

2020

## The Effect of Sample Size and Estimation Methods on the Accuracy of Estimating the Parameters of the Rush Model Preparation

هبة ضعضع  
جامعة حلب، سوريا, hebadada@yahoo.com

منار طومان  
جامعة حلب، سوريا, manarToman@yahoo.com

مصطفى طيفور  
جامعة حلب، سوريا, mustafataifor@yahoo.com

Follow this and additional works at: <https://digitalcommons.aaru.edu.jo/jpu>



Part of the [Arts and Humanities Commons](#), and the [Social and Behavioral Sciences Commons](#)

### Recommended Citation

(طيفور, مصطفى (2020) and ضعضع, هبة; طومان, منار "The Effect of Sample Size and Estimation Methods on the Accuracy of Estimating the Parameters of the Rush Model Preparation," *Jerash for Research and Studies Journal* *الدراسات والبحوث للبحوث والدراسات*: Vol. 21 : Iss. 1 , Article 6.  
Available at: <https://digitalcommons.aaru.edu.jo/jpu/vol21/iss1/6>

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Jerash for Research and Studies Journal *الدراسات والبحوث للبحوث والدراسات* by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact [rakan@aar.edu.jo](mailto:rakan@aar.edu.jo), [marah@aar.edu.jo](mailto:marah@aar.edu.jo), [u.murad@aar.edu.jo](mailto:u.murad@aar.edu.jo).

---

## The Effect of Sample Size and Estimation Methods on the Accuracy of Estimating the Parameters of the Rush Model Preparation

### Cover Page Footnote

جميع الحقوق محفوظة لجامعة جرش 2020. معيدة وطالبة دكتوراه، كلية التربية، جامعة حلب، سوريا. أستاذ مساعد، كلية التربية، جامعة حلب، سوريا.

## أثر حجم العينة وطرائق التقدير في دقة تقدير معالم نموذج راش The Effect of Sample Size and Estimation Methods on the Accuracy of Estimating the Parameters of the Rush Model Preparation

هبة عبد اللطيف ضعيف\*، منار أحمد طومان\*\* و مصطفى أحمد طيفور\*\*

تاريخ القبول 2019/10/2

تاريخ الاستلام 2019/4/30

### ملخص

هدفت الدراسة إلى الكشف عن أثر طرائق التقدير (الارجحية العظمى، تقدير بيبز، طريقة بروكس) عند حجوم عينة مختلفة (2000، 1500، 1000، 500) على تقدير معالم المفردة والأفراد ودقتها في ضوء نظرية الاستجابة للمفردة (نموذج راش). لتحقيق لأهداف الدراسة تم توليد عينة عشوائية من (2000) استجابة على (40) مفردة ثنائية استجابة (0 أو 1) باستخدام برنامج Wingen بقدرة وصعوبة ذات توزيع طبيعي معياري، ومن ثم التحقق من افتراضات نظرية الاستجابة للمفردة (أحادية البعد، الاستقلال المحلي، المنحنى المميز للمفردة) والتحقق من ملاءمة المفردات لنموذج راش، نتيجة لذلك حصلنا على (31) مفردة ملاءمة لنموذج راش.

أظهرت النتائج أن:

1. عدم وجود فروق ذات دلالة إحصائية في تقدير معلم المفردة لنموذج راش (الصعوبة) ومعامل الأفراد (القدرة) تعزى لعامل حجم العينة وطريقة التقدير والتفاعل بينهما.
  2. توجد فروق ذات دلالة إحصائية في دقة تقدير معالم المفردة (الصعوبة) تعزى لعامل حجم العينة وطريقة التقدير والتفاعل بينهما.
  3. توجد فروق ذات دلالة إحصائية في دقة تقدير معالم الأفراد (القدرة) تعزى لعامل حجم العينة وطريقة التقدير والتفاعل بينهما.
  4. تزداد دقة تقدير معالم صعوبة المفردة ومعامل قدرة الأفراد بازدياد حجم العينة.
  5. لا يوجد فروق ذات دلالة إحصائية بين طريقة الأرجحية العظمى وطريقة بيبز من جهة وطريقة بروكس من جهة أخرى.
  6. توجد فروق ذات دلالة إحصائية في دقة تقدير قدرة الأفراد بين الطرائق الثلاث مثنى مثنى.
- الكلمات المفتاحية: حجم العينة، طريقة التقدير، دقة التقدير، نموذج راش.

© جميع الحقوق محفوظة لجامعة جرش 2020.

\* معيدة وطالبة دكتوراه، كلية التربية، جامعة حلب، سوريا.

\*\* أستاذ مساعد، كلية التربية، جامعة حلب، سوريا.

## Abstract

The aim of this study was to investigate the effect of estimation methods (maximum likelihood, Bayes estimation, Proxy method) at different sample sizes (2000, 1500, 1000, 500) on the estimation of Item parameters and Ability Parameter and their accuracy in the light of the item response theory (Rush model). To achieve the objectives of the study, a random sample of (2000) responses were generated on (40) single-response (0 or 1) using Wingen software with difficulty and standard normal distribution, and then verified the assumptions of item response theory (Unidimensionality, local independence, item characteristic curve) and to check the suitability of items for the Rush model. As a result, we got (31) items fits for the Rush model.

The results showed that:

1. There are no statistically significant differences in the estimation of the item parameter of the Rush model (difficult) and Ability Parameter attributable to the sample size factor and the method of estimation and interaction between them.
2. There are statistically significant differences in the accuracy of estimating the item parameters (difficult) due to the factor of sample size and the method of estimation and interaction between them.
3. There are statistically significant differences in the accuracy of estimating the parameters of individuals (power) attributable to the sample size factor and the method of estimation and their interaction.
4. The accuracy of the item parameter (difficult) and the accuracy of the ability parameter factor is more accurate as the sample size increases.
5. There are no statistically significant differences between the maximum odds method and the Bayes method on the one hand and the Bayes method and Proxy method on the other.
6. There are statistically significant differences in the accuracy of estimating the ability of individuals between the three methods.

**Keywords:** Sample Size, Estimation Method, Estimation Accuracy, Rush Model.

## المقدمة

تهدف عملية القياس في العلوم بعامة للحصول على نتائج دقيقة من خلال تطوير أساليب وطرائق القياس، وجعلها أكثر دقة وأقل أخطاء، إضافة إلى معرفة الظروف البحثية الملائمة لكل طريقة، وتحذو العلوم التربوية والنفسية حذو العلوم الأخرى وذلك ببناء مقاييس وفق أسس علمية دقيقة، وذلك باستخدام مجموعة من الأساليب الإحصائية توجت نهاية ببناء أول نظرية للقياس وسميت النظرية الكلاسيكية.

استخدمت هذه النظرية لبناء الاختبارات وتحليل مفرداتها وتفسير نتائجها والتحقق من صدقها وثباتها، سعياً لتحقيق أعلى درجات الموضوعية في هذه الاختبارات، ومع الانتشار الواسع

لاستخدام هذه النظرية في بناء أنواع مختلفة من الاختبارات والمقاييس النفسية، تبين فاعليتها. انتشر العمل في هذه النظرية بشكل واسع واستخدم في القياس النفسي نظراً لسهولة، ولكنها واجهت العديد من الانتقادات، ولعل أهمها التعبير عن قدرة الفرد من خلال الدرجة الظاهرية التي تقدر من خلال أدائه على الاختبار، وبناءً عليه فإنه سيتغير تقدير قدرة الفرد بتغير خصائص الاختبار من حيث السهولة والصعوبة، كما أن خصائص الاختبار والمفردات تتغير بتغير خصائص الأفراد.

الامر الذي أسفر عن ظهور اتجاه حديث في القياس، والذي أطلق عليه في بداية الأمر نظرية السمات الكامنة (LTT) Latent Traits Theory، بينما أطلق عليها لاحقاً نظرية الاستجابة للمفردة (IRT) Item Response Theory، وتعد هذه النظرية من النظريات السيكمومترية المعاصرة التي اتضحت فائدتها في التغلب على الكثير من مشكلات القياس التي عجزت النظرية الكلاسيكية عن مواجهتها، والميزة الأكثر أهمية في هذه النظرية، هي استقلال تقدير معالم المفردة (الصعوبة، التمييز، التخمين) عن قدرة الأفراد المستخدمين لإيجاد هذه الإحصائيات، وكذلك استقلال تقدير قدرات الأفراد عن معالم المفردات المستخدمة في عملية التقدير. هذا ويشير بعض خبراء القياس من أمثال لورد Lord ونوفيك Novick أن النظرية المعاصرة في القياس تُعد امتداداً لمفاهيم ومبادئ النظرية الكلاسيكية<sup>1</sup>، حيث توفر نماذج هذه النظرية تقديراً للقدرة مستقلاً عن العينة، بمعنى آخر يعبر عن تقديرات القدرة بوحدة لا تتعلق بصفات العينة، أو مجموعة معينة من الأفراد، فكما أن وزن فرد لا يتعلق بعمره أو جنسه، فإن تقدير قدرة الفرد المشتق من أي من نماذج نظرية الاستجابة للمفردة لا يتعلق بهذه العوامل أو غيرها من المميزات<sup>2</sup>.

لذا جاءت نظرية الاستجابة للفقرة استجابة للتطور البحثي، للتخلص من العيوب التي شاب استخدام النظرية التقليدية في التقويم التربوي والنفسي، إلا أن الركيزة الأساسية التي يتوقف عليها استخدام نظرية الاستجابة للفقرة هي قضية التقدير الإحصائي لمعالم الفقرات وقدرة الأفراد، حيث تعتمد دقة هذا التقدير على كثير من العوامل الذي اهتم البحث الايومتري بدراساتها.

**1- مشكلة الدراسة:** ان أحد الاساليب التي تؤثر في التباين الحقيقي وتباين الخطأ هي المعاينة الاحتمالية التي تشكل القاعدة الاساسية للإحصاء التحليلي وكما يشكل اختيار عينة عشوائية ممثلة واختيار الاساليب الاحصائية المناسبة اهم التحديات التي تواجه الباحثين وتؤثر في دقة التحليلات الاحصائية والاستنتاجات حول تقدير معالم المجتمع واختبار الفروض على مدى عشوائية العينة وصدق تمثيلها للمجتمع<sup>3</sup>. ومن المشكلات الأساسية التي ترتبط بالعينة هي مشكلة حجمها، فهناك تباين في تحديد حجم العينة المناسب الذي قد يمثل حجم المجتمع، فهناك من

يميل إلى الحجم الكبير بغية تمثيلها للمجتمع، وهناك من يميل إلى الحجم الصغير بغية السيطرة عليها، إلا أنه لا يوجد معيار ثابت أو نسبة مئوية محددة لتحديد حجم العينة المناسب<sup>4</sup>.

از تعد اخطاء المعاينة من المصادر الأساس لأخطاء القياس التي لا يقل تأثيرها عن أخطاء المقياس، وقد تتأتى اخطاء المعاينة من طريقة اختيار العينة أو من حجمها، حيث كما يبدو لا يوجد اتفاق على حجم مناسب للعينة في قياس المفاهيم النفسية، لاسيما في حساب الخصائص القياسية للمقياس التي تعد من اهم خطوات بناء المقاييس النفسية ومن اهم مؤشرات قدرة المقياس على قياس السمة التي اعد لقياسها<sup>5</sup>.

وبالرغم من ان الباحثين في جميع مجالات البحث لديهم فكرة عن العينات وكيف ينبغي الاهتمام باختيارها الا ان الواقع الملموس لحركة البحث العلمي لبعض الدراسات والبحوث يعكس صورة مغايرة لضوابط ومعايير البحث العلمي، از تزخر بالعديد من الممارسات الخاطئة في اختيار العينة والخلل بشروطها وافتراضاتها من عشوائية وتمثيل، وفي نمط آخر من الممارسات الخاطئة نجد ان حجم العينة يحدد بطريقة تميل الى الذاتية والتخمين أو الاعتماد على نسب حددت على مجتمعات قد لا تتناسب مع مجتمع الدراسة، الامر الذي يجعل النتائج يشوبها الغموض والريبة لخللها لشرط التمثيل<sup>6</sup>.

وبما ان موضوعية وصدق نتائج الاختبارات تعتمد على دقة الأساليب التي استخدمت في بناءها واختيار فقراتها وتفسير نتائجها، وكذلك في وصفها للقدرة التي يقيسها الاختبار، لذلك فان الأمر يتطلب ضرورة استخدام التوجهات الحديثة في القياس والتي اثبتت البحوث التجريبية انها تحقق الدقة والموضوعية المنشودة في العلوم النفسية والتربوية<sup>7</sup>. كما درس تأثير طرائق التقدير على دقة تقدير معالم المفردة مثل دراسة بوك وأيتكن<sup>8</sup> (Bock & Aitkin, 1981) ودراسغو<sup>9</sup> (Drasgow, 1989) وسيونغ<sup>10</sup> (Seong, 1990) وستون<sup>11</sup> (Stone, 1992) وين<sup>12</sup> (Yen, 1987) والعبد الله<sup>13</sup> (2012) والعضايلة<sup>14</sup> (2016)، كما اشارت الدراسات دراسة فاريش<sup>15</sup> (Farish, 1984) وهامبلتون وكوك<sup>16</sup> (Hambleton & Cook, 1983) وريبي وجينز<sup>17</sup> (Ree & Jensen, 1983) والصريرة<sup>18</sup> (2008) والمدانات<sup>19</sup> (2008) والقحطاني<sup>20</sup> (2011) ودرويش<sup>21</sup> (2014) والعبيسات<sup>22</sup> (2015) وجبران<sup>23</sup> (2017) والطراونة<sup>24</sup> (2017) والصريرة<sup>25</sup> (2017) لأثر حجم العينة على دقة تقدير معالم المفردة بتدخل عوامل أخرى، الأمر الذي دعى إلى دراسة أثر هذين العاملين معاً على دقة التقدير، كما تباينت نتائج الدراسات كدراسة فان<sup>25</sup> (Fan, 1998) ودراسة عبدالجبار والشافعي<sup>26</sup> (2006) والزهراني<sup>27</sup> (2008) في تحديد حجم العينة الملانم واثر كل من الاسلوب والحجم على الاختبارات التي يتم بنائها او معالجتها احصائياً في ضوء نظرية الاستجابة للمفردة.

وبناءً على ما سبق تتلخص مشكلة الدراسة بالإجابة عن التساؤل الآتي:

ما أثر حجم العينة وطريقة التقدير في دقة تقدير معالم المفردة وفق نموذج راش

ويتفرع عن التساؤل الرئيسي التساؤلات التالية:

- ماهي تقديرات معاملات المفردات (الصعوبة) ومعاملات الأفراد (القدرة) وفق نموذج راش عند حجوم عينة (2000، 1500، 1000، 500) وفق طريقة الارجحية العظمى وطريقة تقدير بيبز وطريقة بروكس؟
- هل توجد فروق ذات دلالة إحصائية في تقدير معالم المفردة (صعوبة المفردة) وفق نموذج راش ودقتها تعزى لحجم العينة وطرائق التقدير والتفاعل بينهما؟
- هل توجد فروق ذات دلالة إحصائية في تقدير القدرة وفق نموذج راش ودقتها تعزى لحجم العينة وطريقة التقدير والتفاعل بينهما؟

2- أهداف الدراسة: تحاول الدراسة الحالية تحقيق الأهداف التالية:

1. التحقق من افتراضات نظرية الاستجابة للمفردة، ومطابقة الاختبار لنموذج راش وتقدير معالم المفردة والأفراد وفق لطرائق التقدير (طريقة الأرجحية، طريقة بيبز، طريقة بروكس) عند حجوم عينة مختلفة.
2. دراسة أثر حجم العينة وطريقة التقدير على صعوبة المفردة ودقتها المقدره لنموذج راش والتفاعل بينهما.
3. دراسة أثر حجم العينة وطريقة التقدير على قدرة الأفراد ودقتها المقدره لنموذج راش والتفاعل بينهما.

3- أهمية الدراسة: تتلخص أهمية الدراسة بـ:

1. الإجابة عن أسئلة الدراسة والتوصل إلى أثر حجم العينة على دقة التقدير في ضوء طريقتي الارجحية وبيبز للتقدير.
2. تعد الدراسة الحالية في مجال نظرية الاستجابة للمفردة من الموضوعات الحديثة التي تحتاج المزيد من التوضيح والفهم، وعلى هذا فإن الدراسة الحالية ستغني المكتبة العربية بأحد الموضوعات محدودة الدراسة حتى الوقت الراهن.

3. ستركز الدراسة الحالية على دقة تقدير معالم المفردة والتي تعتبر الخطوة الأولى والأهم لإعداد الأدوات والتأكد من خصائصها، تمهيداً للاستفادة منها في التطبيقات الواسعة لنظرية الاستجابة للمفردة.
4. تتيح هذه الدراسة فحص تأثير طريقة التقدير وحجم العينة في دقة تقدير معاملات المفردة في نموذج راش، مما يمهد السبيل لتوظيف نتائج الدراسة الحالية في دراسات أخرى.
5. تأتي هذه الدراسة استمراراً لجهود العاملين في مجال القياس والتقويم لتطوير الاختبارات، وانتقاء أفضل الطرائق للكشف عن الأداء التفاضلي للمفردات.
6. إمكانية تعميم نتائج الدراسة واستخدامها مع المقاييس النفسية والاختبارات التحصيلية الأخرى.

#### 4- مصطلحات الدراسة:

**حجم العينة Sample Size:** وتمثل العدد الكلي لأفراد العينة والمأخوذة بأحد طرق تصميمات العينة العشوائية.<sup>28</sup>

#### تعرفه الباحثة اجرائياً:

**حجم العينة:** يبلغ استجابات 2000 مفحوص وعينات عشوائية جزئية منها بحجم (1500، 1000، 500)

**طرائق التقدير Estimation Methods:** هي مجموعة من الأساليب الرياضية الاحتمالية بغية تقدير معالم المفردة والأفراد في ضوء افتراضات نظرية الاستجابة للمفردة، حيث تناولت الدراسة (طريقة التقدير القائمة على الأرجحية العظمى Maximum Likelihood Estimation، طريقة التقدير القائمة على نظرية بيز Bayesian Estimation، طريقة بروكس (Proxy)

**دقة التقدير Accuracy Of Estimation:** تعبير يشير إلى جودة التقدير لمعالم المفردة والأفراد، والتي تتميز بالاحتمال الكبير في أن يكون التقدير قريب من القيمة الحقيقية للمعلم، وذلك باختيار التقدير غير المتحيز Unbiased Estimator، صاحب أقل تباين بين التقديرات الأخرى غير المتحيزة، وذلك باستخدام الخطأ المعياري في التقدير.<sup>29</sup>

نموذج راش **Rasch Model**: وهو أحد نماذج نظرية الاستجابة للمفردة أحادية البعد، يهتم بتحديد موقع المفردة الاختبارية على ميزان صعوبة جميع المفردات التي تشكل الاختبار (معلم الصعوبة)، كما يهتم بتدريج حجوم قدرة الفرد باختبار معين على نفس ميزان تعبير المفردات<sup>30</sup>، يفترض هذا النموذج أن معامل التمييز متساوي لكل مفردات الاختبار وانعدام أثر التخمين.

## 5- الإطار النظري والدراسات السابقة:

### 5-1- الإطار النظري:

نظرية الاستجابة للمفردة **Item Response Theory**: مع بداية الستينيات شقت نظرية الاستجابة للمفردة طريقها بقوة وجدارة في البحوث النفسية والتربوية، وأصبحت ميداناً واسعاً يهيمن على ساحة البحث الفعال، نظراً لأهميتها في تطور اتجاهات القياس واتخاذها مسارا جديداً يفي بشروط الموضوعية، ويضع الحلول المناسبة للمشكلات التي لم تستطع النظرية الكلاسيكية مواجهتها.

حيث تعتبر هذه النظرية الإطار الإحصائي الذي يربط درجات الاستجابة على مفردات الاختبار بالسمات الكامنة التي أعد الاختبار لقياسها<sup>31 32 33</sup> ففي مجال العلوم الإنسانية والسلوكية، يتسم قياس السمات النفسية بأنه غير مباشر، حيث يتم قياس مقدار ما يمتلكه الفرد من السمة من خلال أدائه على مواقف لها علاقة بالسمة المدروسة، أي من خلال السلوك الظاهري القابل للملاحظة على ما يدل عليها ولهذا يعبر عنها بطرق احتمالية<sup>34</sup>.

طرائق التقدير: **Estimate Methods**: لقد انبثق عن نظرية الاستجابة للمفردة طرائق رياضية لتقدير معالم نماذج هذه النظرية والتي منها:

### طريقة الأرجحية العظمى (Maximum likelihood Method (ML)

تقوم هذه الطريقة على أساس إيجاد تقدير المعالم من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها وتعتبر من أكثر طرائق التقدير شهرة. لتكن لدينا  $X_j = (x_{1j}, x_{2j}, \dots, x_{nj})$  والذي يعبر عن استجابة فرد ما على (n) مفردة حيث:

$$x_{ij} = \begin{cases} 0 & \text{إذا كانت إجابة الفرد على المفردة } i \text{ خاطئة} \\ 1 & \text{إذا كانت إجابة الفرد على المفردة } i \text{ صحيحة} \end{cases}$$

حيث يعبر عن دالة الاحتمال عندئذ دالة الأرجحية العظمى تأخذ الشكل:

$$L(U/\theta) = \prod_{i=1}^n \prod_{j=1}^N P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}} \dots \dots \dots (4)$$

لإيجاد القيمة المتوقعة للقدرة يتم أخذ اللوغارتم الطبيعي للمعادلة السابقة واشتقاقها جزئياً بالنسبة للمشتق وعدم هذا المشتق، كالتالي:

$$S(\theta) = \frac{\partial}{\partial \theta} \ln \left( L \left( \frac{U}{\theta} \right) \right) = 0$$

$$\Rightarrow \frac{\partial}{\partial \theta} \sum_{j=1}^n \sum_{k=0}^{m_j} u_{jk} \ln (P_{jk}(\theta)) = \sum_{j=1}^n \sum_{k=0}^{m_j} \frac{u_{jk}}{P_{jk}} \frac{\partial P_{jk}(\theta)}{\partial \theta} = 0^{35}$$

1. طريقة الأرجحية العظمى المشتركة (Joint Maximum likelihood estimation): يمكن تطبيق هذه الطريقة في النماذج اللوجستية الأحادية والثنائية والثلاثية ويتم وفقاً لهذه الطريقة تقدير معالم القدرة والفقرة في آن معاً.
2. طريقة الأرجحية العظمى الشرطية (Conditional Maximum Likelihood Estimation): وتقوم هذه الطريقة بفصل المعالم الاحصائية للمفحوصين أثناء عملية التدرج وتطبق على النموذج اللوجستي الأحادي حيث يكون اقتران الاحتمالية (Likelihood) مشروطاً بعدد الإجابات الصحيحة.
3. طريقة الأرجحية العظمى الهامشية (Marginal Maximum Likelihood Estimation): ويمكن تطبيق هذه الطريقة على النماذج اللوجستية الأحادية والثنائية والثلاثية حيث يتم وفقاً لهذه الطريقة إيجاد اقتران الاحتمالية الهامشي (Marginal Likelihood Function) لمعالم الفقرة من خلال تكامل اقتران الكثافة الاحتمالية على معلم القدرة ثم إيجاد تقديرات معالم الفقرة وعندما تكون تقديرات معالم الفقرة معروفة فإنه يمكن عندئذ تقدير معلمة القدرة<sup>36</sup>.

**طريقة بيبز (Bayesian Estimation):** تتميز هذه الطريقة بالإضافة لاستخدامها معلومات عن العينة بأن تعتمد طريقة بيبز على استخدام معلومات مسبقة عن المعالم غير المعروفة معتبراً هذه المعالم متغيرات عشوائية (Random Variables)، وبافتراض ان لهذه المعالم معلومات مسبقة (Prior Information) التي يمكن صياغتها على شكل توزيع احتمالي يعرف بدالة الكثافة الاحتمالية السابقة (Prior P.d.f) ويجري التعرف على هذه المعلومات من بيانات وتجارب سابقة أو من النظرية التي تحكم تلك الظاهرة، وكذلك تعتمد نظرية بيبز على معلومات العينة الحالية المتمثلة بدالة الإمكان (Likelihood Function) الخاصة بالملاحظات، وعليه بدمج دالة الكثافة الاحتمالية للمعالم  $P(\theta)$  مع دالة الأمكان  $P(Y/\theta)$ ، يتم الحصول على دالة الكثافة الاحتمالية اللاحقة للمعالم (Posterior p.d.f) وهي  $P(\theta/Y)$ <sup>37</sup>.

طريقة بروكس لتقدير معالم نموذج راش: وقد أشار لها علام (2001)<sup>38</sup> تقوم هذه الطريقة على الخطوات التالية:

1. نقوم باستبعاد المفردات التي نجح (أو فشل) فيها جميع أفراد العينة، ونستبعد الأفراد الذين نجحوا (أو فشلوا) في الإجابة على جميع المفردات.
2. نحسب المقدار  $x_i = \ln\left(\frac{m-m_i}{m_i}\right)$  والذي يمثل لوغاريتم نسبة الإجابات الخطأ إلى الإجابات الصحيحة للمفردة ( $i$ ) التي حصل عليها ( $m$ ) من الأفراد. حيث ان  $m_i$ : عدد الأفراد الذين أجابوا أجا به صحيحة على المفردة ( $i$ ) حيث  $i = 1, 2, \dots, n$
3. نحسب  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$  وهو متوسط  $x_i$  لعدد قدره ( $n$ ) من المفردات.
4. نحسب  $Y_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$  وهو تباين  $x_i$  لعدد قدره ( $n$ ) من المفردات.
5. نحسب  $z_k = \ln\left(\frac{k}{n-k}\right)$  وهو لوغاريتم نسبة الإجابة الصحيحة إلى الإجابات الخطأ للمفردات والتي عددها ( $n$ ).
6. نحسب  $\bar{Z} = \frac{\sum_{k=1}^m z_i}{m}$  وهو متوسط  $z_i$  لجميع الأفراد وعددهم ( $m$ ).
7. نحسب  $Y_z^2 = \frac{\sum_{i=1}^m (z_i - \bar{z})^2}{m-1}$  وهو تباين  $z_i$  لجميع الأفراد وعددهم ( $m$ ).
8. نفرض أن  $T_x = \left[ \frac{1 + \frac{Y_x^2}{2.89}}{1 - \frac{Y_x^2 + Y_z^2}{2.89}} \right]^{\frac{1}{2}}$  هو عامل تعديل لتحرير صعوبة المفردات من تشتت الصعوبة، حيث أن المقدار  $(1.7)^2 = 2.89$  حيث أن  $1.7$  هو معامل التدرج الذي يحول المنحنى اللوغاريتمي إلى تطابق تقريبي مع المنحنى الطبيعي.
9. كما أن  $T_z = \left[ \frac{1 + \frac{Y_z^2}{2.89}}{1 - \frac{Y_x^2 + Y_z^2}{2.89}} \right]^{\frac{1}{2}}$  هو عامل تعديل لتحرير قدرة الأفراد من تشتت القدرة
10. تقدير صعوبة المفردة ( $i$ ) هو  $T_x * x_i$  مع خطأ معياري قدره  $T_x * \left[ \frac{m}{m_i(m-m_i)} \right]^{\frac{1}{2}}$ .
11. تقدير قدرة الفرد ( $k$ ) هو  $T_z * z_k$  مع خطأ معياري قدره  $T_z * \left[ \frac{n}{k(n-k)} \right]^{\frac{1}{2}}$ .

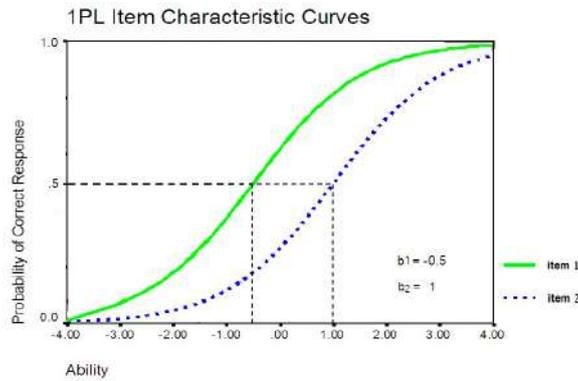
النموذج اللوغاريتمي أحادي المعلم (راش) One-Parameter Logistic Model: يطلق على هذا النموذج بنموذج المعلم الواحد (One-Parameter Model)، ويهتم بتحديد موقع

المفردة الاختبارية على ميزان صعوبة جميع المفردات التي تشكل الاختبار، كما يهتم بتدريج مستويات قدرة الفرد باختبار معين على نفس ميزان تعبير المفردات<sup>39</sup>، حيث أقرح تحليل بيانات الاختبار وفقاً لنظرية الاحتمال (Probability Theory). وكانت المحصلة نموذج لوغاريتمي للمنحنى المميز للمفردة<sup>40</sup>. وهذا النموذج يفترض أن معايرة صعوبة المفردة وقدرة الشخص وحدها كافية وفعالة لتقدير معالم المفردات والقدرة. ويعطى بالعلاقة:

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

$P_i(\theta_j)$  هو احتمال الإجابة الصحيحة للفرد الذي قدرته  $\theta_j$  على المفردة التي صعوبتها  $b_i$ .<sup>41</sup>

ويمكن تمثيل الدالة الاحتمالية لنموذج راش بالشكل التالي:



الشكل (1): منحنى خصائص لمفردتين لنموذج راش

ويتضح من الشكل أن المنحنى خصائص المفردتين متوازيين، ويختلفان فقط في إزاحة المنحنى عن المحورين الاحداثيين، وذلك نابع عن تفاوت في معامل الصعوبة وثبات قيمة معامل التمييز، وانعدام معامل التخمين.

**2-5- الدراسات السابقة:** اهتمت بعض الدراسات بتناول تأثير عوامل معينة على دقة تقدير معالم الفقرات فعلى سبيل المثال هدفت دراسة سرحان (2018)<sup>42</sup> تهدف إلى المقارنة بين طريقتي الأرجحية العظمى والقيمة العظمى للتوزيع البعدي في دقة تقدير القدرة وفقاً لنظرية الاستجابة للفقرة، ولتحقيق أهداف الدراسة هذا الهدف اتبع الباحث خطوات علمية في إجراءات تحليل لاختبار تفسير الرموز الصادر من (مركز التعليم السريع LLC, 2001, Learning Express) على وفق نظرية الاستجابة للفقرة، الذي يضم (3) اختبارات فرعية تقيس التشابهات الرمزية (30) فقرة، فرز وتصنيف الأشكال (30) فقرة، التفكير في سلسلة (30) فقرة، طبق الاختبار على عينة

عشوائية طبقية مكونة من (1000) طالب وطالبة. تكون الاختبار في صورته النهائية من (73) فقرة. أشارت النتائج وجود فرق في تقدير القدرة بين الطريقتين ولصالح طريقة القيمة العظمى للتوزيع البعدي من خلال انخفاض قيم الخطأ المعياري لها.

دراسة العبد الله (2012)<sup>43</sup> الدراسة الكشف عن أثر طريقة التقدير (EAP<sup>45</sup>، WML<sup>44</sup>) ، MLE<sup>46</sup>) على دقة تقدير معلم القدرة لدى الأفراد باستخدام نموذج التقدير الجزئي المعمم GPCM مع مراعاة طول الاختبار وحجم العينة ونوع توزيع البيانات، وتم استخدام بيانات حقيقية ومولدة، حيث تم سحب عينة عنقودية بحجم (903) طالب وطالبة من ست مدارس مثلت ثلاث إدارات تربية (الهرم، الدقي، الجيزة)، كما تم تطبيق اختبار تحصيلي محكي المرجع لوحدين دراسيتين في اختبار الجبر والإحصاء، حيث تكون من (19) مفردة متعددة التدرج ومختلفة في عدد فئات الاستجابة.

كما استخدم بيانات مولدة حاسوبياً باستخدام برنامج WINGEN والذي يمثل نموذج التقدير الجزئي المعمم (GPCM)، حيث تم توليد ثلاثة حجومات مختلفة من البيانات (1000، 1500، 2000)، واعتماد ثلاثة توزيعات مختلفة لحجم العينة وهي (الطبيعي، المنتظم، بيتا)، واستخدام ثلاثة أطوال مختلفة للاختبار وهي (10، 20، 30) مفردة، حيث تكون الاختبار من أربع فئات.

أظهرت النتائج بأن ترتيباً أفضل طرق التقدير هو EAP و WML و MLE وذلك عند تغيير حجم العينة مع تثبيت كل من طول الاختبار ونوع التوزيع وكذلك عند تغيير محك التقييم RMSE عنها في Bias. ترتيباً أفضل طرق التقدير هو EAP ثم WML ثم MLE وذلك بالنسبة لمحك التقييم RMSE عنها في Bias بغض النظر عن المتغيرات الأخرى المتمثلة بحجم العينة وطول الاختبار ونوع التوزيع. وازدادت دقة تقدير معلم القدرة عند الأفراد بازدياد حجم العينة وطول الاختبار ذلك بالنسبة لمحك التقييم RMSE عنها في Bias بغض النظر عن طريقة التقدير المستخدمة، كما أظهرت أن أفضل النتائج في تقدير معالم القدرة كانت عند استخدام التوزيع الطبيعي للبيانات ثم التوزيع المنتظم وأخيراً توزيع بيتا، وذلك بالنسبة لمحك التقييم RMSE عنها في Bias بغض النظر عن طريقة التقدير المستخدمة.

وعلى صعيد آخر اهتمت العديد من الدراسات باستقصاء دقة تقديرات معالم الفقرة والقدرة

عند اختلاف الطريقة المستخدمة في تقدير تلك المعالم، ففي دراسة قام بها كل من جاري وفيرمينت<sup>47</sup> (Garre & Vermunt, 2006) تم استخدام المنحى الأميريقي في جزء من الدراسة من خلال تطبيق بيانات حقيقية بهدف التحقق من نتائج الاشتقاقات الرياضية التي توصل والمتعلقة

بدقة تقديرات المعلمات وذلك عند استخدام من خلال استخدام توزيعات قبلية مختلفة لمعلمة القدرة وقد دلت النتائج أن تقديرات المعلمات التي يتم الحصول عليها من خلال طريقة بيبز تتمتع بثبات أفضل من تقدير المعلمات التي تم الحصول عليها عند استخدام طريقة الأرجحية العظمى وخصوصاً عند تقدير القدرات على أطراف متصل القدرة.

وفي دراسة قام بها وانغ وفيسبول<sup>48</sup> (Wang & Vispoel, 1998) هدفت إلى تقييم النتائج التي يتم الحصول عليها من استخدام أربع طرق في تقدير القدرة وهي: طريقة الأرجحية العظمى وثلاثة طرق مرتبطة بأسلوب بيبز هي: طريقة أوينر (Owen's Method)، وتوقع التوزيع البعدي (EAP)، وطريقة تعظيم التوزيع البعدي: (MAP)، إذ تم استخدام أسلوب المحاكاة لاختبار تكيفي، وقد أظهرت النتائج اختلافات واضحة بين طريقة الأرجحية العظمى وطريقة بيبز من بينها أن طريقة الأرجحية العظمى تعطي أخطاء عالية في تقدير معلمة القدرة مقارنة بطرق بيبز، وفيما يتعلق بطرق بيبز فإن طريقة أوينر كانت أسوأ طريقة من حيث دقة التقديرات التي تقدمها.

وفي دراسة قام كيم<sup>49</sup> (Kim, 2001) هدفت إلى مقارنة دقة التقديرات التي تم الحصول عليها باستخدام طريقة الأرجحية العظمى الشرطية وطريقة الأرجحية العظمى الهاشمية وطريقة الأرجحية العظمى المشتركة وطريقة بيبز التي تعتمد على توقع التوزيع البعدي وذلك من خلال استخدام أسلوب المحاكاة، إذ توصلت الدراسة إلى أن نتائج تقدير معلمة القدرة ومعلمة القدرة كانت متشابهة للطرق الأربعة.

تتفق الدراسات الواردة سابقاً مع الدراسة الحالية باهتمامها بطريقة التقدير كعامل مؤثر في دقة تقدير، وقد راعت بعض الدراسات أخذ حجم العينة بعين الاعتبار، إلا أننا لا يمكننا التنبؤ بأثر طريقة تقدير معالم القدرة على معاملات المفردة نموذج راش، الأمر الذي دعا إلى دراسة أثر كل من حجم العينة وطريقة تقدير للمفردة على دقة تقدير معالم الأفراد على وفق نموذج راش.

## 6- الطريقة والإجراءات:

**1-6- منهج البحث:** يعد المنهج التجريبي من أكثر مناهج البحث العلمي كفاءة ودقة وهذا يرتبط بمجموعة من المميزات التي يتمتع بها<sup>50</sup>. وان من أهم مميزات هذا المنهج تعدد تصميماته وتطور وسائل القياس، مما جعل منه منهجاً مرناً يمكن تكييفه لحالات كثيرة، خصوصاً مع طبيعة الظاهرة الإنسانية التي يصعب معها ضبط المتغيرات الخارجية<sup>51</sup>. والتصاميم التجريبية التامة تشكل الأساس الذي يتم بناء عليه الوصول إلى التصاميم التجريبية التي تتناول عدداً من المعالجات (بدلاً من معالجة واحدة) كالتصاميم العملية، ومنها تصاميم القياسات المتكررة<sup>52</sup>. إذ استخدمت الباحثة هذا المنهج بغية تحقيق أهداف البحث.

**2-6- عينة الدراسة:** هي العينة المستخدمة للمقارنة بين طريقتي التقدير عند حجوم عينة مختلفة، حيث تم توليد استجابة (2000) لـ (40) مفردة ثنائية الاستجابة موزعة توزيع طبيعي معياري وذلك باستخدام برنامج WINGEN، ثم تم أخذ عينات عشوائية جزئية بحجم (1500، 1000، 500).

**3-6- خصائص القياسية للاختبار:** ستقوم الباحثة من التحقق من افتراضات نظرية الاستجابة للمفردة على العينة المولدة، كما يلي:

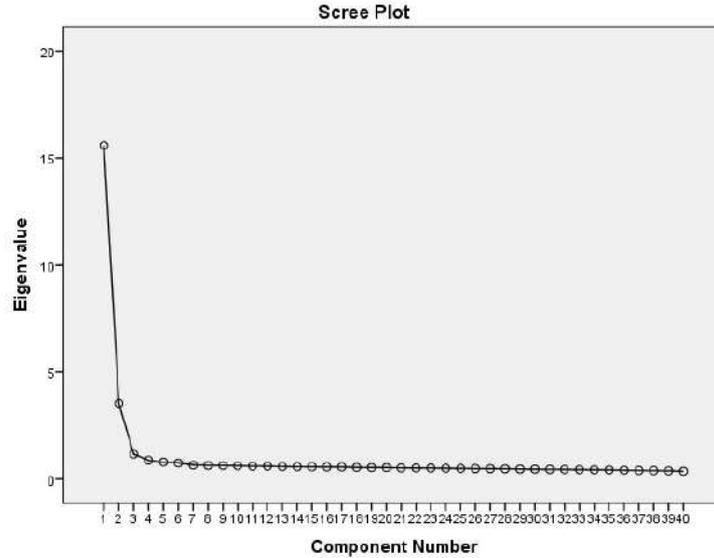
(1) افتراض أحادية البعد **Unidimensionality** للاختبار: تفترض نظرية الاستجابة للمفردة وجود سمة واحدة تُفسر أداء الفرد في المقياس، ولذلك تُسمى بالنماذج أحادية البعد. وللتحقق من هذا الافتراض تم الاعتماد على بعض المؤشرات التي اقترحها هاتي<sup>53</sup> (Hattie, 1985) ومنها معامل ألفا كرونباخ، حيث بلغت قيمته (0.961) وهو مؤشر قوي على أحادية البعد<sup>54</sup>.

وكذلك من المؤشرات المهمة للدلالة على أحادية البعد ما يعتمد على المكونات الرئيسية Indices based on Principal Components، وهي مؤشرات تستند على استخدام التحليل العاملي (Factor Analysis). وقبل البدء في إجراءات التحليل العاملي للتحقق من أحادية البعد لمفردات الاختبار تم التأكد من ملاءمة وكفاية حجم العينة، وذلك من خلال قيمة اختبار Kaiser-Meyer-Olkin (KMO-Test) لكفاية العينة، والتي بلغت (0.988) وهي بحسب محك كايزر تعد نسبة ممتازة. وفيما يتعلق بالشرط الثاني الذي يتطلب أن يكون اختبار بارتليت Bartlett's Test Of Sphericity دالة إحصائياً، ودلالته تعني أن مصفوفة الارتباطات ليس مصفوفة الوحدة Identity Matrix أي (خالية من العلاقات بين المفردات)، فقد أظهرت النتائج أن قيمة اختبار بارتليت بلغت (38411.236) وهي دالة إحصائياً وفقاً لقيمة الدلالة sig=0.000. وبعد التأكد من شروط التحليل العاملي قامت الباحثة باستخدام التحليل العاملي الاستكشافي بطريقة المكونات الأساسية Principle Components لاستجابات الطلاب على مفردات الاختبار، وذلك للتحقق ما إذا كان الاختبار يقيس سمة واحدة كما هو مفترض وبعد إجراء تدوير المحاور وفق Varimax. وقد تم حساب قيمة الجذر الكامن Eigenvalue، ونسبة التباين المفسر Explained Variance، وكذلك التباين المفسر التراكمي لكل عامل من العوامل:

الجدول (1): التباين الكلي المفسر للتحليل العاملي الخاص بالبيانات المولدة

مجموع مربعات التشبعات المستخلصة			الجذور الكامنة الابتدائية			
نسبة التباين المفسر التراكمي	نسبة التباين المفسر	الجذر الكامن	نسبة التباين المفسر التراكمي	نسبة التباين المفسر	الجذر الكامن	العامل
38.961	38.961	15.584	38.961	38.961	15.584	1
47.800	8.839	3.536	47.800	8.839	3.536	2
50.710	2.909	1.164	50.710	2.909	1.164	3
			52.898	2.188	.875	4
			54.882	1.984	.794	5
			56.753	1.871	.748	6
			58.384	1.631	.652	7
			59.976	1.592	.637	8
			61.552	1.576	.630	9
			63.097	1.544	.618	10
			64.601	1.505	.602	11
			66.102	1.500	.600	12
			67.555	1.453	.581	13
			68.993	1.438	.575	14
			70.418	1.424	.570	15
			71.832	1.415	.566	16
			73.237	1.405	.562	17
			74.593	1.356	.542	18
			75.938	1.344	.538	19
			77.277	1.339	.536	20
			78.579	1.302	.521	21
			79.868	1.289	.516	22
			81.150	1.282	.513	23
			82.404	1.254	.501	24
			83.638	1.235	.494	25
			84.851	1.213	.485	26
			86.061	1.209	.484	27
			87.254	1.193	.477	28
			88.416	1.162	.465	29
			89.573	1.156	.463	30
			90.703	1.130	.452	31
			91.820	1.118	.447	32
			92.926	1.105	.442	33
			94.009	1.084	.433	34
			95.079	1.069	.428	35
			96.115	1.036	.414	36
			97.124	1.010	.404	37
			98.118	.994	.397	38
			99.087	.969	.388	39
			100.000	.913	.365	40

ويتضح أن العامل الأول يفسر أكبر نسبة تباين في الدرجات مقارنة ببقية العوامل؛ حيث بلغت قيمة الجذر الكامن للعامل الأول (15.584)، بنسبة تباين (38.961%) وهي أعلى نسبة تباين مفسر، مما يعني أن هذا العامل هو المسيطر على تفسير التباين الكلي لدرجات الاختبار، وبالتالي يعد هذا الاختبار أحادي البعد، أي أن هناك سمة كامنة واحدة يقيسها الاختبار وهي المسؤولة عن تفسير ما يحدث من تباين في درجات الاختبار. والتمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للمقياس يوضح ذلك:



الشكل (2). تمثيل قيم الجذور الكامنة للعوامل

ويلاحظ من الشكل (2) أن الخط المنحني بدأ يغير ميله بشكل مفاجئ بين النقطتين المناظرتين للعاملين الأول والثاني، كما يلاحظ أن قيم الجذور الكامنة للعوامل بدءاً من العامل الثاني أصبحت متقاربة، مما يعد مؤشراً لاعتماد العامل الأول واستبعاد بقية العوامل.

(2) الاستقلال الموضوعي **Local Independence**: الاستقلال المحلي وأحادية البعد متشابهان، ولكنهما غير متكافئين في المفاهيم. فعند تحقق افتراض الأحادية، يتحقق افتراض الاستقلال الموضوعي ولكن العكس غير صحيح. أي يمكن أن يتحقق افتراض الاستقلال المحلي بدون بيانات ذات بعد واحد طالما تؤخذ في الحسبان جميع الجوانب التي تؤثر على نتائج الاختبار<sup>55</sup>. كما أشار علام (2005)<sup>56</sup> أن الاستقلال المحلي يعد مؤشراً لأحادية البعد إذا كان

النموذج المستعمل يحدد قيمة تقديرية لقدرة الفرد على بعد أحادي. وقد اكتفت الباحثة بالتحقق من شرط أحادية البعد للتحقق من شرطي أحادية البعد والاستقلال المحلي.

وكذلك يمكن الاستعانة بنتائج التحليل العاملي لفقرات المقياس حيث يتم فحص مصفوفة الارتباط ويقصد بها التأكد بعدم وجود مفردات في مصفوفة الارتباط لها معامل ارتباط تام أو صفر مع كل أو معظم المتغيرات، أي يتطلب أن تكون القيمة المطلقة لمحدد مصفوفة الارتباطات أكبر من 0,00001 فإذا كانت أكبر من هذه القيمة دل ذلك على عدم وجود ارتباطات مرتفعة، أو عدم وجود اعتماد خطي بين المتغيرات أو الفقرات وتكرار للمعلومات التي تشارك بها كل فقرة<sup>57</sup>، وحيث أن محدد مصفوفة الارتباطات في هذه الأداة هو (0,02215) وأعلى من القيمة المذكورة، وبهذا يمكن الحكم على عدم وجود اعتماد خطي بين فقرات المقياس مما يدل على الاستقلال المحلي لهذه الفقرات. ويمكن تأكيد ذلك كذلك باختبار "بارتليت" الذي يدل عند دلالاته أن مصفوفة الارتباطات ليست مصفوفة الوحدة Identity Matrix وتتوفر على الحد الأدنى من العلاقات.

**(3) المنحنى المميز للمفردة ICC:** الافتراض الثالث لنظرية الاستجابة للمفردة IRT هو المنحنى المميزة للمفردة (ICC). حيث تشير الزيادة المطردة للدالة المميز للمفردة إلى زيادة في احتمال نجاح الأفراد ذوي الدرجات الأعلى بالإجابة على هذه المفردة، باحتمالية أعلى من ذوي الدرجات المنخفضة على السمة<sup>58</sup>. ويشير هذا الافتراض إلى طبيعة المنحنى أو الدالة المميزة لكل مفردة، والتي تصف العلاقة بين القدرة والأداء على المفردة. ويعتمد شكل المنحنى المميز للمفردة على معالم المفردة من صعوبة وتمييز وقدرة الأفراد، حيث تتوازي منحنيات المميزة للمفردة في نموذج راش. ويتحقق هذا الافتراض من ملائمة المفردات لنموذج راش.

للتحقق من ملائمة نموذج راش يجب التأكد من:

**(1) تساوي مؤشرات التمييز Equal Discrimination Indices:** وفقاً لمقترح هاملتون وآخرون (Hambleton et al. 1991) فحص توزيع معاملات ارتباط مفردات الاختبار بالدرجة الكلية المكتسبة في الاختبار، يعطي تصوراً عن مدى تجانس مؤشرات تمييز مفردات الاختبار ومقدار تحقق افتراض تساوي معاملات التمييز، لذا قامت الباحثة بحساب الارتباطات الثنائية المتسلسلة الحقيقية Point-Biserial Correlations، وقد تم حساب هذه المعاملات باستخدام برنامج الـ Bilog-Mg3، وذلك بالطريقة التي وضحتها، والجدول (1) يوضح نتائج هذا الحساب:

الجدول (2). قيم معامل الارتباط الثنائي المتسلسل

R	I	R	I	R	I	R	I
.264**	31	.232**	21	.232**	11	.325**	1
.667**	32	.232**	22	.312**	12	.423**	2
.608**	33	.321**	23	.444**	13	.666**	3
.672**	34	.311**	24	.421**	14	.310**	4
.701**	35	.663**	25	.482**	15	.623**	5
.666**	36	.718**	26	.611**	16	.524**	6
.685**	37	.282**	27	.594**	17	.425**	7
.394**	38	.619**	28	.492**	18	.606**	8
.694**	39	.412**	29	.311**	19	.710**	9
.586**	40	.628**	30	.822**	20	.312**	10

يتضح من خلال الجدول السابق أن معاملات الارتباط المتسلسل الثنائي موجبة وذات دلالة إحصائية، وبالتالي يمكن قبول افتراض تساوي مؤشرات التمييز، وهذا ما يتوافق مع شروط نموذج راش في تساوي معامل التمييز.

(2) ملاءمة المفردات للنموذج: وقد استخدمت الباحثة برنامج Bilog-Mg3 لحساب قيمة الملاءمة لمفردات الاختبار، استبعد البرنامج المفردات (1، 2، 3، 4، 5، 6)، والجدول (2) يوضح قيم ملاءمة المفردات لنموذج راش:

الجدول (3). إحصاء الملائمة وفق نموذج راش

Sig	أحصاء الملائمة	I									
0.2096	10.9	31	0.3775	7.5	21	0.1350	12.4	11	-	-	1
0.5393	7.0	32	0.5258	5.1	22	0.5940	7.4	12	-	-	2
0.9819	2.5	33	0.0747	12.9	23	0.7882	3.9	13	-	-	3
0.1022	13.3	34	0.5442	6.0	24	0.0340	16.6	14	-	-	4
0.441	7.9	35	0.5894	6.5	25	0.1428	12.2	15	-	-	5
0.1711	11.6	36	0.6869	6.5	26	0.4154	8.2	16	-	-	6
0.4125	7.2	37	0.6475	6.0	27	0.4981	8.4	17	0.4064	7.2	7
0.2436	10.3	38	0.8797	4.4	28	0.2832	9.7	18	0.0922	13.6	8
0.0246	19.1	39	0.7962	4.6	29	0.9167	3.3	19	0.3008	10.6	9
0.1683	11.6	40	0.0120	18.0	30	0.9590	2.6	20	0.8954	3.5	10

من الجدول السابق نلاحظ أن جميع المفردات ذات قيمة الدلالة  $SIG > 0.05$  لإحصاء الملائمة، ما عدا المفردات (14، 30، 39) غير ملائمة لنموذج راش وذلك لأن  $SIG < 0.05$  وبالتالي نستبعد المفردات غير الملائمة لنحصل في النهاية على 31 مفردة ملائمة لنموذج راش.

#### 6-2- حدود البحث: تشمل حدود البحث بـ:

- اقتصرت هذه الدراسة على استخدام نموذج أحادي المعلم (نموذج راش) وذلك لأنه من النماذج ثنائية الاستجابة الأكثر استخداماً.
- اقتصرت الدراسة على تناول طريقتين للتقدير القدرة هما طريقة الأرجحية العظمى وطريقة تقدير بيبز وطريقة بروكس.
- طول الاختبار ثابت (40) مفردة ثنائية الاستجابة، تم استبعاد 9 مفردات كونها غير ملائمة لنموذج راش، وعند أربع حجوم للعينة (2000) ومجموعات عشوائية جزئية منها (1500، 1000، 500).

#### 7- الإجابة عن أسئلة الدراسة:

التساؤل الأول: ماهي تقديرات معاملات المفردة والمفردات وفق نموذج راش عند حجوم عينة (2000، 1500، 1000، 500) وفق طريقة الأرجحية العظمى وطريقة تقدير بيبز؟

للإجابة على التساؤل التالي قامت الباحثة باستخراج معاملات صعوبة المفردة وذلك باستخدام BILOG-MG3 وفق طريقة الأرجحية العظمى عند حجم عينة 2000 ومن ثم تم استخراج معاملات الصعوبة وفق طريقة تقدير بيبز ومن ثم تم اختيار عينة عشوائية بحجم 1500 فرد ومن ثم تقدير معاملات الصعوبة وفق الطريقتين، بنفس الشكل لحجم عينة 1000 و5000. الجدول (4) يبين ملخص نتائج تقديرات صعوبة المفردة ودقة التقدير:

**الجدول (4).** الإحصاءات الوصفية لمعاملات صعوبة المفردات وفق نموذج راش وفق طرائق التقدير عند حجوم عينة مختلفة

نوع القياس	N	2000		1500		1000		500	
		دقة	صعوبة	دقة	صعوبة	دقة	صعوبة	دقة	صعوبة
1	Averg	0.038	-0.240	0.0445	-0.227	0.0538	-0.242	0.076	-0.194
	STD	0.007729	1.022	0.0445	1.018	.010430	1.001	0.0169	1.021
2	Averg	0.039	-0.240	0.045	-0.227	0.054	-0.242	0.077	-0.194
	STD	.007930	1.022	0.00908	1.022	.01045	1.001	0.0165	1.020
3	Averg	0.05	-0.241	0.054	-0.290	0.059	-0.350	0.060	-0.390
	STD	0.0081	1.201	0.0083	1.501	0.0085	1.481	0.0089	1.551

نلاحظ من الجدول أن مدى تقديرات الصعوبة ودقة تقديرها عند حجم عينة (2000 و1500 و1000) أن متوسط الصعوبة متطابق في طريقة الأرجحية العظمى وتقدير بيبز إلا أن هنالك فروق صغيرة في الانحراف المعياري، إلا أنه هنالك فرق بسيط في تقدير معلم الصعوبة ودقة التقدير عند حجم عينة (500). نلاحظ في طريقة الأرجحية العظمى وطريقة تقدير بيبز أن متوسط معاملات صعوبة المفردات ومتوسط دقة تقدير يتفاوت تبعاً لتفاوت حجم العينة. تختلف نتائج هذه الدراسة عن دراسة الثوابتة (2010) التي توصلت الدراسة إلى أن زيادة حجم العينة يزيد من قيمة معلمة صعوبة الفقرة، وينقص من الخطأ المعياري في تقدير معلمة الصعوبة وفق نموذج ثلاثي المعلم.

كما أظهرت نتائج التحليل باستخدام برنامج BILOG-MG3 لمتوسطات تقديرات القدرة عند حجوم عينة مختلفة وبطرائق التقدير المعتمدة، الجدول (5) يبين الإحصاءات الوصفية لقدرة أفراد العينة ودقتها:

**الجدول (5).** الإحصاءات الوصفية لقدرة الأفراد وفق نموذج راش وفق طرائق التقدير عند حجوم عينة مختلفة

نوع القياس	N	2000		1500		1000		500	
		دقة	قدرة	دقة	قدرة	دقة	قدرة	دقة	قدرة
1	Averg	0.4086	-0.0120	0.4086	-0.0327	0.4085	-0.0149	0.4046	-0.0163
	STD	0.0954	1.1350	0.0981	1.1455	0.1010	1.1422	0.1098	1.1460
2	Averg	0.4086	-0.0120	0.3702	-0.0143	0.3550	-0.0009	0.3472	-0.0022
	STD	0.0954	1.1350	0.0854	0.9935	0.0821	0.9494	0.0863	0.9527
3	Averg	0.5021	-0.0326	0.3577	-0.0070	0.3540	-0.0149	0.3534	-0.0262
	STD	0.0041	1.1566	0.0815	0.9449	0.0828	0.9547	0.0838	0.9668

نلاحظ من الجدول أن مدى تقديرات قدرة الأفراد ودقة تقديرها عند حجم عينة (2000 و1500 و1000) أن متوسط الصعوبة متطابق في طريقة الأرجحية العظمى وتقدير بيبز إلا أن

هنالك فروق صغيرة في الانحراف المعياري وتختلف عن متوسط القدرة بطريقة بروكس، تتزايد هذه الفروق كما صغر حجم العينة.

**التساؤل الثاني: هل توجد فروق ذات دلالة إحصائية في دقة تقدير معالم المفردة (الصعوبة) بين الطرائق عند حجوم العينة المختلفة؟**

للإجابة عن هذا التساؤل، تم استخدام تحليل التباين الثنائي لمعرفة أثر كل من حجم العينة (2,000، 1,500، 1,000، 500) وطريقة التقدير (طريقة الارجحية العظمى، طريقة بايز، طريقة بروكس) والتفاعل بين هذين العاملين على تقدير معالم المفردة (الصعوبة) ودقة تقديرها وذلك باستخدام برنامج الـ SPSS كما هو موضح في الجدول التالي:

**الجدول (6). نتائج تحليل التباين الثنائي لتقدير صعوبة المفردة وفق نموذج راش ودقتها تبعاً لحجم العينة وطريقة التقدير**

القرار	SIG	F المحسوبة	متوسط المربعات	DF	مجموع المربعات	مصدر التباين	
لا يوجد	.595	.631	.914	3	2.743	حجم العينة	
لا يوجد	.916	.088	.128	2	.255	طريقة التقدير	
لا يوجد	.304	1.201	1.739	6	10.435	حجم العينة*طريقة التقدير	صعوبة المفردة
			1.448	468	677.585	الخطأ	
				480	699.803	التباين الكلي	
يوجد	.000	164.331	.018	3	.053	حجم العينة	دقة تقدير
يوجد	.001	7.249	.001	2	.002	طريقة التقدير	معلم
يوجد	.000	15.926	.002	6	.010	حجم العينة*طريقة التقدير	صعوبة المفردة
			.000	461	.049	الخطأ	
				473	1.462	التباين الكلي	

نلاحظ من الجدول السابق أن قيمة  $sig > 0.05$  في اختبار تحليل التباين لدراسة الفروق في صعوبة المفردة وفق نموذج راش وذلك في حجم العينة وطريقة التقدير والتفاعل بينهما، والذي يشير إلى عدم وجود فروق ذات دلالة إحصائية في تقدير معالم الصعوبة تعزى لحجم العينة وطريقة التقدير والتفاعل بينهما، في حين اظهر نتائج تحليل التباين الثنائي أن قيمة  $sig < 0.05$  في اختبار تحليل التباين لدراسة الفروق في دقة تقدير صعوبة المفردة وفق نموذج راش وذلك في

حجم العينة وطريقة التقدير والتفاعل بينهما، والذي يشير إلى وجود آثار ذات دلالة إحصائية في دقة تقدير معالم الصعوبة تعزى لحجم العينة وطريقة التقدير والتفاعل بينهما، ولمعرفة مصادر الفروق قامت الباحثة بإجراء اختبار توكاي لمعرفة مصدر الفروق، الجدول (7) يوضح نتائج اختبار توكاي:

الجدول (7). المقارنات الثنائية لدقة تقدير معلم صعوبة نموذج راش تبعاً لمتغيري حجم العينة وطريقة التقدير

القرار	sig	الخطأ المعياري	Mean Difference	المقارنات الثنائية		
يوجد	.001	.0013348	-.004967*	1500	2000	تباين
يوجد	.000	.0013348	-.015433*	1000		
يوجد	.000	.0013553	-.028334*	500		
يوجد	.000	.0013348	-.010467*	1000	1500	
يوجد	.000	.0013553	-.023367*	500	1000	
يوجد	.000	.0013553	.012901*	500	1000	
لا يوجد	.932	.0011560	.000413	تقدير بيبز	طريقة الأرجحية	تباين
يوجد	.010	.0011691	-.003422*	طريقة بروكس		
لا يوجد	.932	.0011560	-.000413	طريقة بروكس		

نلاحظ من الجدول السابق أن قيمة  $SIG=0.000 < 0.05$  في اغلب المقارنات الثنائية تعزى الى عامل حجم العينة، كما هو ملاحظ أن دقة التقدير في طريقتي الأرجحية العظمى وطريقة تقدير بيبز وطريقة بروكس تتزايد تبعاً لمتغير حجم العينة، أي أن دقة التقدير في كل طرائق التقدير تزداد بازدياد حجم العينة وذلك لانخفاض الخطأ المعياري بازدياد حجم العينة، الأمر الذي يتفق مع دراسة الطنطاوي (2007) والذي يشير أن زيادة حجم العينة يزيد من دقة التقدير، الأمر الذي يتطابق مع المفاهيم الأساسية في الإحصاء التي تشير أن زيادة حجم العينة يزيد من اقتراب متوسط صعوبة المفردة للمتوسط الحقيقي. تعتبر النتيجة التالية متممة لنتائج دراسة الدرايبيعي (2001) والتي استخدمت العينات (50، 100، 500) أن هنالك فروق ذات دلالة إحصائية في تقدير معلم الصعوبة تعزى لحجم العينة، دون سلوك محدد لمعلم الصعوبة، والتي تجعل من حجم عينة 500 فأكثر معيار لحجم عينة يعطي معاملات مفردة أكثر استقراراً، الأمر الذي يتفق مع متطلبات نظرية الاستجابة للمفردة حجوم عينة كبيرة للحصول على تقديرات لمعلم الصعوبة أفضل، الأمر الذي كان سبب بتباين قيم معاملات الصعوبة والذي أثر بدوره على النتائج تبعاً لحجم العينة. كما أن استقرار معاملات الصعوبة، يعتمد على حجم وبنية المجموعات التي تطبق عليها، وبناءً على ذلك لا يجوز اعتبار معاملات الصعوبة صفات ثابتة لا تتغير.<sup>59</sup>

كما أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية في دقة التقدير بين طريقة بروكس وطريقة تقدير بيبز كما لا يوجد فروق ذات دلالة وطريقة الأرجحية وطريقة تقدير بيبز ووجود فروق ذات دلالة إحصائية في دقة تقدير صعوبة المفردة وفق نموذج راش. في حين أظهرت النتائج وجود فروق ذات دلالة إحصائية بين طريقة الأرجحية العظمى وطريقة بروكس.

## 6-2- هل توجد فروق ذات دلالة إحصائية في تقدير قدرة الأفراد بين الطرائق عند حجم العينة المختلفة؟

للإجابة عن هذا التساؤل، تم استخدام تحليل التباين الثنائي لمعرفة أثر كل من حجم العينة (2000، 1500، 1000، 500) وطريقة التقدير (طريقة الأرجحية العظمى، طريقة بيبز، طريقة بروكس) والتفاعل بين هذين العاملين على تقدير قدرة الأفراد ودقة تقديرها وذلك باستخدام برنامج الـ SPSS كما هو موضح في الجدول التالي:

**الجدول (8).** نتائج تحليل التباين الثنائي لتقدير قدرة الأفراد وفق نموذج راش ودقتها تبعاً لحجم العينة وطريقة التقدير

القرار	SIG	F المحسوبة	متوسط المربعات	DF	مجموع المربعات	مصدر التباين	
لا يوجد	.986	.049	.057	3	.170	حجم العينة	قدرة الأفراد
لا يوجد	.847	.166	.192	2	.383	طريقة التقدير	
لا يوجد	.990	.143	.165	6	.988	حجم العينة*طريقة التقدير	
			1.154	14970	17279.548	الخطأ	
				14982	17285.180	التباين الكلي	
يوجد	.000	701.093	5.083	3	15.248	حجم العينة	دقة التقدير
يوجد	.000	185.778	1.347	2	2.694	طريقة التقدير	
يوجد	.000	307.440	2.229	6	13.373	حجم العينة*طريقة التقدير	
			.007	14970	108.528	الخطأ	
				14982	2548.732	التباين الكلي	

نلاحظ من الجدول السابق أن قيمة  $sig > 0.05$  في اختبار تحليل التباين لدراسة الفروق في تقدير قدرة الأفراد وفق نموذج راش وذلك في حجم العينة وطريقة التقدير والتفاعل بينهما، والذي يشير إلى عدم وجود فروق ذات دلالة إحصائية في متوسط قدرة الأفراد تعزى لحجم العينة وطريقة التقدير والتفاعل بينهما، في حين اظهر نتائج تحليل التباين الثنائي أن قيمة  $sig < 0.05$  في اختبار تحليل التباين لدراسة الفروق في دقة تقدير قدرة الأفراد وفق نموذج راش وذلك في

حجم العينة وطريقة التقدير والتفاعل بينهما، والذي يشير إلى وجود آثار ذات دلالة إحصائية في دقة تقدير معالم الصعوبة تعزى لحجم العينة وطريقة التقدير والتفاعل بينهما، ولمعرفة مصادر الفروق قامت الباحثة بإجراء اختبار توكاي لمعرفة مصدر الفروق، الجدول (9) يوضح نتائج اختبار توكاي:

الجدول (9). المقارنات الثنائية لدقة تقدير قدرة الأفراد وفق نموذج راش تبعاً لمتغيري حجم العينة وطريقة التقدير

القرار	sig	الخطأ المعياري	Mean Difference	المقارنات الثنائية		
				1500	2000	
يوجد	.000	.0016793	.060985*	1500	2000	نوع العينة
يوجد	.000	.0019068	.067259*	1000		
يوجد	.000	.0024588	.071441*	500		
يوجد	.010	.0020092	.006273*	1000	1500	نوع الطريقة
يوجد	.000	.0025391	.010456*	500		
لا يوجد	.406	.0026949	-.004182	500	1000	
يوجد	.000	.0017035	.027965*	تقدير بيبز	طريقة الارحجية	نوع الطريقة
يوجد	.001	.0017043	-.006240*	طريقة بروكس		
يوجد	.000	.0017040	.034205*	طريقة بروكس	طريقة بيبز	

أظهرت النتائج وجود فروق ذات دلالة إحصائية في دقة تقدير معالم القدرة تعزى لعامل القدرة عند جميع مستويات حجوم العينة وذلك بتناقص قيمة الخطأ المعياري بازدياد حجم العينة، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية في دقة تقدير معلم القدرة يعزى لعامل طريقة التقدير. الأمر الذي يؤكد على أهمية حجم العينة بوصفه عاملاً مؤثراً على دقة القياس وموضوعيته، وهذا يتفق مع ما أشار إليه لندن وجلاس (Linden & Glass, 2002) من أن قيم الخطأ المعياري تختلف باختلاف عينة التدريج وحجم تجمع الفقرات، وكذلك تتفق نتائج هذه الدراسة مع دراسة فاريش وستيف (Farish & Stephen , 1984) إلى ضرورة زيادة حجم عينة الدراسة لنموذج راش للحصول على تقديرات أدق.

**التوصيات والمقترحات:**

1. زيادة حجم العينة يؤدي إلا تقديرات دقيقة لمعاملات المفردة وفق نموذج راش وبالتالي يمكن اعتبار زيادة حجم العينة عامل إيجابي في دقة النتائج.
2. في حال تناولنا دراسات تهتم بتقدير معاملات المفردة والافراد وفق نموذج راش فإن استخدام أحد طرائق التقدير (الأرجحية العظمى، طريقة تقدير بيبز، طريقة بروكس) ستعطي نتائج متكافئة وذلك عند أحجام عينة أكبر من 500 فرد.
3. في حال كان الهدف من الدراسة دراسة دقة التقدير فإن زيادة حجم العينة يزيد الدقة وخفض الخطأ المعياري المرتكب.
4. تتشابه طريقة بيبز والارجحية العظمى من جهة وطريقة بيبز وطريقة بروكس في دقة تقدير معالم الصعوبة في حين تتباين في دقة تقدير القدرة.
5. دراسة أثر متغيرات أخرى في دقة تقدير المفردات.
6. العمل على دراسة تنبؤيه لأثر حجم العينة على دقة تقدير معاملات المفردة والأفراد.

**الهوامش**

- 1 علام، صلاح الدين. (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. دار الفكر العربي، القاهرة. ص54.
- 2 الشرقاوى، أنور. الشيخ، سليمان. كاظم، أمينة. عبد السلام، نادية. (1996): اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة، الانجلو المصرية. ص309.
- 3 زكري، علي بن محمد عبد الله (2009). الخصائص السيكمترية لاختبار أوتيس - لينون للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صيبا التعليمية، أطروحة دكتوراه غير منشورة، جامعة أم القرى، كلية التربية. ص4.
- 4 بست، جون.و. (1988). مناهج البحث التربوي، ترجمة عبد العزيز غانم الغانم، ط1، سلسلة الكتب المترجمة، الكويت. ص37.
- 5 المياحي، أمل إسماعيل عايز. (2005). أثر اختلاف حجم العينة وطول المقياس في الخصائص السيكمترية لمقاييس الشخصية، اطروحة دكتوراه غير منشورة، جامعة بغداد كلية التربية ابن رشد. ص9.
- 6 الغامدي، عبد اللطيف بن محمد (2004). اثر اسلوب اختيار العينة وحجمها على دقة تقدير معالم المجتمع الاحصائي، دراسة منشورة، المملكة العربية السعودية، جامعة ام القرى، كلية التربية. ص 12-13.

- 7 أبو هشام، السيد محمد (2006). دراسة مقارنة بين النظرية التقليدية وانموذج راش في اختبار فقرات مقياس مدخل الدراسة لدى طلاب الجامعة، دراسة منشورة، جامعة الزقازيق، كلية التربية.
- 8 Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika*. 46,443-459.
- 9 Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-parameter logistic model. *Applied Psychological Measurement*. 13. 77-90.
- 10 Seong, T.-J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*. 14. 299-311.
- 11 Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement* 16.1-16.
- 12 Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*. 52. 275-291.
- 13 العبد الله، زياد. (2012). أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدرج. أطروحة دكتوراه غير منشورة. معهد الدراسات التربوية. جامعة القاهرة: مصر.
- 14 العضايلة، بسام. (2016). المقارنة بين طريقتي الأرجحية العظمى وبييز في دقة تقدير معالم الافراد والعناتب الفارقة باستخدام نموذج راش للتقدير المتدرج. أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردنية: الأردن.
- 15 Farish, S. J. (1984). Investigating Item Stability: An Empirical Investigation into the Variability of Item Statistical Under of Varying Sample Design and Sample Size. Occasional Paper No. 18. Condition Publication Type: 143; 110, Australia
- 16 Hambleton, R. K, Cook. L. L. (1983). Robutness of Item Response Models and Effects of Test Length and Sample Size in the Precision of Ability Estimates. New York. In D. J. Weiss (Ed), *New Horizons in Testing*. Pp.31-49
- 17 Ree, M. J. Jensen, H. E. (1983). Effects of Sample size on Linear Equating of Item Characteristics Curve Parameters. In D. J. Weiss (Ed). *New Horizons in Testing*. Pp. (135-146).
- 18 الصرايرة، راجي. (2008). مقارنة ثلاثة إحصائيات لفحص جودة مطابقة الفقرة تحت ظروف نوع النموذج الرياضي وطول الاختبار وحجم العينة والتفاعلات بينهما. أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردني: الأردن.
- 19 المدانات، رائد. (2008). أثر طريقة المعادلة باستخدام جذع مشترك وعدد فقراته وحجم العينة على القيم المعادلة والخطأ في المعادلة بين صورتين اختبار في الفيزياء. أطروحة دكتوراه غير منشورة. كلية الدراسات التربوية العليا. جامعة عمان. الأردن.

- 20 القحطاني، سعيد. (2011). مدى اقتران تحقق فحص الاستقلال المحلي بتحقق أحادية البعد في نظرية استجابة الفقرة تحت ظروف تباين حجم العينة وطول الاختبار. رسالة ماجستير غير منشورة. كلية العلوم التربوية. جامعة مؤتة: الأردن.
- 21 درويش، هيثم. (2014). دقة تقدير معالم الفقرة باستخدام نماذج نظرية الاستجابة للفقرة البارامترية المعدلة واللابارمترية باختلاف حجم العينة وطول الاختبار (دراسة مقارنة). أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردنية: الأردن.
- 22 العبيسات، أيمن. (2015). المقارنة بين صيغتي روبن وجيلفورد لتصحيح معامل ارتباط بايسيريال من التضخم تحت ظرف تباين حجم العينة وعدد البدائل الفقرة. رسالة ماجستير غير منشورة. كلية الدراسات العليا. جامعة مؤتة: الأردن.
- 23 الصرايرة، اخلاص. (2017). المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة. رسالة ماجستير غير منشورة. كلية الدراسات العليا. جامعة مؤتة: الأردن.

25 Fan ,X.(1998). \_Item Response Theory And Classical Test Theory: An Empirical Comparison Of Their Item/Person Statistics, Educational And Psychological Measurement, June 1998, Vol58n3,p357(25).

- 26 عبد الجبار، عادل صلاح عمر، والشافعي، محمد منصور محمد (2006): اثر حجم عينة التحليل على مؤشرات الملازمة الاحصائية وتقديرات الصعوبة المتضمنة ببرنامجي (مايكروسكال) و(بابيلوك) للمفردات باستخدام انموذج راش " دراسة محاكاة". رسالة التربية وعلم النفس.
- 27 الزهراني، بندر بن حمدان (2008). أثر اختلاف حجم العينة واتساع مدى القدرة على دقة تقدير الدرجة الحقيقية المقدرة باستخدام النظرية التقليدية والنماذج أحادية البعد في النظرية الحديثة القياس. اطروحة دكتوراه غير منشورة، القياس والتقويم، جامعة أم القرى، المملكة العربية السعودية.
- 28 العبد الله، زياد. (2012). أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدريج. أطروحة دكتوراه غير منشورة. معهد الدراسات التربوية. جامعة القاهرة: مصر. ص8.
- 29 الرحيل، راتب. الدرابسة، رياض. (2014). أثر طريقتي التعامل مع القيم المفقودة وطريقة تقدير القدرة على دقة تقدير معالم الفقرات والأفراد. المجلة الدولية التربوية المتخصصة. المجلد 3. العدد 6. ص 23-47. ص28.
- 30 علام، صلاح الدين. (2000). القياس والتقويم التربوي والنفسي، أساسياته وتطبيقاته وتوجيهاته المعاصرة، ط1، القاهرة، دار الفكر العربي. ص683.

- 31 Cantrell, C. (1997). Item Response Theory: Understanding The One-Parameter Rasch Model. Paper Presented At The Annual Meeting Of The Southwest Educational Research Association (Austin, TX, January 23).
- 32 Henson, R. (1999). Understanding The One-Parameter Rasch Model Of Item Response Theory. Paper Presented At The Annual Meeting Of The Southwest Educational Research Association (San Antonio, TX, January 21-23).
- 33 Rust, J. & Golombok, S. (1999). Modern Psychometrics: The Science Of Psychological Assessment. (2nd ed), U.S.A, Routledge (UK).
- 34 النبهان، موسى. (2004). أساسيات القياس في العلوم السلوكية. عمان، دار الشروق للنشر والتوزيع، ط (1). ص 6
- 35 de Gruijter, D. N. M. & van der Kamp, L. J. Th. (2005). Statistical Test Theory for Education and Psychology. © D. N. M. de Gruijter & L. J. Th. Van der Kamp.
- 36 Hambleton, R. K. & Swaminathan, H. (1985). Item Response Theory, Principles and Applications. Bstn: Kluwer. Nijhoff Publishing amembers of the Kluwer Academic Publishers Group.
- 37 Mislevy, R. T. (1990): Modeling Item Response When Different Subjects Employ Different Solations Strategies, Netherlands: Psychometrica, Vol. 55, No., PP. 195-215.
- 38 علام، صلاح الدين. (2001). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية. القاهرة: دار الفكر العربي.
- 39 علام، صلاح الدين. (2000). القياس والتقويم التربوي والنفسي، أساسياته وتطبيقاته وتوجيهاته المعاصرة، ط1، القاهرة، دار الفكر العربي. ص683.
- 40 Baker, Frank, B. (2001): The Basics of item responses Theory, 2ed, ERIC Clearinghouse on Assessment and Evaluation. P25.
- 41 النقي، احمد محمد. (2009). النظرية الحديثة في القياس. ط1. دار المسيرة، عمان. ص18-19.
- 42 سرحان، محمد. (2018). المقارنة بين طريقتي الأرجحية العظمى والقيمة العظمى للتوزيع البعدي في تقدير القدرة وفقاً لنظرية الاستجابة للفقرة. أطروحة دكتوراه غير منشورة. كلية التربية ابن رشد للعلوم الإنسانية. جامعة بغداد.
- 43 العبد الله، زياد. (2012). أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدرج. أطروحة دكتوراه غير منشورة. معهد الدراسات التربوية. جامعة القاهرة: مصر.
- 44 WML طريقة وورم للأرجحية العظمى الموزونة.
- 45 EAP طريقة التقدير القائمة على نظرية بيز.
- 46 MLE طريقة التقدير القائمة على طريقة الأرجحية العظمى.

- 47 Garre. G. & Vermunt. K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation. *Behaviormentrika*. Vol. 33. No. 1.
- 48 Wang, T, Vispoel. W. (1998). Properties of Ability Estimation Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*, Vol. 35, Issue 22.
- 49 Kim, S. (2001). An Evaluation of A Markov Chain Monte Carlo Method for the Rasch Model. *Applied Psychological Measurement*, Vol. 25, No. 2, 163-176.
- 50 ملحم، سامي محمد. (2010): *مناهج البحث في التربية وعلم النفس*، ط6. دار المسيرة للنشر والتوزيع والطباعة، عمان - الاردن. ص404.
- 51 العساف، صالح بن حمد. (2010). *المدخل الى البحث في العلوم السلوكية*. ط1، الرياض، دار الزهراء. ص279
- 52 التل، سعيد. البطش، محمد وليد. ابو زينة، فريد كامل. (2007). *مناهج البحث العلمي - تصميم البحث والتحليل الإحصائي*. ط1، عمان، دار المسيرة للنشر والتوزيع والطباعة. ص251.
- 53 Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- 54 يعقوب، ابراهيم محمد. (1990). *دراسة مقارنة للخصائص السيكومترية لمقياس مفهوم الذات المبني بالطريقة التقليدية طريقة نموذج راش*. اطروحة دكتوراه غير منشورة، الجامعة الاردنية. ص66.
- 55 Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*; ISSN 2298-0172. P26.
- 56 علام، صلاح الدين. (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. دار الفكر العربي، القاهرة. ص64.
- 57 تيفزة، أمحمد بوزيان. (2012). *التحليل العاملي الاستكشافي والتوكيدي مفاهيمها ومنهجيتها بتوظيف حزمة SPSS وليزرل LISREL*. دار المسيرة للنشر والتوزيع والطباعة، عمان، الأردن. ص89.
- 58 Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*; ISSN 2298-0172. P26.
- 59 مخائيل، أمطانيوس. (2009). *القياس والتقويم في التربية الحديثة*. دمشق: منشورات جامعة دمشق. ص343.

## قائمة المصادر والمراجع:

## المراجع العربية:

- أبو هشام، السيد محمد. (2006). دراسة مقارنة بين النظرية التقليدية وانموذج راش في اختبار فقرات مقياس مدخل الدراسة لدى طلاب الجامعة، دراسة منشورة، جامعة الزقازيق، كلية التربية.
- بست، جون.و. (1988). *مناهج البحث التربوي*، ترجمة عبد العزيز غانم الغانم، ط1، سلسلة الكتب المترجمة، الكويت.
- بني عطا، زايد. (2013). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة: دراسة محاكاة. *مجلة جامعة الشارقة*. المجلد 11. العدد 2.
- بني عطا، زايد. الشرفين، نضال. (2012). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات. *المجلة الاردنية في العلوم التربوية*. مجلد 8.
- التقي، احمد محمد. (2009). *النظرية الحديثة في القياس*. ط1. دار المسيرة، عمان.
- الثل، سعيد. البطش، محمد وليد. ابو زينة، فريد كامل. (2007). *مناهج البحث العلمي - تصميم البحث والتحليل الإحصائي*. ط1، عمان، دار المسيرة للنشر والتوزيع والطباعة.
- تيغزة، أمحمد بوزيان. (2012). *التحليل العاملي الاستكشافي والتوكيدي مفاهيمها ومنهجيتها بتوظيف حزمة SPSS و ليزرل LISREL*. دار المسيرة للنشر والتوزيع والطباعة، عمان، الأردن.
- جبران، نبيل. (2017). أثر حجم العينة وطول الاختبار ونوع الفقرة في دقة تقدير معالم الفقرة والقدرة ودالة معلومات الاختبار. أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردنية. الجامعة الأردنية: الأردن.
- درويش، هيثم. (2014). دقة تقدير معالم الفقرة باستخدام نماذج نظرية الاستجابة للفقرة البارامترية المعدلة واللابارمترية باختلاف حجم العينة وطول الاختبار (دراسة مقارنة). أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردنية: الأردن.

الرحيل، راتب. الدرابسة، رياض. (2014). أثر طريقتي التعامل مع القيم المفقودة وطريقة تقدير القدرة على دقة تقدير معالم الفقرات والأفراد. المجلة الدولية التربوية المتخصصة. المجلد 3. العدد 6. ص 23-47.

الزهراني، بندر بن حمدان. (2008). أثر اختلاف حجم العينة واتساع مدى القدرة على دقة تقدير الدرجة الحقيقية المقدرة باستخدام النظرية التقليدية والنماذج أحادية البعد في النظرية الحديثة القياس. أطروحة دكتوراه غير منشورة، القياس والتقييم، جامعة أم القرى، المملكة العربية السعودية.

سرحان، محمد. (2018). المقارنة بين طريقتي الأرجحية العظمى والقيمة العظمى للتوزيع البعدي في تقدير القدرة وفقاً لنظرية الاستجابة للفقر. أطروحة دكتوراه غير منشورة. كلية التربية ابن رشد للعلوم الإنسانية. جامعة بغداد.

الشرقاوى، أنور. الشيخ، سليمان. كاظم، أمينة. عبد السلام، نادية. (1996). اتجاهات معاصرة في القياس والتقييم النفسي والتربوي، القاهرة، الانجلو المصرية.

الصرايرة، اخلاص. (2017). المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة. رسالة ماجستير غير منشورة. كلية الدراسات العليا. جامعة مؤتة: الأردن.

الصرايرة، راجي. (2008). مقارنة ثلاثة إحصائيات لفحص جودة مطابقة الفقرة تحت ظروف نوع النموذج الرياضي وطول الاختبار وحجم العينة والتفاعلات بينهما. أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردنية: الأردن.

الطراونة، نادية. (2017). أثر حجم العينة في تقدير معامل الثبات في ضوء نظرية التعميم. رسالة ماجستير غير منشورة. كلية الدراسات العليا. جامعة مؤتة: الأردن.

طنطاوي، الشيماء. (2007). أثر كل من حجم العينة وطول الاختبار على دقة نتائج الاختبارات مرجعية المحك في الرياضيات باستخدام نموذج راش. رسالة ماجستير غير منشورة. كلية التربية. جامعة الزقازيق.

عبد الجبار، عادل صلاح عمر، والشافعي، محمد منصور محمد. (2006). اثر حجم عينة التحليل على مؤشرات الملائمة الاحصائية وتقديرات الصعوبة المتضمنة ببرنامجي

(مايكروسكال) و(بايلوك) للمفردات باستخدام انموذج راش "دراسة محاكاة". رسالة التربية وعلم النفس.

العبد الله، زياد. (2012). أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدريج. أطروحة دكتوراه غير منشورة. معهد الدراسات التربوية. جامعة القاهرة: مصر.

العبيسات، أيمن. (2015). المقارنة بين صيغتي روبن وجيلفورد لتصحيح معامل ارتباط بايسيريال من التضخم تحت ظرف تباين حجم العينة وعدد البدائل الفقرة. رسالة ماجستير غير منشورة. كلية الدراسات العليا. جامعة مؤتة: الأردن.

العساف، صالح بن حمد. (2010). المدخل الى البحث في العلوم السلوكية. ط1. الرياض، دار الزهراء.

العضيلة، بسام. (2016). المقارنة بين طريقتي الأرجحية العظمى وبييز في دقة تقدير معالم الافراد والعتبات الفارقة باستخدام نموذج راش للتقدير المتدرج. أطروحة دكتوراه غير منشورة. كلية الدراسات العليا. الجامعة الأردنية: الأردن.

علام، صلاح الدين. (2000). القياس والتقويم التربوي والنفسي، أساسياته وتطبيقاته وتوجيهاته المعاصرة، ط1، القاهرة، دار الفكر العربي.

علام، صلاح الدين. (2001). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية، القاهرة: دار الفكر العربي.

علام، صلاح الدين. (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. دار الفكر العربي، القاهرة.

الغامدي، عبد اللطيف بن محمد. (2004). اثر اسلوب اختيار العينة وحجمها على دقة تقدير معالم المجتمع الاحصائي، دراسة منشورة، المملكة العربية السعودية، جامعة ام القرى، كلية التربية.

القحطاني، سعيد. (2011). مدى اقتران تحقق فحص الاستقلال المحلي بتحقق أحادية البعد في نظرية استجابة الفقرة تحت ظروف تباين حجم العينة وطول الاختبار. رسالة ماجستير غير منشورة. كلية العلوم التربوية. جامعة مؤتة: الأردن.

مخائيل، أمطانيوس. (2009). القياس والتقويم في التربية الحديثة. دمشق: منشورات جامعة دمشق.

المدان، رائد. (2008). أثر طريقة المعادلة باستخدام جذع مشترك وعدد فقراته وحجم العينة على القيم المعادلة والخطأ في المعادلة بين صورتين اختبار في الفيزياء. أطروحة دكتوراه غير منشورة. كلية الدراسات التربوية العليا. جامعة عمان. الأردن.

ملحم، سامي محمد. (2010). مناهج البحث في التربية وعلم النفس، ط6، دار المسيرة للنشر والتوزيع والطباعة، عمان - الاردن.

المياحي، أمل إسماعيل عايز. (2005). أثر اختلاف حجم العينة وطول المقياس في الخصائص السيكومترية لمقاييس الشخصية، اطروحة دكتوراه غير منشورة، جامعة بغداد كلية التربية ابن رشد.

النبهان، موسى. (2004). أساسيات القياس في العلوم السلوكية. عمان، دار الشروق للنشر والتوزيع، ط (1).

يعقوب، ابراهيم محمد. (1990). دراسة مقارنة للخصائص السيكومترية لمقياس مفهوم الذات المبني بالطريقة التقليدية طريقة نموذج راش. اطروحة دكتوراه غير منشورة، الجامعة الاردنية.

#### المراجع الإنجليزية:

- Baker, Frank, B. (2001). *The Basics of item responses Theory*, 2ed, ERIC Clearinghouse on Assessment and Evaluation.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika*.46,443-459.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459.
- Cantrell, C. (1997). Item Response Theory: Understanding The One-Parameter Rasch Model. *Paper Presented At The Annual Meeting Of The Southwest Educational Research Association (Austin, TX, January 23)*.

- Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-parameter logistic model. *Applied Psychological Measurement*, 13, 77-90.
- Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*; ISSN 2298-0172.
- Farish, S. J. (1984). Investigating Item Stability: An Empirical Investigation into the Variability of Item Statistics Under of Varying Sample Design and Sample Size. *Occasional Paper No. 18. Condition Publication Type: 143; 110, Australia*.
- Farish, S. J. (1984). Investigating item stability: An empirical investigation into the variability of item statistics under conditions of varying sample design and sample size. *Occasional paper, No. 18. Publication type: 143; 110, Australia*. [on-line] Available: <http://eric.ed.gov>
- Farish, Stephen.J. (1984). Investigating item stability: An empirical investigation into the variability of item statistics under conditions of varying sample design and sample size. *Occasional paper No. 18. publication type: 143;110,Australia*. <http://eric.ed.gov>[Available on-line].
- Garre. G. & Vermunt. K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation. *Behaviormentrika*. Vol. 33. No. 1.
- Hambleton, R. K. & Swaminathan, H. (1985). Item Response Theory, Principles and Applications. Bston: Kluwer. *Nijhoff Publishing amembers of the Kluwer Academic Publishers Group*.
- Hambleton, R. K, Cook. L. L. (1983). Robutness of Item Response Models and Effects of Test Length and Sample Size in the Precision of Ability Estimates. New York. In D. J. Weiss (Ed), *New Horizons in Testing*. Pp.31-49.
- Hambleton, R. K., & Cook, L. L. (1983). Robustness of item response models and effects of test length and sample size on the precision of ability estimates. In *New horizons in testing* (pp. 31-49).
- Hambleton, R.K.; Swaminthan, H.; Rogers, h.j. (1991). *Fundamentals Of Item Response Theory*.Newbury Park,Ca:Sage. The International Professional Publishers.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Henson, R. (1999). Understanding The One-Parameter Rasch Model Of Item Response Theory. *Paper Presented At The Annual Meeting Of The*

- Southwest Educational Research Association (San Antonio, TX, January 21-23).*
- Kim, S. (2001). An Evaluation of A Markov Chain Monte Carlo Method for the Rasch Model. *Applied Psychological Measurement*, Vol. 25, No. 2, 163-176.
- Linden , W. and Glas ,C. (2002) *Computerized adaptive testing: theory and practice*: Lluwer Academic Publishers.
- Lord, F. Novick, M. (1986). *Statistical Theories of Mental Test Scores*. Reading Mass: Addison Wisely.
- Mislevy, R. T. (1990). Modeling Item Response When Different Subjects Employ Different Solations Strategies, *Netherlands: Psychometrica*, Vol. 55, No., PP. 195-215.
- Pelton, W. (2002). *The Accuracy of Unidimensional Measurement Models in the Presence of Deviations form the Underlying Assumptions*. Unpublished Doctoral Dissertation. Brigham Young University. USA.
- Ree, M. J. Jensen, H. E. (1983). Effects of Sample size on Linear Equating of Item Characteristics Curve Parameters. In D. J. Weiss (Ed). *New Horizons in Testing*. Pp. (135-146).
- Rust, J. & Golombok, S. (1999). *Modern Psychometrics: The Science Of Psychological Assessment*. (2nd ed), U.S.A, Routledge (UK).
- Seong, T.-J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*. 14. 299-311.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement* 16.1-16.
- Tiot, M.D., (2003), *IRT From SSI: BILOG\_MG, MULTILOG*.
- Wang, T, Vispoel. W. (1998). Properties of Ability Estimation Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*, Vol. 35, Issue 22.
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*. 52. 275-291.

**List of Sources & References:**

- Abdel-Jabbar, Adel Salah Omar, and Al-Shafi'i, Mohamed Mansour Mohamed. (2006). *The effect of the size of the analysis sample on the indicators of statistical relevance and difficulty estimates included in the (Microscale) and (Baylock) programs for vocabulary using the Rush model "simulation study"*. Education and Psychology Thesis.
- Abu Hisham, Mr. Muhammad .(2006). *A comparative study between the traditional theory and the Rush model in examining the paragraphs of the study entry scale for university students*, published study, Zagazig University, Faculty of Education.
- Al Sarayrah, Ikhlas. (2017). *The comparison between the marginal method and the correlative methods in calculating the vertebrae discrimination in light of the variance of the test length and the sample size*. A magister thesis that is not published. College of Graduate Studies. Mutah University: Jordan.
- Al Sarayrah, Raji. (2008). *Comparison of three statistics for examining the quality of the vertebra matching under the conditions of the mathematical model type, the length of the test, the sample size, and the interactions between them*. Unpublished doctoral thesis. College of Graduate Studies. The University of Jordan: Jordan.
- Al Sharkawy, Anwar. Sheikh, Solomon. Kazem, Amina. Abdul Salam, Nadia. (1996). *Contemporary trends in measurement, psychological and educational evaluation*, Cairo, Anglo-Egyptian.
- Al-Abdullah, Ziyad. (2012). *The effect of some estimation methods on the accuracy of estimating features within the multi-scalar single response models*. Unpublished doctoral thesis. Institute of Educational Studies. Cairo University: Egypt.
- Al-Adayleh, Bassam. (2016). *Comparison of the two highest odds and Bayes methods of accurate estimation of individual parameters and differential thresholds using the Rush gradient model*. Unpublished doctoral thesis. College of Graduate Studies. The University of Jordan: Jordan.
- Al-Assaf, Saleh bin Hamad. (2010). *Introduction to research in behavioral sciences*. 1<sup>st</sup> Floor, Riyadh, Dar Al-Zahraa.
- Al-Ghamdi, Abd al-Latif bin Muhammad. (2004). *The effect of the sample selection method and its size on the accuracy of estimating the parameters*

- of the statistical community*, published study, Kingdom of Saudi Arabia, Umm Al-Qura University, College of Education.
- Allam, Salahuddin. (2000). *Educational and psychological measurement and evaluation, its basics*, applications and contemporary directives, 1<sup>st</sup> floor, Cairo, Dar Al-Fikr Al-Arabi.
- Allam, Salahuddin. (2001). *Diagnostic tests benchmark reference in the educational and psychological areas*, Cairo: Dar Al-Fikr Al-Arabi.
- Allam, Salahuddin. (2005). *Single-dimensional and multi-dimensional experimental response models and their applications in psychological and educational measurement*. Arab Thought House, Cairo.
- Al-Mayahi, Amal Ismail Ayez. (2005). *The effect of the difference in the sample size and the length of the scale on the psychometric properties of personality measures*, unpublished doctoral thesis, University of Baghdad, College of Education, Ibn Rushd.
- Al-Obeisat, Ayman. (2015). *Comparison of the Robin and Guilford formulas for correction of the BCIER correlation coefficient of inflation under the circumstance of sample size variance and the number of vertebrae variants*. A magister Thesis that is not published. College of Graduate Studies. Mutah University: Jordan.
- Al-Qahtani, Saeed. (2011). *The extent of the association of the verification of local independence with a one-dimensional verification in the paragraph response theory under the circumstances of the sample size variance and the length of the test*. A magister Thesis that is not published. College of Educational Sciences. Mutah University: Jordan.
- Al-Zahrani, Bandar bin Hamdan (2008). *The effect of differing sample size and breadth of ability to accurately estimate the estimated real score using conventional theory and one-dimensional models in modern measurement theory*. Unpublished PhD thesis, Measurement and Evaluation, Umm Al-Qura University, Kingdom of Saudi Arabia.
- Baker, Frank, B. (2001). *The Basics of item responses Theory*, 2ed, ERIC Clearinghouse on Assessment and Evaluation.
- Bani Ata, Zayed. (2013). Investigating the accuracy of estimating the three-parameter logistic model of the parameters of the vertebra and the ability of

- individuals in light of changing test length and sample size: a simulated study. *University of Sharjah Journal*. Volume 11. Issue 2.
- Bani Ata, Zayed. The two honorable people, Nidal. (2012). The effect of the different form of power distribution on the parameters of the paragraph and the information function. *The Jordanian Journal of Educational Sciences*. Volume 8.
- Best, John W. (1988). *Educational Research Methods*, translated by Abdulaziz Al-Ghanem Al-Ghanem, First Edition, Series of Translated Books, Kuwait.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459.
- Cantrell, C. (1997). Item Response Theory: Understanding the One-Parameter Rasch Model. *Paper Presented At The Annual Meeting Of The Southwest Educational Research Association (Austin, TX, January 23)*.
- Convictions, pioneer. (2008). *The effect of the equation method using a common stem, the number of paragraphs, and the sample size on the equation values and the error in the equation between two test images in physics*. Unpublished doctoral thesis. College of Higher Education Studies. Amman University. Jordan.
- Darwish, Haytham. (2014). *The accuracy of estimating the parameters of the paragraph using models of the response theory of the modified Parametric Paragraph with different sample size and test length (comparative study)*. Unpublished doctoral thesis. College of Graduate Studies. The University of Jordan: Jordan.
- de Gruijter, D. N. M. & van der Kamp, L. J. Th. (2005). *Statistical Test Theory for Education and Psychology*. © D. N. M. de Gruijter & L. J. Th. Van der Kamp.
- Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-parameter logistic model. *Applied Psychological Measurement*, 13. 77-90.
- Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*; ISSN 2298-0172.

- Fan ,X.(1998). Item Response Theory And Classical Test Theory: An Empirical Comparison Of Their Item/Person Statistics, *Educational And Psychological Measurement*, June 1998, Vol58 n3, p357(25).
- Farish, S. J. (1984). Investigating Item Stability: An Empirical Investigation into the Variability of Item Statistical Under of Varying Sample Design and Sample Size. *Occasional Paper No. 18. Condition Publication Type: 143; 110, Australia.*
- Garre. G. & Vermunt. K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation. *Behaviormentrika*. Vol. 33. No. 1.
- Gibran, Nabil. (2017). *The effect of sample size, test length, and paragraph type on the accuracy of estimating the parameters of the paragraph, ability and function of the test information*. Unpublished doctoral thesis. College of Graduate Studies. University of Jordan. The University of Jordan: Jordan.
- Hambleton, R. K. & Swaminathan, H. (1985). Item Response Theory, Principles and Applications. Bstom: Kluwer. *Nijhoff Publishing amembers of the Kluwer Academic Publishers Group.*
- Hambleton, R. K, Cook. L. L. (1983). Robutness of Item Response Models and Effects of Test Length and Sample Size in the Precision of Ability Estimates. New York. In D. J. Weiss (Ed), *New Horizons in Testing*. Pp.31-49.
- Hambleton, R. K., & Cook, L. L. (1983). Robustness of item response models and effects of test length and sample size on the precision of ability estimates. In *New horizons in testing* (pp. 31-49).
- Hambleton, R.K.; Swaminthan, H.; Rogers, h.j. (1991). *Fundamentals Of Item Response Theory*. Newbury Park, Ca: Sage. The International Professional Publishers.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Henson, R. (1999). Understanding The One-Parameter Rasch Model Of Item Response Theory. *Paper Presented At The Annual Meeting Of The Southwest Educational Research Association (San Antonio, TX, January 21-23).*

- Kim, S. (2001). An Evaluation of A Markov Chain Monte Carlo Method for the Rasch Model. *Applied Psychological Measurement*, Vol. 25, No. 2, 163-176.
- Leave, salary. Al-Darabsa, riad. (2014). The effect of the two methods of dealing with missing values and the method of estimating the ability to accurately estimate the parameters of paragraphs and individuals. *International Journal of Specialized Education*. Volume 3. Issue 6. pp. 23-47.
- Linden , W. and Glas ,C. (2002). *Computerized adaptive testing: theory and practice*: Lluwer Academic Publishers.
- Lord, F. Novick, M. (1986). *Statistical Theories of Mental Test Scores*. Reading Mass: Addison Wisely.
- Meet, Ahmed Mohamed. (2009). *Modern theory of measurement*. I 1. Dar Al Masirah, Amman.
- Melhem, Sami Muhammad. (2010). *Research Methods in Education and Psychology*, 6<sup>th</sup> Edition, Al Masirah House for Publishing, Distribution and Printing, Amman - Jordan.
- Mikhael, Amtanius. (2009). *Measurement and evaluation in modern education*. Damascus: Damascus University Publications.
- Mislevy, R. T. (1990). Modeling Item Response When Different Subjects Employ Different Solations Strategies, *Netherlands: Psychometrica*, Vol. 55, No., PP. 195-215.
- Nabhan, Moses. (2004). *Fundamentals of measurement in behavioral sciences*. Amman, Dar Al-Shorouk for Publishing and Distribution, 1st edition.
- Pelton, W. (2002). *The Accuracy of Unidimensional Measurement Models in the Presence of Deviations from the Underlying Assumptions*. Unpublished Doctoral Dissertation. Brigham Young University. USA.
- Ree, M. J. Jensen, H. E. (1983). Effects of Sample size on Linear Equating of Item Characteristics Curve Parameters. In D. J. Weiss (Ed). *New Horizons in Testing*. Pp. (135-146).
- Rust, J. & Golombok, S. (1999). *Modern Psychometrics: The Science Of Psychological Assessment*. (2nd ed), U.S.A, Routledge (UK).
- Sarhan, Muhammad. (2018). *Comparison of the two methods of maximum likelihood and maximum value of the dimensional distribution in power estimation according to the theory of response to the paragraph*.

- Unpublished doctoral thesis. Ibn Rushd College of Education for Humanities. Baghdad University.
- Seong, T.-J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*. 14. 299-311.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement* 16.1-16.
- Tantawi, El-Shaimaa. (2007). *The effect of both sample size and test length on the accuracy of the results of the test benchmark tests in mathematics using the Rush model*. A magister thesis that is not published. Faculty of Education. Zagazig University.
- Tarawneh, Nadia. (2017). *The effect of sample size on estimating the stability factor in light of generalization theory*. A magister thesis that is not published. College of Graduate Studies. Mutah University: Jordan.
- The hill, happy. Al-Batsh, Muhammad Walid. Abu Zina, Farid Kamel. (2007). *Scientific Research Methods - Research Design and Statistical Analysis*. 1<sup>st</sup> floor, Amman, Dar Al Masirah for Publishing, Distribution and Printing.
- Tighza, Mohamed Bouziane. (2012). *Exploratory and empirical factor analysis of its concepts and methodology using SPSS and LISREL*. Al Masirah House for Publishing, Distribution and Printing, Amman, Jordan.
- Tiot, M.D., (2003). *IRT From SSI: BILOG\_MG, MULTILOG*.
- Wang, T, Vispoel. W. (1998). Properties of Ability Estimation Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*, Vol. 35, Issue 22.
- Yaqoub, Ibrahim Muhammad. (1990). *A comparative study of the psychometric properties of the self-concept scale built in the traditional way Rush model method*. Unpublished doctoral thesis, University of Jordan.
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*. 52. 275-291.