An-Najah University Journal for Research - B (Humanities)

Volume 29 | Issue 8 Article 3

2015

The Effect of Test Length and Ability Distribution Form on the Estimation of A person's Ability According To Rasch Model in Item Response Theory

Arwa Alhawari
arwaesa@yahoo.com

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/anujr_b

Recommended Citation
Alhawari, Arwa (2015) "The Effect of Test Length and Ability Distribution Form on the Estimation of A person's Ability According To Rasch Model in Item Response Theory," *An-Najah University Journal for Research - B (Humanities)*: Vol. 29 : Iss. 8 , Article 3.
Available at: https://digitalcommons.aaru.edu.jo/anujr_b/vol29/iss8/3

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in An-Najah University Journal for Research - B (Humanities) by an authorized editor. The journal is hosted on Digital Commons, an Elsevier platform. For more information, please contact rakan@aaru.edu.jo, marah@aaru.edu.jo, u.murad@aaru.edu.jo.

مجلة جامعة النجاح للأبحاث (العلوم الانسانية) المجلد 29 (8)، 2015

أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وفق نموذج راش في نظرية استجابة الفقرة

The Effect of Test Length and Ability Distribution Form on the Estimation of A person's Ability According To Rasch Model in Item Response Theory

اروى الحواري

Arwa Alhawari

وزارة التربية والتعليم، اربد، الأردن.

بريد الكتروني: arwaesa@yahoo.com

تاريخ التسليم: (2014/5/8)، تاريخ القبول: (2014/11/19)

ملخص

هدفت الدراسة الحالية للكشف عن أثر طول الاختبار وشكل توزيع القدرة في تقديرات القدرة للأفراد، وتقديرات معالم الصعوبة للفقرات، واقتران المعلومات للفقرات، وفق نموذج راش في نظرية استجابة الفقرة. ومن ضمن إجراءات الدراسة تم توليد البيانات، ومجموعة الفقرات، وقدرات الأفراد؛ إذ تم توليد بيانات بواقع (1000) مفحوص، بطول الاختبار (30,60) فقرة، وشكل توزيعات القدرة (ملتو التواءً سالباً، طبيعي، ملتو التواءً موجباً)، وفق النموذج الأحادي المعلمة. تم تحليل البيانات المولدة باستخدام البرمجيات (BILOG-MG)، (WINGEN) و(SPSS). أظهرت النتائج وجود فروق ذات دلالة إحصائية (α= 0.05) بين متوسطات الأخطاء المعيارية في تقديرات معالم قدرة الأفراد تعزى لشكل توزيع القدرة ولطول الاختبار، حيث كانت تقديرات قدرة الأفراد أكثر دقة لصالح شكل الالتواء السالب والموجب. ولطول الاختبار (30) فقرة. كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية في تقديرات معلم الصعوبة وكانت النتائج لصالح الاختبار بطول (30) فقرة؛ إذ زود بمعلومات أكثر. وتوصي الباحثة بإجراء المزيد من البحوث باعتماد النماذج المعلمية الأخرى، واستخدام بيانات حقيقية.

الكلمات المفتاحية: اقتران المعلومات، معالم الصعوبة، التوزيع الطبيعي، والالتواء الموجب والسالب، تقدير القدرة.

An-Najah University Journal for Research - B (Humanities), Vol. 29 [2015], Iss. 8, Art. 3

Abstract

This study investigated the Effect of Test Length and Ability distribution form on The Estimation of a person's Ability, item difficulty, and the information function of test and its items, According to Rasch Model in Item Response Theory (IRT). Among the procedures of the study was to generate data and items were also generating the examinee's ability, the generated data rate (1000) were examined, for all variables; test length (30, 60) item, and Ability distribution (normal, positive and negative skewed) according to Item Response Theory Models (one-parameter). According to the analysis of data depending on the softwares (BILOG-MG), (WINGEN) and (SPSS), The results showed that there were statistically significant differences at (α=0.05) among the standard error means of item in the estimation of difficulty parameters, such estimations in a person's ability were more accurate in the positive and negative skewed. The test consisted from 30 items. Also The results showed there were statistically significant differences (α=0.05) among the standard error means of item in the estimation of difficulty parameters for the test consisted from 30 items, which provided the most information. The researcher recommends further research depending on other parametric models, also by using real data.

Keywords: Information Function, difficulty parameters, normal distribution, positive and negative skewed, The Ability Estimation.

المقدمة والإطار النظري

تلعب الاختبارات بشكل عام، واختبارات التحصيل بشكل خاص، دورا هاما في العملية التعليمية؛ فهي تؤدي إلى تحسين عملية التعليم ومخرجاتها من خلال العديد من القرارات التي يمكن اتخاذها في ضوء نتائج الاختبارات التي تجرى عبر المراحل المختلفة لعملية التعليم. وتعتمد صحة القرارات المتخذة على نوع ودقة المعلومات والنتائج والتغذية الراجعة التي تقدمها الاختبارات، فهي تقدم نتائج ومعلومات موضوعية واقعية يمكن اعتمادها في صنع القرارات ورسم السياسات التربوية، إلا أنّ كثيراً منها أصبحت تقتصر على قياس قدرة الأفراد على استرجاع الحقائق والمعلومات المتفرقة المتعلقة بالمحتوى الدراسي وبعض التطبيقات البسيطة دون الاستناد إلى نظرية أو نموذج متطور في القياس التربوي .(Sawalmeh, 1994; Sawalmeh & Kawasmeh, 2000)

https://digitalcommons.aaru.edu.jo/anujr_b/vol29/iss8/3

ويزخر علم القياس النفسي والتربوي بالعديد من النماذج الإحصائية السيكومترية التي يستند إليها تصميم الاختبارات السيكولوجية والتربوية، وبناؤها وتفسير نتائجها. إلا أنه يوجد بعض المتغيرات التي يمكن أن تؤثر في دقة تقديرات قدرة الأفراد وصدق قياسها، بالإضافة لتأثر كمية المعلومات المستمدة من الفقرة والتي تتغير بحسب مستوى القدرة، إذ تمثل دالة معلومات الاختبار الحد الأقصى من المعلومات التي يمكن الحصول عليها من إجابات المفحوصين على فقرات الاختبار بغض النظر عن أسلوب التصحيح وإعطاء الدرجات، وبالتالي فإنّ طول الاختبار يؤثر في كمية المعلومات المستمدة من الاختبار. وإجرائيا فإنّ دالة معلومات الاختبار هي محصلة دوال معلومات الفقرات، ومنها يمكن استخلاص إسهام الفقرة في القياس الكلي وذلك عند موازنة معلومات الفقرات المختلفة، وتقسم نماذج القياس النفسي والتربوي في قسمين هما (Lord, 1980):

نظرية القياس الكلاسيكية (Classical Test Theory): ظهرت نظرية القياس الكلاسيكية في بدايات القرن العشرين، وتقوم فكرتها على مفهوم أخطاء القياس (Measurement Errors)، فلكي نقيس خاصية معينة ينبغي تقدير المصادر المتعددة للخطأ والتي تؤثر في تباين درجات الاختبارات، ولها العديد من النماذج والافتراضات والاستخدامات على مستوى الاختبار والفقرة. إلا أنها لم تخل من جوانب القصور في تحليل نتائج الاختبارات (Allam, 2005).

نظرية القياس الحديثة (Measurement Modern Theory): تعرف النظرية الحديثة في القياس باسم السمات الكامنة كما تسمى بنظرية الاستجابة للفقرة (Item Response Theory) لاهتمامها بالربط بين استجابة المفحوص لفقرة اختبار ذات خصائص معينة وقدرته، كما أنها تركز على تحديد مواقع الأفراد على المقاييس النفسية والتربوية، وتتمثل في مجموعة من النماذج الرياضية والإحصائية التي تستخدم في تحليل الفقرات والمقاييس. إذ تعد نظرية الاستجابة للفقرة نموذجا رياضيا يركز على مجموعة من الافتراضات للعلاقة بين الأداء الملاحظ للمفحوص على الاختبار والسمات الكامنة (Hambelton & Swaminathan, 1985)**.**

وتنطلق هذه النظرية من مسلمتين؛ الأولى تفترض التنبؤ بأداء الأفراد أو يمكن تفسير أدائهم على فقرة اختبارية أو في اختبار نفسي أو تربوي في ضوء خاصية أو مجموعة خصائص مميزة تسمى السمات أو القدرات، أي أنها تفترض وجود سمة أو قدرة تحدد استجابة الفرد للفقرات، وهذه السمة يشترك فيها جميع الأفراد، والثانية تصف العلاقة بين أداء الفرد على فقرة اختبارية ومجموعة السمات أو القدرات التي تكمن خلف الأداء باستخدام دالة طردية تزايدية (Alshafei, 2008).

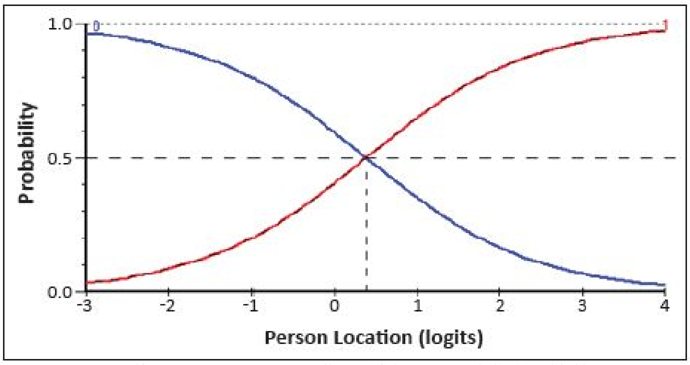


FIGURE 1: Item characteristic curve and complement: probability of 0/1 responses.

شكل (1): المنحنى المميز للفقرة.

يوضح الشكل (1) المنحنى المميز للفقرة عندما يكون هناك سمة واحدة فقط كامنة خلف أداء الفرد على الاختبار، كما يبين الشكل توزيعي القدرة لمجموعتين من الأفراد على نفس الفقرة، حيث تتمحور الإجابة الصحيحة في الدالة التزايدية الاطرادية. وتقوم هذه النظرية على مجموعة من الافتراضات، كما ينبثق عنها مجموعة من النماذج تسمى نماذج السمات الكامنة، ويعبر عن كل نموذج بدالة رياضية تحدد علاقة أداء الفرد على فقرة من فقرات المقياس أو الاختبار بقدرته التي تكمن وراء هذا الأداء وتفسره (Baker, 2001).

افتراضات نظرية السمات الكامنة الأحادية البعد Assumptions of Latent Trait Theory of Unidimesionality: ترتكز نماذج السمات الكامنة الأحادية البعد إلى مجموعة من الافتراضات يجب توافرها في البيانات المستمدة من الاختبار، ومجموعة الافتراضات هي: أحادية البعد حيث يقيس الاختبار سمة واحدة فقط، والاستقلال الموضعي وهو استقلال أداء الفرد على فقرة الاختبار عن أدائه على فقرة أخرى من الاختبار نفسه، منحنى خصائص الفقرة ويمثل مخطط بياني لاحتمالية الإجابة الصحيحة على الفقرة، كما تفترض معظم النماذج المستعملة أن عامل السرعة لا يؤدي دوراً في الإجابة عن الفقرة وأن إخفاق الأفراد في الإجابة عن فقرات الاختبار يرجع إلى انخفاض قدراتهم وليس لتأثير عامل السرعة في إجاباتهم، واللاتباين ويعني أن معالم الفقرة لا تعتمد على التوزيع الإحصائي للسمة المراد قياسها، وأن المعالم التي تصف أداء الأفراد لا تعتمد على فقرات الاختبار (Anastasi & Urbina, 1997; Crocker & algina, 1986; Hambelton & Jones, 1993; Lord, 1980) وفيما يلي توضيح لهذه الافتراضات:

1. أحادية البعد (Unidimesionality)

تفترض نظرية استجابة الفقرة وجود سمة أو قدرة واحدة تفسر أداء المفحوصين على الاختبار، بمعنى أن جميع فقرات الاختبار تقيس بعدا واحدا أي يمكن تفسير إجابة الفرد على أنها

تعود لهذه القدرة التي تقيسها فقرات الاختبار وعندها يشار إلى الاختبار بأنه أحادي البعد (Unidimesionality)، أما النماذج التي تفترض وجود أكثر من قدرة واحدة تفسر أداء الفرد على الاختبار وهي نماذج متعددة الأبعاد (Multidimensionality)، ويرى هامبلتون وسوامينثان (Hambelton & Swaminathan, 1985) أن افتراض أحادية البعد أكثر الافتراضات تعقيدا في نظرية الاستجابة للفقرة؛ لأنه في كثير من الأحيان لا يمكن تحقيقه بصورة صارمة، أو بمعنى أخر من الناحية العملية يصعب تحقيق هذا الافتراض لوجود عوامل مؤثره في أداء المفحوصين مثل قلق الاختبار ومستوى الدافعية والقدرة على الإجابة بسرعة.

ولتحقيق افتراض أحادية البعد يتطلب وجود عامل (بعد) سائد يؤثر في الأداء على الاختبار، وأجمع العلماء على فحص هذا الافتراض باستخدام التحليل العاملي. حيث يتم تحليل استجابات المفحوصين على فقرات الاختبار وملاحظة نتائج التحليل من خلال قيم الجذور الكامنة (Eigen Values) ونسب التباين المفسر لكل من العاملين الأول والثاني حيث يستدل على أحادية البعد أو وجود عامل سائد عندما يكون الفرق بين قيمة الجذر الكامن الأول وقيم الجذور الكامنة للعوامل الأخرى كبيرا نسبيا (Hattie, 1985; Linacre, 2008).

2. الاستقلال الموضعي (Local Independence)

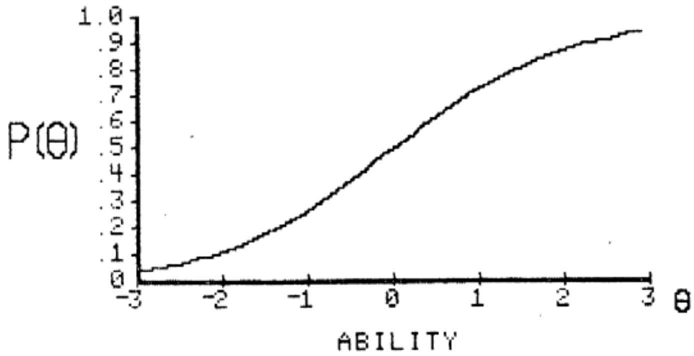
يكافئ افتراض الاستقلال الموضعي افتراض أحادية البعد عندما تكون قدرة المفحوص (θ) أحادية البعد، ويقصد به أن استجابات المفحوص للفقرات المختلفة في الاختبار مستقلة إحصائيا، وحتى يتحقق هذا الافتراض عمليا فإن استجابة المفحوص على فقرة ما لا تؤثر ايجابيا أو سلبيا على استجابته على فقرة أخرى، بمعنى أن المحتوى العلمي لفقرة يجب أن لا يشير إلى الإجابة الصحيحة لفقرة أخرى في نفس الاختبار، ولا بد هنا من الإشارة إلى مفهوم اللاتباين أو اللاتغير (Invariance) بالنسبة لمعالم الفقرات والأفراد (Warm, 1978):

- **اللاتباين لمعالم الأفراد:** تحرر القياس من توزيع العينة المستخدمة وهذا يعني ثبات تقدير كل من قدرة الفرد وصعوبة الفقرة واستقرارها بالرغم من اختلاف عينة الأفراد المستخدمة في تدريج المقياس طالما أنها عينة ملائمة.
- **اللاتباين لمعالم الفقرات:** تحرر القياس من مجموعة الفقرات المستخدمة وهذا يعني ثبات تقدير كل من قدرة الفرد وصعوبة الفقرة واستقرارها بالرغم من اختلاف مجموعة الفقرات المستخدمة في القياس، وطالما أن هذه المجموعات المختلفة من الفقرات تقع على ميزان تدريج واحد.

3. منحنى خصائص الفقرة (Item Characteristic Curve (ICC))

عبارة عن دالة رياضية تربط احتمالية الإجابة الصحيحة على الفقرة بقدرة المفحوص التي تم قياسها بمجموعة من الفقرات في الاختبار الذي تم بناؤه لتلك الغاية، ويعد منحنى خصائص الفقرة الصيغة الرياضية لاحتمالية الاستجابة للفقرة $Pi(\theta)$ حيث يعكس هذا المنحنى مستوى الأداء على مهام معينة تناظر متصلا مستقلا تعبر عن قدرة (سمة نفسية)، ويأخذ هذا المنحنى

الشكل اللوجستي وهو يمثل شكل الحرف (S). ويوضح الشكل (2) منحنى خصائص الفقرة في أبسط أشكاله وفيه يمثل الإحداثي السيني متصل القدرة (θ) المقاسة في الفقرة، ويمثل الإحداثي الصادي احتمالية الاستجابة الصحيحة للمفحوص P(θ) على تلك الفقرة، ويشير هذا المنحنى إلى أن احتمال إجابة الفرد على فقرة إجابة صحيحة يزداد بازدياد قدرة الفرد (Baker, 2001).



شكل (2): منحنى خصائص الفقرة.

ويظهر من منحنى خصائص الفقرة الشكل (2) أن احتمالية الاستجابة الصحيحة للفقرة تقترب من الصفر للمفحوصين ذوي القدرة المتدنية، وتزداد قيمة الاحتمالية بزيادة مستوى القدرة للمفحوص حتى تقترب من الواحد الصحيح للمفحوصين ذوي أعلى مستويات القدرة.

4. السرعة في الإجابة (Speededness)

تفترض نماذج الاستجابة للفقرة، أن عامل السرعة لا يلعب دورا في الإجابة على فقرات الاختبار، بمعنى أن إخفاق الأفراد في إجابة فقرات الاختبار يرجع إلى انخفاض قدرتهم، وليس إلى تأثير عامل السرعة على إجاباتهم (Allam, 2005).

نماذج استجابة الفقرة أحادية البعد: تستند هذه النماذج على عدد من الفروض التي يجب أن تتوافر في البيانات المستمدة من الاختبارات، مستفيدة من ثلاث معالم: هي تمييز الفقرة وصعوبة الفقرة والخط التقاربي الأسفل (Lower- Asymptote) أو التخمين وهو مناظر لاحتمالية الاستجابة الصحيحة التي قد تحدث بالصدفة. وتعددت نماذج السمات الكامنة وتباينت من حيث أشكال المنحنيات المميزة تبعاً لاختلاف عدد معالم الفقرات وهي :(Hulin, Drasgow, & Parsons, 1983)

1. النموذج أحادي المعلمة (نموذج راش) Rasch Model:

ويسمى بالنموذج الترجيح اللوغاريتمي أحادي المعلمة IPL وذلك لأنه يستخدم الدالة الآسية في التنبؤ بالاحتمالات، وتشمل على (معلمة) واحدة فقط، وهي معلمة صعوبة الفقرات

Item Difficulty Parameter لتمثيل الفروق الفردية بين الفقرات. ويعتبر من أبسط 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 متماثلة في قدرتها التمييزية وأن معلمة التخمين تساوي صفراً. ويتم التعبير عن احتمالية الاستجابة الصحيحة للفقرة $P(\theta)$ بدلالة معلمة صعوبة الفقرة (b) والتي تأخذ قيما متغيرة، وقد عبر بيكر (Baker, 2001) عن النموذج الأحادي المعلمة وفق المعادلة الآتية:

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

حيث أن:

θ : قدرة المفحوص، $P_i(\theta)$: احتمال أن المفحوص الذي يمتلك القدرة θ أجاب على الفقرة i إجابة صحيحة، b_i: معلمة صعوبة الفقرة.a_i: ميل دالة الفقرة عند النقطة b، وهي معلمة تمييز الفقرة وهنا لها جميعها قدرة تمييزية واحدة (1).

2. النموذج ثنائي المعلمة (نموذج بيرنبوم) Birnbaum Model

وهذا النموذج يسمح للفقرات بأن تختلف في كل من معلمتي الصعوبة والتمييز، لذلك تضمنت الصيغة الرياضية لهذا النموذج معلمة تمييز الفقرة وأصبح يشتمل هذا النموذج على معلمتي (الصعوبة والتمييز) يمثلان خصائص الفقرة.

3. النموذج ثلاثي المعلمة (نموذج لورد) Lord Model

يسمى النموذج اللوغارتمي ثلاثي المعلمة (3PL)Three – Parameter Logistic Model) إذ أضاف معلمة ثالثة في احتمال توصل الأفراد للإجابة الصحيحة وهي معلمة التخمين إضافة إلى معلمتي (الصعوبة والتمييز)، وبهذا يجب أن يؤخذ ذلك بعين الاعتبار عند مطابقة البيانات المستمدة من الاختبار لهذا النموذج.

دالة المعلومات (Information Function): يعد مفهوم دالة المعلومات من المفاهيم الأساسية في نظرية الاستجابة للفقرة، فهي دالة رياضية تمثل العلاقة بين قدرة الفرد والمعلومات المقدمة لفقرات الاختبار، فهي تعبر عن كمية المعلومات المتمثلة في تمييز الفقرة بين مستويات القدرة للأفراد، وذلك بتحديد أقصى ارتفاع لمنحنى دالة المعلومات، والذي يمثل مقدار المعلومات التي تقدمها الفقرة أو الاختبار ككل، عند تقدير قدرات المفحوصين، ويمكن من خلالها تحديد الخطأ المعياري في التقدير. ويندرج الحديث عن دالة المعلومات في محورين هما:

دالة معلومات الفقرة (Item Information Function): يرى هامبلتون وسومنثيان (Hambelton & Swaminathan, 1985) أن لكل فقرة من فقرات الاختبار منحنى

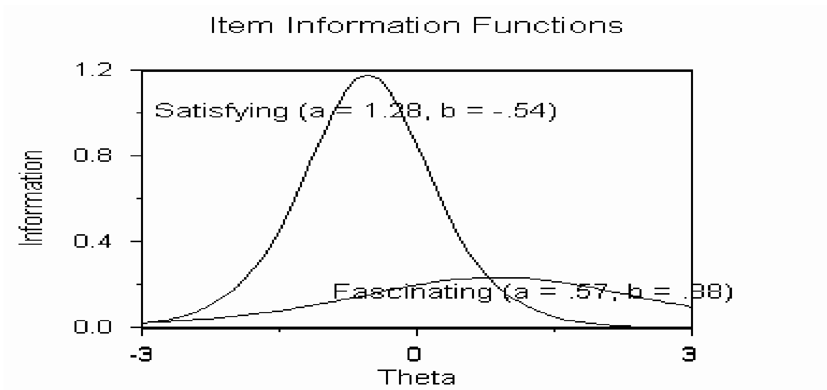
Published by Arab Journals Platform, 2015

معلومات الفقرة، وهو عبارة عن اقتران يبين مدى مساهمة الفقرة في تحديد القدرة. وبشكل عام فإن الفقرات ذات التمييز العالي تساهم بقوة أكبر في تأكيد دقة القياس من تلك ذات التمييز المتدني، حيث تعطي الفقرة أفضل مساهمة لها في تأكيد دقة القياس حول قيمة صعوبتها (b) على متصل القدرة، أي أن دالة المعلومات تعتمد على ميل اقتران الاستجابة للفقرة والتباين عند كل مستوى للقدرة (θ)، فكلما زاد الميل فإن التباين يقل وتزداد المعلومات التي توفرها الفقرة، فإذا ما كان المنحنى مزاحا باتجاه اليمين فهذا يعني أن الفقرة صعبة، وإذا كان ارتفاعه كبيرا فهذا يعني أن تمييز الفقرة عال والعكس صحيح. وتعطى معادلة منحنى معلومات الفقرة في حالة النموذج الأحادي بالعلاقة الآتية:

$$Ii(\theta) = Pi(\theta)\, Qi(\theta)$$

حيث: Ii(θ): دالة المعلومات للفقرة (i)، Pi(θ): احتمالية الاستجابة الصحيحة على الفقرة (i)، **Qi(θ)**: احتمالية الاستجابة الخاطئة على الفقرة (i).

وعادة ما يأخذ المنحنى الشكل الجرسي كما هو موضح في الشكل الأتي:



شكل (3): منحنى دالة المعلومات لفقرتين مختلفتين.

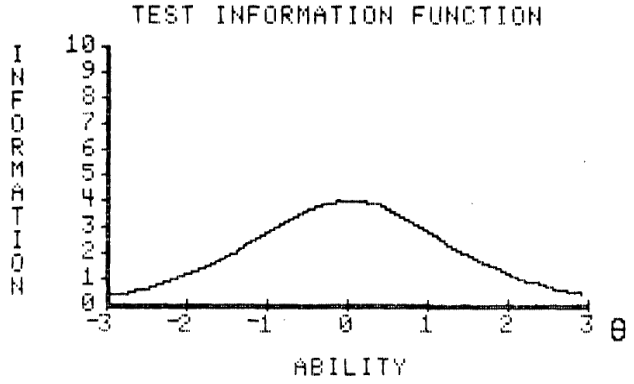
ويشير الشكل (3) إلى منحنى دالة المعلومات لفقرتين مختلفتين، حيث أن الفقرة ذات التمييز العالي قدمت كمية أكبر من المعلومات وان الفقرة ذات التمييز المنخفض لم تعط معلومات كثيرة.

دالة معلومات الاختبار(Test Information Function): تعَرف بأنها المجموع الحسابي لدوال معلومات الفقرات المكونة للاختبار، لذلك فإن دراسة دالة معلومات الفقرة والمتغيرات المؤثرة فيها تعطي فرصة للحصول على اختبار ذي دالة مرغوبة، وتحسب دالة معلومات الاختبار بجمع معلومات الفقرات وفق المعادلة الآتية: $I(\theta) = \sum Ii(\theta)$

حيث : I(θ) : دالة معلومات الاختبار، Ii(θ) : دالة المعلومات للفقرة (i).

Alhawari: The Effect of Test Length and Ability Distribution Form on the Es

ويرى هامبلتون وسوامنثيان (Hambelton & Swaminathan, 1985) أنه كلما زاد عدد الفقرات تزداد قيمة معلومات الاختبار، بالإضافة إلى أنه كلما زاد الميل تزداد المعلومات التي توفرها الفقرة وهذا يعني أن الفقرات التي تتضمن معالم تمييز كبيرة تقدم معلومات أكبر عن قدرة المفحوصين وبالتالي الحصول على دقة أكبر، وعليه يمكن انتقاء فقرات اختبار اعتمادا على كمية المعلومات التي تسهم بها الفقرات في كمية المعلومات الكلية للاختبار.



شكل (4): منحنى دالة المعلومات لاختبار مكون من خمس فقرات.

ويشير الشكل (4) إلى منحنى دالة المعلومات لاختبار مكون من خمس فقرات. فعند بناء الاختبارات باستخدام نظرية الاستجابة للفقرة يمكن استخدام دالة معلومات الفقرة لاختيار الفقرات على اعتبار أن معلومات الفقرة تتغير عبر مستويات القدرة المختلفة، وعليه فانه من الممكن اختيار فقرات تقدم دقة قياس مرتفعة عند نقطة معينة على متصل السمة، حيث تزيد من القدرة على تقدير أخطاء القياس. وأشار بيكر (Baker, 2001) إلى أن منحنى معلومات الاختبار يقدم فوائد كثيرة، ومعلومات عن دقة الاختبار في تقدير قدرات الأفراد. وبين أن المنحنى المفضل هو الذي تقع قمته عند نقطة الفصل أو القطع (Cut Point).

الخطأ المعياري في التقدير: ويعرف وورم (Warm, 1978) الخطأ المعياري في التقدير أنه القيمة المتوقعة للانحراف المعياري لأخطاء تقدير القدرة. ويتم حسابه بسهولة من منحنى معلومات الاختبار. كما تزود دالة معلومات الاختبار بمستوى الخطأ المعياري للاختبار عند كل مستوى قدرة. لذا فإن الاختبار الذي تم تحليل فقراته من الممكن حساب دالة معلوماته من خلال جمع دوال معلومات الفقرات المكونة له، وكمية المعلومات التي تسهم بها مجموعة من الفقرات عند مستوى قدرة معين تتناسب عكسيا مع الخطأ المعياري في تقدير القدرة عند مستوى معين، فإذا كانت كمية المعلومات عالية فإن الخطأ في التقدير يكون منخفضا، بمعنى أن العلاقة عكسية بينهما. وتحسب قيمة الخطأ المعياري في التقدير عند مستوى القدرة (θ)، وتعطى المعادلة الآتية (Wright & Master, 1982):

Published by Arab Journals Platform, 2015

An-Najah University Journal for Research - B (Humanities), Vol. 29 [2015], Iss. 8, Art. 3

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

حيث: SE(θ): الخطأ المعياري في تقدير القدرة عند مستوى معين من القدرة (θ)، (θ)I: كمية معلومات الاختبار.

وتقيس 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؛ فإن المعلومات التي يقدمها الاختبار عند أي مستوى قدرة هي مجموع مقدار معلومات الفقرات عند ذاك المستوى. وبالتالي المستوى العام لاقتران معلومات الاختبار سوف يكون أكبر من المعلومات لفقرة 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. وباستخدام نموذج راش ذي المعلمة الواحدة، فإن 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 (Baker, 2001).

إن تقدير قدرة المفحوص يتم من خلال نمط استجابة المفحوص على مجموعة من فقرات الاختبار وهو ما يهدف الاختبار إلى قياسه، والقدرة هي السمة أو البعد التي تقيسه فقرات الاختبار، ويمكن أن تكون سمة من سمات الشخصية مثل مفهوم الذات، وقد تكون محددة مثل القدرة على إجراء العمليات على الأعداد، وقد تكون استعداداً أو قابلية أو سمة نفسية. ويستخدم الرمز (θ) للدلالة على قدرة المفحوص مع وجود مقدار من الاحتمالية للاستجابة الصحيحة P(θ) عند كل مستوى قدرة (Hambelton, Swaminathan, Rogers, 1991).

ويرى بيكر (Baker, 2001) أن تقدير قدرة المفحوص يمكن أن تحقق هدفين أساسيين: الأول وهو تقييم المفحوص تبعا للقدرة الحقيقية التي تكمن وراء استجابته عن فقرات الاختبار، والثاني المقارنة بين المفحوصين لغايات التصنيف والابتعاث.

وزودت دراسة جريين وساكس وميشيل (Green, Sax & Michael, 1982) بمقارنة 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-ريتشاردسون (KR- 20) لذوي القدرة المنخفضة والمتوسطة فقط بتفضيل أربعة بدائل، بينما لذوي القدرة المرتفعة فلا توجد فروق دالة إحصائياً.

وينسجم ذلك مع ما توصلت إليه دراسة جودوين (Jodoin, 2003) ودراسة الشريفين (Al-Shraifin, 2006) اللتين توصلتا إلى أن قيمة اقتران المعلومات لاختبار الاختيار من متعدد تتزايد تدريجياً بتزايد القدرة حتى تصل إلى أقصى قيمة ممكنة لها عندما يكون

https://digitalcommons.aaru.edu.jo/anujr_b/vol29/iss8/3

مستوى القدرة (θ=b) متوافق مع صعوبة الفقرة، في حين جاء في دراسة ليفين ودراسجو (Levine & Drasgow, 1983) إن زيادة عدد بدائل الفقرة تؤدي إلى الحصول على اكبر قدر ممكن من المعلومات حول الأفراد ذوي القدرة المنخفضة، أما بالنسبة لذوي القدرات المرتفعة ينصح بالتقليل من عدد البدائل.

وفي دراسة بومرك (Pommerich, 2007) بعنوان أثر استخدام معالم الفقرات تمت معايرتها باستخدام الطريقة التقليدية في الاختبارات التكيفية المحوسبة، والتي هدفت لاختبار البرامج التي تستخدم كل من الكمبيوتر والاختبارات الورقية، حيث اهتمت بدراسة مدى الاستقرار في تقدير معالم الفقرات باستخدام اختبارات مختلفة الطول في كلتا الحالتين. وأظهرت النتائج عدم وجود استقرار في معالم الفقرات حيث يقل هذا الاختلاف بزيادة طول الاختبار.

في حين أجرى فيزباترك (Fitzpatric, 2009) دراسة بعنوان أثر اختبارات التهيئة على معدل إتقان الطلبة والتي هدفت لدراسة اثر تخفيض طول الاختبار على معدل الإتقان. واعتمد الباحث على 12 اختبار من نوع الاختيار من متعدد كل منها تتكون من ثلاثة نماذج على النحو (15، 10، 5) فقرات. واعتمد الباحث على النموذج أحادي المعلم في تحليل بياناته. أظهرت النتائج عدم وجود استقرار في تقدير معالم الفقرة حينما يقل حجم الاختبار، وأوصى الباحث باستخدام الحجم 15 فقرة لزيادة الاستقرار في تقدير معالم الفقرات.

ويرى تريفسان وساكس وميشيل (Trevisan, Sax & Michael, 1991) أن التنبؤ بقدرات الأفراد على الاختبار يكون أكثر دقة إذا تم ربط 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.

وأكد ريس (Reis, 1986) أن البيانات غير الطبيعية تؤثر في تقدير معالم الفقرات، حيث أشار ماك دونلاند ومولاك (McDonald & Mulaik, 1979) أن لطول الاختبار أثر كبير في تقدير معالم الفقرات، وأنه على الأقل لا بد من وجود 15 فقرة لتقدير معالم الفقرات.

وذكر كارن ومايكل وكالن (Karon, Michael, & Kallen, 2008) أن طبيعة توزيع البيانات تؤثر على افتراض أحادية البعد، وأن حجم الفقرات الكبير يقاوم انتهاك التوزيع الطبيعي للبيانات، وان نقصان حجم الفقرات في الاختبارات يتسبب في انتهاك هذا الافتراض، كما انه وثيق الصلة بتوازن أو تجانس المحتوى. وبالتالي يقود إلى نتائج غير صحيحة حول معالم الفقرة، ومعلم القدرة للأفراد، ويعطي تقديرا متحيزا للثبات ودالة المعلومات الاختبار.

مشكلة الدراسة وأسئلتها

لا تزال الاختبارات تتمتع بأهميتها في مجالات التقييم في مدارس وزارة التربية والتعليم، وخاصة تلك التي يتم بناؤها في ضوء نماذج رياضية مختلفة ترتبط بالنظرية الحديثة في القياس

التـي يعوّل عليهـا كثيـرا فـي بنـاء الاختبـارات وبنـوك الأسـئلة الحديثـة، بحيث تتمتـع بخصـائص سيكومترية تـوفر لمستخدمي ومطـوري الاختبـارات وصـانعي القـرارات الخيـارات المتعـددة فـي اختيـار فقـرات المقيـاس المحـدد. ونظـرا لأهميتهـا فـي الكشـف عـن القـدرات الحقيقيـة للأفـراد، وخاصـة في المجال التعليمي. كـان لا بـد مـن إيجـاد مقـاييس تتمتـع بنـوع مـن الثبـات والدقـة، علـى اختلاف أشكالها وأنواعها. بالرغم من وجود عدد من المتغيرات التي قد تؤثر على تقديرات قدرة المفحوصـين كطـول الاختبـار وشـكل توزيـع القـدرة. وجـاءت تحديداً لدراسـة **أثـر طـول الاختبـار وشكل توزيـع القدرة فـي تقديرات قدرة الأفراد وفق نمـوذج راش فـي نظريـة اسـتجابة الفقـرة** ولتجيب عن الأسئلة الآتية:

السؤال الأول: هل تختلـف دقـة تقـدير معلمـة صـعوبة الفقـرة المقـدرة باسـتخدام نمـوذج راش باختلاف طول الاختبار (30، 60) فقرة؟

السؤال الثاني: هل تختلف دقة تقدير معلمـة صـعوبة الفقـرة المقدرة باستخدام نمـوذج راش باختلاف شكل توزيع القدرة (التوزيع الطبيعي، ملتوٍ التواءً موجباً، ملتوٍ التواءً سالباً)؟

السؤال الثالث: هل تختلف دقة تقدير قدرة الأفراد المقدرة باستخدام نمـوذج راش بـاختلاف طول الاختبار (30، 60) فقرة؟

السؤال الرابع: هل تختلف دقة تقدير قدرة الأفراد المقدرة باستخدام نمـوذج راش بـاختلاف شكل توزيع القدرة (التوزيع الطبيعي، ملتوٍ التواءً موجباً، ملتوٍ التواءً سالباً)؟

أهداف الدراسة

تهـدف الدراسـة الحاليـة إلـى معرفـة أثـر طـول الاختبـار وشـكل توزيـع القـدرة علـى متغيـري صعوبة الفقرة واقتـران المعلومـات للفقـرات والاختبـار، وعلـى قـدرة الفـرد. وأثرهـا علـى إمكانيـة تقدير كل من معلمة الصعوبة وقدرة الأفراد.

أهمية الدراسة

تأتي أهمية هذه الدراسة من أهمية الاختبارات والمقاييس المختلفة والمستخدمة فـي الكشـف عن السمات المختلفة وتفسيرها، كما يساعد علـى إيجـاد المقيـاس الصـحيح والـذي مـن خلالـه يمكـن الكشف عن السمة بدقة وعناية، ويأتي دور الاختبـارات فـي إمكانيـة بنـاء صـور متعـددة للمقـاييس من خلال معايرتها ومعادلتها، لضمان صدق المقياس وفاعليتـه. وتكمـن أهميتهـا فـي الكشـف عـن أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وأهميـة ذلـك فـي بنـاء اختبـارات ومقاييس تكشف عن السمة المقاسة بدقة وثبات. واثر ذلك على شكل توزيع القدرة على متغيـري صعوبة الفقرة واقتران المعلومات للفقرات والاختبار بالإضـافة إلى قدرة الفـرد، حيث تـم تحليـل البيانـات المولـدة بطريقـة أكثـر موضـوعية وهـي اسـتخدام النمـوذج نمـوذج راش الـذي يكسـب 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.

تعريف المصطلحات

صعوبة الفقرة: القيمة التي تتدرج على متصل السمة ويكون احتمال إجابة الفقرة عندها إجابة صحيحة مساويا 50% حينما يكون التخمين صفر.

قدرة الفرد: قيمة يتم تقديرها بتعظيم أرجحية استجابات الفرد عن فقرات الاختبار استجاب اليها.

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 (Baker, 2001).

منهجية الدراسة وإجراءاتها

المتغيرات المستقلة: وهي: طول الاختبار (60، 30) فقرة، شكل توزيع القدرة (طبيعي، ملتو التواءً موجباً، ملتو التواءً سالباً).

المتغير التابع: وهو دقة تقدير معلم الصعوبة ودقة تقدير قدرة المفحوصين.

مجتمع الدراسة

تكون مجتمع الدراسة من البيانات المولّدة تبعاً للنموذج استجابة الفقرة (أحادي المعلمة) في الفترة الواقعة ما بين 2014/1/1 إلى 2014/3/1. إذ تم توليد مجموعة الفقرات بواقع (1000) فرداً، وتوليد تقديرات لقدرات الأفراد تبعا لمتغيرات الدراسة بالاعتماد على النموذج الأحادي المعلمة وإخضاعها للتحليل الإحصائي. وتم استخدام برنامج (WINGEN) لتوليد تقديرات قدرات (1000) فرد، من توزيع طبيعي بوسط حسابي يساوي صفرا وانحراف معياري يساوي واحد، ولجميع متغيرات الدراسة؛ من حيث: طول الاختبار، شكل توزيع القدرة، تبعا للنموذج أحادي المعلمة. وتسمى هذه القدرات بالقدرات الحقيقية للأفراد، ويوضح الجدول (1) الإحصائيات الوصفية لمعالم القدرة تبعاً لشكل التوزيع.

جدول (1): الإحصاءات الوصفية للقدرة الحقيقية.

توزيع القدرة	القيمة	القيمة	الوسط	الانحراف	الالتواء	التفلطح
ملتو التواءً سالباً	-2.248	2.998	1.7459	1.0003	-1.016	.642
طبيعي	-3.531	3.63	-0.027	0.9885	-.131	.119
ملتو التواءً	-3	2.407	-1.659	1.0686	.977	.465

يتضح من الجدول 1 أن قيمة القدرة الحقيقية قد تراوحت بين 2.248- و2.998 بوسط حسابي 1.7459 وانحراف معياري 1.0003 عندما توزعت القدرة بشكل ملتو التواءً سالبا، أما عندما توزعت القدرة بشكل طبيعي تراوحت قيمة القدرة الحقيقية بين 3.531- و3.63 بوسط

An-Najah University Journal for Research - B (Humanities), Vol. 29 [2015], Iss. 8, Art. 3

حسابي 0.027- وانحراف معياري 0.9885، في حين تراوحت قيمة القدرة الحقيقية بين 3.0- و2.407 بوسط حسابي 1.659- وانحراف معياري 1.0686 عندما توزعت القدرة بشكل ملتو التواءً موجبا.

أداة الدراسة

تكونت من مجموعة الفقرات المولدة تبعا للمتغير المستقل طول الاختبار (30، 60) فقرة، بالاعتماد على معلم القدرة الذي تم توليده مسبقا، وتبعا للنموذج (أحادي المعلمة)، ويمثل الجدول (2) ملخصا للإحصاءات الوصفية لمعالم الفقرات تبعاً لشكل التوزيع.

جدول (2): الإحصاءات الوصفية لمعلمة الصعوبة.

شكل توزيع	عدد	القيمة	القيمة	الوسط	الانحراف	الالتواء	التفلطح
ملتو التواءً سالباً	30	-	2.393	0.178	1.0075	-0.2373	0.145
	60	-	3.125	-	1.1456	0.1239	0.073
طبيعي	30	-	1.701	0.143	0.77	-0.3255	-
	60	-	1.918	0.115	0.9406	-0.1443	-
ملتو التواءً موجباً	30	-1.87	2.166	-	1.0081	0.3574	-
	60	-	2.176	0.217	0.9956	-0.0799	-

يتضح من خلال الجدول 2 أن قيم معلمة الصعوبة تراوحت ضمن المدى المطلوب (- 2, 2)، لمتغيرات الدراسة: طول الاختبار (30، 60)، وشكل توزيع القدرة (ملتو التواءً سالبا، طبيعي، ملتو التواءً موجبا)، بوسط حسابي قريب من الصفر وبانحراف معياري قريب من الواحد صحيح، تبعا للنموذج أحادي المعلمة، وهذا يحقق الغاية من الدراسة المطلوبة.

التحقق من افتراض أحادية البعد: تم التحقق من هذا الافتراض باستخدام فكرة تحليل المكونات الأساسية واعتماد احد الطرق الآتية:

- نسبة التباين المفسر للعامل الأول: أشار ريكاس (Reckase, 1979) أنه إذا كانت نسبة التباين المفسر للعامل الأول (20%) من التباين الكلي دل على أحادية البعد.
- نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني: وقد حدد جلورفيلد (Glorfeld, 1995) النسبة بحيث تكون أكبر أو تساوي 2 للدلالة على أحادية البعد (Linacre, 2008).
- التمثيل البياني لقيم الجذور الكامنة: وهو رسم خاص للجذور الكامنة يوضح من خلاله المكونات الداخلة في التحليل.

https://digitalcommons.aaru.edu.jo/anujr_b/vol29/iss8/3

Alhawari: The Effect of Test Length and Ability Distribution Form on the Es

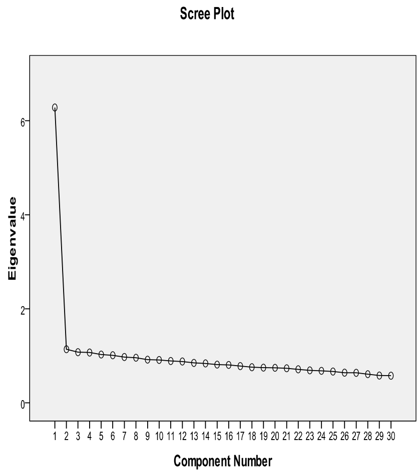
جدول (3): يم الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني عند طول الاختبار 30 فقرة.

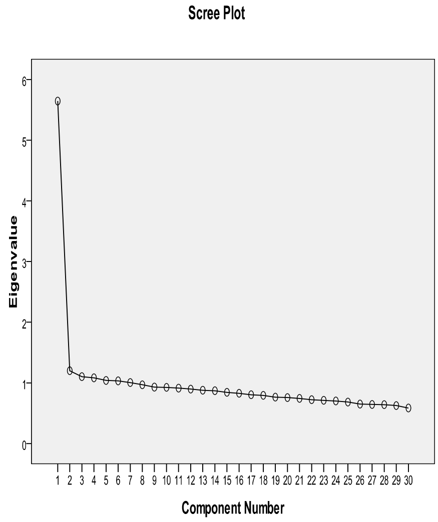
شكل توزيع القدرة	الجذر الكامن نسبة التباين المفسر	العامل الأول	العامل الثاني	نسبة الجذر الكامن للعامل الأول إلى العامل الثاني
ملتو التواء سالب	الجذر الكامن	5.646	1.200	4.705
	نسبة التباين المفسر	18.819%	4.000	
توزيع طبيعي	الجذر الكامن	5.625	1.098	5.122
	نسبة التباين المفسر	18.750%	3.662	
ملتو التواء موجب	الجذر الكامن	6.281	1.140	5.509
	نسبة التباين المفسر	20.938%	3.798	

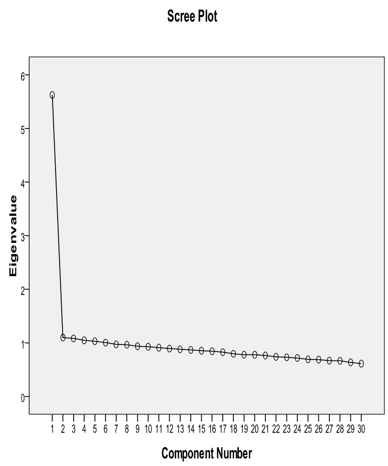
يتضح من الجدول (3) أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر من(2)، وأن ما يفسره العامل الأول من التباين يقارب (18.819%) في حالة الالتواء السالب، أما في حالة التوزيع الطبيعي لشكل القدرة كانت 18.750%، وفي حالة الالتواء الموجب كانت تزيد عن (20%)، هذان مؤشران على أحادية البعد، كما يتضح تحقق افتراض أحادية البعد من خلال التمثيل البياني باستخدام (Scree Plot) لقيم الجذور الكامنة للعوامل المكونة لكل حالة من حالات طول الاختبار وشكل توزيع القدرة.

يتضح من الشكل 5 التالي استقرار قيمة الجذر الكامن بعد العامل الأول تقريبا، وهذا يدل على وجود عامل سائد على العوامل الأخرى. وقد استقرت قيمة الجذر الكامن في حالة اختلاف توزيعات القدرة (ملتو سالب، طبيعي، ملتو موجب) وهذا مؤشر على تحقق افتراض أحادية البعد.

An-Najah University Journal for Research - B (Humanities), Vol. 29 [2015], Iss. 8, Art. 3







ملتو التواء سالب توزيع طبيعي ملتو التواء موجب

شكل (5): التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة لكل شكل من أشكال توزيع الاختبار.

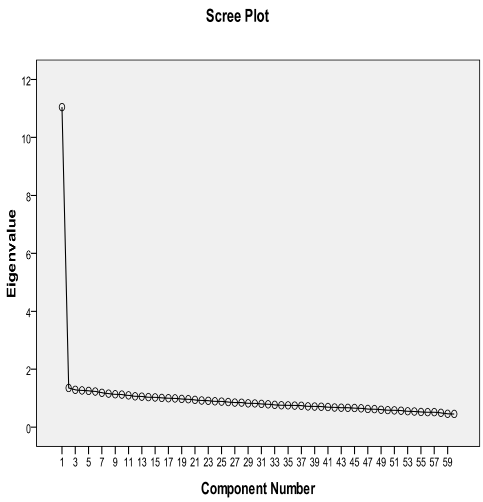
https://digitalcommons.aaru.edu.jo/anujr_b/vol29/iss8/3 16

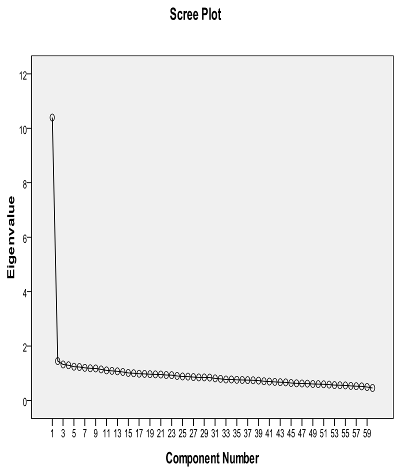
جدول (4): قيم الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني عند طول اختبار 60 فقرة.

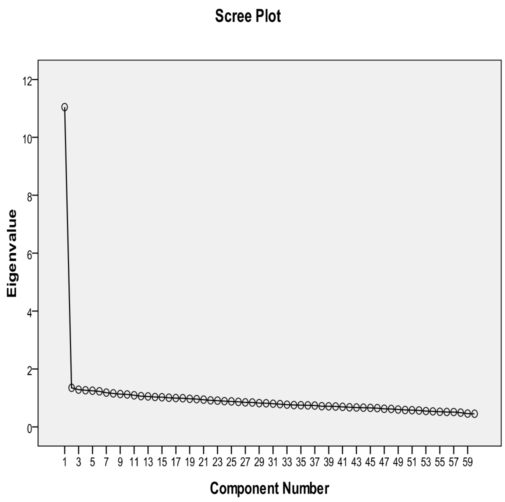
شكل توزيع القدرة	الجذر الكامن نسبة التباين المفسر	العامل الأول	العامل الثاني	نسبة الجذر الكامن للعامل الأول إلى العامل الثاني
ملتو التواء سالب	الجذر الكامن	10.396	1.443	7.204
	نسبة التباين المفسر	17.327%	2.406	
توزيع طبيعي	الجذر الكامن	9.855	1.140	8.644
	نسبة التباين المفسر	16.425 %	3.798	
ملتو التواء موجب	الجذر الكامن	11.046	1.352	8.170
	نسبة التباين المفسر	18.108 %	2.217	

يتضح من الجدول (4) أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر من (2)، وأن ما يفسره العامل الأول من التباين يساوي 17.327%، ما يقارب (20%) في حالة الالتواء السالب، وعند التوزيع الطبيعي لشكل القدرة كانت نسبة التباين المفسر 16.425%، أما في حالة الالتواء الموجب كانت 18.108%، وهذان مؤشران على أحادية البعد، كما يتضح تحقق افتراض أحادية البعد من خلال التمثيل البياني باستخدام (Scree Plot) لقيم الجذور الكامنة للعوامل المكونة لكل حالة من حالات طول الاختبار وشكل توزيع القدرة.

An-Najah University Journal for Research - B (Humanities), Vol. 29 [2015], Iss. 8, Art. 3







ملتو التواء سالب توزيع طبيعي ملتو التواء موجب

شكل (6): التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة لكل شكل من أشكال توزيع القدرة.

يلاحظ من الشكل (6) 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 يعتبر افتراض الاستقلال الموضعي مكافئا لأحادية البعد في حالة النموذج أحادي المعلمة.

عرض النتائج ومناقشتها

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

1. **النتائج المتعلقة بدقة تقدير معلمة صعوبة الفقرة باستخدام نموذج راش لتغير طول الاختبار (30,60) فقرة**

للإجابة عن السؤال الأول والذي ينص على: هل تختلف دقة تقدير معلمة صعوبة الفقرة المقدرة باستخدام نموذج راش باختلاف طول الاختبار (30، 60) فقرة؟ تم تحليل البيانات باستخدام تحليل التباين الأحادي للكشف عن مدى الفروق في المتوسطات الحسابية لمتوسطات الأخطاء المعيارية في تقدير معالم الصعوبة باختلاف طول الاختبار حيث يوضح الجدول 5 النتائج:

جدول (5): نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير معالم الصعوبة باختلاف طول الاختبار.

مصدر التباين S of v	مجموع المربعات Sum of Squares	درجات الحرية df	التباين المقدر Mean Square	قيمة F المحسوبة F	الدلالة الإحصائية Sig.
بين المجموعات	.446	1	.446	151.186	. 000
داخل المجموعات	2.950	998	.00295		
الكلي	**3.396**	**999**			

أشارت نتائج تحليل التباين الأحادي في الجدول (5) لمتوسطات الأخطاء المعيارية في تقدير قدرات معالم صعوبة الفقرة باختلاف طول الاختبار إلى وجود فروق ذات دلالة إحصائية عند مستوى القدرة (α= 0.05) بين متوسطات الأخطاء المعيارية في تقدير معالم الصعوبة تعزى لطول الاختبار ولصالح الاختبار بطول 30 فقرة، هذا يعني أن طول الاختبار كان له اثر على تقدير معلمة الصعوبة، وان نقصان حجم الاختبار يؤثر سلبا على تقدير معلمة الصعوبة. وان لطول الاختبار أثر على توازن او تجانس المحتوى، وهذا ما أكده كارن ومايكل وكالن (Karon, Michael, & Kallen, 2008). وأن كمية المعلومات التي تعطيها تلك الفقرات (30) كافية لتغطية متصل السمة. وأن زيادة حجم الفقرات لا يزيد من كمية معلومات الاختبار بل يبقيها ثابتة.

Published by Arab Journals Platform, 2015

2. النتائج المتعلقة بدقة تقدير معلمة صعوبة الفقرة المقدرة باستخدام نموذج راش باختلاف شكل توزيع القدرة

وللإجابة عن السؤال الثاني والذي ينص على: هل تختلف دقة تقدير معلمة صعوبة الفقرة المقدرة باستخدام نموذج راش باختلاف شكل توزيع القدرة (التوزيع الطبيعي، ملتوٍ التواءً موجباً، ملتوٍ التواءً سالباً)؟ تم استخدام طريقة شفيه (Scheffe's Method) لعقد المقارنات البعدية بين متوسطات الأخطاء المعيارية لتقديرات معلمة الصعوبة لأفراد الاختبار لمعرفة أي الفروق دالة إحصائياً، ويبين الجدول (6) نتائج المقارنات الثنائية بين متوسطات الأخطاء المعيارية:

جدول (6): نتائج المقارنات الثنائية بين متوسطات الأخطاء المعيارية لتقديرات معالم الصعوبة في الاختبار.

مقارنات الأزواج	المتوسط الحسابي	الانحراف المعياري	الأخطاء المعيارية	درجات الحرية	الدلالة الإحصائية
p1 & p 2	.034800	1.306688	.238568	29	.885
P2 & p 3	.313300	1.348866	.246268	29	.213
P1 & p 3	.278500	1.331417	.243082	29	.021
q 1 & q2	-.147267	1.526658	.197091	59	.458
q 1 & q3	-.249550	1.553993	.200620	59	.218
q 2 & q3	-.102283	1.474584	.190368	59	.593

حيث تشير p إلى الاختبار بطول 30 فقرة، وتشير q إلى الاختبار بطول 60 فقرة، في حين أن الأرقام 3 ,2 ,1 تشير إلى شكل توزيع القدرة (ملتو التواء سالب، طبيعي، ملتو التواء موجب) على التوالي. وكانت النتائج لصالح الاختبار 30 فقرة كما هو مشار اليها في الجدول (6) عندما توزعت القدرة بشكل ملتو التواءً موجباً وسالباً. ولتفسير ذلك تعزو الباحثة السبب إلى أن الالتواء السالب والالتواء الموجب يكون لمعلم الصعوبة، وأن انحياز الفقرات في اتجاه واحد دليل على الارتباطات المرتفعة فيما بينها، وأن ترابطها معا يشكل اتساقاً داخلياً. ولم تقدم أي من الدراسات السابقة أثر شكل الالتواء على دقة تقدير معلمة الصعوبة، مما تجدر الإشارة إليه تفوق هذه الدراسة في تحقيق نتائج حول أثر شكل توزيع القدرة على دقة تقدير معلمة الصعوبة.

3. النتائج المتعلقة بدقة تقدير معالم القدرة باختلاف طول الاختبار(30، 60) فقرة

وللإجابة عن السؤال الثالث الذي ينص: هل تختلف دقة تقدير قدرة الأفراد المقدرة باستخدام نموذج راش باختلاف طول الاختبار (30، 60) فقرة؟ تم تحليل البيانات باستخدام تحليل التباين الأحادي للكشف عن مدى الفروق في المتوسطات الحسابية لمتوسطات الأخطاء المعيارية في تقدير معالم القدرة باختلاف طول الاختبار حيث يوضح الجدول 7 النتائج:

جدول (7): نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير معالم القدرة باختلاف طول الاختبار.

مصدر التباين S of v	مجموع المربعات	درجات الحريةdf	التباين المقدر	قيمة F المحسوبةF	الدلالة الإحصائية.Sig
بين المجموعات	.422	1	.422	174.38	. 000
داخل المجموعات	2.420	998	.00242		
الكلي	**2.842**	**999**			

أشارت نتائج تحليل التباين الأحادي في الجدول (7) إلى وجود فروق ذات دلالة إحصائية عند مستوى القدرة ($\alpha = 0.05$) بين متوسطات الأخطاء المعيارية في تقدير قدرات الأفراد تعزى لطول الاختبار ولصالح الاختبار يطول 30 فقرة.

4. النتائج المتعلقة بدقة تقدير معالم القدرة باختلاف شكل توزيع القدرة باستخدام نموذج راش للاختبار

وللإجابة عن السؤال الرابع والذي ينص: هل تختلف دقة تقدير قدرة الأفراد المقدرة باستخدام نموذج راش باختلاف شكل توزيع القدرة (التوزيع الطبيعي، ملتوٍ التواءً موجباً، ملتوٍ التواءً سالباً)؟: تم تحليل البيانات باستخدام تحليل التباين الأحادي للكشف عن مدى الفروق في المتوسطات الحسابية لمتوسطات الأخطاء المعيارية في تقدير معالم القدرة باختلاف شكل توزيع القدرة ويوضح الجدول 8 النتائج:

جدول (8): نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير معالم القدرة باختلاف شكل التوزيع.

شكل توزيع القدرة	مصدر التباين S of v	مجموع المربعات Sum of Squares	درجات الحرية df	التباين المقدر Mean Square	قيمة F المحسوبة F	الدلالة الإحصائية Sig.
ملتو التواء سالب	بين المجموعات	.205	2	.1025	61.377	.000
	داخل المجموعات	1.669	997	.00167		
	الكلي	1.874	999			
طبيعي	بين المجموعات	. 0420	2	.0021	7.36	.302
	داخل المجموعات	2. 850	997	.00285		
	الكلي	2.892	999			
ملتو التواء موجب	بين المجموعات	. 191	2	.0955	47.98	.001
	داخل المجموعات	1. 990	997	.00199		
	الكلي	2.181	999			

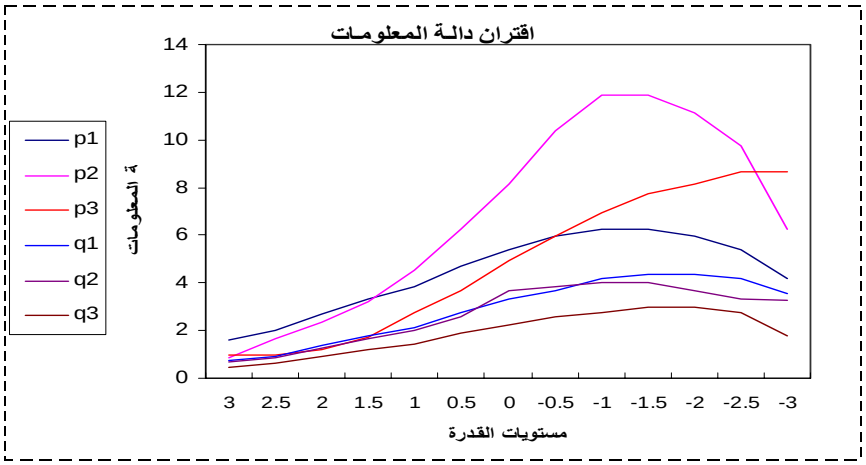
أشارت نتائج تحليل التباين الأحادي في الجدول (8) لمتوسطات الأخطاء المعيارية في تقدير قدرات الأفراد باختلاف شكل توزيع القدرة إلى وجود فروق ذات دلالة إحصائية عند مستوى القدرة ($\alpha= 0.05$) بين متوسطات الأخطاء المعيارية في تقدير قدرات الأفراد تعزى لشكل توزيع القدرة، حيث كان متوسط الأخطاء المعيارية في حالة شكل التوزيع الطبيعي غير دال إحصائيا. في حين كان الاختبار الملتوٍ التواء سالبا والملتوٍ التواء موجبا أكثر فاعلية في تقدير معلم قدرة الأفراد.

تم حساب قيم اقتران المعلومات عند قيم مختارة من القدرة لإيجاد اقتران المعلومات لفقرات الاختبار،باستخدام البرنامج الإحصائي (BILOG-MG) لتحليل البيانات، والجدول (9) يبين تقديرات قدرات أفراد الاختبار لكل مستوى منها بوحدة (اللوجيت)، وقيم اقتران المعلومات لها عند ذاك المستوى.

جدول (9): قيم اقتران المعلومات لكل اختبار عند مستويات مختارة من القدرة.

القدرة(θ)	اقتران المعلومات 30 فقرة			اقتران 60 فقرة المعلومات		
	p1	P2	P3	q 1	q 2	q 3
3	1.60	0.84	1	0.718	0.67	0.48
2.5	1.98	1.64	1.00	0.943	0.87	0.63
2	2.69	2.37	1.21	1.352	1.26	0.91
1.5	3.31	3.19	1.73	1.778	1.64	1.18
1	3.84	4.53	2.78	2.100	1.98	1.42
0.5	4.73	6.25	3.70	2.778	2.60	1.88
0.0	5.41	8.16	4.94	3.306	3.70	2.23
-0.5	5.95	10.41	5.95	3.698	3.84	2.60
-1	6.25	11.89	6.93	4.165	4.00	2.78
-1.5	6.25	11.89	7.72	4.340	4.00	2.97
-2	5.95	11.11	8.16	4.340	3.70	2.97
-2.5	5.41	9.77	8.65	4.165	3.31	2.78
-3	4.16	6.25	8.65	3.560	3.25	1.78

حيث تشير p إلى الاختبار بطول 30 فقرة، q الاختبار بطول 60 فقرة، في حين أن الأرقام 1, 2, 3 تشير إلى شكل توزيع القدرة (ملتو التواء سالب، طبيعي، ملتو التواء موجب) على التوالي، وقد تم رسم النتائج المشار إليها في الجدول (9) وتم الحصول على الشكل (7).



شكل (7): منحنيات اقتران المعلومات.

ويشير الشكل (7) إلى منحنيات اقتران المعلومات لمستوى طول الاختبار التي توضح كمية المعلومات التي يقدمها الاختبار عند مستويات القدرة المختلفة ضمن النتائج المشار إليها في الجدول (9). يتضح من خلال الشكل أن الاختبار عندما يكون بطول 30 فقرة يقدم كمية معلومات اكبر، إذ أن كمية المعلومات المقدمة بطول اختبار 30 فقرة عندما توزعت القدرة بشكل ملتو التواءً موجباً أكبر من المنحنيات الأخرى، أيضا قدمت معلومات أكثر حينما توزعت بشكل ملتو التواءً سالباً، أما عند زيادة طول الاختبار (60) فقرة لم يقدم أي معلومات عن الاختبار، وهذا يعني أن قيمة اقتران المعلومات مؤشراً على ثبات الاختبار كونها تتناسب عكسياً مع خطأ التقدير الذي يزداد ثبات الاختبار بنقصانه (Wright & Master, 1982).

مما سبق يتبين أن طول اختبار 30 فقرة وشكل توزيع القدرة (الملتو التواءً موجباً وسالباً) كان أكثر فاعلية في تقدير قدرة المفحوص، إذ أن طبيعة توزيع البيانات وحجم الفقرات تؤثر في كمية المعلومات التي تقدمها كما أنها تؤثر على تقدير معلمة الصعوبة للأفراد وتقدير قدرة الأفراد. لذا يوصى عند بناء الاختبارات المقننة أو لاختبارات التكيفية اختيار عدد فقرات الاختبار 30 فقرة ليتناسب مع مستوى قدرة الأفراد، فزيادة طول الاختبار لا يقدم أي معلومات عن الاختبار، إضافة إلى ذلك فهو يقلل من ثبات الاختبار، ولا يعطي تقديرا صحيحا لقدرات الأفراد.

التوصيات

توصـي الباحثـة بمزيدٍ مـن الدراسـات حـول أثـر طـول الاختبـار وشـكل توزيـع القـدرة تبعـا لنمـاذج نظريـة الفقـرة الأحاديـة والمتعـددة التـدريج. كمـا توصـي بإتبـاع الطـرق الصـحيحة عنـد صياغة بنوك الاختبار تبعا لنماذج النظرية الحديثة في القياس للكشف عن قدرة الأفراد بصـورة صحيحة. كما توصي الباحثة بإجراء البحث على بيانـات حقيقيـة للحصـول علـى المزيـد مـن الدقـة في القياس.

References (Arabic & English)

- Allam, S. (2005). *Models of uni and multi-dimensional in Item Response Theory, Principles and Application in Educational and Psychological Measurement*. t1, Cairo: Dar Al feker Arabi.
- Alshafei, M. M. (2008). *Effects of violation some IRT assumptions on the calibration of item bank and Equating scores of drown tests*, a worksheet, the regional seminar for Psychology, Faculty of Education, King Saud University, the symposium publications.
- Al-Shraifin, N . (2006). *Psychometric Properties of the Criteria reference test in Educational Measurement and Evaluation according to the modern theory of measurement and evaluation.* Journal of Educational and Psychological Sciences (4) 7.109 to 80.
- Anastasi, A. & Urbina, S. (1997). *Psychological Testing. (7 ed.),* New York: Printice Hall.
- Baker, Frank B. (2001). *The Basic Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation .
- Crocker, L. & Algina, J. (1986). *Introduction to Classical & Modern Test Theory*. Fort Worth, TX: Harcourt Brace Jovanovich.
- 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.

- Jodoin, M. G. (2003). *Measurement efficiency of innovative item Formats in computer-Based testing*. Journal of Educational Measurement, 40(1). 1-*15.*
- Glorfeld, L.W. (1995). *An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain*. Educational and Psychological Measurement. 55, 377-393.
- 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.
- Hattie, J. (1985). *Methodology review: Assessing unidimensionality of tests and items*. Applied Psychological Measurement, 9, 139–164.
- Hambleton, R. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications***.** Boston : Kluwer .
- Hambleton, R.K. & Jones, 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, Ronald K. & Swaminathan, Hariharan & Rogers, H. Jane. (1991). *Fundamentals of Item Response Theory***.** Newbury Park, NJ: SAGE.S.
- Hulin, C. L. Drasgow, F. & Parsons, C.K. (1983). *Item Response Theory: Application to Psychological Measurement*. Illinois: Dow Jones - Irwin.
- Levine, M. V. & Drasgow, F. (1983). *The relation between incorrect option choice and estimation ability*. Educational and Psychological Measurement, 43, 675-685.
- Lord, F.M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
- Karon F. Cook Michael A. Kallen. (2009). *Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption*. Qual Life Res (2009) 18:447–460 DOI 10.1007/s11136-009-9464-4.

An-Najah University Journal for Research - B (Humanities), Vol. 29 [2015], Iss. 8, Art. 3

- Linacre, J. M. (2008). Winsteps Rasch measurement (Version 3.63.2). Chicago, IL: MESA Press. Masters. N.G. (1982). *A Rasch Model for Partial Credit Scoring*. Psychometrika, 47, 149-174.
- McDonald, R. P. & Muliak, S. (1979). *Determinancy of common factors: A non-technical review*. Psychological Bulletin, 86, 297-306.
- 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.
- Reckase, M. D. (1979). *Unifactor latent trait models applied to multifactor tests: Results and implications*. Journal of Educational Statistics, 4, 207–230.
- Reis, R. (1986). *A structural equation analysis of Weiner's attribution-affect model of helping behavior*. Journal of Personality and Social Psychology, 50, 1123-1133.
- Sawalmeh, y. (1994). *The multiple Tests of true and false: characteristics And the relative number of correct alternatives.* Research Yarmouk: Series Human and Social Sciences 10 (2) from 0.573 to 595.
- Sawalmeh, y. & Kawasmeh, A. (2000). *The impact of diversity in the number of alternatives right and wrong in The multiple Tests of true and false in their psychometric characteristics*. Damascus University Journal 16 (1) 0.61 to 88.
- 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.
- Warm, T.A. (1978). *A Primer of Item Response Theory*. Oklahoma: U .S. Coast Guard Institute 73/69.
- Wright, B. D. & Masters, G. N. (1982). *Rating Scale Analysis Rasch Measurement.* MESAPress,Kimbark Avenue**.** Chicago, P. 59.

https://digitalcommons.aaru.edu.jo/anujr_b/vol29/iss8/3