

ومن هذه الأمور، عدد المموهات، حيث أكدت معظم الدراسات مثل: دراسة ستراتون وكاتس (Straton & Catts, 1980) ودراسة بدسكو ونيفو (Budescu & Nevo, 1985) التي تناولت هذا الموضوع بشكل جيد حيث وضحت هذه الدراسة بأن اختلاف عدد المموهات يؤثر على الخصائص السيكومترية للاختبار ولفقراته، وفي هذا الصدد يرى إبل وفريسيبي (Eble & Frisbie, 1986) أن اختلاف عدد المموهات في اختبارات الاختيار من متعدد يؤدي إلى اختلاف عدد المقارنات للإجابة المعطاة، حيث يؤثر ذلك على صعوبة الفقرة ومعامل تمييزها، وباختلاف القدرة التمييزية لها يختلف ثبات أداء المفحوصين على الفقرة وعلى الاختبار ككل، مما يؤثر بالتالي على معامل ثبات الاختبار، فكلما زادت الدالة التمييزية لل فقرات زاد ثبات الاختبار، لأن انخفاض القدرة التمييزية للفقرة يجعل النجاح على الفقرة مرتبطاً بالصدفة، وبذلك فهي لا تقيس القدرة الحقيقية للطلبة، وكذلك مدى صعوبة الفقرة التي تؤثر على ثبات الاختبار، وذلك لأنه عندما تكون الفقرات متوسطة الصعوبة يكون الثبات أعلى منه في حال الفقرات السهلة جداً أو الصعبة جداً.

كما أن زيادة عدد مموهات فقرات الاختبار قد تعمل على تشويش الطلبة وتزيد من أخطائهم مما يستدعي تفضيل الاختبار الذي يحوي على عدد أكبر من الفقرات وبعده مموهات أقل لتلك الفقرات، وهذا يتفق مع نتائج دراسة جريير (Grier, 1975)، والتي بينت أن معامل ثبات الاختبار يزداد كلما قل عدد مموهات الفقرات، وبخلاف نتائج دراسة بدسكو ونيفو (Budescu & Nevo, 1985)، التي بينت أن معامل ثبات الاختبار يزداد بزيادة عدد المموهات للفقرات، وأما بعض الدراسات مثل دراسة كوستين (Costin, 1972) فقد بينت أن عدد المموهات لا يؤثر على معامل ثبات الاختبار، ولتجنب تأثير عدد المموهات على معامل ثبات الاختبار فلا بد أن تكون المموهات متجانسة ومبنية وفق استراتيجية جيدة لنحصل على اختبار قادر على قياس قدرات المفحوصين بدقة وموضوعية، ويزود واضعيه بمعلومات موثوقة لعملية اتخاذ القرارات التدريسية، ويكون له دور ايجابي في عملية التعليم والتعلم.

وأما الأمر الثاني فهو شكل المموهات، عندما تكون تقليدية، والتي تتلخص في وجود إجابة واحدة صحيحة وعدد من المموهات، أو مركبة؛ كأن تحتوي على مموهات على شكل "لا شيء مما ذكر" أو "جميع ما ذكر صحيح" أو الجمع بين المموهات مثل "أ+ب"، وهذا ما نلاحظه في كثير من الاختبارات التحصيلية التي يعدها المعلمون، أو قد يتم استخدام المموهات المركبة لفقرات الاختبار إما لتوفير الوقت والجهد في بناء مموهات جيدة، أو عدم توفر الخبرة لدى مطوري الاختبارات بالمحتوى الدراسي أو بالمفحوصين، أو عدم وجود استراتيجية معينة يتم من خلالها بناء مموهات جذابة ومناسبة لفقرات اختبارات الاختيار من متعدد، ومن ناحية أخرى يرى جرونلاند (Gronlund, 1982) أن استخدام المموه "جميع ما ذكر صحيح" يزيد من احتمالية تخمين الإجابة من بين المموهات، بحيث إذا عرف المفحوص أن مموهين من مموهات

الفقرة صحيحان فإنه سيختار المموه "جميع ما ذكر صحيح" بغض النظر عن عدد المموهات المتبقية للفقرة، وإذا عرف المفحوص أن أحد المموهات على الأقل خاطئ فإنه سيستبعد المموه "جميع ما ذكر صحيح"، ويضيف عودة (2004) إن بعض المعلمين يلجأ إلى وضع المموه "لا شيء مما ذكر" وخاصةً في الأسئلة الحسابية، وذلك تحسباً من عدم وجود إجابة صحيحة بين المموهات دون أن يقصد المعلم ذلك، وفي هذه الحالة يبقى المعلم في حلٍ من أي اعتراض حول عدم وجود إجابة صحيحة لأن المموه "لا شيء مما ذكر" هو المموه الهروبي.

وأما الأمر الثالث فهي الطريقة التي يتم من خلالها اختيار المموهات لفقرات اختبار الاختيار من متعدد، فمن الإرشادات الواردة في أدبيات القياس أن تكون المموهات متجانسة مع بعضها البعض، مما يزيد من قدرة المموهات على جذب انتباه المفحوصين، وكذلك فإن تأثير جاذبية المموهات على خصائص الفقرة عامل من العوامل المحتملة، ومثالاً على ذلك: خصائص الطلاب، وطبيعة المادة، وإن أي إجراء لاختيار المموهات قد يكون له دور في تحديد خصائص الفقرات والاختبار.

وكذلك لا بد من أن يكون متن الفقرة مُصاغاً بشكل جيد، فعندما تكون الفقرة غامضة وضعيفة لغوياً فإن ذلك يؤدي إلى إرباك الطلبة ويزيد من احتمال أن يستجيبوا للفقرة بشكل خاطئ حتى وإن كانوا يمتلكون المعرفة اللازمة لإجابة الفقرة بشكل صحيح، وفي هذا الصدد يرى جرونلاند ولين (Gronland & Linn, 1990)، أن مثل هذا الغموض يؤدي إلى فهم مختلف من مفحوص لآخر لمضمون الفقرة.

إن هذه الإرشادات - التي يتبعها المتخصصون والمطبقون للاختبارات في المدارس والجامعات والمؤسسات الأخرى- يمكن أن ينعكس تطبيقها على خصائص الاختبار كأداة قياس، مما يعني أن مدى الاهتمام بتطوير اختبار من نوع الاختيار من متعدد يعكس مدى الاهتمام بتطبيق القياس العلمي في المجال التربوي والنفسي. وساعد في هذا الاهتمام ظهور النظرية الحديثة في القياس التي بدورها حلت بعض جوانب القصور التي كانت تواجه النظرية الكلاسيكية.

حيث عملت النظرية الحديثة في القياس على تلافي هذا القصور، فعند استخدام نماذج هذه النظرية بوجود مجتمع كبير من الأفراد، يكون تقدير قدرات الأفراد متحررة من الفقرات، ومن هنا أعطي أهمية اللاتغير (Invariance) في تقدير معالم الفقرات بين مختلف مجموعات الأفراد، كما توفر النظرية الحديثة مؤشراً إحصائياً لتقدير درجة الدقة في قياس القدرة لكل فرد (Hambleton & Swaminathan, 1985)

ونظراً لاختلاف أشكال منحنيات خصائص الفقرة، ونتيجة لاستخدام هذه النظرية بشكل واسع، فقد تم تطوير نماذج مختلفة تصف العلاقة بين أداء الأفراد على الاختبار والسمات الكامنة وراء هذا الأداء باقتران رياضي لوغاريتمي. ومن هذه النماذج النموذج ثلاثي المعلمة، حيث يعتبر امتداد للنموذج الثنائي، فهو يضيف معلماً جديداً للفقرة يسمى معلم التخمين (C_i)، فهذا

المعلم يحدد احتمالاً بأن يجيب المفحوص إجابة صحيحة عن الفقرة بالتخمين. وتم الاعتماد على هذا النموذج لأغراض هذه الدراسة، وتعطى المعادلة الرياضية لهذا النموذج كمايلي:

$$P_i(\Theta) = c_i + (1 - c_i)[1 + e^{(-Dai(\Theta - bi))}]^{-1} \quad ; i = 1, 2, 3 \dots, n$$

حيث:

$P_i(\Theta)$: احتمال أن يجيب المفحوص ذو القدرة (Θ) عن الفقرة (i) إجابة صحيحة.

D: عامل التدرج (Scaling Factor)، وهو ثابت لجميع فقرات الاختبار ويساوي (1.7).

a_i : معامل تمييز الفقرة.

b_i : معامل صعوبة الفقرة (i).

e: هي الأساس اللوغاريتمي الطبيعي وتساوي (2.718).

C_i : معلمة التخمين للفقرة (i).

وأما دالة معلومات الاختبار فتلعب دوراً مهماً في نظرية استجابة الفقرة، إذ يمكن من خلالها تحديد الخطأ المعياري في التقدير، وتتمتع دالة معلومات الاختبار بميزة كونها مستقلة عن عينة المفحوصين الخاضعة للاختبار، وبذلك تقدم نظرية استجابة الفقرة مميزات إضافية فيما يتعلق بزيادة القدرة على تقدير أخطاء القياس (Brannick, 2003).

كما أن دالة معلومات الفقرة تبين مدى مساهمة كل فقرة من فقرات الاختبار في دالة الاختبار بشكل مستقل عن الفقرات الأخرى للاختبار، فإذا كان لدينا فكرة جيدة عن قدرات مجموعة من المفحوصين، فيمكن انتقاء فقرات الاختبار التي ترفع من شأن المعلومات التي يقدمها الاختبار في المدى الذي تتوزع فيه قدرات المفحوصين.

وتعطى دالة معلومات الفقرة في النموذج اللوجستي ثلاثي المعلمة كما في المعادلة التالية:

$$I(\theta) = \frac{D^2 a_i^2 Q_i (P_i - C_i)^2}{(1 - C_i)^2 P_i}$$

ويأخذ $I(\theta)$ القيمة العظمى له عند قيمة معلمة الصعوبة (b) على مقياس القدرة (Θ) في النموذجين الأحادي والثنائي، أما في النموذج ثلاثي المعلمة فتكون القيمة العظمى للمعلومات عند قيمة (Θ) كما في المعادلة التالية:

$$\theta_{\max} = bi + \frac{1}{Dai} \ln \left[\frac{1}{2} + \frac{1}{2} \sqrt{1 + 8ci} \right]$$

واقتران معلومات الاختبار $I(\theta)$ ككل يكون مساوياً لمجموع اقترانات دالة المعلومات لفقرات الاختبار جميعها عند مستوى قدرة معين كما في المعادلة الآتية:

$$I(\theta) = \sum_{i=1}^n \frac{D^2 a_i^2 Q_i (P_i - C_i)^2}{(1 - C_i)^2 P_i}$$

وترتبط دقة القياس بحجم المعلومات التي تعطيها الفقرة التي تساهم في الاختبار أو المقياس ككل. ويتغير اقتران معلومات الاختبار بتغير القدرة، وهذا التغير يؤدي -أيضاً- إلى التغير في الخطأ المعياري في التقدير (S.E.E)، والقيمة العظمى لاقتران معلومات الاختبار $I(\theta)$ سوف تقابل أقل قيمة للخطأ المعياري في التقدير (لأن الخطأ المعياري في التقدير يتناسب عكسياً مع

$$\left. \text{الجذر التربيعي لاقتران معلومات الاختبار} \right) \left(S.S.E = \frac{1}{\sqrt{I(\theta)}} \right)$$

كما أن الاختبار السهل من المتوقع أن يقدم تقديرات أكثر دقة ومصداقية عند مستويات القدرة المتدنية، بعكس الاختبار الصعب الذي يمكن أن يقدم معلومات أكثر دقة عند مستويات القدرة العليا، لذا فهو أكثر فائدة للمفحوصين ذوو القدرات العالية (Hamblton & Swaminathan, 1985).

Swaminathan, 1985)

مشكلة الدراسة

إن اختيار الموهبات يتم وفق أسس منطقية أو وفق أسس إحصائية، ويعتمد الكثير من مطوري الاختبارات على الأسس المنطقية في ضوء محددات معينة، مبنية على خبرتهم في محتوى المادة، أو خبرتهم التراكمية في المفحوصين، ويتم الاعتماد على فحص جاذبية الموهبات إحصائياً وخاصةً في الاختبارات المقتنة، إلا أن الأدبيات المتوفرة في القياس لم تشر إلى أن اختيار الموهبات من إجابات المفحوصين أنفسهم مقارنةً بالموهبات التي يتم اختيارها وفق أسس منطقية من مطوري الاختبارات أو الباحثين، أو وفق أسس إحصائية، وفحص انعكاسات هذا الاختيار على دالة معلومات الاختبار يحتاج إلى دراسات تجريبية، والتي تشكل الدراسة الحالية منطلقاً بحثياً في هذا المجال. وبالتحديد ستحاول هذه الدراسة الإجابة عن السؤال التالي:

هل تختلف دالة معلومات اختبار الاختيار من متعدد باختلاف الإستراتيجية المتبعة في اختيار موهبات فقراته؟

أهمية الدراسة

تأتي الأهمية العلمية والعملية لهذه الدراسة من خلال محاولتها تحديد الاستراتيجية المثلى لتحديد موهبات فقرات اختبار الاختيار من متعدد للحصول على تقديرات دقيقة لدالة معلومات

الفقرة ولدالة معلومات الاختبار، حيث يُعد الانتشار الواسع لتطبيقات نظرية استجابة الفقرة من الامور التي ساهمت في تطوير الاختبارات، لأن النظرية الحديثة في القياس تتمتع بدقة وموضوعية في قياس السلوك ومن ثم توفير تقدير صادق للخصائص السيكومترية للاختبار ولفقراته، وتقدير دقيق لدالة معلومات الفقرة ودالة معلومات الاختبار عند مستويات معينة من القدرة.

محددات الدراسة

- اقتصرت هذه الدراسة على إيجاد دالة معلومات الفقرة ودالة معلومات الاختبار.
- اقتصرت عينة الدراسة على طلبة الصف التاسع الأساسي.
- اقتصرت هذه الدراسة على وحدة "تحليل المقادير الجبرية" من منهج الرياضيات المقرر لطلبة الصف التاسع الأساسي.
- استخدم في هذه الدراسة النموذج اللوجستي ثلاثي المعلمة لل فقرات، الذي يتعامل مع جميع معالم الفقرة (الصعوبة، والتمييز، والتخمين).
- البيانات التي تم استخدامها جمعت من تطبيق أداة الدراسة على عينة الدراسة لعام 2011/2010م.
- تم استخدام برنامج (MG - BILOG) وبرنامج (SPSS)، للتحقق من خلالها من افتراضات أحادية البعد والاستقلال الموضوعي لبيانات الدراسة.
- لم يتم إلغاء جميع الفقرات غير المطابقة للنماذج، بل تم إلغاء الفقرات غير المطابقة والمشاركة بين النماذج الثلاثة.

الدراسات السابقة

اهتمت العديد من الدراسات بدراسة خصائص المفحوصين على دقة تقدير معالم الفقرات، من هذه الدراسات دراسة أجراها بيلتون (Pelton, 2002) التي هدفت إلى مقارنة الدقة والاستقرار في تقدير معلمة الصعوبة ومعلمة القدرة باستخدام النظرية الكلاسيكية في القياس والنماذج اللوجستية في نظرية استجابة الفقرة، وذلك من خلال استخدام أسلوب المحاكاة، وتوصلت الدراسة إلى أن تقديرات القدرة يمكن مقارنتها عبر النظرية الكلاسيكية ونظرية استجابة الفقرة، إذ تتباين تقديرات القدرة حسب كمية المعلومات المتوفرة في مجموعة البيانات المولدة، والتي تتأثر بدورها بأحادية البعد ودرجة التخمين ومدى التباين في صعوبة الفقرات مقارنة بقدرات الأفراد، كما توصلت الدراسة إلى أنه بوجود عينة مكونة من (999) مفحوصاً، واختبار مكون من (33) فقرة، فإن النموذج اللوجستي الثنائي يقدم تقديرات أكثر دقة لمعلمة الصعوبة من نموذج راش والنموذج اللوجستي الثلاثي والنظرية الكلاسيكية في القياس وذلك بوجود تخمين قليل في الإجابة على فقرات الاختبار.

كما قام شيزوكي وتاكوشي وآخرين (Shizuka & Takeuchi, et all. 2006) بدراسة هدفت إلى مقارنة بين الخصائص السيكومترية لفقرات اختبار استيعاب المقروء باللغة الإنجليزية، الذي يؤهل الطلبة بالالتحاق بالجامعات اليابانية، ويتكون الاختبار من (38) فقرة ذات الاختيار من متعدد، وكانت تلك الفقرات تحتوي على أربعة موهبات، وفي مرحلة لاحقة تم خفض عدد موهبات من أربعة موهبات إلى ثلاثة موهبات من خلال حذف الموهبة الأقل تمويهاً، وتكون الاختبار من نموذجين: النموذج الأول يتكون من فقرات لها ثلاثة موهبات، والنموذج الثاني من فقرات لها أربعة موهبات، واستخدم الباحث النموذج اللوجستي الأحادي (راش) في هذه الدراسة.

وتم تطبيق النموذجين من الاختبار على عينتين مختلفتين من الطلبة، وتم تقدير معلمة الصعوبة للفقرات في كلا النموذجين، حيث أشارت النتائج عدم وجود فروق ذات دلالة إحصائية في تقديرات معلمة صعوبة الفقرات لكلا النموذجين، وأشارت النتائج إلى عدم وجود تغيير يستحق الذكر في دقة وعدد الموهبات المميزة. وأوصى الباحث باستخدام الاختبارات التي تحتوي على ثلاثة موهبات تجنباً لهدر الوقت والمال في تطوير الاختبارات وإدارتها.

وأجرى فيزبارتك (Fitzpatrick, 2009) دراسة هدفت إلى بيان أثر تقليل عدد فقرات اختبار التهيئة على معدل إتقان الطلبة، ولتحقيق غرض الدراسة تم إعداد (12) اختبار من نوع الاختيار من متعدد، حيث تكونت الاختبارات من (5،10،15) فقرة، وتم تطبيق الاختبار واستخدام النموذج أحادي المعلمة لأغراض التحليل، وأظهرت النتائج أن هناك تباين في معدلات إتقان الطلبة عند استخدام الاختبارات القصيرة، كما أشارت النتائج -أيضاً- إلى أن هناك اختلاف في معدلات الصعوبة، وأوصت الدراسة باستخدام اختبارات تحتوي على عدد أكبر من (15) فقرة لزيادة الاستقرار في معلم الفقرات.

وفي دراسة أخرى أجرت تارانت ووير، ومحمد (Tarrant, Ware, Mohammed,) (2009) التي هدفت إلى معرفة فعالية المموهات معتمدون على نتائج التحليل الوصفي للبيانات ومن أجل تحقيق هدف هذه الدراسة قام الباحثون بتحديد مموهات غير فعالة في اختبارات الاختيار من متعدد والمطور من قبل المعلم والمقدمة في تخصص التمريض في إحدى الجامعات في هونج كونج، وأعد الباحثون سبعة اختبارات من نوع الاختيار من متعدد في تخصص التمريض. وتم تطبيق الاختبارات على طلبة كلية التمريض، وتم تحليل النتائج بناءً على دالة معلومات الفقرة لكل نموذج من نماذج الاختبار لتحديد مموهات غير فعالة في نماذج الاختبار.

وأشارت النتائج إلى أن نسبة المموهات الفعالة في نماذج الاختبار كانت (52.2%)، كما أشارت نتائج الدراسة إلى أن (10.2%) من المموهات كانت نسبة اختيارها كإجابة صحيحة يساوي صفر، وأن الفقرات ذات المموهات الأكثر فاعلية كانت فقراتها تتميز بارتفاع معاملات الصعوبة والتميز لها مقارنةً مع الفقرات الأخرى من الاختبار.

وأجرى باجبيه وامراحي (Baghaei, Amrahi, 2011) دراسة هدفت إلى تحديد العدد الأمثل من المموهات لفقرات اختبارات الاختيار من متعدد وللمقارنة بين إحصائيات الفقرة وثبات الاختبار، ولتحقيق هدف هذه الدراسة قام الباحثان بإعداد ثلاثة نماذج (A, B, C) متكافئة، يحتوي كل نموذج على 30 فقرة، ومختلفة فقط في عدد المموهات للفقرات في كل نموذج، ففي النموذج (A) كانت الفقرة الواحدة تحتوي على خمسة مموهات من ضمنها الإجابة الصحيحة، وللنموذج (B) أربعة مموهات بعد حذف أحد المموهات في النموذج (A) عشوائياً، وللنموذج (C) ثلاثة مموهات بعد حذف أحد المموهات عشوائياً في النموذج (B).

وتم تطبيق نماذج الاختبار الثلاثة على عينة عشوائية مكونة من 78 طالباً و102 طالبة من طلبة مرحلة البكالوريوس في تخصص اللغة الإنجليزية، وتم حساب كل من الأوساط الحسابية لعلامات الطلبة والأوساط الحسابية للأخطاء المعيارية، وحساب قيم ثبات نماذج الاختبار، وتقدير معلمة الصعوبة ومعلمة التمييز لفقرات النماذج.

وأشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية بين الأوساط الحسابية لعلامات الطلبة وبين الأوساط الحسابية للأخطاء المعيارية، وبين قيم معاملات الثبات للنماذج الثلاثة، وكذلك أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين تقديرات معلمة الصعوبة للفقرات في نماذج الاختبار، أما معلمة التمييز فأشارت النتائج إلى وجود فروق ذات دلالة إحصائية، حيث كانت تقديرات معلمة التمييز للنموذج الذي يحتوي فقراته على ثلاثة مموهات هي الأعلى، وأوصى الباحثان الاعتماد على اختبارات اختيار من متعدد التي تحتوي فقراته على ثلاثة مموهات توفيراً للجهد والوقت.

وفي دراسة أجراها بني عطا والرابعي (2013) هدفت إلى التحقق من أثر عدد البدائل وتغيير الموقع المموه القوي في فقرات اختبار الاختيار من متعدد على معالم الفقرات، وقدرة الفرد ودالة معلومات الاختبار، ولتحقيق الغرض من هذه الدراسة، قام الباحثان بإعداد اختبار تحصيلي في مقرر الرياضيات من نوع الاختيار من متعدد مكون من (41) فقرة، واشتمل على

أربعة نماذج حسب عدد البدائل وموقع المموه القوي، وتم تطبيق الاختبار على عينة مكونة من (2111) طالباً وطالبة من طلبة الصف العاشر في الأردن.

وأظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة ومعالم التخمين وبين متوسطات معلمة القدرة للأفراد تعزى إلى عدد البدائل وموقع المموه القوي، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات معالم التمييز تعزى إلى عدد البدائل وموقع المموه القوي، ولقد كانت تلك الفروق لصالح النموذج الذي يحتوي على خمسة بدائل وموقع المموه يكون بعيداً عن الإجابة الصحيحة. كما أشارت النتائج إلى أن دالة معلومات الاختبار تباينت بتباين نماذج الاختبار، وكذلك أشارت النتائج وجود فروق ذات دلالة إحصائية بين قيم معاملات الثبات الامبريقي حيث جاءت لصالح النموذج الذي يحتوي على خمسة بدائل وموقع المموه يكون بعيداً عن الإجابة الصحيحة.

وفي دراسة أجراها الزبون (2013) هدفت لمعرفة أثر حجم العينة على تقديرات دالة المعلومات للاختبار والخطأ المعياري في تقديرها، ولتحقيق أغراض هذه الدراسة قام الباحث باستخدام استجابات (7500) طالباً وطالبة من طلبة الصف الثامن الأساسي على الاختبار الوطني لضبط نوعية التعليم، وقام الباحث بتوزيع الطلبة عشوائياً على خمسة مجموعات (500) طالباً وطالبة للعينة الأولى، 1000 طالباً وطالبة للعينة الثانية، 1500 طالباً وطالبة للعينة الثالثة، 2000 طالباً وطالبة للعينة الرابعة، 2500 طالباً وطالبة للعينة الخامسة). وأظهرت النتائج أن مقدار دالة المعلومات للاختبار يزداد بزيادة حجم العينة، وكذلك بينت النتائج أن الخطأ المعياري في تقدير دالة المعلومات يتناقص بزيادة حجم العينة.

الطريقة والإجراءات

مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلبة الصف التاسع الأساسي الذكور والإناث في المدارس الحكومية التابعة لمديرية تربية الزرقاء/ الأردن، ويبلغ عدد الطلبة الذكور (5245) طالباً، وعدد الإناث (5401) طالبة. حسب إحصائيات قسم التخطيط في مديرية تربية الزرقاء.

عينة الدراسة

تكونت عينة الدراسة من (1500) طالباً وطالبة من طلبة الصف التاسع الأساسي، إذ تم اختيار مجموعة من المدارس عشوائياً من بين المدارس التي تحوي الصف التاسع الأساسي.

أداة الدراسة

لتحقيق أهداف الدراسة، تم إعداد اختبار تحصيلي من نوع الاختيار من متعدد بهدف قياس تحصيل طلبة الصف التاسع في مبحث الرياضيات (وحدة المقادير الجبرية).

بناء المموهات لفقرات الاختبار باستخدام ثلاثة طرق مختلفة

الطريقة الأولى: تم توزيع نموذج فقرات الاختبار (بدون مموهات) على مجموعة تتكون من (10) معلمين ممن يدرسون مادة الرياضيات للصف التاسع الأساسي في المدارس الحكومية التابعة لتربية الزرقاء، وتم الطلب من كل معلم كتابة جميع المموهات الممكنة لكل فقرة من فقرات الاختبار، بالإضافة إلى الإجابة الصحيحة للفقرة، وبعد ذلك تم حصر جميع المموهات لكل فقرة ووضعها في نموذج واحد فقط، وتم عرض هذا النموذج على ثلاثة معلمين من ذوي الخبرة لاختيار أفضل ثلاثة مموهات بالإضافة إلى الإجابة الصحيحة للفقرة بناءً على خبرتهم، لتشكيل النموذج (A).

الطريقة الثانية: بعد حصر مموهات كل فقرة من فقرات الاختبار، التي تم الحصول عليها من خلال الطريقة الأولى، تم تكوين عدة نماذج للاختبار لكل نموذج ثلاثة مموهات بالإضافة إلى الإجابة الصحيحة، من خلال التوزيع العشوائي لمموهات كل فقرة على فقرات النماذج وتغيير موقع الإجابة الصحيحة لنماذج الاختبار بعد ذلك تم تطبيق نماذج الاختبار على عينة عشوائية بواقع (50) طالباً وطالبة لكل نموذج.

تم تطبيق نماذج الاختبار في نفس الوقت، وبشكل عشوائي على طلبة الشعبة الواحدة، ثم تم تصحيح أوراق الاختبار، وإجراء التحليلات الإحصائية لتحديد أفضل ثلاثة مموهات (الأكثر تمييزاً) لكل فقرة من فقرات الاختبار، بالإضافة إلى الإجابة الصحيحة للفقرة، لتشكيل النموذج (B).

الطريقة الثالثة: تم توزيع فقرات الاختبار (أي على شكل تكميل)، على عينة عشوائية تتكون من (200) طالباً وطالبة من طلبة الصف التاسع الأساسي، بعد أن تم إبلاغ الطلبة عن الاختبار بالاتفاق مع إدارة كل مدرسة، ومدرسي ومدرسات الرياضيات للصف التاسع الأساسي، وعلى أن تكون علامة هذا الاختبار جزءاً من علامة الاختبار الأول للطلبة، ذلك أن فقرات الاختبار تغطي معظم الوحدة الأولى من كتاب الرياضيات المقرر للصف التاسع الأساسي، حيث يكون الاختبار الأول للطلبة في المدارس في الوحدة الأولى والثانية تقريباً، ثم تم تصحيح أوراق الاختبار من قبل الباحث، حيث تم رصد أخطاء الطلبة أثناء إجاباتهم على فقرات الاختبار، وتم إجراء مقارنات بين أخطاء الطلبة في كل فقرة من فقرات الاختبار، وتحديد الأخطاء الأكثر تكراراً، لتضمين تلك الأخطاء لمموهات فقرات الاختبار، ليتم تشكيل النموذج (C) للاختبار.

تطبيق نماذج الاختبار على المفحوصين

تم تطبيق نماذج الاختبار الثلاثة على عينة الدراسة كلاً في مدرسته، وبشكل جماعي داخل غرفة الصف، وفق البرنامج الزمني المتفق عليه مع معلمي الرياضيات في كل مدرسة، وقد تم توزيع نماذج الاختبار (النموذج A، النموذج B، النموذج C)، إذ يأخذ الطالب الأول في الجلوس النموذج A، والطالب الذي بجانبه النموذج B، والطالب الذي يليه النموذج C وهكذا، إلى أن يتم توزيع نماذج الاختبار على جميع الطلبة المفحوصين داخل الشعبة المختارة، وبعد إنهاء الطالب

الأول الإجابة على النموذج A، يتم إخضاعه للنموذج B، وبعد إنهائه الإجابة على النموذج B، يتم إخضاعه للنموذج C، وهكذا حتى يتم إخضاع كل طالب من الطلبة المفحوصين للإجابة على نماذج الاختبار الثلاثة.

وتم تصحيح أوراق الاختبار بحيث تعطى للفقرة الصحيحة العلامة (1) وصفر إذا كانت خاطئة، وتم إجراء الإحصائيات اللازمة للإجابة عن سؤال الدراسة.

المعالجات الإحصائية

صدق المحتوى

للتأكد من صدق محتوى الاختبار، قام الباحث بعرض الصورة الأولية ل فقرات الاختبار، وجدول المواصفات، وتحليل المحتوى، والنتائج التعليمية، على مجموعة من المحكمين والمختصين في تدريس الرياضيات من مدرسين، ومشرفين تربويين، ينقسمون كمايلي: أربعة من مشرفي الرياضيات في مديرية الزرقاء الأولى والثانية، أحدهم يحمل درجة الدكتوراه في أساليب تدريس الرياضيات، والباقي يحملون درجة الماجستير في أساليب تدريس الرياضيات، وسبعة معلمين ممن يدرسون مادة الرياضيات للصف التاسع الأساسي، واقتراح ما يروونه مناسباً من تعديل على بعض الفقرات، من خلال استبيان تم أعداده لهذا الغرض.

وتم تعديل بعض فقرات الاختبار الغامضة، وقد تراوحت الأوساط الحسابية لتقديرات المحكمين ضمن تدرج ليكرت على فقرات الاستبيان المتعلقة بالحكم على فقرات الاختبار ما بين (3.89 - 4.42)، وللحكم على جدول المواصفات ما بين (4.45 - 4.51)، أما الوسط الحسابي لتقديرات المحكمين للحكم على مدى تغطية فقرات الاختبار ككل لجدول المواصفات فقد كان 4.43.

صدق المحك

تم التأكد من صدق المحك وذلك من خلال حساب معامل ارتباط بيرسون بين أداء العينة الاستطلاعية على النماذج الثلاثة (A, B, C)، وعلاماتهم في الاختبار الأول في مادة الرياضيات، إذ كانت قيمته (0.84) للنموذج (A)، و(0.67) للنموذج (B)، و(0.64) للنموذج C.

معامل الثبات

تمّ تقدير معامل ثبات الاختبار للنماذج المعده، باستخدام النظرية التقليدية في القياس، ونظرية استجابة الفقرة والجدول (1) يبين ذلك.

جدول (1): معاملات الثبات للنماذج الثلاثة وفق النظرية التقليدية والنظرية الحديثة في القياس.

النموذج	وفق النظرية التقليدية Cronbach's alpha	وفق النظرية الحديثة Empirical Reliability
A	0.94	0.94
B	0.85	0.88
C	0.85	0.88

التحقق من افتراض أحادية البعد (Unidimensionality)

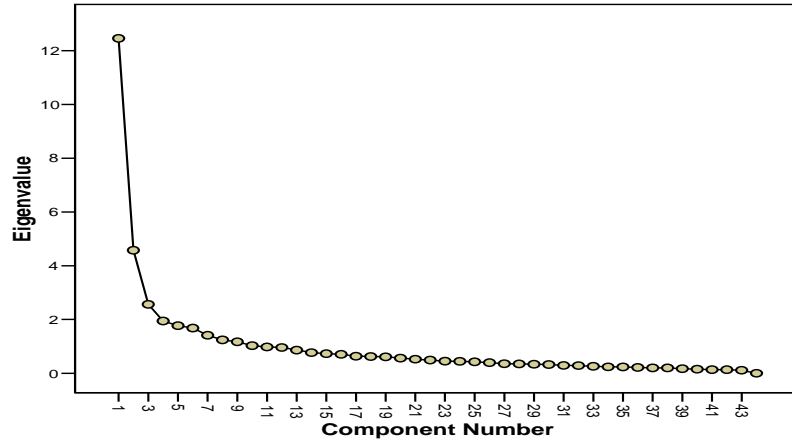
تم التحقق من افتراض أحادية البعد من خلال استخدام التحليل العاملي، بحيث تم استخدام طريقة المحاور الأساسية (Principal Axis Factoring)، مع استخدام طريقة (Quartimax Rotation) للعوامل التي تزيد جذورها الكامنة عن (1) للبيانات الناتجة من استجابات العينة الاستطلاعية على نماذج الاختبار الثلاثة حيث أستخدم برنامج (SPSS) في تحليل البيانات. وكانت نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر من (2)، وكانت نسبة الفرق بين الجذريين الكامنين الأول والثاني إلى نسبة الفرق بين الجذريين الكامنين الثاني والثالث عالية (3.92 – 70.45)، وتعد هذه القيم كما يشير هاتي (Hattie, 1985) مؤشر على أحادية البعد. ويمثل الجدول (2) نتائج التحليل العاملي لنماذج الاختبار الثلاثة.

جدول (2): نتائج التحليل العاملي لنماذج الاختبار (A, B, C).

نموذج الاختبار	عدد العوامل	نسبة التباين التي يفسرها العامل الأول %	نسبة التباين التي يفسرها العامل الثاني %	نسبة التباين التي يفسرها العامل الثالث %	نسبة الجذر الكامن الأول / (نسبة الجذر الثاني - الجذر الثالث)	نسبة الجذر الكامن الأول / (نسبة الجذر الثاني)
A	10	28.318	10.401	5.827	3.917	2.722
B	15	15.170	5.973	5.681	31.496	2.539
C	15	15.245	5.805	5.671	70.447	2.626

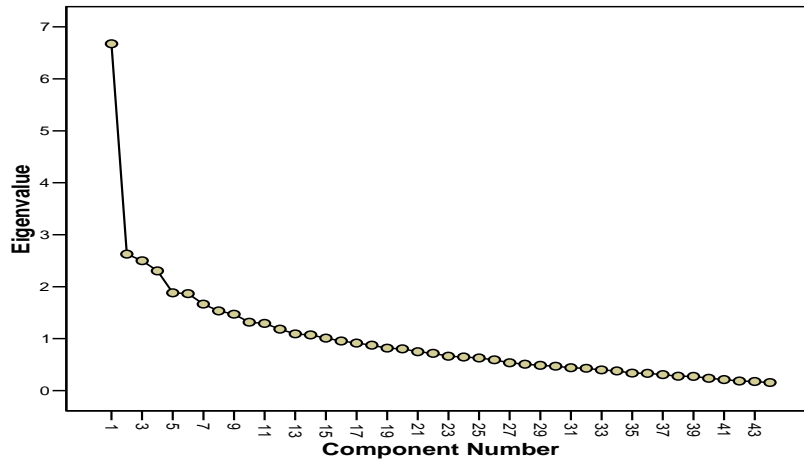
وتبين الأشكال (1)، و(2)، و(3)، التمثيل البياني للجذور الكامنة الخاصة بنماذج الاختبار الثلاثة بعد تطبيقها على عينة الدراسة.

Scree Plot

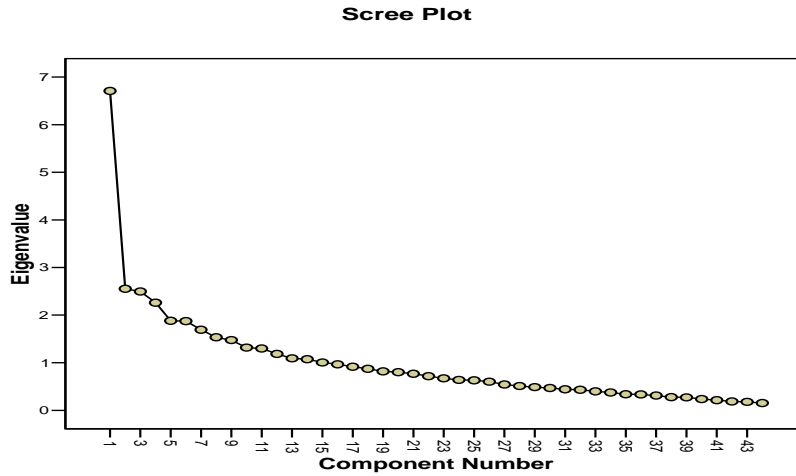


شكل (1): التمثيل البياني لقيم الجذور الكامنة للعوامل للنموذج (A) من الاختبار.

Scree Plot



شكل (2): التمثيل البياني لقيم الجذور الكامنة للعوامل للنموذج (B) من الاختبار.



شكل (3): التمثيل البياني لقيم الجذور الكامنة للعوامل للنموذج (C) من الاختبار.

افتراض الاستقلال الموضعي

يعد تحقق افتراض الاستقلال الموضعي، نتيجة محصلة من تحقق افتراض أحادية البعد (Hulin, Drasgow & Parsons, 1983)، في حين ينظر إليه بعض الباحثين على أنه افتراض مكافئ لافتراض أحادية البعد (Hambleton & Swaminathan, 1985)، وبالتالي إذا تحقق الباحث من افتراض أحادية البعد، فإن افتراض الاستقلال الموضعي يتحقق أيضاً، وعليه فقد اكتفى الباحث بالتحقق من افتراض أحادية البعد، للإشارة إلى تحقق افتراض الاستقلال الموضعي.

إجراءات المطابقة للأفراد والفقرات

تم إجراءات المطابقة للأفراد والفقرات، بإدخال استجابات الأفراد المفحوصين والبالغ عددهم (1500) طالباً وطالبة، على نماذج الاختبار (A, B, C)، إذ أدخلت استجابات الطلبة على كل نموذج من نماذج الاختبار الثلاثة على حدا. وقد أفرزت نتائج التحليل عند استخدام برنامج (BILOG – MG) لأول مرة على البيانات الخام، باستخدام اختبار كأي تربيع عند مستوى الدلالة ($\alpha=0.01$)، عدم مطابقة استجابات مجموعة من المفحوصين، وكانت قيمة الاحتمالية لهم اقل من (0.01)، ويبين الجدول (3) أعداد الطلبة وأرقام المفحوصين غير المطابقين للنموذج اللوجستي ثلاثي المعلمة.

جدول (3): أعداد المفحوصين غير المطابقين للنموذج اللوجستي ثلاثي المعلمة.

أرقامهم	أعداد المفحوصون غير المطابقين	عدد المفحوصين	نموذج الاختبار
987، 817 620، 176، 150	5	1500	A
،182، 128، 106، 104، 102، 62 830، 1344، 776، 561	10	1500	B
،662، 550، 287، 230، 214، 62، 48 ،776، 754، 751، 750، 748، 746 830، 1119	15	1500	C

وبعد ذلك قام الباحث بحذف استجابات الأفراد التي لم تتطابق استجاباتهم مع توقعات النموذج، حيث بلغ عدد الأفراد غير المطابقين للنموذج (27) مفحوصاً.

ومن ثم قام الباحث بإعادة التحليل لاختبار مدى مطابقة الفقرات للنموذج اللوجستي ثلاثي المعلمة، باستخدام اختبار كأي تربيع عند مستوى الدلالة ($\alpha=0.01$)، وتبين عدم مطابقة بعض الفقرات في النماذج الثلاثة (A, B, C) للنموذج اللوجستي ثلاثي المعلمة، إذ كانت قيمة الاحتمالية أقل من (0.01) لهذه الفقرات، وبين الجدول (4) أرقام الفقرات غير المطابقة للنموذج اللوجستي ثلاثي المعلمة، حسب تسلسل أرقامها في نماذج الاختبار في مجموعة الدراسة.

جدول (4): أرقام الفقرات غير المطابقة للنموذج اللوجستي ثلاثي المعلمة.

أرقامها	أعداد الفقرات غير المطابقة	عدد الفقرات	النموذج
40، 36، 27، 11، 9	6	44	A
27، 19، 14، 12، 11، 9	7	44	B
27، 15، 14، 13، 11، 9، 5	9	44	C

ويلاحظ من الجدول السابق أن الفقرات التي تحمل الأرقام (9، 11، 27) في نماذج الاختبار الثلاثة (A, B, C)، كانت غير مطابقة للنموذج اللوجستي ثلاثي المعلمة، ونتيجة للتغاير الواضح في وضع بقية الفقرات غير المطابقة للنموذج المستخدم، وكون هذه الفقرات تمثل العمود الفقري لهذه الدراسة، كان لا بد من الإبقاء على أكبر قدر ممكن من فقرات الاختبار في كل نموذج من نماذج الاختبار الثلاثة، وبالتالي فقد اكتفى الباحث بحذف الفقرات الثلاثة المشتركة فقط في النماذج الثلاثة، دون أن يتم حذف الفقرات الأخرى غير المطابقة للنموذج اللوجستي ثلاثي المعلمة، وبناءً على ذلك تم اعتماد عينة المفحوصين المكونة من (1473) طالباً وطالبة، وعينة الفقرات (41) فقرة في كل نموذج، كأساس في إجراء التحليلات الإحصائية المختلفة.

إذ قام الباحث بإعادة التحليل لحساب التقديرات النهائية لكل من معالم الفقرات ومعلمة القدرة للمفحوصين، وفق النظرية الحديثة في القياس، للبيانات الخاصة باستجابات أفراد مجموعة الدراسة على النماذج الثلاثة (A, B, C)، للإجابة عن سؤال الدراسة.

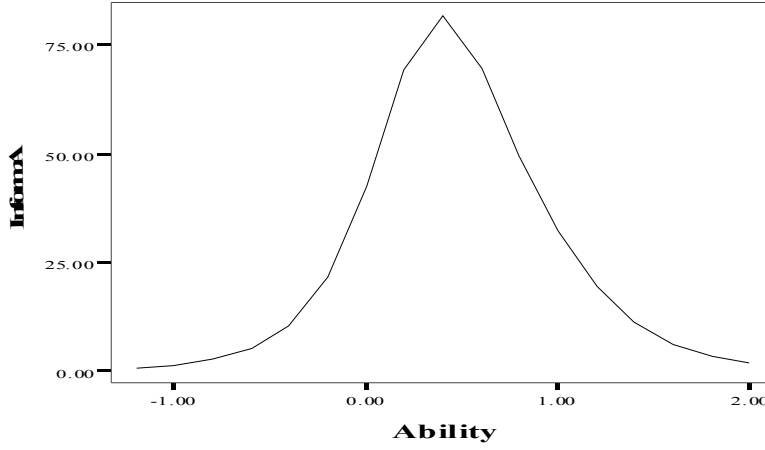
نتائج الدراسة ومناقشتها

تم استخدام النموذج اللوجستي الثلاثي لتقدير معالم الصعوبة، ومعالم التمييز، ومعالم التخمين لكل فقرة في نماذج الاختبار الثلاثة (A, B, C)، وتحديد قدرات المفحوصين باستخدام برمجية (BILOG –MG)، وتم إيجاد دالة المعلومات لكل فقرة في كل النماذج عند مستويات مختارة من القدرة، وإيجاد دالة معلومات الاختبار لكل نموذج عند مستويات مختارة من القدرة، وتم رسم دالة المعلومات الاختبار لكل نموذج من نماذج الاختبار (A, B, C)، والمقارنة بينها عند مستويات قدرة محددة، ويبين الجدول (5) قيم دالة المعلومات عند مستويات مختارة من القدرة.

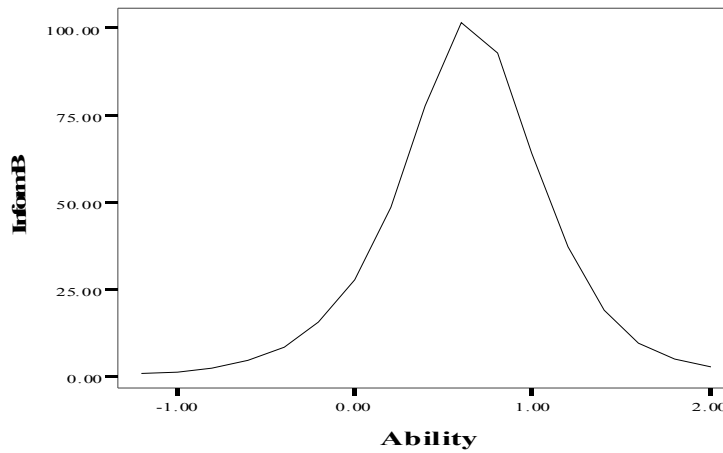
جدول (5): قيم دالة المعلومات التي تقدمها نماذج الاختبار (A, B, C) عند مستويات مختارة من القدرة.

دالة المعلومات			القدرة
النماذج			
C	B	A	
0.45	0.56	0.57	1.2-
1.10	1.15	1.24	1.0-
2.50	2.26	2.56	0.8-
5.35	4.35	5.12	0.6-
10.48	8.35	10.39	0.4-
18.01	15.53	21.55	0.2-
28.04	27.57	42.49	0.0
43.63	48.25	69.38	0.2
69.31	77.40	81.70	0.4
96.88	101.55	69.61	0.6
100.66	92.68	49.51	0.8
72.37	63.90	32.16	1.0
40.36	36.99	19.44	1.2
19.30	18.89	11.14	1.4
9.12	9.44	6.20	1.6
4.51	4.91	3.42	1.8
2.34	2.67	1.89	2.0

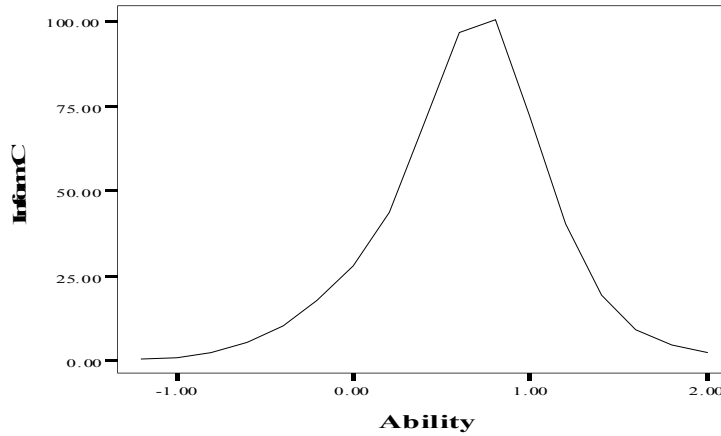
وتم رسم دالة معلومات الاختبار عند مستويات القدرة المختارة الناتجة من تطبيق نماذج الاختبار الثلاثة (A, B, C)، والموضحة بالأشكال (4)، و(5)، و(6).



شكل (4): رسم دالة معلومات الاختبار للنموذج (A) والقدرة في مجموعة الدراسة.



شكل (5): رسم دالة معلومات الاختبار للنموذج (B) والقدرة في مجموعة الدراسة.



شكل (6): رسم دالة معلومات الاختبار للنموذج (C) والقدرة في مجموعة الدراسة.

أن العلاقة بين كمية المعلومات والقدرة هي دالة المعلومات للفقرة، وأن كمية المعلومات التي تقدمها الفقرة الواحدة تكون قليلة، وكذلك تقيس الفقرة القدرة بدرجة عالية من الدقة عندما يتوافق مستوى القدرة مع مستوى معلمة الصعوبة للفقرة، كما تتأثر دالة المعلومات للاختبار بخصائص الفقرات وعددها.

ويلاحظ من خلال الجدول (3) والأشكال المعروضة سابقاً، أن أكبر قيمة لدالة معلومات الاختبار التي يقدمها نموذج الاختبار (B)، الذي بنيت مموهات فقراته باستخدام التحليل الاحصائي (الفقرات الأكثر تمييزاً) تساوي (101.55) عند مستوى قدرة (0.6)، يليه نموذج الاختبار (C)، الذي بنيت فقراته من خلال اجابات الطلبة، فكانت أكبر قيمة لدالة معلومات الاختبار عند مستوى القدرة (0.8) وكانت قيمتها (100.66)، ونلاحظ تقارب قيمتي دالة معلومات الاختبار للنموذجين (B, C)، ولكن عند مستويين مختلفين من القدرة، وكانت أقل قيمة لدالة معلومات الاختبار في نموذج الاختبار (A)، الذي بنيت فقراته من خلال المعلمين، فقد كانت أكبر قيمة لدالة المعلومات عند مستوى القدرة (0.4)، إذ بلغت قيمتها (81.70)، ويفسر تقارب قيمة دالة الفقرات عند مستويات معينة من القدرة إلى أن كل مفحوص من عينة الدراسة التي خضعت للاختبارات، كان يخضع للإجابة عن النماذج الثلاثة في نفس جلسة الاختبار، وبالتالي فإن قدرة المفحوص للإجابة عن إحدى النماذج تكون مساوية للإجابة عن بقية النماذج، وهذا ما تدل عليه نتائج التحليل.

ونلاحظ أن نماذج الاختبار (A, B, C)، قد قدمت معلومات أكثر فاعلية عند المفحوصين ذوي القدرات المتوسطة؛ وكان أفضلها النموذج (B) الذي بنيت فقراته إحصائياً، ويمكن تفسير ذلك إلى ارتفاع قيم معلمة التمييز للفقرات عند هذه المستويات من القدرة، وبالتالي ستضخم من

دالة المعلومات للاختبار، إضافة إلى المدى الواسع من التشتت لتوزيع القدرة لاستجابات الأفراد على نماذج الاختبار الثلاثة، فالمدى الواسع من التشتت جعل نماذج الاختبار تعطي أكبر قيمة لدالة المعلومات عند مستويات القدرة المتوسطة، بينما كانت قيم دالة المعلومات التي تقدمها نماذج الاختبار، أقل ما يمكن عند مستويات القدرة العالية أو المتدنية.

وبناءً على هذه النتائج فإن الباحث ينصح باستخدام نموذج الاختبار الذي تم بناء فقراته إحصائياً، وذلك لأن فقرات هذا النموذج قدمت معلومات عالية عند مستويات القدرات المتوسطة للطلبة، ومثل هذه النموذج يتناسب مع معظم الطلبة في المراحل الدراسية والعمرية المختلفة، لأن قدراتهم في الغالب متوسطة وتتبع التوزيع الطبيعي.

المقترحات

1. مراعاة الطريقة التي يتم من خلالها بناء الموهبات لفقرات اختبارات الاختيار من متعدد، حيث أن الطريقة التي يتم من خلالها بناء الموهبات يؤثر على تقديرات معالم الفقرات وتقديرات القدرة للمفحوصين، وبالتالي يؤثر على دالة معلومات الفقرة ودالة معلومات الاختبار.
2. إجراء دراسات أخرى للتأكد من نتائج هذه الدراسة لمقررات دراسية مختلفة، ومراحل عمرية مختلفة.
3. الاعتماد على دالة معلومات الاختبار كمعيار عند بناء بنوك الاسئلة.

References (Arabic & English)

- Audeh, A. (2004). *Evaluation and measurement in the educational system*. Irbid, Jordan, Alamal Publishers.
- Alzboon, H. (2013). The Effect of sample in estimating the information functions of the test and its estimating standard error depending on the modern test theory of measurement. *Alnjah University Journal of Social Research*, 27(6).
- Bani Ata, Z. & Alrabbaie, I. (2013). The effect of alternative numbers and changing the position of the strong distractor on items parameters, person ability and information function. *Jordanian Journal of Educational Sciences*. 9(3). 319-333.

- Baghaei, P. & Amrahi, N. (2011). The effects of number of options on the psychometric characteristic of multiple choice. *Psychological Test and Assessment Modeling*, 53(2). 192-211.
- Brannick, M. (2003). “Basics of IRT one – linefile”: [//a:/item response theory.htm](#).
- Budescu, D. & Nevo, B. (1985). Optimal number of options: an investigation of the assumption of proportionality. *Journal of Educational Measurement*, 22(3). 183-196.
- Costin, F. (1972). Three-choice versus four-choice items implications for reliability and validity of objective achievement tests. *Educational and Psychological Measurement*, 32(2). 1035-1038.
- Ebel, R. & Frisbie, D. (1986). *Essentials of Educational Measurement*. (Fourth edition), New Jersey: Prentice-Hall.
- Grier, J. (1975). The number of alternatives for optimum test reliability. *Journal of Educational Measurement*, (12)2. 109-113.
- Gronlund, N. & Linn, R. (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan publishing co. Inc.
- Gronlund, N. (1982). *Constructing Achievement Tests*. (Third edition), New Jersey: Prentice - Hall.

- Hamilton, R. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer – Nijhoff publishing.
- Hulin, L. Drasgow, F. & Parsons, K. (1983). *Item response theory: Applications to Psychological measurement*. Homewood IL: Dow & Jones Irwin.
- Pellton, W. (2002). *The accuracy of unidimensional measurement model in the presence of deviations from the underlying assumptions*. Unpublished doctoral dissertation, Brigham University, USA.
- Straton, R. & Catts, R. (1980). *A comparison of two, three and four choices item tests given a fixed total number of choices*. *Educational and Psychological Measurement*, 40. 357-365.
- Shizuka, T. Takeuchi, O. Yashima, T. & Yoshizawa, K. (2006). *A comparison of three- and four-option English tests for university entrance selection purposes in Japan*. *Language Testing*, 23(1). 35-57.
- Fitzpatrick, R. (2009). *The Impact of Anchor Test Configuration on Student Proficiency Rates*. *Educational Measurement: Issues and Practice*. 27(4). 34-40 Win 2008.
- Tarrant, M. Ware, J. & Mohammed, M. (2009). *An assessment of functioning and non functioning distracters in multiple-choice questions: A descriptive analysis*. *BMC Medical Education*, 9, 2-11.