Jerash for Research and Studies Journal مجلة جرش للبحوث والدراسات

Volume 22 | Issue 2 Article 22

2021

Comparing Ability Estimation Accuracy Using Test Length and Ability Parameter Distribution Based on Three Parameter Logistic Model (3PLM) Using Real and Generated Test Scores

AMER MAHMOUD AL-SOBUH M.A
Jordanian Ministry of Education, amersobuh84@gmail.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

AL-SOBUH, AMER MAHMOUD M.A (2021) "Comparing Ability Estimation Accuracy Using Test Length and Ability Parameter Distribution Based on Three Parameter Logistic Model (3PLM) Using Real and Generated Test Scores," *Jerash for Research and Studies Journal مجلة جرش للبحوث والدراسات*: Vol. 22: Iss. 2, Article 22.
Available at: https://digitalcommons.aaru.edu.jo/jpu/vol22/iss2/22

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 مجلة جرش للبحوث والدراسات 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 جرش للبحوث والدراسات المجلد 22 العدد 2، 2021، ص ص 1219- 1250

المقارنة بين دقة تقدير القدرة بإختلاف طول الإختبار وشكل توزيع معلمة القدرة تبعا للنموذج اللوجستي ثلاثي المعلمة بإستخدام بيانات حقيقية ومولدة

عامر محمود سليم الصبح*

تاريخ الاستلام 2020/1/28 تاريخ القبول 2020/3/8

ملخص

هدفت الدراسة الحالية إلى المقارنة بين دقة تقدير القدرة باختلاف طول الاختبار وشكل توزيع معلمة القدرة للبيانات الحقيقية والمولدة، وذلك للإجابة على التساؤل الذي يدور حول صلاحية استخدام البيانات المولدة في الفحص والمقارنة بين الإحصائيات والنماذج التي قدمتها نظرية الاستجابة للفقرة، ولتحقيق هدف الدراسة تم استخدام بيانات الاختبار الوطني لمادة العلوم للصف الثامن لعام 2013/2012 م كمصدر للعينات التي تعتمد على بيانات حقيقية، ثم تم توليد استجابات مماثلة لخصائص العينات التي تعتمد على البيانات الحقيقية باستخدام برمجية WINGEN، وتم تقدير قدرات الأفراد والخطأ المعياري للتقدير باستخدام طريقة تقدير الأرجحية العظمى وطريقة توقع التوزيع البعدي والمعمول بها في برمجية BILOG-MG، لكل من العينات التي تعتمد على البيانات الحقيقية والمولدة، وبعد إجراء التحليل الإحصائي والمتمثل باستخدام اختبار t للعينات المترابطة وحساب حجم الأثر، توصلت الدراسة إلى عدم وجود فروق عملية في دقة التقديرات لمعلمة القدرة باختلاف طبيعة البيانات (حقيقية، مولدة) وذلك باختلاف طول الاختبار (50,25) فقرة، واختلاف شكل التوزيع لمعلمة القدرة (طبيعي، ملتوي نحو اليمين، ملتوي نحو اليسار).

الكلمات المفتاحية: دقة تقدير القدرة، التوزيع الطبيعي، التوزيع موجب الالتواء، التوزيع سالب الالتواء، الارجحية العظمى، النموذج اللوجستي ثلاثي المعلمة.

© جميع الحقوق محفوظة لجامعة جرش 2021.

* وزارة التربية والتعليم، الأردن. Email: amersobuh84@gmail.com

Comparing Ability Estimation Accuracy Using Test Length and Ability Parameter Distribution Based on Three Parameter Logistic Model (3PLM) Using Real and Generated Test Scores

Amer Mahmoud Salim Al-Sobuh, *Ministry of Education, Jordan.*

Abstract

The study aimed to compare ability estimation accuracy using test length and ability parameter distribution based on Three Parameter Logistic Model (3PL) using real and generated test scores. to answer the question, which is about the validity of the use of the data generated in the examination and comparison between the statistics and models provided by item response theory , and to achieve the aim of the study, eighth grade national science scores for the school year 2012\2013 were used as the real scores in the study and generate scores were obtained using the real scores for the same sample via WINGEN 3.0 software. Examinees' abilities and the estimation standard error were computed via maximum likelihood and expected a posterior used in BILOG-MG 3.0 software for the real and generated samples. After the suitable statistical analysis procedures were performed (paired sample t-test and effect size statistics for the dependent sample t-test), it was found that there is no practical differences in the estimation accuracy for the ability parameter based on the nature of scores (real, generated) based on the test length (25 and 50) items, and the distribution of the ability parameter (normal distribution , positively skewed, negatively skewed).

Keywords: Accuracy of Ability estimation, Normal distribution, Positive and negative skewed, Maximum Likelihood, Three parameter logistic model.

مقدمة الدراسة والإطار النظري.

يعنى القياس النفسي بالتعرف على الفروق بين درجة امتلاك الأفراد لسمة أو مجموعة سمات أو ما يشار إليه بالفروق بين الأفراد (Inter Individual)، كما يهتم بدرجة تفاوت امتلاك الفرد للسمات المختلفة وهو ما يشار إليه بالفروق داخل الفرد (Intra Individual) بهدف تصنيف الأفراد، الأمر الذي يؤدي إلى إتخاذ قرارات مناسبة لهم وذلك بوضع الأفراد في أماكن وبرامج وتخصصات مناسبة لإحتياجاتهم ومهاراتهم (الطريري، 1997).

ويهتم العاملين في مجال القياس النفسي والتربوي بالوصول إلى مستوى مقبول من الدقة والموضوعية في عمليات القياس، وقد نشأت نظريات القياس النفسي التي حاولت الوصول إلى مؤشرات إحصائية لقياس درجة الدقة فيما يراد قياسه. ولعل إستخدام نظرية القياس الكلاسيكية

(Classical Test Theory – CTT) ساعد في تقدير قدرات الأفراد والتي يعبر عنها بالعلامة الحقيقية (True Score)، إلا أن إعتماد الظواهر المقاسة وفق النظرية الكلاسيكية على خصائص عينة الأفراد والأداة المستخدمة على هذه العينة حد من التوصل إلى نتائج دقيقة وموضوعية في القياس، لذا وجه المختصون في القياس جهودهم لإيجاد نظام قياس أكثر موضوعية، للوصول في عملية القياس إلى قيم تعكس بشكل أكثر دقة مقدار ما لدى الفرد من السمة موضوع القياس (قدرة الفرد)، مما أدى ذلك إلى التوصل إلى نظرية السمات الكامنة (Latent Trait Theory-LTT)، وأشار كروكر والجينا (Croker & Algina, 1986) إلى أن الإفتراض الأساسي في نظرية الاستجابة للفقرة هو أن هنالك سمة كامنة أو قدرة ويعبر عنها بالرمز اللاتيني (θ) تكمن وراء استجابة المفحوص على أي فقرة من فقرات الإختبار. وأضاف كلاين (Kline, 2005) إلى أن العلاقة بين مستوى قدرة الفرد θ واستجابة المفحوص الصحيحة عن أي فقرة من فقرات الإختبار هي علاقة طردية، فالفرد الذي له مستوى قدرة عال في الرياضيات له احتمالية عالية في الإجابة عن فقرات اختبار الرياضيات إجابة صحيحة والعكس كذلك.

يمكن وصف العلاقة بين أداء الفرد على الفقرة ومقدار ما لديه من سمة من خلال معادلات أو نماذج رياضية يمكن تمثيلها بيانيا أطلق عليها منحنى خصائص الفقرة (Item Characteristic Curve-ICC)، وتختلف هذه النماذج بإختلاف المعالم التي يتضمنها النموذج ونوع الإجابة على الفقرة، ثنائية الاستجابة (Dichotomous) أو متعددة الاستجابة (Polytomous)، وطبيعة السمة المقاسة سواء أكانت أحادية البعد أم متعددة الأبعاد، وتكمن الفكرة الأساسية لنماذج الاستجابة للفقرة الإختبارية في أن هذه النماذج تحاول اشتقاق قياسات أو قيم تقديرية للسمة أو السمات التي تنطوي عليها مجموعة من الاستجابات لمجموعة من الفقرات الإختبارية (علام، 2002).

ومن أهم النماذج اللوجستية للفقرات ثنائية الاستجابة النموذج اللوجستي أحادي المعلمة والنموذج اللوجستي ثنائي المعلمة، والنموذج اللوجستي ثلاثي المعلمة. والخطوة الأولى والأهم في تطبيق نظرية الاستجابة للفقرة على بيانات الإختبار هي في تقدير المعالم التي يشتمل عليها نموذج معين يتم اختياره، ولذلك ظهر العديد من طرق التقدير لمعالم الفقرات والأفراد (قدرة الأفراد) الكامنة وراء نمط الاستجابة الظاهرة على فقرات الإختبار، ومنها: طريقة الأرجحية العظمى (Maximum Likelihood Estimation-MLE)، وطريقة توقع التوزيع البعدي Expected A Posteriori-EAP).

وفيما يتعلق بتقدير القدرة بإستخدام طريقة الارجحية العظمى، تتصف التقديرات الناتجة بالاتساق Consistency بحيث انه كلما كبر حجم العينة فإن التقديرات تقترب من القيم الحقيقية،

كما تتصف تقديرات هذه الطريقة بالفعالية Efficient بمعنى أن لهذه التقديرات اقل تباين، وتتوزع هذه التقديرات على شكل توزيع طبيعي بوسط يساوي القيمة الحقيقية لـθ وبتباين يساوي $I(\theta)^{-1}$ (Hambleton & swaminathan, 1985). إلا أنه يمكن أن لا يكون لاقتران الأرجحية قيم نهائية عظمى، ويحدث ذلك عندما يجيب المفحوص إجابة صحيحة على كل الفقرات ويكون تقدير الأرجحية العظمى $\theta = \infty$ أو إجابة خاطئة على كل الفقرات وفي هذه الحالة يكون تقدير الأرجحية العظمى $\theta = -\infty$، وفي هذه الحالة يتم إلغاء مثل هذه الحالات قبل إجراء عملية التقدير بإستخدام طريقة الأرجحية العظمى، فلا نستطيع تقدير قدرات مثل هؤلاء الأفراد فعليا. كما انه في بعض الإجابات الخاصة التي لا تكون مميزة كالبديهيات فهذه لا يكون لها قيم قصوى في اقتران الأرجحية، وفي بعض الحالات يكون نمط الاستجابة مضللا Aberrant مثل أن يجيب المفحوص إجابة صحيحة عن الفقرات الصعبة والمميزة ويجيب عن الفقرات السهلة إجابة غير صحيحة، وعندها فان الطرق العددية المستخدمة لإيجاد قيم قصوى سوف تنحرف، وغالبا ما تحدث هذه المشكلة في النموذج اللوجستي ثلاثي المعلمة وليس مع النموذج أحادي وثنائي المعلمة، كما أنه بالإمكان حدوث هذه المشكلة في الإختبارات التي تحوي أكثر من أربعين فقرة (Hambleton, Swaminthan & Rogers, 1991). وحيث أن طريقة الأرجحية العظمى لا تقدم معلومات حول قدرة المفحوصين الذين أجابوا بشكل كامل على فقرات الإختبار، أو حتى الذين لم يجيبوا عن أي فقرة إجابة صحيحة، وبتوفر المعلومات حول التوزيع القبلي لقدرات المفحوصين يصبح بالإمكان إستخدام الأساليب التي تعتمد على نظرية بيز للحصول على تقديرات جيدة للقدرة (علام، 2005). ومن هذه الأساليب التقدير بإستخدام أسلوب توقع التوزيع البعدي وفي التقدير بإستخدام أسلوب توقع التوزيع البعدي ((Expected A Posterior (EAP) يتم إستخدام التوزيع الطبيعي المعياري حيث يتم تقسيم السمة الكامنة θ التي تمثل هذا التوزيع والتي تقع بين [3 ,3-] في العادة إلى 61 تجزيئا بفترة طولها 0.1 يطلق على كل منها ربيعا ويعطى لكل ربيع وزنا يرمز له بالرمز $W_r(Q_r)$ (التقي،2009). والميزة الرئيسية لهذا الأسلوب هو عدم الحاجة إلى عمليات تقريب تستخدم طرق عددية مثل طريقة نيوتن رافسن والمستخدمة في أسلوب تقدير الأرجحية العظمى وأساليب أخرى، حيث يظهر من صيغة Gauss-Hermite أن التقدير في أسلوب توقع التوزيع البعدي عبارة عن سلسلة من المقادير المجمعة (keller, 2000)إلا إن عدم وجود عدد كافي من الربيعيات يؤثر على دقة تقدير أسلوب توقع التوزيع البعدي Hou & Dodd, (Chen, 1997). كما أنه وفي دراسات المحاكاة فإن تقديرات أسلوب توقع التوزيع البعديEAP لم تكن بدقة تقديرات أسلوب القيمة العظمى للتوزيع البعديMAP، وذلك بغياب عدد كبير من

الفقرات ذات التمييز العالي، وذلك لكل مستويات القدرة (Wang & Vispoel, 1998).كما أن طريقة توقع التوزيع البعدي تعاني من مشكلة إقتراب التقديرات من متوسط التوزيع القبلي، وإستخدام توزيع قبلي غير مناسب سيؤثر على دقة التقديرات، كما وجد بأنه فقط عندما يقترب عدد الفقرات من عدد كبير جدا فإن التقدير يقترب من القيمة الحقيقية لـ θ (Bock & Mislevy, 1982).

مشكلة الدراسة وأسئلتها:

إن إستخدام الأسلوب المناسب لتقدير معلمة القدرة أمر هام وضروري بهدف الحصول على أفضل تقدير، كما أن طبيعة البيانات المستخدمة في تقدير معلمة القدرة تلعب دورا رئيسا في دقة النتائج المحصلة، ومن هنا كانت مشكلة الدراسة والتي تهدف للوقوف حول التساؤل الذي يدور حول مدى الدقة في إستخدام البيانات المولدة للفحص والمقارنة بين النماذج والإحصائيات المختلفة التي قدمتها نظرية الاستجابة للفقرة تحت ظروف مختلفة. وللإجابة على هذا التساؤل الرئيسي، فإن الدراسة الحالية تسعى للإجابة على السؤالين الآتيين:

السؤال الأول: هل تختلف دقة تقديرات القدرة بإختلاف عدد فقرات الإختبار في ظل النموذج اللوجستي ثلاثي المعلمة للبيانات الحقيقة والمولدة؟

السؤال الثاني: هل تختلف دقة تقديرات القدرة بإختلاف شكل توزيع معلمة القدرة وطول الإختبار في ظل النموذج اللوجستي ثلاثي المعلمة للبيانات الحقيقية والمولدة؟

أهداف الدراسة:

إنطلاقا من أن توخي الدقة في عملية القياس يؤدي إلى الحيلولة دون اتخاذ قرارات خاطئة على مستوى الفرد والمجتمع، فقد كان هدف الدراسة يرتكز على إستخدام طريقة الارجحية العظمى وطريقة توقع التوزيع البعدي بهدف الكشف عن دقة تقدير القدرة، في ظل تباين طول الإختبار وشكل توزيع معلمة القدرة تبعا للنموذج اللوجستي ثلاثي المعلمة، وذلك بإستخدام بيانات حقيقية وبيانات تحاكي تلك البيانات الحقيقية، بهدف الوقوف على التساؤل الذي يدور حول مدى دقة إستخدام البيانات المولدة في فحص النماذج الرياضية والإحصائيات التي قدمتها نظرية الاستجابة للفقرة في ظل تباين عدد من الظروف المختلفة.

أهمية الدراسة:

تحاول هذه الدراسة الكشف عن مدى صلاحية إستخدام البيانات المولدة في فحص الإحصائيات المختلفة والتي قدمتها نظرية الاستجابة للفقرة، وذلك بالاعتماد على طريقتين شائعتي

الإستخدام في تقدير القدرة للمفحوصين وهما طريقة الأرجحية العظمى وطريقة توقع التوزيع البعدي، مع الأخذ بعين الاعتبار طول الإختبار وشكل توزيع معلمة القدرة، وعليه فمن المتوقع أن يكون للدراسة مساهمة في الجانبين التاليين:

أولا: من الناحية النظرية:

جاءت هذه الدراسة للكشف عن مدى التطابق بين التقديرات لمعلمة القدرة والتي تعتمد على بيانات حقيقية وتقديرات القدرة الناتجة عن إستخدام بيانات مولدة في ظل تباين طول الإختبار وشكل توزيع معلمة القدرة بالاعتماد على النموذج اللوجستي ثلاثي المعلمة.

ثانيا: من الناحية العملية:

تأتي الأهمية العملية لهذه الدراسة من مدى التكاليف المادية التي قد يتكلفها الباحث في الحصول على بيانات حقيقية من مجتمع معين، أضف إلى ذلك الوقت والجهد المبذول للحصول على هذه البيانات، كما أنه قد يتعذر الحصول على بيانات حقيقية من مجتمع معين لسبب ما، لذلك وبالإعتماد على الفائدة النظرية فإن التطابق بين تقديرات القدرة للبيانات الحقيقية والبيانات المولدة بشكل يحاكي تلك البيانات الحقيقية سيسهم في توفير الوقت والجهد والتكلفة، كما سيعطي الباحث فكرة عن طبيعة المجتمع الذي بصدد دراسته، مما يسهم في انتقاء الطرق والأساليب المناسبة لتقدير معالم ذلك المجتمع قبل الشروع بالعمل الميداني.

فرضيات الدراسة:

تتحدد فرضيات الدراسة كما يأتي:

1- لا توجد فروق ذات دلالة احصائية عند مستوى الدلالة ($\alpha = 0.05$) بين الاخطاء المعيارية لتقديرات القدرة بإختلاف عدد فقرات الإختبار في ظل النموذج اللوجستي ثلاثي المعلمة للبيانات الحقيقة والمولدة.

2- لا توجد فروق ذات دلالة احصائية عند مستوى الدلالة ($\alpha = 0.05$) بين الاخطاء المعيارية لتقديرات القدرة بإختلاف شكل توزيع معلمة القدرة وطول الإختبار في ظل النموذج اللوجستي ثلاثي المعلمة للبيانات الحقيقية والمولدة.

حدود الدراسة:

1- تتحدد الدراسة بإستخدام طريقة الارجحية العظمى (MLE) وأسلوب توقع التوزيع البعدي(EAP) المعمول بها في برمجية Bilog-mg3.

2- كما أن الدراسة ستقتصر بالنسبة للبيانات الحقيقية على فقرات الإختبار الوطني لمادة العلوم للصف الثامن الأساسي للعام الدراسي2013/2012 والتي تطابق إفتراض النموذج اللوجستي ثلاثي المعلمة في هذه الدراسة.

3- إستخدام التوزيع الطبيعي والتوزيع الملتوي نحو اليمين والتوزيع الملتوي نحو اليسار للعينات المسحوبة بشكل عشوائي.

4- اقتصرت الدراسة على إستخدام حجم عينة مقداره (1000) مفحوص للعينات المستخدمة في مختلف المواقف البحثية.

5- اقتصرت الدراسة على إستخدام طول الإختبار(25) فقرة و(50) فقرة في مختلف المواقف البحثية.

الدراسات السابقة:

قام ريكاس ومارك (Reckas & Mark, 1978) بإجراء دراسة هدفها مقارنة دقة تقدير معالم القدرة والفقرة للنموذج اللوجستي أحادي المعلمة والنموذج اللوجستي ثلاثي المعلمة وأظهرت النتائج أن النموذج اللوجستي ثلاثي المعالم قد طابق بيانات الإختبار بشكل أفضل من النموذج أحادي المعلمة، كما كان تباين تقدير معالم القدرة للنموذج أحادي المعلمة اكبر من تباين تقدير معلم القدرة للنموذج ثلاثي المعلمة، وان النموذج ثلاثي المعلمة يحتاج إلى حجم عينة اكبر لمعايرة فقرات الإختبار من النموذج أحادي المعلمة كما يفضل إستخدام النموذج أحادي المعلمة في حالة العينات الصغيرة، وأظهرت النتائج أن هنالك ارتباطا عاليا بين تقديرات القدرة وفق النموذجين لمعظم مجموعة البيانات.

وأجرى هارول (Harwell, 1996) دراسة هدفت لتفحص أخطاء تقدير معلم القدرة بإختلاف حجم العينة بإستخدام أسلوب المحاكاة وفقا للنموذج اللوجستي ثنائي المعلمة، حيث تم توليد استجابات على 25 فقرة ثنائية الاستجابة بأحجام عينات مختلفة (500،1000،1500،2000) مفحوص، وحساب الخطأ المعياري لتقديرات الصعوبة والقدرة، حيث أظهرت النتائج أن الخطأ المعياري لتقديرات معلم الصعوبة يقل مع زيادة حجم العينة، بينما كان الخطأ المعياري لتقدير معلم القدرة متقارباً في كل الظروف.

كما قام كل من شن وهو ودود (Chen, Hou & Dodd, 1997) بعمل دراسة محاكاة بهدف التحقق من تطبيق طريقة توقع التوزيع البعدي EAP لتقدير القدرة على الإختبارات التكيفية المحوسبة CAT بالاعتماد على نموذج التقدير الجزئي ومقارنة التقديرات الناتجة من هذه

الطريقة، مع التقديرات الناتجة من إستخدام طريقة الارجحية العظمى MLE، حيث تم تقييم أداء طريقة توقع التوزيع البعدي EAP في ظل ظروف مختلفة، وذلك بإختلاف عدد الربيعيات Quadrature (80,40,20,10)، واختلاف التوزيع القبلي (طبيعي، منتظم)، وجرى تقييم الأداء النسبي لطريقة الارجحية العظمى وطريقة توقع التوزيع البعدي تحت شكلين من أشكال التوزيع للسمة الكامنة (طبيعي، ملتوي نحو اليسار)، أظهرت النتائج أنه بغض النظر عن شكل توزيع السمة الكامنة فإن طريقة الارجحية العظمى وطريقة توقع التوزيع البعدي مع التوزيع القبلي الطبيعي أو المنتظم ضمن إستخدام 80,40,20 ربيعيا تقدم كلتا الطريقتين تقديرات دقيقة في الإختبارات التكيفية المحوسبة بالاعتماد على نموذج التقدير الجزئي، كما أن الزيادة في عدد الربيعيات المستخدمة من 20 إلى 80 لا يزيد في دقة التقدير لطريقة توقع التوزيع البعدي، وأظهرت النتائج أن إستخدام التوزيع القبلي الطبيعي لطريقة توقع التوزيع البعدي أفضل من إستخدام التوزيع القبلي الملتوي نحو اليمين أو اليسار، كما أن إستخدام التوزيع القبلي الملتوي نحو اليمين لم يقدم تقديرات أدق من إستخدام التوزيع الملتوي نحو اليسار، إضافة إلى ذلك فإن إستخدام طريقة توقع التوزيع البعدي بإستخدام التوزيع الطبيعي أو المنتظم كتوزيع قبلي اظهر أن هذه الطريقة تعتبر بديل جيد لطريقة الارجحية العظمى للفقرات متعددة التدريج المستخدمة في الإختبارات التكيفية المحوسبة.

أما دراسة لندن (Linden, 1998) فقد هدفت إلى فحص طرق تقدير المعلمات المعتمدة على أسلوب بيز في انتقاء الفقرات في الإختبارات التكيفية من حيث دقة التقديرات التي تقدمها، إذ اقترحت هذه الدراسة عدة طرق لتحديد معايير انتقاء الفقرات للاختبارات التكيفية، والتي تعتمد على التوزيع البعدي الحقيقي (true posterior)، حيث تم إستخدام مجمع فقرات مؤلف من 300 فقرة، وتم معايرتها بإستخدام النموذج اللوجستي ثنائي المعلمة، وتوصلت الدراسة إلى أن طرق انتقاء الفقرات تبعاً لأسلوب بيز تتفوق على طريقة الأرجحية العظمى في تقدير معلمة القدرة، خصوصاً عند إستخدام الإختبارات القصيرة واعتماد معيار متوسط مربعات الأخطاء mean-squared error كمحك لدقة التقديرات.

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

كما قام كيم (Kim, 1998) بإجراء دراسة تم فيها إستخدام أسلوب المحاكاة لتقدير معالم الفقرات والأفراد بإستخدام النموذج اللوجستي أحادي المعلمة حيث تم فحص أربعة مجموعات من المعلومات بإستخدام طريقة الأرجحية العظمى الشرطية وطريقة الأرجحية العظمى الهامشية وطريقة الأرجحية العظمى المشتركة وطريقة بيز التي تعتمد على توقع التوزيع البعدي حيث توصلت الدراسة إلى أن نتائج تقدير معلمات الفقرة ومعلمة القدرة كانت متشابهة للطرق الأربعة.

وقام كل من ليو وسينيكا ويو (Liou, Sinica & Yu,1999) بإجراء دراسة هدفت لتفحص دقة التقديرات لمعلم القدرة بإستخدام أسلوب المحاكاة تبعا للنموذج اللوجستي ثلاثي المعالم حيث تم توليد فقرات ثنائية الاستجابة لتصحيح ثلاثة اختبارات تكونت من (60,40,20) فقرة وبحجم عينة 500 مفحوص، وتم إستخدام طريقة الأرجحية العظمى (MLE) لتقدير معلم القدرة، وأظهرت النتائج أن الخطأ المعياري للتقدير كان مرتفعا عند المستويات المرتفعة والمنخفضة من القدرة.

وقد أجرى **عبابنة (2004)** دراسة هدفت إلى فحص اثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية بإستخدام نظرية الاستجابة للفقرة، حيث تم بناء اختبار قدرة عقلية مكون من أربعة اختبارات فرعية وبلغ عدد الفقرات بالصورة النهائية (71) فقرة وطبق الإختبار على 1000 طالب وطالبة من الصف السابع الأساسي، وقد تم إستخدام برنامج Bilog3.11 لتقدير قدرة المفحوصين، ومعالم فقرات الإختبار، والأخطاء المعيارية في التقدير، وإحصائيات مطابقة البيانات للنموذج اللوجستي الثلاثي الذي تم اعتماده لهذه الدراسة وقد تم التوصل إلى نتائج منها: تزداد الدقة في تقدير معالم الفقرة بزيادة حجم عينة المفحوصين، وتتسم تقديرات القدرة للمفحوصين بالاستقرار عندما يتم إستخدام عينات معايرة بأحجام كبيرة، كما تتأثر الدقة في تقدير القدرة للمفحوصين بوضع مطابقة البيانات للنموذج ويكون عندئذ من غير المؤكد الحصول على دقة اكبر عند زيادة أحجام العينات المستخدمة في المعايرة عن مستوى معين، وتزداد الدقة في تقدير معلم القدرة بزيادة عدد فقرات الإختبار أو نسبتها إلى الإختبار الكلي.

كما أجرى **جاري وفيرمينت (Garre & Vermunt, 2006)** دراسة بتطبيق بيانات حقيقية بهدف التحقق من نتائج الاشتقاقات الرياضية التي توصل إليها، والتي تخص دقة تقديرات المعلمات، حيث تم إستخدام توزيعات قبلية مختلفة لمعلمة القدرة، وقد دلت النتائج أن تقديرات المعلمات التي يتم الحصول عليها من خلال طريقة بيز تتمتع بثبات أفضل من تقدير المعلمات التي تم الحصول عليها عند إستخدام طريقة الارجحية العظمى، وبخاصة عند تقدير القدرات عند أطراف متصل القدرة.

أما **الطراونة (2011)** فقد قامت بدراسة هدفت إلى المقارنة بين طرق تقدير القدرة بإستخدام النموذج المناسب في ضوء الخطأ المعياري، تم إستخدام نتائج الإختبار الوطني لضبط نوعية التعلم لعام 2010 المعد من قبل وزارة التربية والتعليم الأردنية لطلبة الصف الثامن الأساسي في مادتي الرياضيات والعلوم، حيث استخدمت الباحثة برمجية BILOG-MG للإجابة عن أسئلة الدراسة، وقد تم التوصل إلى التالي: إن طريقة توقع التوزيع البعدي (EAP) أعطت أعلى درجة في دقة التقدير للقدرة تليها طريقة القيمة العظمى للتوزيع البعدي (MAP)، ثم تليها طريقة الارجحية العظمى (MLE) حيث كانت الأخيرة الأقل درجة في دقة تقدير معلمة القدرة، وتبين أن طريقة الأرجحية العظمى (MLE) تعمل على زيادة في الخطأ المعياري للتقدير عند تقدير القدرة، وذلك عند نهاية طرفي التوزيع.

كما أجرى الشريفين (2012) دراسة هدفت للكشف عن أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغيّر حجم العينة. لتحقيق هدف الدراسة تم بناء اختبار تحصيلي في الفيزياء من نوع الاختيار من أربعة بدائل، تكوّن بصورته النهائية من (33) فقرة حيث طبق الإختبار على عينة الدراسة المكونة من (1000) طالبٍ وطالبة من طلبة الصف الثاني الثانوي العلمي، وحللت النتائج وفق النموذج الثلاثي المعلمة بإستخدام البرمجية BILOG-MG-3.0، أظهرت النتائج وجود فروق ذات دلالة إحصائية ($\alpha \leq 0.05$) في متوسطات الأخطاء المعيارية لتقديرات معالم الفقرات تعزى للتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة أو طريقة التقدير، كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية ($\alpha \leq 0.05$) في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لمتغير حجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لطريقة التقدير. كما بيّنت النتائج عدم وجود فروق دالة إحصائياً بين معاملات الثبات المقدرة وفق نظرية الاستجابة للفقرة عند أحجام العينة المختلفة (10 ,500 ,1000). وأشارت النتائج إلى أن دقة تقديرات معلمة القدرة تزداد في حالة عينة الأفراد ذوي القدرة العالية وعينة الأفراد ذوي القدرة المتدنية عند إستخدام طريقة بيز التوقع (EAP) ، في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسطة بإستخدام طريقة الأرجحية العظمى(MLE) بغض النظر عن حجم العينة.

مما سبق يظهر التفاوت في دقة تقدير طريقة الارجحية العظمى وطريقة بيز بتفاوت بعض العوامل مثل حجم العينة أو شكل توزيع قدرات الأفراد، كما تناولت الدراسات السابقة بيانات حقيقية أو مولدة في الوصول إلى استنتاجاتها المتعلقة بدقة تقدير القدرة لطريقة الارجحية العظمى وطريقة بيز،ومن هنا تكمن أهمية الدراسة الحالية وذلك بعمل مقارنة بين تقديرات القدرة

الناتجة عن بيانات حقيقية، والتقديرات الناتجة عن إستخدام بيانات تحاكي خصائص تلك البيانات الحقيقية،وذلك بتفاوت طول الإختبار وشكل توزيع معلمة القدرة بإستخدام طريقة MLE وطريقة EAP، وذلك للكشف فيما إذا كانت دقة التقديرات لمعلمة القدرة تختلف بتفاوت طبيعة البيانات حقيقية أم مولدة، والى أي مدى يمكن التعويل على البيانات المولدة في استنتاج خصائص المجتمع.

التعقيب على الدراسات السابقة:

مما سبق يظهر التفاوت في دقة تقدير طريقة الارجحية العظمى وطريقة بيز بتفاوت بعض العوامل مثل حجم العينة أو شكل توزيع قدرات الأفراد، كما تناولت الدراسات السابقة بيانات حقيقية أو مولدة في الوصول إلى استنتاجاتها المتعلقة بدقة تقدير القدرة لطريقة الارجحية العظمى وطريقة بيز،ومن هنا تكمن أهمية الدراسة الحالية وذلك بعمل مقارنة بين تقديرات القدرة الناتجة عن بيانات حقيقية، والتقديرات الناتجة عن إستخدام بيانات تحاكي خصائص تلك البيانات الحقيقية،وذلك بتفاوت طول الإختبار وشكل توزيع معلمة القدرة بإستخدام طريقة MLE وطريقة EAP، وذلك للكشف فيما إذا كانت دقة التقديرات لمعلمة القدرة تختلف بتفاوت طبيعة البيانات حقيقية أم مولدة، والى أي مدى يمكن التعويل على البيانات المولدة في استنتاج خصائص المجتمع.

المنهجية والتصميم

يتناول هذا الفصل وصفا شاملا لمجتمع الدراسة وعينتها، والإجراءات التي تم إتباعها للتحقق من أحادية البعد وجودة مطابقة البيانات والأفراد للنموذج المستخدم والطرق الإحصائية المستخدمة في الإجابة عن أسئلة الدراسة.

مجتمع الدراسة:

تألف مجتمع الدراسة من بيانات جميع طلبة الصف الثامن في مدارس المملكة الأردنية الهاشمية الحكومية، والخاصة، والمدارس التابعة لوكالة الغوث، والمدارس التابعة لمديرية التعليم والثقافة العسكرية للعام الدراسي 2013/2012م، والبالغ عددهم (25151) طالب وطالبة والذين تقدموا للاختبار الوطني لمادة العلوم وذلك حسب إحصائيات وزارة التربية والتعليم الاردنية (إدارة الامتحانات والإختبارات في وزارة التربية والتعليم).

عينة الدراسة:

تم اختيار عينه عشوائية لاستجابة 1000 مفحوص تتوزع درجاتهم طبيعيا على (25) فقرة تم اختيارها عشوائيا من بين فقرات الإختبار والبالغة (50) فقرة بالاعتماد على بيانات الإختبار الوطني، وتوليد عينة تمثل إستجابة (1000) مفحوص على (25) فقرة تحاكي خصائص التوزيع الطبيعي للعينة التي تعتمد على بيانات الإختبار الوطني، واختيار عينه عشوائية لاستجابة 1000 مفحوص تتوزع درجاتهم توزيعا ملتويا نحو اليسار على (25) فقرة والتي سبق وأن تم اختيارها عشوائيا من بين (50) فقرة بالاعتماد على بيانات الإختبار الوطني، وتوليد عينة تمثل استجابة (1000) مفحوص على (25) فقرة تحاكي خصائص التوزيع الملتوي نحو اليسار للعينة التي تعتمد على بيانات الإختبار الوطني المناظرة لها، واختيار عينه عشوائية لاستجابة 1000 مفحوص تتوزع درجاتهم توزيعا ملتوياً نحو اليمين على (25) فقرة، والتي سبق وأن تم اختيارها عشوائيا من بين (50) فقرة بالاعتماد على بيانات الإختبار الوطني، وتوليد عينة تمثل استجابة (1000) مفحوص على (25) فقرة تحاكي خصائص التوزيع الملتوي نحو اليمين للعينة التي تعتمد على بيانات الإختبار الوطني المناظرة لها، وإستخدام عينة عشوائية لاستجابة 1000 مفحوص تتوزع درجاتهم طبيعيا على (50) فقرة بالاعتماد على بيانات الإختبار الوطني، وتوليد عينة تمثل استجابة (1000) مفحوص على (50) فقرة تحاكي خصائص التوزيع الطبيعي للعينة التي تعتمد على بيانات الإختبار الوطني المناظرة لها، وإستخدام عينه عشوائية لاستجابة 1000 مفحوص تتوزع درجاتهم توزيعا ملتويا نحو اليسار على (50) فقرة بالاعتماد على بيانات الإختبار الوطني، وتوليد عينة تمثل استجابة (1000) مفحوص على (50) فقرة تحاكي خصائص التوزيع الملتوي نحو اليسار للعينة التي تعتمد على بيانات الإختبار الوطني المناظرة لها، واختيار عينه عشوائية لاستجابة 1000 مفحوص تتوزع درجاتهم توزيعا ملتويا نحو اليمين على (50) فقرة بالاعتماد على بيانات الإختبار الوطني وتوليد عينة تمثل استجابة (1000) مفحوص على (50) فقرة تحاكي خصائص التوزيع الملتوي نحو اليمين للعينة التي تعتمد على بيانات الإختبار الوطني المناظرة لها، وذلك بعدما تم ترتيب درجاتهم على الإختبار ترتيبا تنازليا بالاعتماد على برمجية (SPSS)، والجدول (1) يبين الإحصاءات الوصفية للعينات التي تم اختبارها عشوائيا بالاعتماد على بيانات الإختبار الوطني لمادة العلوم للصف الثامن لعام 2013/2012.

الجدول (1): الإحصاءات الوصفية للعينات التي تعتمد على البيانات الحقيقية لمختلف التوزيعات

شكل التوزيع وعدد الفقرات	الوسط	الوسيط	المنوال	الانحراف المعياري	العلامة الصغرى	العلامة العظمى	عدد الأفراد
التواء سالب 25 فقرة	17.00	18.00	19.00	4.317	2	25	1000
التواء سالب 50 فقرة	31.18	34.00	38.00	9.791	2	49	1000
طبيعي 25 فقرة	9.06	9.00	00.9	3.785	0	21	1000
طبيعي 50 فقرة	18.87	19.00	19.00	6.292	4	34	1000
التواء موجب 25فقرة	11.00	10.00	7.00	4.758	1	25	1000
التواء موجب 50 فقرة	20.00	18.00	16.00	8.266	0	48	1000

أداة الدراسة:

لتحقيق هدف الدراسة والمتمثل في المقارنة بين دقة تقدير القدرة بإختلاف طول الإختبار وشكل توزيع معلمة القدرة تبعا للنموذج اللوجستي ثلاثي المعلمة بإستخدام بيانات حقيقية ومولدة، تم إستخدام نتائج الإختبار الوطني لمادة العلوم والمعد من قبل إدارة الامتحانات والإختبارات في وزارة التربية والتعليم الأردنية للصف الثامن الأساسي للعام الدراسي 2013/2012، حيث تكون الإختبار من (50) فقرة من نوع الاختيار من متعدد، يجيب عنها الطالب في زمن مقداره ساعة ونصف، وقام المعنيون في وزارة التربية والتعليم الأردنية بتشكيل لجان من مشرفي إدارة الامتحانات والإختبارات، وعدد من مشرفي الميدان ببناء فقرات الإختبار، حيث قامت هذه اللجان بدراسة نتاجات التعلم للصفوف (الخامس والسادس والسابع والثامن) لمبحث العلوم، وتحديد مهارات التعلم الأساسية التي سيقيسها الإختبار وتدريج النتاجات وفق مستويات الأداء الأربعة التالية (الأساسي، الإتقان الجزئي، الإتقان التام المتقدم).

صدق الأداة: عملت اللجان على إعداد جدول مواصفات وبناء تجمع من الفقرات الإختبارية وطباعتها، ومن ثم تجريب هذه الفقرات على عينة من الطلبة وذلك للتحقق من خصائصها السيكومترية مثل: مستوى صعوبة الفقرات وتمييزها وفاعلية البدائل ومن ثم عدلت صياغة بعض الفقرات وحذفت الفقرات التي كان تمييزها سالب أو قريبا من الصفر، وانتقيت الفقرات النهائية من تجمع الفقرات التي تم بناؤها بحيث كان الانتقاء ممثلا للأوزان والنسب في جدول المواصفات الذي تم تطويره ليشمل مؤشرات الأداء.

ثبات الأداة: تم حساب معامل ثبات الإختبارات بإستخدام معادلة "كرونباخ ألفا"، حيث بلغت قيمة معامل الثبات لاختبار العلوم النسخة الورقية (0.857) ومعامل ثبات النسخة الإلكترونية (0.844) وذلك بحسب إحصائيات وزارة التربية والتعليم.

إجراءات الدراسة:

تم التحقق من إفتراض أحادية البعد والاستقلال الموضعي ومن مطابقة الأفراد والفقرات وفق النموذج ثلاثي المعلمة حسب نظرية الاستجابة للفقرة، وذلك على النحو الآتي:

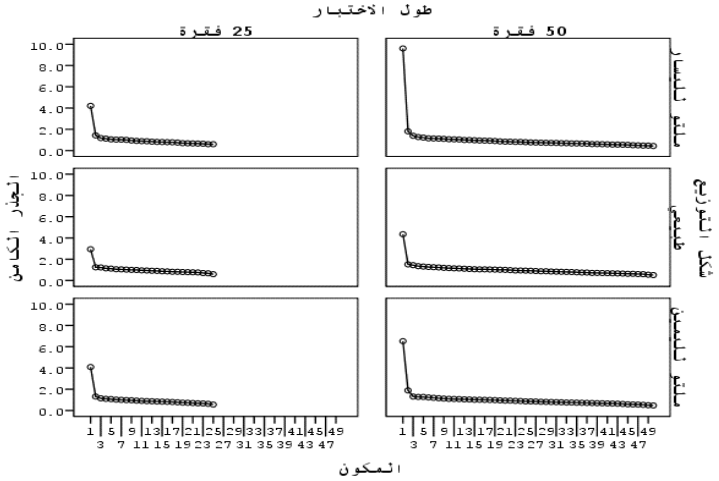
أ- التحقق من أحادية البعد Unidimensionality لبيانات الإختبار لدى أفراد العينة: تمّ التحقق من أحادية البعد بإستخدام التحليل العاملي الإستكشافي في برنامج SPSS لمعالجة بيانات العينة، والجدول(2) يوضح نتائج التحليل العاملي الإستكشافي لبيانات العينة.

الجدول (2): نتائج التحليل العاملي الإستكشافي لفقرات الإختبار.

شكل التوزيع	طول الإختبار	المكون	الجذر الكامن المدور	التباين المفسر المدور	التباين المفسر المدور التراكمي	الجذر الكامن الأول / الجذر الكامن الثاني	(الجذر الكامن الأول – الجذر الكامن الثاني) / (الجذر الكامن الثاني – الجذر الكامن الثالث)
ملتو لليسار	25	1	12.05	**48.19**[1]	48.19	**8.43**[2]	**41.17**[3]
		2	1.43	5.72	53.91		
		3	1.17	4.69	**58.60**[4]		
		4					
	50	1	22.94	**45.88**[1]	45.88	**12.68**[2]	**49.74**[3]
		2	1.81	3.62	49.50		
		3	1.38	2.77	**52.26**[4]		
		4					
طبيعي	25	1	11.81	**47.25**[1]	47.25	**9.46**[2]	**448.72**[3]
		2	1.25	4.99	52.25		
		3	1.22	4.90	**57.14**[4]		
		4					
	50	1	24.82	**49.63**[1]	49.63	**16.36**[2]	**312.04**[3]
		2	1.52	3.03	52.67		
		3	1.44	2.88	**55.55**[4]		
		4					
ملتو لليمين	25	1	10.73	**42.91**[1]	42.91	**8.21**[2]	**56.64**[3]
		2	1.31	5.23	48.13		
		3	1.14	4.56	**52.70**[4]		
		4					

شكل التوزيع	طول الإختبار	المكون	الجذر الكامن المدور	التباين المفسر المدور	التباين المفسر المدور التراكمي	الجذر الكامن الأول / الجذر الكامن الثاني	(الجذر الكامن الأول – الجذر الكامن الثاني) / (الجذر الكامن الثاني – الجذر الكامن الثالث)
	50	1	25.33	[1]**50.66**	50.66	[2]**13.46**	[3]**41.26**
		2	1.88	3.76	54.42		
		3	1.31	2.63	[4]**57.05**		
		4					

يظهر الجدول (2) أن كافة قيم التباين المفسر (المؤشر الأول) لجميع المواقف البحثية قد تخطت قيمة 20% كمؤشر أول لأحادية البعد، وتحقق المؤشر الثاني لأحادية البعد الناتج عن حاصل قسمة الجذر الكامن الأول على الجذر الكامن الثاني حيث تخطت القيم العدد 2 لجميع المواقف البحثية، وكذلك أظهرت عملية قسمة حاصل طرح الجذر الكامن الثاني من الجذر الكامن الأول على حاصل طرح الجذر الكامن الثالث من الجذر الكامن الثاني نتائج ضخمة، مما يشير إلى تحقق إفتراض أحادية البعد كمؤشر ثالث لجميع المواقف البحثية، ويلاحظ أن قيم التباين المفسر التراكمي تزيد عن 52.26 % مما يشير إلى تحقق إفتراض أحادية البعد كمؤشر رابع لجميع المواقف البحثية (Hatti, 1984). والشكل (1) يبين نتائج التحليل العاملي لبيانات الدراسة وفقاً لكافة المواقف البحثية المشمولة بالدراسة بإستخدام الجذور الكامنة وعدد العوامل كمؤشر على تحقق أحادية البعد.



الشكل (1): تمثيل Scree Plot لنتائج التحليل العاملي لبيانات الدراسة وفقاً لكافة المواقف البحثية المشمولة بالدراسة.

ب- التحقق من إفتراض الاستقلال الموضعي Local Independence: بما أن أحادية البعد متحققة، فإن الاستقلال الموضعي متحقق (Hambleton & swaminathan, 1985).

ج- نتائج التحقق من مطابقة الأفراد والفقرات وفقًا للنموذج ثلاثي المعلمة في نظرية الاستجابة للفقرة: فيما يتعلق بمطابقة أفراد العينة للنموذج ثلاثي المعلّمة وفق نظرية الاستجابة للفقرة، تم القيام بإجراء مطابقة لأفراد العينة بإستخدام برنامج (BILOG–MG 3.0)، بالاعتماد على استجابات أفراد العينة على نموذج الإختباري المعدّ (25 فقرة، 50 فقرة)؛ إذ يُعدّ هذا الإجراء الخطوة الأولى من خطوات معايرة (تقدير) معلمة القدرة ومعالم الفقرات في نظرية الاستجابة للفقرة، وذلك لاختيار الأفراد المطابقين للنموذج اللوجستي ثلاثي المعلّمة، حيث تم التحقق من مدى مطابقة الأفراد للنموذج ثلاثي المعلمة عند مستوى الدلالة ($\alpha \geq 0.01$)، والجدول (3) يبين عدد الأفراد غير المطابقين وأرقامهم في مختلف المواقف البحثية المشمولة بالدراسة والمترتبة عن تفاعل متغيرات (شكل التوزيع، طول الإختبار، طريقة التقدير).

الجدول (3): عدد الأفراد غير المطابقين وأرقامهم في مختلف المواقف البحثية المشمولة بالدراسة.

ملتو لليمين				طبيعي				ملتو لليسار			
50		25		50		25		50		25	
MLE	EAP	MLE	EAP	MLE	EAP	MLE	EAP	MLE	EAP	MLE	EAP
27	19	21	3	23	17	15	1	42	10	47	3
1	70	85	85	41	41	94	133	475	364	2	423
70	77	193	155	76	62	523		563	475	4	428
77	97	532	193	79	76	616		606	563	5	759
97	101	549		93	79	675		681	681	8	
101	124	590		153	141	717		682	682	10	
124	140	636		159	159	862		769	684	14	
140	151	677		172	172	868		807	705	17	
164	164	728		175	175	873		887	738	35	
166	166	754		225	225	880		920	807	39	
192	171	764		272	245	883		935	902	76	
209	177	783		290	272	888		938		99	
292	192	787		317	290	893		945		125	
319	209	789		402	317	896		950		289	
346	292	794		430	402	903		961		423	
350	319	800		627	477	909		966		640	
399	332	812		885	907			968		711	
542	346	826		891	911			970		713	
837	350	843		893				971		759	

ملتو لليسار				طبيعي				ملتو لليمين			
50		25		50		25		50		25	
MLE	EAP	MLE	EAP	MLE	EAP	MLE	EAP	MLE	EAP	MLE	EAP
42	10	47	3	23	17	15	1	27	19	21	3
973		771		907				840	399	872	
975		773		911				859		886	
976		776		993				868		887	
977		780		994				874			
978		787		1000				887			
979		790						890			
980		791						891			
981		797						902			
982		803						914			
983		804									
984		806									
985		807									
986		815									
987		818									
990		819									
991		820									
992		821									
993		824									
994		826									
996		827									
997		829									
998		830									
999		832									
1000		833									
		834									
		835									
		889									
		951									
		958									

علمًا أنَّ الأفراد غير المطابقين قد تم تصنيفهم على أنهم غير مطابقين وفقًا لمعيارين؛ هما: أن تكون احتمالية مطابقة الفرد أقل من مستوى الدلالة المعتمد أثناء عملية تقدير معلمة القدرة وهو (0.01)، وأن يكون قد تم تعذر حساب الخطأ المعياري لقدرة الفرد عند أي مستوى قدرة، بما يفيد أن الفرد غير المطابق وفقًا للمعيار الثاني لم يقدم أية معلومات كأن يكون تحصيله كاملًا أو صفرًا وذلك في حالة إستخدام طريقة MLE.

وفيما يتعلق بمطابقة فقرات الإختبار للنموذج ثلاثي المعلّمة وفق نظرية الاستجابة للفقرة، تم القيام بإجراء مطابقة لفقرات الإختبار بإستخدام برنامج(Bilog-MG 3.0)، بالإعتماد على استجابات أفراد العينة على نموذج الإختبار المعدّ؛ إذ يُعدّ هذا الإجراء الخطوة الثانية من خطوات تقدير معلمة القدرة ومعالم الفقرات في نظرية الاستجابة للفقرة، وذلك لاختيار الفقرات المطابقة للنموذج ثلاثي المعلمة، حيث أفرزت نتائج التحليل مطابقة جميع فقرات الإختبار للنموذج ثلاثي المعلمة عند مستوى الدلالة ($\alpha \geq 0.01$)، باستثناء الفقرات التي انخفضت احتمالية مطابقة دون مستوى (0.01)، والجدول (4) يبين عدد الفقرات غير المطابقة وأرقامها في مختلف المواقف البحثية المشمولة بالدراسة والمترتبة عن تفاعل متغيرات (شكل التوزيع، طول الإختبار، طريقة التقدير).

الجدول (4): عدد الفقرات غير المطابقة وأرقامها في مختلف المواقف البحثية المشمولة بالدراسة.

شكل التوزيع	طول الإختبار	طريقة التقدير	الفقرات غير المطابقة	
			العدد	الرقم
ملتوٍ لليسار	25	EAP	1	17
		MLEL		
	50	EAP		
		MLEL	1	23
طبيعي	25	EAP	2	12 ،2
		MLEL	2	22 ،2
	50	EAP	2	21 ،13
		MLEL	2	21 ،13
ملتوٍ لليمين	25	EAP	6	25 ،22 ،20 ،12 ،10 ،2
		MLEL	3	25 ،22 ،20
	50	EAP	6	50 ،45 ،40 ،36 ،31 ،13
		MLEL	6	50 ،29 ،48 ،45 ،36 ،13

وللوقوف على قيم مؤشر ($\chi 2$) لحسن مطابقة الفقرة للنموذج ثلاثي المعلمة من عدمها وفقًا لتقديرات برنامج (BILOG-MG3) لمعالم التمييز والصعوبة والتخمين لفقرات الإختبار.

ولأغراض توليد البيانات بما يتناسب مع كل موقف بحثي مشمول بالدراسة؛ فقد تم تزويد برنامج WINGEN 3 بقيم قدرات الأفراد المقدرة للبيانات الحقيقية بشكل يتلازم مع تزويد البرنامج بقيم معالم الصعوبة والتمييز والتخمين المقدرة لكل فقرة من فقرات البيانات الحقيقية.

والجدولان (5) و(6) يبينان الإحصائيات الوصفية لمعلمة القدرة ومعالم الفقرات بعد حذف الأفراد غير المطابقين وبعد حذف الفقرات غير المطابقة في مختلف المواقف البحثية المشمولة بالدراسة والمترتبة عن تفاعل متغيرات (شكل التوزيع، طول الإختبار، طريقة التقدير).

الجدول (5): قيم الإحصائيات الوصفية لمعلمة القدرة بعد حذف الأفراد غير المطابقين في مختلف المواقف البحثية المشمولة بالدراسة.

شكل التوزيع	طول الإختبار	طريقة التقدير	البيانات	إحصائيات القدرة				
				العدد	القيمة الصغرى	المتوسط الحسابي	الانحراف المعياري	القيمة العظمى
ملتو لليسار	25	EAP	حقيقية	997	-2.62	0.03	0.78	1.81
			مولدة	997	-2.47	0.03	0.88	1.85
		MLE	حقيقية	953	-3.83	0.08	0.85	2.76
			مولدة	953	-3.66	0.03	1.08	2.97
	50	EAP	حقيقية	990	-2.56	0.05	0.87	2.24
			مولدة	990	-2.62	0.04	0.90	2.44
		MLE	حقيقية	958	-3.99	0.13	0.90	3.30
			مولدة	958	-3.89	0.07	0.98	3.33
طبيعي	25	EAP	حقيقية	999	-1.44	0.03	0.73	1.91
			مولدة	999	-1.28	0.01	0.80	2.43
		MLE	حقيقية	985	-2.38	0.14	0.78	1.93
			مولدة	985	-3.91	-0.07	1.26	2.73
	50	EAP	حقيقية	983	-1.63	0.05	0.79	1.66
			مولدة	983	-1.68	0.02	0.85	2.24
		MLE	حقيقية	977	-3.96	0.01	1.07	1.85
			مولدة	977	-3.25	0.05	0.99	2.41
ملتو لليمين	25	EAP	حقيقية	997	-1.34	0.02	0.84	2.51
			مولدة	997	-1.10	0.00	0.79	3.01
		MLE	حقيقية	979	-3.71	-0.03	1.18	3.03
			مولدة	979	-3.62	-0.03	1.22	3.33
	50	EAP	حقيقية	981	-1.52	0.01	0.85	2.44
			مولدة	981	-1.35	0.00	0.85	2.77
		MLE	حقيقية	973	-3.88	-0.06	1.15	2.19
			مولدة	973	-3.93	-0.05	1.15	2.60

الجدول (6): قيم الإحصائيات الوصفية لمعالم الفقرات بعد حذف الفقرات غير المطابقة في مختلف المواقف البحثية المشمولة بالدراسة.

شكل التوزيع	طول الإختبار	طريقة التقدير	الإحصائي	معالم الفقرة					
				التمييز	الخطأالمعياري للتميز	الصعوبة	الخطأالمعياري للصعوبة	التخمين	الخطأالمعياري للتخمين
ملتو لليسار	25	EAP	عدد الفقرة المطابقة	24					
			القيمة الصغرى	0.808	0.134	-2.337	0.058	0.143	0.035
			المتوسط الحسابي	1.526	0.220	-0.424	0.182	0.219	0.056
			الانحراف المعياري	0.56	0.07	1.01	0.11	0.04	0.01
			القيمة العظمى	3.148	0.390	2.557	0.528	0.304	0.085
		MLE	عدد الفقرة المطابقة	25					
			القيمة الصغرى	0.763	0.112	-2.946	0.053	0.150	0.031
			المتوسط الحسابي	1.425	0.223	-0.417	0.231	0.273	0.062
			الانحراف المعياري	0.58	0.07	1.18	0.15	0.06	0.02
			القيمة العظمى	2.985	0.358	2.964	0.593	0.387	0.101
	50	EAP	عدد الفقرة المطابقة	50					
			القيمة الصغرى	0.714	0.129	-1.488	0.032	0.080	0.022
			المتوسط الحسابي	1.744	0.230	-0.049	0.118	0.167	0.040
			الانحراف المعياري	0.64	0.08	0.68	0.06	0.05	0.01
			القيمة العظمى	3.899	0.490	1.832	0.332	0.289	0.077
		MLE	عدد الفقرة المطابقة	49					
			القيمة الصغرى	0.682	0.133	-1.826	0.035	0.093	0.023
			المتوسط الحسابي	1.664	0.230	-0.107	0.136	0.200	0.045
			الانحراف المعياري	0.63	0.08	0.70	0.07	0.05	0.01
			القيمة العظمى	3.597	0.493	1.564	0.402	0.321	0.085
طبيعي	25	EAP	عدد الفقرة المطابقة	23					
			القيمة الصغرى	0.884	0.158	0.004	0.029	0.108	0.016
			المتوسط الحسابي	1.878	0.399	1.347	0.256	0.170	0.029
			الانحراف المعياري	1.02	0.20	0.87	0.35	0.05	0.01
			القيمة العظمى	4.836	0.847	3.641	1.553	0.251	0.053
		MLE	عدد الفقرة المطابقة	23					
			القيمة الصغرى	0.902	0.177	0.066	0.048	0.102	0.017
			المتوسط الحسابي	1.795	0.396	1.283	0.240	0.169	0.030
			الانحراف المعياري	0.82	0.20	0.83	0.32	0.04	0.01
			القيمة العظمى	3.638	0.847	3.592	1.498	0.243	0.053

شكل التوزيع	طول الإختبار	طريقة التقدير	الإحصائي	معالم الفقرة					
				التمييز	الخطأالمعياري للتميز	الصعوبة	الخطأالمعياري للصعوبة	التخمين	الخطأالمعياري للتخمين
ملتو لليمين	50	EAP	عدد الفقرة المطابقة	48					
			القيمة الصغرى	0.715	0.185	0.085	0.035	0.125	0.020
			المتوسط الحسابي	1.507	0.357	1.525	0.325	0.196	0.032
			الانحراف المعياري	0.63	0.14	1.00	0.38	0.03	0.01
			القيمة العظمى	4.149	0.639	4.249	1.831	0.258	0.068
		MLE	عدد الفقرة المطابقة	48					
			القيمة الصغرى	0.766	0.187	0.083	0.035	0.134	0.020
			المتوسط الحسابي	1.578	0.385	1.528	0.324	0.207	0.032
			الانحراف المعياري	0.64	0.15	0.99	0.39	0.03	0.01
			القيمة العظمى	4.159	0.704	4.461	1.881	0.274	0.064
	25	EAP	عدد الفقرة المطابقة	19					
			القيمة الصغرى	1.292	0.207	0.409	0.062	0.134	0.014
			المتوسط الحسابي	2.068	0.369	1.296	0.110	0.233	0.026
			الانحراف المعياري	0.63	0.15	0.49	0.04	0.06	0.01
			القيمة العظمى	3.398	0.784	2.141	0.216	0.368	0.042
		MLE	عدد الفقرة المطابقة	22					
			القيمة الصغرى	1.281	0.200	0.198	0.061	0.144	0.014
			المتوسط الحسابي	2.108	0.372	1.172	0.107	0.239	0.028
			الانحراف المعياري	0.69	0.18	0.53	0.04	0.06	0.01
			القيمة العظمى	3.690	0.969	2.070	0.203	0.366	0.049
	50	EAP	عدد الفقرة المطابقة	44					
			القيمة الصغرى	1.137	0.196	0.086	0.042	0.121	0.015
			المتوسط الحسابي	2.577	0.501	1.178	0.102	0.227	0.024
			الانحراف المعياري	1.16	0.35	0.51	0.06	0.05	0.01
			القيمة العظمى	7.205	1.890	2.396	0.382	0.316	0.039
		MLE	عدد الفقرة المطابقة	44					
			القيمة الصغرى	0.973	0.168	-0.030	0.044	0.151	0.015
			المتوسط الحسابي	2.191	0.437	1.257	0.124	0.230	0.027
			الانحراف المعياري	0.94	0.30	0.63	0.08	0.05	0.01
			القيمة العظمى	5.558	1.768	2.914	0.571	0.324	0.051

وللوقوف على قيم معالم التمييز والصعوبة والتخمين لفقرات الإختبار للنموذج ثلاثي المعلمة بعد حذف الأفراد والفقرات غير المطابقة وفقًا لتقديرات برنامج (BILOG-MG3).

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 22 [2021], Iss. 2, Art. 22

نتائج الدراسة ومناقشتها:

هدفت الدراسة إلى المقارنة بين دقة تقدير القدرة بإختلاف طول الإختبار وشكل توزيع معلمة القدرة تبعا للنموذج اللوجستي ثلاثي المعلمة بإستخدام بيانات حقيقية تمثل نتائج الإختبار الوطني لمادة العلوم للصف الثامن الأساسي والمعد من قبل إدارة الامتحانات والإختبارات في وزارة التربية والتعليم الأردنية، وبيانات مولدة تحاكي خصائص تلك البيانات الحقيقية.

أولًا. للإجابة عن سؤال الدراسة الأول الذي نصَّ على: هل تختلف دقة تقديرات القدرة بإختلاف عدد فقرات الإختبار في ظل النموذج اللوجستي ثلاثي المعلمة للبيانات الحقيقة والمولدة؟ فقد تم حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP)، وفقًا للبيانات الحقيقية والمولدة، عندما يكون توزيع القدرة طبيعيًا وطول الإختبار (25 فقرة، 50 فقرة)، بالإضافة إلى إستخدام اختبار t للعينات المترابطة للكشف عن جوهرية الفرق الظاهري بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP) وفقًا للبيانات الحقيقية والمولدة عندما يكون توزيع القدرة طبيعيًا وطول الإختبار (25 فقرة، 50 فقرة)، وذلك كما هو مبين في الجدول (7).

الجدول (7): نتائج اختبار t للعينات المترابطة للمتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP) وفقًا للبيانات الحقيقية والمولدة عندما يكون توزيع القدرة طبيعيًا وطول الإختبار (25 فقرة، 50 فقرة).

طول الإختبار	طريقة التقدير	البيانات	العدد	المتوسط الحسابي	الانحراف المعياري	الارتباط	قيمة ت المحسوبة	درجة الحرية	الدلالة الإحصائية	حجم الأثر
25	EAP	حقيقية	999	**0.443**	0.15	***0.57**	***-33.125**	998	0.000	-0.033
		مولدة	999	0.577	0.11					
	MLE	حقيقية	985	**0.537**	0.46	***0.48**	***-11.988**	984	0.000	-0.012
		مولدة	985	1.399	2.44					
50	EAP	حقيقية	983	**0.382**	0.17	***0.63**	***-19.178**	982	0.000	-0.020
		مولدة	983	0.466	0.14					
	MLE	حقيقية	977	0.840	1.63	***0.46**				
		مولدة	977	**0.628**	0.57		***4.549**	976	0.000	0.005
	* دالة عند مستوى الدلالة ($\alpha = 0.05$).									

https://digitalcommons.aaru.edu.jo/jpu/vol22/iss2/22

يتضح من الجدول (7) وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP)،وفقًا للبيانات الحقيقية والمولدة عندما يكون توزيع القدرة طبيعيًا وطول الإختبار (25 فقرة)؛ لصالح عملية التقدير بالاعتماد على البيانات الحقيقية مقارنة بعملية التقدير بالاعتماد على البيانات المولدة. كما يتضح من الجدول (7) وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (EAP)، وفقًا للبيانات الحقيقية والمولدة عندما يكون توزيع القدرة طبيعيًا وطول الإختبار (50 فقرة)؛ لصالح عملية التقدير بالاعتماد على البيانات الحقيقية مقارنة بعملية التقدير بالاعتماد على البيانات المولدة. وكذلك يتبين وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (MLE)، وفقًا للبيانات الحقيقية والمولدة عندما يكون توزيع القدرة طبيعيًا وطول الإختبار (50 فقرة)؛ لصالح عملية التقدير بالاعتماد على البيانات المولدة مقارنة بعملية التقدير بالاعتماد على البيانات الحقيقية.

وتشير النتائج السابقة إلى أن البيانات الحقيقية قدمت أدق التقديرات لمعلمة القدرة في حالة توزيع القدرة توزيعا طبيعيا وطول الإختبار (25 فقرة) مع إستخدام طريقتي (MLE, EAP)، بينما تفاوتت دقت التقديرات لمعلمة القدرة والتي تقدمها البيانات الحقيقية والمولدة في حالة توزيع القدرة توزيعا طبيعيا وطول الإختبار (50 فقرة) مع إستخدام طريقتي (MLE, EAP)، كما يتبين مما سبق بأن البيانات الحقيقية قدمت أدق التقديرات حول معلمة القدرة في حالة توزيع القدرة توزيعا طبيعيا وذلك مع إستخدام طريقة (EAP) بإختلاف طول الإختبار (25،50) فقرة، بينما تفاوتت دقت التقديرات حول معلمة القدرة والتي تقدمها البيانات الحقيقية والمولدة عندا تتوزع القدرة توزيعا طبيعيا وذلك مع إستخدام طريقة (MLE).

كما ويظهر من الجدول (7) بان طريقة (EAP) قدمت تقديرات لمعلمة القدرة بشكل أدق من طريقة (MLE) وذلك لانخفاض قيمة المتوسط الحسابي لخطأ التقدير في حالة (25،50) فقرة لكل من البيانات الحقيقية والمولدة مقارنة مع قيمة المتوسط الحسابي لخطا التقدير في حالة (25،50) فقرة لكل من البيانات الحقيقية والمولدة بإستخدام طريقة (MLE)، كما وازدادت دقة التقديرات لمعلمة القدرة بزيادة عدد فقرات الإختبار بإستخدام طريقة (EAP) لكل من البيانات الحقيقية والمولدة، وتحقق ذلك فقط بالنسبة للبيانات المولدة مع إستخدام طريقة (MLE)، كما يلاحظ أن الأخطاء المعيارية للتقدير متقاربة لطريقة (EAP) لكل من البيانات الحقيقية والمولدة في حالة (25،50) فقرة، بينما لم يتحقق ذلك مع طريقة (MLE) خاصة في حالة (25) فقرة.

وعلى الرغم من أن الدلالة الإحصائية للفروق بين متوسطات الأخطاء المعيارية في تقدير القدرة والتي تعبر عن دقة التقديرات لمعلمة القدرة والناتجة عن إستخدام بيانات حقيقية وبيانات مولدة، عندما تتوزع القدرة توزيعا طبيعيا تختلف بإختلاف طول الإختبار، خاصة مع إستخدام عدد كبير من الفقرات (50) فقرة، كما وتتفاوت دقة التقديرات الناتجة عن بيانات حقيقية ومولدة بإختلاف الطريقة المستخدمة في تقدير القدرة خاصة مع إستخدام طريقة (MLE)، إلا أن حجم الأثر أشار إلى عدم وجود دلالة عملية لهذا الاختلاف بين دقة التقديرات الناتجة عن بيانات حقيقية ومولدة، وعليه يمكن الاستنتاج بعدم وجدود فروق ذات دلالة عملية بين دقة التقديرات لمعلمة القدرة والناتجة عن إستخدام بيانات حقيقية ومولدة بإختلاف طول الإختبار.

وتتفق هذه النتائج مع دراسة لندن (Linden, 1998) والتي أشارت إلى أفضلية طريقة بيز على طريقة الأرجحية العظمى في تقدير القدرة خاصة مع إستخدام عدد قليل من الفقرات؛ ودراسة وانغ وفيسبول (Wang & Vispoel, 1998) والتي أشارت إلى طريقة الارجحية العظمى تعطي أخطاء عالية في التقدير مقارنة بطرق بيز؛ ودراسة (عبابنة، 2004) والتي أشارت إلى ازدياد دقة التقدير لمعلم القدرة بزيادة عدد الفقرات.

ثانيًا. للإجابة عن سؤال الدراسة الثاني الذي نصَّ على: هل تختلف دقة تقديرات القدرة بإختلاف شكل توزيع معلمة القدرة وطول الإختبار في ظل النموذج اللوجستي ثلاثي المعلمة للبيانات الحقيقية والمولدة؟ فقد تم حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 50 فقرة على اختلاف توزيع القدرة (ملتوٍ لليسار، طبيعي، ملتوٍ لليمين)، بالإضافة إلى إستخدام اختبار t للعينات المترابطة للكشف عن جوهرية الفرق الظاهري بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 50 فقرة على اختلاف توزيع القدرة (ملتوٍ لليسار، طبيعي، ملتوٍ لليمين)، وذلك كما هو مبين في الجدول (8).

الجدول (8): نتائج اختبار t للعينات المترابطة للمتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقتي (MLE, EAP) وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 25، 50 فقرة على اختلاف توزيع القدرة (ملتوٍ لليسار، طبيعي، ملتوٍ لليمين).

توزيع القدرة	طول الإختبار	طريقة التقدير	البيانات	العدد	المتوسط الحسابي	الانحراف المعياري	الارتباط	قيمة ت المحسوب	درجة الحرية	الدلالة الاحصائية	حجم الاثر
ملتوي لليسار	25	EAP	حقيقية	997	**0.376**	0.09	**0.341**	**-25.121**	996	0.000	-0.02
			مولدة	997	0.410	0.12					
		MLE	حقيقية	953	**0.449**	0.23	***0.266**	**13.989**	952	0.000	0.014
			مولدة	953	0.566	0.18					
	50	EAP	حقيقية	990	0.214	0.16	**0.47**	0.255	989	0.799	0.008
			مولدة	990	0.212	0.15					
		MLE	حقيقية	958	**0.303**	0.37	**0.76**	**-3.179**	957	0.002	-0.09
			مولدة	958	0.329	0.22					
طبيعي	25	EAP	حقيقية	999	**0.443**	0.15	**0.57**	**-23.151**	998	0.000	-0.03
			مولدة	999	0.577	0.11					
		MLE	حقيقية	985	**0.537**	0.46	**0.48**	**-11.988**	984	0.000	-0.01
			مولدة	985	1.399	2.44					
	50	EAP	حقيقية	983	**0.382**	0.17	**0.63**	**-19.17**	982	0.000	-0.02
			مولدة	983	0.466	0.14					
		MLE	حقيقية	977	0.840	1.63	**0.46**	**4.549**	976	0.000	0.005
			مولدة	977	**0.628**	0.57					
ملتوي لليمين	25	EAP	حقيقية	997	**0.472**	0.17	**0.556**	**36.552**	996	0.000	0.036
			مولدة	997	0.638	0.11					
		MLE	حقيقية	979	**1.237**	1.87	**0.139**	**-3.110**	978	0.002	0.00
			مولدة	979	1.753	5.08					
	50	EAP	حقيقية	981	**0.363**	0.22	**0.69**	**-12.70**	980	0.000	-0.01
			مولدة	981	0.431	0.20					
		MLE	حقيقية	973	1.366	4.12	**0.17**	1.014	972	0.311	0.001
			مولدة	973	1.215	2.91					

* دالة عند مستوى الدلالة ($\alpha \geq 0.05$).

يتضح من الجدول (8) عدم وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (EAP)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 50 فقرة وتوزيع القدرة (ملتوٍ لليسار). كما يتضح من الجدول (8) عدم وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (MLE)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 50 فقرة وتوزيع القدرة ملتوٍ لليمين، وكذلك وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (EAP)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 25,50 فقرة على اختلاف توزيع القدرة (طبيعي، ملتوٍ لليمين، ملتوي لليسار)؛ لصالح عملية التقدير بالاعتماد على البيانات الحقيقية مقارنة بعملية التقدير بالاعتماد على البيانات المولدة إلا في حالة 25 فقرة في التوزيع الملتوي لليسار.

في حين يتضح من الجدول(8) وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (MLE)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 50,25 فقرة وتوزيع القدرة (ملتوٍ لليسار)؛ لصالح عملية التقدير بالاعتماد على البيانات الحقيقية مقارنة بعملية التقدير بالاعتماد على البيانات المولدة، وكذلك الحال في حالة 25 فقرة وتوزيع القدرة ملتوي لليمين.

وأخيرًا؛ يتضح من الجدول(8) وجود فرق دال إحصائيًا عند مستوى الدلالة $\alpha \leq 0.05$ بين المتوسطين الحسابيين للأخطاء المعيارية في تقدير معلمة القدرة للطلبة بإستخدام طريقة (MLE)، وفقًا للبيانات الحقيقية والمولدة عندما يكون طول الإختبار 50 فقرة وتوزيع القدرة (طبيعي)؛ لصالح عملية التقدير بالاعتماد على البيانات المولدة مقارنة بعملية التقدير بالاعتماد على البيانات الحقيقية.

تشير النتائج السابقة إلى أنه عندما يكون طول الإختبار كبيراً نسبيا 50 فقرة فإن البيانات الحقيقية قدمت أدق التقديرات حول معلمة القدرة مع إستخدام طريقة (EAP) في حالة التوزيع (الطبيعي، ملتوي لليمين) لمعلمة القدرة، لكن تفاوتت دقة التقديرات التي تقدمها البيانات الحقيقية والمولدة حول معلمة القدرة مع إستخدام (MLE) في حالة التوزيع (الطبيعي، ملتوي لليسار) لمعلمة القدرة، كما أنه لا يوجد فرق دال إحصائيا بين دقة التقديرات الناتجة عن بيانات حقيقية وبين تلك الناتجة عن بيانات مولدة مع إستخدام طريقة (EAP) في حالة التوزيع الملتوي لليسار لمعلمة القدرة، وكذلك مع إستخدام (MLE) في حالة التوزيع الملتوي لليمين، وهذا يشير إلى

عدم انتظام في الأخطاء المعيارية والناتجة عن إستخدام البيانات الحقيقية والمولدة بإختلاف شكل التوزيع مع إستخدام طريقتي (MLE, EAP) مع إستخدام (50,25) فقرة.

وعلى الرغم من أن الدلالة الإحصائية أشارت إلى تفاوت دقة تقديرات القدرة تبعا للبيانات الحقيقية والمولدة بإختلاف شكل توزيع معلمة القدرة خاصة مع إستخدام طريقة (MLE)، إلا أن حجم الأثر أشار إلى عدم وجود دلالة عملية لهذا الاختلاف بين دقة التقديرات الناتجة عن بيانات حقيقية ومولدة، حيث كانت قيم الدلالة العملية اقل من 0.2 والتي تشير الى دلالة ضعيفة (Cohen, 1977)، وعليه يمكن الاستنتاج بعدم وجدود فروق بين دقة التقديرات لمعلمة القدرة والناتجة عن إستخدام بيانات حقيقية ومولد بإختلاف شكل التوزيع لمعلمة القدرة مع اختلاف عدد الفقرات (50,25) فقرة.

ويظهر مما سبق بان طريقة (EAP) قدمت أدق التقديرات حول معلمة القدرة بإختلاف شكل توزيع معلمة القدرة وهذا يتفق مع دراسة (الطراونة، 2012)؛ ودراسة (عبابنة، 2006)؛ ودراسة جاري وفيرمنت (Garre & Vermunt, 2006).

التوصيات:

إجراء المزيد من الدراسات حول المقارنة بين دقة التقديرات لمعلمة القدرة ومعالم الفقرة، والناتجة من إستخدام بيانات حقيقية ومولدة بإستخدام تصاميم مختلفة لمعرفة أثر تغير بعض العوامل على دقة التقديرات الناتجة، وإستخدام طريقة (EAP) في تقدير القدرة نظرا لانخفاض متوسط الأخطاء المعيارية في التقدير.

المصادر والمراجع:

المراجع العربية:

التقي، احمد، (2009). **النظرية الحديثة في القياس**. عمان: دار المسيرة للنشر.

الشريفين، نضال، (2012). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغيُّر حجم العينة. **المجلة التربوية**- مجلس النشر العلمي، 26 (104)،177-238، جامعة الكويت، الكويت.

Jerash for Research and Studies Journal ???? ??? ?????? ?????????, Vol. 22 [2021], Iss. 2, Art. 22

الطراونة، أرياف احمد، (2011). **المقارنة بين طرق تقدير القدرة بإستخدام النموذج المناسب في ضوء الخطأ المعياري في تقديرها**. رسالة ماجستير غير منشورة، جامعة مؤتة، الكرك، الأردن.

الطريري، عبد الرحمن، (1997). **القياس النفسي والتربوي نظرياته أسسه وتطبيقاته.** السعودية، الرياض: مكتبة الرشد للنشر والتوزيع.

طشطوش، سليمان محمد، (2001). **أساسيات المعاينة الإحصائية**. الأردن، عمان دار الشروق للنشر والتوزيع.

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

عبابنة، عماد، (2006). مقارنة فاعلية طريقة الارجحية العظمى وطريقة بيز في تقدير معلمة القدرة عند إستخدام النموذج اللوجستي. **مجلة الاكاديمية العربية المفتوحة في الدنمارك**، (3)، 5-22.

علام، صلاح الدين، (2002). **القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجيهاته المعاصرة**. القاهرة: دار الفكر العربي.

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

عوده، احمد، (2004). **القياس والتقويم في العملية التدريسية**. اربد: دار الأمل للنشروالتوزيع.

كاظم، أمينة، (1988). **دراسة نظرية نقدية حول القياس الموضوعي للسلوك "نموذج راش"**. الكويت: سلسلة الكتب المتخصصة.

وزارة التربية والتعليم، (2013). **نتائج الإختبار الوطني لضبط نوعية التعليم**. إدارة الامتحانات والإختبارات، الأردن، عمان.

https://digitalcommons.aaru.edu.jo/jpu/vol22/iss2/22 28

المراجع الأجنبية:

Allen, M. J., and Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Baker, F. (2001). *The basic of item response theory*, (2nded). College Park, MD: EARIC Clearing House on Assessment and Evaluation.

Bock, R. D. and Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*. 6(4), 431- 444.

Chen, S. Hou & Dodd. (1997). the effect of population distribution and method of theta estimation on computerized adaptive testing. *Educational and Psychological Measurement*, 57(3), 422- 439.

Cohen, J. (1977). *Statistical Power analysis for the behavioral science*, New Yourk: Academic Press.

Crocker, L. and Algina, J. (1986*). Introduction to classical and modern test theory.*New Yourk: Holt Pine hart and Winston.

Garre, G. and Vermunt, K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation, *Behaviormentrika*, 33(1), 43-59.

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***,** 38-74

Hambleton, R.K. and Swaminathan, H. (1985). *Item response theory principles and application.* Boston: Kluwer-Nigh off Publishing

Hambleton, R.K., swaminathan, H. and Rogers, H. (1991). *Fundamental of item response theory*. New York: Stage publication the international professional publishers.

Harwell, M., Stone, Clement A., Hsu, T. C. and Kirisci, L. (1996). Monte Carlo studies item response theory. *Applied Psychological Measurement*, 20(2), 101-125.

Hatti, J. (1984). An empirical study of various indices for determining unidimentionality. *Multivariate Behavioral Research*, 19, 49-78.

Hulin, C. L., Drasgow, F. and Parsons, C.K. (1983). *Item response theory*. Homewood III. Dow Jon's-Irwin.

Kim, H. S. (1998). *An evaluation of a Markova chain Monte Carlo method for the rasch model*. Paper presented at the joint annual meeting of the

Psychometric Society and the classification Society of North America (Urbana, IL, June19, 1998).

keller, L.A. (2000). Ability estimation procedures in computerized adaptive testing. *American Institute of Certified Public Accountant* Newyourk, America.

Kline,T.J.B. (2005). *Psychological testing.* New Delhi: Sag Publications', Inc.

Linden, W. (1998). Bayesian item selection criteria of adaptive testing. *Psychometrika*, 63(2).

Liou, M., Sinica, A. and Yu, L.C. (1999).Assessing statistical accuracy in ability estimation, *Educational and Psychological Measurement*, 58(4), 569-595.

Pagano, R.R. (2009). *Understanding statistics in the behavioral sciences*, (9th ed).Belmont: wads worth engage learning.

Reckase, M.D. (1978). *A Comparison of the one- and three- parameter logistic models for item calibration [microform]*. Paper presented at the annual meeting of the American Educational Research Association. Ontario, Canada, march 27-31, 1978.

Wang, T. and Vispoel, W.P. (1998). Properties of ability estimate methods in computerized adaptive testing. *Journal of Educational Measurement***,** 35(2), 109-135.

List of Sources and references:

Ababneh, Imad. (2004). *The effect of sample size, method of selection, number of items and method of selecting them on the accuracy of item parameter estimation and ability to test mental ability using item response theory.* Unpublished PhD thesis, Amman Arab University for Graduate Studies, Amman, Jordan.

Ababneh, Imad. (2006). Comparing the effectiveness of the maximum likelihood method and Bayes method in estimating the ability parameter when using the logistic model. *The Arab Open Academy Journal in Denmark*, (3), 5-22.

Allam, Salah al-Din. (2002). *Educational and psychological measurement and evaluation: its basics, applications, and contemporary guidelines*. Cairo: Arab Thought House.

Allam, Salah al-Din. (2005). *One-dimensional and multi-dimensional test response models and their applications in psychological and educational measurement.* Cairo: Arab Thought House.

Allen, M.J. and Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Al Taki, Ahmed. (2009). *Modern theory of measurement*. Amman: Dar Al Masirah Publishing.

Al-Sharifin, Nidal. (2012). The effect of the method of estimating the parameters of the paragraph and the capabilities of individuals on the values of the paragraph parameters, and the psychometric properties of the test, in light of the change in the sample size**.** *Educational Journal - Scientific Publication Council*, 26 (104), 177-238, Kuwait University, Kuwait.

Al-Tarawneh, Aryaf Ahmed. (2011). *Comparison between the methods of estimating capacity using the appropriate model in light of the standard error in its estimation.* Unpublished Master's Thesis, Mutah University, Karak, Jordan.

Al-Tariri, Abdul Rahman. (1997). *Psychometric and educational theories, foundations and applications.* Saudi Arabia, Riyadh: Al-Rushd Library for Publishing and Distribution.

Baker, F. (2001). *The basic of item response theory*, (2nded). College Park, MD: EARIC Clearing House on Assessment and Evaluation.

Bock, R. D. and Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement.* 6(4), 431- 444.

Chen, S. Hou & Dodd. (1997). the effect of population distribution and method of theta estimation on computerized adaptive testing. *Educational and Psychological Measurement*, 57(3), 422- 439.

Cohen, J. (1977). *Statistical Power analysis for the behavioral science*, New Yourk: Academic Press.

Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*.New Yourk: Holt Pine hart and Winston.

Garre, G. and Vermunt, K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation, *Behaviormentrika*, 33(1), 43-59.

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***,** 38-74.

Hambleton, R. K. and Swaminathan, H. (1985). *Item response theory principles and application.* Boston: Kluwer-Nigh off Publishing

Hambleton, R.K., swaminathan, H. and Rogers, H. (1991). *Fundamental of item response theory***.** New York: Stage publication the international professional publishers.

Harwell, M., Stone, Clement A., Hsu, T. C. and Kirisci, L. (1996). Monte Carlo studies item response theory. *Applied Psychological Measurement*, 20(2), 101-125.

Hatti, J. (1984). An empirical study of various indices for determining unidimentionality. *Multivariate Behavioral Research*, 19, 49-78.

Hulin, C. L., Drasgow, F. and Parsons, C. K. (1983*). Item response theory*. Homewood III. Dow Jon's-Irwin.

Kazem, Amina. (1988). *A critical theoretical study on the objective measurement of behavior "Rush model"*. Kuwait: Specialized Book Series.

keller, L. A. (2000). Ability estimation procedures in computerized adaptive testing. *American Institute of Certified Public Accountant*, Newyourk, America.

Kline,T.J.B. (2005). *Psychological testing.* New Delhi: Sag Publications', Inc.

Linden, W. (1998). Bayesian item selection criteria of adaptive testing. *Psychometrika*, 63(2).

Liou, M., Sinica, A. and Yu, L.C. (1999).Assessing statistical accuracy in ability estimation, *Educational and Psychological Measurement*, 58(4), 569-595.

Ministry of Education. (2013). *The results of the national test to control the quality of education.* Administration of Exams and Tests, Jordan, Amman.

Odeh, Ahmed. (2004). *Measurement and evaluation in the teaching process.* Irbid: Dar Al-Amal for Publishing and Distribution.

Pagano, R. R. (2009). *Understanding statistics in the behavioral sciences*. (9th ed).Belmont: wads worth engage learning.

Reckase, M. D. (1978). *A Comparison of the one- and three- parameter logistic models for item calibration [microform]*. Paper presented at the annual meeting of the American Educational Research Association. Ontario, Canada, march 27-31, 1978.

Tashtoush, Suleiman Muhammad. (2001). *Basics of statistical sampling.* Jordan, Amman, Dar Al-Shorouk for Publishing and Distribution.

Wang, T. and Vispoel, W. P. (1998). Properties of ability estimate methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135.