أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس

The Effect of Sample in Estimating the Information Function of the Test and its Estimating Standard Error Depending on the Modern Test Theory of Measurement

حابس الزبون

Habis Alzboon

قسم علم النفس، كلية التربية، جامعة الطائف، المملكة العربية السعودية.

بريد الكتروني: Habes_zbn77@yahoo.com

تاريخ التسليم: (٢٠١٢/٤/١٧)، تاريخ القبول: (٢٠١٣/٣/١٧)

ملخص

هدفت هذه الدراسة إلى معرفة أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس، ولأغراض الدراسة تمّ استخدام استجابات (٧٥٠٠) طالبا وطالبة من طلبة الصف الثامن الأساسي، تمّ اختيارهم عشوائيا موزعين على خمس مجموعات بواقع (٥٠٠ طالب للعينة الأولى، ١٠٠٠ طالب وطالبة للعينة الثانية، ١٥٠٠ طالب وطالبة للعينة الثالثة، ٢٠٠٠ طالب وطالبة للعينة الرابعة، ٢٥٠٠ طالب وطالبة للعينة الخامسة) على الاختبار الوطني لضبط نوعية التعليم لمادة الرياضيات، والمكون من (٤٠) فقرة من نوع الاختيار من متعدد. وباستخدام برنامج Bilog-mg3 لتحليل استجابات الأفراد، والبرنامج الاحصائي spss، تم إيجاد معالم الفقرات بالاعتماد على النظرية الحديثة في القياس حسب النموذج الثلاثي وتقدير دالة المعلومات والخطأ المعياري في تقدير دالة المعلومات لخمسة نماذج من الاختبار. وقد أظهرت النتائج أن مقدار تقديرات دالة المعلومات يتغير بتغير حجم العينة، إذ يزداد بزيادة حجم العينة، كما أظهرت النتائج ان الخطأ المعياري في تقدير دالة المعلومات يتغير بتغيير حجم العينة، إذ يتناقص بزيادة حجم العينة.

الكلمات المفتاحية: دالة المعلومات، الخطأ المعياري في تقدير دالة المعلومات، النظرية الحديثة في القياس.

Abstract

The study aims to investigate the Effect of sample size in Estimate the Information Function of the test and it's Estimating standard Error depending on the modern test theory of measurement. To achieve the purposes of the study, responses of (7500) students from eighth grade students were are randomly selected divided into five groups (The first consist of 500 students, The second consist of 1000 students, The there'd consist of 1500 students, The fourth consist of 2000 students, and The last consist of 2500 students) subjected to the 40- item national multiple choice test for mathematics education quality adjustment. Using the Bilog-mg3 and spss programs to Estimate the Item parameters, information function, and standard Error in Estimating the information function for the five forms of the test according to three parameter logistic IRT model. The results showed that, the amount of information function estimates varies increasingly according to the increasing sample, Yet, the results show that the standard Error in estimating the information function varies decreasingly according to the increasing sample.

Key words: information function, the standard error in estimating the information function, the modern test theory of measurement.

المقدمة

تمثل النظرية الحديثة في القياس الاتجاة المعاصر في القياس النفسي والتربوي، ويطلق عليها اسم نظرية استجابة الفقرة (Item Response Theory-IRT)، ونظرية المنحنى المميز للفقرة (Item Characteristic Curve-ICC)، ونظرية السمات الكامنة، ويعدّ لازرسفيلد (Lazersfeld, 1968) أول من استخدم مصطلح السمة الكامنة بهذا المعنى (Hambleton & Swaminathan,1985).

ويرجع الفضل في تقديم أسس النظرية الحديثة في القياس للمهتمين بالقياس النفسي والتربوي إلى لورد (Lord, 1952, 1953). وقد تطورت هذه النظرية، وما يتعلق بها من نماذج سيكومترية منذ ذلك الوقت بهدوء وثبات، وظلت أعمال لورد وغيره من علماء القياس في هذا المجال حتى السبعينيات من هذا القرن لا يلتفت إليها إلا عدد محدود من خبراء القياس في الولايات المتحدة الأمريكية وبعض الدول المتطورة. غير أنه حدث تقدم سريع في الحركة البحثية والتطبيقية لهذه النظرية ابتداء من عام ١٩٦٨، وهو العام الذي نشر فيه لورد كتابه"النظريات الإحصائية لدرجات الاختبارات العقلية

(Statistical Theories of Mental Test Scores)، الذي جمع بين غلافية لأول مرة أسس النظرية السيكومترية الكلاسيكية، وأسس نظرية النظرية الحديثة في القياس. وتتميز هذة النظرية بما يأتي:

١. وجود مجموعة كبيرة من فقرات الاختبار التي تقيس نفس السمة، ويكون تقدير قدرة الفرد مستقلا عن عينة الفقرات التي تطبق عليه (Item Free).

٢. وجود مجتمع كبير من الأفراد، حيث تكون الخصائص السيكومترية للفقرات (مثل معاملات الصعوبة والتمييز) مستقلة عن عينة الأفراد التي استخدمت في تقدير هذه الخصائص (Person Free).

وتطرح النظرية الحديثة في القياس نمذجه للعلاقة بين متغير غير ملحوظ يستخدم لقياس هذة القدرات (مفاهيميا يطلق علية اسم القدره التي يقيسها الاختبار)، واحتمالية الاستجابة الصحيحة على فقرة معينة؛ وذلك من خلال دوال لوغاريتمية تربط قدرة المفحوص، ومعالم الفقرة باحتمالية الاستجابة الصحيحة عليها، وقد انبثق عنها نماذج متعددة، جميعها تفترض أن قدرة واحدة تقيس الأداء على الاختبار، والقدره يمكن تمثيلها على متصل مستمر لا حدود له، ولكنها تتغير في خصائصها التي تصفها الفقرات، وتمتد درجات الصعوبة والقدرة نظريا على متصل يتراوح بين (-∞) إلى (+∞)، ولكنه عمليا يتراوح بين (-٣) و(+٣)؛ وذلك لأنه من النادر توافر قيما تزيد عن (+٣) أو تقل عن (-٣) (Hambleton and Swaminathan,1985).

وقد طورت نماذج رياضية مختلفة؛ وذلك للتعبير عن العلاقة بين القدرة واحتمال الإجابة الصحيحة على الفقرة وتختلف هذه النماذج في صيغتها الرياضية وعدد معالمها، كما وتختلف من حيث عدد الافتراضات اللازم توافرها في البيانات المستخدمة.

وتقسم نماذج نظرية استجابة الفقرة إلى قسمين:-

- النماذج ثنائية التدريج (dichotomous IRT models) منها: النموذج أحادي المعلمة (one parameter model)، والنموذج ثنائي المعلمة (two parameters model)، والنموذج ثلاثي المعلمة (three parameters model).
- النماذج متعددة التدريج (polytomous IRT models)، وتميز هذه النماذج بين المفحوصين الذين لديهم جزء من المعرفة، والذين ليس لديهم معرفة، ومنها: نموذج الاستجابة المتدرجة (graded response models GRM)، وهو امتداد للنموذج ثنائي المعلمة، ونموذج التقدير الجزئي (Partial credit model, pcm)، وسلم التقدير (Rating scale models RSM)، وهما امتداد للنموذج أحادي المعلمة ونموذج الاستجابة الاسمية (Nominal response model NRM)، الذي يناسب الاستجابات التي لايمكن تصنيفها في فئات رتبية (Embreston &rise, 2000).

وليس بالضرورة أن يلائم نموذج رياضي كل البيانات، فقد يلائم بعضها ولا يلائم بعضها الآخر، لذلك عند استخدام أي نموذج رياضي، فمن الضروري فحص مدى ملاءمة النموذج للبيانات، وإذا لاءم النموذج الرياضي البيانات، عندها فقط يمكن تقدير معالم الفقرة وقدرة الأفراد (Hambleton and Swaminathan,1985;& Traub and Wolf, 1981) .

وتتميز النظرية الحديثة في القياس (Modern Test Theory of Measurement) عن النظرية الكلاسيكية في القياس (Classical Theory) في أن النظرية الحديثة في القياس، كما يراها لورد وُضِعت تحت افتراضات قوية حول سلوك الأفراد عند استجابتهم لفقرات أسئلة محددة، وهذه الافتراضات هي (Hambleton and Swaminathan,1985; Crocker and Algena,1986;Embretson,2000) :

١. **افتراض أحادية البعد (Unidimentionality)**

يشير هدا الافتراض إلى أن هناك قدرة أو سمة واحدة فقط ضرورية لشرح اداء المفحوص على الاختبار، وهذا الافتراض لا يمكن تحقيقه في كثير من الأحيان؛ لان هناك عوامل معرفية واختبارية وشخصية تتعلق بالمفحوص تؤثر على أدائه على الاختبار مثل القلق والدافعية، وعادةً نعتبر أن هذا الافتراض قد يتحقق إذا وجد هناك عامل طاغي (Dominant Factor) وحيد يؤثر على اداء المفحوص على الاختبار، وهذا العامل الطاغي عادةً ما يسمى بالقدرة التي يقيسها الاختبار.

وللكشف عن افتراض أحادية البعد يتم ذلك من خلال طريقتين :

أ. إجراء التحليل العاملي: حيث يتم اخضاع العلامات على الفقرات للتحليل العاملي، لمعرفة البنية العاملية لهذه الفقرات، أي لمعرفة هل يوجد عامل رئيسي طاغي؟ أي وجود عامل يكون المؤثر الرئيسي على الأداء، ويسمى هذا العامل بالعامل المسيطر والمؤثر على الأداء بالقدرة المقاسة بالاختبار.

ب. التوزيعات المشروطة لعلامات الاختبار عند كل مستوى من مستويات القدرة $F(X/\theta)$، وتكون لمجموعات جزئية منتهية من المجتمع الكلي، وتكون متشابهة (Identical). أما إذا كانت هذه التوزيعات متغيرة عبر المجتمعات الجزئية للمجتمع الكلي، فهذا يعني أن الاختبار يقيس شيئاً آخر غير القدرة الوحيدة التي يقيسها أصلا.

٢. **افتراض الاستقلال الموضعي (Local Independence)**

ويقصد بهذا الافتراض أن استجابات المفحوصين من نفس القدرة للفقرات المختلفة مستقلة عن بعضها بعضاً وبالتالي فإن استجابة المفحوص على فقرة لا تؤثر على أدائه على فقرة أخرى سلباً أو ايجاباً. ويمكن الكشف عن هذا الافتراض بما يأتي:

أ. إذا وجد في ترتيب الفقرات تأثير على الأداء، فمعنى ذلك أن افتراض الاستقلال الموضعي لا يتحقق.

ب. عدم استيفاء أحادية البعد أي إذا لم تكن فقرات الاختبار تقيس قدرة واحدة، فإن افتراض الاستقلال الموضعي لا يتحقق، اي إذا تحقق افتراض أحادية البعد، فإن افتراض الاستقلال الموضعي يتحقق أيضاً.

٣. منحنى خصائص الفقرة (Item Characteristic Curve)

وهو عبارة عن دالة تربط بين احتمال النجاح في الإجابة على الفقرة، والقدرة التي يقيسها الاختبار الذي يحتوي تلك الفقرة، وان هذه الدالة أو العلاقة تتخذ شكل المنحنى اللوغارتمي، الذي يفترض أن يكون شكله كحرف S حيث يمثل قدرة المفحوص على المحور السيني، واحتمال حصول المفحوص على الإجابة الصحيحة للفقرة تمثل الصادات, أو هو معادلة انحدار غير خطي لاحتمال إجابة الفقرة إجابة صحيحة على القدرة، أو السمة التي يقيسها الاختبار .(Crocker and Algina,1986)

والفرق الرئيس بين نماذج استجابة الفقرة الشائعة، هو في الصيغة الرياضية لمنحنى خصائص الفقرة (ICC)، ويستطيع باني الاختبار أن يختار الدالة الرياضية التي يراها مناسبة بعد مطابقة النموذج للبيانات. وإذا كان التوزيع الشرطي لعلامات الفقرة عند مستوى محدد من القدرة متماثل عند المجتمعات المختلفة، فهذا يعني أن منحنى خصائص الفقرة (ICC) سوف يكون متماثل عند المجتمعات المختلفة، وبالتالي تتحقق خاصة اللاتغير (Invariant) لمنحنى خصائص الفقرة (ICC) بتغير من يأخذ الاختبار.

٤. افتراض السرعة (Speedeness)

ينص هذا الافتراض على إن إجابة المفحوص على الفقرة سواء أكانت صحيحة أم خاطئة يجب أن تعتمد على مقدار القدرة التي يقيسها الاختبار، وليس على سرعته في الإجابة.فإذا اعتمد الاختبار على السرعة هذا يؤدي إلى انتهاك افتراض أحادية البعد؛ لأنه سوف يكون هنالك قدرتان تؤثر على الاداء على الاختبار، وهي السرعة والسمة التي يقيسها الاختبار.

نماذج استجابة الفقرة ثنائية التدريج

أولاً: النموذج اللوجستي أحادي المعلمة (نموذج راش) (one parameter logistic model)

يعد هذا النموذج ابسط نماذج استجابة الفقرة ثنائية التدريج، ويفترض هذا النموذج أن جميع الفقرات لها القدرة التمييزية نفسها للتمييز بين المفحوصين، أي تكون جميع المنحنيات المميزة لفقرات الاختبار متوازية (متساوية في الميل)، وتختلف عن بعضها البعض في نقاط التقائها بالمحور الافقي، الذي يمثل متصل القدرة أو السمة الكامنة، كما يفترض عدم لجوء المفحوصين للتخمين العشوائي عند اجابتهم عن فقرات الاختبار، وهذا يعني إن الفقرات تختلف فقط في صعوبتها، وتكون معادلة هذا النموذج كما يأتي:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

حيث $P_i(\theta)$: احتمال إجابة الفقرة (i) إجابة صحيحة من قبل مفحوص له القدرة θ تم اختياره عشوائياً من بين المفحوصين الذين لهم القدرة نفسها.

D: ثابت مقداره ١,٧(Sampling Factor)، وهو عامل تدريج يجعل القدرة المستنتجة من استعمال هذا النموذج مساوية للقدرة المستنتجة من استعمال النموذج، الذي يعتمد على المنحنى الطبيعي التراكمي.

b_i : معلمة الصعوبة للفقرة i

e : الأساس اللوغاريتمي الطبيعي (٢,٧١٨)

ثانياً: النموذج اللوجستي ثنائي المعلمة (two parameter logistic model)

اقترحة عالم الاحصاء بيرنبيوم (birnbum) مع مجموعة من زملائة بجامعة كولومبيا الاميريكية، ويفترض هذا النموذج أن الفقرات تختلف في صعوبتها، وأيضاً تختلف في درجة التمييز بين المفحوصين وأن التخمين قليل جداً، وتكون معادلة هدا النموذج كما يأتي:

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}}$$

a_i : معامل التمييز للفقرة i .

ثالثاً: النموذج اللوجستي ثلاثي المعلمة (three parameter logistic model)

يفترض هدا النموذج أن الفقرات تختلف في صعوبتها وايضاً تختلف في درجة التمييز بين المفحوصين، وأن مستويات التخمين لدى المفحوصين تختلف من فقرة إلى أخرى، وتكون معادلة هذا النموذج كما يأتي:

$$P_i(\theta) = c_i + (1+c_i)\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}}$$

c_i : معلمة التخمين للفقرة i، ويسمي الخط التقاربي الادنى (lower asymptote) للمنحى المميز للفقرة، ويمثل احتمال الاجابة الصحيحة للافراد ذوي القدرة المنخفضة (Hambelton & swaminathan,1985).

والنظرية الحديثة في القياس لا تتفق مع النظرية الكلاسيكية في افتراضها تساوي احتمالات الإجابة الصحيحة عن جميع فقرات الاختبار لفرد ذي قدرة معينة, فمن الممكن أن تكون بعض الفقرات أكثر صعوبة من غيرها بالنسبة لهذا الفرد؛ لذلك فإن احتمال الإجابة الصحيحة عن مجموعة من الفقرات يعدّ دالة لموقع كل من الفرد، والفقرة على متصل السمة الكامنة التي تقيسها هذه الفقرات. أي أن هذه النظرية تفترض وجود دالة مميزة

(Characteristic Function) تعتمد على موقع الفرد والفقرة معاً على متصل سمة كامنة معينة، ويتحدد هذا المنحنى أو هذه الدالة إذا علمنا هذا الموقع؛ لذلك يطلق على هذا القياس: القياس المتحد أو المقترن (Conjoint Measurement) (علام، ٢٠٠٢).

وتُعدّ دالة المعلومات (item information function, IIF) من الاحصاءات المهمة في النظرية الحديثة في القياس، حيث يمكن من خلالها تحديد الخطأ المعياري في التقدير، حيث أنه يمكن من ايجادها تحديد الخطأ المعياري في التقدير، وذلك بالاعتماد على الارجحية العظمى لمعلم القدرة، حيث أن تباين الخطا في تقدير القدرة يساوي معكوس دالة المعلومات، وكما أن مصفوفة التباين المشترك (variance covariance matrix) للتقديرات تساوي معكوس مصفوفة المعلومات (information matrix) لتقديرات معالم الفقرة (Hambleton & Swaminathan,1993).

وتلعب دالة معلومات الاختبار دوراً مهماً في النظرية الحديثة في القياس، إذ يمكن من خلالها تحديد الخطأ المعياري في التقدير، وتتمتع دالة معلومات الاختبار، التي تمثل مجموع دوال معلومات الفقرات عند مستوى معين من القدرة بميزة، وهي كون دالة معلومات الاختبار مستقلة عن عينة المفحوصين، وبذلك تقدم النظرية الحديثة في القياس مميزات إضافية، فيما يتعلق بزيادة القدرة على تقدير أخطاء القياس (Brannick, 2003).

كما أن الخطأ المعياري في التقدير يرتبط عكسيا مع الجذر التربيعي لدالة معلومات الاختبار وفق العلاقة الآتية:

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

وهذا يعني أن الخطأ المعياري في التقدير، يكون اقل ما يمكن عند مستويات القدرة التي تناظر أقصى معلومات.

ومن الاستخدامات المفيدة لدالة معلومات الفقرة، إمكانية معرفة مدى مساهمة كل فقرة في دالة معلومات الاختبار بشكل مستقل عن الفقرات الأخرى للاختبار، فإذا كان لدينا فكرة جيدة عن قدرات مجموعة المفحوصين، فيمكن انتقاء فقرات الاختبار التي تعظم المعلومات التي يقدمها الاختبار في المدى الذي تتوزع فيه قدرات المفحوصين في الاختبار.

ولقد أشار هامبلتون وسوامينثان (Hambleton,Swaminthan,1985) إلى أن الاختبار السهل من المتوقع أن يقدم تقديرات أكثر دقة عند مستويات القدرة المتدني، كما أن الاختبار الصعب من المتوقع أن يقدم معلومات أكثر دقة عند مستويات القدرة العليا؛ لذا فهو أكثر فائدة للمفحوصين ذوي القدرة العالية.

وتستخدم دالة معلومات الفقرة لاختيار الفقرات عند بناء الاختبارات، بالاعتماد على النظرية الحديثة في القياس، على افتراض ان معلومات الفقرة تتغير عبر مستويات السمة

المختلفة؛ وعليه فإنه من الممكن اختيار فقرات تقدم دقة قياس مرتفعة عند نقطة معينة على متصل السمة، كما أن الفقرات التي تتضمن معالم تمييز كبيرة، تقدم معلومات أكبر عن قدرة المفحوصين؛ وبالتالي الحصول على دقة أكبر، وتعطى دالة المعلومات للاختبار من خلال المعادلة الآتية:

$$I(\theta) = \sum_{i=1}^{n} \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)}$$

حيث أن:

$I(\theta)$: دالة معلومات الاختبار.

(θ): معلم القدرة للمفحوصين.

$P_i(\theta)$: دالة استجابة الفقرة $Q_i(\theta) = 1 - P_i(\theta)$

$P_i'(\theta)$: المشتقة الأولى لدالة استجابة الفقرة.

وعليه يمكن انتقاء فقرات الاختبار اعتماداً على كمية المعلومات التي تساهم بها الفقرات في كمية المعلومات الكلية للاختبار.

وهناك عدد من الخطوات المهمة التي من الممكن الاعتماد عليها عند بناء الاختبار وهي:

١. وصف شكل دالة المعلومات المستهدفة في الاختبار (target information).
٢. اختيار فقرات تغطي دوال معلوماتها مستويات الصعوبة المناظرة لدالة المعلومات المستهدفة.
٣. بعد إضافة كل فقرة إلى الاختبار يتم حساب دالة معلومات الاختبار لفقرات الاختبار التي يتم اختبارها.
٤. الاستمرار في اختيار فقرات الاختبار، حتى تقترب دالة معلومات الاختبار من دالة المعلومات المستهدفة.

وتتأثر دالة المعلومات بمعالم الفقرة، ففي النماذج اللوجستية الأحادية والثنائية تكون أعلى كمية معلومات عند معلم الصعوبة للفقرة (عند تساوي القدرة مع الصعوبة)؛ وذلك لان شكل دالة معلومات الفقرة بشكل عام قريب من الشكل الجرسي (bell shape)، بينما في حالة النموذج اللوجستي الثلاثي يمكن ملاحظة تأثير معالم الفقرة على دالة معلومات الاختبار من خلال تبسيط المعادلة الرياضية لدالة معلومات الاختبار على النحو الآتي (Hambleton & Swaminathan,1993).

$$I_i(\theta)=\frac{2.89{a_i}^2(1-c_i)}{(c_i+e^{(1.7a_i(\theta-b_i))})(1+e^{(1.7a_i(\theta-b_i)})^2}$$

وبالاعتماد على المعادلة الرياضية السابقة يمكن ملاحظة ما يأتي:

١. تكون كمية المعلومات أكبر بشكل عام عندما تكون معلمة التمييز كبيرة.

٢. تزداد كمية المعلومات كلما اقتربت معلمة التخمين من الصفر.

٣. تكون كمية المعلومات عندما تكون (θ) قريبة من (b_i) أكبر من كمية المعلومات عندما تكون عندما تكون (θ) بعيدة من (b_i).

وقد بين ستوكنج (stocking,1990) أن تقديرات الفقرة من القضايا المهمة في النظرية الحديثة في القياس، وأنه من خلال ايجاد القيم القصوى لتقديرات القدرة من خلال دالة المعلومات لمعالم الفقرة باستخدام اجراءات المشتقة الأولى للنماذج اللوجستية الاحادية والثنائية والثلاثية يمكن الحصول على تقدير لمعالم الفقرة يؤدي إلى تعظيم دالة المعلومات لمعلم الفقرة.

ولقد أجريت العديد من الدراسات التي تناولت تقدير معالم الفقرات ودالة المعلومات بالاعتماد على النظرية الحديثة في القياس، ففي دراسة لهامبلتون وجونز (Hambleton & jones, 1994)، والتي هدفت إلى التحقق من الآثار الناتجة عن اختيار الفقرات ذات القدرة التمييزية العالية على دقة دالة المعلومات؛ لأن تقدير معالم تلك الفقرات يتم بنسبة من الخطأ، مما يؤثر على تقدير دالة معلومات الفقرة ودالة معلومات الاختبار، بحيث نحصل على قيم أكبر من قيمها الحقيقية، وقد تّم أخذ متغيرين لهما أهمية في تحديد مقدار التأثير، وهما حجم عينة المفحوصين المستخدمة في معايرة الفقرات، ونسبة بنك الفقرات إلى طول الاختبار، وقد تمّ استخدام طرق المحاكاة (simulation)، وكان حجم العينة في المرة الاولى (٥٠٠) مفحوص، وفي المرة الثانية (٢٠٠٠) مفحوص وتضمن بنك الفقرات (٨٠) فقرة، وقد توصلت الدراسة إلى أن دالة معلومات الاختبار تغالي في التقدير؛ وذلك بسبب الزيادة المفرطة في اختيار الفقرات ذات القدرة التمييزية العالية، كما بينت نتائج الدراسة أن حجم عينة المفحوصين أكثر تأثرا من نسبة بنك الفقرات إلى طول الاختبار على دقة المعلومات للاختبار.

وبين هيندرسون وردياني (Hendrson & Dianne,2001) من خلال دراسته التي هدفت إلى دراسة أثر الخطأ في تقدير معالم الفقرات على تطوير الاختبار، حيث تكونت عينة الدراسة من ٣٤٠٠ مفحوص مقسمين إلى ثلاث عينات (٤٠٠، ١٠٠٠، ٢٠٠٠) مفحوص واستخدم أسلوب المحاكاة لتوليد ثلاثة بنوك منفصلة للفقرات، وتمّ الاعتماد على ثلاث طرق لاختيار الفقرات حيث اعتمدت الطريقة الأولى على اختيار الفقرات التي تعمل تغطية منحنى دالة معلومات الاختبار المستهدفة (target test information function) وأما الطريقة الثانية فأعتمدت على اختيار الفقرات التي تمتلك أعلى معلومات بغض النظر عن موقعها على مقياس

القدرة (Θ)، وبحيث تغطي دالة معلومات الاختبار المستهدفة، واعتمدت الطريقة الثالثة على تحديد موقع الفقرة على مقياس القدرة (Θ)، بحيث تغطي أعلى معلومات للفقرة؛ أي يتم تحديد (Θ) التي تعطي أعلى دالة معلومات للفقرة، وقد تمّ الاعتماد على النموذج اللوجستي ثنائي المعلمة، والنموذج اللوجستي ثلاثي المعلمة، وقد توصلت الدراسة إلى أن الاختبارات التي اعتمدت على الطريقة الأولى والطريقة الثانية تعملان بشكل متسق على المغالاة في تقدير دالة معلومات الاختبار، وأن الاختبارات التي اعتمدت على الطريقة الثالثة تعطي تقديرات ثابتة لدالة معلومات الاختبار.

وفي دراسة قام بها واجنر وهافي (Wagner & Harvey,2003)؛ والتي هدفت إلى استخدام النظرية الحديثة في القياس لتطوير اختبار التفكير الناقد من خلال مبادىء اختبار واطسون وجليسر للتفكير الناقد، حيث تم فحص درجة تأثر اختبار واطسون وجليسر للتفكير الناقد بمستوى التخمين، ومقارنته باختبار واجنر المطور من اختبار واطسون وجليسر بعدد أكبر من البدائل، وقد تكونت عينة الدراسة من (٤٠٧) مفحوص طبق عليهم اختبار واجنر والمكون من (١١١) فقرة، وقد توصلت نتائج الدراسة إلى ان اختبار واجنر الكامل ولد كمية معلومات أكبر من اختبار واطسون وجليسر وخطأ معياري اقل، كما توصلت نتائج الدراسة إلى أن زيادة التخمين في اطار النظرية الحديثة في القياس يؤدي إلى تقليل كمية معلومات الفقرة، وكمية معلومات الاختبار؛ وبالتالي زيادة أخطاء التقدير.

أما سوامنثان وزملاؤه (Swaminathan et al.2003) فقد أجروا دراسة لتحديد أثر الحكم المسبق على دقة تقديرات معالم الفقرات في الاختبارات ثنائية التدريج باستخدام النظرية الحديثة في القياس، وقد هدفت الدراسة إلى فحص أثر الطريقة البابيزية (Bayesian Estimation) في تقدير معالم الفقرات، واستقصاء العلاقة بين حجم عينة المفحوصين، ومواصفات المعلومات القبلية على دقة تقدير معالم الفقرات، وقد توصلت الدراسة إلى أن التحسن في تقدير معالم الفقرات يختلف باختلاف النموذج اللوجستي، حيث تبين بأن النموذج اللوجستي أحادي المعلمة كان الأكثر تحسنا عند زيادة حجم العينة مقارنة بالنموذجين الثنائي والثلاثي.

كما توصلت دراسة ستون ويوموتو (Stone & yamoto,2004)، والتي كانت بعنوان أثر حجم العينة على تقديرات معالم الفقرات ثنائية التدريج باستخدام نموذج راش ونماذج نظرية استجابة الفقرة، حيث استخدم نتائج اختبار (Knox's Cube Test Revised) في سحب (٣٠) عينة عشوائية لتقدير معالم الفقرات، وقد وجدا بأن نموذج راش يعطي أقل تقدير لمعلمة الصعوبة، وأن العينات الصغيرة تكون أقل في مطابقة نموذج حسن المطابقة.

وفي دراسة عبابنة (٢٠٠٤) بعنوان أثار حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة، ولتحقيق أهداف هذه الدراسة تمّ إعداد اختبار قدرة عقلية مؤلف من أربعة اختبارات فرعية هي اختبار المفردات، اختبار المتشابهات، اختبار المتضادات، واختبار

الحساب، حيث بلغ عدد فقرات الاختبار النهائية (٧١) فقرة، طبقت فقرات الاختبار على عينة مكونة من طلاب وطالبات الصف السابع من مدارس التربية والتعليم، وقد بلغ عدد أفراد عينة الدراسة (١٠٠٠) طالب وطالبة، وقد أستخدمت البرمجية biolog 3.11 لتقدير قدرات المفحوص ومعالم فقرات الاختبار والاخطاء المعيارية في التقدير واحصائيات مطابقة السياقات للنموذج اللوجستي الثلاثي، الذي تمّ اعتمادها لإغراض هذه الدراسة، وقد توصلت الدراسة إلى أنه تزداد الدقة في تقدير معلمة الصعوبة والقدرة عندما يكون مدى القدرة للمفحوص متوافقاً مع مدى صعوبة الفقرات، وكذلك تزداد الدقة في تقدير معلمة التخمين عند استخدام عينة من ذوي القدرة المتدنية في معايرة الفقرات، كما أظهرت النتائج بأن تقديرات معالم الفقرة لعينة مختارة من الفقرات تتسم بالاستقرار النسبي عند معايرتها ضمن المقياس ككل أو بصورة مستقلة باستخدام العينة نفسها من الأفراد، وأن الدقة في تقدير معالم الفقرة تزداد بزيادة حجم عينة المفحوصين.

وفي دراسة بومرك (pommerich, 2007) بعنوان أثر استخدام معالم الفقرات التي تمت معايرتها باستخدام الطريقة التقليدية في الاختبارات التكيفية المحوسبة، والتي هدفت لاختبار البرامج التي تستخدم كل من الكمبيوتر والاختبارات الورقية، حيث اهتمت الدراسة بدراسة مدى الاستقرار في تقدير معالم الفقرات عند استخدام الاختبارات الورقية واستخدام الكمبيوتر باستخدام اختبارات مختلفة الطول، وقد أظهرت نتائج الدراسة عدم وجود استقرار في معالم الفقرات وأن هذا الاختلاف في المعالم يقل عند استخدام اختبارات أطول.

وفي دراسة لفيزباترك (Fitzpatric, 2009) بعنوان أثر اختبارات التهيئة على معدل اتقان الطلبة، والتي هدفت إلى دراسة أثر تخفيض طول اختبارات التهيئة على معدل اتقان الطلبة وذلك بالاعتماد على (١٢) اختبار من نوع الاختيار من متعدد، وتكونت الاختبارات في الدراسة من (١٥، ١٠، ٥) فقرة، وبعد تطبيق هذه النماذج من الاختبارات، تمّ التحليل بالاعتماد على النموذج أحادي المعلمة، وقد أشارت نتائج الدراسة إلى أن هناك زيادة في تباينات الاختلافات في معدلات اتقان الطلبة عند استخدام الاختبارات القصيرة، كما كان هناك اختلاف وعدم استقرار في معاملات الصعوبة، وقد أوصت الدراسة باستخدام اختبارات تحتوي على أكبر من (١٥) فقرة لزيادة الاستقرار في تقدير معالم الفقرات.

أما دراسة الثوابية (٢٠١٠) فقد هدفت إلى استقصاء أثر حجم العينة في تقدير معلمة صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة، ولتحقيق أهداف الدراسة تمّ اشتقاق معلمة الصعوبة، والخطأ المعياري في تقديرها باستخدام اختبار تحصيلي في الرياضيات للصف العاشر الأساسي، تكون في صورته النهائية من (٨٠) فقرة من نوع الاختبار من المتعدد.

وطبق الاختبار على عينات عشوائية طبقية تراوح حجمها ما بين(٢٠٠) إلى (١١٢٩٢) طالباً وطالبة، وقد استخدمت برمجية (BILOG - MG) لتقدير معلمة الصعوبة، والخطأ المعياري في تقديرها، وقد توصلت الدراسة إلى أن قيمة معلمة صعوبة الفقرة تزداد بزيادة حجم

العينة، كما توصلت إلى أن الخطأ المعياري في تقدير معلمة صعوبة الفقرة يتناقص بزيادة عدد أفراد العينة.

مشكلة الدراسة

الغرض من هذه الدراسة هو الكشف عن أثر حجم عينة المفحوصين على تقدير دالة المعلومات والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس، ونظراً لأن دالة المعلومات للاختبار ذات أهمية في النظرية الحديثة في القياس، حيث تزودنا بالمعرفة عن مدى كفاءة الاختبار في تقدير القدرة عند مستوياتها المختلفة، فإن هناك حاجة للاطمئنان على مدى تأثر دالة المعلومات للاختبار والخطأ المعياري في تقديرها عند تغيير حجم العينة، إذ أن الاختبارات قد تطبق على عدد كبير من المفحوصين وأحيانا على عدد قليل، وهذا يستدعي توفير بيانات عن خصائص الفقرات ودالة المعلومات في هذه السياقات المختلفة.

أهمية الدراسة

تعتبر النظرية الحديثة في القياس من النظريات التي أحدثت ثورة في نظرية الاختبارات، حيث تمّ الاعتماد عليها من قبل مطوري الاختبارات؛ وذلك لتقديمها حلولا لمشاكل عملية في الاختبارات، وتعتمد النظرية الحديثة في القياس على الفقرات المكونة للاختبار، اذ تفترض ان كل فقرة من فقرات الاختبار تقيس السمة الكامنة نفسها، وفي ضوء ذلك يمكن حساب كمية المعلومات عند كل مستوى من مستويات القدرة ولكل فقرة، والغرض من الاختبار هو تقدير قدرة المفحوص، لذا فان كمية المعلومات التي توفرها فقرات الاختبار عند أي مستوى من مستويات القدرة تلعب دورا أساسيا في دقة تقدير هذه القدرة.

وتُعدُّ دالة المعلومات للاختبار ذات أهمية في النظرية الحديثة في القياس، حيث تزودنا بالمعرفة عن مدى كفاءة الاختبار في تقدير القدرة عند مستوياتها المختلفة، لذلك تكمن أهمية هذه الدراسة في البحث عن أثر حجم العينة على تقدير دالة المعلومات والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس.

أسئلة الدراسة

جاءت هذه الدراسة للإجابة عن الأسئلة الآتية

١. هل تختلف تقديرات دالة المعلومات باختلاف حجم العينة المستخدمة في تقديره؟

٢. هل يختلف الخطأ المعياري في تقدير دالة المعلومات باختلاف حجم العينة المستخدمة في تقديره؟

متغيرات الدراسة

المتغيرات المستقلة في هذه الدراسة

- حجم عينة

المتغيرات التابعة

- دالة المعلومات للاختبار
- الخطأ المعياري في تقدير دالة المعلومات.

تعريف المصطلحات

الخطأ المعياري للتقدير (standard error of estimate): الانحراف المعياري للخطأ في تقدير معلم، وسوف يتمّ الاعتماد عليه للدلالة على التشتت أو التجانس في تقدير دالة المعلومات، وفي هذه الدراسة يعرف الخطأ المعياري للتقدير بالقيمة التي يتم الحصول عليها من تطبيق المعادلة الآتية:

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

دالة معلومات الفقرة (item information function): هو اقتران رياضي يعبر عن كمية المعلومات المتمثلة في تمييز الفقرة بين مستويات القدرة، ويتم التوصل إليه من بيانات التطبيق. وفي هذه الدراسة تعرف دالة معلومات الفقرة بالقيمة التي يتم الحصول عليها من تطبيق المعادلة الآتية:

$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta)$$

دالة معلومات الاختبار (test information function): هو اقتران رياضي يعبر عن مجموع دوال المعلومات لفقرات الاختبار، وفي هذه الدراسة تعرف دالة معلومات الاختبار بالقيمة التي يتم الحصول عليها من تطبيق المعادلة الآتية:

$$I(\theta) = \sum_{i=1}^{n} \frac{P_i'(\theta)^2}{P_i(\theta) Q_i(\theta)}$$

الاختبار الوطني لضبط نوعية التعلم: (national test to monitor and control the quality of education): اختبار تُعدّه وزارة التربية والتعليم الأردنية لقياس مدى امتلاك طلبة الصف الثام الأساسي لمهارات التعلم الأساسية في مبحث الرياضيات.

مجتمع الدراسة

تكوّن مجتمع الدراسة من جميع طلبة الصف الثامن الأساسي من الذكور والإناث المنتظمين في المدارس الحكومية في المملكة الأردنية الهاشمية والبالغ عددهم (١٢٩٥٠٩) طالب وطالبة.

عينة الدراسة

تكونت عينة الدراسة من طلبة الصف الثامن الأساسي للعام الدراسي ٢٠٠٩-٢٠١٠ والتي بلغ حجمها (٧٥٠٠) طالبا وطالبة، موزعين على خمس مجموعات بواقع (٥٠٠ طالب للعينة الأولى، ١٠٠٠ طالب وطالبة للعينة الثانية، ١٥٠٠ طالب وطالبة للعينة الثالثة، ٢٠٠٠ طالب وطالبة للعينة الرابعة، ٢٥٠٠ طالب وطالبة للعينة الخامسة)، حيث تمّ اختيار هذه العينة بالاعتماد على الطريقة العينة العشوائية البسيطة اعتماداً على الاختيار العشوائي من مجتمع الدراسة.

أداة الدراسة

لغرض تحقيق أهداف الدراسة، اعتمد الباحث على الاختبار الوطني لضبط نوعية التعليم لمادة الرياضيات، والمُعدّ من قبل وزارة التربية والتعليم والمكون من (٤٠) فقرة من نوع الاختيار من متعدد والذي مر بالاجراءات التالية الآتية:

١. تم تشكيل ست لجان لكتابة فقرات الاختبار، وقد شكلت اللجان من مشرفي القياس والتقويم وعدد من مشرفي الميدان، وأعضاء من ادارة المناهج وادارة التدريب.
٢. مراجعة نتاجات التعلم من قبل اللجان، وتحديد مهارات التعلم الأساسية التي ستقاس في الاختبار، وذلك بعد استخلاصها من النتاجات العامة والخاصة.
٣. اعداد جدول مواصفات للاختبار.
٤. بناء عدد كبير من الفقرات الاختبارية (item pool).
٥. طباعة الفقرات الاختبارية ونسخها.
٦. تجريب الفقرات على عينة تجريبية من الطلبة.
٧. ادخال بيانات الفقرات وتصحيحها.
٨. التحليل الاحصائي لاستجابات الطلبة على الفقرات، للتعرف على الخصائص السيكومترية لها مثل: فاعلية البدائل، درجة صعوبة الفقرات ودرجة تمييزها.
٩. اختيار الفقرات الملائمة التي تقيس نتاجات التعلم والمتعلقة بمهارات التعلم الأساسية.
١٠. بناء الاختبار وفقا لجدول المواصفات الذي تم اعداده سابقا.

وقد تكون الاختبار بصورته النهائية من (٤٠) فقرة من نوع الاختيار من متعدد موزعة على المحاور الآتية: الأعداد والعمليات، والأنماط والجبر، والقياس، والهندسة والاحصاء والاحتمالات.

وتقيس هذه الفقرات أداء الطلبة على المهارات الآتية: إجراء العمليات الحسابية الأساسية على الأسس للأعداد الصحيحة، وحل وتمثيل المتباينات من الدرجة الأولى، وتمييز وتمثيل العلاقات والاقترانات، واستخدام تشابه المثلثات وخواص الاشكال الهندسية في حل المسألة، وتحليل وتركيب المقادير الجبرية حتى الدرجة الثانية، وتوظيف العمليات الحسابية الأساسية في حل المسألة الحسابية.

صدق الأداة

تم تحليل الاختبار للتأكد من صدق المحتوى (content validity) بتحديد النتاج الذي تقيسه كل فقرة، والوحدة أو الفصل الذي ينتمي له النتاج، ووزن العلامة المخصصة لتلك الفقرة، ومن ثم مطابقة جدول تحليل الاختبار بجدول المواصفات، والذي وجد مطابقا بدرجة عالية مما يدلل على صدق المحتوى.

ثبات الأداة

للتحقق من ثبات الاختبار، تمّ ايجاد معامل ثبات الاختبار باستخدام معادلة كرونباخ الفا والذي بلغ (٠,٨٦).

حدود الدراسة

تحددت هذه الدراسة بما يأتي:

١. اقتصرت هذه الدراسة على إيجاد دالة المعلومات للفقرة والاختبار.
٢. تطبيق الدراسة على طلبة الصف الثامن الأساسي.
٣. دراسة مبحث واحد وهو الرياضيات.
٤. البيانات التي تم جمعها، جمعت من تطبيق الاختبار على عينة الدراسة للعام الدراسي ٢٠٠٩-٢٠١٠.

اجراءات الدراسة

للإجابة على أسئلة الدراسة قام الباحث بعمل الإجراءات التالية الآتية:

بعد القيام بالإجراءات الرسمية من موافقات وزيارات لادارة مديرية الامتحانات في وزارة التربية والتعليم، والالتقاء مع الاشخاص القائمين على اعداد الاختبار، تمّ اخذ البيانات الخاصة بتطبيق الاختبار، وتم اخذ (٥) عينات عشوائية من الطلبة، بحيث يختلف حجم العينة في كل مرة من مرات التحليل، وقد تراوح حجم العينة من (٥٠٠) إلى (٢٥٠٠) مفحوص موزعين على العينات الخمسة بواقع (٥٠٠) طالب للعينة الأولى، (١٠٠٠) طالب وطالبة للعينة الثانية، (١٥٠٠) طالب وطالبة للعينة الثالثة، (٢٠٠٠) طالب وطالبة للعينة الرابعة، (٢٥٠٠) طالب

وطالبة للعينة الخامسة)، وقد تم حساب معالم الفقرات ودالة المعلومات والخطأ المعياري في تقديرها بالاعتماد على النموذج ثلاثي المعلمة.

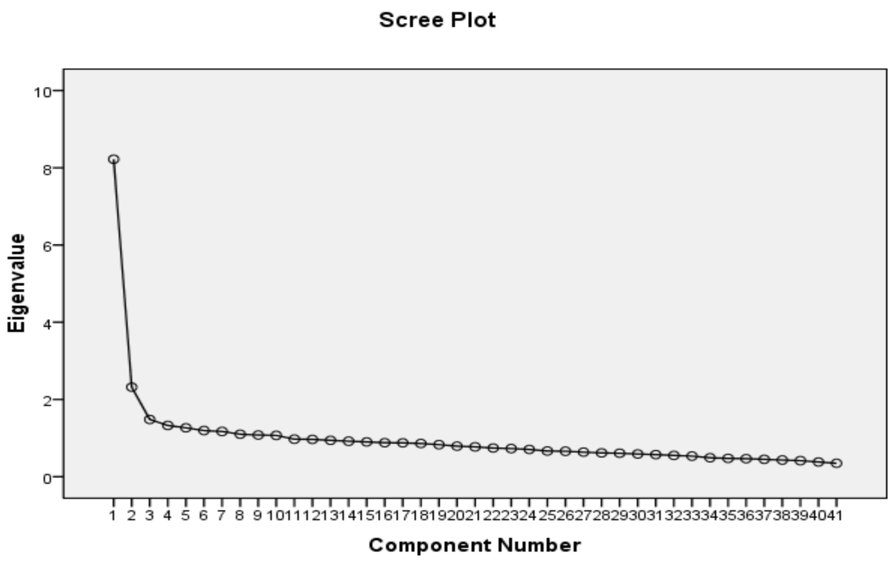
المعالجة الإحصائية

تم إجراء التحليل العاملي لاستجابات الأفراد باستخدام برنامج SPSS، وقد وفّر ذلك مؤشرات على أحادية البعد للاختبار في حالة تطبيقه على العينات الخمسة، فقد كانت نسبة الجذر الكامن للعامل الاول إلى الجذر الكامن للعامل الثاني أكثر من (٢)، وكانت نسبة الفرق بين الجذرين الكامنين الأول والثاني إلى الفرق بين الجذرين الكامنين الثاني والثالث عالية (بين ٨,٢٢ للاختبار المطبق على العينة الثالثة وللاختبار المطبق على العينة الثانية (٣,٨٩)، وتُعد هذه القيم كما يشير هاتي (Hattie, 1985) مؤشرات على أحادية البعد، مما يسمح باستخدام نماذج نظرية الاستجابة على الفقرة في تقدير معالم الفقرات وقدرات الأفراد، ويبين جدول (١) خلاصة لبعض نتائج التحليل العاملي للاختبارات الخمسة.

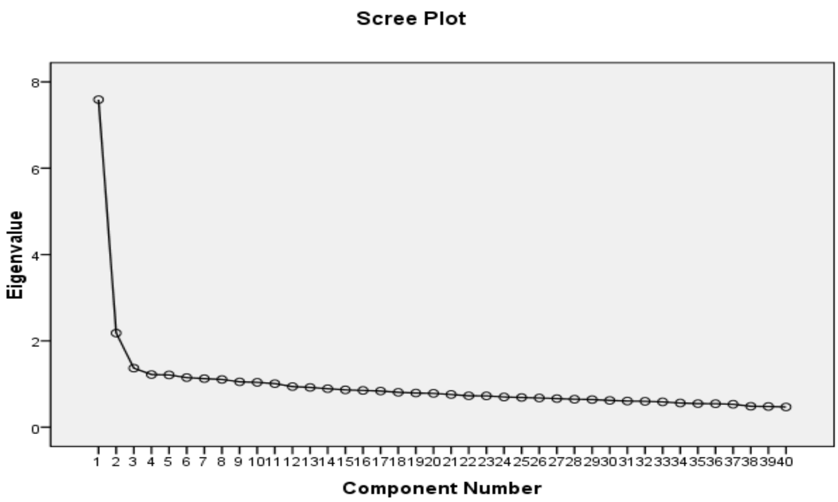
جدول (١): نتائج التحليل العاملي للاختبارات الخمسة.

الاختبار	عدد العوامل	نسبة التباين التي يفسرها العامل الأول	الجذر الكامن الأول/ الجذر الكامن الثاني	(الجذر الكامن الأول- الجذر الكامن الثاني) / (الجذر الثاني- الجذر الثالث)
المطبق على العينة الاولى	٩	٢٠,٥٥	٣,٦١	٧,٤٣
المطبق على العينة الثانية	١١	١٨,٩٧	٣,٤٨	٣,٨٩
المطبق على العينة الثالثة	٩	١٨,٨١	٣,٩٧	٨,٢٢
المطبق على العينة الرابعة	٩	١٥,١٥	٣,٣٢	٧,٢١
المطبق على العينة الخامسة	٧	١٨,٨٢	٣,٨٦	٧,٥٨

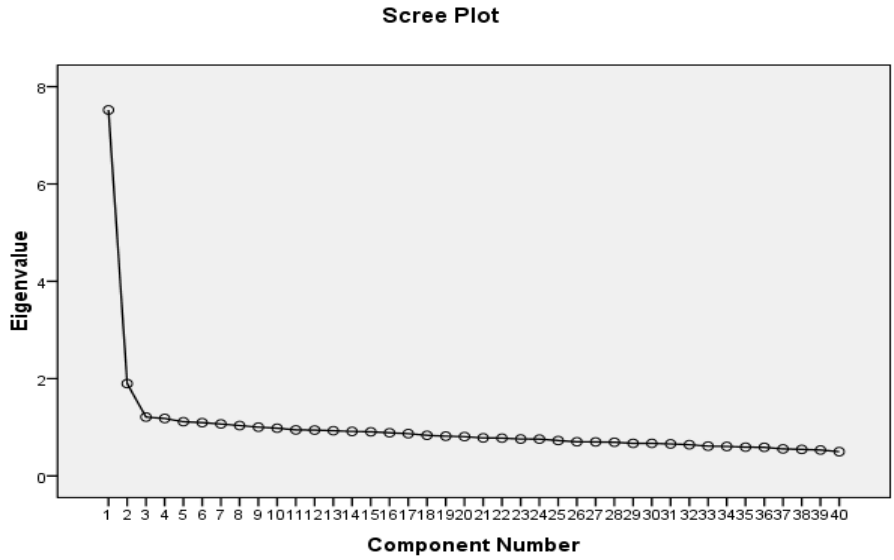
وتبين الاشكال (١، ٢، ٣، ٤، ٥) التمثيل البياني للجذور الكامنة الخاصة بالاختبار عند تطبيقه على العينات الخمسة، أو ما يعرف باختبار فرز العوامل (Scree Test). وهو يؤكد على وجود عامل سائد يعبر عن السمة المقاسة، حيث أن هناك انحناء في الرسم بعد العامل الأول.



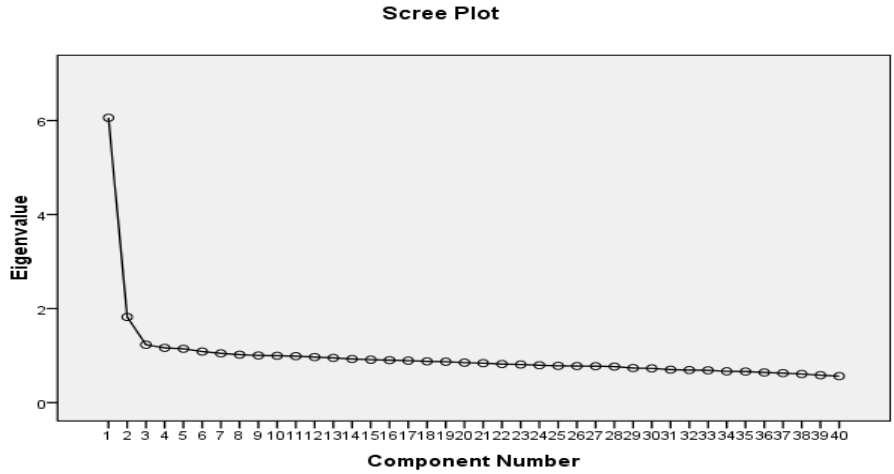
شكل (١): التمثيل البياني للجذور الكامنة الخاصة بالاختبار المطبق على العينة الاولى.



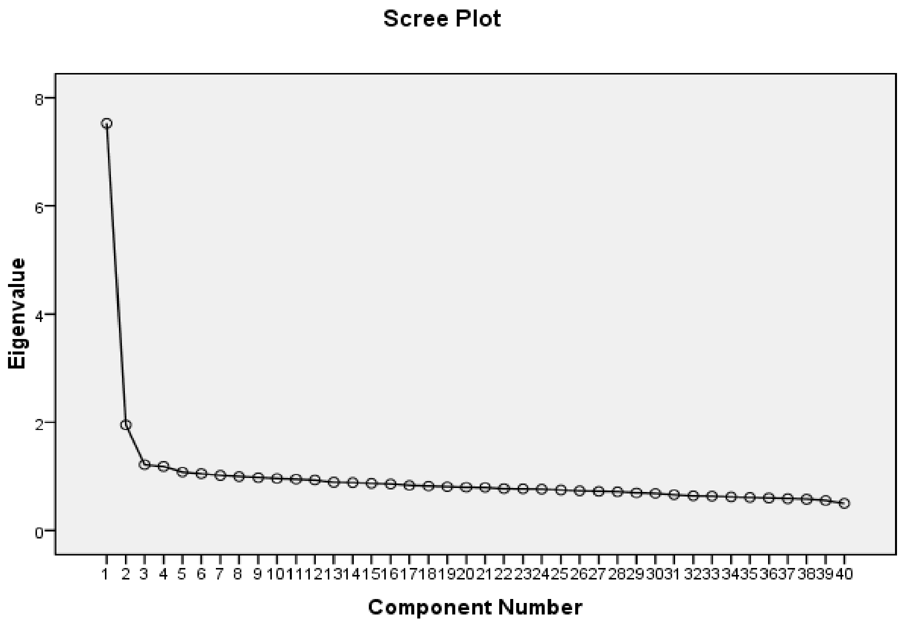
شكل (٢): التمثيل البياني للجذور الكامنة الخاصة بالاختبار المطبق على العينة الثانية.



شكل (٣): التمثيل البياني للجذور الكامنة الخاصة بالاختبار المطبق على العينة الثالثة.



شكل (٤): التمثيل البياني للجذور الكامنة الخاصة بالاختبار المطبق على العينة الرابعة.



شكل (٥): التمثيل البياني للجذور الكامنة الخاصة بالاختبار المطبق على العينة الخامسة.

كما تم الاعتماد على برنامج Bilog-mg3 لتحليل استجابات الأفراد، والبرنامج الاحصائي spss حيث تضمنت عملية التحليل إيجاد معالم الفقرات بالاعتماد على النظرية الحديثة في القياس حسب النموذج الثلاثي، وتقدير دالة المعلومات والخطأ المعياري في تقدير دالة المعلومات لخمسة نماذج من الاختبار.

نتائج الدراسة ومناقشتها

للإجابة على السؤال الأول: هل تختلف تقديرات دالة المعلومات باختلاف حجم العينة المستخدمة في تقديره؟

للإجابة عن هذا السؤال تم تقدير دالة المعلومات لجميع فقرات الاختبار عند تطبيقه على كل عينة من العينات الخمسة، وحساب المتوسط الحسابي لدوال المعلومات الخاصة بكل اختبار والجدول (٢) يبين متوسطات تقديرات دالة المعلومات في حالة تطبيق الاختبار على الخمسة عينات.

جدول (٢): متوسطات تقديرات دالة المعلومات باختلاف حجم العينة.

حجم العينة	دالة المعلومات
٥٠٠	٠,٥٢
١٠٠٠	٠,٦٥
١٥٠٠	٠,٦٧
٢٠٠٠	٠,٧٩
٢٥٠٠	١,٣٨

يتبين من الجدول (٢) أن مقدار تقديرات دالة المعلومات تتغير بتغير حجم العينة حيث تغير متوسط دالة المعلومات من (٠,٥٢) عندما كان حجم العينة (٥٠٠) طالب وطالبة حتى وصل إلى (١,٣٨) عندما كان حجم العينة (٢٥٠٠) طالب وطالبة، كما يلاحظ أن الفرق زاد عندما زاد حجم العينة من (٢٠٠٠) طالب وطالبة إلى (٢٥٠٠) طالب وطالبة وربما تكون الزيادة في دالة المعلومات بزيادة حجم العينة تعود إلى الزيادة في صعوبة الفقرات وتمييزها، أو لأنه بزيادة حجم العينة يقترب متوسط معلمة التخمين من الصفر وهذا يؤدي إلى زيادة في دالة معلومات الاختبار حيث أن كمية المعلومات تزداد كلما قلت قمية التخمين، وقد يكون سبب ذلك أنه عند الاعتماد على النموذج الثلاثي يتم أخذ معلمة التخمين بعين الاعتبار مما يقلل من أثر التخمين وبالتالي زيادة دالة معلومات الاختبار.

للإجابة على السؤال الثاني: هل يختلف الخطأ المعياري في تقدير دالة المعلومات باختلاف حجم العينة المستخدمة في تقديره؟

للإجابة عن هذا السؤال تم ايجاد الخطأ المعياري في تقدير دالة المعلومات لجميع فقرات الاختبار عند تطبيقه على كل عينة من العينات الخمسة، وحساب المتوسط الحسابي للخطا المعياري لتقدير دوال المعلومات الخاصة بكل اختبار، والجدول (٣) يبين متوسطات الخطا المعياري لدالة المعلومات في حالة تطبيق الاختبار على الخمسة عينات

جدول (٣): متوسطات الاخطاء المعيارية لتقديرات دالة المعلومات باختلاف حجم العينة.

حجم العينة	الخطأ المعياري
٥٠٠	١,٣٤
١٠٠٠	١,٢٣
١٥٠٠	١,٢٠
٢٠٠٠	١,١٢
٢٥٠٠	٠,٨٧

يتبين من الجدول (٣) أن مقدار الخطا المعياري لتقديرات دوال المعلومات يتغير بتغير حجم العينة، حيث يتناقص بزيادة حجم العينة، حيث تغير متوسط الاخطاء المعيارية لدوال

المعلومات من (١,٣٤) عندما كان حجم العينـة (٥٠٠) طالب وطالبـة حتـى وصـل إلـى (٠,٨٧) عندما كان حجم العينة (٢٥٠٠) طالب وطالبة، حيث أن الخطأ المعياري يتناسب عكسيا مع حجم العينة حيث انه بزيادة حجم العينة تقل قمية الخطأ المعياري وهذا منطقي؛ إذ أن كمية المعلومـات عند اي مستوى من مستويات القدرة تتناسب عكسيا مع الخطأ المعيـاري، أو لأنـه بزيـادة حجم العينة يقترب متوسط معلمة التخمين من الصفر وهذا يؤدي إلى التقليل من قيمة الخطأ المعيـاري، حيث أن قيمة الخطأ المعياري تقل كلما قلت قمية التخمين، وقد يكون سبب ذلك أنـه عند الاعتمـاد على النموذج الثلاثي يتم أخذ معلمـة التخمين بعين الاعتبـار ممـا يقلل من أثر التخمين وبالتـالي التقليل من قيمة الخطأ المعياري.

المقترحات

١. مراعـاة حجـم عينـة المفحوصـين عنـد بنـاء الاختبـارات بالاعتمـاد علـى النظريـة الحديثـة فـي القياس، وذلك بزيادة عددها قدر المسطاع؛ للتقليـل مـن أخطـاء القيـاس وبالتـالي زيـادة الدقـة في القياس.

٢. اجراء دراسة للمقارنـة بين دالة المعلومات في حالة حسابها بالاعتمـاد علـى النمـوذج الثنـائي والثلاثي المعلمة.

٣. توظيف تقدير دالة المعلومات عند بناء بنوك الاسئلة.

المراجع العربية والأجنبية

- الثوابية، أحمد (٢٠١٠). "أثر حجم العينة على تقدير صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة". مجلة جامعة دمشق. ٢٦ (١).
- عبابنة، عماد. (٢٠٠٤). "اثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة". رسالة دكتوراه غير منشورة. جامعة عمان العربية. الاردن.
- علام، صلاح الدين. (٢٠٠٢). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجيهاته. دار الفكر العربي. القاهرة.
- Brannick, M. (2003). “Basics of IRT one – linefile”: //a:/item response theory.htm.
- Crocker, L. &. Algina, J. (1986). **“**Introduction to classical and modern test theory**”.** University of Florida. U.S.A.
- Embretson, S. & Reiase, S. (2000). “Item Response Theory for Psychologists”. New jersey*:* Lawrence Erlbaum Associates. Inc.
- Fitzpatrick, Ann. R. (2009). “The Impact of Anchor Test Configuration on Student Proficiency Rates”. Educational Measurement: Issues and Practice. 27(4). 34-40 Win 2008.

- Hambleton, R.k. & Swaminathan, H. & Rogers, H.j. (1991). "Fundamentals of Item Response Theory: International Educational and Professional". Publisher Newbury park.
- Henderson, G.M. & Dianne, J.K. (2001). "Minimizing the influence of item parameter Estimation Errors in Test Development: A Comparison of Three Selection Procedures". Journal of Experimental Education. 69(3).
- Hambleton, R.k. & Swaminathan, H. (1985). "Item Response Theory: Principles and applications". Boston MA: Kluwer-Nyjhoff.
- Hamleton, R. H & Jones, R. W. (1994). "Item Parameter estimation errors and their Influence on test information function". Applied Measurement in education. 7 (3). 171 - 186.
- Lazarsfeld, P .F. (1950). "The Logical and Mathematical Foundation of Latent structure Analysis". Measurement and prediction. Presentation University Press
- Pommerich, M. (2007). "The Effect of Using Item Parameters Calibrated from Paper Administrations in Computer Adaptive Test Administrations". Journal of Technology. Learning. and Assessment. 5(7). Mar 2007.
- Traub, R.E. & wolfe, R.G. (1981). "Laten trait theories and the assessment of educational achievement". In D.C Berliner (ED). Review of Research in education: vol.9/woshing-ton. DC. American educational research Association.
- Swaminathan, H et al. (2003). "Small sample Estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates". Journal of Applied Psychological Measurement. 27 (1). 27-51.
- Stocking, M.K. (1990). "Specifying optimum examinees for item parameters estimation in item response theory". Psyehometrika. 55(3). 361-475.
- Wanger, T. A. & Harvey, R. J. (2003). "Developing anew critical thinking test using item response theory". Paper presented at the 2003 annual Conference of society for industrial and organizational psychology. Orlando.