تقصي دقة تقدير قدرة الأفراد باستخدام الطرق المعتمدة في برنامج Bilog_Mg وفق النموذج الأحادي البارامتر (نموذج راش)

عبد الكريم شرفاوي[1،*] نصيرة بن نابي[2]،
abdelkarim.cherfaoui@univ-alger2.dz
[1] مخبر التربية والصحة النفسية، جامعة الجزائر2، الجزائر [2] جامعة الجزائر2، الجزائر

Investigating of the accuracy of estimating a Person's ability by using methods in the program Bilog_Mg according to One- Parameter Model (Rasch Model)

Abdelkarim Cherfaoui[1,*] Nacera Benabi[2],
[1] Laboratory of Education and Mental Health, Algiers 2 University (Algeria),
[2] Algiers 2 University (Algeria),

تاريخ الاستلام: 15/يونيو/2025؛ تاريخ القبول: 28/أغسطس/2026؛ تاريخ النشر: 31/أغسطس/2026

ملخص. هدفت هذه الدراسة إلى تقصي مدى دقة تقدير قدرة الأفراد باستخدام طرق التقدير: طريقة الأرجحية العظمى (ML)، طريقة توقع التوزيع البعدي (EAP) وطريقة تعظيم الاقتران البعدي (MAP) المعتمدة في برنامج Bilog_Mg وفق النموذج الأحادي البارامتر (نموج راش) اعتمدا على الخطأ المعياري للتقدير، ولتحقيق أهداف الدراسة تم استخدام أسلوب المحاكاة للبيانات لتوليد بيانات ثنائية الإجابة (1،0) لاختبار مكون من 20 فقرة وفق نموذج راش، وثلاث مستويات لحجم عينة (1000،500،200) باستخدام برنامج (WinGen) بعد ذلك تم تقدير قدرة الأفراد والأخطاء المعيارية للتقدير تحت كل موقف من المواقف السابقة باستخدام طرق التقدير، وقد تم التوصل إلى النتائج التالية :

أعطت طريقة تعظيم الاقتران البعدي (MAP) أعلى درجة في دقة التقدير للقدرة، تليها طريقة توقع التوزيع البعد (EAP)، ثم طريقة الأرجحية العظمى (ML) وهذا بالنسبة للمستويات الثلاثة لحجم العينة (1000،500،200).

الكلمات الدالة. دقة التقدير، طرق التقدير، نموذج راش، الخطأ المعياري للتقدير، القدرة.

Abstract. This study set out to examine how accurately ability of persons can be estimated using three commonly applied statistical methods: Maximum Likelihood (ML), Expected a Posteriori (EAP), and Maximum A Posteriori (MAP). These methods, implemented through the Bilog_MG software, were assessed within the framework of the Rasch model, also known as the One-Parameter Model. To carry out the analysis, simulated data were generated to produce binary responses (0 or 1) for a 20-item test, based on three sample sizes—200, 500, and 1000—using the WinGen program. Each estimation method was then applied to these datasets to estimate ability of persons and evaluate the precision of these estimates through standard errors. The results showed that the MAP method provided the most precise ability estimates across all sample sizes, followed by EAP and then ML.

Keywords. Accuracy of Estimation; Estimation Methods; Rasch Model; Standard Error of Estimation; Ability.

* abdelkarim.cherfaoui@univ-alger2.dz

1. مقدمة

قام علماء القياس النفسي بعدة محاولات في سبيل تحقيق الموضوعية والدقة في تقدير السلوك، باعتباره خطوة مهمة في فهم الظاهرة السلوكية وضبطها والتحكم فيها، لذا فإن بناء الأداة، التي تصل بنا إلى التقدير الأمثل والموضوعي للسلوك من أهم أهداف العاملين في مجال القياس النفسي والتربوي، وفي سبيل تحقيق هذه الأهداف ظهرت عدة نظريات حاولت الوصول في تقديرها للقدرات أو السمات إلى أعلى درجات الدقة من أجل اتخاذ قرارات سليمة على مستوى الأفراد والمجتمع وهذا بالاعتماد على نماذج رياضية وطرق مختلفة لتقدير القدرة من بينها نجد كل من طريقة توقع التوزيع البعدي (EAP) وطريقة تعظيم الاقتران البعدي(MAP) اللتان تعتمدان على أسلوب بييز (Bayes) في التقدير، وطريقة الأرجحية العظمى (ML) وهي الطرق التي سيتم تسليط الضوء عليها في هذه الدراسة، والوصول الى مؤشرات عن دقة هذه التقديرات والقياسات تندرج تحت إطاران يستند كل اطار على أسس وافتراضات محددة في القياس (عبابنة، 2008، ص.394) هما النظرية الكلاسيكية Classical Test Theory(CTT)والنظرية الحديثة أو نظرية الاستجابة للمفردة Item Response Theory (IRT)، حيث تعتبر هذه الأخيرة تطورا مهما في ميدان القياس النفسي، فهي تعالج الكثير من قضايا القياس بشكل أكثر فعالية من النظرية الكلاسيكية وقد أشار كل من هاملتون وسوامنثان (Hambleton & Swaminathan,1985) إلى بعض المشكلات والقصور الذي تعاني منه النظرية الكلاسيكية كاعتمادها في استخراج إحصائيات الفقرة (الصعوبة، التمييز) على خصائص العينة المطبق عليها إذ تختلف هذه الإحصائيات باختلاف متوسط أفراد العينة، واعتماد درجات المفحوصين على خصائص المفردات التي يتكون منها الاختبار(أي غير مستقلة عن عينة فقرات الاختبار)، إضافة إلى افتراض أن تباين أخطاء القياس هو نفسه لمستويات القدرة جميعا، وقد حاولت نظرية الاستجابة للمفردة معالجة هذه المشكلات حيث تقوم على افتراض أن أداء الاشخاص على الاختبارات والمقاييس النفسية والتربوية يمكن تفسيره في ظل خاصية أو مجموعة من الخصائص مميزة لهذا الأداء يطلق عليها السمات(Traits)، وما يميز هذه النظرية في أن تقديرات المحصل عيها لمعالم المفردات من صعوبة وتمييز وتخمين أنها مستقلة عن مجموعة الأفراد التي تم استخدامه في عملية تقدير هذه المعالم، كذلك تقدير قدرة الاشخاص تكون هي الاخرى مستقلة عن مجموعة الفقرات (Hambleton and Jones, 1993,p.40) ولتحديد وتفسير العلاقة بين أداء الأفراد على مفردات الاختبار وبين السمات أو القدرات التي تقف وراء هذا الأداء ظهرت عدت نماذج من هذه النظرية فحسب عدد الابعاد التي تفسر الاداء نجد نماذج الاستجابة للمفردة أحادية البعد (Unidimensional) ونماذج متعددة الأبعاد (Multidimensionl)، ومن الناحية العملية نجد أن النماذج الأحادية البعد أكثرها شيوعا واستخداما وهذه النماذج تنقسم إلى قسمين: نماذج ثنائية التدريج (Dichotomous Models) ونماذج متعددة التدريج(Polytomous Models) فالنماذج ثنائية التدريج تعالج مفردات من نوع صفر

(0) واحد (1) أي إما أن تكون الإجابة صحيحة أو خاطئة ومن أكثر النماذج انتشارا نجد نموذج راش (Rasch Model)هو نموذج أحادي المعلمة(الصعوبة)، نموذج بيرنبوم (Birnbaum Model) الثنائي المعلم (الصعوبة، التمييز) ونموذج لورد(Lord's Model) الثلاثي المعلم (الصعوبة، التمييز، التخمين)، ويتوقف اختيار النموذج على نوع البيانات المراد تحليلها إضافة إلى تدخل اعتبارات أخرى مثل خبرة الباحث بالنموذج وتوفر البرامج الحاسوبية المناسبة ويعتبر برنامج Bilog_Mg من بين أكثر البرامج شهرة في تحليل النماذج ثنائية التدريج، والخطوة الأولى في استخدام نظرية الاستجابة للمفردة في بناء وتصميم الاختبارات هي تقدير بارامترات (Parameter estimations)التي يشتمل عليها النموذج وبما أن هذه البارامترات (بارامتر الصعوبة، التمييز، التخمين) وقدرة الأفراد غير معلومة والشيء الوحيد المعلوم هو إجابات الأفراد على المفردات فقط لذلك فان أساليب التقدير تهدف إلى تحديد قيمة القدرة لكل فرد مختبر وقيم بارامترات المفردة وهي عملية يصعب القيام بها يدويا لذلك نلجأ إلى البرامج الحاسوبية ومن بينها برنامج Bilog_Mg، ولمعرفة دقة هذه التقديرات يستخدم مؤشر الخطأ المعياري للتقدير كمؤشر لدقة القياس لأنه يبين مدى اقتراب درجة الفرد على المقياس أو الاختبار من درجته الحقيقية، حيث أنه كلما كانت قيمة القدرة دقيقة كان الخطأ المعياري في تقديرها صغيرا، ويعرف بأنه مقلوب الجذر التربيعي لدالة معلومات الاختبار (Baker,2001.P115)، وقد توصل بني عطا(2017) في دراسته عن وجود فروق في دقة التقدير بين الطرق المعتمدة في برنامج Bilog_Mg كما أن هذه التقديرات تتأثر بعدة عوامل أخرى مثل طول الاختبار، وحجم العينة، وبالرجوع إلى البيئة الجزائرية نجد أن الأبحاث والدراسات التي استخدمت برنامج Bilog_Mg في تدريج مفردات الاختبارات والمقاييس هي قليلة جدا، سعت هذه الدراسة إلى تقصي مدى دقة تقدير قدرة الأفراد باستخدام طرق التقدير المعتمدة في برنامج Bilog_Mg المتمثلة في طريقة الأرجحية العظمى (ML)، طريقة توقع التوزيع البعدي(EAP) وطريقة تعظيم الاقتران البعدي(MAP)، لاختبار مكون من 20 فقرة وفق النموذج الأحادي البارامتر (نموذج راش) عند حجم عينة مختلف إذ تم اختيار ثلاثة أحجام مختلفة عن طريق بيانات تحاكي الواقع للمواقف الاختبارية بهدف إعطاء أدلة إمبريقية عن أي طريقة من الطرق المستخدمة في البرنامج تعطي أدق التقديرات عند استخدام حجم عينة معين وبالتحديد فان هذه الدراسة هدفت للإجابة عن الأسئلة التالية:

1- هل توجد فروق بين المتوسطات الحسابية للأخطاء المعيارية لتقدير قدرة الأفراد تعزى لطريقة التقدير (ML، EAP، MAP) المستخدمة في برنامج Bilog_Mg عند حجم العينة 200 وفق النموذج الأحادي البارامتر (نموذج راش)؟.

2- هل توجد فروق بين المتوسطات الحسابية للأخطاء المعيارية لتقدير قدرة الأفراد تعزى لطريقة التقدير (ML، EAP، MAP) المستخدمة في برنامج Bilog_Mg عند حجم العينة 500 وفق النموذج الأحادي البارامتر (نموذج راش)؟.

3- هل توجد فروق بين المتوسطات الحسابية للأخطاء المعيارية لتقدير قدرة الأفراد تعزى لطريقة التقدير (ML، EAP، MAP) المستخدمة في برنامج Bilog_Mg عند حجم العينة 1000 وفق النموذج الأحادي البارامتر (نموذج راش)؟.

1.1. أهداف الدراسة.

سعت الدراسة إلي التعرف عن أي طريقة من طرق التقدير التي يعتمد عليها برنامج Bilog_Mg في تقدير قدرة الافراد (الأرجحية العظمى (ML)، توقع التوزيع البعدي (EAP)، تعظيم الاقتران البعدي (MAP)) تعطي أقل خطأ معياري للتقدير (أكثر دقة)، عند حجم العينة 200 ،500 و1000 فرد وفق نموذج راش.

2.1. أهمية الدراسة.

تكمن أهمية الدراسة الحالية في محاولة تقصي دقة تقدير قدرة الأفراد عند استخدام ثلاث طرق مختلفة للتقدير المعتمدة في برنامج Bilog_Mg في ضوء مستويات مختلفة من حجم العينة، مما قد يساهم في توفير أدلة إمبريقية تساعد متخذي القرارات، الباحثين ومصممي الاختبارات النفسية والتربوية في تحديد طريقة التقدير المناسبة للحصول على دقة اكبر في تقدير القدرة أو السمة المقاسة وبالتالي اتخاذ قرارات صائبة تتعلق بتصنيف لأفراد أو توجيههم أو إرشادهم، كما جاءت هذه الدراسة لتلقي الضوء على طرق القياس الحديثة في بناء الاختبارات في إطار نظرية الاستجابة للمفردة واستكمالا للجهود المبذولة للتحكم في المتغيرات التي تؤثر على دقة عملية القياس النفسي والتربوي .

3.1. حدود الدراسة.

- اقتصرت الدراسة على استخدام النموذج الأحادي البارامتر (نموذج راش).

- اقتصرت الدراسة على استخدام طريقة الأرجحية العظمى (ML)، طريقة توقع التوزيع البعدي (EAP)، وطريقة تعظيم الاقتران البعدي (MAP) المستخدمة في برنامج Bilog-Mg.

- أجريت هذه الدراسة باستخدام طريقة المحاكاة (Simulation)، بيانات مولدة باستخدام برنامج WinGenبتوزيع طبيعي($\mu=1$، $\sigma=0$) لمعلمة القدرة ومعلمة الصعوبة.

- اقتصرت الدراسة على استخدام أحجام عينات (200،500،1000) لاختبار واحد مكون من 20 فقرة.

4.1. تحديد مفاهيم الدراسة.

– **نموذج راش (Rasch Model)** : أحد النماذج اللوجستية والتي تتعلق بالفقرات ثنائية الاستجابة اقترحه عالم الرياضيات الدنماركي جورج راش (Georg Rash) سنة 1960، ويفترض هذا النموذج أن جميع المفردات لها نفس القدرة التمييزية بين مستويات القدرة، ولكن تختلف في مستوي صعوبتها. (علام، 2005، ص. 69)

– **دقة التقدير (Accuracy of Estimation) :** هو تعبير يشير إلى جودة التقدير من خلال اقتراب قيم المقدرة من القيم الحقيقية ويقاس بالخطأ المعياري للتقدير فكلما قل الخطأ المعياري زادت دقة التقدير والعكس صحيح.

– **الخطأ المعياري لتقدير (Standard Error of Estimation):** هو مؤشر لدقة التقدير ويساوي مقلوب الجذر التربيعي لدالة معلومات الاختبار (Baker,2001,P.115).

– **طرق تقدير القدرة(Ability Estimation Methods):** هي الطرق المستخدمة في برنامج Bilog-Mg لتقدير معلمة القدرة للفرد (طريقة الأرجحية العظمى (ML)، طريقة توقع التوزيع البعدي (EAP) وطريقة تعظيم الاقتران البعدي (MAP)).

– **معلم القدرة (Ability Parametre):** هو مستوى ما يمتلكه الافراد من القدرة أو السمة المقاسة من خلال استجاباتهم على مفردات الاختبار ويتم تقديرها وفقا للنموذج الأحادي البارامتر (نموذج راش) المستخدم في الدراسة.

2. الإطار النظري والدراسات السابقة.

تعتبر نظرية الاستجابة للمفردة (Item Response Theory) من التطورات الحديثة في مجال القياس والتقويم التربوي (علام،2000، ص.682) حيث عالجت الكثير من المشاكل في بناء المقاييس والاختبارات النفسية والتربوية مقارنة مع نظرية القياس التقليدية، وانبثق عنها مجموعة من النماذج التي تحدد العلاقة بين أداء الفرد على الفقرة والقدرة التي تقف وراء هذا الأداء ويمكن أن تحدد هذه العلاقة من خلال ما يسمى منحنى خاصية الفقرة (Item Characteristic curve, ICC)، حيث كلما زادت قدرة الفرد زاد احتمال الإجابة الصحيحة عن الفقرة ، وتصنف نماذج الاستجابة للمفردة ثنائية التدريج إلى:

1.2. النموذج الأحادي البارامتر (نموذج راش): One- Parameter Model (The Rasch Model)

نشر هذا النموذج سنة 1960 من طرف عالم الرياضيات الدنماركي جورج راش (1901– 1980) (Georg Rash)، ويفترض هذا النموذج أن جميع المفردات لها نفس القدرة التمييزية بين مستويات القدرة ولكن تختلف في مستوي صعوبتها (علام، 2005، ص. 69) ويعد هذا النموذج من أبسط نماذج الاستجابة للمفردة أحادية البعد وأكثرها شهرة ويطلق على هذا النموذج نموذج المعلم الواحد لأنه يشتمل على معلم الصعوبة فقط (item

(difficulties parameter، يعتمد نجاح أي شخص (S) في استجابته على أي مفردة اختباريه (i)على الفرق بين قدرة(θ) هذا الفرد (القدرة الكامنة وراء استجابته) ودرجة صعوبة المفردة (β) (تقي الدين،2013، ص. 19)، فإذا كانت استجابة الفرد على البند إجابة صحيحة عندها تمنح له درجة واحدة (01) ويرمز لها بالرمز ($X_{is}=1$) ويعبر عنها احتماليا بالعلاقة التالية:

$$P(X_{is}=1/\theta_s,\beta_i)=\frac{e^{(\theta_s-\beta_i)}}{1+e^{(\theta_s-\beta_i)}}\cdots\cdots\cdots\cdots\cdots\cdots(01)$$

وتعتمد أرجحيه حدوث أي من الاستجابتين (صحيحة أو خاطئة) على القدرة التي يمتلكها الفرد (θ_S) وصعوبة البند أو الفقرة (β_i) الذي قام بالإجابة عليه (كاظم، 1988، ص. 23 ; Hambleton, 1991. P 12).

2.2. النموذج اللوجستي الثنائي البارامتر (نموذج بيرنبوم):

Two- Parameter Logistic Model (Birnbaum's Model)

تم اقتراح هذا النموذج في جامعة كلومبيا الأمريكية من طرف عالم الإحصاء بيرنبوم (Birnboaum,1968) مع مجموعة من زملائه (علام، 2005،ص. 70) ، يعتمد هذا النموذج على معلمتين هما معلمة الصعوبة (Difficulty) ومعلمة التمييز (Discrimination) يمثلان خصائص المفردة حيث يفترض هذا النموذج أن المفردات تختلف في صعوبتها وتمييزها، يقوم هذا النموذج على الفرق بين القدرة التي يمتلكها الفرد (θ_S) ودرجة صعوبة البند أو الفقرة الذي قام بالإجابة عليها (β_i) مضروبا في درجة تمييز الفقرة (α_i) (تقي الدين، 2013، ص. 22) ويعبر عن احتمال الإجابة الصحيحة عن الفقرة في هذا النموذج بالصيغة التالية:

$$P(X_{is}=1/\theta_s,\beta_i,\alpha_i)=\frac{e^{\alpha_i(\theta_s-\beta_i)}}{1+e^{\alpha_i(\theta_s-\beta_i)}}\cdots\cdots\cdots\cdots\cdots\cdots(03)$$

3.2.النموذج اللوجستي الثلاثي البارامتر (نموذج لورد)

Three- Parameter Logistic Model (Lord's Model)

لم يأخذ النموذجين السابقين بعين الاعتبار ظاهرة التخمين حيث أن المفحوصين سيجيبون على بعض الفقرات إجابة صحيحة نتيجة التخمين (Guessing) وبالتالي فان احتمال الإجابة الصحيحة تشتمل على جزء صغير يعود إلى التخمين (Baker, 2001, p. 28) ، لذلك أضاف هذا النموذج معلما ثالثا(c_i) الخط التقاربي الأدنى (Lower Asymptote) للمنحنى المميز للمفردة، يمثل احتمال توصل هؤلاء المفحوصين إلى الإجابة الصحيحة على الفقرة عن طريق التخمين (التوصل إلى الحل الصحيح بالتخمين غير مرتبط بارتفاع أو انخفاض مستوى القدرة) والصيغة الرياضية للنموذج التي تعبر عن احتمال الإجابة الصحيحة عن الفقرة هي:

$$P(X_{is} = 1/\theta_s, \beta_i, \alpha_i, c_i) = c_i + (1 - c_i)\frac{e^{\alpha_i(\theta_s - \beta_i)}}{1 + e^{\alpha_i(\theta_s - \beta_i)}} \cdots\cdots\cdots (04)$$

4.2.أساليب تقدير قدرات الأفراد.

يعتمد برنامج Bilog_Mg لتقدير قدرة الأفراد على الطرق التالية:

أولا: طريقة الأرجحية العظمى: Maximum Likelihood (ML)

احدى الطرق الكثيرة الاستخدام حيث تقوم على تقدير قدرة الافراد من خلال تعظيم الاحتمالية للمعلمة المراد تقديرها، وشكل دالة الارجحية Likelihood Function (L) اذا افترضنا ان $P_i(\theta_s)$ احتمال استجابة فرد ذو قدرة (θ_s) علي الفقرة (i)اجابة صحيحة و $Q_i(\theta_s)$ لاحتمال الاجابة الخاطئة عن الفقرة، فان الصيغة الرياضية لدالة الارجحية (L) لأنماط الاستجابات لعدد من الافراد(N) بواسطة مجموعة من الفقرات(n) ذوي مستويات قدرة ثابتة كما يلي(علام، 2005،ص.93):

$$L = \prod_{s=1}^{N}\prod_{i=1}^{n} P_i(\theta_s)^{u_{si}}\ Q_i(\theta_s)^{(1-u_{si})}$$

حيث أن:

$$Q_i(\theta_s) = 1 - P_i(\theta_s)$$

$$u_i = \begin{cases} 1 & \text{اذا كانت الاجابة صحيحة} \\ 0 & \text{اذا كانت الاجابة خاطئة} \end{cases}$$

والرمز $\prod$: هو ناتج الضرب

ومن خلال تطبيق خوارزمية نيوتن–رافسون (Newton–Raphson) نجد تقديرات القدرة θ_s للأفراد التي تجعل لوغاريتم الارجحية عندها قيمة عظمى (شرفاوي، 2022،ص.122)

ثانيا: طريقة توقع التوزيع البعدي Expected A Posteriori (EAP):

تعتمد هذه الطريقة على نظرية بييز (Bayes THeorem) والتي تربط بين الاحتمالات الشرطية (Conditional) والاحتمالات الهامشية (Marginal) وتتطلب كذلك افتراضات قبلية (Probabilities Prior) للبارامترات (علام, 2005, ص.101)، تم اقترح هذه الطريقة في التقدير من طرف بوك ومسليفي (Bock & Mislevy) وتقوم هذه الطريقة على استخدام التوزيع الطبيعي المعياري حيث يتم تقسيم قيم السمة الكامنة (θ) التي تمثل هذا التوزيع الذي يقع في العادة بين (−3،+3) بفترات طولها 0،1 (أي تجزأ إلى61 جزءا)

تسمى ربيعا ((Quadrant (Q_r)) ويعطى لها وزنا ($W(Q_r)$) واستخدام الاقتران اللوغارتمي للأرجحية عند كل ربيع($L(Q_r)$، في هذه الطريقة يتم تقدير القدرة مرة واحدة دون اجراء عمليات التقريب المتتابع كما هو الحال في طريقة الأرجحية العظمى (تقي الدين، 2013، ص. 106)، ويتم تقدير قدرة الفرد من خلال المعادلة التالية (Du Toit, 2003,p.561) :

$$\theta=\frac{\sum_{r=1}^{61}[Q_r\ L(Q_r)\ W(Q_r)]}{\sum_{r=1}^{61}[L(Q_r)\ W(Q_r)]}$$

ثالثا: طريقة القيمة العظمى التوزيع البعدي (Maximum A Posteriori (MAP:

هي الطريقة الثانية التي تعتمد على أسلوب بييز(Bayes) في تقدير قدرات الأفراد، وإجراءات هذه الطريقة في التقدير هي نفسها في طريقة الأرجحية العظمى باستثناء استخدام التوزيع القبلي للسمة الكامنة (معلومات سابقة عن توزيع القدرة) وغالبا ما يستخدم التوزيع الطبيعي المعياري، ويتم تعريف التوزيع البعدي من خلال ضرب اقتران الأرجحية باقتران التوزيع القبلي (تقي الدين،2013، ص. 102)، وفي هذا الأسلوب يتم استخدام خوارزمية التوقع والتعظيم وتكرار مرحلة التوقع والتعظيم حتى يتقارب التقدير، يمتاز هذا الأسلوب في تقدير القدرة لجميع الأفراد عكس طريقة الأرجحية العظمى التي لا يمكنها تقدير معلمة القدرة للأفراد الذين أجابوا إجابة صحيحة أو خاطئة على كل الفقرات (حمادنة،2011، ص.11)، كما تجدر الإشارة إلى أن هذه الطريقة تعتمد على المنوال في اختيار قيمة القدرة الأكثر تكرار في التوزيع القبلي في بداية عملية التقدير أما طريقة توقع التوزيع البعدي (EAP) فتعتمد على حساب متوسط معلم القدرة من خلال التوزيع القبلي (عبدالله،2012،ص. 49).

5.2.الدراسات السابقة:

1- دراسة بني عطا (2017)، بعنوان تقصي أثر طول الاختبار وحجم العينة على دقة طرق تقدير معالم الفقرات وقدرات الأفراد في برنامج Bilog، حيث سعت الدراسة إلى توضيح دقة التقدير باستخدام الطرق المعتمدة في برنامج Bilog-Mg لمعالم الفقرة والقدرة، لا سيما في ظل الاختلاف في كل من طول الاختبار وحجم العينة، اعتمادا على برنامج (WinGen) الذي يستخدم أسلوب المحاكاة من أجل توليد أربعة أطوال اختبار متباينة (80، 60، 30، 10) فقرة وأربعة مستويات لحجم العينة (1500، 1000، 500، 250)، حيث تم تقدير معالم الفقرة والقدرة الافراد وفقًا للنموذج الثلاثي المعلم باستخدام برنامجBilog-Mg، مع حساب جذر التربيعي لمتوسطات مربعات الانحرافات للفروق بين قيم المعالم الحقيقية والمولدة كمقياس لتقييم دقة هذه التقديرات، كشفت نتائج الدراسة المتعلقة بتدريج فقرات الاختبار، وفقًا للطرق الثلاث المستخدمة، عن وجود أثر دال إحصائيا لكل من

عامل طول الاختبار وعامل حجم العينة على دقة تقديرات القدرة، وتحديدا أظهرت طريقة توقع الاقتران (EAP) دقة عالية.

2- دراسة سليم الصباح (2014)، بعنوان المقارنة بين دقة تقدير القدرة باختلاف طول الاختبار وشكل توزيع معلمة القدرة وفقًا للنموذج الثلاثي باستخدام البيانات الحقيقية والمولدة، والتي سعت إلى إجراء تقييم ومقارنة لدقة تقديرات القدرة فيما يتعلق بطول الاختبار وشكل توزيع القدرة لكل من البيانات الحقيقية والمحاكاة.(المولدة) استندت هذه المقارنة في التقدير الى كل من طريقة الارجحية القصوى (MLE) وطريقة توقع الاقتران (EAP) المستخدمة في برمجية Bilog-Mg، ولتحقيق أهداف الدراسة تم تحليل البيانات من اختبار العلوم الوطني لطلاب الصف الثامن، جنبًا إلى جنب مع توليد بيانات تحاكي خائص مجموعة البيانات الحقيقية من خلال برنامج (WinGen) تم استخلاص تقديرات قدرة الافراد والأخطاء المعيارية المقابلة لها، والتي تعتبر كمؤشرات لدقة التقدير باستخدام برمجية Bilog-Mg، مع أطوال اختبار تضم كل من 50 و25 مفردة، مع حجم عينة مكونة من 1000 شخص في جميع الظروف البحثية، وخلصت الدراسة إلى أن طريقة EAP أسفرت عن تقديرات أكثر دقة لمعلم القدرة عبر التوزيعات المختلفة للقدرة (عادي، ملتوي نحو اليمين، ملتوي نحو اليسار) بالإضافة إلى ذلك، أشارت الدراسة إلى عدم وجود اختلافات في دقة التقديرات بين مجموعات البيانات الحقيقية والمولّدة.

3- دراسة الشواورة (2013) بعنوان دقة تقدير معالم الفقرات بطريقة الأرجحية العظمى الهامشية وبييز (Bayes) في ظروف مختلفة في عدد الفقرات وحجم العينة والنموذج أللوغاريتمي، سعت هذه الدراسة إلى بيان أي من الطريقتين أكثر دقة في تقدير معالم الفقرات وفقا للنموذج اللوغارتمي المستخدم (الأحادي، الثنائي) بحجم عينة (250،500،1000،1500)، وعدد فقرات (10،40) باستخدام المنهج التجريبي وللإجابة عن تساؤلات الدراسة تم توليد البيانات بالاعتماد على برنامج (WinGen) وعلى برنامجي (Bilog-Mg) و(WinBUGS) لتقدير معالم الفقرات والاعتماد على الجذر التربيعي لمتوسطات مربعات الخطأ (RMSE) كمؤشر لدقة التقدير وتوصلت الدراسة إلى أفضلية طريقة بييز (Bayes) من طريقة الأرجحية العظمى الهامشية (MML) في معظم مواقف الدراسة.

4- دراسة الشريفين (2012) هدفت إلى توضيح تأثير طرق التقدير على قيم بارامترات الفقرة والخصائص السيكومترية للاختبار، في ظل التغيرات في حجم العينة، وسعياً لتحقيق هذا الهدف، تم اعداد اختبار تحصيلي فقراته من نوع اختيار من أربعة بدائل في وحدة الفيزياء، يشمل 33 فقرة، وتم إجراؤه على مجموعة مكونة من 1000 طالب من طلاب الصف الثاني من المدارس الثانوية العلمي، تمت عملية التدريج واستخراج نتائج باستخدام برنامج (Bilog-Mg) حيث تم الكشف عن وجود فروق ذات دلالة إحصائية في متوسط الأخطاء المعيارية

لتقديرات القدرة ترجع إلى متغير حجم العينة وللتفاعل بين طريقة التقدير وحجم العينة، بينما لم تظهر فروق ذات دلالة إحصائية تعزي لطريقة التقدير، علاوة على ذلك كشفت النتائج عن تحسن دقة تقديرات القدرة في عينات الأفراد ذوي القدرات العالية وذوي القدرات المنخفضة عند استخدام الطريقة اللاحقة المتوقعة(EAP) في حين تتحسن الدقة عند مستويات الأفراد ذوي القدرات المتوسطة عند استخدام طريقة تقدير الارجحية القصوى (MLE)بغض النظر عن مستويات حجم العينة.

5- دراسة الطراونة (2011) بعنوان «مقارنة طرق تقدير القدرة باستخدام النموذج المناسب في ضوء الخطأ المعياري في تقدير القدرة»، التي سعت إلى تقييم الطرق المختلفة لتقدير القدرة (الارجحية القصوى(MLE) ، القيمة العظمى التوزيع البعدي (MAP)، طريقة التوقع التوزيع البعدي (EAP))، ولتحقيق الهدف الأساسي لهذه الدراسة وهو المقارنة بين طرق التقدير المذكورة أعلاه تم استخدام نتائج الاختبار الوطني الذي أجري في عام 2011 (بإدارة وزارة التربية والتعليم الأردنية)، في مادة الرياضيات بحجم عينة يبلغ 8862 ممتحن وفي مادة العلوم بحجم عينة من 7003 ممتحن ولأغراض التحليل تم استخدم برنامج البايلوغ لتقدير قدرات الافراد وكذا الأخطاء المعيارية المقابلة لها، وأشارت النتائج إلى أن طريقة التوقع التوزيع البعدي ظهرت باعتبارها الطريقة الأكثر دقة لتقدير القدرة، تليها طريقة القيمة العظمى التوزيع البعدي، بينما جاءت طريقة الارجحية القصوى في المرتبة الأخيرة بسبب الخطأ المعياري المرتفع.

6- دراسة الدرابيع (2001) بعنوان «فعالية النموذج اللوغاريتمي أحادي المعلمة «نموذج راش» في دقة تقدير قدرة الافراد ومعامل صعوبة الفقرات باختلاف حجم العينة وطول الاختبار»، حيث سعت إلى تقييم فعالية نموذج Raschفي دقة تقدير كل من القدرة والصعوبة، باستخدام أحجام عينات مختلفة (500، 100، 50) وأطوال الاختبار (300، 50، 25)، تم الحصول عليها بأسلوب المحاكاة، ولعملية التقدير تم استخدام برنامج-Bilog Mg، بالاعتماد على الجذر التربيعي لمتوسطات مربع الخطأ Root Mean Square Error (RMSE) كمؤشر لدقة التقدير، كشفت الدراسة عن وجود تفاعل بين حجم العينة وطول الاختبار فيما يخص تقدير قدرة الافراد، كما أظهرت النتائج عن وجود فروق دالة احصائيا في دقة التقدير تعزى فقط إلى طول الاختبار بينما لم تظهر أي فروق بين مستويات حجم العينة في دقة التقدير.

7- دراسة جاو وجن(Gao & Chen, 2005) هدفت الدراسة إلى المقارنة في دقة التقدير بين طريقة الأرجحية القصوى (MLE) وطريقة تعظيم الاقتران البعدي (MAP) التي تعتمد على أسلوب بييز(Bayes) ، ولتحقيق هدف الدراسة تم توليد بيانات بأحجام (100،500،200)، واختبارات بأطوال (10،30،60) فقرة، توصلت الدراسة إلى

أن دقة التقدير بين الطريقتين عند حجم عينة (500،2000) كانت متشابهة ، كذلك عن تغير عدد الفقرات أما عند حجم عينة 100 فرد كانت الأفضلية لطريقة تعظيم الاقتران البعدي.

8- دراسة وانغ وفيسبول (Wang & Vispoel, 1998) بعنوان «خصائص طرق تقدير القدرة في الاختبارات التكيفية المحوسبة»، سعت الدراسة إلى مقارنة خصائص أربع طرق مختلفة لتقدير قدرة الافراد وهي طريقة الاحتمالية القصوى (MLE) وثلاث طرق تعتمد على الاسلوب البييزي (طريقة أوين Owen's، والطريقة توقع التوزيع البعدي(EAP) ، طريقة القيمة العظمى للتوزيع البعدي(MAP))، وباستخدام المحاكاة لتوليد البيانات لاختبار تكيفي محوسب، أكدت نتائج الدراسة المحصل عليها على تفوق الطرق البييزية الثلاثة حيث أظهرت أقل الخطأ بينما عادت الأفضلية في أسلوب بييز لطريقة توقع التوزيع البعدي(EAP).

3 . إجراءات الدراسة الميدانية .

1.3.المرحلة الأولى: تصميم الدراسة:

تم في هذه المرحلة تصميم القياسات المتكررة لتقصي دقة تقدير قدرة الأفراد باستخدام الطرق المعتمدة في برنامج Bilog-Mg (طريقة الأرجحية العظمى، طريقة توقع التوزيع البعدي، وطريقة تعظيم الاقتران البعدي) حيث تم قياس الأخطاء المعيارية للتقدير وهو مؤشر للحكم على دقة التقدير (متغيرا تابعا) لكل طريقة من الطرق الثلاثة (متغير مستقل) عند كل مستوى من المستويات الثلاثة (03) من حجم العينة (200،500،1000)، حيث تتقارب هذه المستويات من الواقع العملي لتطبيق الاختبارات مما أتاح الحصول على 3 قياسات متكررة عند كل مستوى.

2.3.المرحلة الثانية: توليد البيانات:

توليد بيانات من الاستجابات الثنائية (0،1) باستخدام طريقة المحاكاة (Simulation) (لان هذا الأسلوب يمكننا من ضبط المتغيرات) لعينات تحاكي عينات المجتمع الأصلي بطريقة المونتي كارلو (Monte Carlo Methods (MCM)) بحجم عينات (200،500،1000) باستخدام برنامج (WinGen V.3) من تصميم هان وهامبيلتون (Han & Hambleton, 2007) وقد تمت عملية توليد البيانات تحت افتراض التوزيع الطبيعي للقدرة بمتوسط حسابي مقداره (0) وانحراف معياري مقداره (1)، لاختبار مكون من 20 سؤالا (وتم اختيار هذا الطول لمناسبته لكثير من الظروف التربوية) وفق النموذج الأحادي البرامتر (نموذج راش) تحت افتراض التوزيع الطبيعي لمعلمة الصعوبة بمتوسط حسابي مقداره (0) وانحراف معياري مقداره (1).

3.3.الأساليب الإحصائية المستعملة:

- الوسط الحسابي و الانحراف المعياري.
- تحليل التباين الأحادي للقياسات المتكررة، اختبار بنفيروني (Bonferroni) للمقارنات البعدية.
- التحليل العاملي الاستكشافي باستخدام طريقة المكونات الرئيسية (Principal Component Analysis) وتدوير العوامل باستخدام طريقة الفاريماكس (Varimax).
- برنامج WinGen، برنامج Bilog_Mg، برنامج SPSS V25.

4. عرض نتائج الدراسة ومناقشتها.

قبل الإجابة عن أسئلة الدراسة وعرض النتائج ومناقشتها تم التحقق من افتراضات نظرية لاستجابة للمفردة (أحادية البعد، الاستقلال الموضعي) للبيانات المولدة عند حجم العينة (200،500،1000).

- **افتراض أحادية البعد** (Unidimensionality)**:** تم فحص افتراض أحادية البعد (Unidimensionality) باستخدام التحليل العاملي الاستكشافي اعتمادا على طريقة المكونات الرئيسية (Principal Component Analysis) وباستخدام طريقة الفاريماكس (Varimax) لتدوير العوامل التي تعدت قيمة الجذر الكامن لها الواحد (01) والجدول (01) يبن نتائج التحليل العاملي:

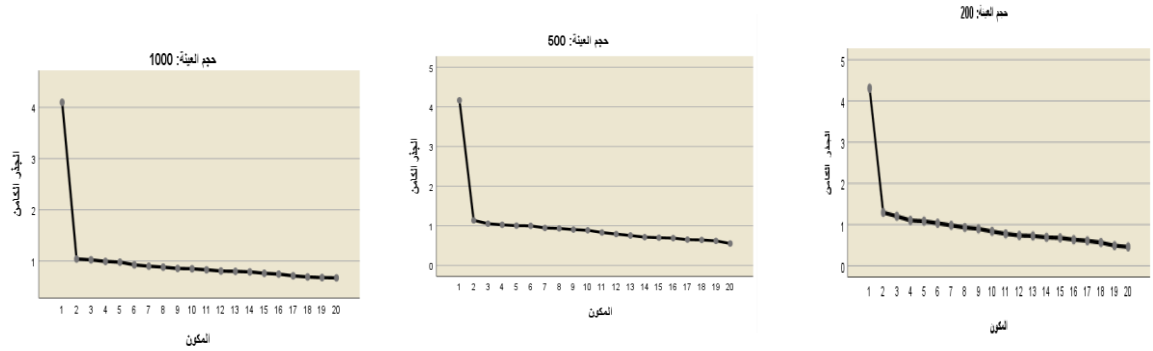
الجدول (1) : نتائج التحليل العاملي الاستكشافي للبيانات المولدة باختلاف حجم العينة (200،500،1000)

حجم العينة	المكون	الجذر الكامن	العامل 2	العامل 3	نسبة جك1 إلى جك2	*جك1-جك2 جك2- جك3
200	الجذر الكامن	4,307	1,292	1,201	3,33	33,13
	نسبة التباين المفسر	21,535	6,462	6,007		
500	الجذر الكامن	4,165	1,137	1,053	3,66	36,05

* جك1: الجذر الكامن الأول، جك2: الجذر الكامن الثاني، جك3: الجذر الكامن الثالث.

	نسبة التباين المفسر	20,823	5,687	5,267		
1000	الجذر الكامن	4,099	1,039	1,022	3,94	180
	نسبة التباين المفسر	20,493	5,195	5,112		

من خلال نتائج الجدول (1) نجد أن قيم التباين المفسر عند مختلف مستويات حجم العينة (200، 500، 1000) قد فاقت قيمة 20% كمؤشر على أحادية البعد (تراوحت بين 20,49 و 20,82)، ونسبة الجذر الكامن الأول إلى الجذر الكامن الثاني تجاوزت كلها قيم العدد (2)، وهو مؤشر ثاني على أحادية البعد، وكذلك أعطت نتائج عملية قسمت حاصل طرح الجذر الكامن الثاني من الجذر الكامن الأول على حاصل طرح الجذر الكامن الثالث من الجذر الكامن الثاني (33,13، 36,05، 180) قيما أكبر من العدد (7) وهو دليل إضافي على أحادية البعد (Hatti,1985) يتبين من نتائج التحليل ألعاملي تحقق افتراض أحادية البعد لجميع بيانات الدراسة كما تم تمثيل الجذور الكامنة بيانيا (Scree Plots) كما يوضحه الشكل التالي:



الشكل (1) : رسم بياني لقيم الجذور الكامنة لبيانات الدراسة

يتبين من الشكل (1)، أن الجذر الكامن الأول يتميز بشكل واضح (سائد) عن بقية الجذور الكامنة لبقية العوامل في جميع مستويات حجم العينة و هو ما يعزز تحقق افتراض أحادية البعد في البيانات.

– **افتراض الاستقلال الموضعي Local Independence:** التحقق من افتراض الاستقلال الموضعي فقد أشار هامبلتون وسوامنيثان (Hambleton, Swaminatan& Rogers, 1991.p 10) إلى أن هذا الافتراض يكافئ افتراض أحادية البعد أي إذا تحقق افتراض أحادية البعد في الاختبار يحقق الاستقلال الموضعي.

أولا: النتائج المتعلقة بالإجابة عن السؤال الأول : هل توجد فروق بين المتوسطات الحسابية للأخطاء المعيارية لتقدير قدرة الأفراد تعزى لطريقة التقدير (ML، EAP، MAP) المستخدمة في برنامج Bilog_Mg عند حجم العينة 200 وفق النموذج الأحادي البارامتر (نموذج راش)؟.

للإجابة عن هذا السؤال تم استخدام برنامج Bilog-Mg لتحليل البيانات والحصول على تقديرات الخطأ المعياري لتقدير القدرة وفق نموذج راش باستخدام كل من طريقة الأرجحية العظمى (ML)، طريقة توقع التوزيع البعدي(EAP)، تعظيم الاقتران البعدي (MAP)، بعدها تمت عملية معالجة شكل البيانات (مخرجات المرحلة الثالثة من برنامج Bilog-Mg في التحليل (Phase3)) لتتوافق مع البرنامج الإحصائي SPSS V25 حيث تم حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير قدرات الأفراد باستخدام الطرق الثلاثة عند حجم عينة يساوي 200 فرد، وذلك كما هو مبين في الجدول التالي:

الجدول (2): المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري لتقدير القدرة لكل طريقة تقدير عند حجم عينة 200 فرد.

طريقة التقدير	الخطأ المعياري للتقدير	
	المتوسط الحسابي	الانحراف المعياري
طريقة الأرجحية العظمى	0,598	0,789
طريقة توقع التوزيع البعدي	0,501	0,021
طريقة القيمة العظمى للتوزيع البعدي	0,497	0,019

يتضح من الجدول (2) عند استخدام الطرق المعتمدة في برنامج Bilog_Mg في تقدير القدرة عند حجم عينة يساوي 200، أن أعلى متوسط ظاهريا للأخطاء المعيارية للتقدير كان عند استخدام طريقة الأرجحية العظمى بمتوسط (0,598) ويليه متوسط الأخطاء المعياري للتقدير لطريقة التوقع البعدي (EAP) بـ (0,501)، ثم متوسط طريقة القيمة العظمى للتوزيع البعدي (MAP) بـ (0,497).

من خلال النتائج يتبين أن طريقة القيمة العظمى للتوزيع البعدي (MAP) أعطت تقديرات ظاهريا أكثر دقة في تقدير من الطرق الأخرى، وفي ضوء ما تقدم، تم إجراء تحليل التباين الأحادي للقياسات المتكررة لقيم الخطأ المعياري لتقدير القدرة للكشف عن جوهرية الفرق الظاهري بين متوسطات الخطأ المعياري في التقدير ونظرا لعدم تحقق شرط الكروية في البيانات من خلال نتائج اختبار (Mauchly's Test of Sphericity) تم تعديل درجات الحرية باستخدام الإحصائي Greenhouse-Geisser كما يظهر في الجدول (3):

الجدول (3): نتائج اختبار تحليل التباين الأحادي للقياسات المتكررة لقيم الأخطاء المعيارية لتقديرات القدرة تبعا لطريقة التقدير عند حجم عينة 200 فرد.

المتغير	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	حجم الأثر

طريقة التقدير Greenhouse-Geisser (0,521)	الخطأ المعياري للتقدير	1,310	1,043	1,256	500,74	0,00	0,716
	الخطأ	0,521	207,528	0,0025			
	المجموع الكلي	1,831	208,571	1,2585			

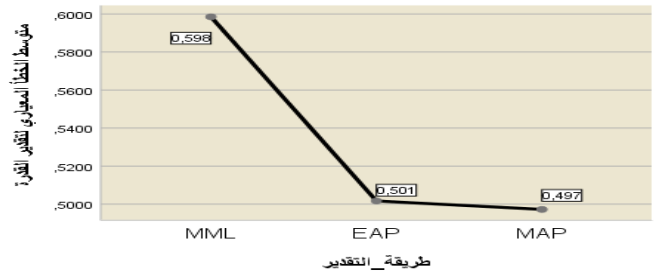
أظهرت النتائج الواردة في الجدول (3) أن قيمة F تساوي 500,74 وهي دالة إحصائيا عند مستوى الدلالة ($\alpha = 0,01$) بين الاوساط الحسابية لقيم الخطأ المعياري لتقدير القدرة باستخدام الطرق الثلاثة **(**ML, EAP, MAP**)** عند استخدام عينة مكونة من 200 فرد، وبناء عليه يمكن القول أن طريقة التقدير المستخدمة لها اثر على دقة التقدير القدرة اعتمادا على مؤشر الخطأ المعياري للتقدير، كذلك نجد أن مؤشر حجم الأثر (Effect size)أو الدلالة العملية (Practical significance) (مربع ايتا η^2) قد بلغ 0,716 وهو مرتفع حسب معايير كوهين(Cohen, 1988) (أكبر من 0,14) أي أن طريقة التقدير أسهمت بـ 71,6 % في تفسير تباين قيم دقة تقدير القدرة مما يؤكد على أهمية طريقة التقدير و للوقوف على مواقع الفروق في دقة تقدير تم استخدام اختبار بنفيروني (Bonferroni) للمقارنات البعدية كما هو مبين في الجدول(4):

الجدول (4): نتائج اختبار Bonferroni للمقارنات البعدية بين متوسطات الأخطاء المعيارية لتقدير القدرة تبعا لطريقة التقدير عند حجم عينة 200.

طريقة التقدير	متوسط الخطأ المعياري	طريقة توقع التوزيع البعدي	طريقة القيمة العظمى للتوزيع البعدي
	0,501	**0,501**	**0,497**
طريقة الأرجحية العظمى	**0,598**	**0,097***	**0,101***
طريقة توقع التوزيع البعدي	**0,501**	–	**0,004***

*دال إحصائيا عند مستوى الدلالة **(α= 0,05)**

يتضح من الجدول (4) أن نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير قدرات الأفراد باستخدام اختبار بنفيروني (Bonferroni) الناتجة عن تطبيق طرق التقدير الثلاثة (ML, EAP, MAP) كانت دالة إحصائيا عند مستوي دلالة **(**α= 0,05**)** وكانت هذه الفروق لصالح طريقة القيمة العظمى للتوزيع البعدي مما يدل على أن هذه الطريقة الأكثر دقة، كما بينت النتائج أفضلية طريقة توقع التوزيع البعدي من طريقة الأرجحية العظمى في دقة التقدير وهذا عند استخدام عينة مكونة من 200 فرد وهو ما يوضحه الشكل (2) :



الشكل (2): العلاقة بين طريقة التقدير ومتوسط الخطأ المعياري لتقدير القدرة – حجم عينة 200 فرد–

ثانيا: النتائج المتعلقة بالإجابة عن السؤال الثاني : هل توجد فروق بين المتوسطات الحسابية للأخطاء المعيارية لتقدير قدرة الأفراد تعزى لطريقة التقدير (ML، EAP، MAP) المستخدمة في برنامج Bilog_Mg عند حجم العينة 500 وفق النموذج الأحادي البارامتر (نموذج راش)؟.

للإجابة عن هذا السؤال تم استخدام برنامج Bilog-Mg للحصول على قيم الخطأ المعياري لتقدير القدرة وفق نموذج راش لكل فرد من أفراد العينة المكونة من 500 فرد باستخدام طرق التقدير المعتمدة في الدراسة (ML, EAP, MAP)، بعد ذلك تمت عملية معالجة شكل البيانات (مخرجات المرحلة الثالثة من برنامج Bilog_Mg في التحليل (Phase 3)) لتتوافق مع البرنامج الإحصائي SPSS V25 ومنه حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير قدرات الأفراد باستخدام الطرق الثلاثة المستخدمة في الدراسة كما هو مبين في الجدول التالي:

الجدول (5): المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري لتقدير القدرة وفقا لطريقة التقدير عند حجم عينة 500 فرد.

طريقة التقدير	الخطأ المعياري للتقدير	
	المتوسط الحسابي	الانحراف المعياري
طريقة الأرجحية العظمى	0,539	0,121
طريقة توقع التوزيع البعدي	0,456	0,037
طريقة القيمة العظمى للتوزيع البعدي	0,449	0,028

يتضح من الجدول (5) عند استخدام طرق تقدير القدرة المعتمدة في الدراسة عند حجم عينة يساوي 500 فرد أن متوسط الأخطاء المعيارية للتقدير لطريقة الأرجحية العظمى(ML) بلغ (0,539) هو أعلى من متوسطات الأخطاء المعيارية للتقدير لطريقة التوقع التوزيع البعدي بـ(0,456) وطريقة القيمة العظمى للتوزيع البعدي بـ(0,449)، وبالتالي فان الطريقة الأخيرة ظاهريا أكثر دقة في التقدير ولمعرفة إن كانت هذه الفروق ذات دلالة تم إجراء تحليل التباين الأحادي للقياسات المتكررة لقيم الخطأ المعياري لتقدير معلمة القدرة، ونظرا لعدم تحقق

شرط الكروية في البيانات من خلال نتائج اختبار (Mauchly's Test of Sphericity) تم تعديل درجات الحرية باستخدام الإحصائي Greenhouse-Geisser، كما في الجدول التالي:

الجدول (6): نتائج اختبار تحليل التباين الأحادي للقياسات المتكررة لقيم الأخطاء المعيارية لتقديرات القدرة تبعا لطريقة التقدير عند حجم عينة 500 فرد.

المتغير	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	حجم الأثر
طريقة التقدير Greenhouse-Geisser (0,545)	الخطأ المعياري للتقدير	2,498	1,090	2,292	395,23	0,000	0,442
	الخطأ	3,154	543,815	0,006			
	المجموع الكلي	5,652	544,905	2,298			

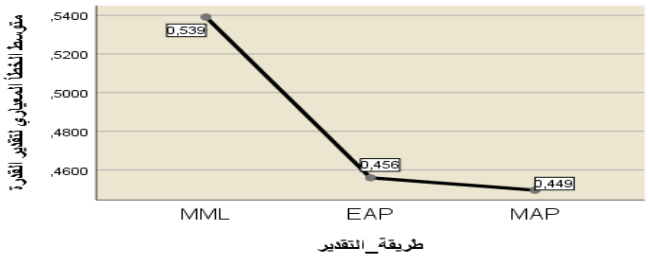
تضح من النتائج الواردة في الجدول (6) أعلاه أن قيمة F تساوي (395,23) وهي دالة إحصائيا عند مستوى الدلالة ($\alpha = 0,01$) بين الاوساط الحسابية لقيم الخطأ المعياري لتقدير القدرة باستخدام الطرق الثلاثة (ML, EAP, MAP) عند استخدام حجم عينة يساوي 500 فرد، وبالتالي يمكننا القول أن طريقة التقدير المستخدمة لها اثر على دقة تقدير القدرة اعتمادا على مؤشر الخطأ المعياري للتقدير، كما يبين مؤشر حجم الأثر (Eeffect size)(مربع ايتا η^2) الذي بلغ 0,442 وهو مؤشر يقع في المستوى المرتفع (أكبر من 0,14) أي أن المتغير المستقل (طريقة التقدير) أسهمت بـ44,2 % في تفسير تباين قيم الخطأ المعياري لتقدير القدرة (دقة تقدير) مما يؤكد على أهمية طريقة التقدير في دقة التقدير، ولمعرفة اتجاه الفروق بين متوسطات الأخطاء المعيارية للتقدير تم استخدام اختبار Bonferroni للمقارنات البعدية والجدول التالي يظهر هذه المواقع:

الجدول (7): نتائج اختبار Bonferroni للمقارنات البعدية بين متوسطات الأخطاء المعيارية لتقدير القدرة تبعا لطريقة التقدير عند حجم عينة 500 فرد.

طريقة التقدير	متوسط الخطأ المعياري	طريقة توقع التوزيع البعدي EAP	طريقة القيمة العظمى للتوزيع البعدي MAP
		0,456	0,449
طريقة الأرجحية العظمى ML	0,539	0,083*	0,090*
طريقة توقع التوزيع البعدي EAP	0,456	–	0,007*

*دال إحصائيا عند مستوى الدلالة ($\alpha = 0,05$)

يظهر الجدول (7) نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير قدرات الأفراد باستخدام اختبار Bonferroni والناتجة عن استخدام طرق التقدير الثلاثة المعتمدة في الدراسة (ML,EAP,MAP) كانت كلها دالة إحصائيا عند مستوي دلالة (**α= 0,05**) ولصالح طريقة القيمة العظمى للتوزيع البعدي (MAP)، وأشارت النتائج عند تطبيق طريقة (ML) في تقدير قدرة الأفراد إلى متوسط أخطاء معيارية يبلغ (0,539) وهو أكبر مقارنة مع متوسطي الأخطاء المعيارية عند استخدام كل من طريقة توقع التوزيع البعدي(EAP) (0,456) وطريقة القيمة العظمى للتوزيع البعدي (MAP) (0,449) ، وهذا يعني أن طريقة (MAP) أكثر دقة من باقي الطرق الأخرى وأفضلية طريقة توقع التوزيع البعدي (EAP) من طريقة الأرجحية العظمى (ML) في دقة عند استخدام عينة مكونة من 500 فرد وفق نموذج راش والشكل (3) يوضح مواقع الفروق بين متوسطات الأخطاء المعيارية لطرق التقدير.



الشكل (3): العلاقة بين طريقة التقدير ومتوسط الخطأ المعياري لتقدير القدرة – حجم عينة 500 فرد–

ثالثا: النتائج المتعلقة بالإجابة عن السؤال الثالث : هل توجد فروق بين المتوسطات الحسابية للأخطاء المعيارية لتقدير قدرة الأفراد تعزى لطريقة التقدير (ML، EAP، MAP) المستخدمة في برنامج Bilog_Mg عند حجم العينة 1000 وفق النموذج الأحادي البارامتر (نموذج راش)؟.

للإجابة عن هذا السؤال تم الحصول على قيم الخطأ المعياري لتقدير القدرة لكل فرد من أفراد العينة المكونة من 1000 فرد باستخدام طرق التقدير المعتمدة في الدراسة (ML, EAP, MAP) من خلال برنامج Bilog_Mg، ثم معالجة شكل البيانات (مخرجات المرحلة الثالثة من برنامج Bilog_Mg في التحليل (Phase3)) لتتوافق مع البرنامج الإحصائي SPSS V25 ثم حساب المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير قدرات الأفراد باستخدام طرق التقدير الثلاثة كما هو مبين في الجدول(8).

الجدول (8): المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري لتقدير القدرة وفقا لطريقة التقدير عند حجم عينة 1000 فرد.

طريقة التقدير	الخطأ المعياري للتقدير	
	المتوسط الحسابي	الانحراف المعياري
طريقة الأرجحية العظمى ML	0,546	0,119
طريقة توقع التوزيع البعدي EAP	0,461	0,036
طريقة القيمة العظمى للتوزيع البعدي MAP	0,454	0,028

يلاحظ من الجدول (8) أن متوسط الأخطاء المعيارية للتقدير عند استخدام طريقة الأرجحية العظمى(ML) بلغ (0,546) وهو أعلى من متوسطات الأخطاء المعيارية للتقدير لطريقة التوقع التوزيع البعدي(EAP) بـ (0,461) وطريقة القيمة العظمى للتوزيع البعدي(MAP) بـ (0,454)، وبالتالي فان هذه الطريقة الأخيرة (MAP) أكثر دقة ظاهريا في التقدير تأتي بعدها طريقة (EAP) ثم طريقة (ML) ولمعرفة دلالة هذه الفروق تم إجراء تحليل التباين الأحادي للقياسات المتكررة لقيم الخطأ المعياري لتقدير معلمة القدرة ولعدم تحقق شرط الكروية في البيانات من خلال نتائج اختبار (Mauchly's Test of Sphericity) تم تعديل درجات الحرية باستخدام الإحصائي Greenhouse-Geisser، كما يظهر في الجدول (9) :

الجدول (9): نتائج اختبار تحليل التباين الأحادي للقياسات المتكررة لقيم الأخطاء المعيارية لتقديرات القدرة تبعا لطريقة التقدير عند حجم عينة 1000 فرد.

المتغير	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	حجم الأثر
طريقة التقدير Greenhouse-Geisser (0,540)	الخطأ المعياري للتقدير	5,252	1,080	4,865	880,75	0,000	0,469
	الخطأ	5,958	1078,481	0,006			
	المجموع الكلي	11,237	1079,561	4,871			

بالنظر إلى النتائج الواردة في الجدول (9) نجد أن قيمة F تساوي (880,75) وهي دالة إحصائيا عند مستوى الدلالة ($\alpha = 0,01$) بين الاوساط الحسابية لقيم الخطأ المعياري لتقدير القدرة باستخدام الطرق الثلاثة (ML, EAP, MAP) المعتمدة في الدراسة عند حجم عينة تساوي 1000 فرد وفق نموذج راش، وبالتالي يمكننا القول أن طريقة التقدير المستخدمة لها اثر على دقة تقدير القدرة اعتمادا على مؤشر الخطأ المعياري للتقدير، كما أظهر مؤشر حجم الأثر مربع ايتا (η^2) الذي بلغ 0,469 وهو مؤشر يقع في المستوى المرتفع (أكبر من 0,14) أي

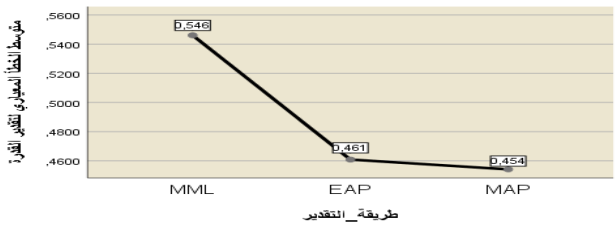
أن طريقة التقدير أسهمت بـ 46,9 % في تفسير تباين قيم الخطأ المعياري لتقدير القدرة (دقة تقدير) مما يبين أهمية طريقة التقدير، وللوقوف على اتجاه الفروق في بين متوسطات الأخطاء المعيارية للتقدير تم استخدام اختبار Bonferroni للمقارنات البعدية والجدول التالي يظهر هذه المواقع:

الجدول (10): نتائج اختبار Bonferroni للمقارنات البعدية بين متوسطات الأخطاء المعيارية لتقدير القدرة تبعا لطريقة التقدير عند حجم عينة 1000 فرد.

طريقة التقدير	متوسط الخطأ المعياري	طريقة توقع التوزيع البعدي EAP	طريقة القيمة العظمى للتوزيع البعدي MAP
	0,461	0,461	0,454
طريقة الأرجحية العظمى ML	0,546	0,085*	0,092*
طريقة توقع التوزيع البعدي EAP	0,461	-	0,007*

*دال إحصائيا عند مستوى الدلالة ($\alpha = 0,05$)

يظهر الجدول (10) نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير قدرات الأفراد باستخدام طرق التقدير الثلاثة المعتمدة في الدراسة (ML,EAP,MAP) كانت كلها دالة إحصائيا عند مستوي دلالة ($\alpha = 0,05$) ولصالح طريقة القيمة العظمى للتوزيع البعدي (MAP)، وأشارت النتائج إلى أن متوسط أخطاء معيارية للتقدير في طريقة (ML) الذي بلغ (0,546) وهو أكبر مقارنة مع متوسطي الأخطاء المعيارية عند استخدام طريقتي توقع التوزيع البعدي(EAP) بـ (0,461) وطريقة القيمة العظمى للتوزيع البعدي (MAP) بـ(0,454)، والشكل (4) يوضح مواقع الفروق بين متوسطات الأخطاء المعيارية لطرق التقدير.



الشكل رقم(4): العلاقة بين طريقة التقدير ومتوسط الخطأ المعياري لتقدير القدرة – حجم عينة 1000 –

يظهر الشكل (4) عند استخدام عينة مكونة من 1000 فرد أن طريقة (MAP) أكثر دقة من باقي الطرق الأخرى، وأن طريقة (EAP) أكثر دقة من طريقة (ML) في دقة تقدير القدرة وفق نموذج راش.

مناقشة النتائج:

من خلال النتائج التي خلصت إليها هذه الدراسة والتي سعت إلى تقصي مدى دقة تقدير قدرة الأفراد عند استخدام طرق التقدير المعتمدة في برنامج Bilog-Mg (MAP, EAP, ML) في ضوء مستويات مختلفة من حجم العينة (200،500،1000) فرد لاختبار واحد مكون من 20 فقرة وفق نموذج راش حيث أظهرت جميع نتائج أسئلة الدراسة الثلاثة أن قيم مؤشر الأخطاء المعيارية للتقدير كمؤشر لدقة تقدير قدرات الأفراد حسب طريقة الأرجحية العظمى (ML) باختلاف مستويات حجم العينة قد كانت أكبر من نظيراتها التي تعتمد على أسلوب بييز (Bayes) (طريقة توقع التوزيع البعدي(EAP)، طريقة تعظيم الاقتران البعدي(MAP))، وبالتالي فان زيادة حجم العينة لم يؤثر على ترتيب أفضلية هذه الطرق في دقة تقدير قدرة الافراد، كما بينت نتائج هذه الدراسة أهمية طرق التقدير ومدى اسهامها في تفسير تباين قيم دقة التقدير(الخطأ المعياري للتقدير) من خلال مؤشر الدلالة العلمية (مربع ايتا $\boldsymbol{\eta^2}$) الذي بلغ نسبة 71,6%، 44,2%، 46,9%، عند حجم العينة 200، 500، 1000 على التوالي وهي قيم تقع في المستوى المرتفع، ويعزى أفضلية الطرق التي تعتمد على أسلوب بييز (Bayes) في التقدير على طريقة الأرجحية العظمى (ML) إلى أثر مراعاة التوزيعات القبلية للسمة الكامنة وإلى البناء الرياضي الذي تعتمده كل طريقة حيث نجد طريقة الأرجحية تغالي في مواجهة مشكلة تقدير قدرة المفحوصين الذين أجابوا إجابة صحيحة على كل الفقرات أو العكس مما يجعل الخطأ المعياري للتقدير أكبر من الخطأ المعياري لتقديرات قدرات الأفراد باستخدام طرق بييز (Bayes)، وبشكل عام فان النتائج الموصل اليها في هذه الدراسة تتفق مع الدراسات التي قام بها كل من الشواورة (2013) وفقا للنموذج اللوغاريتمي الأحادي و الثنائي، دراسة الصبح (2014)، دراسة وانغ وفيسبول(Wang & Vispoel, 1998) في أفضلية طرق التي تعتمد على أسلوب بييز (Bayes) في التقدير من طريقة الأرجحية العظمى وتختلف هذه الأخيرة مع دراسة بني عطا (2017) ودراسة الطراونة (2011) عن الدراسة الحالية في أفضلية طريقة توقع التوزيع البعدي(EAP) عن طريقة القيمة العظمى للتوزيع البعدي (MAP) حسب النموذج الثلاثي، كما تتفق الدراسة الحالية مع دراسة جاو جن (Gao & Chen, 2005) في أفضلية طريقة القيمة العظمى للتوزيع البعدي (MAP) عن طريقة عن طريقة الأرجحية العظمى، كما تتفق مع دراسة الشريفين(2012) في أفضلية طريقة توقع التوزيع البعدي(EAP) عن طريقة الأرجحية العظمى.

5.الخلاصة:

أظهرت النتائج تقصي دقة تقدير قدرة الأفراد وفقا للنموذج الأحادي البارامتر (نموذج راش) بالاعتماد على برنامج Bilog_Mg، أن الطرق التي تعتمد على أسلوب بييز (EAP، MAP) في التقدير كانت أفضل من طريقة الأرجحية العظمى (ML) عند استخدام اختبار مكون من 20 فقرة وعند أحجام عينات مختلفة مكونة من (200،500،1000) فرد بينما عادت الأفضلية في التقدير بين الطرق التي تعتمد على أسلوب بييز إلى طريقة القيمة العظمى للتوزيع البعدي (MAP) ويمكن تلخيص ترتيب طرق التقدير من حيث دقة التقدير كما يلي:

1- القيمة العظمى للتوزيع البعدي (MAP).

2- توقع التوزيع البعدي(EAP).

3- الأرجحية العظمى (ML).

بالرغم من أهمية النتائج التي توصلت إليها الدراسة الحالية في تقصي دقة تقدير قدرة الأفراد إلا انه هناك بعض العوامل المؤثرة الأخرى قد تلعب دورا مهما في دقة تقدير قدرات الأفراد ولم يتم ضبطها في الدراسة الحالية مما قد تحد من تعميم النتائج بشكل عام لذا توصي الدراسة الحالية بما يلي:

- دراسة أثر طرق التقدير باختلاف طول الاختبار وحجم العينة وشكل التوزيع على دقة تقدير قدرة الأفراد.

- دراسة أثر طرق التقدير باستخدام النماذج اللوجستية (الثنائي، الثلاثي) الأخرى على دقة تقدير قدرة الأفراد.

- إجراء دراسات تتناول طرق تقدير المستخدمة في الدراسة باستخدام برامج أخرى.

- استخدام الطرق التقدير التي تعتمد على نظرية بييز في تقدير قدرة الأفراد وفق نموذج راش حيث أنها أكثر دقة في ذلك.

قائمة المراجع:

التقي، أحمد.(2013). *النظرية الحديثة في القياس*،ط2،عمان: دار المسيرة.

الشريفين، نضال كمال. (2013). تقصي أثر عدد الخطوات الأسئلة متعددة التدريج وشكل التوزيع لصعوباتها على تقديرات القدرة للأفراد، والصعوبة للأسئلة، ودالة المعلومات للاختبار وفق نموذج التقدير الجزئي. *المجلة التربوية*، 28 (109)، 213-275

الشواورة، شادي يوسف. (2013). *دقة تقدير معالم الفقرات بطريقتي الأرجحية العظمى الهامشية وبييز في ظروف مختلفة في عدد الفقرات وحجم العينة والنموذج اللوغاريتمي المستخدم*. رسالة دكتوراه ، الجامعة اليرموك ، الأردن.

صلاح مراد وأمين سليمان. (2002). *الاختبارات والمقاييس في العلوم النفسية والتربوية – خطوات إعدادها وخصائصها*، القاهرة، دار الكتاب الحديث.

عبابنة، عماد غصاب (2008). جدوى التحول من النظرية الكلاسيكية في القياس إلى نظرية الاستجابة للفقرة. المؤسسة العربية للاستشارات العلمية وتنمية الموارد البشرية ، 9(27) ، 392-412

عبدالله، زياد أحمد. (2012). *أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدريج*. رسالة دكتوراه، جامعة القاهرة، القاهرة.

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

كاظم، أمينة.(1988). *استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس*، الكويت: مطبوعات جامعة الكويت.

الطراونة، أرياف أحمد. (2011). *المقارنة بين طرق تقدير القدرة باستخدام النموذج المناسب في ضوء الخطأ المعياري في تقديرها* . رسالة ماجستير، جامعة مؤتة، الأردن.

حمادنة، مروان عبد الله ذياب. (2011). *فاعلية أسلوب تحسين مطابقة الفرد القائم على تصحيح تقدير القدرة وتوزيعها المرجعي عند الاختلاف في حجم العينة والنموذج اللوجستي*. رسالة دكتوراه، جامعة اليرموك, اربد.

الصبح، عامر محمود سليم. (2014). *المقارنة بين دقة تقدير القدرة باختلاف طول الاختبار وشكل توزيع معلمة القدرة تبعا للنموذج اللوجستي ثلاثي المعلمة باستخدام بيانات حقيقية ومولدة*. رسالة ماجستير، جامعة مؤتة، مؤتة.

علام, صلاح الدين محمود. (2014). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*، القاهرة، دار الفكر العربي.

الشريفين، نضال كمال محمد. (2012). *أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. المجلة التربوية*، 26(104)، 177-238.

الدرابيع، ماهر يونس. (2001). *فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة نموذج راش في دقة تقدير قدرة الفرد ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار. مجلة دراسات العلوم التربوية*، 28(1)، 197-208.

شرفاوي، عبد الكريم (2022). طرق تقدير معالم الفقرة والقدرة وأثرها في دقة التقدير باستخدام نظرية الاستجابة للمفردة (راش نموذجا) (رسالة دكتوراه). جامعة الجزائر 2، الجزائر.

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

Hambleton, R. K, & Swaminathan, H. (1985). *Item Response Theory Principles and Applications*. Boston Kluwer – Nijhoff Publishing.

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, 12(3), 535-556.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*.Newbury Park : Sage Publication.

Han, K. T., & Hambleton, R. K. (2007). *User's Manual WinGen*. Center for Educational Assessment Report No. 642, Amherst, MA: University of Massachusetts, School of Education.

Baker, Frank B. (2001). *The basics of item response theory* (2ed Ed). ERIC Clearinghouse on Assessment and Evaluation: USA.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Du Toit, M. (2003). *IRT from SSI: Bilog-MG, multilog, parscale, testfact*: Scientific Software International.

Gao, F., & Chen, L. (2005). *Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model*. Applied Measurement in Education, 18(4), 351-380.

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