المجلة العربية للقياس والتقويم

"العنوان

أثر طريقة التقدير وعدد فقرات الاختبار وحجم العينة على دقة تقديرات القدرة للأفراد باستخدام نموذج التقدير الجزئي

اعداد

نادر سالم العنزي

باحث دكتوراة في القياس والتقويم ، قسم علم النفس ، كلية التربية ، جامعة الملك سعود

أ.د / إسماعيل سلامه البرصان

أستاذ القياس والتقويم ، قسم علم النفس ، كلية التربية ، جامعة الملك سعود

أثر طريقة التقدير وعدد فقرات الاختبار وحجم العينة على دقة تقديرات القدرة للأفراد باستخدام نموذج التقدير الجزئي

الملخص

هدفت الدراسة الى فحص إثر كل من طرق التقدير الارجحية العظمى (ML) والطريقة البايزية (BE) وحجم العينة وعدد فقرات الاختبار على دقه تقدير القدرة للأفراد باستخدام نموذج التقدير الجزئي (PCM)، تم استخدم منهج المحاكاة التجريبي المقارن اذ تم توليد بيانات افتراضيه لأحجام عينات متفاوتة (٢٥٠ – ٥٠٠ – ١٠٠٠ – ٥٠٠٠)، واطوال فقرات (١٠ – ٢٠ – ٣٠)، فقره وتم تحليل البيانات باستخدام برنامج (Excalibur) كما اجرى تحليل التباين المختلط (Mixed ANOVA) باستخدام برنامج (SPSS) للكشف عن دلاله الفروق في دقه تقدير القدرة للأفراد لكل ظرف من ظروف الدراسة وقد اظهرت نتائج الدراسة، ان اثر طريقه التقدير ذو آثر دال احصائيا على دقه تقدير القدرة حيث كان حجم الاثر متوسط و حجم العينة وطول الفقرات ذو اثر دال احصائيا على دقه تقدير القدرة ولكنه ذو حجم اثر ضعيف لحجم العينة وذو حجم اثر كبير لطول الفقرات وبالنسبة للتفاعل الثنائي والثلاثي فكان ذو اثر دال احصائيا على دقه تقدير القدرة ولكنه ذو حجم اثر ضعيف.

الكلمات المفتاحية: طرق التقدير، معالم المفردات، معالم الأفراد، الدقة، نموذج التقدير الجزئي.

The Effect of Estimation Method Sample Size, and Number of Questions on The Accuracy of the Estimation of Person Parameters Using the Partial Credit Model

Abstract

This study aimed to examine the impact of estimation methods—Maximum Likelihood (ML) and Bayesian Estimation (BE) in addition to sample size and test length, on the accuracy of ability estimation using the Partial Credit Model (PCM). A comparative simulation approach was employed by generating synthetic datasets for varying sample sizes (250, 500, 1000, and 5000) and test lengths (10, 20, and 30 items). Data were analyzed using Excalibur software, and a mixed-design ANOVA was conducted in SPSS to determine the statistical significance of differences in ability estimation accuracy across conditions. The results indicated that the estimation method had a statistically significant effect on the accuracy of ability estimation, with a medium effect size. Both sample size and test length also showed statistically significant effects; however, the effect size was small for sample size and large for test length. As for the two-way and three-way interactions, they were statistically significant but demonstrated small effect sizes.

Keywords: Estimation Methods, Item Parameters, Ability Estimation, Accuracy, Partial Credit Model (PCM)

المقدمة:

تقوم نظرية الاستجابة للمفردة على فرضية أساسية أن القيمة الاحتمالية لاستجابة الفرد لمفردة اختبارية تكون دالة لكل من السمة أو القدرة التي يقيسها الاختبار لدى الفرد وخصائص الفقرة التي يحاول الإجابة عنها الفرد حيث جاءت نظرية استجابة الفقرة ونماذجها ثنائية التدريج (Dichotomous) ومتعددة التدريج (Polytomous) كنتيجة لمحاولات علماء القياس النفسي والتربوي مثل (ساميجيما ١٩٧٢، Samejima ورايت ١٩٧٧ ، Wright ، لورد ١٩٨٠ ، Lord، هامبلتون ١٩٨٧، Hambleton) للتوصل إلى تحقيق مطالب القياس الموضوعي ولتطوير مقاييس أكثر دقة في قياسها للسمات النفسية والتربوية ، وتفترض هذه النظرية إمكانية التنبؤ بأداء الفرد في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات . وهي تفترض وجود واحد أو أكثر من هذه السمات تكمن خلف استجابات الفرد لبنود الاختبار، وهذه السمات لا تلاحظ مباشرة، وإنما يكون ذلك من خلال إجابة الفرد على المفردات. ومن خلال نماذج استجابة الفقرة تجري محاولة تقدير للسمة أو السمات التي تنطوي عليها مجموعة من الاستجابات لمجموعة من المفردات. وحيث تتأثر معلم الأفراد والمفردات بعدد من المتغيرات مثل حجم العينة، وعدد المفردات، ونوع النموذج المستخدم وغيرها من المتغيرات. حيث يجب دراسة وفحص المتغيرات المؤثرة في دقة تقدير المعالم للوصول إلى الدقة المطلوبة للقياس الموضوعي.

وتكمن الفكرة الأساسية لنماذج الاستجابة للمفردة من أنها تحاول اشتقاق قياسات أو قيم تقديرية للسمة، التي تنطوي عليها مجموعة من الاستجابات لمجموعة من الفقرات، وتعد هذه النماذج معادلات رياضية تتطلب الحل لتقدير قيم المتغيرات فيها، والمتمثلة بتقدير معلمة قدرة الأفراد (θ) ومعلمة الفقرات المتمثلة في الصعوبة والتمييز والتخمين، ولحل هذه المعادلات ظهر العديد

من الطرق التي تقدر قدرة الأفراد ومعالم الفقرة الكامنة، وراء نمط الاستجابة الظاهرة ومنها طريقة الأرجحية القصوى Maximum Likelihood Estimation (MLE) ، وتعد من الطرق الرئيسية في تقدير معالم المفردات والأفراد من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها عند توفر المعلومات عن العينة (Frank & Ragmen, 2009). حيث تتعامل هذه الطريقة التي طورها بوك وايتكن (Bock & Atiken) عام ١٩٨١ مع مستويات القدرة غير المعلومة وذلك بالتعبير عن احتمالات انماط الاستجابات بتوقعات (Expectaions) مع توزيع مجتمع معين، حيث تستخدم توزيع افتراضيا قبليا لكل من المعالم التي يفترض تقديرها ويتم في عمليه التقدير استخراج الوسط الحسابي للتوزيع البعدي اضافه الى المنوال، واقترح بوك استخدام التوزيع الطبيعي كتوزيع قبلي للقدرة الكامنة وراء استجابات الافراد، واعتبر ان اي نمط من الاستجابات التي يقدمها الافراد في اي اختبار ترتبط مع عينه يتم اختيارها عشوائيا من توزيع قبلي معتمد وبذلك تختلف هذه الطريقة عن غيرها باعتمادها على توزيعات قبليه يتم الاعتماد عليها في ايجاد توقعات لكل من اعداد الافراد الذين اجابوا عن اي من الفقرات اجابه صحيحه وتكون قدرتهم مساويه لأي من القدرات التي يمكن الافراد اخذها في توزيع القدرة المعتمدة من جهة ولأي من اعداد الافراد المقابلة لمختلف القدرات التي تتضمنها التوزيعات الاخرى(التقي،٢٠٠٩). وتعد هذه الطريقة من أكثر طرق التقدير شهره وشيوعا وخاصه في الاختبارات التكيفية المحوسبة وفي بنوك الفقرات وغيرها من تطبيقات نظرية الاستجابة للفقرة حيث يتم تقدير معلمه القدرة من خلال نمط استجابة المفحوص على مجموعة من فقرات الاختبار وتأخذ هذه الإجابة القيم واحد إذا كانت صحيحه وصفر إذا كانت خطا وتحسب داله الأرجحية العظمى من خلال ايجاد مجموع حاصل ضرب نمط اجابه المفحوص عن فقرات الاختبار (Bellhouse,2004: Fo,2010). وتتمثل محددات هذه الطريقة انه عندما يجيب المفحوص عن جميع الفقرات اجابه صحيحة فان قيمة (θ) المقدرة تساوي (∞) وعندما يجيب اجابه خطا عن جميع الفقرات فان قيمة (θ) المقدره تساوي ($-\infty$) وبالتالي يصعب الحصول على قيمه قصوى للأرجحية في هذه الحالة ويتم معالجه هذه المشكلة في الاختبارات المبنية على القياس التكيفي المحوسب Computerized Adaptive Testing (CAT) من خلال وضع تقديرات عليا ودنيا لمثل هذه الانماط من الإجابة (Thompson,2009 &

.(Baker,2004

والطريقة الأخرى طريقة بييز (Bayes Estimation)، التي تعد من الأساليب الإحصائية المستخدمة لتقدير معالم الفقرات وقدرة الأفراد (θ) مع الأخذ بالحسبان المعلومات السابقة للمعلمة غير المعروفة في عملية التقدير معتمدا على خبرة الباحث، أومن خلال الخصائص الإحصائية لذلك المعلم الذي يتم تقديره (Swiminithan & Gifford, 1982). التي اقترحها كل من بوك ومسليفي (Bock & Mislevy) في عام ١٩٨٢ وهي طريقه لا تعتمد على استخدام التقدير الدوراني الذي يعتمد طرق تقديريه تكراريه كطريقه نيوتن رافسون وانما بطريقة مباشرة يستخدم التوزيع الطبيعي المعياري حيث يتم تقسيم السمة الكامنة التي تمثل هذا التوزيع وتقع بين (3، 3−) بالعادة الى(61) جزء بفترات طولها (0.1) يطلق على كل منها ربيعا (Quadrant) كما يعطى لكل من هذه الربيع الربيعات وزنا (Weight)، ويمثل إما المساحة الواقعة فوق التجزئة (r , r+1) ، او قيمة الاحتمال في التوزيع المعياري الاعتدالي. وتمثل قيمه القدرة المستخرجة الوسط الحسابي بينما تمثل القدرة المستخرجة باستخدام طريقه الأرجحية العظمى وطريقة القيمة العظمى للتوزيع البعدي المنوال للتوزيع في كل منهما(التقي،٢٠٠٩).

مشكلة البحث

يعد تحديد موقع الافراد والفقرات على متصل السمة المشترك هو الغرض الاساسي لتطبيق المقياس وفقا لنظرية الاستجابة للفقرة (Baker,2001/2010). كما أبدى علماء القياس اهتماما كبيرا بمفهوم تقدير المعالم لما له من اهمية في مجال القياس النفسي والتربوي وقد استخدم كثير من الباحثين نماذج نظرية الاستجابة للفقرة وطرق التقدير المختلفة في تقدير هذه المعالم سعيا للوصول الى افضل دقه تقدير لهذه المعالم فتباينت نتائج هذه الدراسات حول تأثير طرق التقدير على دقه تقدير معالم الفقرات والافراد، فبعض الدراسات اكدت على وجود فروق في دقة تقدير معالم الفقرات والافراد تعزى لطريقة التقدير مثل دراسة كوريسو واخرين(Kurisu, et al., 2022)، ودراسة القرني (٢٠٢٢)، ودراسة صن واخرين (Sun, et al., 2012)،

ودراسة (الطراونة ، ٢٠١١)، ودراسة وانغ وانغ (Wang & Wang, 2001). فيما اكدت نتائج دراسات اخرى على عدم وجود فروق ذات دلاله إحصائية بين طرق تقدير القدرة كدراسة ديوانتي واخرون (Dewanti, et al., 2021)، ودراست هوو(Ho, 2010)، ودراسة كيم(Kim, 2001)، حيث تتحدد إشكالية الدراسة في دقة تقدير معالم المفردات والأفراد لنظرية الاستجابة للمفردة من خلال عدد من المتغيرات؛ مثل حجم العينة وعدد الفقرات وطريقة التقدير حيث تمت دراسة هذه المتغيرات في ضوء النماذج ثنائية الاستجابة. في حين انه لا توجد الكثير من الدراسات التي اهتمت في دراسة هذه العوامل مجتمعة في النماذج متعددة الاستجابة لاسيما نموذج التقدير الجزئي Partial Credit Model. (PCM) بناء على ذلك تبنت الدراسة الحالية مشكلة البحث التي تتقصى فعالية دقة تقدير معالم المفردات والأفراد لنموذج التقدير الجزئي (PCM) في ضوء عدد من المتغيرات (حجم العينة وعدد الفقرات وطريقة التقدير) والتي لم يتم التطرق اليها في البحوث والدراسات السابقة. نظرا لحاجة مجال القياس العلمي والتطبيقي والاكاديمي لهذا النوع من الدراسات الداعمة والمعززة لميدان القياس والتقويم النفسي والتربوي؛ ونظرًا لقلة الأبحاث والدراسات التي تناولت دقة تقدير معالم المفردات والأفراد لنماذج نظرية الاستجابة للمفردة عديدة الاستجابة ؛ بالإضافة الى تحديد أي طرق التقدير دقة في تقدير المعالم لنموذج التقدير الجزئي (PCM) ، حيث اشتملت طرق تقدير المعالم على طريقتين الأولى اعتمدت على الأرجحية القصوى Maximum Likelihood Estimation والثانية اعتمدت على طريقة بييز Bayes Estimation فعند تناول أي نموذج من نماذج الاستجابة للمفردة بالبحث عبر دقة تقدير معالمه وما يؤثر عليها فإن جميع نتائج البحث تكون محكومة بطريقة التقدير التي تم اختيارها وبالتالي قد نحصل على نتائج أكثر دقة عند استخدام طريقة أخرى في تقدير المعالم ؛ ولذلك جاءت الحاجة لإجراء مثل هذه الدراسة التي سوف تتناول متغير "حجم العينة" و عدد الفقرات" و طريقة التقدير" ودورهم في تقدير دقة تقدير معالم المفردات والأفراد لنموذج التقدير الجزئي (PCM).

أسئلة الدراسة:

١) ما أثر طريقة التقدير وعدد المفردات وحجم العينة على دقة تقدير معالم الأفراد باستخدام نموذج التقدير الجزئي؟

٢) ما أثر التفاعل بين طريقة التقدير وعدد المفردات وحجم العينة على دقة تقدير معالم الأفراد باستخدام نموذج التقدير الجزئي؟

أهداف الدراسة: تهدف الدراسة الى تحقيق الأهداف التالية:

١) التعرف على أثر حجم العينة وعدد المفردات وطريقة التقدير على دقة تقدير معالم الأفراد باستخدام نموذج التقدير الجزئي.

٢) التعرف على أثر التفاعل بين حجم العينة وعدد المفردات وطريقة التقدير على دقة تقدير معالم الأفراد باستخدام نموذج التقدير الجزئي.

أهمية الدراسة:

للدراسة الحالية أهمية من الناحتين النظرية والتطبيقية.

تتمثل الأهمية النظرية للدراسة في أنها تتعلق بدقة تقدير معالم الأفراد لنموذج التقدير الجزئي وهو ما يسعى إليه القياس للوصول إلى نتائج أكثر دقة وموضوعية للتعبير عن السمة المقاسة كما أنها تتمثل في أنها إضافة إلى دراسات طرق التقدير لنماذج نظرية الاستجابة للمفردة عديدة الاستجابة وذلك لندرة الدراسات العربية التي تناولت هذا الموضوع وعليه فإن الدراسة تعمل على توجيه انتباه الباحثين والمهتمين بقضايا القياس في البيئة العربية إلى أهمية هذا الجانب من الموضوعات.

أما الأهمية التطبيقية فتتمثل في الكشف عن أفضل طريقة لتقدير معالم الأفراد لنموذج التقدير الجزئي عند أحجام وأطوال مختلفة للمفردات والأفراد، وكذلك لتشجيع الباحثين لاستخدام نظرية الاستجابة للفقرة في تطوير المقاييس النفسية وتحليل الاختبارات وذلك من خلال توليد عدد كبير من البيانات وتحليلها باستخدام البرامج الحاسوبية الحديثة، لإزالة الخوف والرهبة من تلك البرامج الكمبيوترية. وفتح المجال للبحوث والدارسات العربية التي تستخدم المحاكاة للتحقق من القضايا في الجوانب السيكومترية المتعلقة باستخدام نظرية الاستجابة للفقرة والعوامل التي تؤثر في دقة التقديرات لندرتها في البحوث والدارسات في الوطن العربي – في حدود علم الباحث –

حدود البحث:

١. تقتصر الدراسة على استخدام طريقة الأرجحية القصوى (ML) وطريقة بايز (BE) في التقدير.
٢. تقتصر الدراسة على استخدام نموذج التقدير الجزئي (PCM) لماسترز.
٣. سوف تناول الدراسة بيانات متعددة التدريج بخمس فئات استجابة.
٤. تقتصر الدراسة على استخدام بيانات مولدة بأسلوب المحاكاة (Simulation Study) لـ (١٠، ٢٠، ٣٠) فقرة متعددة التدريج، وفق لأحجام عينات مختلفة تتراوح بين (٢٥٠، ٥٠٠، ١٠٠٠، ٥٠٠٠).

مصطلحات البحث:

١. طرق التقدير (Estimation Method): مجموعة من الأساليب الرياضية الاحتمالية التي تهدف لتقدير معالم الفقرات وقدرات الأفراد في ضوء افتراضات نظرية الاستجابة للفقرة (ضعضع وآخرون، 2020).
٢. معالم الأفراد (Person Parameter): وهي معلمة القدرة وتعرف بموقع الفرد، وتعرف إجرائيا بأنها قدرات الأفراد التي نحصل عليها من نتائج تحليل البيانات المولدة في ضوء نموذج التقدير الجزئي.
٣. الدقة (Accuracy): يشير إلى جودة التقدير لمعالم الفقرات والأفراد، والتي تتميز بالاحتمال في أن يكون التقدير قريبا من القيمة الحقيقية للمعلم، ويعرف إجرائيا بأنه التقدير غير المتحيز صاحب أقل تباين بين التقديرات الأخرى غير المتحيزة، ويحدد باستخدام الخطأ المعياري في التقدير.
٤. نموذج التقدير الجزئي (Partial Credit Model): أحد نماذج نظرية الاستجابة للفقرة الذي يعد امتدادا لنموذج راش ويعرف إجرائيا بأنه: النموذج اللوغاريتمي الذي يفترض أن معامل الصعوبة هو البارامتر الوحيد المؤثر على أداء الأفراد.

الإطار النظري:

نظرية الاستجابة للمفردة: Item Response Theory

يوضح هامبلتون وسوامينثان (Hambleton & Swiminathan, 1985)، إلى مجموعة من المزايا التي تتميز بها نظرية الاستجابة للفقرة وهي:

تقدم نظرية الاستجابة للفقرة مؤشرا إحصائيا يقيس درجة الدقة في قياس القدرة (θ) لكل فرد ، وكذلك تقدير قدرة الفرد (θ) يتم بشكل مستقل عن عينة الفقرات التي تطبق على الفرد، وتعني أن الفقرات المستخدمة في تقديرات الأفراد متحررة عن خصائص الفقرات (Sample Free) ، وأيضا تقدير معالم الفقرات يكون مستقلا عن عينة الأفراد (Sample Free)

افتراضات نظرية الاستجابة للمفردة:

تقوم نظرية الاستجابة للمفردة على عدد من الافتراضات ذكرها هامبلتون وسواميناثان(1985) Hambleton & Swaminathan وهي:

١. أحادية البعد (Unidimensionality): إي أن الاستجابات على المقياس تعزى الى سمة مفردة (جميع الفقرات تقيس سمة واحدة).
٢. الاستقلال الموضعي (Local Independence): ويعني الاستقلال الاحصائي لاستجابات الفرد للفقرات المختلفة في المقياس.
٣. منحنى خصائص الفقرة (Item Characteristic Curve): وجود دالة مميزة لكل فقرة من فقرات المقياس تربط بين احتمال الإجابة الصحيحة على فقرات المقياس والقدرة التي تقيسها فقرات المقياس.
٤. التحرر من السرعة (Speedness): يعني ان عامل السرعة لا يلعب دوراً في الإجابة على الفقرة.

نماذج نظرية الاستجابة للفقرة

وهي ثلاثة نماذج (علام، ٢٠٠٥):

١) النموذج اللوغاريتمي آحادي المعلمة One Parameter Logistic Model

يقدر هذا النموذج احتمالية (P) إجابة الفرد(J) بدلالة كل من قدرة الفرد الكامنة (θ) وصعوبة الفقرة (b_i) من خلال نمذجتها في صيغة رياضية احتمالية هي دالة الترجيح اللوغاريتمي غير

الخطية التالية:

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

وتسمي الصيغة السابقة نموذج الترجيح اللوغاريتمي أحادي المعلمة (1PLM) لأنها تشمل على معلمة واحدة فقط هي صعوبة الفقرة.

٢) النموذج اللوغاريتمي ثنائي المعلمة Two Parameter Logistic Model

وهو النموذج الذي يسمح للفقرات بأن تختلف في كل من معلمتي الصعوبة (b) والتمييز (a) وبذلك أصبح النموذج يشتمل على معلمتي هي الصعوبة والتمييز:

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

٣) النموذج اللوغاريتمي ثلاثي المعلمة Three Parameter Logistic Model

في هذا النموذج تم إضافة معلماً ثالثاً يمثل احتمال توصل الأفراد للإجابة الصحيحة على الفقرة عن طريق التخمين (C) وصيغته كالتالي:

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

نموذج التقدير الجزئي Partial Credit Model (PCM)

ابتكر ماسترز(Masters) عام ١٩٨٢ هذا نموذج التقدير الجزئي وهو ينتمى إلى نماذج راش اللوغاريتمي لذلك فهو يكتسب نفس مزايا نماذج الاستجابة للفقرة بالنسبة من حيث التحرر من معالم الفقرات والأفراد كما أنه يقدر المعارف الجزئية مثل المهارات التي تتطلب خطوات متعددة مثل حل المسائل الحسابية كما انه يناسب تحليل الاستجابات على مقياس الاتجاهات الشخصية التي تعتمد على موازين التقدير(Reeve, 2004). ويختلف نموذج التقدير الجزئي(PCM)، عن نموذج الاستجابة المتدرجة(GRM)، في انه يعد نموذجا مباشرا من نماذج الاستجابة للفقرة لان احتمال الاستجابة في قسم معين يتم التعبير عنه مباشره كحد جبري اسي مقسوما على مجموع الحدود الجبرية الأسية (علام، 2005).

معادلة منحنيات اقسام الاستجابة لنموذج التقدير الجزئي

$$p(x_{j\theta}.\delta_{jh}) = \frac{\exp \sum_{h=0}^{xj}(\theta-\delta_{jh})}{\sum_{k=0}^{mj} exp \sum_{h=0}^{k}(\theta-\delta_{jh})}$$

فاذا كانت اقسام الاستجابة للفقرة (n) من الاستجابات فان الفقرة يتم الإجابة عنها من خلال عدد (n−1) من الخطوات بحيث يتم الانتقال من القسم الثاني الى القسم الثالث من اقسام الاستجابة وذلك خلال خطوه صعوبتها (δ_2) وهكذا ويعبر عن (δ) بانها نقطة على ميزان السمة الكامنة يتقاطع عندها منحنى الاستجابة لقسمين متتاليين ويطلق عليها صعوبة الخطوة و تشير الى مدى اتمام الفرد لكل خطوة (Akkermans & Muraki,1997).

وتتمثل خطوات الفقرة في الانتقال بين البدائل او اقسام الاستجابة في فقرات الاستبانات او مقاييس الاتجاهات فمثلا اذا كانت اقسام الاستجابة اوافق اوافق بشده لا اوافق لا اوافق بشده فان عدد حل هذه الفقرة عباره عن ثلاث خطوات ولا يلزم في هذا النموذج ان يكون الفرق بين خطوات الفقرة ثابت لدى جميع الفقرات وقد يكون الفرق كبيرا في فقره وصغير في فقره اخرى وتكون بعض الخطوات داخل كل فقره اقل سهوله او اكثر صعوبة نسبيا عن غيرها من الخطوات (علام، ٢٠٠٥).

طرق أو أساليب تقدير المعالم

طريقة الارجحية العظمى Maximum Likelihood Method (ML)

تعتبر من الطرق الواسعة الانتشار حيث يتم تقدير المعالم فيها من خلال تعظيم الاحتمالية للمعلمة المراد تقديرها ويكون شكل داله الأرجحية (Likelihood Function)، (L) اذا افترضنا ان $P_i(\theta)$ احتمال استجابة فرد ذو قدره (θ) على الفقرة (i) اجابة صحيحة ومنه يمكن اعطاء العلاقة التالية كتعبير عن احتمال الاستجابة على الفقرة كالتالي يرمز $Q_i(\theta)$ لاحتمال الإجابة الخاطئة للفرد ذوي القدرة (θ) على الفقرة (i) حيث ان$Q_i(\theta) = 1 - P_i(\theta)$ يأخذ القيمة (١) اذا كانت الإجابة صحيحة والقيمة (٠) اذا كانت الإجابة خاطئة على الفقرة، بصورة عامه فان الصيغة الرياضية لدالة الأرجحية لأنماط الاستجابات لعدد من الافراد بواسطه

مجموعه من الفقرات ذوي مستويات قدره ثابتة.

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

من خلال هذه الصيغة التي تبين احتمال اجابات الافراد من ذوي مستوى قدرة معين هو حاصل ضرب الاحتمالات لجميع المفردات (علام، ٢٠٠٥).

طرق التقدير القائمة على نظرية بييز Bayesian Estimation (BE)

تستخدم هذه الطريقة عندما لا يكون بالإمكان تطبيق الأرجحية العظمى وذلك بوجود الاستجابات المتطرفة او عندما يجيب المفحوص عن جميع فقرات الاختبار إجابه صحيحه او إجابه خاطئة حيث تعتمد هذه الطريقة في التقدير على نظريه بييز والتي تربط بين الاحتمالات الشرطية والاحتمالات الهامشية وتتطلب هذه الاساليب افتراض احتمالات قبليه للمعالم في ضوء اعتبارات نظريه او امبريقية بمعنى انه اذا توافرت معلومات عن توزيع القدرة لدى مجموعه من المفحوصين فان الاساليب التي تعتمد على نظريه بييز يمكن ان تساعد في الحصول على تقديرات جيده لهذه القدرة حيث يمكن افتراض اي شكل من اشكال التوزيعات (علام، ٢٠٠٥) ومن هذه الأساليب أو الطرق:

طريقة التوقع البعدي Expected A Posterior

اقترح هذه الطريقة في التقدير بوك ومسليفي (Bock & Mislevy) عام ١٩٨٢ وهي طريقة لا تعتمد على استخدام التقدير الذي يعتمد الطرق التقديرية التكرارية كطريقة نيوتن رافسن دون اللجوء الى تقريب متتابع كما في طريقة الارجحية العظمى حيث يتم التقدير بطريقة مباشرة، حيث يستخدم التوزيع الطبيعي المعياري و يتم تقسيم السمه الكامنة التي تمثل هذا التوزيع وتقع بين −٣ الى ٣ الى ٦١ جزء بفترات طولها ٠,١ يطلق على كل منها ربيع (Quadrant) كما يعطي لكل من هذه الربيعات وزنا ويمثل اما المساحة الواقعة فوق التجزئة (r+1 , r)، او قيمة الاحتمال في التوزيع المعياري الاعتدالي وتمثل قيمه القدرة المستخرجة المتوسط الحسابي للتوزيع (التقي،٢٠٠٩).

يعمل هذا الأسلوب على إيجاد تقديرات القدرة للأفراد بالاعتماد على تقديرات معالم الفقرات من

خلال طريقة الارجحية العظمى الهامشية ويحسب التقدير للتوقع البعدي (EAP) من خلال المتوسط الحسابي للتوزيع البعدي للقدرة دون اللجوء الى تقريب متتابع كما في طريقة الارجحية العظمى

طريقة القيمة العظمى للتوزيع البعدي (Maximum A Posteriori):

يحتاج اسلوب الأرجحية العظمى الى عينة كبيره من الفقرات لأنه من الاساليب التي تستخدم مفهوم التقاربية في تقدير السمه الكامنة، كما ان طريقة القيمة العظمى للتوزيع البعدي تستخدم نفس الاسلوب باستثناء استخدام التوزيع القبلي للسمه الكامنة والذي هو في العادة التوزيع الطبيعي المعياري ويعرف التوزيع البعدي من خلال ضرب دالة الارجحية بدالة التوزيع القبلي (التقي،٢٠٠٩)، وفي هذه الطريقة يوجد افتراض مسبق للقدرة يتعلق بشكل التوزيع ويعتبر نقطه الانطلاق لدالة الأرجحية بحيث تتضاعف كل التقديرات المتتالية من خلال المنحنى الطبيعي وتتشابه هذه الطريقة مع طريقه الأرجحية العظمى في انهما يقدمان قيم الأرجحية العظمى لدالة الأرجحية وتختلف عنها في ان دالة الأرجحية في طريقه الأرجحية العظمى تكون بمضاعفه التوزيع المسبق وانتاج ما يعرف بالتوزيع البلدي(حبيب وعزيز،٢٠١٨).

الدارسات السابقة:

على الرغم من حداثة نشأة دراسات المحاكاة نسبيا في الوطن العربي إلا أنها لقيت اهتماما كبيرا من قبل الباحثين والمهتمين في الآونة الأخيرة ونتج عن ذلك إجراء العديد من الدراسات النظرية والميدانية في محاولة التحقق من بعض المتغيرات مثل حجم العينة وعدد الفقرات وطريقة التقدير، التي تؤثر على تقديرات معالم الفقرات والأفراد لنماذج نظرية الاستجابة للفقرة، وأظهرت العديد من النتائج تباين في دقة التقدير المعالم وفقاً للعوامل المؤثرة لنماذج نظرية الاستجابة للفقرة. الامر الذي يدعو الى أجراء المزيد من الدراسات للوصول إلى دقة التقدير لمعالم الفقرات والأفراد.

المحور الأول: الدراسات التي تناولت تأثير حجم العينة وطول الاختبار على دقة تقدير المعالم

هدفت دراسة بني عطا (٢٠١٧) إلى تقصي أثر طول الاختبار وحجم العينة على دقة طرق تقدير معالم الفقرات وقدرات الأفراد في برنامج بايلوج. تم توليد بيانات ثنائية التدريج من خلال النموذج 3PL لأربعة اختبارات طول كل منها (10، 30، 60، 80) وأربعة مستويات لحجم العينة (٢٥٠، 500، ١٠٠٠، 5000) باستخدام برنامج (WINGEN). وباستخدام طرق التقدير (الأرجحية العظمى، توقع الاقتران، تعظيم الاقتران) المستخدمة في برنامج(Bilog–Mg3). أظهرت نتائج الدراسة وجود أثر ذي دلالة إحصائية لكل من طول الاختبار وحجم العينة والتفاعل بينهما في دقة تقديرات معالم المفردات (التمييز، الصعوبة، التخمين) عند استخدام طريقة الأرجحية العظمى في معايرة الفقرات، وأن دقة تقديرات معالم الفقرات تزداد بزيادة طول الاختبار وحجم العينة. وبينت كذلك النتائج بأنه عند معايرة قدرة الأفراد بطرق التقدير الثلاثة المستخدمة في برنامج بايلوج عن وجود أثر ذي دلالة إحصائية لكل من طول الاختبار وحجم العينة وطريقة التقدير في دقة تقديرات معلمة القدرة للفرد، حيث تفوقت طريقة توقع الاقتران (EAP) على بقية الطرق في دقتها لتقدير معلمة القدرة، وأظهرت النتائج بشكل عام أن طرق التقدير المستخدمة في برنامج بايلوج أنتجت تقديرات دقيقة لمعالم الفقرات وقدرات الأفراد عندما كان طول الاختبار (٣٠) فقرة وأعلى وحجم العينة (٥٠٠) فأعلى عند استخدام النموذج اللوجستي ثلاثي المعلمة.

وفي دراسة تجريبية حديثة لشاهين وانيل (Şahin& Anıl, 2019) لتفحص تأثير حجم العينة وطول الاختبار على دقة تقدير معلمات الفقرة باستخدام (MMLE) في ظل النماذج ثنائية التدريج ١PL و٢PL و٣PL من خلال اختبار لغة مكون من ٥٠ مفردة تم تطبيقه على ٦٢٨٨ فردًا. تم تدريج العينة الى تسعة مستويات مختلفة (١٥٠ و٢٥٠ و٣٥٠ و٥٠٠ و٧٥٠ و١٠٠٠ و٢٠٠٠ و٣٠٠٠ و٥٠٠٠) وثلاثة اختبارات أطوال كل منها (١٠، ٢٠، ٣٠) من الاختبار الأصلي. تم استخدام طريقة MMLE لتقدير معلمات المفردات. أشارت النتائج إلى أنه عندما يكون حجم العينة يبلغ ١٥٠ فردًا، يُظهر نموذج ١PL تقديرًا دقيقًا لمعلمة صعوبة الفقرة حتى عندما يتكون الاختبار من ١٠ فقرات فقط. في المقابل، نموذج ٢PL، يتطلب حجم عينة كبير (٧٥٠) عندما يتكون الاختبار من ١٠ فقرات فقط؛ اما إذا كان الاختبار مكون من ٢٠ مفردة فلابد من توفر عينة بحجم ٥٠٠؛ وعندما يكون الاختبار مكون من ٣٠ مفردة، فأن حجم

العينة المطلوب هو ٢٥٠ لإنتاج تقديرات دقيقة لصعوبة المفردات ومعلمات التمييز بينها. اما عندما يكون الاختبارات مكون من ١٠ و٢٠ مفردة، يتطلب نموذج 3PL حجم عينة يبلغ ٧٥٠ مفردة و٣٥٠ مفردة إذا كان الاختبار مكون من ٣٠ مفردة لإنتاج تقديرات دقيقة للصعوبة، التمييز والتخمين بين الفقرات.

وفي دراسة اجراها الثوايبة (2010) هدفت الى استقصاء اثر حجم العينة في تقدير معلمه صعوبة الفقرات والخطأ المعياري في تقديرها ولتحقيق ذلك تم اعداد اختبار لطلبة الصف العاشر الاساسي في الرياضيات وكان عدد فقراته (80) فقره وحجم عينة (11292) فرد وتم توزيعهم الى عدد من العينات وتم استخدام طريقة الارجحية العظمى الهامشية من خلال برنامج (Bilog–MG) لتقدير معلمة الصعوبة والخطأ المعياري في التقدير وقد توصلت الدراسة إلى ان صعوبة الفقرة تزداد بزيادة حجم العينة فقد كانت متوسط الصعوبة (31) لوجيت عندما كان حجم العينة (200) فرد وبسبب زيادة حجم العينة ارتفع متوسط صعوبة الفقرة إلى (1.1) لوجيت وذلك عند حجم عينه بلغ (11292) فرد اما الخطأ المعياري في التقدير تناقص بزياده عدد افراد العينة فبلغ متوسط الاخطاء المعيارية في التقدير(32) لوجيت عندما كان حجم العينة (200) فرد وعند زياده حجم العينة (11292) فرد اصبح الخطأ المعياري في التقدير (0.07) لوجيت.

المحور الثاني: الدراسات التي تناولت نموذج التقدير الجزئي في دقة تقدير المعالم

بينت دراسة مونت كارلو لداي وآخرون(Dai et al., 2021) التي تحققت من إداء بعض النماذج المتعددة التدريج polytomous وعلى وجه الخصوص نموذجي (GRM و GPCM)في دقة تقدير المعالم المفردات والأفراد عندما تكون الاختبارات قصيرة وحجم العينة صغير بالإضافة الى وجود بيانات مفقودة؛ تم توليد بيانات متعددة التدريج لخمسة أطوال للاختبارات وهي كالتالي (٣، ٥، ٧، ١٠، ١٥) وأربعة مستويات لحجم العينة (١٥٠، ٣٠٠، ٦٠٠، ١٢٠٠). كما تم تعيين ثلاث مستويات عشوائية لتمييز حيث تشير قيمة (٠,٤٤، ٠,٧٥) = α للمفردات الضعيفة، (٠,٥٨، ٠,٩٨ = α) للمفردات المتوسطة، و (٠,٧٥، ١,٣٣= α) للمفردات الجيدة. باستخدام طريقة تقدر (MMLE) في برامج (R). أظهرت النتائج

انه في حال كان حجم العينة لا يقل عن ٣٠٠ وطول أداة لا يقل عن خمس مفردات يكون أداء GPCMمستقرًا عبر الاختبار بينما يتحسن أداء GRM بشكل ملحوظ مع زيادة طول لاختبار. اما بالنسبة لمعلمات الفرد، تكشف GRM عن تقديرات أكثر دقة عندما تكون نسبة البيانات المفقودة صغيرة، في حين يتم تفضيل GPCM في وجود كمية كبيرة من البيانات المفقودة.

دراسة لنكير(Linacare, 2002) والتي هدفت الى دراسة تأثير طول الاختبار وحجم العينة على دقة تقدير معالم نموذج التقدير الجزئي (PCM)، حيث اعتمدت الدراسة على تحليل بيانات اختبار تم تطبيقه على (100, 300, 500)، فرد واحتوى على اطوال مختلفة (10, 30, 50)، فقرة وتم استخدام برنامج WINSTEPS لتحليل البيانات وفق نموذج التقدير الجزئي (PCM) واشارت نتائج الدراسة إلى أن طول الاختبار يؤثر على دقة التقدير حيث زادت دقة تقديرات القدرة ومعالم الفقرات عند زيادة عدد الفقرات حيث اظهرت الاختبارات القصيرة ١٠ فقرات مستوى اعلى من خطا التقدير مقارنه بالاختبارات الأطول وان حجم العينة له تأثير ملحوظ على استقرار التقديرات حيث كانت تقديرات المعالم اقل استقرارا في العينة الصغيرة ١٠٠ فرد مقارنه بالعينات الاكبر مما يدل على ان زياده عدد الافراد يؤدي الى تقديرات اكثر دقة.

المحور الثالث: الدراسات التي تناولت طرق التقدير واثرها على دقة تقدير المعالم

كما هدفت دراسة المحاكاة لكل من الطراونة عيسى (٢٠١١) إلى مقارنة تأثير طول الاختبار على درجة تقدير معلمة القدرة في النموذج اللوجستي ثنائي المعلمة وثلاثي المعلمة، باستخدام منهجيتين تقدير وهما: الطريقة البايزية والتقدير بالأرجحية القصوى؛ بإتباع المنهج التجريبي. أظهرت النتائج أن دقة تقدير معلمة القدرة في النموذج اللوجستي ثنائي المعلمة وفق منهجية التقدير بالأرجحية القصوى وطريقة البايزية تزداد مع زيادة عدد عناصر الاختبار. كما أظهرت النتائج أنه في الاختبارات الطويلة والمتوسطة الطول، تزداد دقة تقديرات معلم القدرة مع أسلوب التقدير بالأرجحية القصوى وبجميع شروط حجم العينة، بينما في الاختبارات القصيرة تفوقت الطريقة البايزية في جميع الظروف. أما فيما يتعلق بدقة تقدير معلمة القدرة في النموذج اللوجستي ثلاثي المعلمة تشير النتائج إلى أن دقة تقدير معلمة القدرة تزداد مع زيادة عدد

عناصر الاختبار؛ بينما تتفوق طريقة البايزية فيما يتعلق بدقة التقدير في جميع ظروف حجم العينة، أما في الاختبارات الطويلة تتفوق طريقة الاحتمالية القصوى في جميع الظروف المختلفة.

وفي دراسة أخرى لضعضع وآخرين (٢٠٢٠) التي هدفت إلى المقارنة بين طرق التقدير (الأرجحية العظمى، وتقدير بييز، وطريقة بروكس) عند أحجام عينات مختلفة (٥٠٠، ١٠٠٠، ١٥٠٠، ٢٠٠٠) على تقدير معالم المفردو والأفراد ودقتها في ضوء نظرية الاستجابة للمفردة (نموذج راش)، تم توليد البيانات لـ (٢٠٠٠) فرد و (٤٠) فقرة ثنائية الاستجابة بقدرة وصعوبة ذات توزيع طبيعي وبعد مطابقة البيانات تم مطابقة (٣٠) مفردة، حيث أظهرت النتائج أن هناك فروقا في دقة تقدير معلم الصعوبة وتقدير القدرة للأفراد بازدياد حجم العينة، بينما لم تكن هناك فروق بين طرق التقدير الثلاث، الأرجحية العظمى وطريقة بييز وطريقة بروكس.

وأجرى جيو وشين (Geo & Chen, 2005). دراسة هدفت الى تقدير معالم الفقرات ، الصعوبة و التمييز و التخمين بحجم عينة (١٠٠، ٥٠٠، ٢٠٠٠) فرد واطوال الاختبارات (١٠، ٣٠، ٦٠) فقره ثنائية الإجابة وبعد ان تم توليد البيانات وفق الشروط السابقة على النموذج اللوغاريتمي الثلاثي تم تقدير معالم الفقرات وفق الاسلوبين (MML, Bayes) وتم استخدام برنامج (ETIRM) لتقدير معالم الفقرات وفق طريقة بييز، اما الفقرات المقدرة وفق طريقه الأرجحية العظمى الهامشية قدرت من خلال برنامج (Bilog–MG) وطبقت الاختبارات الثلاثة على عينة مكونه من(١٠٠) فرد ثم طبقت الاختبارات على عينة (٥٠٠) فرد ثم طبق الاختبارات بأطواله الثلاثة على عينه من (٢٠٠٠) فرد واشارت النتائج الى ان تقديرات طريقه بييز اكثر دقة من تقديرات طريقه الأرجحية العظمى اعتمادا على معاملات الارتباط بين القيم المقدرة والقيم الحقيقية عندما يكون حجم العينة صغير، واظهر التوزيع اثرا على نتائج طريقة بييز أما في حالة العينات الكبيرة تظهر التوزيعات السابقة أثرا معتدلا ، وبشكل عام كانت تقديرات ترجيحيه العظمى وبييز تميل الى اعطاء نتائج متشابهة ودقيقة عندما يكون حجم العينة كبيرا وعندما يصبح حجم العينة كبيرا يعطيان نتائج لا يمكن فصلهما حول المعالم الثلاثة في النموذج الثلاثي.

التعقيب على الدراسات السابقة:

ومن خلال استعراض الدراسات السابقة يتضح بشكل عام أن هناك توجها عاما نحو البحث عن دقة تقديرا معالم الفقرات وقدرات الأفراد خصوصا عند استخدام نظرية الاستجابة للفقرة، لما لهذه الدقة للتقديرات من أثر في التطبيقات العملية للاختبارات والمقاييس النفسية في المواقف المختلفة، حيث أجمعت الدراسات على وجود عوامل كثيرة تؤثر في دقة تقديرات معالم الفقرات والأفراد من أهمها طول الاختبار وحجم العينة وطريقة التقدير ويظهر التباين في نتائج الدراسات من حيث الطريقة الأفضل في تقدير المعالم، حيث تتفق الدراسة الحالية مع الدراسات السابقة في سعيها لتقدير دقة تقدير معالم الفقرة والقدرة باستخدام نظرية الاستجابة للفقرة، في حين تختلف الدراسة الحالية مع الدراسات السابقة في أنها تفحص دقة التقدير باختلاف حجم العينة وطول الاختبار وطريقة التقدير لنموذج التقدير الجزئي(PCM).

فرضيات البحث:

١. لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة $(\propto = 0.05)$ بين متوسطات تقديرات معالم الأفراد تبعاً لمتغيري حجم العينة (٥٠٠، ١٠٠٠، ٥٠٠٠، ١٠٠٠٠) وعدد فقرات الاختبار (١٠، ٢٠، ٣٠) وطريقة التقدير (ML, BM) باستخدام نموذج التقدير الجزئي.

٢. . لا توجد فروق ذات دلالة إحصائية عند مستوى الدلالة $(\propto = 0.05)$ بين متوسطات تقديرات معالم الأفراد تبعاً لمتغيري حجم العينة (٥٠٠، ١٠٠٠، ٥٠٠٠، ١٠٠٠٠) وعدد فقرات الاختبار (١٠، ٢٠، ٣٠) وطريقة التقدير (ML, BM) باستخدام نموذج التقدير الجزئي والتفاعل بينهم.

4.منهج البحث وإجراءاته:

- **منهج البحث:** سوف تتبنى الدراسة منهج المحاكاة التجريبي المقارن لمناسبته لأهداف الدراسة وأسئلتها وذلك لكونه ملائمًا لطبيعة الدراسة وأهدافها؛ إذ يُعدّ هذا المنهج الأنسب لاختبار الفروق والتأثيرات بين مجموعات

مختلفة في بيئات مضبوطة، مما يسهم في تعميق الفهم حول كيفية تأثير العوامل المختلفة على النتائج.

- **بيانات الدراسة:** الدراسة الحالية دراسة محاكاة "Simulation Study" سوف تعتمد على بيانات مولدة من الحاسب الآلي. وبالتالي سيتم تجاور بناء الأداة إلى الحصول على البيانات، والعينة لأن الهدف من الدراسة سيتم تجاوز هاتين النقطتين إلى معالجات تتم بعد هاتين النقطتين وستتألف بيانات الدراسة من بيانات مولدة من حيث عدد الأفراد (٢٥٠، ٥٠٠، ١٠٠٠، ٥٠٠٠)، وكذلك من حيث عدد المفردات (١٠، ٢٠، ٣٠) وبالتالي سيكون في العينة ثلاثة اختبارات متنوعة في مفرداتها، وكل اختبار يجيب عنه أربع مجموعات من الأفراد.
- **التصميم:** تتألف الدراسة الحالية من ثلاث متغيرات مستقلة وهم حجم العينة وطول الفقرات وطريقة التقدير؛ بالتالي أصبح لدينا تصميم 2x3x4 وسوف يتم تناولهم حسب الترتيب التالي:
  - حجم العينة: سوف يتم دراستهما في ثلاث مستويات (٢٥٠، ٥٠٠، ١٠٠٠، ٥٠٠٠).
  - عدد الاسئلة: سوف يتم تناولها في ثلاث (١٠، ٢٠، ٣٠).
  - طريقة تقدير معالم الفقرات والأفراد: سيتم تقدير معالم المفردات والأفراد مره باستخدام طريقة الأرجحية القصوى (ML) ومره باستخدام طريقة بايز (BM).

إجراءات الدراسة:

١. **توليد البيانات:** توليد استجابات لاربعة احجام من العينات مكونه من (٥٠٠، ١٠٠٠، ٥٠٠٠، ١٠٠٠٠) فرد؛ تتبع قدراتهم التوزيع الطبيعي ($M = 0, SD = 1$) ثم يتم توليد معالم فقرات لاختبار مكون من ثلاث أطوال (١٠، ٢٠، ٣٠) فقرة تبعا لنموذج التقدير الجزئي (PCM)؛ باستخدام برنامج (WinGen v1.4).

٢. التحقق من مناسبة البيانات المولدة في كل ظرف من ظروف الدراسة لافتراضات نموذج التقدير الجزئي (PCM). لم يتم التحقق من الافتراضات الأساسية لنظرية الاستجابة للفقرة،

مثل أحادية البعد والاستقلال الموضعي، وذلك لان البيانات قد تم توليدها باستخدام المحاكاه تم التحم فيها بدقة لتحقق هذه الافتراضات، وبهذا فان خصائص البيانات المولدة مثل عدد الأبعاد وطبيعة العلاقة بين المتغيرات قد تم تثبيتها مسبقا ضمن النموذج المولد، ونذكر على سبيل المثال وليس الحصر (Reise, 1990; Kang & Chen, 2008; Ames & Samonate, 2015 ;Fu.,2010 ; Gorter et al., 2020) وبالتالي، فإن التحقق من هذه الفرضيات يصبح غير ضروري في هذه المرحلة من الدراسة، بخلاف الدراسات التطبيقية التي تعتمد على بيانات حقيقية قد لا تتحقق تلك الافتراضات.

٣. **تقدير المعالم:** تقدير معالم الأفراد (القدرة) ومعالم المفردات (الصعوبة) باستخدام برنامج (Excliber) وفقا لنموذج التقدير الجزئي (PCM).
٤. **الاجابة عن أسئلة الدراسة:** الإجابة عن أسئلة الدراسة بناء على ما تم التوصل اليه من نتائج.
٥. **صياغة النتائج:** استخراج النتائج ومناقشتها وصياغة التوصيات.

الأساليب الإحصائية:

للوصول الى نتائج الدراسة سوف تعتمد الدراسة على الخطوات التالية:

١. تقدير معالم فقرات الاختبار المكونة من (١٠، ٢٠، ٣٠) مفردة وتقدير استجابة (٢٥٠، ٥٠٠، ١٠٠٠،٥٠٠٠) فرد عليها تبعا لنموذج التقدير الجزئي (PCM) باستخدام برنامج (Excliber)؛ من خلال المتوسطات الحسابية والانحرافات المعيارية لكل ظرف من ظروف الدراسة (2X3X4).
٢. حساب الخطأ المعياري في دقة تقدير معالم الأفراد والفقرات تبعا لنوع النموذج المستخدم والمقارنة بينهما باستخدام تحليل التباين المختلط (Mix design ANOVA) باستخدام برنامج (SPSS)، لدلالة الفروق في دقة التقدير لكل ظرف من ظروف الدراسة والتفاعل بينهما، حيث المتغيرات المستقلة حجم العينة (٢٥٠، ٥٠٠، ١٠٠٠،٥٠٠٠)، وعدد المفردات (١٠، ٢٠، ٣٠) حيث يمثل كل منهما تصميم بين المجموعات حيث أنه في دراسات المحاكاة لكل

ظرف من ظروف الموقف البحثي ينتج فيه بيانات جديدة data set ، ومتغير طريقة التقدير تصميم داخل المجموعات حيث يتم التقدير بطريقتين لكل موقف بحثي.

٣- التعرف على أثر طريقة التقدير وطول الاختبار وحجم العينة على دقة تقدير معالم الأفراد باستخدام نموذج التقدير الجزئي.

وذلك من خلال صياغة عدد من التساؤلات والفروض، التي تم التحقق من صحتها بإجراء دراسة محاكاة تم من خلالها توليد بيانات ذات خصائص معروفة مسبقا تحقق أهداف الدراسة. حيث قدرت معلمة القدرة والصعوبة لنموذج التقدير الجزئي (PCM) باستخدام طريقتي التقدير الأرجحية القصوى (ML) والطريقة البايزية (BE)، وحُسبت المتوسطات الحسابية والانحرافات المعيارية والخطأ المعياري للتقدير(SEE)، وطُبّق تحليل التباين المختلط الثلاثي؛ للكشف عن وجود تأثيرات تفاعلية بين المُتغيّرات مع حساب حجم الأثر. كما أُجريت المقارنات البعدية بين المجموعات في دقة تقدير مَعْلَمَة القدرة والصعوبة لنموذج التقدير الجزئي (PCM)، باستخدام اختبار بينفروني؛ مما ساعد على تفسير النتائج بشكل دقيق.

وفيما يأتي عرض النتائج وفقًا لأسئلة الدراسة:

أولًا: النتائج المُتَعَلِّقَة بالإجابة عن السؤال الأول: " ما أثر طريقة التقدير وعدد فقرات الاختبار وحجم العينة على دقة تقدير معلم القدرة باستخدام نموذج التقدير الجزئي؟"

للإجابة على هذا السؤال قدرة القدرة باستخدام طريقة الارجحية القصوى (ML)، والطريقة البايزية (BE)، ويبين الجدول التالي المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري (SEE)؛ لتقدير القدرة وفقًا لطريقة التقدير باختلاف حجم العينة وطول الفقرات.

جدول (١) المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري لتقدير القدرة وفقًا لطريقة التقدير، باختلاف طول الاختبار وحجم العينة.

الإحصاءات الوصفية

			المتوسط	الانحراف المعياري
دقة تقدير الارجحية القصوى	**250**	**10**	0.3630776	0.10857692
		20	0.2654008	0.06976687

		30		0.04420131
	500	10		0.13056327
		20		0.06831039
		30		0.05195465
	1000	10		0.09871377
		20		0.06257044
		30		0.04935941
	5000	10		0.10137258
		20		0.06066705
		30		0.06003716
دقة تقدير البايزية	250	10		0.06303385
		20		0.04221153
		30		0.03036319
	500	10		0.05621176
		20		0.04380556
		30		0.03509957
	1000	10		0.04797664
		20		0.03864129
		30		0.03471022
	5000	10		0.04487982
		20		0.03550947
		30		0.03823834

يتضح من الجدول (١)، أنه بشكل عام تظهر طريقة تقدير القدرة باستخدام الارجحية القصوى (ML)، تباينا أعلى في دقة تقدير القدرة، مقارنة بالطريقة البايزية في تقدير القدرة عبر جميع أحجام العينات وطول الفقرات، وبالنسبة لطريقة الأرجحية في التقدير (ML)، يكون دقة التقدير (SEE) أكبر عند زيادة عدد الفقرات ماعد عند عدد فقرات (١٠)، وبالنسبة لحجم العينة (٢٥٠) و (٥٠٠) فيظهر اختلاف في دقة تقدير معالم القدرة لصالح حجم عينة (٢٥٠) وعند حجم عينة (٢٥٠) تظهر طريقة الأرجحية القصوى (ML) متوسطات وانحرافات معيارية لقيم الخطأ المعياري أعلى نسبيا.

وفي المقابل تعطي الطريقة البايزية (BE) في التقدير متوسطات وانحرافات معيارية أقل لقيم الخطأ

المعياري في تقدير القدرة عبر جميع أحجام العينة وطول الفقرات مقارنة بطريقة التقدير باستخدام الارجحية القصوى (ML)،

وبالنسبة لطريقة التقدير البايزية (BE)، يكون الخطأ المعياري في التقدير (SEE) أدق عند زيادة عدد الفقرات لكل حجم عينة، ولكن يكون الخطأ المعياري في التقدير (SEE)، غير منتظم فيكون عند كل طول فقرة عبر أحجام العينات غير مستقر فيكون تارة لصالح حجم عينة على لأخرى ويظهر ذلك واضح عند حجم عينة (٥٠٠٠) ولكل طول الفقرات.

وفي ضوء ما تقدّم؛ أُجري تحليل التباين المختلط الثلاثي لاختبار الفروق والتفاعلات بين طرق التقدير، وأحجام العينة، وطول الفقرات في دقة تقديرات القدرة متبوعًا بحساب حجم الأثر؛ حيث كُشف عن الفروق بين متوسطات الخطأ المعياري للتقدير باختلاف طريقة تقدير معلمة القدرة (الارجحية القصوى – ML والطريقة البايزية BE)، وحجم العينة (٢٥٠ – ٥٠٠ – ١٠٠٠ – ٥٠٠٠)، وطول الفقرات (١٠ – ٢٠ – ٣٠). حيث تم التحقق من الافتراضات الأساسية لاختبار تحليل التباين المختلط، والتي تمثلت في افتراض اعتدالية توزيع البيانات لكل طريقة تقدير وافتراض تجانس تباين الفرق بين مستويات العامل داخل المجموعات وافتراض تجانس التغاير، ونظرا لأن بيانات الدراسة قد تم توليدها من مجتمع طبيعي بمتوسط صفر وانحراف معياري واحد فإن شرط الاعتدالية يعتبر متحققا، وتشير نظرية النهاية المركزية إذا كانت حجم العينة كبير فإن البيانات تتبع التوزيع الطبيعي، وبالنسبة لافتراض تجانس التغاير يعتبر طريقة التقدير (BE – ML) متغير داخل المجموعات يتكون من مستويين فقط ، فإن شرط تجانس التغاير يعد متحققا تلقائيا مما يدعم إمكانية تطبيق تحليل التباين المختلط بثقة في النتائج (Filed,2018).

ويوضّح الجدول (٢) نتائج تحليل التباين المختلط الثلاثي الاتجاه لدقة تقديرات القدرة باختلاف طريقة التقدير، وحجم العينة، وطول الفقرات.

نتائج تحليل التباين المختلط الثلاثي الاتجاه لدقة تقديرات القدرة باختلاف طريقة التقدير، وحجم العينة، وتوزيع البيانات

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	قيمة ف (F)	الدلالة الإحصائية (P)	مربع ايتا الجزئي (η^2)

طرق التقدير	1.663	1	1.663	1779.345	0.000	0.081
الحجم	0.055	3	0.018	2.739	0.042	0.000
طول الفقرات	47.707	2	23.854	3568.972	0.000	0.261
طرق التقدير × الحجم	0.008	3	0.003	2.866	0.035	0.000
طرق التقدير× طول الفقرات	0.648	2	0.324	346.925	0.000	0.033
الحجم× طول الفقرات	0.213	6	0.35	5.329	0.000	0.001
طريقة التقدير× الحجم× طول الفقرات	0.018	6	0.003	3.184	0.004	0.001

ويتضح من الجدول (٢) نتائج تحليل التباين المختلط ثلاثي الاتجاهات، وتأثيرات ثلاثة مُتغيّرات – طرق التقدير، وحجم العينة، وطول الفقرات، إلى جانب تفاعلاتها في المُتغيّر التابع (دقة تقدير مَعْلَمَة القدرة).

فبالنسبة لطرق التقدير؛ أظهرت النتائج وجود فروق ذات دلالة إحصائيًّة في متوسطات دقة تقدير مَعْلَمَة القدرة عند مستوى ($\alpha \leq .05$)، تبعًا لطرق التقدير (الارجحية العظمى ML، والطريقة البايزية BE)، حيث بلغت قيمة (ف) والدلالة الإحصائية (= F (1, 20238) 1779.345 , $p < 0.000$). كما أشارت قيمة مربع إيتا الجزئي (η^2) إلى طرق التقدير البالغة (0.081) إلى أن التأثير ذو أثر متوسط حسب المحكات المحددة في (الضجعي، ٢٠٠٥). ويُفسّر (8.1٪) من التباين في دقة تقدير القدرة.

وبالنسبة لحجم العينة؛ فتوجد فروق ذات دلالة إحصائية في متوسطات دقة تقدير مَعْلَمَة القدرة عند مستوى ($\alpha \leq .05$) تبعًا لحجم العينة (٢٥٠-٥٠٠- ١٠٠٠-٥٠٠٠)، حيث بلغت قيمة (ف)، والدلالة الإحصائية ($F (1, 20238) = 2.739, p < 0.042$). وبلغت قيمة مربع إيتا الجزئي (η^2) لتأثير حجم العينة (0.000)؛ مما يُشير إلى تأثير ذو أثر صغير حسب المحكات المحددة في (الضجعي، ٢٠٠٥). حيث يُفسّر (0.000٪) من التباين في دقة تقدير القدرة.

كما أُجريت المقارنات البعدية بين المجموعات في دقة تقدير مَعْلَمَة القدرة باستخدام اختبار شيفيه بين المتوسطات الحسابية للأخطاء المعيارية لتقدير مَعْلَمَة القدرة، تبعًا لحجم العينة (٢٥٠-٥٠٠-

١٠٠٠–٥٠٠٠) ؛ بهدف تحديد اتجاه الفروق، والجدول (٣) يوضّح ذلك

جدول (٣)

نتائج اختبار شيفيه للمقارنة البعدية بين المجموعات في دقة تقدير مَعْلَمَة القدرة، تبعًا لحجم العينة

حجم العينة		الفرق بين المتوسطات	مستوى الدلالة
250	500	0.0001381	1.000
	1000	0.0044050	0.323
	5000	0.0035096	0.452
500	250	-0.0001381	1.000
	1000	0.0042669	0.142
	5000	0.0033716	0.200
1000	250	-0.0044050	0.323
	500	-0.0042669	0.142
	5000	-0.0008954	0.896
5000	250	-0.0035096	0.452
	500	-0.0033716	0.200
	1000	0.0008954	0.896

يلاحظ من النتائج المعروضة في جدول (٣) عدم وجود فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq .05$)

عند جميع أحجام العينات (٢٥٠ – ٥٠٠ – ١٠٠٠ – ٥٠٠٠) وان كانت تظهر فروق بين المتوسطات حيث يلاحظ أن الفرق بين المتوسطات الحسابية للأخطاء المعيارية بين أحجام العينات يقترب من الصفر، وقد أدى هذا إلى عدم وجود فروق دالة إحصائية في دقة تقدير معلمة القدرة عند مستوى ($\alpha \leq .05$).

أما بالنسبة لطول الفقرات؛ فتوجد فروق ذات دلالة إحصائية في متوسطات دقة تقدير مَعْلَمَة القدرة عند مستوى ($\alpha \leq .05$)، تبعًا لعدد الفقرات (١٠ – ٢٠ – ٣٠)، حيث بلغت قيمة (ف) والدلالة الإحصائية ($F\ (1, 20238) = 3568.972, p < 0.000$) ؛ مما يُشير إلى أن المستويات المختلفة لعدد الفقرات لها تأثير دال إحصائيًّا في دقة تقدير القدرة. كما تُشير قيمة مربع إيتا الجزئي

(η^2) البالغة (0.261) إلى أنه يمكن إرجاع حوالي (26٪) من التباين الإجمالي في دقة تقدير القدرة إلى تأثير عدد الفقرات، الذي يُعدّ حجم تأثير كبير حسب المحكات المحددة في (الضجعي، ٢٠٠٥).

كما أُجريت المقارنات البعدية بين المجموعات في دقة تقدير مَعْلَمَة القدرة باستخدام اختبار شيفيه بين المتوسطات الحسابية للأخطاء المعيارية لتقدير مَعْلَمَة القدرة، تبعًا لطول الفقرات (١٠ – ٢٠ ٣٠)؛ بهدف تحديد اتجاه الفروق، والجدول (٤) يوضّح ذلك

جدول (٤)

نتائج اختبار شيفيه للمقارنة البعدية بين المجموعات في دقة تقدير مَعْلَمَة القدرة، تبعًا لطول الفقرات

عدد الفقرات (I)		الفرق بين المتوسطات (I-J)	مستوى الدلالة
10	20	.0952024*	0.000
	30	.1412518*	0.000
20	10	-.0952024-*	0.000
	30	.0460494*	0.000
30	10	-.1412518-*	0.000
	20	-.0460494-*	0.000

ويوضّح الجدول (٤) وجود فروق دالة إحصائيًّا في دقة تقدير مَعْلَمَة القدرة عند مستوى ($\alpha \leq .05$)، بين أطول الفقرات في دقة تقدير القدرة؛ حيث كان الفرق بين المتوسطات عند طوال فقرة (١٠) و طول فقرة (٢٠) يساوي (.0952024) وبذلك يعني أن طول الفقرة (٢٠) تعطى أقل قيمة لتقدير الخطأ المعياري في التقدير حيث الرقم الأقل هو الأدق في تقدير القدرة، والفرق بين طول الفقرة (٢٠) و طول الفقرة (٣٠) يساوي (.0460494) وبذلك يعني أن طول الفقرة (٣٠) حيث الرقم الأقل هو الأدق في تقدير القدرة ، ، وحيث الفرق بين طول الفقرة (١٠) و طول الفقرة (٣٠) يساوي (.1412518) وبذلك يعني أن طول الفقرة (٣٠) هي الأكثر دقة في تقدير القدرة حيث الرقم الأقل هو الأدق في تقدير القدرة ، ومما سبق؛ يمكن استنتاج ان طول الفقرة عامل مهم في دقة تقدير القدرة حيث كانت الفقرات الأعلى هي الأكثر دقة في دقة تقدير معلمة القدرة.

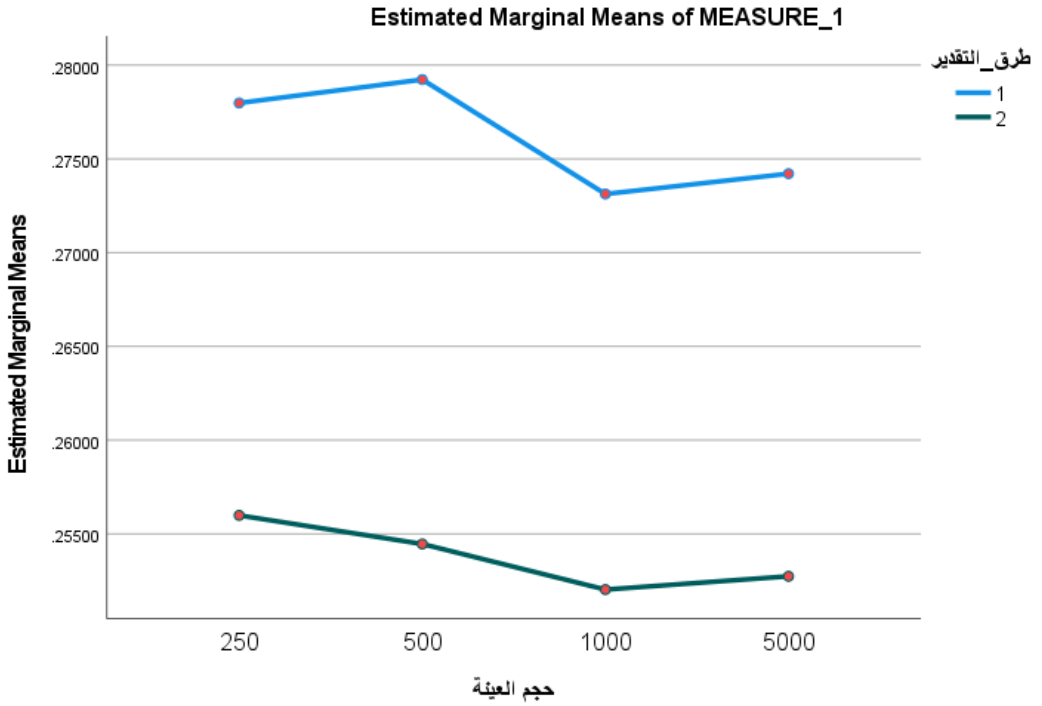
ثانيا: النتائج المُتَعَلِّقَة بالإجابة عن السؤال الثاني: " ما أثر طريقة التقدير وعدد فقرات الاختبار وحجم العينة على دقة تقدير معلم القدرة باستخدام نموذج التقدير الجزئي والتفاعل بينهم؟"

وبالنسبة للتفاعلات؛ فقد أظهرت النتائج وجود فروق ذات دلالة إحصائيَّة في دقة تقدير مَعْلَمَة القدرة عند مستوى دلالة ($\alpha \leq .05$) للتفاعل بين طريقة تقدير معلمة القدرة (الارجحية الصوى ML و الطريقة البايزية BE)، وحجم العينة (٢٥٠-٥٠٠-١٠٠٠- ٥٠٠٠)؛ إذ بلغت قيمة (ف) لهذا التفاعل والدلالة الإحصائية ($F(3, 20238) = 2.866$, $P = 0.035$). وتُشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.035) إلى أنه يُفسّر (3.5٪) من التباين في دقة تقدير مَعْلَمَة القدرة؛ مما يُشير إلى أن التأثير المشترك لهذه المُتغيّرات على دقة التقدير ذو أثر صغير حسب المحكات المحددة في (الضجعي، ٢٠٠٥). ويُلاحظ أن حجم العينة كان له تأثير دال إحصائيًّا في دقة تقدير معلمة القدرة بشكل مستقل، كما كان لتفاعله مع حجم العينة تأثير دال إحصائيًّا في دقة تقدير معلمة القدرة، ولكن حجم التأثير كان ضعيف بشكل مستقل وعند التفاعل مع حجم العينة أيضا.

ويوضّح الشكل (٣) التمثيل البياني لتأثير التفاعل بين طريقة تقدير القدرة وحجم العينة على الخطأ المعياري لتقديرات القدرة.

شكل ٣

التمثيل البياني لتأثير التفاعل بين طريقة تقدير القدرة وحجم العينة على الخطأ المعياري لتقديرات القدرة.

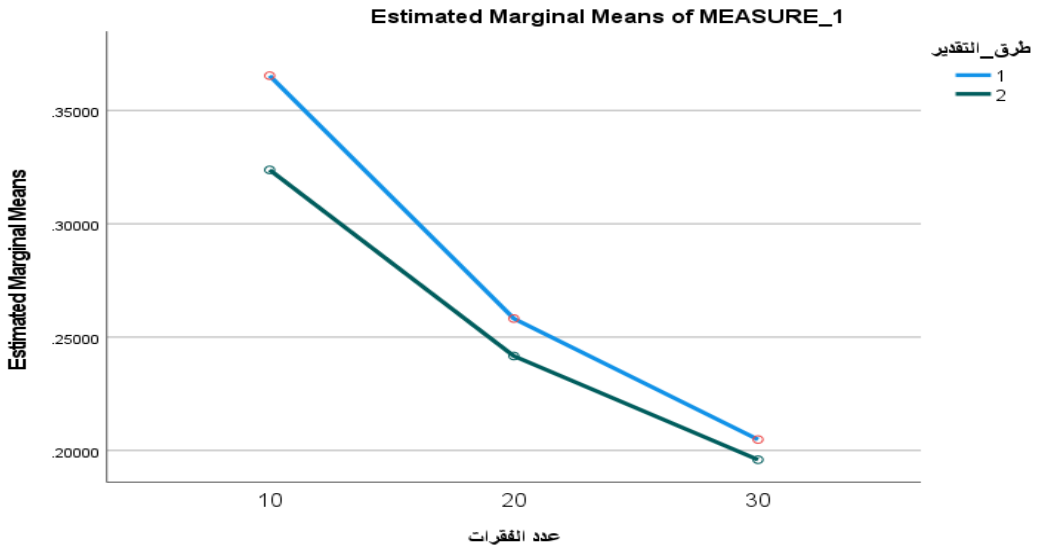


كما توجد فروق ذات دلالة إحصائيَّة في دقة تقدير مَعْلَمَة القدرة عند مستوى (α ≤ .05) للتفاعل بين طريقة تقدير معلمة القدرة (الارجحية الصوى ML والطريقة البايزية BE)، وطول الفقرات (١٠ – ٢٠ – ٣٠)، حيث بلغت قيمة (ف) لهذا التفاعل والدلالة الإحصائية F (2,20238)) = 346.925, P = 0.000). وتُشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.033) إلى أنه يُفسّر (٣,٣٪) من التباين في دقة تقدير مَعْلَمَة القدرة؛ مما يُشير إلى أن التأثير المشترك لهذه المُتغيّرات على دقة التقدير ذو إثر صغير حسب المحكات المحددة في (الضجعي، ٢٠٠٥). ونلاحظ أن طول الفقرات كعامل مستقل كان حجم التأثير كبير على دقة تقدير القدرة وعند تفاعل طول الفقرات مع طرق التقدير أصبح حجم التأثير ذو إثر ضعيف.

ويوضّح الشكل (٤) التمثيل البياني لتأثير التفاعل بين طريقة تقدير القدرة وطول الفقرات على الخطأ المعياري لتقديرات القدرة.

شكل ٤

التمثيل البياني لتأثير التفاعل بين طريقة تقدير القدرة وطول الفقرات على الخطأ المعياري لتقديرات القدرة.



كما توجد فروق ذات دلالة إحصائيًّة في دقة للتفاعل بين حجم العينة (٢٥٠ – ٥٠٠ – ١٠٠٠ – ٥٠٠٠) وطول الفقرات (١٠ – ٢٠ – ٣٠) عند مستوى ($\alpha \leq .05$) ، حيث بلغت قيمة (ف) لهذا التفاعل والدلالة الإحصائية(F (6,20238) = 5.329, P = 0.000). وتُشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.001) إلى أنه يُفسّر(٠,٠٠١٪) من التباين في دقة تقدير مَعْلَمَة القدرة؛ مما يُشير إلى تاثير ذو أثر صغير حسب المحكات المحددة في (الضجعي، ٢٠٠٥).

وأخيرًا، يلاحَظ وجود فروق ذات دلالة إحصائيًّة في دقة تقدير مَعْلَمَة القدرة عند مستوى ($\alpha \leq .05$) للتفاعل الثلاثي، بين طرق تقدير معلمة القدرة (الارجحية الصوى ML والطريقة البايزية BE)، و حجم العينة (٢٥٠ – ٥٠٠ – ١٠٠٠ – ٥٠٠٠) وطول الفقرات (١٠ – ٢٠ – ٣٠)، حيث بلغت قيمة (ف) لهذا التفاعل الثلاثي والدلالة الإحصائية F (6,20238) = 3.184, P = (0.004). وتُشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل الثلاثي، (0.001)؛ إلى أنه يُفسّر(0.001٪) من التباين في دقة تقدير مَعْلَمَة القدرة؛ مما يُشير إلى أن التأثير المشترك لهذه المُتَغَيِّرَات ذو اثر ضعيف على دقّة التقدير ، حسب المحكات المحددة في (الضجعي، ٢٠٠٥).

وبالنسبة للتفاعل الثلاثي ، فقد أظهرت النتائج وجود فروق ذات دلالة إحصائية في دقة تقدير معلمة القدرة عند مستوى دلالة ($\alpha \leq .05$) للتفاعل بين طريقة تقدير معلمة القدرة (الارجحية الصوى ML

و الطريقة البايزية BE)، وحجم العينة (٢٥٠-٥٠٠-١٠٠٠- ٥٠٠٠)؛ وعدد فقرات الاختبار (١٠ ، ٢٠ ، ٣٠) فقرة، إذ بلغت قيمة (ف) لهذا التفاعل والدلالة الإحصائية (3.184 , P = 0.004 F (6, 20238) =) وتُشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.01) إلى أنه يُفسّر (1٪) من التباين في دقة تقدير مَعْلَمَة القدرة؛ مما يُشير إلى أن التأثير المشترك لهذه المُتغيّرات على دقة التقدير ذو أثر صغير حسب المحكات المحددة في (الضجعي، ٢٠٠٥).

نتائج الدراسة:

١) أظهرت النتائج تفوق طريقة التقدير البايزية (BE)، على طريقة الارجحية العظمى (ML)، لمتوسطات الأخطاء المعيارية في دقة تقدير القدرة لنموذج التقدير الجزئي (PCM)، بينما لم يظهر لحجم العينة تأثير ذو دلالة عملية فقد فسر (0.00 %) من التباين في دقة تقدير القدرة لنموذج التقدير الجزئي (PCM)، يمكن تفسير ذلك بأن تأثير طول الفقرات وطريقة التقدير كان أقوى لدرجة غطّى على أثر حجم العينة، أو أن النماذج الحديثة كـ BE أقل تأثرًا بحجم العينة. وأظهرت أن عدد الفقرات تأثيرا ذو دلالة إحصائية حيث كانت الدلالة العملية (26%) من التباين المفسر في دقة تقدير القدرة لنموذج التقدير الجزئي (PCM) مما يعني أن عدد الفقرات يعد من العوامل الحاسمة في دقة التقدير.

٢) أظهرت نتائج تحليل التباين المختلط الثنائي أن التفاعل بين طريقة التقدير وحجم العينة ليس له أثر ذو دلالة عملية فقد فسر (0.00) من التباين المفسر في دقة تقدير القدرة ، إلا أن التفاعل بين طريقة التقدير وطول الفقرات له أثر ذو دلالة إحصائية وكان حجم الأثر متوسط وهذا يدل على أن تأثير عدد الفقرات في تحسين الدقة قد يختلف بشكل طفيف بناء على طريقة التقدير وتؤكد هذه النتائج أهمية النظر إلى كل من التأثيرات الرئيسية والتفاعلات في تحليل دقة تقدير المعلمات حيث يوفر عدد الفقرات تأثيرا واضحا وعالي القوة، في حين أن طريقة التقدير تلعب دورا صغيرا في هذا التأثير، وكذلك التفاعل بين حجم العينة وطول الفقرات فقد كان حجم الأثر ضعيف، وبالنسبة للتفاعل الثلاثي بين طريقة التقدير وحجم العينة وطول الفقرات كان حجم الأثر ضعيف.

التوصيات:

من خلال النتائج التي تم التوصل إليها في هذه الدراسة يمكن التوصية بالتالي:

١) حيث نتائج الدراسة أن طريقة تقدير القدرة عامل مهم في دقة تقدير قدرة الأفراد، لذا توصي الدراسة باستخدام طريقة التقدير البايزية (BE)، في تقدير معلمة القدرة للأفراد حيث كانت ادق في دقة تقدير القدرة عند جميع الفقرات وعلى طول أحجام العينات وان كانت عند حجم العينة الكبيرة متقاربه إلا نها ادق.

٢) توضح النتائج أن طول الفقرات هو العامل الأقوى تأثيرًا على دقة التقدير، يليه طريقة التقدير، في حين أن حجم العينة لم يكن له أثر عملي مهم. كما أظهرت النتائج أن التفاعل بين طول الفقرات وطريقة التقدير له دلالة عملية، في حين أن التفاعلات الأخرى أقل أهمية. بناءً على ذلك، توصي الدراسة بـ التركيز على تحسين تصميم الاختبارات من خلال زيادة عدد الفقرات واختيار طريقة تقدير مناسبة، مع التقليل من الاعتماد المفرط على حجم العينة في هذا النوع من النماذج..

٣) حيث أظهرت الدراسة أن حجم عينة ٢٥٠ مناسب للوصول إلى نتائج دقيقة وموثوقة لذا توصي الدراسة أن هذا الحجم كافي ومناسب في ظل محددات الدراسة.

الدارسات المستقبلية:

١) دراسة أثر طرق تقدير أخرى على دقة تقدير المعالم باستخدام نماذج عديدة الاستجابة

٢) إجراء المزيد من الدارسات للبحث عن العوامل والمتغيرات الأخرى التي تؤثر على دقة تقدير المعالم باستخدام نماذج عديدة الاستجابة

- **قائمة المراجع**

المراجع العربية:

بني عطا، زايد صالح إبراهيم .(2017) .تقصي أثر طول الاختبار وحجم العينة على دقة طرق تقدير معالم الفقرات وقدرات الأفراد في برنامج بايلوج .***المجلة الدولية للبحث في التربية وعلم النفس***.606–579 , (2)*5* ,

دي أيالا، آر .(2017) .***النظرية والتطبيق في نظرية الاستجابة للفقرة*** .دار جامعة الملك سعود.

التقي، أحمد محمد (2009). ***النظرية الحديثة في القياس***. دار المسيرة للنشر والتوزيع.

ضعضع، هبة عبداللطيف، طومان، منار أحمد، وطيفور، مصطفى أحمد. (2020). أثر حجم العينة وطرائق التقدير في دقة تقدير معالم نموذج راش. ***جرش للبحوث والدراسات***، مج21، ع1، 170–131.

الجضعي، خالد سعد (٢٠٠٥). ***تقنيات صنع القرار، تطبيقات حاسوبية***. دار الأصحاب للنشر والتوزيع: الرياض

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

حبيب، صفاء وعزيز، صادق (2018)، ***بناء المقاييس النفسية وفقا لنظرية الاستجابة للفقرة باستعمال الانموذج الكشفي المعمم***. الدار المنهجية.

بيكر، فرانك. (2010). ***أسس نظرية الاستجابة للمفردة*** (عبدالرحمن الطريري ؛ السيد هاشم وسوسن شلبي ، مترجم). جامعة الملك سعود. (العمل الأصلي نشر في 2001).

الثوايبة، أحمد محمود. (٢٠١٠). أثر حجم العينة على تقدير صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة، ***مجلة دمشق***.

القرني، أحمد سعيد. (٢٠٢٢). أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة .***المجلة العربية للعلوم التربوية والنفسية***، ٦(30)، ٢-٢٢.

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

الطراونة، عيسى عبد الوهاب سليم .(2011) ***مقارنة فاعلية طريقة الأرجحية القصوى وطريقة بييز في تقدير مَعْلَم القدرة عند استخدام النموذج اللوغاريتمي*** *(دكتوراه). جامعة القاهرة. معهد الدراسات والبحوث التربوية، مصر .*

المراجع الأجنبية:

Ackerman, W., & Murky, E. (1997). Item information and discrimination functions for ternary PCM items. *Psychometrika, 62*(4), 569-578. https://doi.org/10.1007/BF02295309

Bellhouse, D. R. (2004). The Reverend Tomes Bayes, FRS: A biography to celebrate the tercentenary of his birth. *Statistical Science, 19*(1), 3-43. https://doi.org/10.1214/088342304000000017

Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of polytomous IRT models with rating scale data: An investigation over sample size, instrument length, and missing data. *Frontiers in Education, 6*, 1-15. https://doi.org/10.3389/feduc.2021.656934

De Ayala, R. J. (2022). *The theory and application of item response theory* (2nd ed.). Guilford Press.

Dewanti, S. S., Hadi, S., & Numan, M. (2021). The application of item response theory in analysis of characteristics of mathematical literacy test items. *Ilkogretim Online, 20*(1), 1226-1237. https://doi.org/10.17051/ilkonline.2021.01.105

Fu, Q. (2010). *Comparing accuracy of parameter estimation using IRT model in the presence of guessing* (Unpublished PhD dissertation). Illinois University, USA.

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. https://doi.org/10.1207/s15324818ame1804_2

Hambleton, R. K., & Cook, L. L. (1983). Robustness of item response models and effects of test length and sample size in the precision of ability estimates. *New Horizons in Testing*, 31-49.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer-Nijhoff Publishing.

Kurisu, K., Hashimoto, M., Ishizawa, T., Shibayama, O., Inada, S., Fujisawa, D., & Yoshiuchi, K. (2022). Development of computer adaptive testing for measuring depression in patients with cancer. *Scientific Reports, 12*(1), 1-7.

https://doi.org/10.1038/s41598-022-04818-7

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Rijmen, F. (2009). Efficient full information maximum likelihood estimation for multidimensional IRT models. *ETS Research Report Series, 2009*(1), i-31. https://doi.org/10.1002/j.2333-8504.2009.tb02177.x

Şahin, A., & Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Kuram ve Uygulamada Egitim Bilimleri, 17*(1), 321–335. https://doi.org/10.12738/estp.2017.1.0270

Stocking, M. (1990). Specifying optimum examinees for item parameter estimation in item response theory. *Psychometrika, 55*(3), 559-573. https://doi.org/10.1007/BF02294712

Sun, S. S., Tao, J., Chang, H. H., & Shi, N. Z. (2012). Weighted maximum-a-posterior estimation in tests composed of dichotomous and polytomous items. *Applied Psychological Measurement, 36*(4), 271-290. https://doi.org/10.1177/0146621612438795

Swaminathan, H., & Gifford, J. A. (1982). Bayesian in the Rasch model. *Journal of Educational Statistics, 7*(3), 175-191. https://doi.org/10.3102/10769986007003175

Thompson, N. (2009). *Ability estimation with item response theory*. St Paul Assessment System Corporation. https://assess.com/

Wang, S., & Wang, T. (2001). Precision of Warm weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement, 25*(4), 317-331. https://doi.org/10.1177/01466210122032091

A **mes, A. J., & Samonte, K.** (2015). Using SAS PROC MCMC for item response theory models. *Educational and Psychological Measurement, 75*(4), 585–609. https://doi.org/10.1177/0013164414554610

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications Limited.

Fu, Q. (2010). *Comparing accuracy of parameter estimation using IRT models in the presence of guessing* [Doctoral dissertation, University of Illinois at Chicago]. ProQuest Dissertations Publishing.

Gorter, R., Fox, J. P., Riet, G. T., Heymans, M. W., & Twisk, J. W. R. (2020). Latent growth modeling of IRT versus CTT measured longitudinal latent variables. *Statistical Methods in Medical Research, 29*(4), 962–986. https://doi.org/10.1177/0962280219849175

Kang, T., & Chen, T. T. (2008). Performance of the Generalized S-X^2 item fit index for polytomous IRT models. *Journal of Educational Measurement, 45*(4), 391–406. https://doi.org/10.1111/j.1745-3984.2008.00072.x

Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model–data fit in IRT. *Applied Psychological Measurement, 14*(2), 127–137. https://doi.org/10.1177/014662169001400203