المجلة الأردنية في العلوم التربوية، مجلد 5، عدد 4، 2009، 309-335

أثر عدد البدائل في اختبار الاختيار من متعدد في تقديرات القدرة للأفراد والخصائص السيكومترية للفقرات والاختبار وفق نموذج راش في نظرية الاستجابة للفقرة

نضال الشريفين* و إيمان طعامنة**

تاريخ تسلم البحث 2009/1/8 تاريخ قبوله 2009/8/6

ملخص: هدفت هذه الدراسة إلى الكشف عن أثر عدد البدائل في اختبار الاختيار من متعدد في تقديرات القدرة للأفراد، وتقديرات الصعوبة للفقرات، واقتران المعلومات للفقرات والاختبار. ولتحقيق أهداف الدراسة تم بناء اختبار تحصيلي من نوع الاختيار من متعدد في الرياضيات لطلبة الصف العاشر الأساسي، تكوّن الاختبار بصورته النهائية من (40) فقرة. وتم تشكيل ثلاث صور للاختبار تختلف فقط في عدد بدائل فقراتها؛ فكانت ثلاثة بدائل لفقرات الصورة الأولى، وأربعة بدائل لفقرات الصورة الثانية، وخمسة بدائل لفقرات الصورة الثالثة. طبقت صور الاختبار على عينة مكونة من (600) طالبٍ وطالبةٍ بواقع (200) طالبٍ وطالبةٍ لكل صورة، ثم حللت الاستجابات على كل صورة من صور الاختبار بشكل مستقل باستخدام برنامجي (BIGSTEPS) و(BILOG-MG). أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية (α=0.05) بين متوسطات الأخطاء المعيارية في تقديرات معالم الصعوبة للفقرات، وعدم وجود فروق ذات دلالة إحصائية (α=0.05) بين معاملات ثبات الأفراد تعزى لعدد البدائل، بينما كانت معاملات الثبات للفقرات متساوية. كما أظهرت وجود فروق ذات دلالة إحصائية (α=0.05) بين متوسطات الأخطاء المعيارية في تقديرات معالم القدرة للأفراد؛ حيث كانت التقديرات أكثر دقة لقدرات أفراد الصورة الأولى للاختبار منها لقدرات 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 اكبر من المعلومات عند مستويات القدرة المتوسطة والمرتفعة من الصورتين الأخريين. كما بينت النتائج وجود فروق دالة إحصائياً بين معاملات صدق المحك ولصالح الصورة الثانية للاختبار. (**الكلمات المفتاحية:** عدد البدائل، اختبار الاختيار من متعدد، نموذج راش، اقتران المعلومات، الخصائص السيكومترية، نظرية الاستجابة للفقرة).

The Effect of Multiple Choice Test Number of Alternatives on The Estimation of a person's Ability and the Psychometric Properties of a Test and its Items According to Rasch Model in Item Response Theory (IRT)

Nedal al-Shraifin, *Faculty of Education, Yarmouk University, Irbid, Jordan.*
Eman Ta'mnah, *Faculty of Educational studies, Arab Open University, Amman, Jordan.*

***Abstract**: This study investigated the effect of multiple choice test number of alternatives on the estimations of: a Person's ability, item difficulty, and the information function of test and its items. To achieve the study objectives, a multiple choice achievement test consisting of 40 items in tenth grade maths was constructed. The test took three different forms in terms of the number of alternatives: Form One with three alternatives, Form Two with four alternatives, and Form Three with five alternatives. The test forms were applied on a total sample of 600 male and female students, with 200 students assigned to each form. Data obtained for each form of the test were analyzed separately using (BIGSTEPS) and (BILOG-MG). The study findings revealed the following: there were no statistically significant differences (α=0.05) among the standard error means of item in the estimation of difficulty parameters, and no statistically significant differences (α=0.05) among a person's reliability coefficients due to number of alternatives.However, item reliability coefficients were equal and there were statistically significant differences (α=0.05) among the standard error means of a person's ability parameters, and such estimations in a person's ability were more accurate in Form One of the test than in Form Three, and in Form Two of the test than in Form Three whereas such estimations were similar in Forms One and Two. In addition to that, Form One of the test yielded more information at the low ability level but at the medium and high ability levels, the most information was in Form Three of the test, and there were statistically significant differences (α=0.05) among criterion validity coefficients in favor of Form Two of the test. **(keywords:** Number of Alternatives, Multiple-choice Test, Rasch Model, Information Function, Psychometric Properties, Item Response Theory (IRT)).*

مقدمة: تعد الاختبارات من أهم الأدوات التي تزودنا بالبيانات التي يستند إليها في اتخاذ العديد من القرارات المهمة التي تخص الفرد والمجتمع، فالاختبار إجراء منظم لقياس عينة من سلوك الأفراد، وانتشر استخدام الاختبارات انتشاراً واسعاً في العديد من المجالات، حيث تصمم هذه الاختبارات لأهداف متنوعة مثل اختيار شخص لوظيفة ما من بين مجموعة من المتقدمين لهذه الوظيفة، أو لأغراض التصنيف كتحديد مسار الطلاب بما يتناسب مع قدراتهم ومهاراتهم، وفي تقويم تحصيل الطلبة من خلال الدرجات التي يحصلون عليها في الاختبارات الصفية وغير ذلك الكثير (Allen &Yen, 1979). 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 و المجتمع إلى حد قد يصعب معالجته، أو قد يحتاج علاجه إلى وقت طويل مما يعيق عملية التنمية ومواكبة تطور المجتمعات الأخرى، وللحصول على قرارات أكثر دقة يجب توفر معلومات صادقة ودقيقة من خلال التخطيط و الإعداد الجيد للاختبار (عودة، 2002).

وهناك أشكال عدة لهذه الاختبارات منها الاختبارات ذات الإجابة المختارة والاختبارات ذات الإجابة المنتجة. ولفقرات الإجابة المختارة أشكال متعددة تستخدم لقياس تحصيل الطلبة، ولكن فقرات الاختيار من متعدد تتمتع بانتشار واسع، وتستخدم لأغراض

* كلية التربية، جامعة اليرموك، إربد، الأردن.
** كلية الدراسات التربوية، الجامعة العربية المفتوحة/فرع الأردن.
© حقوق الطبع محفوظة لجامعة اليرموك، 2009، إربد، الأردن.

متنوعة، كما أشارت نتائج الأبحاث التي قارنت فقرات الاختيار من متعدد مع غيرها من أنواع الفقرات مثل فقرات الصواب والخطأ وفقرات التكميل أن فقرات الاختيار من متعدد تستطيع قياس التحصيل بشكل أكثر صدقاً وثباتاً وأكثر كفاءة وفاعلية (Frisbie & Sweeney, 1982). ومما لاشك فيه أن فقرات الاختيار من متعدد من أكثر أشكال الأسئلة التي لها استخدام واسع في الاختبارات التحصيلية واختبارات الاستعداد، وعند بناء مثل هذه الاختبارات ينتج تساؤل هام عن عدد البدائل التي يجب أن تتضمنها كل فقرة، حيث أن زيادة عدد البدائل يؤدي إلى زيادة صعوبة الفقرة وثبات الاختبار كما أن فرصة التخمين تقل. (Budescu & Nevo, 1985)

وقد تنبأ إيبل (Ebel,1969) من خلال الصيغة الرياضية التي طورها بأن التحسن في ثبات الاختبار الموضوعي يزداد بزيادة عدد البدائل من بديلين إلى ثلاثة، ولكن يكون هذا التحسن صغيراً عند زيادة عدد البدائل إلى أربعة أو أكثر. ومن خلال الدراسات التجريبية التي قام بها كل من كوستين (Costin,1972) وستراتون وكاتس (Straton & Catts,1980) وجدوا بأن الاختبار المكون من ثلاثة بدائل أكثر ثباتاً من بديلين أو أربعة أو خمسة بدائل لكل فقرة من فقرات الاختبار. وفي دراسة أخرى لكوستين (Costin,1970) التي تم فيها حساب درجة الصعوبة ودرجة التمييز للفقرات وكذلك ثبات الاختبار بالاعتماد على مفاهيم النظرية التقليدية (الكلاسيكية) في القياس ومقارنتها باختباري الاختيار من متعدد (ذي الثلاثة بدائل، وذي الأربعة بدائل)، كانت درجة الصعوبة ودرجة التمييز للفقرات ذات الثلاثة بدائل أعلى منها للفقرات ذات الأربعة بدائل، وكذلك وجد أن ثبات الاختبار ذي الثلاثة بدائل أعلى منه للاختبار ذي الأربعة بدائل.

وتوصلت دراسة كريهان وهالدينا وبريور (Crehan, Haladyna & Brewer, 1993) إلى أن الفقرات ذات البدائل الثلاثة أصعب من الفقرات ذات البدائل الأربعة، بالإضافة إلى عدم وجود أثر ذي دلالة لعدد البدائل على تمييز الفقرات. أما دراسة تريفيسان وساكس وميشيل (Trevisan ,Sax & Micheal, 1994) فلم تجد فروق ذات دلالة إحصائية بين معاملات الثبات تعزى لعدد البدائل، بينما أوصت دراسة الشبلي (1995) باستخدام فقرات بثلاثة بدائل لبناء أسئلة الاختيار من متعدد.

فقد أظهرت الدراسات السابقة أنه لا يوجد اتفاق على العدد المثالي للبدائل، كما أنها لم تدرس أثر عدد البدائل لكل فقرة، وقدرة الفرد على ثبات الاختبار، كما أن هذه الدراسات قامت بتحليل البيانات المستمدة من الاختبارات في ضوء النظرية التقليدية التي خدمت المتخصصين في القياس ومستخدمي الاختبار بشكل جيد لفترة طويلة، ولكن هذه النظرية عانت من بعض جوانب القصور، ومنها: اعتماد إحصائيات الفقرات الشائع استخدامها، مثل: الصعوبة والتمييز على عينة الأفراد المستخدمة للحصول على هذه الإحصائيات والتي تؤثر بدورها على متوسط مستويات القدرة، كما أن درجات الأفراد في الاختبار تعتمد على عينة الفقرات التي اشتمل عليها الاختبار، بالإضافة إلى أن مفهوم ثبات الاختبار يعرّف على أساس أن صور الاختبارات متوازية والتي يصعب الحصول عليها عملياً، كما أنها تفترض تساوي تباين خطأ القياس لجميع الأفراد رغم أنه يختلف باختلاف مستوى القدرة (Hambleton & Swaminathan, 1985).

وقارنت دراسة هامبلتون وجونز (Hambleton & Jonse, 1993) بين النظريتين التقليدية ونظرية الاستجابة للفقرة، وبينت جوانب القصور في النظرية التقليدية، وكيف جاءت نظرية الاستجابة للفقرة لتعالج هذه الجوانب، وأشارت إلى التفوق الذي تظهره نظرية الاستجابة للفقرة في حل المشكلات التي تعجز عن حلها النظرية التقليدية. أما دراسة هوانج (Hwang, 2002) التي قامت خلالها بالمقارنة التحليلية والتجريبية بين إحصائيات الفقرة والأفراد خلال النظرية التقليدية ونظرية الاستجابة للفقرة مستخدمةً اختبار رياضيات من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة، حيث أظهرت النتائج أن افتراضات نظرية الاستجابة للفقرة قوية، وأكثر قوة في القياس ويتمثل ذلك في تحرر القياس من قدرة العينة التي تؤدي الاختبار، وتحرر القياس من الفقرات التي يجيب عنها الأفراد، إضافة إلى خاصية اللاتغير لتقديرات معالم الفقرة والقدرة والتي تفتقر إليها النظرية التقليدية، وما لها من أهمية في التطبيق للاختبارات التكيفية، وبنوك الفقرات، وتحيز الفقرات.

وتقوم نظرية الاستجابة للفقرة (Item Response Theory) على مجموعة من الافتراضات الأساسية وهي: افتراض أحادية البعد (Unidimensionality)، وافتراض الاستقلال الموضعي (Local Independence)، والعلاقة الوتيرية بين القدرة والأداء على الفقرة (Item Characteristic Curve)، وافتراض التحرر من السرعة في الأداء (Speededness). وقد انبثق عنها مجموعة من النماذج منها النموذج اللوجستي أحادي المعلمة (نموذج راش) (Rasch Model)، الذي استخدم في تحليل استجابات أفراد عينة الدراسة. حيث يعتبر هذا النموذج من أبسط هذه النماذج وأكثرها شهرة، والدالة الرياضية التي تعبر عنه تربط بين احتمال إجابة الفرد إجابة صحيحة على الفقرة ومعلمة صعوبة الفقرة، ومنحنيات خصائص الفقرة لهذا النموذج تختلف فقط في موقعها على متصل السمة. كما يفترض أن جميع الفقرات تتساوى في قوتها التمييزية وأن معلمة التخمين تساوي (صفراً). والدالة الرياضية التي تمثل هذه الاحتمالية يعبر عنها كالتالي:

$$p_i(\theta) = \frac{1}{1+e^{-Da(\theta-b_i)}} \quad \ldots\ldots\ldots\ldots (i = 1,2,\ldots..n)$$

حيث:

$P_i(\theta)$: احتمال إجابة الفرد الذي مستوى قدرته (θ) بشكل صحيح عن الفقرة (i).

a: ميل منحنى خصائص الفقرة عند نقطة انعطاف المنحنى (معلم تمييز الفقرة)، ويفترض في نموذج راش أن تكون قيمة التمييز ثابتة لجميع الفقرات وفي الغالب تكون قيمتها مساوية الواحد الصحيح.

b_i: النقطة التي تقع على متصل القدرة عندما يكون ميل منحنى خصائص الفقرة أكبر ما يمكن (معلم الصعوبة للفقرة).

ويشير هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) أن إحدى الطرق التي تمكننا من خلال نماذج السمات الكامنة بناء اختبارات أفضل، بحيث تعطي معلومات أكثر حول السمة الكامنة المقدّرة هو استخدام اقتران المعلومات، ويمكن استخدامها أيضا للتأكد من دقة تقدير معالم الفقرة ومعلمة القدرة، وتأتي أهمية اقتران معلومات الفقرة من كون مساهمة كل فقرة باقتران معلومات الاختبار يتحدد بشكل مستقل عن باقي الفقرات في الاختبار، أما في النظرية التقليدية فان مساهمة الفقرة بثبات الاختبار وصدقه لا يتحدد بشكل مستقل عن باقي فقرات الاختبار. واقتران معلومات الاختبار هي مقدار يتناسب عكسيا مع الخطأ المعياري في التقدير، و بذلك تعتبر تقديراً لثبات الاختبار لأنه بزيادة الخطأ المعياري في التقدير يقل الثبات والعكس صحيح. وفي حالة نموذج راش يمكن كتابة اقتران المعلومات للفقرة كالآتي:

$$I_i(\theta) = P_i(\theta)[1 - P_i(\theta)]$$

حيث:

Pi(θ):احتمال الإجابة عن الفقرة(i) إجابة صحيحة.

أما اقتران معلومات الاختبار فهي عبارة عن ناتج جمع اقترانات المعلومات للفقرات جميعها و يمكن التعبير عنها بالصيغة الآتية:

$$I(\theta) = \sum I_i(\theta)$$

حيث:

Ii(θ): اقتران معلومات الفقرة .(i)

I(θ): اقتران معلومات الاختبار.

والمعلومات التي يقدمها الاختبار ليس من الضروري أن تكون ثابتة باختلاف قيمة (θ)، وهي عند مستوى قدرة معين تعتمد على خصائص الفقرات المستخدمة لتقدير(θ) فعلى سبيل المثال المعلومات التي تزودنا بها عند المستويات العليا من القدرة تكون قليلة باستخدام فقرات سهلة جداً لأن كل ذوي القدرة المرتفعة تقريباً سيجيبون عنها إجابة صحيحة، ويمكن للفقرات نفسها أن تزودنا بكمية أكبر من المعلومات عن القدرة لذوي القدرة المنخفضة (Hulin, Drasgow & Parsons,1983). حيث تقيس الفقرة القدرة بدرجة عالية من الدقة عند مستوى القدرة المتوافق مع معلمة صعوبة الفقرة، وتنخفض كمية معلومات الفقرة مع ابتعاد مستوى القدرة عن صعوبة الفقرة، ويقترب من الصفر عند أطراف تدريج القدرة. ونظراً لأن الاختبار مكون من مجموعة من الفقرات، لذلك فان المعلومات التي يقدمها الاختبار عند أي مستوى قدرة هي مجموع مقدار معلومات الفقرات عند ذلك المستوى. لذا فان المستوى العام لاقتران معلومات الاختبار سوف يكون أكبر من المعلومات لفقرة مفردة. وبناءاً عليه فإن الاختبار يقيس القدرة بدرجة عالية من الدقة أكثر من فقرة مفردة. وكلما زاد عدد فقرات الاختبار تزداد كمية المعلومات التي يقدمها الاختبار، لهذا يتوقع أن الاختبارات المحتوية على عدد أكبر من الفقرات تقيس قدرة الفرد بدرجة عالية من الدقة أكثر من الاختبارات القصيرة. إن اقتران معلومات الاختبار مفيد جداً في نظرية الاستجابة للفقرة، فهو يخبرنا عن قدرة الاختبار في تقدير قدرة الأفراد.

وباستخدام نموذج راش ذي المعلمة الواحدة، فان دالة معلومات الفقرة تزداد مع زيادة مستوى القدرة حتى تصل إلى القيمة العظمى والتي تكون عند مستوى القدرة المتوافق مع صعوبة الفقرة، وبعد هذه النقطة يبدأ اقتران المعلومات بالانخفاض، وينسجم ذلك مع ما توصلت إليه دراسة جودوين (Jodoin, 2003) ودراسة الشريفين (2006) واللتين توصلتا إلى أن قيمة اقتران المعلومات لاختبار الاختيار من متعدد تتزايد تدريجياً بتزايد القدرة حتى تصل إلى أقصى قيمة ممكنة لها عندما يكون مستوى القدرة متوافق مع صعوبة الفقرة (θ=b). كما جاء في توصيات دراسة ليفين ودراسجو (Levine & Drasgow, 1983) أنه بزيادة عدد بدائل الفقرة يمكن الحصول على أكبر قدر ممكن من المعلومات حول الأفراد ذوي القدرة المنخفضة، أما بالنسبة لذوي القدرات المرتفعة فانه ينصح بالتقليل من عدد بدائل الفقرة.

ويشير كروكر و ألجينا (Crocker & Algina,1987) أن اقتران المعلومات له فوائد متعددة مثل مقارنة اختبارات مختلفة تقيس السمة نفسها، أو مقارنة طريقتين مختلفتين لتصحيح نفس الاختبار، بالإضافة إلى مقارنة استراتيجيات مختلفة في تشكيل اختبارات مختلفة مثل المقارنة بين الاختبار التقليدي والاختبار التكيفي وذلك من خلال كمية المعلومات التي يعطيها كل اختبار منها. كما أشار هامبلتون وسوامينثان (Hambleton & Swaminathan,1985) إلى أن كمية المعلومات التي تعطيها مجموعة من الفقرات عند مستوى قدرة معين تتناسب عكسياً مع الخطأ المعياري في تقدير القدرة، ويمكن توضيح ذلك من خلال العلاقة الآتية:

$$SE(\theta) = 1/[I(\theta)]^{1/2}$$

حيث SE(θ): الخطأ المعياري في التقدير.

I(θ): اقتران المعلومات للاختبار.

ويمكن الاستفادة من اقتران المعلومات للاختبار في حساب الكفاءة النسبية للاختبار مقارنة باختبار آخر. وقد عرف لورد (Lord,1980) الكفاءة النسبية للاختبار Y بالنسبة للاختبار X بأنها النسبة بين اقتران المعلومات للاختبارين ويعبر عنها كالتالي:

$$RE\{Y,X\} = \frac{I\{\theta,Y\}}{I\{\theta,X\}}$$

حيث I{θ,Y}، I{θ,X}: هي اقتران المعلومات للاختبارين X، Y على الترتيب.

ودرجات الاختبارين X, Y يمكن أن تكون الدرجات في اختبارين مختلفين عند مستوى القدرة نفسها، أو نحصل على هذه الدرجات من الاختبار نفسه بطريقتين. وتعرف الكفاءة النسبية فقط عندما نأخذ النسبة بين اقتران المعلومات للاختبارين X, Y عند مستوى القدرة نفسه، وتختلف الكفاءة النسبية للاختبارين باختلاف مستوى القدرة.

ويرى تريفسان وساكس وميشيل (Trevisan, Sax & Michael, 1991) أن التنبؤ بقدرات الأفراد على الاختبار يكون أكثر دقة إذا ربطنا ذلك بمستويات القدرة، حيث وجدوا أن الاتجاه المفضل لعدد البدائل هو أربعة بدائل، تليه ثلاثة بدائل، ثم خمسة بدائل للطلبة ذوي القدرة المنخفضة، أما بالنسبة لذوي القدرة المرتفعة وذوي القدرة المتوسطة لا يوجد تفضيل لعدد معين من البدائل، حيث لم توجد فروق دالة إحصائياً، وكان العدد المثالي للبدائل عبر مجموعات القدرة هو ثلاثة بدائل وذلك باستخدام النموذج اللوجستي ثلاثي المعلمة. أما في دراسة جريين وساكس وميشيل (Green, Sax & Michael, 1982) تم مقارنة اختبارات الاختيار من متعدد ذات الثلاثة بدائل, وذات الأربعة بدائل, وذات الخمسة بدائل, حيث وجدت فروق دالة إحصائياً لمعاملات ثبات الاختبار المقدرة باستخدام معادلة كودر-ريتشاردسو (KR-20) لذوي القدرة المنخفضة والمتوسطة فقط بتفضيل أربعة بدائل، بينما لذوي القدرة المرتفعة فلا توجد فروق دالة إحصائياً.

وتأتي أهمية هذه الدراسة في أنها تهدف إلى معرفة أثر عدد البدائل على متغيري صعوبة الفقرة واقتران المعلومات للفقرات والاختبار بالإضافة إلى قدرة الفرد، حيث تم تحليل البيانات المستمدة من صور الاختبار بطريقة أكثر موضوعية وهي استخدام النموذج اللوجستي أحادي المعلمة (نموذج راش) الذي يكسب الدراسة أهمية أخرى لما يتميز به من إمكانية تقدير صعوبة الفقرات بطريقة مستقلة عن قدرات الأفراد، وكذلك تقدير قدرات الأفراد بطريقة مستقلة عن صعوبة الفقرات. كما أنه يلاحظ من الدراسات السابقة الذكر أنه لا يوجد اتفاق على عدد مثالي للبدائل.

مشكلة الدراسة وأسئلتها:

نظراً للانتشار الواسع لاستخدام اختبارات الاختيار من متعدد، وخاصة في المجال التعليمي؛ حيث يفضّل المعلمون استخدام اختبارات الاختيار من متعدد بأربعة أو خمسة بدائل لفقراتها، لاعتقادهم أنها تميّز بشكل أفضل بين الطلبة، بينما لم تتفق الدراسات التي بحثت في العدد المثالي للبدائل على عدد البدائل الأفضل، كما أشارت إلى أن عدد البدائل الأفضل يتباين بتباين قدرات الطلبة، كما أن هذه الدراسات استخدمت النظرية الكلاسيكية في تحليل تلك الاختبارات والتي تعاني من عدة أوجه قصور أبرزها اختلاف قيم معالم الفقرات مثل الصعوبة والتمييز باختلاف عينة الأفراد رغم أن هذه الدراسات اعتبرت متوسطات الصعوبة ومتوسطات التمييز مؤشراً على أفضلية عدد بدائل معين، وأما هذه الدراسة فقد استخدمت نموذج راش في تحليل الاستجابات على صور الاختبار الثلاث، والذي يتميز بأن تقديرات قدرات الأفراد متحررة من صعوبة الفقرات، وتقديرات صعوبة الفقرات متحررة من قدرات عينة الأفراد، كما هدفت هذه الدراسة لمعرفة أثر عدد البدائل في اختبار الاختيار من متعدد على تقديرات القدرة للأفراد والصعوبة للفقرات المقدّرة وفقاً لنموذج راش، وكذلك أثرها على اقتران المعلومات للفقرة والاختبار الذي يعد مؤشراً على ثبات الاختبار، ثم تحديد أي من صور الاختبار الثلاث أكثرها كفاءة، وذلك بعد التحقق من افتراضات نموذج راش. وبالتحديد فان هذه الدراسة سعت للإجابة عن الأسئلة الآتية:

1. هل تختلف دقة تقدير معلمة صعوبة الفقرة المقدرة باستخدام نموذج راش باختلاف عدد بدائل فقرات الاختبار؟
2. هل تختلف دقة تقدير معلمة قدرة الفرد المقدرة باستخدام نموذج راش باختلاف عدد بدائل فقرات الاختبار؟
3. هل تختلف معاملات صدق المحك باختلاف عدد بدائل فقرات الاختبار؟
4. هل تختلف قيم اقتران المعلومات للفقرة والاختبار عند كل مستوى من مستويات القدرة باختلاف عدد بدائل فقرات الاختبار؟
5. هل تختلف الكفاءة النسبية للصورتين الثانية والثالثة للاختبار بالنسبة للصورة الأولى عند كل مستوى من مستويات القدرة باختلاف عدد بدائل فقراته؟

تعريف المصطلحات

اختبار الاختيار من متعدد:

اختبار تتكون كل فقرة من فقراته من أرومة وله بدائل عدة، يعطى الفرد العلامة (1) عند اختيار البديل الصحيح و العلامة (0) عند اختيار البديل الخاطئ. وله ثلاثة صور تختلف فقط في عدد البدائل: ثلاثة، أربعة، خمسة بدائل لكل فقرة من فقرات الاختبار.

صعوبة الفقرة:

القيمة التي تندرج على متصل السمة ويكون احتمال إجابة الفقرة عندها إجابة صحيحة مساويا 50%.

قدرة الفرد:

قيمة يتم تقديرها تعظم أرجحية استجابات الفرد عن فقرات الاختبار.

اقتران المعلومات للفقرة:

اقتران رياضي يبين مدى مساهمة الفقرة في اقتران معلومات الاختبار بشكل مستقل عن الفقرات الأخرى للاختبار.

اقتران المعلومات للاختبار:

اقتران رياضي يعبر عن مجموع اقترانات المعلومات لجميع فقرات الاختبار عند مستوى معين من القدرة.

منهجية الدراسة وإجراءاتها:

مجتمع الدراسة وعينتها:

يتكون مجتمع الدراسة من طلاب وطالبات الصف العاشر الأساسي للعام الدراسي 2006-2007 التابعين لمدارس مديرية التربية والتعليم لمنطقة اربد الأولى في الأردن.

ومن خلال الرجوع إلى التقرير الإحصائي السنوي في مديرية التربية والتعليم لمنطقة إربد الأولى تم الحصول على قائمة بأسماء مدارس الذكور وأخرى بأسماء مدارس الإناث، بعد ذلك تم تحديد المدارس التي توجد فيها شعب للصف العاشر الأساسي، وكان عدد هذه المدارس (76) مدرسة منها (35) مدرسة للذكور و(41) مدرسة للإناث، وتم اختيار عينة الدراسة بالطريقة العشوائية

العنقودية حيث اشتملت عينة الدراسة على (3) مدارس يوجد فيها شعب للصف العاشر، وتم تحديد عدد الشعب والطلبة في كل مدرسة، وطبق الاختبار على جميع الشعب الموجودة في كل مدرسة من المدارس التي تم اختيارها والتي بلغ عددها (15) شعبة لأن علامة الاختبار سيتم احتسابها ضمن علامات الطلبة وذلك لضمان جدية الطلبة في الإجابة على فقرات الاختبار، وكان عدد أفراد العينة (602) طالباً وطالبة ولكن تم حذف ورقة عشوائياً من كلتا الصورتين ذوات الأربعة بدائل وذوات الخمس بدائل لأغراض التحليل وذلك للحصول على أعداد متساوية من الطلبة لكل نموذج حيث أصبح عدد أفراد العينة التي حللت استجاباتهم (200) طالباً وطالبة لكل صورة من صور الاختبار الثلاث.

أداة الدراسة:

تم بناء الاختبار التحصيلي في مبحث الرياضيات وفقاً للخطوات الآتية، وبالاعتماد على خطوات بناء الاختبار التحصيلي التي أوردها غرونلند ولين (Gronlund & Linn, 1990) كما يلي:

أولاً: تحديد الغرض من الاختبار وهو قياس تحصيل الطلبة في وحدتي الهندسة التحليلية والإحصاء والاحتمالات لمادة الرياضيات، بافتراض أن المعارف أو المهارات في هاتين الوحدتين تقيسان سمة أو قدرة أحادية البعد (Unidimensional)، حيث يتم تدريس هذه المعارف في كتاب الرياضيات المقرر لطلبة الصف العاشر، وهو ما تم التحقق منه عند التحقق من افتراضات نموذج راش.

ثانياً: كتابة أهداف المحتوى و عمل جدول مواصفات، تم فيه ربط مستويات الأهداف بمحتوى المادة الدراسية موضوع الاختبار، ويوضح الملحق (1) جدول المواصفات الذي بني عليه الاختبار.

ثالثاً: تم كتابة (60) فقرة من نوع الاختيار من متعدد لكل فقرة منها خمسة بدائل أحدها الإجابة الصحيحة، وقد روعيت فيها الأسس الفنية في كتابة هذا النوع من الفقرات، وملاءمتها للهدف الذي تقيسه من حيث المحتوى والمستوى المعرفي، كما روعي أن لا تعتمد إجابة إحدى هذه الفقرات على إجابة الفقرات الأخرى.

رابعاً: عرض الاختبار بصورته الأولية (60 فقرة) على مجموعة من المحكمين من أصحاب الاختصاص والخبرة للتأكد من صدق المحتوى للاختبار، وطلب من كل منهم على حدة أن يبدي رأيه بالإجابة عن فقرات الاستبيان المعدّ لذلك، حول طول الاختبار ودقة الصياغة ومدى تمثيل الفقرات للأهداف وبناءً على آرائهم تم إجراء التعديلات اللازمة على فقرات الاختبار، وتم حذف الفقرات التي أجمع المحكمون على أن هناك فقرات أخرى تؤدي الهدف نفسه كما تم الأخذ برأي المحكمين بان الاختبار طويل ولا يناسب الزمن المخصص له وبناء على ذلك تم حذف (15) فقرة، و تعديل الفقرات المتبقية وصياغتها بشكل دقيق، حيث بلغ عدد فقرات الاختبار في صورته النهائية (45) فقرة.

خامساً: التجريب الأولي للاختبار على عينة استطلاعية. حيث تم تطبيق الاختبار بصورته الأولية على مجموعة مكونة من (45) طالباً وطالبة من طلبة المدرسة النموذجية لجامعة اليرموك موزعين على شعبتين؛ شعبة للذكور وشعبة للإناث تم اختيارهما بشكل عشوائي من شعب الصف العاشر الأساسي في المدرسة، وبالاعتماد على التجريب الأولي للاختبار تم تحديد الزمن المناسب للإجابة عن فقرات الاختبار. ثم أجريت عملية التحليل لبيانات التجريب الأولي باستخدام برنامج التحليل الإحصائي (SPSS) حيث تم حساب درجة صعوبة الفقرات بإيجاد نسبة الطلبة الذين أجابوا عن الفقرة إجابة صحيحة من بين الطلبة الذين حاولوا الإجابة عن هذه الفقرة، وتم حساب القدرة التمييزية للفقرات من خلال حساب معاملات الارتباط بوينت بايسيريال بين الفقرة والاختبار، وكان ذلك بالاعتماد على مفاهيم النظرية التقليدية في القياس وذلك لأن عدد أفراد العينة الاستطلاعية (45) طالباً وطالبة وهو غير كافٍ لكي نستخدم نموذج راش في التحليل حيث أن استخدامه يتطلب حجم عينة (200) فرد كحد أدنى كما أشار بيجر (Bejar,1980). والجدول(1) يبين معاملات الصعوبة ومعاملات التمييز لفقرات الاختبار للعينة الاستطلاعية.

الجدول(1): معاملات الصعوبة ومعاملات التمييز لفقرات الاختبار للعينة الاستطلاعية

أرقام الفقرات	معامل الصعوبة	معامل التمييز	أرقام الفقرات	معامل الصعوبة	معامل التمييز	أرقام الفقرات	معامل الصعوبة	معامل التمييز
1	0.76	0.49	16	0.56	0.24	31	0.22	0.19
2	0.89	0.26	17	0.76	0.24	32	0.29	0.03
3	0.42	0.49	18	0.67	0.17	33	0.33	0.49
4	0.62	0.55	19	0.27	0.36	34	0.58	0.49
5	0.82	0.6	20	0.42	0.3	35	0.31	0.53
6	0.7	0.54-	21	0.44	0.64	36	0.42	0.56
7	0.49	0.38	22	0.31	0.39	37	0.33	0.41
8	0.58	0.59	23	0.38	0.17	38	0.36	0.58
9	0.64	0.4	24	0.2	0.28	39	0.29	0.13-
10	0.31	0.46	25	0.38	0.46	40	0.33	0.63
11	0.58	0.36	26	0.27	0.59	41	0.49	0.52

أرقام الفقرات	معامل الصعوبة	معامل التمييز	أرقام الفقرات	معامل الصعوبة	معامل التمييز	أرقام الفقرات	معامل الصعوبة	معامل التمييز
12	0.49	0.56	27	0.24	0.14	42	0.36	0.16
13	0.62	0.62	28	0.78	0.3	43	0.42	0.33
14	0.73	0.41	29	0.22	0.27	44	0.27	0.19
15	0.78	0.3	30	0.24	0.52	45	0.42	0.43

يتضح من الجدول (1) أن معاملات الصعوبة تتراوح بين (0.2 و 0.89) بوسط قدره (0.47)، أمّا معاملات التمييز فقد تراوحت بين (-0.54 و 0.64) بوسط قدره (0.36) وقد تم حذف الفقرات (2، 5، 6، 32، 39)؛ وذلك نظراً لتدني معاملات التمييز للفقرات (6، 32، 39) فقد كانت معاملات التمييز لهذه الفقرات (-0.54، 0.03، -0.13) على التوالي، أما الفقرتان (2، 5) فقد تم حذفهما نظراً لارتفاع معامل الصعوبة لهاتين الفقرتين (فقرات سهلة) فقد كانت معاملات الصعوبة لهما على التوالي (0.89، 0.82). وبعد حذف هذه الفقرات أصبح الاختبار بصورته النهائية مكون من (40) فقرة. وتم حساب معامل ثبات الاتساق الداخلي للاختبار باستخدام معادلة (كودر-ريتشاردسون-20) (KR-20)، وكانت قيمة معامل الثبات للاختبار(0.84).

سادساً: تم حذف أحد المموهات عشوائياً من بدائل كل فقرة من فقرات الاختبار ذي الخمسة بدائل لتكوين الاختبار ذي الأربعة بدائل (الصورة الثانية للاختبار)، وكذلك حذف أحد المموهات عشوائياً من بدائل كل فقرة من فقرات الاختبار ذي الأربعة بدائل لتكوين الاختبار ذي الثلاثة بدائل (الصورة الثالثة للاختبار)، ويوضح الملحق (2) فقرتين من فقرات الاختبار كما وردت في صور الاختبار الثلاث.

سابعاً: إعداد ورقة تعليمات للاختبار وورقة الإجابة النموذجية لكل صورة من صور الاختبار الثلاث.

ثامناً: إجراءات جمع البيانات: بعد تطبيق صور الاختبار الثلاث بصورتها النهائية على عينة الدراسة، تم جمع استجابات الطلبة عن صور الاختبار، وتمت مراجعة أوراق الطلبة للكشف عن الاستجابات ذات النمطية الواحدة. وبناءً على ذلك لم تستثنى أي ورقة من التحليل.

تاسعاً: تم التحقق من افتراضات نموذج راش (Rasch Model) كما يلي:

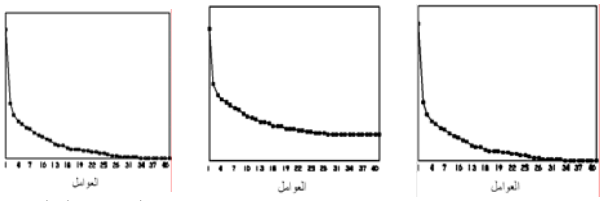
1. **افتراض أحادية البعد (Unidimensionality)**: وقد تم التحقق من هذا الافتراض باستخدام برنامج (SPSS)، وذلك لكل صورة من صور الاختبار الثلاث. حيث تم استخدام أسلوب التحليل العاملي (Factor Analysis) باستخدام تحليل المكونات الأساسية (Principle Component Analysis) لاستجابات الأفراد عن فقرات صور الاختبار الثلاث، وقد جرى التدوير باستخدام طريقة التدوير المتعامد(Varimax Rotation) للعوامل التي كانت قيم الجذر الكامن لها أكبر من واحد، وكان عددها (13) عاملاً لكل صورة من صور الاختبار، وقد تم حساب قيم الجذور الكامنة (Eigen Value) ونسبة التباين المفسر (Explained Variance) لكل عامل من العوامل، ويبين الجدول(2) قيم الجذر الكامنة ونسبة التباين المفسر للعاملين الأول والثاني، وكذلك نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني، وذلك لصور الاختبار الثلاث.

الجدول (2): قيم الجذر الكامن ونسبة التباين المفسّر للعاملين الأول والثاني، ونسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني لصور الاختبار الثلاث

الصورة		العامل		نسبة الجذر الكامن للعامل الأول إلى الثاني
		الأول	الثاني	
الأولى	الجذر الكامن	8.83	3.63	2.43
	نسبة التباين المفسر	%21.53	%8.86	
الثانية	الجذر الكامن	8.23	3.91	2.1
	نسبة التباين المفسر	%20.07	%9.53	
الثالثة	الجذر الكامن	8.82	3.74	2.36
	نسبة التباين المفسر	%21.52	%9.12	

يلاحظ من الجدول (2) أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر من (2)، وأن ما يفسره العامل الأول من التباين أكبر من (20%)، وهذان مؤشران على أحادية البعد كما أشار هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985). كما يتضح تحقق افتراض أحادية البعد من خلال التمثيل البياني باستخدام (Scree Plot) لقيم الجذور

الكامنة للعوامل المكونة لكل صورة من صور الاختبار الموضح في الشكل (1).



الصورة الأولى الصورة الثانية الصورة الثالثة

الشكل (1): التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة لكل صورة من صور الاختبار

يلاحظ من الشكل (1) استقرار في قيمة الجذر الكامن بعد العامل الأول تقريباً وهذا يدل على وجود عامل سائد على بقية العوامل وهذا مؤشر آخر على تحقق افتراض أحادية البعد كما أشار هامبلتون وسواميناثان (Hambleton & Swaminathan , 1985). كما بلغت قيمة معامل الثبات الذي تم تقديره باستخدام طريقة كرونباخ α لصور الاختبار الثلاثة (0.90، 0.91، 0.93) على الترتيب، وهي مؤشر جيد على أحادية البعد كما أشار هتي (Hattie, 1985).

2. افتراض الاستقلال الموضعي (Local Independence): وهذا الافتراض مكافئ لافتراض أحادية البعد (Hambleton & Swaminathan , 1985).

3. افتراض التحرر من السرعة في الأداء (Speededness): أنهى جميع الطلبة الإجابة عن فقرات الاختبار ضمن الوقت المحدد للإجابة، ولذلك كان إخفاق الطلبة في الإجابة عن فقرات الاختبار بسبب انخفاض قدرتهم وليس بسبب تأثير عامل السرعة.

4. افتراض تساوي القدرة التمييزية: أشار هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) إلى أنه لكي يتحقق تكافؤ مؤشرات التمييز ومطابقتها للنموذج يجب أن تكون قيمتها واقعة ضمن حدود المدى (متوسط معاملات التمييز ±0.15)، ويوضح الجدول (3) متوسطات معاملات التمييز وأدنى قيمة وأعلى قيمة لمعاملات التمييز لصور الاختبار الثلاث.

الجدول (3): متوسطات معاملات التمييز لصور الاختبار الثلاث

	الصورة الأولى	الصورة الثانية	الصورة الثالثة
متوسط معاملات التمييز	0.40	0.44	0.42
أدنى قيمة	0.30	0.32	0.31
أعلى قيمة	0.51	0.52	0.51

ويلاحظ من الجدول (3) تقارب معاملات التمييز لفقرات كل صورة من صور الاختبار، حيث كانت واقعة ضمن المدى (0.25، 0.55) لفقرات الصورة الأولى، وضمن المدى (0.29، 0.59) لفقرات الصورة الثانية، وضمن المدى (0.27، 0.57) لفقرات الصورة الثالثة. وبالتالي يمكن اعتبار معاملات التمييز متكافئة في كل نموذج، وهذا مؤشر على تحقق هذا الافتراض. ومن الجدير بالذكر أن افتراضات نظرية الاستجابة للفقرة لن تتحقق بشكل مطلق في أي مجموعة من البيانات (Albanese & Foresyth, 1984).

5. افتراض انخفاض مؤشر التخمين:

فيما يتعلق بانخفاض مؤشر التخمين، فلا توجد طريقة مباشرة لتحديد ما إذا كان الفرد أجاب عن الفقرة بالتخمين (عشوائية الاختيار) أم لا. ولكن عند رسم خطوط الانحدار غير الخطي للدرجات على الاختبار (منحنيات خصائص الفقرة)، كما في الملاحق (3، 4، 5) والتي تمثل منحنيات الخصائص لبعض الفقرات في كل صورة من صور الاختبار الثلاث على الترتيب، لوحظ أن الخط التقاربي (الأسفل) (Lower Asymptote) يقترب من الصفر في الفقرات جميعها، وهذا يحدث عندما لا يلجأ الفرد إلى الإجابة عشوائياً. وهو مؤشر على حسن مطابقة الفقرات للنموذج.

عاشراً: تم التحقق من مطابقة الاستجابات عن فقرات كل صورة من صور الاختبار الثلاث لنموذج راش (Rasch Model) كما يلي:

تم تحليل استجابات أفراد عينة الدراسة الذين أجابوا عن فقرات كل صورة من صور الاختبار الثلاث باستخدام برنامج (BIGSTEPS) المصمم لتحليل البيانات وفقاً لنموذج راش. حيث يحذف البرنامج استجابات الأفراد الحاصلين على العلامة الكاملة (40)، أو الحاصلين على العلامة (صفر)، وكذلك يحذف الفقرات التي يجيب عنها أفراد العينة جميعهم إجابة صحيحة، أو تلك التي يخفق جميع الأفراد في الإجابة عنها. ولم يكن هناك أي من الحالات الأربعة السابقة في استجابات أفراد عينة الدراسة، لذلك دخلت (40) فقرة جميعها في التحليل لاستجابات (200) فرد لكل صورة من صور الاختبار بشكل منفصل.

ولمطابقة الاستجابات عن فقرات الاختبار بصوره الثلاث مع نموذج راش، تم استخدام بعض المؤشرات الإحصائية التي يعتمدها برنامج (BIGSTEPS) لاختيار الأفراد المطابقين للنموذج، وكذلك الفقرات المطابقة للنموذج وفق ما أشار إليه (Linacre, 2002) وهي: إحصائي المطابقة الكلية للأفراد/ للفقرات (Total (ZSTD) Fit Statistic) ويعتمد هذا الإحصائي في حسابه على متوسط مربعات البواقي المعيارية (Mean Square Statistic) (MNSQ)، حيث تكون القيمة المتوقعة لهذا المتوسط أصغر أو تساوي واحد، وذلك عند تمام المطابقة. ولما كان متوسط مربعات البواقي المعيارية لا يكون إلا موجباً لذا فهو توزيع ذو ذيل واحد، وبالمثل يكون توزيع (t) للمطابقة الكلية. وينبغي نظرياً أن يكون التوزيع التقريبي لهذا الإحصائي التائي اعتدالياً، له متوسط (صفر) وانحراف معياري يساوي (واحد). ولهذا الإحصائي مؤشران؛ إحصائي المطابقة الموزون للفرد/للفقرة (The standardized Information

(Weighted Fit Statistic for Person\Item (ZSTD) الخارجية (Outfit)، والداخلية (Infit).

ويستخدم هذا الإحصائي لمطابقة الأفراد والفقرات للنموذج؛ فيما يتعلق بالأفراد يعتبر إحصائي المطابقة الداخلية (ZSTD) مؤشر إحصائي للسلوكات غير المتوقعة التي تؤثر في استجابات الأفراد عن الفقرات التي تكون قريبة من مستوى قدراتهم. وبالتالي فهو ينبئ عما إذا كان هناك شيء غير متوقع عند استجابة الفرد عن الفقرة. بينما يعتبر إحصائي المطابقة الخارجية (ZSTD) مؤشر إحصائي بديلاً عن السابق وله صفات مشابهة ولكنه أكثر حساسية للسلوكات غير المتوقعة التي تؤثر في استجابات الأفراد عن الفقرات التي تبتعد عن مستوى قدراتهم. وذلك لكل تقدير من تقديرات القدرة. أما فيما يتعلق بالفقرات يقوم إحصائي المطابقة الخارجية (ZSTD) باختبار ملاءمة الفقرة للنموذج، وذلك بوجه عام من فرد إلى فرد. فإذا حدث اتساق بين الاستجابات الملاحظة للأفراد على الفقرة واحتمال نجاحهم كما يتوقع النموذج، كان معنى هذا أن هناك اتساق بين الاستجابات الملاحظة للأفراد على هذه الفقرة ودرجاتهم الكلية على الاختبار، أي أن استجاباتهم على باقي فقرات الاختبار. وهذا يدل على الاتفاق بين السمة التي تعبر عنها الفقرة والسمة التي تعبر عنها باقي الفقرات وذلك عبر العينة كلها، ومعنى هذا ملاءمة الفقرة بوجه عام لمتطلبات النموذج. بينما يعتبر إحصائي المطابقة الداخلية (ZSTD) اختبار لمدى استقرار الصعوبة النسبي للفقرات، عبر مستويات القدرة المختلفة؛ أي اختبار مدى ثبات تدرج صعوبة الفقرات عند كل مستوى من مستويات القدرة.

ويتم حساب الإحصائيات السابقة من خلال البرنامج لكل تقدير من تقديرات القدرة للأفراد، وكذلك لكل تقدير من تقديرات الصعوبة للفقرات. حيث يعد الفرد غير مطابق للنموذج عندما تزيد قيمة إحصائي المطابقة الخارجية أو الداخلية (ZSTD)عن (2)، أو أن يكون متوسط مربعات البواقي المعيارية (MNSQ) الداخلية أو الخارجية أكبر من (1)؛ أي يقع خارج حدود المطابقة وهي (0.7-1.3). ولمعرفة المزيد عن هذه الإحصائيات والاطلاع على المعادلات الخاصة بها، يمكن الرجوع إلى (Wright, Mead & Bell,1980, P13-34; Wright & Stone, 1979, P77).

تم تحليل الاستجابات للمرة الأولى وذلك لحذف الأفراد غير المطابقين للنموذج، حيث تم تقدير قدرة كل فرد والخطأ المعياري في تقدير هذه القدرة، وكذلك قيم إحصائي المطابقة الداخلية والخارجية (ZSTD)، ومتوسط مربعات البواقي المعيارية الداخلية والخارجية (MNSQ). والجدول (4) يبين المتوسط الحسابي والانحراف المعياري لكل تقدير من تقديرات القدرة مقدرة بوحدة اللوجيت والخطأ المعياري في قياس هذه القدرة، وقيم إحصائي المطابقة الداخلية والخارجية، ومتوسط المربعات للمطابقة الداخلية والخارجية لصور الاختبار الثلاث.

الجدول (4): المتوسط الحسابي والانحراف المعياري لكل تقدير من تقديرات القدرة لأفراد كل صورة مقدرة بوحدة اللوجيت والخطأ المعياري في قياس هذه القدرة وقيم إحصائي المطابقة الداخلية والخارجية (عدد الأفراد=200، عدد الفقرات=40)

الصورة		القدرة	الخطأ المعياري	Infit إحصائي المطابقة الداخلية		Outfit إحصائي المطابقة الخارجية	
				متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
الأولى	المتوسط الحسابي	0.1	0.36	1	0	1	0
	الانحراف المعياري	0.76	0.03	0.14	1.1	0.2	1.1
الثانية	المتوسط الحسابي	0.32-	0.36	1	0.2-	1.03	0
	الانحراف المعياري	0.77	0.02	0.12	0.8	0.21	1
الثالثة	المتوسط الحسابي	0.57-	0.37	1	0.1-	0.99	0.2-
	الانحراف المعياري	0.88	0.03	0.15	1	0.22	1

ويتضح من الجدول (4) اقتراب المتوسط الحسابي لمتوسطات المربعات الداخلية والخارجية من الواحد، وهو الوضع المثالي كما يتوقعه النموذج، كما يلاحظ أن متوسط قيم إحصائي المطابقة الداخلية والانحراف المعياري لها تقترب من القيم المثالية التي يفترضها النموذج وهي (0، 1) على الترتيب، وكذلك متوسط قيم إحصائي المطابقة الخارجية والانحراف المعياري لها أيضاً تقترب من القيم المثالية التي يفترضها وهي (0، 1) على الترتيب. وبناءً على قيم إحصائي المطابقة الخارجية للأفراد، تم استبعاد استجابات (10، 12، 13) فرداً من أفراد كل صورة من صور الاختبار الثلاث على الترتيب لم تتطابق استجاباتهم الملاحظة مع الاستجابات المتوقعة تبعاً لقدراتهم، أي أنها تبتعد عن توقعات النموذج.

بعد استبعاد الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج من استجابات كل صورة، وإعادة التحليل لاختبار مدى ملاءمة الفقرات للنموذج، فقد تم تقدير معلمة الصعوبة لكل فقرة، بالإضافة إلى الخطأ المعياري في قياس الصعوبة وقيم إحصائي المطابقة الداخلية والخارجية للفقرات (ZSTD) ومتوسط المربعات الداخلية والخارجية للفقرات (MNSQ)، وذلك لكل صورة من صور الاختبار. ويبين الجدول (5) المتوسط الحسابي والانحراف المعياري لكل تقدير من تقديرات الصعوبة للفقرات مقدرة بوحدة اللوجيت والخطأ المعياري في قياس الصعوبة، وكذلك قيم إحصائي المطابقة الداخلية والخارجية.

الجدول (5): المتوسط الحسابي والانحراف المعياري لكل تقدير من تقديرات الصعوبة لفقرات كل صورة مقدرة بوحدة اللوجيت والخطأ المعياري في قياس الصعوبة وكذلك قيم إحصائي المطابقة الداخلية والخارجية (عدد الأفراد = 190، 188، 187 على الترتيب، عدد الفقرات= 40)

الصورة		الصعوبة	الخطأ المعياري	Infit إحصائي المطابقة الداخلية		Outfit إحصائي المطابقة الخارجية	
				متوسط المربعات MNSQ	قيمة الإحصائي ZSTD	متوسط المربعات MNSQ	قيمة الإحصائي ZSTD
الأولى	المتوسط الحسابي	0	0.16	1	0.2-	1	0.3-
	الانحراف المعياري	0.84	0.01	0.13	2.2	0.2	2.2
الثانية	المتوسط الحسابي	0	0.17	1	04-	1.03	0.2-
	الانحراف المعياري	0.78	0.01	0.17	2.7	0.26	2.8
الثالثة	المتوسط الحسابي	0	0.17	1	0.3-	1	0.3-
	الانحراف المعياري	0.86	0.02	0.22	3	0.3	2.8

يلاحظ من الجدول(5) اقتراب المتوسط الحسابي لمتوسطات المربعات الداخلية والخارجية من الواحد وهو الوضع المثالي كما يتوقعه النموذج، كما يلاحظ أن متوسط قيم إحصائي المطابقة الداخلية والانحراف المعياري بعيدانِ نوعاً ما عن القيم المثالية التي يفترضها النموذج وهي (0، 1) على الترتيب، وكذلك متوسط قيم إحصائي المطابقة الخارجية والانحراف المعياري لها أيضاً بعيدانِ نوعاً ما عن القيم المثالية التي يفترضها وهي (0، 1) على الترتيب.

وبناء على قيم متوسطات المربعات تم استبعاد الفقرات الواقعة خارج حدود المطابقة، وهي (0.7 - 1.3) من فقرات كل صورة، وكانت (3، 2، 7) فقرات على الترتيب من صور الاختبار الثلاث، والتي أشارت نتائج التحليل أنها غير متطابقة مع توقعات النموذج، وكانت الاستجابات الملاحظة عنها تبتعد عن الاستجابات التي يتوقعها النموذج. وأعيد التحليل مرة ثالثة بعد حذف الأفراد غير المطابقين، والفقرات غير المطابقة من كل صورة وذلك

للحصول على النتائج النهائية المتحررة من صعوبة الفقرات، وقدرات الأفراد. ويوضح الجدول (6) المتوسط الحسابي والانحراف المعياري لكل من القيم المتحررة لمعالم قدرة الأفراد والخطأ المعياري في قياس القدرة مقدرة بوحدة اللوجيت لكل صورة من صور الاختبار، ويقصد بالقيمة المتحررة لقدرة الفرد بأن قدرة الفرد لا تختلف سواء استخدمت جميع فقرات الاختبار في قياسها أو أي مجموعة فرعية منها؛ أي أنها متحررة من عينة الفقرات (Hambleton & Swaminathan, 1985).

الجدول (6): المتوسط الحسابي والانحراف المعياري لكل من القيم المتحررة لمعالم قدرة الأفراد والخطأ المعياري في قياس القدرة مقدرة بوحدة اللوجيت لكل صورة من صور الاختبار

صورة الاختبار	الأولى		الثانية		الثالثة	
	القدرة	الخطأ المعياري	القدرة	الخطأ المعياري	القدرة	الخطأ المعياري
المتوسط الحسابي	0.18	0.37	0.21-	0.37	0.48-	0.41
الانحراف المعياري	0.81	0.05	0.83	0.02	0.95	0.04

ويلاحظ من الجدول (6) أنه بالانتقال من الصورة الأولى إلى الثالثة يتناقص متوسط تقديرات القدرة للأفراد بشكل مطرد؛ حيث بلغ الفرق لمتوسط التقدير بين الصورتين الأولى والثانية (0.39) لوجيت، وبين الصورتين الثانية والثالثة (0.27) لوجيت، بينما كانت أكبر قيمة لمتوسطات الخطأ المعياري للصورة الثالثة، أما للصورتين الأولى والثانية فكانت متساوية حيث بلغت (0.37) لوجيت. كما يلاحظ أن الانحراف المعياري لتقديرات قدرات الأفراد يتزايد بشكل مطرد بالانتقال من الصورة الأولى إلى الثالثة حيث بلغ الفرق لمتوسط التقدير بين الصورة الأولى، والثانية (0.02)، وبين الصورتين الثانية، والثالثة (0.12). وتوضح الملاحق(6، 8، 10) نتائج التحليل المتحررة لقدرات أفراد الصور الثلاث على الترتيب مقدرة بوحدة اللوجيت.

كما يوضح الجدول (7) المتوسط الحسابي والانحراف المعياري لكل من القيم المتحررة لصعوبة الفقرات والخطأ المعياري في قياس الصعوبة مقدرة بوحدة اللوجيت لكل صورة من صور الاختبار، ويقصد بالقيمة التقديرية المتحررة لصعوبة الفقرات بأن تقدير معلمة صعوبة الفقرة لا يتأثر باختلاف عينة الأفراد التي تؤدي الاختبار؛ أي أن تدرج درجة صعوبة الفقرة في باقي فقرات الاختبار يبقى ثابتاً مهما اختلف الأفراد الذين يؤدون الاختبار؛ أي أنها متحررة من عينة الأفراد (Hambleton & Swaminathan, 1985).

الجدول (7): المتوسط الحسابي والانحراف المعياري لكل من القيم المتحررة لصعوبة الفقرات والخطأ المعياري في قياس الصعوبة مقدرة بوحدة اللوجيت لكل صورة من صور الاختبار

صورة الاختبار	الأولى		الثانية		الثالثة	
	الصعوبة	الخطأ المعياري	الصعوبة	الخطأ المعياري	الصعوبة	الخطأ المعياري
المتوسط الحسابي	0	0.16	0	0.17	0	0.17
الانحراف المعياري	0.78	0.01	0.75	0.01	0.82	0.02

يلاحظ من الجدول (7) أن متوسطات الصعوبة لفقرات الصور الثلاث متساوية، وتساوي (صفر) لوجيت لكل منها، كما أن متوسطات الأخطاء المعيارية لتقديرات الصعوبة متقاربة لجميع صور الاختبار. وتوضح الملاحق (7، 9، 11) نتائج التحليل المتحررة لصعوبة فقرات الصور الثلاث على الترتيب مقدرة بوحدة اللوجيت.

وللتأكد من جودة مطابقة البيانات الناتجة عن استجابات أفراد العينة على صور الاختبار الثلاث، والتحقق من موضوعية تفسير النتائج لصور الاختبار الثلاث بصورتها النهائية حيث تكونت كل صورة من (37، 38، 33) فقرة على الترتيب، أي بعد حذف الأفراد غير المتطابقين، والفقرات غير المطابقة، وكذلك التحقق من افتراضات نموذج راش، حيث يعد ذلك تحقيقاً لمدى صدق نموذج راش في تحقيقه لموضوعية القياس، ولقد أشارت النتائج إلى تحقق شروط الموضوعية في صور الاختبار والتي تتلخص بما يلي:

إن متوسط تقدير قدرات الأفراد المستجيبين على الصورة الأولى للاختبار قد نقص بمقدار (0.03) لوجيت بعد حذف الأفراد غير المطابقين وازداد بمقدار (0.11) بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج معاً. وقد بلغ متوسط قدرات الأفراد (0.18) لوجيت أي أن مستوى الاختبار يصبح أكثر ملاءمة لقدرات الأفراد نظراً لاقتراب قيمة متوسط القدرة للأفراد من قيمة متوسط الصعوبة للفقرات والبالغ (صفراً) لوجيت. وهذا يعني حسن مطابقة الفقرات للنموذج والذي تؤكده قيم إحصائيات المطابقة للفقرات حيث اقتربت من الوضع المثالي، وهو أن يكون متوسط المربعات للقدرات يساوي (1) والانحراف المعياري لهذه الأخطاء يساوي (صفراً). وكذلك متوسط تقدير قدرات الأفراد المستجيبين على الصورة الثانية للاختبار قد ازداد بمقدار (0.06) لوجيت بعد حذف الأفراد غير المطابقين و ازداد بمقدار (0.05) بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج معاً. وقد بلغ متوسط قدرات الأفراد (-0.21) لوجيت؛ أي أن مستوى الاختبار يصبح أكثر ملاءمة لقدرات الأفراد نظراً لاقتراب قيمة متوسط القدرة للأفراد من قيمة متوسط الصعوبة للفقرات والبالغ (صفراً) لوجيت. وهذا يعني حسن مطابقة الفقرات للنموذج والذي تؤكده قيم إحصائيات المطابقة للفقرات، حيث اقتربت من

الوضع المثالي وهو أن يكون متوسط المربعات للقدرات يساوي (1) والانحراف المعياري لهذه الأخطاء يساوي (صفراً). وفيما يتعلق بمتوسط تقدير قدرات الأفراد المستجيبين على الصورة الثالثة للاختبار قد ازداد بمقدار (0.05) لوجيت بعد حذف الأفراد غير المطابقين وبمقدار (0.04) بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج معاً. وقد بلغ متوسط قدرات الأفراد (-0.48) لوجيت أي أن مستوى الاختبار يصبح أكثر ملاءمة لقدرات الأفراد نظراً لأن قيمة متوسط القدرة للأفراد أصبح أقرب إلى قيمة متوسط الصعوبة للفقرات والبالغ (صفراً) لوجيت من متوسط قدرات الأفراد قبل حذف الأفراد والفقرات. وهذا يعني حسن مطابقة الفقرات للنموذج والذي تؤكده قيم إحصائيات المطابقة للفقرات حيث اقتربت من الوضع المثالي وهو أن يكون متوسط المربعات للقدرات يساوي (1) والانحراف المعياري لهذه الأخطاء يساوي (صفراً). ويلاحظ أن الصورة الأولى هو أكثرها مطابقة لأن متوسط تقديرات القدرة لأفراد النموذج أكثرها قرباً من متوسط تقديرات الصعوبة لفقراته، تليها الصورة الثانية في حسن المطابقة ثم الصورة الثالثة.

كما أن الزيادة في متوسط الخطأ المعياري في القياس لتقديرات القدرة لأفراد الصور الثلاث كانت (0.01، 0.01، 0.04) لوجيت على الترتيب، ويعود السبب في ذلك إلى نقصان عدد أفراد العينة بمقدار (10، 12، 13) فرداً من أفراد كل صورة على الترتيب، وكذلك نقصان عدد فقرات الاختبار بمقدار (7.5%، 5%، 17.5%) من فقرات كل صورة على الترتيب. ويلاحظ أن أكبر زيادة في متوسط الخطأ المعياري في القياس كانت للصورة الثالثة، ويعود السبب في ذلك إلى أن نسبة الفقرات المحذوفة من الصورة الثالثة هي أكبر نسبة. وقد بلغ متوسط الأخطاء المعيارية لتقديرات قدرة الأفراد لصور الاختبار (0.37، 0.37، 0.41) على الترتيب، وهي قيم مرتفعة ولعل السبب في ذلك صغر حجم العينة. حيث أشار العبابنة (2004) في دراسته إلى أن دقة تقديرات معالم الفقرات تزداد بزيادة حجم العينة، وكذلك تزداد دقة تقدير معلم القدرة بزيادة عدد فقرات الاختبار.

وبَينت النتائج أن متوسط قيم إحصائي المطابقة الداخلية (ZSTD) لفقرات كل صورة على الترتيب (0.99، 1.02، 0.99) والانحراف المعياري لها (0.16، 0.23، 0.21) وهي تقترب من القيم المثالية التي يفترضها النموذج وهي (1، 0) على الترتيب. وهذا الإحصائي يشير إلى مدى استقرار مستوى الصعوبة النسبي للفقرات عبر مستويات القدرة المختلفة، مما يدل على تدرج صعوبة الفقرات عند كل مستوى من مستويات القدرة. وهذا مؤشر على تحقق أحد افتراضات نموذج راش وهو استقلال صعوبة الفقرات عن عينة الأفراد، وبالتالي تكون الفقرات مطابقة للنموذج (Wright, & Mead, 1980).

كما أن متوسط قيم إحصائي المطابقة الخارجية (ZSTD) لأفراد كل صورة قد بلغ (0، 0، -0.1) على الترتيب والانحراف المعياري له (1، 0.9، 0.8) على الترتيب وهي تقترب من القيم المثالية التي يفترضها النموذج وهي (0، 1) على الترتيب. وهذا مؤشر على عدم اختلاف الصعوبة النسبية لفقرات كل صورة عند معظم الأفراد المستجيبين عن فقراته؛ وهذا يعني تحرر تقديرات قدرات الأفراد لكل صورة من صعوبة الفقرات.

وأشارت النتائج إلى أن القيم التقديرية المتحررة لصعوبة فقرات الصورة الأولى قد تراوحت بين (-1.81، 1.66) وحدة لوجيت، بمتوسط قدره (صفر) لوجيت، وانحراف معياري قدره (0.78)، وقد بلغ متوسط الأخطاء المعيارية لصعوبة الفقرات (0.16) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تقديرات صعوبة الفقرات. وأن القيم التقديرية المتحررة لصعوبة فقرات الصورة الثانية قد تراوحت بين (-1.66، 1.43) وحدة لوجيت، بمتوسط قدره (صفر) لوجيت وانحراف معياري قدره (0.75)، وقد بلغ متوسط الأخطاء المعيارية لصعوبة الفقرات (0.17) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تقديرات صعوبة الفقرات. كما أن القيم التقديرية المتحررة لصعوبة فقرات الصورة الثالثة قد تراوحت بين (-1.77، 2) وحدة لوجيت، بمتوسط قدره (صفر) لوجيت وانحراف معياري قدره (0.82)، وقد بلغ متوسط الأخطاء المعيارية لصعوبة الفقرات (0.17) وهي قيمة متدنية، الأمر الذي يشير إلى دقة تقديرات صعوبة الفقرات.

عرض النتائج ومناقشتها

بعد التحقق من مطابقة الأفراد والفقرات لنموذج راش، تم عرض النتائج المتعلقة بالتساؤلات التالية ومناقشتها على النحو الاتي:

1. النتائج المتعلقة بدقة تقدير معلمة صعوبة الفقرة باستخدام نموذج راش لصور الاختبار الثلاث.

لمعرفة قيم معالم الفقرات ومعالم الأفراد لكل صورة من صور الاختبار بصورتها النهائية، تم تقدير هذه المعالم كما ذكر سابقاً باستخدام برنامج (BIGSTEPS)، حيث كان المتوسط الحسابي لصعوبة فقرات كل صورة (صفراً) لوجيت وهو ثابت لجميع صور الاختبار. وكانت المتوسطات الحسابية للأخطاء المعيارية لتقديرات معالم الصعوبة لفقرات كل صورة (0.16، 0.17، 0.16) على الترتيب، وهي قيم متدنية تشير إلى دقة تقدير معالم الصعوبة لفقرات كل صورة. وللكشف عن الفروق في دقة تقدير معالم صعوبة الفقرات لصور الاختبار الثلاث تم إجراء تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير هذه المعالم باستخدام برنامج (SPSS) كما هو موضح في الجدول (8).

الجدول (8): نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير معالم الصعوبة لصور الاختبار الثلاث

	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الإحصائية
بين المجموعات	0.0009	2	0.00043	2.53	0.082
داخل المجموعات	0.02	105	0.00017		
الكلي	0.0209	107			

ويلاحظ من الجدول (8) عدم وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (α=0.05) بين تقديرات معالم الصعوبة تعزى لعدد البدائل. وهذا يتعارض مع دراسة بيدسكو ونيفو (Budescu & Nevo, 1985) التي أشارت إلى تزايد معلمة الصعوبة بزيادة عدد بدائل فقرات الاختبار، وكذلك تعارضت مع: دراسة الشبلي (1995)، ودراسة كريهان وهالدينا وبريور (Crehan, Haladyna & Brewer, 1993)، التي أشارت إلى أن الفقرات ذات الثلاثة بدائل هي الأكثر صعوبة.

2. النتائج المتعلقة بدقة تقدير معلمة قدرة الفرد المقدّرة باستخدام نموذج راش لصور الاختبار الثلاث.

تم تقدير القيم المتحررة لقدرات الأفراد والأخطاء المعيارية في تقديرها لكل صورة من صور الاختبار. حيث بلغت متوسطات معالم القدرة لصور الاختبار الثلاث (0.18، -0.21، -0.48) على الترتيب، والمتوسط الحسابي للأخطاء المعيارية (0.37، 0.37، 0.41) على الترتيب. وكانت قيم متوسطات الأخطاء المعيارية مقبولة لجميع صور الاختبار وهذا يشير إلى دقة تقدير السمة لمعظم الأفراد في كل صورة. وكانت الأخطاء المعيارية تقل كلما اقتربت تقديرات معالم القدرة من الصفر؛ أي تزداد الدقة في التقدير لمعالم القدرة كلما اقتربت متوسطات تقديرات القدرة للأفراد من متوسطات تقديرات الصعوبة للفقرات، وذلك لأن مستوى الصعوبة لفقرات الاختبار تصبح أكثر ملاءمة لمستوى قدرات الأفراد المستجيبين عنها.

وأشارت نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير قدرات الأفراد لصور الاختبار الثلاث الموضحة في جدول (9)، إلى وجود فروق ذات دلالة إحصائية (α=0.05) بين متوسطات الأخطاء المعيارية في تقدير قدرات الأفراد تعزى لعدد البدائل، حيث كان متوسط الأخطاء المعيارية للصورة الثالثة أعلى منه للصورتين الأولى والثانية.

الجدول (9): نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير معالم القدرة لصور الاختبار الثلاث

	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الإحصائية
بين المجموعات	0.209	2	0.104	67.331	0.00
داخل المجموعات	0.869	561	0.00155		
الكلي	1.078	563			

وتم استخدام طريقة شفيه (Scheffe's Method) لعقد المقارنات البعدية بين متوسطات الأخطاء المعيارية لتقديرات القدرة لأفراد الصور الثلاث لمعرفة أي الفروق دالة إحصائياً، ويبين الجدول (10) نتائج المقارنات الثنائية بين متوسطات الأخطاء المعيارية.

الجدول (10): نتائج المقارنات الثنائية بين متوسطات الأخطاء المعيارية لتقديرات القدرة لصور الاختبار الثلاث

صورة الاختبار	متوسط الخطأ	الفروق في المتوسطات	الدلالة
الأولى	0.37	0.00317	0.736
الثانية	0.37		
الثانية	0.37	* 0.0424	0.00
الثالثة	0.41		
الأولى	0.37	* 0.0392	0.00
الثالثة	0.41		

* ذات دلالة إحصائية عند (α =0.05)

ويلاحظ من الجدول (10) أن الفروق بين متوسطات الأخطاء المعيارية كانت دالة إحصائياً بين الصورتين الأولى والثالثة ولصالح الصورة الثالثة؛ أي أن فقرات الصورة الأولى (الفقرات ذات الثلاثة بدائل) أكثر دقة في تقدير قدرات الأفراد من فقرات الصورة الثالثة (الفقرات ذات الخمسة بدائل)، كما أشارت إلى أن الفروق بين متوسطات الأخطاء المعيارية كانت دالة إحصائياً بين الصورتين الثانية والثالثة، ولصالح الصورة الثالثة؛ أي أن فقرات الصورة الثانية (الفقرات ذات الأربعة بدائل) أكثر دقة في تقدير قدرات الأفراد من فقرات الصورة الثالثة (الفقرات ذات الخمسة بدائل)، كما أشارت إلى أن الفروق بين متوسطات الأخطاء المعيارية بين الصورتين الأولى، والثانية لم تكن دالة إحصائياً؛ أي أن دقة تقدير قدرات الأفراد لا تختلف بين الصورتين الأولى والثانية باختلاف عدد البدائل من ثلاثة إلى أربعة بدائل على الترتيب.

3. النتائج المتعلقة بمعاملات صدق المحك لصور الاختبار الثلاث.

تم باستخدام برنامج (SPSS) إيجاد معاملات صدق المحك لكل صورة بإيجاد معامل الارتباط بيرسون بين العلامات الكلية للأفراد على فقرات صور الاختبار، والعلامات الكلية للأفراد في نهاية

الفصل الدراسي الأول بافتراض أنها تتمتع بدرجة مقبولة من الصدق والثبات، حيث يصعب تقدير ثبات هذه العلامات من البيانات التي توفرت للباحثين لذا حرص الباحثان وبالتنسيق مع المشرفين في مديرية التربية والتعليم لمنطقة اربد الأولى أن يكون الاختبار في نهاية الفصل الدراسي السابق موحداً لأفراد عينة الدراسة، ومنسجماً ذلك مع حرص وزارة التربية والتعليم على أن تكون الاختبارات المطبقة موحده لطلبة الصف العاشر؛ لأنه يتم تصنيف الطلبة عند الانتهاء من هذا الصف إلى أفرع التعليم المختلفة: علمي، أدبي، إدارة معلوماتية،....الخ. والجدول (11) يبين معاملات صدق المحك لصور الاختبار الثلاث.

الجدول (11): معاملات صدق المحك لصور الاختبار الثلاث

صورة الاختبار	الأولى	الثانية	الثالثة
معامل الصدق	0.815	0.913	0.861

ويلاحظ من الجدول (11) أن قيم معاملات الارتباط بين أداء الطلبة على اختبار الرياضيات المعدّ من قبل الباحثين كمتنبئ، وعلامة المحك المستخدم (التحصيل في الرياضيات) كانت جميعها موجبة ومرتفعة وهذا يعني أن نسبة التباين المشترك بين الاختبارين كانت عالية، وربما يعود ذلك إلى تشابه المحتوى بين الاختبارين، حيث بلغت أعلى قيمة(0.913) للصورة الثانية للاختبار(الفقرات ذات الأربعة بدائل)، وأدنى قيمة (0.815) للصورة الأولى للاختبار (الفقرات ذات الثلاثة بدائل)، وكانت جميعها ذات دلالة إحصائية (α =0.05).

وللإجابة عن سؤال الدراسة " ما أثر الفروق بين دلالة معاملات الصدق التنبؤية للصور الثلاث " باستخدام الإحصائي (V) الذي أشار إليه هيز (Hays, 1972, P664)، يبين الجدول (12) نتائج التحليل المتعلقة بمعاملات صدق المحك.

الجدول (12): نتائج التحليل المتعلقة بمعاملات صدق المحك

الصورة	معامل الصدق	قيمة V	(χ^2)	درجات الحرية
الأولى	0.815	15.323	5.991	2
الثانية	0.913			
الثالثة	0.861			

ويلاحظ من الجدول (12) أن قيمة (V) أكبر من قيمة (χ^2) وبالتالي يوجد فروق ذات دلالة إحصائية بين معاملات صدق المحك تعزى لعدد البدائل. حيث كان أعلى معامل صدق للصورة الثانية تليها الصورة الثالثة، ثم الصورة الأولى؛ أي أن الصورة الثانية (ذات الأربعة بدائل) كانت أكثرها دقة في قياس محتوى الاختبار.

و للكشف عن دلالة الفروق بين معاملات صدق المحك تم استخدام اختبار Z، ويبين الجدول (13) نتائج المقارنات الثنائية بين معاملات صدق المحك.

الجدول(13): نتائج المقارنات الثنائية بين معاملات صدق المحك

صورة الاختبار	معامل الارتباط	Z_j**	قيمة Z المحسوبة	قيمة Z الحرجة
الأولى	0.815	1.142	*3.912	1.96
الثانية	0.913	1.545		
الثانية	0.913	1.545	*2.401	1.96
الثالثة	0.861	1.297		
الأولى	0.815	1.142	1.502	1.96
الثالثة	0.861	1.297		

* ذات دلالة إحصائية عند (α = 0.05)

** Z_j العلامة الفشرية المقابلة لمعامل صدق المحك.

ويتضح من الجدول (13) وجود فروق ذات دلالة إحصائية في معاملات صدق المحك بين الصورتين الأولى، والثانية لصالح الصورة الثانية، وكذلك بين الصورتين الثانية، والثالثة لصالح الصورة الثانية بينما لا يوجد فرق ذو دلالة إحصائية بين الصورتين الأولى والثالثة.

4. النتائج المتعلقة بقيم اقتران المعلومات للفقرات والاختبار، والكفاءة النسبية لصور الاختبار الثلاث.

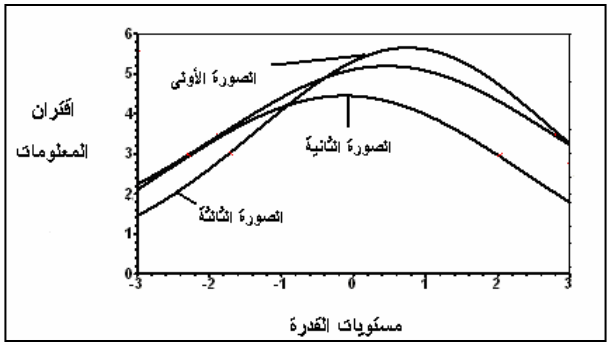
لإيجاد اقتران المعلومات للفقرات، والاختبار لصور الاختبار الثلاث تم استخدام البرنامج الإحصائي (BILOG-MG) المصمم لتحليل البيانات وفق نماذج نظرية الاستجابة للفقرة. والجدول (14) يبين تقديرات قدرات أفراد صور الاختبار الثلاث بوحدة اللوجيت وقيم اقتران المعلومات لكل صورة عند مستويات مختارة من القدرة.

الجدول (14): قيم اقتران المعلومات لكل صورة من صور الاختبار الثلاث عند مستويات مختارة من القدرة

القدرة (θ)	اقتران المعلومات صورة (1)	صورة (2)	صورة (3)	القدرة (θ)	اقتران المعلومات صورة (1)	صورة (2)	صورة (3)
-3	2.08	1.63	1.45	0.2	3.62	4.07	5.47
-2.8	2.23	1.79	1.65	0.4	3.58	4.09	5.58
-2.6	2.39	1.96	1.86	0.6	3.51	4.09	5.64
-2.4	2.55	2.14	2.10	0.8	3.43	4.06	5.65
-2.2	2.70	2.33	2.35	1	3.33	4.00	5.62
-2	2.85	2.52	2.62	1.2	3.21	3.91	5.53
-1.8	2.99	2.71	2.90	1.4	3.09	3.81	5.40
-1.6	3.13	2.90	3.19	1.6	2.95	3.68	5.22
-1.4	3.25	3.08	3.49	1.8	2.81	3.54	5.00
-1.2	3.36	3.26	3.79	2	2.65	3.37	4.75
-1	3.45	3.43	4.09	2.2	2.50	3.20	4.48
-0.8	3.53	3.59	4.37	2.4	2.34	3.02	4.18
-0.6	3.59	3.73	4.65	2.6	2.18	3.83	3.88
-0.4	3.63	3.85	4.90	2.8	2.02	2.64	3.57

القدرة (θ)	اقتران المعلومات			القدرة (θ)	اقتران المعلومات		
	صورة (1)	صورة (2)	صورة (3)		صورة (1)	صورة (2)	صورة (3)
-0.2	3.64	3.95	5.13	3	1.87	3.45	3.26
0	3.7	4.3	5.4				

ويلاحظ من الجدول (14) أن قيمة اقتران المعلومات لكل صورة اختبار تتزايد تدريجياً بتزايد القدرة (θ) حتى تصل إلى أقصى قيمة ممكنة لها عندما تكون قيمة القدرة (صفراً) لوجيت، وهي تقابل متوسط الصعوبة للفقرات (صفراً) لوجيت، وهذا يتطابق مع ما هو متوقع من نموذج راش حيث أن قيمة اقتران المعلومات تكون أقصى ما يمكن عندما تكون قيمة القدرة (θ = 0)، وهذا ما تؤكده قيمة الخطأ المعياري لتقدير صعوبة الفقرات التي كانت أقل ما يمكن عند متوسط الصعوبة لفقرات كل صورة حيث بلغت قيمته (0.15، 0.16، 0.16) لصور الاختبار الثلاث على الترتيب، وهذا يتفق مع ما أشارت إليه دراسة جودوين (Jodoin,2003)، ودراسة الشريفين (2006). ويلاحظ أن الصور الثلاث تعطي أكبر قدر من المعلومات عند مستويات القدرة المتوسطة، وتعطي أقل كمية من المعلومات عند مستويات القدرة المرتفعة والمنخفضة، بمعنى أن الصور الثلاث تعطي معلومات قليلة عن الأفراد ذوي القدرات المرتفعة والمنخفضة. كما كانت الصورة الأولى هي الأعلى في كمية المعلومات تليها الصورة الثانية، ثم الصورة الثالثة وذلك عند مستويات القدرة المنخفضة؛ أي أن الصورة الأولى (ذات الثلاثة بدائل) تعطي معلومات أكثر عن ذوي القدرة المنخفضة، أما عند مستويات القدرة المتوسطة فإن الصورة الثالثة هي الأعلى في كمية المعلومات تليها الصورة الثانية ثم الصورة الأولى؛ أي أن الصورة الثالثة (ذات الخمسة بدائل) تعطي معلومات أكثر عن ذوي القدرة المتوسطة، في حين كانت كمية المعلومات تتناقص تدريجياً بنقصان عدد البدائل عند مستويات القدرة المرتفعة؛ أي أن الصورة الثالثة (ذات الخمسة بدائل) تعطي معلومات أكثر عن ذوي القدرة المرتفعة. وبناءً على ذلك إذا كانت مجموعة الأفراد المراد قياس السمة لديها من ذوي القدرات المتوسطة أو المرتفعة فيفضل استخدام فقرات الاختيار من متعدد التي لكل فقرة من فقراتها خمسة بدائل (الصورة الثالثة)، أما إذا كانت مجموعة الأفراد من ذوي القدرة المنخفضة فيفضل استخدام فقرات الاختيار من متعدد التي لكل فقرة من فقراتها ثلاثة بدائل (الصورة الأولى)، وذلك للحصول على أكبر قدر من المعلومات. كما يوضح الشكل (2) منحنيات اقتران المعلومات لصور الاختبار الثلاث التي توضح كمية المعلومات التي يقدمها الاختبار عند مستويات القدرة المختلفة.



الشكل (2): منحنيات اقتران المعلومات لصور الاختبار الثلاث

كما أن قيمة اقتران المعلومات تعد مؤشراً على ثبات الاختبار كونها تتناسب عكسياً مع خطأ التقدير الذي يزداد ثبات الاختبار بنقصانه، وبعد التحقق من جودة مطابقة فقرات الاختبار لنموذج راش واستبعاد الفقرات التي لم تحقق هذه المطابقة، وتحديد القيم التقديرية لكل من صعوبة الفقرات وقدرات الأفراد، ينبغي الحصول على نوعين من الثبات يطلق عليهما الثبات المتعلق بالعينة (Sample Reliability) والثبات المتعلق بالمقياس (Test Reliability). فمفهوم الثبات في إطار نموذج راش يشير إلى مدى الدقة في تقدير موقع كل من الأفراد، والفقرات على متصل السمة التي نهدف لقياسها (أي المعارف الأساسية في الرياضيات)، ويمكن تحديد مدى دقة الفقرات في تعريف متصل السمة بإيجاد النسبة بين الانحراف المعياري للقيم التقديرية المحررة لصعوبة الفقرات، ومتوسط الخطأ المعياري لهذه القيم. ويطلق على هذه النسبة معامل الفصل بين الفقرات (Item Separation Index)، والذي يجب أن تكون قيمته اكبر ما يمكن. فإذا لم تزد هذه القيمة عن (2) يصعب قياس السمة بواسطة هذه الفقرات (Wright & Masters, 1981).

ويتم الحصول على معامل ثبات العينة باستخدام الصيغة الآتية: ($R=G^2/1+G^2$) (Wright & Master,1981). حيث G: ترمز إلى النسبة التي أشرنا إليها. أما إذا كانت G: ترمز إلى النسبة بين الانحراف المعياري للقيم التقديرية المحررة لقدرات الأفراد، ومتوسط الخطأ المعياري لهذه القيم، فان R: تسمى عندئذ بمعامل ثبات الاختبار وهو يشير إلى مقدرة الفقرات على الفصل بين أفراد العينة أو التمييز بينهم اعتماداً على القيم التقديرية للقدرات التي تم الحصول عليها.

وبلغ الثبات المتعلق بالأفراد لصور الاختبار الثلاث باستخدام الصيغة الرياضية السابقة (0.80، 0.81، 0.84) على الترتيب. وهي قيم مقبولة، حيث تدل على كفاية عينة الأفراد لكل صورة في الفصل بين فقرات هذه الصورة. كما بلغ الثبات المتعلق بفقرات صور الاختبار الثلاث (0.95) لجميع صور الاختبار، وهي قيمة مرتفعة تدل على كفاية فقرات كل صورة في الفصل بين الأفراد المستجيبين على هذه الصورة، وبالتالي التميز بين مستوى المعارف الأساسية في الرياضيات لدى هؤلاء الأفراد. وكانت أعلى قيمة لمعامل ثبات الأفراد للصورة الثالثة تليها الصورة الثانية ثم الصورة الأولى أي أن معامل ثبات الأفراد يزداد بزيادة عدد البدائل، ولعل

السبب في ذلك هو نقصان فرصة تخمين الإجابة الصحيحة من قبل الأفراد، وبذلك يتوصلون إلى الإجابة الصحيحة بسبب ارتفاع قدرتهم وليس بسبب عامل التخمين. وللكشف عن دلالة الفروق بين معاملات الثبات للأفراد تم استخدام الإحصائي(V)، ويبين الجدول (15) نتائج التحليل المتعلقة بمعاملات ثبات الأفراد.

الجدول (15): نتائج التحليل المتعلقة بمعاملات ثبات الأفراد

الصورة	معامل الثبات	قيمة V	(χ^2)	درجات الحرية
الأولى	0.8	1.508	5.991	2
الثانية	0.81			
الثالثة	0.84			

ويلاحظ من الجدول (15) أن قيمة (V) أقل من قيمة (χ^2) وبالتالي لا توجد فروق دالة إحصائياً بين معاملات ثبات الأفراد، وهذا ما تؤكده قيم اقتران المعلومات المتقاربة لصور الاختبار الثلاث عند مستويات القدرة المختلفة. وتم تقدير قيم معامل الفصل بين الفقرات، ومعامل ثبات الاختبار وعدد الطبقات الإحصائية المتمايزة لصعوبة الفقرات وذلك لدراسة أثر عدد البدائل على كل منها، ويوضح الجدول (16) هذه القيم لصور الاختبار الثلاث، حيث إن معامل الفصل بين الفقرات هو مقياس يبين مدى دقة الفقرات في تعريف متصل السمة التي يهدف الباحثان إلى قياسها وهي القدرة الرياضية، ويستخدم في حساب عدد الطبقات الإحصائية المتمايزة لصعوبة الفقرات (Difficulty Strata) حسب المعادلة التالية:

$$H_i = \frac{\left(4G_i + 1\right)}{3}$$

حيث:

G_i: معامل الفصل بين الفقرات أو بين الأفراد.

H_i (Difficulty Strata): عدد الطبقات الإحصائية لصعوبة الفقرات (المستويات المتمايزة إحصائيا لتقديرات قيم معالم الصعوبة للفقرات وتبتعد مراكزها عن بعضها البعض بمقدار ثلاث وحدات من وحدات الأخطاء المعيارية في التقدير)، أو Ability) (Strata: عدد الطبقات الإحصائية لقدرات الأفراد (المستويات المتمايزة إحصائيا لتقديرات قيم معالم قدرات الأفراد وتبتعد مراكزها عن بعضها البعض بمقدار ثلاث وحدات من وحدات الأخطاء المعيارية في القياس) (Linacre & Wright, 1993).

الجدول(16): معامل الفصل بين الفقرات ومعامل الثبات للاختبار وعدد الطبقات الإحصائية لكل صورة من صور الاختبار

صور الاختبار	معامل الفصل بين الفقرات	معامل الثبات للاختبار	عدد الطبقات الإحصائية
الأولى	4.5	0.95	6.333
الثانية	4.28	0.95	6.04
الثالثة	4.45	0.95	6.267

يلاحظ من الجدول (16) أن قيم معامل الفصل بين الفقرات في صور الاختبار الثلاث كانت (4.5، 4.28، 4.45) على الترتيب، ونظراً لأن هذه القيم تزيد عن (2) فإن فقرات الاختبار تعد كافية لتعريف متصل السمة التي نرغب في قياسها كما أشار رايت وماستر (Wright& Master,1982). حيث كانت فقرات كل صورة من صور الاختبار قادرة على إظهار ست طبقات إحصائية متمايزة لتقدير قيم الصعوبة للفقرات. وهذا يتعارض مع دراسة بيدسكو ونيفو (Budescu & Nevo, 1985) التي أشارت إلى أن الثبات يزداد بزيادة عدد البدائل، وكذلك مع دراسة كوستن (Costin, 1972) التي توصلت الى أن الاختبار الذي لفقراته ثلاثة بدائل أكثر ثباتاً من الاختبارين ذي الأربعة والخمسة بدائل، كما تعارضت مع دراسة جريين، وساكس، ومشيل (Green, Sax & Michael, 1982) التي أشارت إلى وجود فروق بين معاملات الثبات تعزى لعدد البدائل.

كما تم تقدير قيم معامل الفصل بين الأفراد، ومعامل ثبات الأفراد، وعدد الطبقات الإحصائية المتمايزة للأفراد وذلك لدراسة أثر عدد البدائل على كل منها، ويوضح الجدول (17) هذه القيم لصور الاختبار الثلاث، حيث أن معامل الفصل بين الأفراد هو ناتج قسمة الانحراف المعياري المعدل من الخطأ على الجذر التربيعي لمتوسط مربعات الخطأ المعياري في القياس، ويستخدم في حساب عدد الطبقات الإحصائية المتمايزة لقدرات الأفراد حسب المعادلة السابقة.

الجدول(17): معامل الفصل بين الأفراد ومعامل الثبات للأفراد وعدد الطبقات الإحصائية لكل صورة من صور الاختبار

صورة الاختبار	معامل الفصل بين الأفراد	معامل الثبات للأفراد	عدد الطبقات الإحصائية
الأولى	1.99	0.8	2.987
الثانية	2.04	0.81	3.053
الثالثة	2.26	0.84	3.347

ويلاحظ من الجدول (17) أن قيم معامل الفصل بين الأفراد في صور الاختبار الثلاث كانت (1.99، 2.04، 2.26) على الترتيب، ونظراً لأن هذه القيم تزيد عن (2) للصورتين الثانية والثالثة وقريبة جداً من (2) للصورة الأولى فإن عينة الأفراد تعد كافية في الفصل بين الفقرات وبالتالي تعريف متصل السمة الذي تقيسه هذه الفقرات كما أشار رايت وماستر (Wright& Master,1982). كما وجد أن قيم معاملات الثبات المتعلقة بالأفراد (0.8، 0.81، 0.84) على الترتيب، وهي قيمة مرتفعة، وتدل على أن فقرات كل صورة قادرة على إظهار ثلاثة طبقات إحصائية لقدرات الأفراد، أي إظهار الفروق الفردية بين الأفراد.

ولإيجاد الكفاءة النسبية للصورتين الثانية والثالثة بالنسبة للصورة الأولى عند مستويات القدرة المختلفة، تم قسمة اقتران المعلومات للصورة الثانية على اقتران المعلومات للصورة الأولى، وكذلك قسمة اقتران المعلومات للصورة الثالثة على اقتران

المعلومات للصورة الأولى. والجدول (18) يبيّن قيم الكفاءة النسبية لصورتي الاختبار الثانية والثالثة بالنسبة للصورة الأولى للاختبار عند مستويات مختارة من القدرة.

الجدول (18): قيم الكفاءة النسبية لصورتي الاختبار الثانية والثالثة بالنسبة للصورة الأولى للاختبار عند مستويات مختارة من القدرة

القدرة(θ)	الكفاءة النسبية		
	صورة (1)	صورة (2)	صورة (3)
-3	1	0.784	0.697
-2.5	1	0.830	0.802
-2	1	0.884	0.919
-1.5	1	0.937	1.05
-1	1	0.994	1.186
-0.5	1	1.050	1.324
0	1	1.162	1.459
0.5	1	1.152	1.583
1	1	1.201	1.688
1.5	1	1.242	1.758
2	1	1.272	1.792
2.5	1	1.296	1.783
3	1	1.844	1.743
المتوسط	1	1.127	1.368

ويلاحظ من الجدول (18) أن متوسط الكفاءة النسبية للصورة الثالثة للاختبار مقارنة بالصورة الأولى للاختبار تساوي (1.368) وهذا يعني أن قيمة اقتران المعلومات للصورة الثالثة تساوي(1.368) قيمة اقتران المعلومات للصورة الأولى، أي أن الصورة الثالثة للاختبار تعطي معلومات أكبر من الصورة الأولى وعند مستويات القدرة جميعها باستثناء مستويات القدرة (θ = -3, -2.5, -2) لوجيت، وهذا يعني أن الأفراد عند مستويات القدرة جميعها تقريبا يحتاجون إلى (30) فقرة من الصورة الثالثة للاختبار لإعطاء القدر نفسه من المعلومات التي تنتج عن (41) فقرة من الصورة الأولى للاختبار.

كما أن متوسط الكفاءة النسبية للصورة الثانية للاختبار مقارنة بالصورة الأولى للاختبار تساوي (1.127) وهذا يعني أن قيمة اقتران المعلومات للصورة الثانية تساوي(1.127) قيمة اقتران المعلومات للصورة الأولى، أي أن الصورة الثانية للاختبار تعطي معلومات أكبر من الصورة الأولى وعند مستويات القدرة جميعها باستثناء مستويات القدرة (θ=-3,-2.5,-2,-1.5) لوجيت، وهذا يعني أن الأفراد عند مستويات القدرة جميعها تقريبا يحتاجون إلى (30) فقرة من الصورة الثانية للاختبار لإعطاء القدر نفسه من المعلومات التي تنتج عن (31) فقرة من الصورة الأولى للاختبار.

بينما يلاحظ أن الصورة الأولى للاختبار هي الأكثر كفاءة عند مستويات القدرة المنخفضة، فمثلاً عند مستوى القدرة (θ=-3) لوجيت نحتاج إلى (30) فقرة من الصورة الثانية للاختبار لإعطاء القدر نفسه من المعلومات التي تنتج عن (24) فقرة من الصورة الأولى للاختبار، وإلى (30) فقرة من الصورة الثالثة للاختبار لإعطاء القدر نفسه من المعلومات التي تنتج (21) فقرة من الصورة الأولى للاختبار.

لذا فإنه من الضروري وعند بناء الاختبارات المقننة أو الاختبارات التكيفيّة اختيار عدد بدائل فقرات الاختيار من متعدد تبعاً لمستوى قدرة الأفراد، حيث أن الاختبار ذو الثلاثة بدائل يتناسب، وبدرجة اكبر مع الأفراد ذوي القدرات المنخفضة كونه يعطي معلومات أكبر من الاختبارين ذوي الأربعة أو الخمسة بدائل عند مستوى القدرة نفسه. أما لمجموعة الأفراد ذوي القدرة المرتفعة أو المتوسطة يفضل استخدام الاختبار ذو الخمسة بدائل كونه يعطي معلومات اكبر عن أفراد هذه الفئة.

المصادر والمراجع

الشبلي، علي مصطفى. (1995). **أثر عدد البدائل في اختبار تحصيلي من نوع الاختيار من متعدد على تجانسه الداخلي والخصائص السيكومترية لفقراته.** رسالة ماجستير غير منشورة. جامعة اليرموك، اربد، الأردن.

الشريفين، نضال كمال. (2006). **الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس والتقويم التربوي والنفسي. مجلة العلوم التربوية والنفسية، 7**(4)، 80-109.

العبابنة، عماد غصاب. (2004). **أثر حجم العينة وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة.** رسالة دكتوراه غير منشورة. جامعة عمان للدراسات العليا، عمان، الأردن.

علام، صلاح الدين. (1995).**الاختبارات التشخيصية مرجعية المحك في المجالات التربوية و النفسية و التدريبية.** القاهرة: دار الفكر العربي.

عودة، أحمد. (2002). **القياس و التقويم في العملية التدريسية.** الأردن: دار الأمل.

Albanese, M. A., & Forsyth, R. A. (1984). The one - two and modified two - Parameter latent trait models: An empirical study of relative fit. **Education and Psychological Measurement, 44**, 229 - 246.

Allen, M.J, and Yen, W.M. (1979). **Introduction to Measurement Theory.** California: Cole publishing company.

Bejar, I. I. (1980). Aprocedure for investigating the Unidimensionality of Achievement test based on item parameter estimates. **Journal of Educational Measurement**,**17**(4), 283

Budescu, D. V. & Nevo, B. (1985). Optimal number of options: An investigation of the Assumption of proportionality. **Journal of Educational Measurement. 22**(3), 183 - 196.

Cerhan, K., Haladyna, T. and Brewer, B. (1993). Use of an inclusive option and the optimal number of

Linacre, J. M., & Wright, B. D. (1993). **A user Guide to BIGSTEPS. (computer program)**. Version 2.4,Chicago: MESA Press.

Linacre, J. M. (2002). **A user Guide to WINSTEPS\MINISTEPS. (computer program)**. Chicago: ISBN Press.

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

Straton, R.G. and Catts, R.M. (1980).A comparison of two, three and four choice items tests given a fixed total number of choices. **Educational and Psychological Measurement.40.**

Trevisan, M.S., Sax, G. & Michael, W.B. (1991). The effect of number of options per item and student ability on test validity and reliability. **Educational and Psychological Measurement. 51**, 829-837.

Trevisan, M.S., Sax, G. & Michael, W.B. (1994). Estimating the optimum number of options per item using an incremental option paradigm. **Educational and Psychological Measurement. 54**(1), 86 - 91.

Wright, B., Mead, R., & Bell, S. (1980). **BICAL, Calibrating Items With Rasch Modal. Research Memorandum**. Chicago: Statistical Laboratory.

Wright, B. D., & Masters, G. N. (1982). **Rating Scale Analysis Rasch Measurement**. MESA Press,Kimbark Avenue**.** Chicago, P. 59.

Wright, D. & Stone, M. (1979). **Best Test Design A handbook for Rasch Measurement**, Chicago: MESA Press.

options for multiple Choice items. **Educational and Psychological Measurement. 53**, 241-247.

Costin, F. (1970). The Optimal number of alternative in multiple Choice achievement tests: Some empirical evidence for a mathematical Proof. **Educational and Psychological Measurement. 30**, 353-358.

Costin, F. (1972). Three-Choice versus four-Choice items: Implications for reliability and validity of objective achievement tests. **Educational and Psychological Measurement. 32**, 1035 - 1038.

Crocker, L., & Algina, J. (1987). **Introduction to Classical and Modern Test Theory**. N. Y.: Harcourt Jovanovich.

Ebel, R. L. (1969). Expected reliability as a function of choices per item. **Educational and Psychological Measurement. 29**, 565 - 570.

Frisbie.,D. A., and Sweeney, D.S. (1982). The relative merits of Multiple true-False achievement test. **Journal of Educational Measurement**. **19**(1), 29-35.

Green, K., Sax, K. & Michael, W. B. (1982). Validity and reliability of tests having differing number of options for students of differing levels of a ability. **Educational and Psychological Measurement. 42**, 239 - 245.

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

Hambleton, R.K., and Jonse, R.W. (1993). Comparison of Classical Test Theory and Item Response Theory and their Applications to test development. **Educational Measurement.Issues and Practice,** 38-47.

Hambleton, R.K., and Swaminathan, H. (1985). **Item Response Theory Principle and Application**. Kluwer: Nijhoff Publishing.

Hashway, R. M. (1978). **Objective Mental Measurement: Individual and Program Evaluation Using the Rasch Model**. U. S. Of America: Praeger.

Hattie, J. (1985). Methodology Review: Assessing unidimensionality of tests and items. **Applied Psychological Measurement.9** ,139-164.

Hays, W. L. (1973). **Statistic for The Social Sciences.** 2nd ed. New York: Holt, Rinehart, and Winston, Inc.

Hulin, C. L., Drasgow, F. & Parsons, C.K. (1983). **Item Response Theory: Application to Psychological Measurement**. Illinois: Dow Jones _ Irwin..

Jodoin, M. G. (2003). Measurement efficiency of innovative item Formats in computer-Based testing. **Journal of Educational Measurement, 40**(1). 1-15.

Julian, E. R. (1988). Using computerized patient simulations to measure the clinical competence of physicians**. Applied Measurement in Education**. 4(1), 299 - 318.

Levine, M. V., and Drasgow, F. (1983). The relation between incorrect option choice and estimation ability. **Educational and Psychological Measurement**, 43, 675-685.

ملحق (1)

جدول مواصفات الاختبار

	الموضوع (النسبة)	مستويات الأهداف (النسبة)				عدد الفقرات
		معرفة (20%)	استيعاب (30%)	تطبيق (40%)	تحليل (10%)	
الهندسة التحليلية (40%)	التوازي والتعامد(10%)	1	2	2	1	6
	بعد نقطة عن مستقيم(10%)	1	2	2	1	6
	خصائص المثلث(10%)	1	2	2	1	6
	خصائص متوازي الأضلاع(10%)	1	2	2	1	6
الإحصاء والاحتمالات (60%)	مقاييس التشتت(10%)	1	2	2	1	6
	تعديل البيانات وأثره على مقاييس التشتت (10%)	1	2	2	1	6
	مفهوم الاحتمال وقوانين الاحتمالات(20%)	2	4	5	1	12
	الاحتمال المشروط واستقلال الحوادث(20%)	2	4	5	1	12
	عدد الفقرات	10	20	22	8	60

ملحق(2)

أمثلة على فقرات الاختبار كما وردت في صوره الثلاث

الصورة الأولى

1. إذا كان التباين لمجموعة من المشاهدات يساوي (9)، وعدّلت كل من هذه المشاهدات حسب العلاقة ص= 34 -2س، حيث س: المشاهدة قبل التعديل، ص: المشاهدة بعد التعديل، فما قيمة الانحراف المعياري للمشاهدات بعد التعديل:
   أ) 6 ب) 3 جـ) 18
2. أطلق صيادان كل منهما طلقة واحدة نحو هدف معين، فإذا كان احتمال إصابة الأول للهدف (0.5)، واحتمال إصابة الثاني للهدف (0.4) فما احتمال إصابة الهدف من الصيادين معا:
   أ) 1 ب) صفر جـ) 0.2

الصورة الثانية

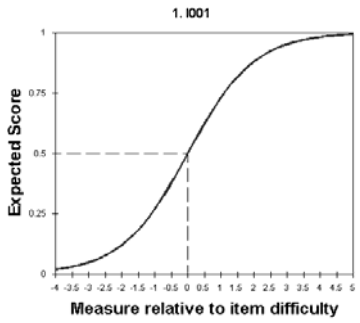
1. إذا كان التباين لمجموعة من المشاهدات يساوي (9)، وعدّلت كل من هذه المشاهدات حسب العلاقة ص= 34 -2س، حيث س: المشاهدة قبل التعديل، ص: المشاهدة بعد التعديل، فما قيمة الانحراف المعياري للمشاهدات بعد التعديل:
   أ) 6 ب) 3 جـ) 18 د) 40
2. أطلق صيادان كل منهما طلقة واحدة نحو هدف معين، فإذا كان احتمال إصابة الأول للهدف (0.5)، واحتمال إصابة الثاني للهدف (0.4) فما احتمال إصابة الهدف من الصيادين معا:
   أ) 1 ب) 0.1 جـ) صفر د) 0.2

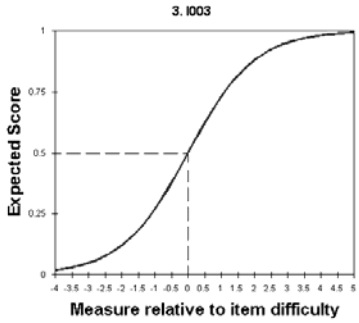
الصورة الثالثة

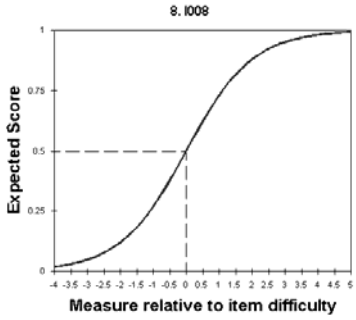
1. إذا كان التباين لمجموعة من المشاهدات يساوي (9)، وعدّلت كل من هذه المشاهدات حسب العلاقة ص= 34 -2س، حيث س: المشاهدة قبل التعديل، ص: المشاهدة بعد التعديل، فما قيمة الانحراف المعياري للمشاهدات بعد التعديل:
   أ) 6 ب) 16 جـ) 3 د) 18 هـ) 40
2. أطلق صيادان كل منهما طلقة واحدة نحو هدف معين، فإذا كان احتمال إصابة الأول للهدف (0.5)، واحتمال إصابة الثاني للهدف (0.4) فما احتمال إصابة الهدف من الصيادين معا:
   أ) 0.9 ب) 1 جـ) 0.1 د) صفر هـ) 0.2

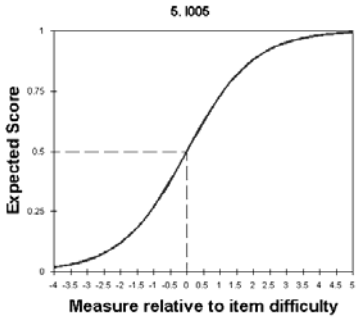
ملحق (3)

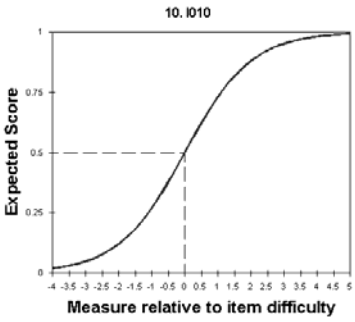
منحنيات الخصائص لبعض فقرات الصورة الأولى للاختبار

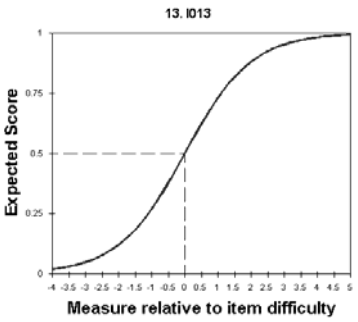


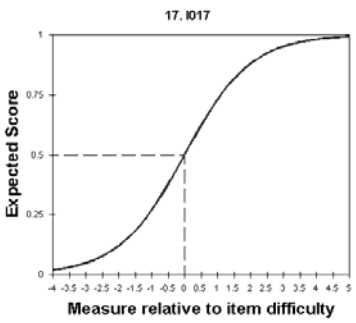


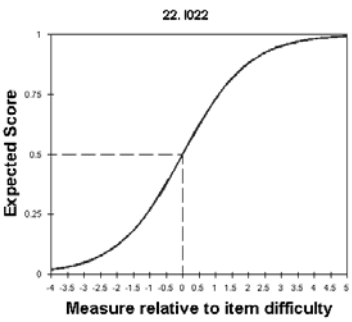






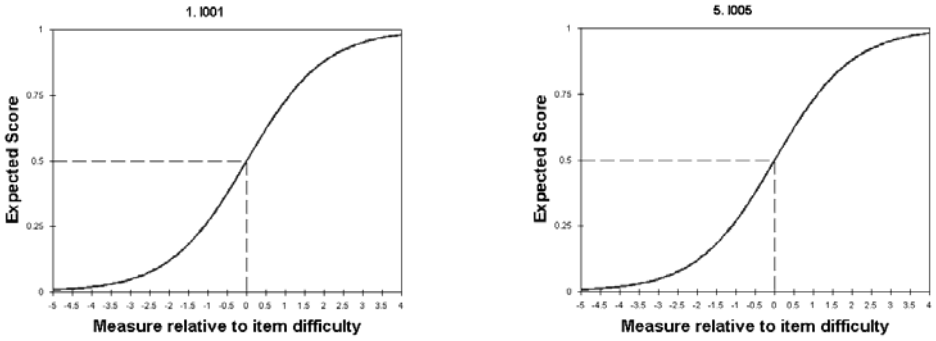


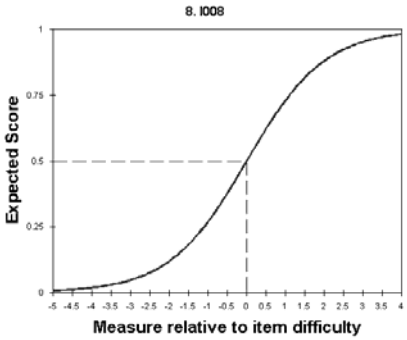


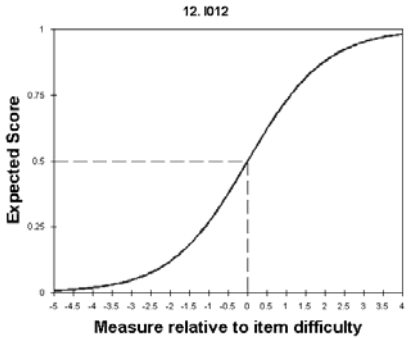


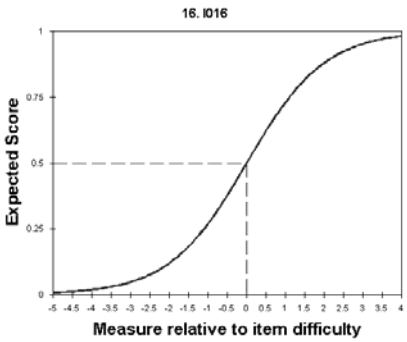
ملحق(4)

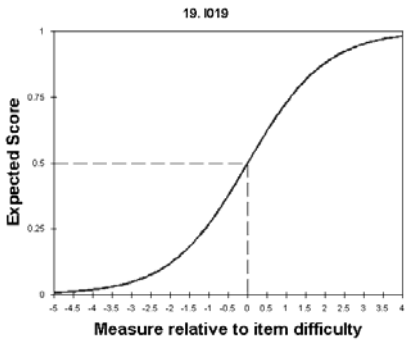
منحنيات الخصائص لبعض فقرات الصورة الثانية للاختبار

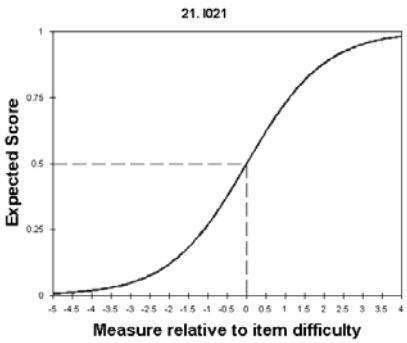


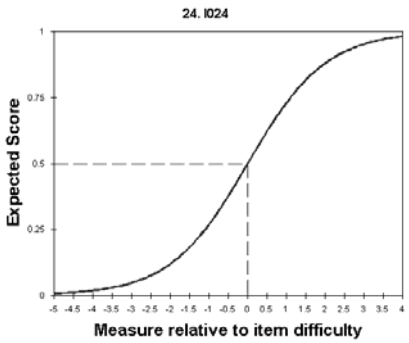






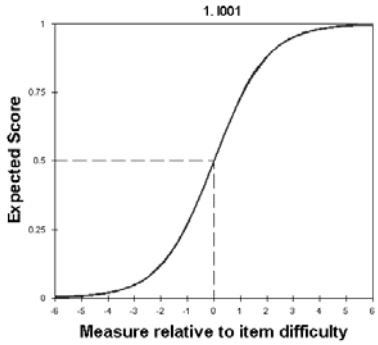


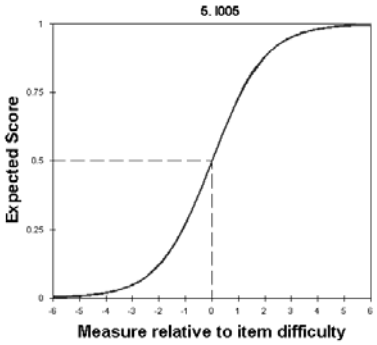


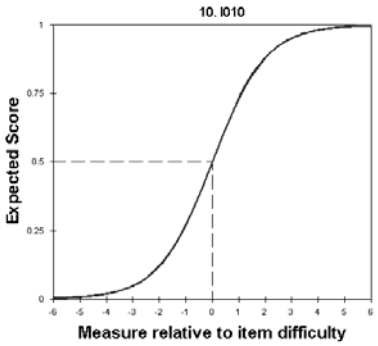


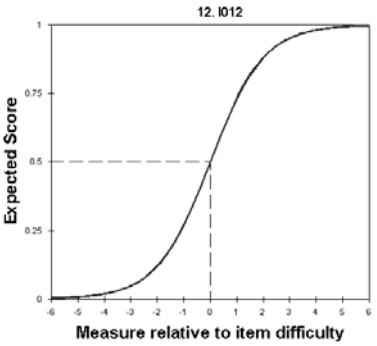
ملحق(5)

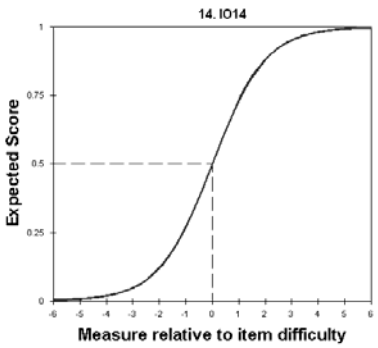
منحنيات الخصائص لبعض فقرات الصورة الثالثة للاختبار

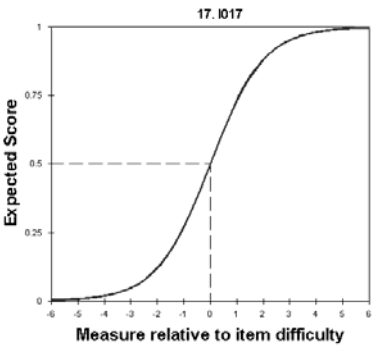


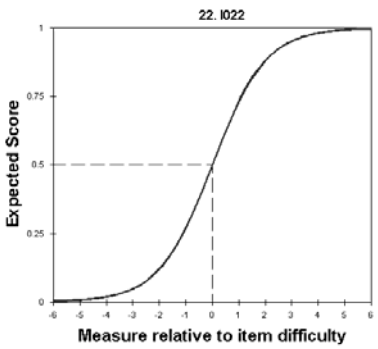


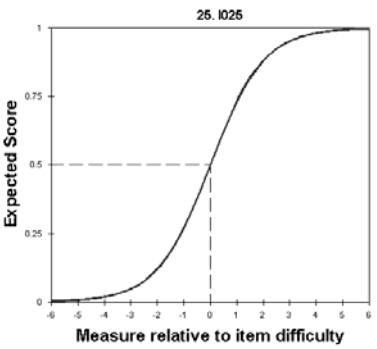












ملحق (6)

قدرات الأفراد والأخطاء المعيارية في قياسها لأفراد الصورة الأولى بصورتها النهائية مرتبة تنازلياً (عدد الأفراد = 190)

	الدرجات الكلية للأفراد	تكرار الأفراد	القيم المتحررة لقدرات الأفراد بوحدة اللوجيت	الخطأ المعياري	الدرجات الكلية للأفراد	تكرار الأفراد	القيم المتحررة لقدرات الأفراد بوحدة اللوجيت	الخطأ المعياري
	34	2	2.66	0.61	20	21	0.19	0.35
	33	5	2.32	0.54	19	20	0.07	0.35
	32	5	2.06	0.5	18	4	0.05-	0.35
	31	0	-	-	17	10	0.18-	0.35
	30	0	-	-	16	9	0.3-	0.35
	29	0	-	-	15	20	0.43-	0.36
	28	9	1.28	0.4	14	13	0.56-	0.36
	27	4	1.12	0.39	13	4	0.69-	0.37
	26	5	0.98	0.38	12	8	0.83-	0.37
	25	10	0.84	0.37	11	5	0.97-	0.38
	24	4	0.7	0.36	10	0	-	-
	23	6	0.57	0.36	9	0	-	-
	22	10	0.44	0.36	8	3	1.44-	0.42
	21	11	0.32	0.35	7	1	1.63-	0.44
الوسط الحسابي			0.18	0.37				
الانحراف المعياري			0.81	0.05				
مجموع التكرارات	190							

ملحق (7)

صعوبة الفقرات والخطأ المعياري في تقدير هذه الصعوبة لكل فقرة من فقرات الصورة الأولى بصورتها النهائية مرتبة تنازلياً (عدد الفقرات = 37 فقرة)

رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري
39	1.66	0.19	33	0.63	0.16	29	0.12-	0.16	26	0.7-	0.16
19	1	0.17	10	0.53	0.16	11	0.14-	0.16	35	0.73-	0.16
9	0.79	0.17	4	0.46	0.16	17	0.14-	0.16	25	0.75-	0.17
22	0.95	0.17	34	0.43	0.16	18	0.22-	0.16	15	0.78-	0.17
38	0.95	0.17	3	0.31	0.16	32	0.24-	0.16	21	0.92-	0.17
13	0.81	0.16	20	0.22	0.16	2	0.36-	0.16	1	1.04-	0.17
37	0.74	0.16	24	0.19	0.16	6	0.36-	0.16	14	1.33-	0.19
16	0.71	0.16	31	0.17	0.16	30	0.54-	0.16	36	1.33-	0.19
5	0.63	0.16	8	0.07	0.15	40	0.54-	0.16	12	1.81-	0.21
23	0.63	0.16									
المتوسط	0	0.16									
الانحراف المعياري	0.78	0.01									

ملحق (8)

قدرات الأفراد والأخطاء المعيارية في قياسها لأفراد الصورة الثانية بصورتها النهائية مرتبة تنازلياً (عدد الأفراد = 188)

الخطأ المعياري	القيم المتحررة لقدرات الأفراد بوحدة اللوجيت	تكرار الأفراد	الدرجات الكلية للأفراد	الخطأ المعياري	القيم المتحررة لقدرات الأفراد بوحدة اللوجيت	تكرار الأفراد	الدرجات الكلية للأفراد	
0.35	0.11-	10	18	0.43	1.65	5	31	
0.35	0.23-	5	17	-	-	0	30	
-	-	0	16	-	-	0	29	
0.35	0.48-	5	15	-	-	0	28	
0.36	0.6-	15	14	0.38	1.01	15	27	
0.36	0.73-	10	13	-	-	0	26	
0.37	0.87-	2	12	0.36	0.74	5	25	
0.38	1-	13	11	0.36	0.61	9	24	
0.39	1.15-	13	10	0.35	0.49	21	23	
0.4	1.31-	12	9	-	-	0	22	
0.42	1.47-	13	8	0.35	0.24	15	21	
0.44	1.65-	1	7	0.35	0.12	14	20	
				0.35	0	5	19	
				0.37	0.21-			الوسط الحسابي
				0.02	0.83			الانحراف المعياري
						188		مجموع التكرارات

ملحق (9)

صعوبة الفقرات والخطأ المعياري في تقدير هذه الصعوبة لكل فقرة من فقرات الصورة الثانية بصورتها النهائية مرتبة تنازلياً (عدد الفقرات = 38 فقرة)

رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري
9	1.34	0.19	16	0.55	0.16	37	0.09-	0.16	14	0.61-	0.16
10	1.08	0.18	4	0.52	0.16	18	0.14-	0.16	12	0.74-	0.16
24	1.08	0.18	20	0.42	0.16	8	0.19-	0.16	40	0.74-	0.16
6	0.98	0.18	29	0.31	0.16	11	0.19-	0.16	38	0.84-	0.16
28	0.95	0.18	2	0.26	0.16	21	0.24-	0.16	30	0.87-	0.16
39	0.95	0.18	23	0.19	0.16	19	0.26-	0.16	1	1.03-	0.17
7	0.8	0.17	17	0.16	0.16	5	0.46-	0.16	15	1.03-	0.17
26	0.8	0.17	32	0.04-	0.16	22	0.56-	0.16	25	1.5-	0.18
31	0.72	0.17	34	0.04-	0.16	35	0.58-	0.16	36	1.66-	0.19
3	0.63	0.17	33	0.09-	0.16						
المتوسط	0	0.17									
الانحراف المعياري	0.75	0.01									

ملحق (10)

قدرات الأفراد والأخطاء المعيارية في قياسها لأفراد الصورة الثالثة بصورتها النهائية مرتبة تنازلياً (عدد الأفراد = 187)

الخطأ المعياري	القيم المتحررة لقدرات الأفراد بوحدة اللوجيت	تكرار الأفراد	الدرجات الكلية للأفراد	الخطأ المعياري	القيم المتحررة لقدرات الأفراد بوحدة اللوجيت	تكرار الأفراد	الدرجات الكلية للأفراد	
0.37	0.06-	9	16	0.55	2.21	8	29	
0.38	0.2-	17	15	-	-	0	28	
0.38	0.34-	9	14	-	-	0	27	
0.38	0.48-	4	13	-	-	0	26	
0.39	0.63-	5	12	-	-	0	25	
0.4	0.79-	13	11	0.41	1.12	4	24	
0.4	0.95-	20	10	-	-	0	23	
0.42	1.11-	15	9	0.39	0.8	16	22	
0.43	1.29-	15	8	-	-	0	21	
0.45	1.49-	17	7	-	-	0	20	
0.48	1.7-	11	6	-	-	0	19	
0.51	1.94-	3	5	0.37	0.22	9	18	
				0.37	0.08	12	17	
				0.41	0.48-			الوسط الحسابي
				0.04	0.85			الانحراف المعياري
						187		مجموع التكرارات

ملحق (11)

صعوبة الفقرات والخطأ المعياري في تقدير هذه الصعوبة لكل فقرة من فقرات الصورة الثالثة بصورتها النهائية مرتبة تنازلياً (عدد الفقرات = 33 فقرة)

رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري	رقم الفقرة	صعوبة الفقرة	الخطأ المعياري
28	2	0.25	7	0.55	0.18	31	0.02-	0.16	14	0.71-	0.16
5	1.03	0.19	33	0.46	0.17	24	0.05-	0.16	35	0.74-	0.16
39	0.99	0.19	17	0.43	0.17	1	0.07-	0.16	40	0.74-	0.16
9	0.88	0.19	34	0.40	0.17	8	0.07-	0.16	22	0.91-	0.16
10	0.71	0.18	4	0.34	0.17	38	0.2-	0.16	12	1.22-	0.16
2	0.68	0.18	16	0.23	0.17	21	0.33-	0.16	36	1.5-	0.17
19	0.65	0.18	18	0.23	0.17	11	0.46-	0.16	15	1.53-	0.17
20	0.65	0.18	3	0.03	0.16	32	0.53-	0.16	25	1.77-	0.18
13	0.58	0.18									
	المتوسط الحسابي		0		0.17						
	الانحراف المعياري		0.82		0.02						