تقدير ثبات الاختبار وتأثره بعدد أبعاده والعلاقة بينها وتوزيع قدرة المفحوصين

إعـداد

د.يوسف عبدالقادر علي ابوشندي

استاذ مساعد في القياس والتقويم

د. ايهاب محمد نجيب عماره

استاذ مساعد في القياس والتقويم

د.راشد سيف المحرزي

استاذ مشارك في القياس والتقويم

الملخص:

هدفت الدراسة تفحص أثر بعدية الاختبار والارتباط بين الأبعاد و توزيع قدرة المفحوصين على تقدير ثبات الاختبار باستخدام ألفا كرونباخ، والتجزئة النصفية، واستجابة الفقرة. ولتحقيق غرض الدراسة تم توليد استجابات ست مجموعات من المفحوصين (١٠٠٠ مفحوص لكل مجموعة) على ستة اختبارات.

توصلت الدراسة إلى أن تقدير الثبات للاختبارات أحادية البعد كان أعلى من ثنائية البعد حسب الطريقتين: الفا كرونباخ والتجزئة النصفية، ولم يختلف ثبات استجابة الفقرة، وزادت قيمة معامل الثبات بزيادة الارتباط بين الأبعاد عند التوزيع الطبيعي لقدرات المفحوصين.

وخلصت النتائج إلى أن تقدير ثبات الفا كرونباخ كان أعلى عندما توزعت القدرات بشكل متسق منها عند التوزيع الطبيعي، وكذلك عند حالتي الارتباط (0.6, 1.0) لثبات التجزئة النصفية. وعلى العكس في ثبات استجابة الفقرة، وحالة قوة الارتباط 0.0 لثبات التجزئة النصفية.

وزاد ثبات الاختبار بزياده طوله، وتأثر ثبات التجزئة النصفية بطريقة ترتيب الفقرات. كما كشفت النتائج أن تقدير ثبات استجابة الفقرة كان أعلى من ثبات النظرية التقليدية.

الكلمات المفتاحية: ثبات الاختبار، ألفا كرونباخ، التجزئة النصفية، نظرية استجابة الفقرة، بعدية الاختبار، قوة الارتباط بين الأبعاد، توزيع قدرة المفحوصين.

Estimation of the stability of the test and its impact on the number of dimensions and the relationship between them And the distribution of the capacity of examinees

Dr. yousf Abdelkader Aly Abo shendy

Dr.Ehab Mohamed Nageb Emarh **Dr.Rashed Seif Elmhrzy**

ABSTRACT:

This study aimed to investigating the effect of test dimensionality and the correlation between dimensions and ability distribution on reliability coefficient. To achieve the aim of the study, responses for four groups (1000 examinee for each group) were generated on six tests.

The results show that unidimensional tests reliability (alpha crounbach and split- half) greater than bidimensional tests, and no differences in IRT reliability. And reliability estimated were increased when correlation between dimensions increased.

The results also revealed that alpha coefficient greater than when uniform ability distribution than normal, also when (r= 0.6, 1.0) for split-half coefficient. And inverse otherwise. A reliability coefficient varies increasingly according to the increasing test lengths, and split- half coefficient is affected by items forming. Finaly, IRT reliability coefficient is greater than classical coefficient.

Key words: Test reliability, Alpha crounbach, Split-half, Item response theory, Test dimensionality, Correlation between dimensions, Ability distribution.

المقدمة :

تعتمد صحة القياس على مدى ثبات نتائجه وصدقها، فالمقياس الثابت يعطي النتائج نفسها اذا قاس نفس السمة مرات متتالية، وتعتمد فكرة الثبات على مدى انحراف درجة كل فرد في التطبيق الأول للمقياس عنها في التطبيق الثاني، ويمكن الاعتماد على التباين كمقياس لهذا الانحراف، وبذلك يعرف الثبات بأنه الجزء الحقيقي من التباين العام للمقياس (السيد،١٩٧٩).

ويعرف عودة (١٩٩٣) الثبات احصائيا بأنه نسبة التباين الحقيقي إلى التباين الكلي؛ أي كم من التباين الكلي في الدرجات يمكن أن يكون تباينا حقيقيا، و يعبر عن هذه النسبة احصائيا بمربع معامل الارتباط بين العلامة الحقيقية والعلامة الظاهرية. ويؤكد على عدم القدرة للوصول إلى علامات حقيقية، لذلك لا يمكن الاعتماد على هذه الطريقة في حساب الثبات، لذلك يلجأ إلى العلامة الظاهرية في تقدير الثبات، اي في الحصول على مؤشر احصائي نحكم من خلاله على دقة القياس، يسمى هذا المؤشر بمعامل الثبات *Reliability Coefficient*.

وبين الين وين (Allen & Yen,1979) ست صيغ مختلفة لحساب معامل ثبات الاختبار هي:

- $\rho_{xx'}$: الارتباط بين العلامات على الاختبار، والعلامات على صورة مكافئة للاختبار.
- $\rho^2_{xx'}$: نسبة التباين في علامات الاختبار التي يمكن تفسيرها من خلال علاقة خطية مع العلامات على صورة مكافئة للاختبار.
- $=\frac{\sigma^2_T}{\sigma^2_X}\rho_{xx'}$: نسبة تباين العلامة الحقيقية إلى تباين العلامة الظاهرية.
- $=\rho^2_{XT}\rho_{xx'}$: مربع معامل الارتباط بين العلامات الحقيقية والعلامات الظاهرية.
- $=1-\rho^2_{XE}\rho_{xx'}$: ١- مربع معامل الارتباط بين العلامات الظاهرية وعلامات الخطأ.
- $=1-\frac{\sigma^2_E}{\sigma^2_X}\rho_{xx'}$: نسبة تباين العلامات الخطأ إلى تباين العلامات الظاهرية.

غير أنه لا يمكن استخدام الصيغ السابقة لحساب معامل الثبات؛ لأن الدرجات الحقيقية في الاختبار تكون غير معلومة، لذلك يمكن تقدير قيمة معامل ثبات الاختبار باستخدام طرائق منها: الصيغ

المتوازية، الاختبار وإعادة الاختبار،التجزئة النصفية المصحح (معادلة سبيرمان براون، معادلة رولون، معادلة جتمان، معادلة جلكسون)، تحليل التباين، والاتساق الداخلي (الفا كرونباخ) وهي الطريقة الأكثر شيوعا.

وتتقيد هذه الطرائق بعينة المفحوصين؛ أي أن تقديراتها تختلف باختلاف عينة المفحوصين كحال بقية التقديرات في النظرية التقليدية في القياس مما يقلل من منفعتها في تقدير ثبات المقاييس. كما أن تقدير الثبات بهذه الطرق يعتمد على دقة تقدير الخطأ المعياري للقياس.

في نظرية استجابة الفقرة يرتبط مفهوم الثبات بدالة معلومات الفقرة $I_i(\Theta)$ ودالة معلومات الاختبار $I_T(\Theta)$ وفي الخطأ المعياري لتقدير قدرات المفحوصين S_{EE} ، حيث بين ثيسن (Thissen, 2000) أن الطريقة الأفضل لتقدير معامل الثبات بالاعتماد على دالة معلومات الاختبار، وبين العلاقة بين الثبات ودالة

معلومات الاختبار حسب المعادلة التالية:

$$R_{xx} = 1 - \frac{1}{\sum_{i=1}^{n} I(\Theta)} \qquad (1)$$

تشير R_{xx} إلى معامل ثبات الاختبار، وتشير $I(\Theta)$ إلى دالة معلومات الفقرة . وتؤكد هذه المعادلة على ان العلاقة بين دالة معلومات الاختبار وثبات الاختبار هي علاقة طردية.

كما قام لورد (Lord, 1980) باشتقاق المعادلة ٢ لتقدير ثبات الاختبار

$$\rho_{xx'} = \frac{\sum_1^N (\sum_1^n P_{ia})^2 - (\sum_1^N \sum_1^n P_{ia})^2/N}{\sum_1^N \sum_1^n P_{ia} Q_{ia} + \sum_1^N (\sum_1^n P_{ia})^2 - (\sum_1^N \sum_1^n P_{ia})^2/N} \qquad (2)$$

n:عدد الفقرات،N: عدد المفحوصين، P_{ia} :احتمال استجابة المفحوص i على الفقرة a استجابة صحيحة.

وحاول هامون ومصباح (Hamon & Mesbah, 2002) التوصل إلى صيغة لتقدير معامل الثبات

مشابهة لمعامل ألفا كرونباخ باستخدام نموذج راش وتوصلا إلى اشتقاق الصيغة التالية لتقدير معامل الثبات:

$$\alpha_R = 1 - \frac{\frac{1}{n}\sum_1^n \frac{1}{I(\bar{E},k)}}{S_{\bar{E}}^2} + \sigma(k^{-1}) \qquad (3)$$

تمثل α_R: معامل الفا راش، $S_{\bar{\theta}}^2$: التباين في تقدير معلمة القدرة، $I(\Theta,k)$: دالة معلومات الفقرة، $\sigma(k^{-1})$: الانحراف المعياري لمعلمة الصعوبة.

وبين السيد (1979) أن من العوامل التي تؤثر على ثبات الاختبار:

- عدد اسئلة الاختبار: ترتفع القيمة العددية لمعامل الثبات تبعا لزيادة عدد أسئلة الاختبار.
- زمن الاختبار: يزداد الثبات تبعا لزيادة الزمن حتى يصل إلى الحد المناسب للاختبار ثم يقل بزيادة الزمن.
- تباين علامات الاختبار: يزداد الثبات تبعا لزيادة تباين درجات الاختبار.
- التخمين: ينقص ثبات الاختبار تبعا لزيادة التخمين.
- صياغة الأسئلة: الأسئلة الغامضة والخادعة والعاطفية والطويلة تقلل الثبات.
- حالة المفحوص: يتأثر الثبات بحالة الفرد الصحية والنفسية وبمدى تدربه على الموقف الاختباري، لذا يؤدي المرض والتعب والتوتر الانفعالي إلى نقصان الثبات.

اهتمت الدراسات المتعلقة بثبات المقاييس بتفحص تأثير بعض الخصائص المتعلقة بالاختبار والمفحوصين على تقدير ثبات ودالة معلومات الاختبار، وبعضها حاول المقارنة بين النظريتين التقليدية والحديثة في تقدير معامل الثبات. وقارنت هذه الدراسات بين معامل ألفا كرونباخ ونموذج راش. ومن هذه الدراسات دراسة كورتينا (Cortina, 1993) التي هدفت إلى توضيح النقص في فهم وتطبيق معامل الفا كرونباخ في البحوث التربوية، وهدفت إلى تفحص درجة تغير معامل الفا كرونباخ بتغير عدد الفقرات المشكلة للاختبار(6, 12, 18) ومعدل الارتباط بين الفقرات (0.3, 0.5, 0.7) وبعدية الاختبار (أحادي، ثنائي، ثلاثي). توصلت الدراسة إلى أن قيمة الفا كرونباخ تزداد بزيادة عدد الفقرات المشكلة للاختبار. وتزداد بزيادة معدل الارتباط بين الفقرات المشكلة للاختبار. ويقل بزيادة عدد أبعاد الاختبار.

وهدفت دراسة جمحاوي (2000) المقارنة بين النظرية التقليدية ونظرية الاستجابة للفقرة في تقديرات معالم الفقرات، وتقديرات معامل ثبات الاختبار، حيث قامت الباحثة بتقدير معامل الثبات باستخدام ألفا كرونباخ، وحسب نماذج نظرية استجابة الفقرة اللوجستية. ولتحقيق الهدف من الدراسة استخدمت اختبار للقدرة العقلية، وطبقت الاختبار على عينة مؤلفة من 1061 طالباً. أشارت نتائج الدراسة

المتعلقة بتقديرات معامل ثبات الاختبار إلى عدم وجود فروق بين معاملي ثبات الاختبار الناتج من النظرية التقليدية (r_{xx}=0.841) والنموذج ثنائي المعلمة (r_{xx}=0.843)، بينما وجدت فروق ذات دلالة إحصائية بين معامل الثبات للاختبار الناتج من النظرية التقليدية والناتج من نموذج راش (r_{xx}=0.743) والنموذج ثلاثي المعلمة(r_{xx}=0.836).

في دراسة هومان ومصباح (Homan & Mesbah, 2002) والتي هدفت إلى اشتقاق صيغة جديدة لتقدير معامل الثبات تحت شرط نموذج راش(نظرية استجابة الفقرة) مشابه لمعامل الفا كرونباخ. ومن ثم قاما بمقارنة التقديرات باستخدام المعاملين من خلال بيانات مولدة وبيانات حقيقية. توصلت الدراسة إلى اشتقاق الصيغة (٣ ص ٤) لتقدير معامل ثبات الفا راش، ومن خلال مقارنة التقديرين على البيانات المولدة توصلا إلى أن قيم معاملات الثبات الفا كرونباخ والفا راش زادت بازدياد الانحراف المعياري للسمة الكامنة. وتوصلت النتائج إلى أنه عند تثبيت تباين السمة الكامنة لا تختلف قيمة معامل الثبات. خلصت النتائج إلى ان معامل ثبات الفا كرونباخ كان قريبا جدا من معامل ثبات الفا راش، وكان الفا كرونباخ دائما اقل. وهذا يؤشر على أن الفا كرونباخ طريقة جيدة لتقدير الثبات في حالة أن البيانات تطابق نموذج راش.

وقارنت دراسة إبراهيم والعزابي (2011) بين الفا كرونباخ ونموذج راش في تقدير صدق وثبات المقاييس. طبق الباحثان لأغراض الدراسة مقياس مارش لتقييم الطلبة في جودة التعليم المكون من ٢٩ فقرة على عينة مكونة من ٧٥٢ طالباً وطالبة من أربع جامعات ماليزية. خلصت الدراسة إلى ان نموذج راش أفضل في تقدير ثبات المقاييس من ألفا كرونباخ، وبيَنا أن تقدير الثبات باستخدام نموذج راش لا يتأثر بتغير طول الاختبار كما يتأثر ألفا كرونباخ.

وتفحص العمري (2011) تأثير طريقة ترتيب الفقرات في المقاييس النفسية على خصائص المقياس وخصائص فقراته، ورتب الباحث الفقرات حسب: تمييزها، ثباتها، صدقها، وترتيب عشوائي. لتحقيق أغراض الدراسة طبق الباحث مقياس للثقة بالنفس مكون من ٤٠ فقرة على عينة مكونة من ٢٦٤ طالباً. توصلت الدراسة إلى أن أعلى قيمة لمعامل الفا كرونباخ عندما تم ترتيب الفقرات بالاعتماد على ثبات الفقرات، ومن ثم ترتيب الفقرات حسب تمييز الفقرات، وأقلها كان عندما رتبت الفقرات بشكل عشوائي.

وهدفت دراسة بني عطا والشريفين (2012) تفحص أثر اختلاف شكل توزيع قدرة المفحوصين(طبيعي، ملتو نحو اليمين، ملتو نحو اليسار، ثنائي المنوال) على تقديرات: معالم الفقرات، و دالة معلومات الاختبار والخطأ المعياري للقياس. وذلك من خلال بيانات مولدة وفق النوذج اللوجستي ثلاثي المعلمة على اختبار مكون من ٦٠ فقرة ثنائية الاستجابة. كشفت نتائج الدراسة عن وجود أثر لاختلاف توزيع القدرات على تقديرات معاملات صعوبة وتمييز الفقرات. وفيما يتعلق بدالة معلومات الاختبار أظهرت نتائج الدراسة أن هناك تأثير لاختلاف شكل توزيع قدرات المفحوصين على تقدير دالة معلومات الاختبار وتوصلت إلى أن أكبر قيمة لدالة المعلومات كانت عندما كان توزيع القدرات ثنائي المنوال، وأقل قيمة عندما توزعت القدرات بشكل طبيعي. ولأن مفهوم الثبات يرتبط بدالة المعلومات فقد قام الباحثان بمقارنة تقديرات معامل الثبات باستخدام طريقتين: الفا كرونباخ، والامبريقى(نظرية استجابة الفقرة) عند الحالات المختلفة من توزيع قدرات المفحوصين، وتوصلت الدراسة إلى أن تقديرات الفا كرونباخ اختلفت باختلاف توزيع قدرات المفحوصين، حيث تماثلت للتوزيعين الطبيعي والملتو نحو اليمين. والتوزيعين الملتو نحو اليسار وثنائي المنوال. أما الثبات الامبريقي فقد كانت قيمته الأقل بالنسبة للتوزيع الطبيعي والاعلى للتوزيع ثنائي المنوال. كمابينت النتائج وجود اختلافات في تقدير معاملات الثبات باختلاف طريقة التقدير(الفا كرونباخ، الامبريقي) وكانت التقديرات أعلى باستخدام الطريقة الامبريقية ولجميع الحالات من توزيع قدرات المفحوصين.

وفي دراسة الزبون (2013) التي هدفت معرفة أثر اختلاف حجم عينة المفحوصين على تقدير دالة معلومات الاختبار والخطأ المعياري في القياس، وذلك من خلال تطبيق الاختبار الوطني لضبط نوعية التعليم في الرياضيات على عينة مكونة من ٧٥٠٠ طالب من طلبة الصف الثامن الأساسي تم اختيار خمس مجموعات بطريقة عشوائية منهم (500, 1000,1500,2000,2500)، وتكون الاختبار من ٤٠ فقرة من نوع الاختيار من متعدد. وتوصلت الدراسة إلى أن حجم العينة يؤثر على تقدير دالة معلومات الاختبار، حيث زادت دالة معلومات الاختبار بزيادة حجم عينة المفحوصين. وعلى العكس بالنسبة للخطأ المعياري للقياس حيث كان يقل بزيادة حجم عينة المفحوصين.

من خلال مراجعة الدراسات السابقة تبين أن الدراسات التي تفحصت تأثير بعض المتغيرات على تقدير معامل الثبات لم تقارن بين التوزيعين الطبيعي والمتسق لمعلمة قدرة المفحوصين. كما أن أي من

هذه الدراسات لم تتفحص تأثير قوة الارتباط بين الأبعاد وإنما تناولت معدل الارتباط بين الفقرات. كما لم تبحث الدراسات في تأثير عدد أبعاد الاختبار وقوة الارتباط بينها على تقدير الثبات بالاعتماد على النظرية الحديثة، واكتفت في البحث في تأثيرها على ثبات الفا كرونباخ والتجزئة النصفية.

دراسات المقارنة بين تقديرات الثبات بالاعتماد على النظريتين التقليدية والحديثة اكتفت بالمقارنة بين ثبات ألفا كرونباخ ونموذج راش، أو النموذج اللوجستي ثنائي وثلاثي المعلمة، ولم تهتم بمقارنة ثبات التجزئة النصفي، ولا مقارنة النماذج التراكمية الصاعدة Normal Ogive Models. الدراسات التي اهتمت بتفحص تأثير ترتيب الفقرات على تقدير الثبات لم تقارن بين الترتيبين: (فردي وزوجي) وهي الطريقة الشائعة لترتيب الفقرات، وحسب البعد الذي تنتمي له الفقرة.

مشكلة الدراسة وأسئلتها

يتخذ المعلمون وغيرهم من المسؤولين التربويين قرارات مختلفة تتعلق بالطالب، تستند هذه القرارات إلى درجات أدوات القياس والتقويم الصفي وخاصة الاختبارات، لذلك يجب ان يحرصوا على ان تتوافر في هذه الأدوات خصائص أساسية جيدة لكي يثقوا في هذه القرارات، ومن هذه الخصائص ثبات الاختبار.

يؤشر ثبات درجات الاختبار إلى مدى دقة درجات أدوات القياس كدليل للتحصيل الحقيقي للطلبة، ويأمل المعلمون أن تكون درجات اختباراتهم ثابتة ومتسقة، ألا أنهم لا يتحققوا من ثبات اختباراتهم، ويرجعون السبب في ذلك إلى قلة الوقت وانشغالهم بالتدريس، كما ويرجعون ذلك إلى عدم درايتهم بالأساليب الاحصائية اللازمة لتقدير الثبات (علام، ٢٠٠٥).

وتعاني الطرق التقليدية في القياس من عدم الدقة في التقديرات المختلفة لخصائص الاختبارات الناتجة عن عدم أخذ الارتباط بين فقرات الاختبار بالحسبان، فكما يشير زينسكي وآخرون (Zenisky et al, 2002) إلى أن هناك عدة باحثين (Anastasi, 1961; Guilford, 1936; Kelly, 1927; Thorndike, 1951) أكدوا على أهمية تجميع الفقرات التي تنتمي إلى مثير مشترك في نفس نصف الاختبار عند تقدير ثبات الاختبار باستخدام طريقة التجزئة النصفية، وعدم فعل ذلك سيؤدي إلى تضخيم الثبات المقدَر لأن ذلك سيضخم معامل الارتباط بين نصفي الاختبار.

في نظرية استجابة الفقرة يساعد معامل الثبات في تقدير الارتباط الموضعي بين الفقرات، حيث توصل النعيمي (2006) إلى ضرورة أخذ الارتباط بين الفقرات والأبعاد بعين الاعتبار عند تقدير معامل ألفا كرونباخ.

ويؤكد علام (٢٠١١) على أن ثبات درجات الاختبارات الصفية يتأثر بعدد الأسئلة المشكلة للاختبار، ويبين أن الاختبار الذي يحتوي على عدد فقرات قليل يحتمل أن يكون ثبات درجاته منخفضاً، وينصح بزيادة طول الاختبار بغرض زيادة درجة ثباته. ولكن امبرتسون وريس (Embretson & Reise, 2000) وضحَا العلاقة بين طول الاختبار ودرجة ثباته، وأكدا ما بينه علام (٢٠١١) أن العلاقة طردية بين طول الاختبار ودرجة ثباته وذلك حسب النظرية التقليدية، ولكن وحسب نظرية استجابة الفقرة بينَا ان الاختبارات القصيرة يمكن أن تكون أكثر ثباتا من الاختبارات الطويلة.

جاءت هذه الدراسة لتحاول تفحص تغير تقدير ثبات الاختبار بتغير عدد أبعاد الاختبار (أحادي البعد، ثنائي البعد)، وتغير توزيع معلمة قدرة المفحوصين (توزيع طبيعي بوسط حسابي صفر وانحراف معياري ١، وتوزيع متسق [4+ ,4-])، وتغير قوة العلاقة الارتباطية بين الأبعاد في حالة الاختبارات ثنائية البعد (0.0 ,0.6 ,1.0). وستحاول تفحص تأثير ترتيب الفقرات في جزأي الاختبار ثنائي البعد (فردي زوجي، حسب البعد الذي تنتمي إليه الفقرة) على تقدير الثبات، وذلك من خلال بيانات مولدة. وبشكل أكثر تحديدا ستحاول الدراسة الإجابة عن الأسئلة التالية:

١. هل يختلف تقدير معامل الثبات باستخدام طريقة ألفا كرونباخ باختلاف:عدد ابعاد الاختبار(بعد واحد، بعدان)، و قوة الأرتباط بين الأبعاد (0.0, 0.6, 1.0)، شكل توزيع قدرة المفحوصين (طبيعي، متسق)؟ وهل يختلف باختلاف عدد الفقرات المشكلة للاختبار؟
٢. هل يختلف تقدير معامل الثبات باستخدام طريقة التجزئة النصفية باختلاف:عدد أبعاد الاختبار(بعد واحد، بعدان)، و قوة الارتباط بين الأبعاد (0.0, 0.6, 1.0)، شكل توزيع قدرة المفحوصين (طبيعي، متسق)؟ وهل يختلف باختلاف عدد الفقرات المشكلة للاختبار؟
٣. هل يختلف تقدير معامل الثبات باستخدام نظرية استجابة الفقرة باختلاف:عدد أبعاد الاختبار(بعد واحد، بعدان)، و قوة الارتباط بين الأبعاد (0.0, 0.6, 1.0)، شكل توزيع قدرة المفحوصين (طبيعي، متسق)؟ وهل يختلف باختلاف عدد الفقرات المشكلة للاختبار؟

٤. هل يؤثر ترتيب الفقرات في جزأي الاختبار على تقدير الثبات باستخدام: كرونباخ الفا، التجزئة النصفية، نظرية استجابة الفقرة؟

٥. هل يختلف تقدير ثبات الاختبار باختلاف طريقة التقدير: ألفا كرونباخ، التجزئة النصفية، نظرية استجابة الفقرة؟

أهمية الدراسة

يعدَ الثبات من أهم خصائص الاختبارات، فكلما زاد ثبات الاختبار كلما قلت أخطاء القياس، وبالتالي زاد الاقتراب من العلامة الحقيقية والوثوق بنتيجة القياس وبالتالي الدقة في اتخاذ القرارات وإختيار المعالجات المناسبة. لذلك تنبع أهمية هذه الدراسة من الحاجة الملحة للتوصل إلى العوامل التي قد تؤثر على دقة تقدير ثبات الاختبار، مما قد يساعد معدو الاختبارات والباحثين في اختيار الطريقة المناسبة لتقديره.

وتظهر أهمية هذه الدراسة في أنها قارنت بين طرائق تقدير الثبات حسب النظرية الكلاسيكية(ألفا كرونباخ، التجزئة النصفية) وطريقة تقدير الثبات حسب نظرية استجابة الفقرة بالاعتماد على دالة معلومات الاختبار.

وتنفرد هذه الدراسة في تفحصها لتأثير شكلين من أشكال توزيع معلمة قدرة المفحوصين (الطبيعي والمتسق) على تقدير معامل الثبات . وكذلك في تفحصها تأثير لمستويين من بعدية الاختبار (أحادي البعد، ثنائي البعد)، وثلاثة مستويات من الارتباط بين الأبعاد(0.0, 0.6, 1.0) وليس معدل الارتباط بين الفقرات . وتتجلى أهمية هذه الدراسة في مقارنتها بين طريقتين من طرائق ترتيب الفقرات(فردي زوجي، حسب البعد) وتفحص تأثير كل منها على تقدير معامل ثبات الاختبار.

مصطلحات الدراسة

ثبات ألفا كرونباخ: حساب الارتباطات بين العلامات لمجموعة الثبات على جميع الفقرات الداخلة في الاختبار، ويتم قسمة الاختبار إلى عدد من الأجزاء يساوي عدد فقراته(كل فقرة تشكل اختباراً فرعياً)، ومن ثم يحسب متوسط معاملات الارتباط بين الفقرات.

ثبات التجزئة النصفية: معامل الارتباط بين علامات مجموعة الثبات على صورتين متكافئتين يتم تكوينهما بقسمة فقرات الاختبار إلى قسمين، وتشير قيمة معامل الارتباط المحسوب إلى معامل ثبات نصف الاختبار لذلك لا بد من تصحيح معامل الثبات النصفي بمعادلة تأخذ بعين الاعتبار الزيادة المتوقعة لمعامل الثبات بزيادة عدد فقراته وهي معادلة سبيرمان– براون.

ثبات نظرية استجابة الفقرة: مقلوب دالة معلومات الاختبار مطروح من واحد صحيح.

الطريقة والاجراءات

إجراءات توليد البيانات

توفر المحاكاة Simulation أحيانا بيانات تخدم الدراسات العلمية لاتوفرها البيانات الواقعية؛ وتمتاز هذه البيانات بوجود القيم الحقيقية لمعالم الفقرات والمفحوصين مما يتيح فرصة للباحثين للتحقق من دقة تقدير هذه المعالم والكشف عن أخطاء التقدير وبالتالي التدخل للتقليل من هذه الأخطاء. كما وتوفر دراسات المحاكاة نتائج نظرية مما قد يشجع الباحثين على التحقق من هذه النتائج بطرق إمبريقية. ومن المنافع التي تقدمها دراسات المحاكاة الحد من أخطاء القياس التي ترجع إلى ظروف تطبيق أدوات القياس مثل الغش وقلق الاختبار وغيرها (Davey, Nering & Thompson, 1997).

وفي هذه الدراسة تم توليد البيانات باستخدام برمجية RESGEN الذي صممها موراكي (Muraki,2000)، وتتطلب هذه البرمجية تحديد المعالم الحقيقية للفقرات، وخصائص توزيع قدرات المفحوصين، والنموذج المستخدم في توليد البيانات.

واستخدم النموذج التراكمي الصاعد ثنائي المعلم التعويضي (Compensatory Normal Ogive Model - 2PL) لتوليد البيانات وتقدير المعالم.

$$P(U_{ij} = 1|\Theta_j, a_i, d_i)=\frac{1}{\sqrt{2\pi}} \int_{ai\boxdot j+di}^{\infty} e^{\frac{-t^2}{2}} dt \qquad (4)$$

ويحتاج هذا النموذج إلى قيمة واحدة لتشير إلى صعوبة الفقرة. وبما أن الاختبارات ثنائية البعد هي من نوع الاختبارات متعددة الأبعاد بين الفقرات Between Items Multidimensional Test؛ أي أن الفقرات أحادية بعد تشكل اختباراً متعدد الأبعاد، وبناء على ذلك فإن كل فقرة من فقرات الاختبار يمكن تمثيلها

بقيمة واحدة لمعلمة التمييز على البعد الذي تنتمي له الفقرة. وقد استخدم النموذج نفسه في تقدير معالم الفقرات والمفحوصين مما يساعد على الوصول إلى نتائج دقيقة.

بعد اختيار المعالم الحقيقية للفقرات تناولت الدراسة توزيعين لمعلمة قدرة المفحوصين: التوزيع الطبيعي المعياري Standard Normal Distribution بوسط حسابي صفر وإنحراف معياري ١، والتوزيع المتسق Uniform Distribution بقيمة ابتدائية -١ وقيمة نهائية +١.

وبعد ذلك تم إدخال القيم الحقيقية لمعالم الفقرات (الصعوبة b_i، التمييز a_i) وتوزيعات معلمة قدرة المفحوصين (Θ) ، وتم توليد استجابات ست مجموعات من المفحوصين على اختبار يتكون من ٦٠ فقرة ثنائية الاستجابة مرة على أساس أنه اختبار أحادي البعد، ومرة على أساس أنه اختبار ثنائي البعد كل منهما مكون من ٣٠ فقرة. واختلفت مجموعات المفحوصين فيما بينها في توزيع القدرة(طبيعي، متسق)، وقوة الارتباط بين الأبعاد (0.0, 0.6, 1.0). و تكونت كل مجموعة من ١٠٠٠ مفحوص.

ومن فقرات الاختبار الأول(٦٠ فقرة) تم الاختيار وبطريقة عشوائية اختبار يتكون من ٣٠ فقرة لتشكل الاختبار الثاني، ومن فقرات الاختبار الثاني تم الاختيار وبطريقة عشوائية اختبار يتكون من ١٠ فقرة.

وبعد الحصول على بيانات محاكاة تتمثل في استجابات ست مجموعات من المفحوصين على ثلاثة اختبارات أحادية البعد (٦٠ فقرة، ٣٠ فقرة، ١٠ فقرات) وثلاثة اختبارات ثنائية البعد ((٣٠،٣٠)، (١٥،١٥)، (٥،٥)). استخدمت هذه الاستجابات في تقدير الثبات بالاعتماد على النظرية الكلاسيكية. ومن ثم استخدمت برمجية NOHARM لتقدير معالم الفقرات وذلك لاستخدامها في تقدير ثبات الاختبار بالاعتماد على نظرية استجابة الفقرة.

إجراءات التحقق من بعدية البيانات

أكد هاتي(Hattie,1984) على أنه عندما يتطابق النموذج مع البيانات، فإن المؤشرات من خلال نماذج السمة الكامنة تبدو من أكثر المؤشرات منطقية في تحديد بعدية البيانات، ويتم اختبار هذه المؤشرات بواسطة تحليل البواقي الذي تحسبه برمجية NOHARM.

في هذه الدراسة تم الاعتماد على أسلوب البواقي Residuals في تفحص بعدية الاختبار وتوافقها مع البيانات. حيث تقوم برمجية NOHARM بحساب مصفوفة البواقي Residuals Matrix، ومنها يتم حساب مجموع مربعات البواقي(Sum of Squares of Residuals(SSR، وحساب الجذر التربيعي لمتوسطات مربعات البواقي Root Mean Squares of Residuals(RMSR). كما وتقوم البرمجية بحساب مؤشر تناكا Tanaka

Index . ويبين جدول١ RMSR و Tanaka Index للاختبارات

جدول١ نتائج تفحص بعدية الاختبارات من خلال مؤشرات RMSR و Tanaka Index

بعدية الاختبار	توزيع قدرة المفحوصين Θ	قوة الارتباط بين الأبعاد	المؤشرات	الاختبار الاول (٦٠ فقرة)	الاختبار الثاني (٣٠ فقرة)	الاختبار الثالث (١٠فقرات)
	طبيعي		RMSR*	٠.٠٠٤٧	٠.٠٠٤٦	٠.٠٠٥١
			TANAKA**	٠.٩٨٩٩	٠.٩٩٢٤	٠.٩٩٥٦
	متسق		RMSR	٠.٠٠٧١	٠.٠٠٧٨	٠.٠٠٧٨
			TANAKA	٠.٩٩٦٦	٠.٩٩٦١	٠.٩٩٦٩
ثنائي البعد	طبيعي	0.0	RMSR	٠.٠٠٤٧	٠.٠٠٤٢	٠.٠٠٣١
			TANAKA	٠.٩٨٤٥	٠.٩٩١٧	٠.٩٩٨١
		0.6	RMSR	٠.٠٠٤٦	٠.٠٠٤٠	٠.٠٠٣٦
			TANAKA	٠.٩٨٦٨	٠.٩٩٢٦	٠.٩٩٧٧
		1.0	RMSR	٠.٠٠٤٦	٠.٠٠٤٢	٠.٠٠٣٤
			TANAKA	٠.٩٩٠٩	٠.٩٩٣٥	٠.٩٩٧٩
	متسق	0.0	RMSR	٠.٠٠٥٦	٠.٠٠٥٩	٠.٠٠٥٨
			TANAKA	٠.٩٩٦١	٠.٩٩٦٣	٠.٩٩٧٢
		0.6	RMSR	٠.٠٠٥٥	٠.٠٠٦٣	٠.٠٠٥٧
			TANAKA	٠.٩٩٦٢	٠.٩٩٥٩	٠.٩٩٧٥
		1.0	RMSR	٠.٠٠٥٥	٠.٠٠٦١	٠.٠٠٤٧
			TANAKA	٠.٩٩٦٢	٠.٩٩٦١	٠.٩٩٨٢

*RMSR: Root mean square of residuals **TANAKA: Tanaka index of goodness of fit

من النتائج الموضحة في جدول١ يتبين أن قيمة الجذر التربيعي لمربعات البواقي كانت منخفضة(قريبة من الصفر) حيث انحصر بين 0.0031 و0.0078. أما مؤشر تناكا فقد كان مرتفعاً وانحصر بين 0.9845 و0.9982(قريباً من ١). وهذه مؤشرات على مطابقة النموذج المستخدم للبيانات، وبالتالي توافق بعدية الاختبار للبيانات.

إجراءات تقدير ثبات الاختبارات

استخدمت الرزمة الاحصائية SPSS لتقدير ثبات الاختبار بالاعتماد على الطريقة التقليدية؛ حيث تم تقدير ثبات الفا كرونباخ للاختبارات أحادية البعد والاختبارات ثنائية البعد ولجميع حالات الارتباط بين الأبعاد، وتوزيع قدرة المفحوصين. ومن ثم تم تقدير ثبات التجزئة النصفية لجميع الحالات من الارتباط بين الأبعاد وحالات قدرة المفحوصن، بالاضافة إلى تقدير ثبات التجزئة النصفية مرة عندما توزعت الفقرات كما هو شائع بحيث تشكل الفقرات الفردية النصف الأول، و تشكل الفقرات الزوجية البعد الثاني. ومرة على أساس توزع الفقرات حسب أبعاد الاختبار. ومن ثم أدخلت معالم الفقرات التي تم تقديرها باستخدام برمجية NOHARM إلى Microsoft EXCEL وذلك لتقدير ثبات الاختبار بالاعتماد على العلاقة[١] (ص٤).

نتائج الدراسة ومناقشتها

١. النتائج المتعلقة بالسؤال الأول(هل يختلف تقدير معامل الثبات باستخدام طريقة ألفا كرونباخ باختلاف كل من: عدد أبعاد الاختبار(بعد واحد، بعدان)، وقوة الارتباط بين الأبعاد (0.0, 0.6, 1.0)، وتوزيع قدرة المفحوصين (طبيعي، متسق)؟) وهل يختلف باختلاف عدد الفقرات المشكلة للاختبار؟

للإجابة عن هذا السؤال تم تقدير ثبات الاختبار باستخدام طريقة ألفا كرونباخ للاختبار المشكل من ٦٠ فقرة، مرة عندما كان أحادي البعد، ومرة عندما تكون من بعدين كل منهما مكون من ٣٠ فقرة أحادية البعد، واختلف الارتباط بين الأبعاد حيث كان: مرة عديم الارتباط (r=0.0) ومرة ارتباط متوسط(r=0.6) ومرة ارتباط تام (r=1.0)، وباختلاف أيضا شكل توزيع معلمة قدرة المفحوصين (طبيعي، متسق).

ولتفحص إن كان تقدير الثبات باستخدام الفا كرونباخ يتأثر بطول الاختبار، تم اختيار ٣٠ فقرة من الفقرات الستين المشكلة للاختبار الأول لتشكل اختباراً جديداً مكوناً من ٣٠ فقرة، ومن ثم تم اختيار ١٠ فقرات من الفقرات الثلاثين المشكلة للاختبار الثاني لتشكل اختباراً ثالثاً مكوناً من ١٠ فقرات، ومن ثم تم تقدير

معامل الفا كرونباخ لهذه الاختبارات، ويظهر جدول٢ هذه التقديرات

جدول٢ نتائج تقدير معامل ثبات الفا كرونباخ

بعدية الاختبار	توزيع قدرة المفحوصين Θ	قوة الارتباط بين الأبعاد	الاختبار الاول (٦٠ فقرة)	الاختبار الثاني (٣٠ فقرة)	الاختبار الثالث (١٠فقرات)
احادي البعد	طبيعي		0.93	0.87	0.68
	متسق		0.98	0.97	0.91
ثنائي البعد	طبيعي	0.0	0.84	0.76	0.46
		0.6	0.90	0.82	0.59
		1.0	0.93	0.87	0.63
	متسق	0.0	0.95	0.91	0.74
		0.6	0.95	0.92	0.74
		1.0	0.95	0.91	0.75

يلاحظ من النتائج في جدول٢ أن تقديرات معامل الثبات باستخدام ألفا كرونباخ على أساس أن الاختبار أحادي البعد كانت أعلى من التقديرات على أساس أن الاختبار ثنائي البعد لجميع الحالات من توزيع قدرة المفحوصين (طبيعي، متسق) وقوة الارتباط بين الأبعاد ولا تختلف التقديرات عندما توزعت قدرات المفحوصين بشكل طبيعي والارتباط بين البعدين تام (r=1.0). ربما يرجع ذلك إلى أن معالم الفقرات للنماذج أحادية البعد نفسها التي تدخل في تقديرات النظرية التقليدية، أما في النماذج ثنائية البعد فيوجد قيمة لكل معلمة على كل بعد. واتفقت هذه النتيجة مع دراسة كورتينا(Cortina, 1993) حيث توصلت إلى أن قيمة ثبات ألفا كرونباخ تقل بزيادة عدد الأبعاد.

ويبين جدول٢ أن قيمة الفا كرونباخ زادت بزيادة الارتباط بين الأبعاد عندما توزعت قدرات المفحوصين بشكل طبيعي، وهذا ليس غريبا لأن حساب الثبات في النظرية التقليدية يعتمد على حساب معامل الارتباط، فثبات ألفا كرونباخ يقوم على حساب الارتباطات بين علامات مجموعة الثبات على جميع فقرات الاختبار. ووافق ذلك ما توصلت إليه دراسة كورتينا(Cortina, 1993) . ولم تختلف عندما توزعت قدرات المفحوصين بشكل متسق. كما ويتضح من النتائج في جدول٢ أن قيمة ألفا كرونباخ عندما توزعت قدرات المفحوصين بشكل متسق أعلى منها عندما توزعت بشكل طبيعي، ولجميع الحالات من البعدية والارتباط بين الأبعاد وطول الاختبار، وقد يكون السبب في ذلك انه عندما توزعت قدرات المفحوصين بشكل طبيعي اقتربت قيمة التباين لاستجاباتهم على كل فقرة (j) من فقرات الاختبار من تباين استجاباتهم على الاختبار الكلي الذي يحتوي n من الفقرات($\sum_1^n S_j^2 \cong S_{Total}^2$) في معادلة حساب ثبات الفا كرونباخ التالية

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_1^n s_j^2}{s_{Total}^2}\right] \qquad (5)$$

وهذا قد يؤدي إلى انخفاض قيمة ألفا كرونباخ α. ووافقت هذه النتيجة ما توصل إليه بني عطا والشريفين (2012) في أن تقدير ثبات ألفا كرونباخ يختلف باختلاف شكل توزيع قدرة المفحوصين.

وفيما يتعلق بتأثير عدد الفقرات المشكلة للاختبار(طول الاختبار) على تقدير ألفا كرونباخ تبين أن قيمة ألفا كرونباخ تزداد بزيادة عدد الفقرات ولجميع الحالات من البعدية، وتوزيع قدرة المفحوصين، وقوة الارتباط بين الأبعاد، وقد يرجع ذلك إلى أن العلاقة بين عدد فقرات n الاختبار وتباين الاختبار s_{Total}^2 علاقة طردية (ينقص التباين بنقصان عدد الفقرات)، ومن المعادلة أعلاه واضح أن العلاقة بين معامل الثبات α وتباين الاختبار s_{Total}^2 ايضا علاقة طردية، وهذا يؤدي إلى أن العلاقة بين عدد فقرات الاختبار ومعامل ألفا كرونباخ علاقة طردية. واتفقت هذه النتيجة أيضا مع نتائج دراسة كورتينا (Cortina, 1993) .

٢. النتائج المتعلقة بالسؤال الثاني(هل يختلف تقدير معامل الثبات باستخدام طريقة التجزئة النصفية باختلاف كل من: عدد أبعاد الاختبار(بعد واحد، بعدان)، قوة الارتباط بين الأبعاد (0.0, 0.6, 1.0)، توزيع قدرة المفحوصين
(طبيعي، متسق)؟ وهل يختلف باختلاف عدد الفقرات المشكلة للاختبار؟)

جدول٣ نتائج تقدير معامل ثبات التجزئة النصفية

بعدية الاختبار	توزيع قدرة المفحوصين Θ	قوة الارتباط بين الأبعاد	الاختبار الأول (٦٠ فقرة)	الاختبار الثاني (٣٠ فقرة)	الاختبار الثالث (١٠فقرات)
أحادي البعد	طبيعي		0.93	0.87	0.71
	متسق		0.99	0.97	0.92
ثنائي البعد	طبيعي	0.0	-0.12	0.35	0.31
		0.6	0.67	0.71	0.55
		1.0	0.93	0.86	0.63
	متسق	0.0	-0.04	0.47	0.53
		0.6	-0.05	0.46	0.53
		1.0	-0.10	0.43	0.50

تشير النتائج في جدول٣ إلى أن تقديرات معامل الثبات باستخدام طريقة التجزئة النصفية على أساس أن الاختبار أحادي البعد كانت أعلى من التقديرات على أساس أن الاختبار ثنائي البعد لجميع الحالات من توزيع قدرة المفحوصين (طبيعي، متسق) وقوة الارتباط بين الأبعاد، واقتربت التقديرات أحادية البعد من التقديرات ثنائية البعد عندما توزعت قدرات المفحوصين بشكل طبيعي والارتباط بين البعدين تام (r=1.0).

وفيما يتعلق بتأثير قوة الارتباط بين الأبعاد في حالة الاختبارات ثنائية البعد فقد لوحظ أن قيمة معامل ثبات التجزئة النصفية تزداد بزيادة قوة الارتباط بين البعدين في حالة التوزيع الطبيعي لقدرات المفحوصين، ويعود ذلك إلى أن ثبات التجزئة النصفية هو ارتباط بين نصفي الاختبار، وفي هذه الحالة يحتوي كل نصف من نصفي الاختبارات على فقرات تنتمي لبعد واحد، وبالتالي كانت العلاقة بين قيمة الثبات ومعامل الارتباط طردية. ولم تختلف التقديرات في حالة التوزيع المتسق لقدرات المفحوصين ولجميع حالات عدد الفقرات المشكلة للاختبار(10, 30, 60).

أما بالنسبة لتأثير اختلاف توزيع قدرات المفحوصين (طبيعي، متسق) فقد تبين من جدول٣ أن ثبات التجزئة النصفية في حالة التوزيع الطبيعي والارتباط 0.0 بين البعدين كانت أعلى من الحالة المناظرة من الارتباط في حالة التوزيع المتسق، أما حالات الارتباط الأخرى (1.0, 0.6) فقد انعكست النتائج فكانت في حالة التوزيع المتسق لقدرات المفحوصين أعلى منها للتوزيع الطبيعي ولجميع حالات عدد الفقرات المشكلة للاختبار(10, 30, 60). وذلك بسبب زيادة معامل الثبات بزيادة الارتباط بين الأبعاد في حالة التوزيع الطبيعي وعدم زيادتها في حالة التوزيع المتسق.

وفيما يتعلق بتأثر تقدير ثبات التجزئة النصفية بعدد الفقرات المشكلة للاختبار(طول الاختبار)، فيبين جدول٣ أن قيمته تزداد بزيادة عدد الفقرات عندما كانت الاختبارات أحادية البعد، ووضح ذلك (السيد، ١٩٧٩، ص٥٣٩؛ علام، ٢٠١١)، حيث بينا أن في طريقة التجزئة النصفية يحسب ثبات الاختبار الكلي r_{xx} من ثبات أجزاء الاختبار r باستخدام معادلة سبيرمان– براون ($r_{xx}=\frac{nr}{1+(n-1)r}$)، مع ملاحظة أن الاختبار الاول (٦٠ فقرة) هو ضعفا الاختبار الثاني (٣٠ فقرة)، والاختبار الثاني هو ثلاثة أضعاف الاختبار الثالث (١٠ فقرات)، وبهذا الخصوص بيَن عوده (2005) أن هذه العلاقة منطقية لأن زيادة عدد

فقرات الاختبار يزيد في تمثيل الاختبار لمحتوى المادة الدراسية، أي زيادة في صدق المحتوى، وبالتالي زيادة في دقة قياس الاختبار للسمة المقيسة، وفي حالة واحدة من حالات ثنائية البعد وهي عندما توزعت قدرات المفحوصين بشكل طبيعي وقوة الارتباط بين الأبعاد تامة (r=1.0). وكانت على العكس تماما أي زاد تقدير الثبات باستخدام التجزئة النصفية بنقصان عدد الفقرات عندما توزعت قدرات المفحوصين بشكل متسق ولجميع الحالات من قوة الارتباط بين الأبعاد.

٣. النتائج المتعلقة بالسؤال الثالث(هل يختلف تقدير معامل الثبات بالاعتماد على نظرية استجابة الفقرة باختلاف كل من: عدد ابعاد الاختبار(بعد واحد، بعدان)، قوة الأرتباط بين الأبعاد (0.0, 0.6, 1.0)، توزيع قدرة المفحوصين (طبيعي، متسق)؟ وهل يختلف باختلاف عدد الفقرات المشكلة للاختبار؟)

الجدول٤ تقديرات معامل الثبات بالاعتماد على نظرية استجابة الفقرة

بعدية الاختبار	توزيع قدرة المفحوصين (Θ)	قوة الارتباط بين الأبعاد	الاختبار الأول (٦٠ فقرة)	الاختبار الثاني (٣٠ فقرة)	الاختبار الثالث (١٠فقرات)
أحادي البعد	طبيعي		0.98	0.96	0.91
	متسق		1.00	0.99	0.99
ثنائي البعد	طبيعي	0.0	0.95	0.94	0.69
		0.6	0.96	0.96	0.76
		1.0	0.98	0.99	0.85
	متسق	0.0	0.99	0.99	0.96
		0.6	0.99	0.99	0.96
		1.0	0.99	0.99	0.95

يلاحظ من النتائج الموضحة في جدول٤ أن تقديرات معامل الثبات باستخدام نظرية استجابة الفقرة على أساس أن الاختبار أحادي البعد كانت قريبة من التقديرات على أساس أن الاختبار ثنائي البعد لجميع الحالات من توزيع قدرة المفحوصين (طبيعي، متسق) وقوة الارتباط بين الأبعاد، باستثناء عندما كانت عدد الفقرات المشكلة للاختبار ١٠ فقرات فقد اختلفت التقديرات باختلاف بعدية الاختبار، ويرجع السبب في ذلك إلى أن الثبات حسب نظرية استجابة الفقرة في هذه الدراسة تم حسابه بالاعتماد على دالة معلومات الاختبار، وقد بين ابوشندي (2008) أن دالة معلومات الاختبار لا تختلف باختلاف بعدية الاختبار، وبالتالي ثبات الاختبار لا يختلف باختلاف بعدية الاختبار.

وفيما يتعلق بتأثير قوة الارتباط بين الأبعاد في حالة الاختبارات ثنائية البعد فقد لوحظ أن قيمة معامل الثبات تزداد بزيادة قوة الارتباط بين البعدين في حالة التوزيع الطبيعي لقدرات المفحوصين، ولم تختلف التقديرات باختلاف الارتباط بين الأبعاد في حالة التوزيع المتسق لقدرات المفحوصين ولجميع حالات عدد الفقرات المشكلة للاختبار (10, 30, 60).

أما بالنسبة لتأثير اختلاف توزيع قدرات المفحوصين (طبيعي، متسق) فقد تبين من جدول٤ أن معامل الثبات المعتمد على نظرية استجابة الفقرة في حالة التوزيع الطبيعي كان أقل منه في حالة التوزيع المتسق، ولجميع الحالات من البعدية والارتباط بين الأبعاد وعدد الفقرات المشكلة للاختبار. وأعلى فرق كان 0.263 في حالة ثنائية البعد والارتباط 0.0 وعدد الفقرات ١٠، وتفسير ذلك يعود إلى أن العلاقة بين قيمة الثبات المحسوب بالاعتماد على نظرية استجابة الفقرة في هذه الدراسة ودالة معلومات الاختبار هي علاقة طردية، ودالة معلومات الاختبار تشير إلى اقصى معلومات يوفرها الاختبار التي تم حسابها بالاعتماد على أقصى معلومات للفقرة ($I_{max}(\Theta)=\frac{1}{4}D^2a_i^2$)، أي أن اقصى دالة معلومات تعتمد فقط على معلمة تمييز الفقرة a_i، وقد دلل ابوشندي (2010) أن تقديرات معلمة تمييز الفقرات عندما تتوزع قدرات المفحوصين بشكل متسق أقل منها عندما توزعت القدرات بشكل طبيعي.

وفيما يتعلق بتأثر تقدير الثبات باستخدام نظرية استجابة الفقرة بعدد الفقرات المشكلة للاختبار(طول الاختبار)، فيبين جدول٤ أن قيمته تقل بنقصان عدد الفقرات قي جميع الحالات، واختلفت هذه النتيجة مع دراسة هومان ومصباح (Homan & Mesbah, 2002) حيث بينا أن تقديرات نظرية استجابة الفقرة للثبات لا تتغير بتغير طول الاختبار.

٤. النتائج المتعلقة بالسؤال الرابع(هل يؤثر ترتيب الفقرات في جزأي الاختبار على تقدير الثبات باستخدام: كرونباخ الفا، التجزئة النصفية، نظرية استجابة الفقرة؟)

للإجابة عن هذا السؤال أجريت التقديرات لمعامل الثبات بالطرائق الثلاث(الفا كرونباخ، التجزئة النصفية، ونظرية استجابة الفقرة) وللحالات المختلفة من بعدية الاحتبار(أحادي البعد، ثنائي البعد)، ولاختلاف توزيع معلمة المفحوصين (طبيعي، متسق)، وباختلاف قوة الارتباط بين الأبعاد، واستخدمت طريقتان لترتيب الفقرات: الطريقة الأولى ترتيب الفقرات فردي وزوجي، والطريقة الثانية ترتيب الفقرات

حسب بعد الاختبار. أظهرت النتائج عدم اختلاف تقديرات الثبات باختلاف طريقة ترتيب الفقرات باستخدام الطريقتين: الفا كرونباخ، ونظرية استجابة الفقرة، واختلفت التقديرات باستخدام طريقة التجزئة النصفية، ويظهر جدول٥ هذه الاختلافات.

جدول٥ تقدير الثبات باستخدام طريقة التجزئة النصفية وباختلاف ترتيب الفقرات

ترتيب الفقرات	البعدية	توزيع القدرة	الارتباط	الاختبار الأول (٦٠ فقرة)	الاختبار الثاني (٣٠ فقرة)	الاختبار الثالث (١٠فقرات)
فردي زوجي	أحادي	طبيعي		0.93	0.88	0.67
		متسق		0.98	0.97	0.90
	ثنائي	طبيعي	٠	0.87	0.79	0.47
			0.6	0.90	0.84	0.60
			١	0.93	0.88	0.63
		متسق	٠	0.97	0.95	0.79
			0.6	0.97	0.95	0.80
			١	0.97	0.95	0.81
حسب البعد	أحادي	طبيعي		0.93	0.87	0.71
		متسق		0.99	0.97	0.92
	ثنائي	طبيعي	٠	-0.12	0.35	0.31
			0.6	0.67	0.71	0.55
			١	0.93	0.86	0.63
		متسق	٠	-0.04	0.47	0.53
			0.6	-0.05	0.46	0.53
			١	-0.10	0.43	0.50

يظهر من النتائج المبينة في جدول٥ أنه في حالة أحادية البعد لم تكن الفروق بين التقديرات كبيرة سواء كان توزيع قدرة المفحوصين طبيعي أم متسق(حيث وصل أكبر فرق إلى 0.04 في حالة ١٠ فقرات وتوزيع القدرة طبيعي).

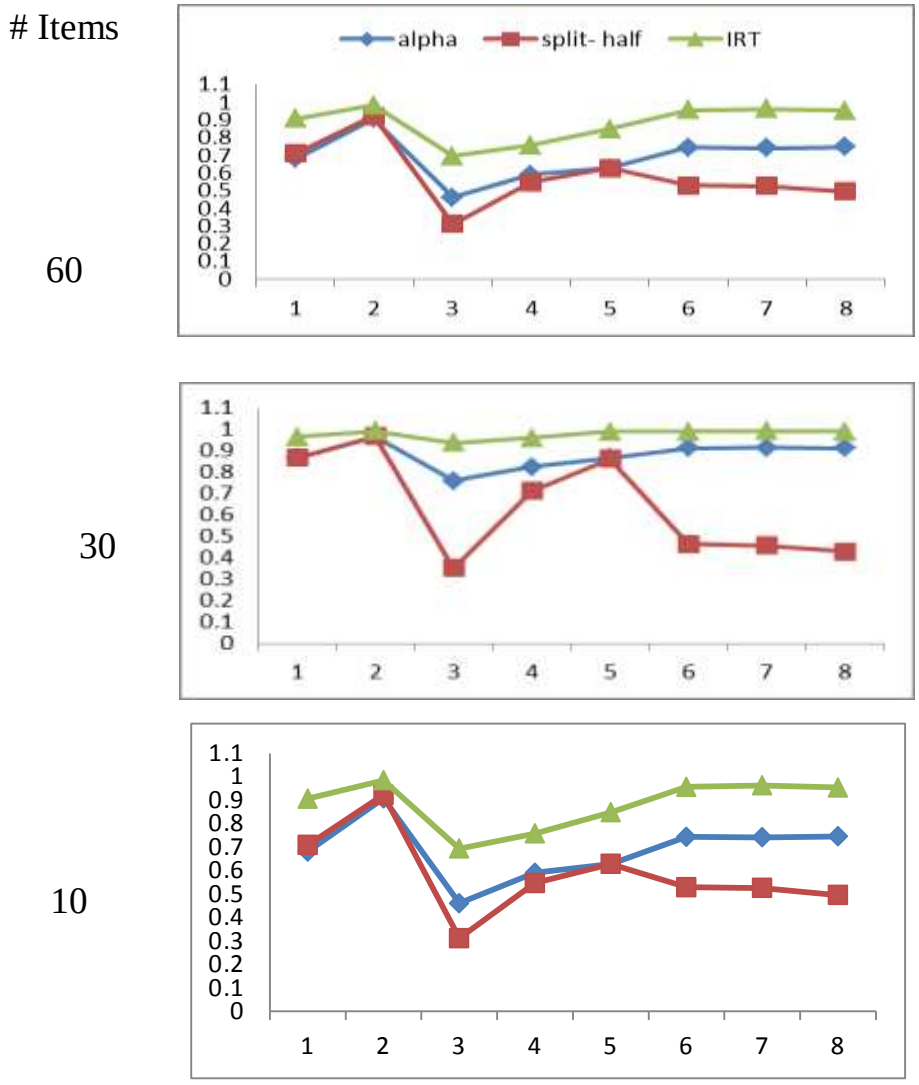
أما في حالة الاختبارات ثنائية البعد فتبين أن تأثير ترتيب الفقرات واضح على تقدير ثبات التجزئة النصفية(انحصرت الفروق بين 0.001 - و1.065)، ويظهر من جدول٥ أن معامل الثبات عندما توزعت الفقرات فردي زوجي على نصفي الاختبار كانت أعلى من الحالات المقابلة لها عندما توزعت الفقرات

حسب البعد. كما يتبين أن الفروق في التقديرات عندما توزعت قدرات المفحوصين بشكل طبيعي أقل منها عندما توزعت بشكل متسق. كما لوحظ أن الفروق بين التقديرات ازدادت بزيادة عدد الفقرات المشكلة للاختبار.

النتائج المتعلقة بالسؤال الخامس(هل يختلف تقدير ثبات الاختبار باختلاف طريقة التقدير: ألفا كرونباخ، التجزئة النصفية، نظرية استجابة الفقرة؟)

استخدمت التمثيلات البيانية لتسهيل عرض نتائج هذا السؤال، حيث يوضح شكل ١ تقديرات معامل الثبات للاختبارات الثمانية، وباستخدام طرائق التقدير الثلاث

شكل ١ تقديرات معامل الثبات باستخدام الفا كرونباخ والتجزئة النصفية ونظرية استجابة الفقرة



توضح من شكل أن تقديرات معامل الثبات بالاعتماد على نظرية استجابة الفقرة كانت أعلى منها بالاعتماد على النظرية الكلاسيكية ولجميع الحالات من بعدية الاختبار وقوة الارتباط بين الأبعاد وتوزيع

قدرة المفحوصين وعدد الفقرات، وهذا يعود إلى ما تم الإشارة إليه سابقا بأن الثبات بالاعتماد على نظرية استجابة الفقرة يعتمد على أقصى معلومات يوفرها الاختبار. كما تبين ان التقديرات باستخدام الفا كرونباخ كانت أكبر من أو يساوي التقديرات بالاعتماد على التجزئة النصفية. واتفقت هذه النتيجة مع دراسات (إبراهيم والعزابي، ٢٠١١؛ بني عطا والشريفين، ٢٠١٢؛ Hamon & Mesbah, 2000) في أن تقدير الثبات باستخدام نظرية استجابة الفقرة أعلى من ثبات الفا كرونباخ، ولم تتفق مع دراسة جمحاوي (2000) التي توصلت إلى عدم اختلاف التقديرات في حالة النموذج ثنائي المعلمة.

الاستنتاجات والتوصيات

خلصت الدراسة إلى أن تقديرات معامل الثبات بالاعتماد على النظرية الكلاسيكية (الفا كرونباخ، والتجزئة النصفية) عندما كانت الاختبارات أحادية البعد كانت أعلى منها عندما كانت الاختبارات ثنائية البعد، وكذلك عندما قدر الثبات بالاعتماد على نظرية اسجابة الفقرة وتشكل الاختبار من ١٠ فقرات. ولم تتأثر التقديرات لبقية الحالات بتغير بعدية الاختبار عندما استخدمت نظرية استجابة الفقرة.

وفيما يتعلق بتأثير الارتباط بين الأبعاد وفي حالة التوزيع الطبيعي لقدرة المفحوصين فقد اتفقت التقديرات باستخدام جميع الطرائق في أن تقدير الثبات يزداد بزيادة الارتباط بين الأبعاد ولم تختلف التقديرات عند التوزيع المتسق باستخدام الفا كرونباخ واستجابة الفقرة، وقل بزيادة الارتباط عند استخدام التجزئة النصفية.

وكانت التقديرات باستخدام الطرائق الثلاث وعلى أساس أن قدرات المفحوصين توزعت بشكل طبيعي أقل منها على أساس التوزيع المتسق، إلا في حالة التجزئة النصفية والارتباط 0.0 بين الأبعاد. وزادت قيمة الثبات بزيادة عدد فقرات الاختبار عند استخدام ألفا كرونباخ ونظرية استجابة الفقرة. وفي طريقة التجزئة النصفية اتفقت النتائج مع ما سبق في حالة التوزيع الطبيعي والارتباط 1.0 وانعكست في بقية الحالات. ولم تختلف التقديرات باختلاف ترتيب الفقرات عند استخدام الطريقتين ألفا كرونباخ واستجابة الفقرة. ولأن معامل الثبات تغير بتغير: بعدية الاختبار، وقوة الارتباط بين الأبعاد، وتوزيع قدرة المفحوصين، وعدد فقرات الاختبار يوصي الباحث بضرورة تحديد البعدية السليمة للاختبار، وقوة الارتباط بين الأبعاد، وتوزيع قدرة المفحوصين قبل تقدير ثبات الاختبار. كما يوصي بإجراء دراسات تتناول مستويات أخرى للارتباط بين الأبعاد، وتوزيعات مختلفة لقدرة المفحوصين، وتوزيعات لمعالم الفقرات. ولان

الباحث استخدم النموذج التعويضي يوصي باجراء دراسات تتناول النماذج غير التعويضية. ويوصي الباحث بزيادة عدد فقرات الاختبار لزيادة ثبات الاختبار.

المراجع :

- إبراهيم، ميكائيل و العزابي، أسامة. (٢٠١١). ثبات المقياس وصدقه بين الطرق التلقيدية ونموذج راش: البديل
- الإحصائي في تطوير المقياس النفسي. مجلة الدراسات التربوية والنفسية، جامعة السلطان قابوس، ٥(١)، ٢٢-٤١ .
- ابوشندي، يوسف. (٢٠١٠). اثر استخدام معايرات أحادية البعد لبيانات متعددة الأبعاد على التقديرات المختلفة
- لنظرية استجابة الفقرة: دراسة محاكاة. مجلة جامعة دمشق(مقبول للنشر).
- ابوشندي، يوسف.(٢٠٠٨). تأثير تعدد الأبعاد للاختبار والعلاقة بينها على تقديرات معالم فقراته:دراسة محاكاة.
- رسالة دكتوراه، جامعة اليرموك، اربد، الأردن.
- بني عطا، زايد والشريفين، نضال.(٢٠١٢).أثر شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. المجلة
- الاردنية في العلوم التربوية.٨(٢)، ١٥١-١٦٦.
- جمحاوي، ايناس.(٢٠٠٠)مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة
- العقلية .رسالة ماجستير غير منشورة، جامعة اليرموك، الأردن.
- الزبون، حابس. (٢٠١٣).أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها بإستخدام

- النظرية الحديثة في القياس. مجلة جامعة النجاح للأبحاث (العلوم الانسانية)، ٢٧(٦)، ١٣٢٣- ١٣٤٤.
- العمري، حسان. (٢٠١١). أثر ترتيب الفقرات في المقاييس النفسية على التجانس الداخلي للمقياس وخصائص
- فقراته. مجلة جامعة دمشق، ٢٧(٣+٤)، ١١١-١٢٩.
- عوده، احمد.(٢٠٠٥). القياس والتقويم في العملية التدريسية ط٣. اربد: دار الامل للنشر والتوزيع.
- السيد، فؤاد.(١٩٧٩). علم النفس الاحصائي وقياس العقل البشري ط٣. القاهرة: دار الفكر العربي.
- النعيمي، عزالدين.(٢٠٠٦). أثر انتهاك افتراض الاستقلال الموضعي على التقديرات المختلفة لنظرية استجابة الفقرة.
- رسالة دكتوراه، جامعة اليرموك، اربد، الأردن
- Allen, Mary J. and yen, Wendy M. (1979). Introduction to Measurement Theory.Wadsworth inc.
- Cortina,J.(1993).What is the Coefficient Alpha? An Examination of Theory and Applications.
- Journal of Applied Psychology. 78(1), 98-104.
- Davey, T., Nering, M., & Thompson, T. (1997). Realistic simulation of item response data
- (ERIC Document Reproduction Service No. ED 414297).
- Embretson, S., & Reise, S .(2000). Item response theory for psychologists. New Jersey:
- Lawrence Erlbaum Associates,Inc.
- Hamon,A& Mesbah, M.(2002). Questionnaire Reliability under The Rasch Model. Statistical

- Methods for Quality of Life Studies, Klwwer Academic Puplisher155-168.
- Hattie, J. (1984). An empirical study of various indices for determining unidimensionality.
- Multivariate Behavioral Research, 19, 49-78.
- Lord, F.M. (1980). Applications of item response theory to practical testing problems.
- Hillsdale, NJ: Lawrence Erlbaum Associates.
- Muraki, E.(2000).RESGEN: Item response generator.princeton, NJ: Educational Testing Service.
- Thissen, D. (2000). Reliability and measurement precision. In H. Wainer (Ed), Computerized
- adaptive testing: A primer (2nd ed., pp. 159-183). Mahwah, NJ: Lawrence Erlbaum Associates.