مجلة جامعة الأقصى- سلسة العلوم التربوية والنفسية، المجلد 6، العدد 1 يناير2023، ص:202-224
ISSN 2791 - 3783

ما وراء التحليل لفاعلية طريقتي بييز والأرجحية العظمى في تقدير القدرة

أ. صدام محمد أبو رواق
وزارة التربية والتعليم الأردنية

د. أحمد سليمان عوده
جامعة اليرموك

(تاريخ الاستلام 2022/08/16، تاريخ القبول 2022/09/13)

Self-efficacy and their relationship to positive thinking For high school students

Mr. Saddam Muhammad Abu Rawaq
Jordanian Ministry of Education

Dr. Ahmed Suleiman Odeh
Yarmouk University

(Received 16/08/2022, Accepted 13/09/2022)

أ. صدام أبو رواق – E-mail address: sdamabwrwaq11@gmail.com

الملخص:

هدفت الدراسة الحالية إلى الكشف عن حجم الأثر الكلي واختلاف الفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022). ولتحقيق ذلك اتبعت الدراسة المنهج الوصفي التحليلي، إذ تكون مجتمع الدراسة من جميع الدراسات والأبحاث التي تم إجراؤها في الفترة (2000 – 2022) والتي تناولت طريقتي التقدير (بييز والأرجحية العظمى) المستخدمتان في تقدير قدرة الأفراد في نظرية الاستجابة للفقرة، وخصوصا الدراسات التي قارنت بين الطريقتين من خلال مؤشرات دقة التقدير، كما تم الاستعانة بنموذج لترميز البيانات والمتمثلة في نتائج الدراسات في ضوء متغيرات الدراسة، وقد خلصت الدراسة إلى أن حجم الأثر الكلي بالاعتماد على أحجام الأثر وأوزان كل من الدراسات المضمنة في عينة الدراسة قد بلغت قيمته (0.187) بخطأ معياري (0.0348)، كما توصلت الدراسة إلى عدم وجود اختلافات في معدل حجوم الأثر للفروق في دقة التقدير العائدة لطريقة التقدير (بييز والأرجحية العظمى) للدراسات باختلاف كل من حجم العينة وطول الاختبار ونوع البيانات المستخدمة. في حين اختلف معدل حجم الأثر باختلاف النموذج اللوجستي المستخدم. كما أوصت الدراسة باستخدام طريقة الأرجحية العظمى لتقدير القدرة في جميع الظروف باستثناء التقدير وفق النموذج الأحادي للاختبارات القصيرة جدًا.

الكلمات المفتاحية: ما وراء التحليل، بييز، الأرجحية العظمى، دقة التقدير، القدرة.

Abstract:

The current study aimed to reveal the size of the total effect and the difference in the differences in the accuracy of estimating the ability in the light of the two estimation methods (Bayes and the MLE: Maximum likelihood). For the studies conducted in the period from (2000-2022). In addition, to achieve this, the study followed the descriptive analytical approach. As the study population consisted of all Studies and research conducted in the period (2000-2022). That dealt with the two estimation methods (Bayes and Maximum Likelihood) used in estimating the ability of individuals in the Item response theory. Especially studies that compared the two methods through indicators of estimation accuracy. In addition, a data-coding model was used, which is represented in the results of the studies in the light of the study variables. The results of the study concluded that the total effect size depending on the effect sizes and the weights of each of the studies included in the study sample had a value of (0.187) with a standard error of (0.0348). In addition, the study concluded that there are no differences in the average volumes the effect of the differences in the accuracy of the estimate due to the estimation method (bayes and maximum likelihood). For the studies according to the difference in the sample size, length of the test and the type of data used. While the average effect size differed according to the logistic model. The study also recommended the use of the maximum likelihood method for estimating ability in all circumstances except for estimation according to the univariate model for very short tests.

Keywords: meta-analysis, Bayes, maximum likelihood, estimation accuracy, ability.

مقدمة:

تعد نظرية الاستجابة للفقرة (Item Response Theory) من التطورات الحديثة في مجال القياس النفسي والتربوي، لما قدمته من طرق سيكومترية ذات فاعلية كبيرة في بناء المقاييس النفسية والتربوية، وطريقة تفسير هذه الدرجات على هذه المقاييس مقارنة بالنظرية التقليدية في القياس، حيث عملت على تقدير معالم الفقرات بمعزل عن قدرات الأفراد وقدرت قدرات الأفراد بمعزل عن فقرات الاختبار من خلال العديد من الطرق الإحصائية.

إذ تعتبر دقة تقدير معالم الفقرة وقدرة الأفراد من أهم القضايا في نظرية الاستجابة للفقرة، حيث تعتمد هذه الدقة على مجموعة من العوامل التي أهتم البحث السيكومتري في دراستها، والتي تتمحور حول تحقيق افتراضات نظرية الاستجابة للفقرة واختيار النموذج اللوجستي المناسب والاهتمام بأسلوب التقدير (Rakkapo, Prasitpong & Arayathanitkul, 2016).

يتم تقدير معالم الفقرات وقدرات الأفراد من خلال أسلوبين، إما بشكل مشترك (joint)؛ أي بتقدير معالم الفقرات والأفراد معًا، بالاعتماد على تقديرات الأرجحية العظمى (Maximum Likelihood)، أو من خلال نظرية بييز (Bayes Theory) التي تأخذ بعين الاعتبار التوزيع القبلي للعينة، ومن ثم تعظيم اقتران التوزيع البعدي لإيجاد تقديرات بييز لمعالم الفقرات (Lord, 1986).

إلا أن كلا الأسلوبين لا يخلوان من العيوب، وتظهر بينهما الكثير من المفارقات ففي حين يصلح أسلوب الأرجحية العظمى مع العينات ذات الحجوم الكبيرة أو العينات المتجانسة، نجد أن أسلوب بييز أكثر دقة في حالة العينات الصغيرة والاختبارات القصيرة، وفي حال كان حجم العينة كبيرًا والاختبار طويلًا فإن كلا الأسلوبين يقودان لنفس النتائج.

مشكلة الدراسة:

مع البحث المستمر في نظرية الاستجابة للفقرة والحرص على الحصول على تقديرات دقيقة للمعالم بشكل عام، وعلى اختبارات تتسم بالثبات والدقة والموثوقية تم إجراء مقارنات ومفاضلات متعددة بين طريقتي التقدير (بييز والارجحية العظمى)، حيث أن استخدام الطريقة الأقل دقة في التقدير ينتج اختبارات غير موثوقة وغير دقيقة ويقود إلى اتخاذ قرارات غير صحيحة تربويًا ونفسيًا، ظهر العديد من الدراسات التي تبحث في هذا الموضوع، وقد تبين وجود تناقضات في نتائج بعضها واتفاق في بعضها الآخر في ضوء ظروف مختلفة كحجم العينة وطول الاختبار (الشريفين، 2012).

وبناءً على ما تم طرحه في الأدب النظري والذي بين وجود تباين في نتائج الدراسات التي بحثت في كل من طريقتي بييز والأرجحية العظمى لتقدير قدرة الأفراد تولد لدى الباحث الشعور بالحاجة إلى إجراء دراسة ذات طابع تكاملي تعمل على التوفيق بين نتائج هذه الدراسات والخروج بمفاضلة نهائية وتقدير قيمة عامة لحجم أثر الفروق في دقة تقدير القدرة باستخدام الطريقتين، إذ أن الحصول على مثل هذه النتائج لا يستدعي إجراء دراسات جديدة تحت ظروف مختلفة بقدر الحاجة إلى أسلوب احصائي يسهم في دراسة نتائج هذه الدراسات واستخلاص المعنى الحقيقي منها، لذا جاءت هذه الدراسة للإجابة عن الأسئلة الآتية:

أسئلة الدراسة:

1. ما حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022)؟

2. هل يختلف حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022) والتي تعزى لكل من طول الاختبار، حجم العينة، النموذج اللوجستي ونوع البيانات؟

أهداف الدراسة:

إن الهدف الرئيس من هذه الدراسة الخروج بتعميم حول أفضلية طرق التقدير باختلاف الظروف التجريبية في تقدير قدرات الأفراد، من خلال العمل على تجميع أكبر قدر ممكن من نتائج الدراسات التي عملت على المقارنة بين طريقتي التقدير (بييز والأرجحية العظمى) للفترة الزمنية الممتدة من (2000–2022)، ومن ثم إجراء ما وراء التحليل لنتائج هذه الدراسات، وتقديم ادلة إحصائية تدعم تصورات عامة للمفاضلة بينهما، من خلال تحقيق الأهداف الفرعية الآتية:

1. الكشف عن حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022).

2. الكشف عن الاختلاف في حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022) والتي تعزى لكل من طول الاختبار، حجم العينة، النموذج اللوجستي ونوع البيانات.

أهمية الدراسة:

تظهر أهمية الدراسة الحالية في جانبين، الجانب النظري والجانب التطبيقي؛ ففي الجانب النظري تعد هذه الدراسة متماشية مع التوجهات البحثية الحديثة في تناول أسلوب ما وراء التحليل كأسلوب إحصائي يعمل على دمج نتائج الدراسات السابقة وصولًا إلى نتائج أكثر شمولية، وكذلك فإنها ستعتبر مرجع مهم في إنها الجدل المتعلق بمقارنة طريقتي التقدير بييز والأرجحية العظمى والخروج بنتيجة متكاملة تأخذ بعينها الظروف الاختبارية المختلفة، كما تظهر أهمية هذه الدراسة في ضرورة إعطاء أهمية أكبر لاعتماد الدلالة العملية في فحص واختبار الفرضيات الإحصائية والذي يقود إلى موثوقية أعلى في نتائجها.

وتظهر الأهمية التطبيقية لهذه الدراسة من خلال تزويد مقدمي الاختبارات بأنواعها المختلفة بطريقة التقدير الأفضل للحصول على اختبارات دقيقة وموثوقة، تقود إلى اتخاذ قرارات تربوية وتعليمية ونفسية صادقة، في ضوء الظروف الاختبارية الراهنة، كما أنها ستزود باني الاختبار بالطول الأنسب للاختبار عند التقدير بأي من الطريقتين، كما ستوفر هذه الدراسة أداة رصد للبيانات اللازمة لما وراء التحليل المتعلق بفحص الفروق في دقة التقدير في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022)، وفقًا لحجم العينة وطول الاختبار.

حدود الدراسة:

1. الحد الزماني: تم إجراء هذه الدراسة على الدراسات التي بحثت في دقة تقدير القدرة في الفترة الممتدة من (2000–2022).

2. تتحدد طرق التقدير في طريقتي بييز والأرجحية العظمى.

3. تتحدد دقة التقدير في مؤشر الخطأ المعياري.

4. تتحد المتغيرات الوسيطة في حجم العينة وطول الاختبار والنموذج اللوجستي المستخدم ونوع البينات (حقيقي، افتراضية).

التعريفات الاصطلاحية والإجرائية للدراسة:

1. ما وراء التحليل: هو عبارة عن أسلوب إحصائي تكاملي يعمل على تجميع نتائج مجموعة من الدراسات تبحث في المشكلة ذاتها أو تهدف إلى ذات الأهداف تحوي متغيرات متشابهة، ذلك للوصول إلى قيمة تمثل معدل النتائج التي ظهرت في الدراسات السابقة (Gay, 2000).

2. طريقة الأرجحية العظمى: هي إحدى الطرق المستخدمة في برنامج (BILOG) لتقدير معالم الفقرة (الصعوبة، التخمين، التمييز) وقدرة الأفراد والتي تعمل على تعظيم اقتران الاحتمالية (Garee & Vermont, 2006).

3. طريقة بييز: هي إحدى الطرق المستخدمة في برنامج (BILOG) لتقدير معلمة القدرة للفرد والتي تتم من خلال

طريقتين هما: طريقة توقع الاقتران البعدي؛ وطريقة تعظيم الاقتران البعدي (Lewy & Nielsen, 2002).

4. دقة التقدير: وهو عبارة عن مصطلح يشير إلى جودة التقدير وذلك بالاعتماد على الجذر التربيعي لمتوسطات مربعات الانحرافات للفروق بين المعالم المقدرة والمعالم الحقيقية (Rakkapo, Prasitpong & Arayathanitkul, 2016).

الإطار النظري:

تستند نظرية الاستجابة للفقرة في تقدير احتمالية استجابة الفرد على الفقرة والتي تسمى الاستجابة المتوقعة الى نماذج احتمالية لوجستية تختلف باختلاف تدرج الاستجابة على الفقرة وعدد معالم الفقرة وتعتبر النماذج ثنائية التدريج هي النماذج الأشهر وهي:

1. النموذج أحادي المعلمة ويسمى بنموذج راش ويتضمن معلمة واحدة للفقرة وهي معلمة الصعوبة ويشترط تساوي تمييز فقرات الاختبار وتمثل الصيغة () احتمالية الاستجابة على الفقرة وفق هذا النموذج.

$$P_i(\theta) = \frac{1}{1+e^{-D(\theta i - bi)}} \quad (1)$$

حيث:

b: صعوبة الفقرة.

$P_i(\theta)$: احتمال نجاح الفرد على الفقرة.

D: عامل التدريج البالغ (1.7) تقريباً.

θ: قدرة الفرد.

2. النموذج ثنائي المعلمة ويسمى بنموذج بيرنوم ويتضمن معلمة التمييز بالإضافة الى معلمة الصعوبة وتختلف معلمة التمييز باختلاف الفقرات وتمثل الصيغة () احتمالية الاستجابة على الفقرة وفق هذا النموذج.

$$P_i(\theta) = \frac{1}{1+e^{-Dai(\theta i - bi)}} \quad (2)$$

حيث:

a: تمييز الفقرة.

3. النموذج ثلاثي المعلمة ويسمى بنموذج لورد ويتضمن معلمة التخمين بالإضافة الى معلمتي الصعوبة والتمييز وتشير معلمة التخمين الى الحد الأدنى لاحتمالية الاستجابة على الفقرة للأفراد من ذوي القدرة المتدنية وتمثل الصيغة () احتمالية الاستجابة على الفقرة وفق هذا النموذج.

$$P_i(\theta) = c_i + (1 - c_i)\frac{1}{1+e^{-D(\theta i - bi)}} \quad (3)$$

حيث:

c: تخمين الفقرة.

هناك عدة طرق تستخدم لتقدير قدرة المفحوصين والمعالم اللوجستية للفقرات؛ وهذه الطرق هي:

أولًا: طريقة الأرجحية العظمى للتقدير (Maximum Likelihood Estimation)؛ وهي طريقة لتقدير معالم الفقرة وقدرة المفحوص وذلك من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها عندما يتوافر لدينا معلومات عن العينة، ومن أساليب هذه الطريقة:

1. **الأرجحية العظمى المشتركة (Joint Maximum Likelihood):** يتم وفقا لهذه الطريقة تقدير معالم القدرة للأفراد والفقرات معا في آن واحد، إذ تعمل هذه الطريقة على نمذجة احتمالات الاستجابات باستخدام تقديرات مبدئية لمستويات قدرة الأفراد، غير أن هذه الطريقة يشوبها بعض القصور؛ فالتقديرات متحيزة وغير متسقة وفق هذه الطريقة، وبخاصة في النموذجين: ثنائي وثلاثي المعلمة، وكذلك في الاختبارات ذات الطول الثابت، والسبب في ذلك هو أن معالم الفقرات والأفراد يتم تقديرها معا في آن واحد (Burns & wagner, 2008).

2. **طريقة الأرجحية العظمى الشرطية (Conditional Maximum likelihood):** تقوم هذه الطريقة على فصل المعالم الإحصائية للمفحوصين أثناء عملية التدريج، اذ يكون اقترن الاحتمالية (Likelihood) مشروطا بعدد الإجابات الصحيحة للمفحوصين على فقرات الاختبار، والعلامة الكلية للمفحوص تعبر إحصاءات كافية لتقدير مستوى قدرة الفرد (θ) ولذلك يمكن اعتبار الاحتمال المشروط لجميع إجابات المفحوص عن فقرات الاختبار في حال معرفة علامته الكلية على أنه دالة الأرجحية، وبتعظيم هذه الدالة (Maximization Function) يمكننا الحصول على تقديرات الأرجحية العظمى الشرطية لمعالم الصعوبة بصرف النظر عن معالم القدرة (θ) (Lewy & Nielsen, 2002).

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} \qquad (4)$$

حيث:

(A): الشرط؛ (B): المشروط؛ p(A): الاحتمال الشرطي؛ p(B): الاحتمال المشروط؛

.

$$p(\theta a|u) = \frac{p(u|\theta a)p(\theta a)}{p(u)} \qquad (5)$$

٣. **طريقة الأرجحية العظمى الهامشية (Marginal Maximum likelihood):** يتم في هذه الطريقة إيجاد اقتران الاحتمالية الهامشي (Marginal likelihood Function) لمعالم الفقرة، من خلال تكامل اقتران الكثافة الاحتمالية (Density likelihood Function) على معالم القدرة ثم إيجاد تقديرات معالم الفقرات. وعندما تكون تقديرات معالم الفقرات معروفة، فإنه يمكن عندئذ تقدير معالم القدرة للأفراد (Hambelton & Swaminathan, 1985).

ثانيا: الطريقة البييزية (Bayesian Method Estimation–BME)؛ تعتمد هذه الطريقة على تقديرات بييز (Bayesian Estimation)، وتستخدم في العادة عندما لا يكون بالإمكان تطبيق طريقة الأرجحية العظمى، وذلك بوجود الاستجابات المتطرفة عندما يجيب المفحوص عن جميع فقرات الاختبار إجابة صحيحة أو إجابة خاطئة، وتحتاج هذه الطريقة والمتضمنة كذلك في برنامج (BILOG–MG3)، بالإضافة إلى استخدامها معلومات عن العينة، تحتاج إلى استخدام معلومات أولية مسبقة (Prior–Information) وتكون هذه المعلومات متوفرة من خبرات سابقة، وسميت بهذا الاسم لأنها تستخدم نظرية بييز في حساب الاحتمال (Hambelton & Swaminathan,1985)، وبالاعتماد على معادلة الاحتمال الشرطي (1)، والتي تنص على.

p(A/B): احتمال الشرط (A) بشرط (B).

يكون احتمال الاستجابة (u) لذوي القدرة (θ*a*) بشرط قبلي لتوزيع القدرة

وفي طريقة بييز يتم اعتبار (θ) كمية ذات تباين معروف من خلال وصف توزيع احتمالي لـ (θ) يطلق عليه التوزيع القبلي (Prior–Distribution)، ويعتمد تحديد هذا التوزيع الاحتمالي على معرفة الباحث والخبرة السابقة عن (θ) وذلك قبل أن يتم جمع البيانات، ولهذا يطلق عليه التوزيع القبلي، ثم يتم سحب العينة من المجتمع المعبر عنه بدلالة (θ)، ويتم بعد ذلك مراجعة التوزيع القبلي استنادا للمعلومات التي توفرها العينة، حيث يطلق على التوزيع الجديد اسم التوزيع البعدي (Lord, 1986).

هناك أكثر من أسلوب من أساليب التقدير البييزية لتقدير قدرات الأفراد منها: (Swaminathan,1985 & Hambelton):

1. أسلوب بييز للتوقع البعدي: Expected A posteriori(EAP)

يعمل هذا الأسلوب على إيجاد تقديرات القدرة للأفراد بالاستناد إلى تقديرات معالم الفقرات من خلال طريقة الأرجحية العظمى الهامشية، ويحسب التقدير للتوقع البعدي (EAP) من خلال المتوسط الحسابي للتوزيع البعدي للقدرة (θ) دون اللجوء إلى تقريب متتابع كما في طريقة الأرجحية العظمى (MLE). وتقاس درجة الدقة من خلال الخطأ المعياري المحسوب، وذلك من خلال الانحراف المعياري للتوزيع البعدي ل (Posterior Standard Deviation–PSD)، ويكون لتقديرات بييز للتوقع البعدي (EAP) أقل معدل خطا معياري من التقديرات الأخرى بما فيها الأرجحية العظمى وتعظيم الاقتران. وتمتاز هذه الطريقة بأنها غير متكررة (Non–Iterative)، سريعة الحساب، وتعطي تقديرات للقدرة (θ) لجميع أنماط الإجابات، ومن عيوبها عندما يكون عدد فقرات الاختبار قليلا، فإن التقدير يكون متحيزا نحو وسط التوزيع، وذلك لأنها تعتمد على معلومات أولية عن وسط المجتمع وانحرافه المعياري (akkapo, Prasitpong & Arayathanitkul, 2016).

2. أسلوب التقدير البعدي الأعظم: Maximum A posteriori (MAP)

يطلق على هذا الأسلوب اسم (التعظيم البعدي)، وإجراءاتها مشابهه لإجراءات طريقة الأرجحية العظمى، حيث يتم القيام بالتقريب المتتالي لإعطاء تقديرات ثابتة لمعلمة القدرة (θ)، ويتم تقدير (θ) عن طريق تعظيم (θ) نفسها. ويستخدم في هذا الأسلوب معلومات سابقة عن توزيع القدرة (θ)، حيث يتم افتراض هذا التوزيع بناء على معلومات سابقة، وغالبا ما يتم استخدام التوزيع الطبيعي المعياري، بمعنى أن المفحوصين قد تم اختيارهم من توزيع متوسطه يساوي (صفر) وانحرافه المعياري يساوي (1) (Cooper, 2017).

إن الطرق البييزيه تقدم حلولا لمشكلات التقدير لمعالم نماذج السمات الكامنة، وذلك عندما يكون حجم عينة الأفراد صغيرا، وعدد فقرات الاختبار قليلا، ففي مثل هذه الحالات يصبح للمعتقدات القبلية أهمية كبيرة، أما إذا كان حجم العينة كبيرا، فإن هذه المعتقدات تبدو قليلة الأهمية في مثل هذه التقديرات (علام، 2005).

أبرز ما يميز الطرق البييزية هو أنها أوجدت حلا للمشكلة المتمثلة في حصول المفحوص على علامة كاملة، أي في حال قيام المفحوص بالإجابة عن جميع الفقرات بشكل صحيح، وكذلك في حال حصول المفحوص على علامة الصفر، أي أن يجيب المفحوص عن جميع فقرات الاختبار بشكل خاطئ. وعلى الرغم من المآخذ على أساليب بييز، والمتمثلة في طبيعة المعلومات الأولية المتوفرة من خبرات سابقة لدى الباحثين، إلا أن أساليب بييز يتوقع منها تخمين دقة التقدير وخاصة إذا كان حجم العينة صغيرا، وفي الحالة التي يكون فيها تقدير الأرجحية العظمى لا يعمل بشكل جيد (Gao & Chen, 2005).

ثالثا: طريقة التقدير الموزونة: Estimation Method Biweight

تعتبر تقديرات الأرجحية العظمى لقدرات الأفراد، ذات حساسية مفرطة للتشويش، والاختلال، (Disturbance) مثل اللامبالاة والتخمين العشوائي، والذي يحدث غالبا في الاختبارات بأنواعها المختلفة. وتم تطوير هذه الطريقة لتقدم تقديرات قوية لقدرات الأفراد باستخدام الأرجحية العظمى، وفي هذه الطريقة فإن الأرجحية يتم تعديلها (Modified) لتوزين الفقرات التي تكون قريبة من كفاءة الفرد أو المفحوص، إن عملية توزين الفقرات (Weighting) على نحو ملائم يمكن أن يقدم متوسطات مكافئة للتحيزات المتوقعة، وبالتالي الحصول على تقديرات للقدرة أكثر دقة من تلك التي يمكن الحصول عليما، وذلك من خلال استخدام أسلوب الأرجحية العظمى غير الموزونة (Lewy & Nielsen, 2002). وقد أدخل ميسلفي و بوك (Bock & Mislevy, 1982) تعديلات جوهرية على تقدير الأرجحية العظمى بقصد التقليل من حساسيتها لاستجابات المفحوصين التي تتعارض مع نماذج نظرية الاستجابة للفقرة، ومن الأمثلة على ذلك اعتبار الاستجابة المشوشة (Response Disturbance) على الفقرة السهلة إجابة غير صحيحة (incorrect)، بالنسبة للمفحوصين ذوي القدرات العالية، وهذا التعديل استخدم الطريقة الموزونة (Biweight Method) لغايات تقدير القدرة للأفراد المفحوصين.

ما وراء التحليل:

يتضمن الأدب التربوي كم هائل من الدراسات المتشابهة في محتواها من حيث طبيعة الأسئلة والفرضيات والأهداف، إلا أنه من الملاحظ عليها وجود اختلاف وتباين في النتائج التي يتم التوصل بسبب اختلاف حجم العينات او طول الاختبار أو غيرها من المتغيرات المستقلة، الأمر الذي لا يمكننا من الاعتماد على نتائج هذه الدراسات للوصول إلى صورة شاملة وكلية حول طبيعة العلاقة بين تلك المتغيرات قيد المعالجة والتجريب.

من هنا نشأت الحاجة إلى ابتكار أساليب إحصائية تمتلك القدرة على تلخيص تلك النتائج المجمعة من دراسات منفردة، والخروج بنتيجة شاملة وكلية عن تلك المتغيرات، إذ تلخصت هذه الأساليب في طرق ما وراء التحليل (Cohen, 1977).

من هنا يمكن القول بأن أسلوب ما وراء التحليل جاء نتيجة الحاجة الماسة إلى طريقة علمية وعملية لتحليل التحليلات الإحصائية ودمج النتائج التي تم التوصل إليها في مجموعة كبيرة من الدراسات السابقة، للخروج بقرارات أو تدخلات أو معالجات عامة يتم تبنيها من قبل المختصين وصناع القرار (سكران، 2006).

مفهوم ما وراء التحليل:

هناك اتفاق عام على المعنى والهدف الذي يحققه أسلوب ما وراء التحليل، إذ يرى بانجرت ورندنر (Bangert & Rndner, 1991) أن ما وراء التحليل هو طريقة بحث تشتمل على أسلوب منظم لتحويل نتائج الدراسات السابقة إلى نتائج تقاس على مقياس معياري، بحيث تصبح المقارنة بين النتائج ممكنة، مما يساعد في حل التناقضات المحتملة بين نتائج تلك الدراسات وتلخيص النتائج في قيمة واحدة تشكل معدلًا لتلك النتائج.

وعرفه كوهين وآخرون (Cohen, et al, 2007: P291) بأنه: "وسيلة لتقييم نتائج مجموعة من الدراسات المختلفة من خلال تجميعها ومعالجتها إحصائيا على مقياس مشترك"، في حين يعرفه بيرنس وواغنر (Burns & wagner, 2008) على أنه أداة كمية تستخدم لجمع ومقارنة دراسات مماثلة من خلال استخدام القرارات الإحصائية لقياس حجم الأثر عبر دراسات مختلفة.

أما جاي (Gay, 2000) فيرى أن ما وراء التحليل يهدف لتجميع ودمج وتلخيص نتائج مجموعة من الدراسات التي اهتمت بمشكلة بحثية معينة، والتعبير عن تلك النتائج من خلال قيمة رقمية تمثل معدل قيم أحجام الأثر المحسوبة لنتائج الدراسات السابقة.

وفي هذه الدراسة فإن أسلوب ما وراء التحليل يمثل أسلوبًا إحصائيًا يهدف إلى تجميع ودمج وتلخيص نتائج مجموعة من الدراسات ورسائل الماجستير وأطروحات الدكتوراه التي اهتمت بمقارنة فاعلية أسلوبي تقدير بييز والأرجحية العظمى في تقدير القدرة للأفراد للفترة الممتدة من (2000-2022).

أهمية ما وراء التحليل:

شهد أسلوب ما وراء التحليل اهتمامًا متزايدًا من قبل الباحثين، إذ أن هناك عدد من النقاط التي تبين أهمية هذا الأسلوب وهي كما يأتي:

1. تقويم نتائج الدراسات والأبحاث الكمية المشتركة في بحث مشكلات بحثية محددة في جميع مجالات البحث العلمي، مع إتاحة الفرصة للنظر في نتائج تلك الدراسات على اختلاف طرق تحليلها ومؤشراتها الإحصائية، من خلال مقياس معياري موحد يتمثل في حجم الأثر يقدم فرصة للمقارنة بين نتائج هذه الدراسات والمفاضلة بينها.

2. يسهم استخدام أسلوب ما وراء التحليل في استمرارية البحث العلمي وتشكيل مساراته المستقبلية، باعتبار أنه من الأساليب الإحصائية الحديثة نسبيًا والذي ظهر في نهاية السبعينات من القرن الماضي على يد العالم جلاس (عبد الحميد، 2016).

3. يحسن أسلوب ما وراء التحليل من دقة وموثوقية نتائج البحوث التكاملية من خلال إجراءات إحصائية محددة وواضحة من حيث تضمين الدراسات السابقة أو استبعادها من العينة قيد الدراسة، فهو يتعدى ما تقدمه الدراسات المنفردة، ويتيح الفرصة للباحثين لبحث التأثيرات عبر الدراسات المتعددة وفحص درجة اتساقها.

4. يقود أسلوب ما وراء التحليل إلى دمج المعلومات التي يتم الحصول عليها من عدد من الدراسات، الأمر الذي يقود إلى زيادة قوة الاختبارات الإحصائية وتقدير مؤشرات أكثر استقرارًا من التقديرات المشتقة من أي دراسة منفردة (الشرايدة والصمادي، 2018).

خطوات ما وراء التحليل:

تندرج دراسات ما وراء التحليل تحت مظلة الدراسات التركيبية، والتي تعرف بأنها مجموعة من الإجراءات التي تضمن تجميع الدراسات المتعلقة بموضع محدد، وجمع المعلومات منها وتقييم جودتها، ومن ثم العمل على دمج تلك النتائج في نتيجة مركبة لتفسيرها ومناقشتها وتلخيصها، وفق الخطوات الآتية (Cooper, 2017):

1. تحديد مشكلة البحث، ومن ثم بحث ومراجعة الأدب النظري السابق وفق محكات محددة.

2. جمع المعلومات من الدراسات السابقة التي تم جمعها، وتقييم نوعية تلك الدراسات.

3. تحليل ودمج نتائج تلك الدراسات، وتفسير النتائج المركبة لها.

4. إعداد تقرير البحث وعرض النتائج بعد تلخيصها بصورة شمولية متكاملة.

الدراسات السابقة:

هناك العديد من الدراسات التي عملت على مقارنة دقة تقدير الصعوبة والقدرة وفقًا لطريقتي بييز والأرجحية العظمى ومن هذه الدراسات ما يأتي:

هدفت دراسة البنوي وزريقات (2022) التعرف على أثر نسبة القيم المفقودة على دقة التقدير وفق نموذج راش في ضوء طريقتي التقدير بييز والأرجحية العظمى، ولتحقيق ذلك اتبع الباحثان المنهج التجريبي حيث تم استخدام بيانات افتراضية تحاكي الظروف التجريبية المتمثلة في نسبة القيم المفقودة ولها مستويين (10%، 20%) وطريقة التقدير ولها مستويين (بييز، الأرجحية العظمى)، على عينة مكونة من (1000) مفحوص افتراضي واختبار مكون من (50) فقرة تم توليدها من خلال برنامج

(WinGen)، وقد توصلت الدراسة إلى أن الفروق في دقة التقدير تعزى لنسبة القيم المفقودة ولا تعزى لطريقة التقدير. كما أن الفروق في دقة تقدير قدرة الأفراد تعود لطريقة بييز ولكل من النسبتين حيث أن خطا التقدير عند استخدام طريقة الأرجحية كان أعلى على الدوام. مع الانتباه لأن الفروق بين الطريقتين كانت أعلى عند نسبة الفقد (20%).

كما هدفت دراسة ضعضع وطومان وطيفور (2020) إلى الكشف عن أثر طرائق التقدير (الأرجحية العظمى، تقدير بييز، طريقة بروكس) على تقدير معالم المفردة والأفراد ودقتها في ضوء نظرية الاستجابة للمفردة (نموذج راش). ولتحقيق أهداف الدراسة تم توليد عينة عشوائية من (2000) استجابة على (40) مفردة ثنائية استجابة (0 أو1) باستخدام برنامج (WinGen)، وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية في دقة تقدير معالم الأفراد (القدرة) تعزى لعامل حجم العينة وطريقة التقدير والتفاعل بينهما، وكذلك وجود فروق ذات دلالة إحصائية في دقة تقدير قدرة الأفراد بين الطرائق الثلاث.

وهدفت دراسة حمدان (2019) إلى المقارنة بين عدة طرائق لتقدير المعالم في نظرية الاستجابة للمفردة، اعتماداً على الخطأ المعياري للتقدير، ولتحقيق أهداف الدراسة تم استخدام بيانات حقيقية (مكتملة) وبيانات مفقودة حيث أن هذه البيانات هي نتائج تطبيق اختبار اللغة الإنكليزية (2) المطبق على طلاب جامعة حلب كلية التربية بأقسامه للفصل الدراسي الثاني للعام الدراسي (2017/2018) المكون من (60) مفردة ثنائية الاستجابة من نوع الاختيار من متعدد، أما البيانات المفقودة تم الحصول عليها بحذف نسبة (15%) من البيانات الحقيقية باستخدام برنامج (kutools) الذي يعمل ضمن بيئة برنامج (EXCEL)، وقد وتبين أن طرق التقدير القائمة على نظرية بييز هي الأفضل في حالة البيانات المكتملة والمفقودة، كما تبين أن النموذج الثنائي المعلم أعطى تقديرات للقدرة بأقل خطأ معياري مقارنة مع بقية النماذج مهما كانت الطريقة المستخدمة في التقدير في حالة البيانات المكتملة والمفقودة، كما أن هناك أثر إيجابي لتطبيق طريقة (EM) في التعويض عن القيم المفقودة حيث لم نحصل على فروق ذات دلالة إحصائية في تقديرات القدرة والصعوبة بين البيانات المكتملة والمفقودة.

وهدفت دراسة بني عطا (2017) إلى تقصي أثر طول الاختبار وحجم العينة على دقة طرق تقدير معالم الفقرات وقدرات الأفراد في برنامج بايلوج. ولتحقيق الهدف من الدراسة تم توليد بيانات ثنائية التدريج من خلال النموذج اللوجستي ثلاثي المعلمة بواقع (100) مرة لأربعة اختبارات طول كل منها (10، 30، 60، 80) وأربعة مستويات لحجم العينة (250، 500، 1000، 1500) باستخدام برنامج (WINGEN) وباستخدام طرق التقدير(الأرجحية العظمى، توقع الاقتران، تعظيم الاقتران) المستخدمة في برنامج (Bilog– Mg3)، وقد أظهرت نتائج الدراسة وجود أثر ذي دلالة إحصائية لكل من طول الاختبار وحجم العينة وطريقة التقدير في دقة تقديرات معلمة القدرة للفرد، حيث تفوقت طريقة توقع الاقتران على بقية الطرق في دقتها لتقدير معلمة القدرة، و أظهرت النتائج بشكل عام أن طرق التقدير المستخدمة في برنامج بايلوج أنتجت تقديرات دقيقة لمعالم الفقرات وقدرات الأفراد عندما كان طول الاختبار (30) فقرة وأعلى وحجم العينة (500) فأعلى عند استخدام النموذج اللوجستي ثلاثي المعلمة.

كما هدفت دراسة الصبح (2014) إلى المقارنة بين دقة تقدير القدرة باختلاف طول الاختبار وشكل توزيع معلمة القدرة للبيانات الحقيقية والمولدة، ولتحقيق هدف الدراسة تم استخدام بيانات الاختبار الوطني لمادة العلوم للصف الثامن لعام (2013/2012) كمصدر للعينات التي تعتمد على بيانات حقيقية، ثم تم توليد استجابات مماثلة لخصائص العينات التي تعتمد على البيانات الحقيقية باستخدام برمجية (WINGEN) وتم تقدير قدرات الأفراد والخطأ المعياري للتقدير باستخدام، طريقة تقدير الأرجحية العظمى وطريقة توقع التوزيع البعدي والمعمول بها في برمجية (BILOG–MG) لكل من العينات التي تعتمد على

البيانات الحقيقية والمولدة، وقد توصلت الدراسة إلى عدم وجود فروق عملية في دقة التقديرات لمعلمة القدرة باختلاف طبيعة البيانات (حقيقية، مولدة) وذلك باختلاف طول الاختبار (25، 50) فقرة ، واختلاف شكل التوزيع لمعلمة القدرة (طبيعي، ملتوي نحو اليمين، ملتوي نحو اليسار).

كما هدفت دراسة الرحيل والدرابسة (2014) إلى بيان أثر طريقة معالجة القيم المفقودة، وطريقة تقدير قدرات الأفراد على دقة تقدير معالم الفقرات والأفراد. لتحقيق ذلك، تم استخدام بيانات مُولَّدة، باستخدام برنامج (WINGEN) لتوليد استجابات (1500) مفحوص على اختبار مكون من (80) فقرة ثنائية الاستجابة، ومطابقة للنموذج اللوجستي ثلاثي المَعْلَمة، حيث أظهرت النتائج وجود اختلاف في دقة تقدير قدرات الأفراد يُعزى لطريقة معالجة القيم المفقودة، ولصالح طريقة تعظيم التوقعات، ووجود اختلاف في دقة تقدير قدرات الأفراد يُعزى لطريقة التقدير المُستخَدَمة، ولصالح طريقة بييز للتوقع البَعدي، وكذلك وجود اختلاف في دقة تقدير قدرات الأفراد يُعزى للتفاعل بين متغيري طريقة المعالجة للقيم المفقودة، وطريقة التقدير، ولكن لصالح طريقة دالة الاستجابة في المعالجة، في حال استخدام طريقة بييز للتوقع البَعدي في التقدير.

وهدفت دراسة الشريفين (2012) إلى الكشف عن أثر طريقة تقدير معالم الفقرات وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تقدير حجم العينة، ولتحقيق هدف الدراسة تم بناء اختبار في الفيزياء من نوع الاختيار من متعدد أربعة بدائل، تكون بصورته النهائية من (33) فقرة، طبق على عينة قوامها (1000) طالب وطالبة من طلبة الصف الثاني ثانوي العلمي، وقد توصلت الدراسة إلى وجود فروق دالة إحصائيًا في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لمتغير حجم العينة، والتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية لطريقة التقدير، كما اشارت النتائج إلى أن تقديرات معلمة القدرة تزداد في حالة عينة الأفراد ذوي القدرة العالية وعينة الأفراد ذوي القدرة المتدنية عند استخدام طريقة بييز في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسطة باستخدام طريقة الأرجحية العظمى بغض النظر عن حجم العينة.

كما هدفت دراسة الطراونة (2011) إلى مقارنة طرق تقدير القدرة: طريقة الأرجحية العظمى، وطريقة القيمة العظمى للتوزيع البعدي، وطريقة التوقع البعدي، باستخدام النموذج اللوجستي المناسب، اعتماداً على الخطأ المعياري للتقدير، ولأغراض هذه الدراسة فقد تم استخدام نتائج الاختبار الوطني لضبط نوعية التعليم لعام 2010 المعد من قبل وزارة التربية والتعليم الأردنية لطلبة الصف الثامن الأساسي في مادتي الرياضيات والعلوم، إذ استخدمت برمجية (BILOG-MG) للإجابة عن أسئلة الدراسة، وقد تم التوصل الى النتائج التالية: أن طريقة توقع التوزيع البعدي (EAP) أعطت أعلى درجة في دقة التقدير تليها طريقة القيمة العظمى للتوزيع البعدي (MAP) للقدرة، حيث كانت الأخيرة الأقل درجة في دقة تقدير معلمة القدرة، تعاني طريقة الأرجحية العظمى(ML) من زيادة الخطأ المعياري للتقدير عند تقدير القدرة، وذلك عند نهاية طرفي التوزيع.

كما هدفت دراسة عبابنة (2007) إلى مقارنة فاعلية طريقة الأرجحية العظمى وطريقة بييز في تقدير معلمة القدرة عند استخدام النموذج اللوجستي الثلاثي، ولأغراض هذه الدراسة تم استخدام اختبار قدرة عقلية طور في دراسة سابقة، إذ استخدمت برمجية Bilog-Mg للإجابة على أسئلة الدراسة، وقد تم التوصل إلى النتائج التالية: تزداد دقة تقديرات معلمة القدرة في حالة عينة ذوي القدرة العالية وعينة ذوي القدرة المتدنية عند استخدام طريقة بييز مقارنة بدقة تقديرات القدرة عند استخدام طريقة الأرجحية العظمى بغض النظر عن عدد الفقرات، كما تزداد دقة تقديرات معلمة القدرة للمفحوصين عند استخدام طريقة الأرجحية العظمى مقارنة بدقة تقديرات معلمة القدرة عند استخدام طريقة بييز وذلك عند المعايرة بجميع فقرات المقياس، بينما يحدث العكس في حالة الاقتصار على عينة فقرات من المقياس، وتزداد دقة تقديرات معلمة القدرة عند استخدام طريقة بييز

مقارنة بدقة تقديرات معلمة القدرة عند استخدام طريقة الأرجحية العظمى عندما يكون هناك مزاوجة بين متوسط صعوبة فقرات الاختبار ومتوسط قدرة المفحوصين، وفي حالة وجود تباين بين متوسط قدرات المفحوصين ومتوسط صعوبة الفقرة فإن دقة تقديرات معلمة القدرة عند استخدام طريقة الأرجحية العظمى تكون أعلى من دقة تقديرات معلمة القدرة عند استخدام طريقة بييز، وذلك عندما تقترب قدرة المفحوص من متوسط صعوبة الفقرة.

التعليق على الدراسات السابقة:

من خلال مراجعة نتائج الدراسات السابقة والتي عملت على مقارنة دقة وتقدير قدرة الأفراد من خلال طريقة الأرجحية العظمى وطريقة بييز، نجد أن هناك عدم استقرار في النتائج التي يتم التوصل إليها في تقدير قدرات الأفراد من خلال طريقتي التقدير بييز والأرجحية العظمى، وهو الأمر الذي يجب دراسة بشكل متفحص وموسع على كم كبير من الدراسات الأخرى التي استخدمت كلا الأسلوبين وذلك لإزالة الغموض والخروج بنتيجة عامة يمكن تعميمها على نطاق واسع، ومن ثم توفير مرجعية علمية واضحة تبين أي من الطريقتين يجب استخدامها في ظل عدد فقرات الاختبار أو حجم العينة، وهذا هو الهدف الذي ستسعى له الدراسة الحالية والذي يعد تفريدًا واضحًا لها عن باق الدراسات.

منهج الدراسة: تم اتباع المنهج الوصفي التحليلي لتوظيف أسلوب ما وراء التحليل (Meta-analysis)، كمجموعة من الإجراءات الإحصائية التي تعمل على توليف ودمج النتائج من دراسات عدة منفردة، للوصول إلى قيمة تلخص تلك النتائج وتساعد في فهمها بصورة أشمل، إذ أن مثل هذا الأسلوب يتناسب مع الهدف العام للدراسة والمتمثل في تجميع أكبر قدر ممكن من نتائج الدراسات التي عملت على المقارنة بين طريقتي التقدير (بييز والأرجحية العظمى) لتحديد أيهما أفضل في ظل العديد من الظروف التجريبية المتعلقة بالاختبار والعينة للفترة الزمنية الممتدة من (2000-2022).

مجتمع الدراسة: تكون مجتمع الدراسة من جميع الدراسات والأبحاث التي تم إجراؤها في الفترة (2000 – 2022) والتي تناولت طريقتي التقدير (بييز والأرجحية العظمى) المستخدمتان في تقدير قدرة الأفراد في نظرية الاستجابة للفقرة، وخصوصا الدراسات التي قارنت بين الطريقتين من خلال مؤشرات دقة التقدير.

عينة الدراسة: تم اختيار الدراسات والأبحاث التي توفرها مواقع ومحركات البحث العلمية المتخصصة في الأبحاث والرسائل وهي (دار المنظومة، باحث جوجل العلمي، سكوبس (scopus)، ابسكو (ebsco)، المنهل) وغيرها من محركات البحث المتاحة، ضمن الفترة الزمنية (2000 – 2022م)، والتي تنطبق مع المعايير الخاصة بهذه الدراسة مثل الفترة الزمنية والمتغيرات وطريقة المقارنة. وقد بلغ عددها (12) دراسة وبتفرعات بلغ عددها (55) دراسة منفصلة. وقد وقعت الدراسات التي تم تضمينها حسب معايير التضمين في هذه الدراسة ضمن الفترة (2007 – 2022).

أداة الدراسة: تم عمل نموذج لترميز البيانات والمتمثلة في نتائج الدراسات في ضوء متغيرات الدراسة، حيث تم ترميز كل دراسة بكل فروعها وحسب ظروفها التجريبية كوحدة تحليل، تحتوي سنة إجراء الدراسة والمعلمة المقدرة ونوع البيانات وطول الاختبار والمتوسط الحسابي لخطأ التقدير والانحراف المعياري لخطأ التقدير والنموذج اللوجستي المستخدم وحجم الأثر والاوزن، وتم ترتيب بيانات الدراسات في النموذج المبين في الجدول (1).

جدول (1): النموذج الترميزي لبيانات الدراسات

المعلمة المقدرة:	1. القدرة

حجم العينة:	1. =< 990	2. >990	
النموذج اللوجستي:	1. أحادي المعلمة	2. ثنائي المعلمة	3. ثلاثي المعلمة
نوع البيانات:	1. افتراضية	2. حقيقية	

رقم الدراسة المرجعي (حسب الظرف التجريبي) ID	سنة إجراء الدراسة	المعلمة المقدرة	نوع البيانات	النموذج اللوجستي المستخدم	بييز			الأرجحية العظمى			حجم الأثر	الوزن
					حجم العينة	المتوسط الحسابي لخطأ التقدير	الانحراف المعياري لخطأ التقدير	طول الاختبار/حجم العينة ***	المتوسط الحسابي لخطأ التقدير	الانحراف المعياري لخطأ التقدير		
1												
2												
.												

وكما هو مبين في النموذج أن كل دراسة تحتوي حجم العينة ووصف لتوزيع دقة تقدير كل من قدرة الأفراد والصعوبة باستخدام طريقتي التقدير (بييز والأرجحية العظمى) من خلال الوسط الحسابي والانحراف المعياري والعدد بالإضافة إلى النموذج المعلمي اللوجستي المستخدم ونوع البيانات المستخدمة (حقيقية وافتراضية).

صدق أداة الدراسة:

تم التحقق من صدق أداة الدراسة من خلال عرض نموذج الترميز المستخدم على مجموعة محكمين من أصحاب الاختصاص والخبرة في القياس والتقويم التربوي، لإبداء الراي والملاحظات للأخذ بما يمكن منها. من حيث مناسبة النموذج وطريقة الترميز والتصنيف وشمولية النموذج وقدرته على الإيفاء بالنتائج المطلوبة للتحليل والمؤشرات

(6)

حيث:P_o : الاتفاق المرصود نسبيًا بين المقيمين أو المرمزين.P_e : الاحتمال الافتراضي لاتفاق الصدفة. المعيار الخاص بمعامل كابا:1 = kappa الحكمان القادرة على المقارنة بين الطريقتين. حيث تم التحقق من مدى مناسبة المتغيرات الوسيطة لهذه الدراسة واقتراح (النموذج اللوجستي ونوع البيانات) كمتغيرات وسيطة بالإضافة إلى حجم العينة وطول الاختبار.

ثبات أداة الدراسة:

تم التحقق من ثبات أداة الدراسة من خلال ترميز البيانات من قبل مختص في القياس بالإضافة للترميز الذي أجراه الباحث وذلك للتأكد من مدى التوافق بين الترميزين ودقة وموثوقية الترميز وقد تم التوفيق بين الترميزين، باستخدام معامل كابا من خلال الصيغة الرياضية الآتية؛ والذي بلغ (0.78)، وهي قيمة تدل على الموافقة المرضية في نسبة التوافق، بالاعتماد على المعيار الخاص بمعامل كابا.

$$k = \frac{p_o - p_e}{1 - p_e} =$$

$$k = \frac{0.89 - 0.5}{1 - 0.5} = 0.78$$

متماثلان تمامًا.0.75 = kappa> درجة موافقة مرضية. كابا 0.4>تناسق غير مرض (عبابنة، 2009).

إجراءات الدراسة:

لتحقيق أهداف الدراسة تم اتباع الإجراءات التالية:

أولا: تحديد الدراسات السابقة بالاعتماد على المتغيرات محور الدراسة وهي (طريقتي التقدير بييز والأرجحية العظمى) حيث تم البحث عنها في محركات البحث على الشبكة العنكبوتية المتخصصة وقواعد البيانات البحثية المعروفة والتي أجريت ضمن السنوات المحددة (2000 – 2022) وذلك من خلال الكلمات المفتاحية (بييز، الأرجحية العظمى، تقدير، قدرة).

ثانيا: تضمين الدراسات التي تحقق المعايير التي تتم في ضوئها هذه الدراسة والمتمثلة في:

دراسات تجريبية أو شبه تجريبيه. دراسات قارنت بين طريقتي التقدير (بييز والأرجحية العظمى). دراسات تقارن دقة تقدير قدرة الأفراد. دراسات تم إجراؤها في الفترة الزمنية (2000 – 2022). دراسات اعتمدت على الخطأ المعياري كمؤشر لدقة التقدير. دراسات محددة النموذج اللوجستي المستخدم أي أنها اعتمدت أحد النماذج ثنائية التدريج الخاصة بنظرية الاستجابة للفقرة لتقدير قدرة الأفراد أو صعوبة الفقرات. دراسات تتمتع بدرجة مقبولة من الصدق والثبات. دراسات تم التحقق من مطابقة بياناتها للنموذج اللوجستي المستخدم وكذلك تم التحقق من مطابقة الفقرات والأفراد للنموذج.

ثالثا: تحديد متغيرات الدراسة وهي كالاتي:

- المتغيرات المستقلة: وتتمثل في طريقة التقدير ولها مستويان (طريقة بييز، طريقة الأرجحية العظمى).
- المتغيرات التابعة: وهي دقة تقدير قدرة الفرد (θ) وتتمثل في الخطأ المعياري لدقة التقدير.
- المتغيرات الوسيطة: وهي المتغيرات التي تساعد على إجراء الدراسة بشكل تفصيلي أكثر حيث تم اعتماد المتغيرات الآتية: النموذج اللوجستي وله ثلاثة مستويات (النموذج أحادي المعلمة، النموذج ثنائي المعلمة، النموذج ثلاثي المعلمة). طول الاختبار وله مستويان (أقل من أو يساوي 33 فقرة، وأكبر من 33 فقرة). حجم العينة وله مستويان (أقل من أو يساوي 990، وأكبر من 990). نوع البيانات وله مستويان (حقيقي، افتراضي).

رابعا: ترميز البيانات؛ تم تناول كل دراسة من الدراسات التي تم قبول عضويتها في هذه الدراسة وترميز بياناتها وفق النموذج الترميزي. والتأكد من الحصول على معلومات المتغيرات الوسيطة لكل من تلك الدراسات.

خامسًا: الإحصائيات المستخدمة

1. تم استخدام معادلات وصيغ مختلفة لحساب حجم الأثر وذلك مراعاة لظروف كل دراسة من الدراسات.

2. تم الاعتماد على الخطأ المعياري في تقدير كل من معالم فقرات الاختبار والقدرة بالطريقتين لحساب حجم الأثر، حيث تم اعتبار قيم خطأ التقدير المعيارية لمعلمة القدرة بطريقة بييز هي العينة الأولى، واعتبار قيم خطأ التقدير المعيارية لمعلمة القدرة بطريقة الأرجحية العظمى هي العينة الثانية، وتم حساب المتوسطات الحسابية لخطأ التقدير المعياري لكل من معالم قدرة الأفراد.

3. تم حساب الوزن لكل من الدراسات (w_i) من خلال من خلال قيمة معكوس التباين لكل من الدراسات من خلال المعادلة (7):

$$w_i = \frac{1}{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}} \quad (7)$$

حيث: s_t^2 : تباين المجموعة التجريبية (الأولى). s_c^2 : تباين المجموعة الضابطة (الثانية). n_t : حجم العينة التجريبية. n_c : حجم العينة الضابطة.

4. تم حساب حجم الأثر العام من خلال استخدام الأوساط الموزونة كما هو مبين في الصيغة (8).

$$d^* = \frac{\sum_{i=1}^{m} w_i d_i}{\sum_{i=1}^{m} w_i} \quad (8)$$

حيث: d^* : حجم الأثر العام. m : عدد الدراسات.

5. تم حساب حجم الأثر الكلي للفروق في دقة تقدير كل من القدرة باستخدام طريقتي التقدير بييز والارجحية العظمى في الدراسات المضمنة باستخدام برمجية (SPSS.28).

سابعًا: تم إجراء تحليل البيانات بأسلوب ما وراء التحليل من خلال برمجية (SPSS.28) وتمثيل أحجام الأثر بيانياً.

عرض نتائج الدراسة ومناقشتها:

للإجابة عن السؤال الأول للدراسة والذي نص على: "**ما حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022)؟**"

تم تقدير حجم الأثر ووزن كل من الدراسات التي قارنت بين طريقتي بييز والأرجحية العظمى في دقة تقدير القدرة ولكل من ظروفها التجريبية، وذلك بالاعتماد على الوسط الحسابي والانحراف المعياري لتوزيع قيم الخطأ المعياري لتقدير قدرة الأفراد بطريقتي التقدير كما هو مبين في الجدول (2).

جدول (2): الأوساط الحسابية والانحرافات المعيارية لتوزيع قيم الخطأ المعياري لتقدير قدرة الأفراد بطريقتي التقدير

الترميز الظرفي	الدراسة	حجم الأثر	المعياري الخطأ	Z	القيمة الاحتمالية	فترة الثقة		الوزن	النسبي الوزن
						الحد الأدنى	الحد الأعلى		
5	ضعضع، 2020	-.057	0.0062	-9.190	0	-.070	-.045	15.29	1.9
6	ضعضع، 2020	-.053	0.0041	-12.998	0	-.062	-.045	15.295	1.9
7	ضعضع، 2020	-.038	0.0034	-11.435	0	-.045	-.032	15.297	1.9
8	ضعضع، 2020	0	0.003	0	1	-.006	0.006	15.297	1.9
9	عبابنة، 2007	-.061	0.0021	-28.856	0	-.065	-.057	15.298	1.9
10	عبابنة، 2007	-.459	0.0047	-98.073	0	-.468	-.450	15.294	1.9
11	عبابنة، 2007	0.055	0.0047	11.807	0	0.046	0.064	15.294	1.9
12	عبابنة، 2007	0.069	0.0018	38.52	0	0.066	0.073	15.299	1.9
13	عبابنة، 2007	-.092	0.0032	-28.906	0	-.098	-.085	15.297	1.9
14	عبابنة، 2007	-.218	0.0035	-61.702	0	-.225	-.211	15.297	1.9
15	عبابنة، 2007	1.131	0.0026	435.135	0	-1.136	-1.126	15.298	1.9
16	عبابنة، 2007	-.276	0.0032	-85.766	0	-.283	-.270	15.297	1.9
18	حمادنه والرحيل 2014	-.051	0.0018	-28.512	0	-.055	-.048	15.299	1.9
19	الطراونة، 2011	-.121	0.0024	-51.444	0	-.126	-.117	15.298	1.9
20	الطراونة، 2011	-.094	0.0028	-33.615	0	-.100	-.089	15.298	1.9
26	البنوي وزريقات، 2022	-.040	0.0028	-14.234	0	-.046	-.034	15.298	1.9
27	البنوي وزريقات، 2022	-.052	0.0026	-20.321	0	-.057	-.047	15.298	1.9
45	حمدان، 2019	-.059	0.0051	-11.660	0	-.069	-.049	15.293	1.9
47	حمدان، 2019	-.054	0.008	-6.828	<.001	-.070	-.039	15.285	1.9
49	حمدان، 2019	-.238	0.0361	-6.606	<.001	-.309	-.168	15	1.8
51	حمدان، 2019	-.052	0.0054	-9.610	0	-.063	-.042	15.293	1.9
53	حمدان، 2019	-.058	0.0083	-7.006	<.001	-.075	-.042	15.283	1.9
55	حمدان، 2019	-.176	0.024	-7.328	<.001	-.223	-.129	15.165	1.8
62	الصبح، 2014	-.073	0.008	-9.151	0	-.089	-.057	15.285	1.9
63	الصبح، 2014	-.089	0.013	-6.851	<.001	-.114	-.064	15.26	1.8
64	الصبح، 2014	-.094	0.0154	-6.102	<.001	-.124	-.064	15.244	1.8
65	الصبح، 2014	-.458	0.0524	-8.736	0	-.561	-.355	14.682	1.8
66	الصبح، 2014	-.765	0.06	-12.748	0	-.883	-.647	14.501	1.8
67	الصبح، 2014	1.003	0.1323	-7.583	<.001	-1.262	-.744	12.069	1.5
68	الصبح، 2014	-.156	0.007	-22.414	0	-.170	-.142	15.288	1.9
69	الصبح، 2014	-.117	0.0086	-13.671	0	-.134	-.100	15.282	1.8
70	الصبح، 2014	-.822	0.0778	-10.562	0	-.975	-.669	14.002	1.7
71	الصبح، 2014	-.162	0.0188	-8.629	0	-.199	-.125	15.217	1.8
72	الصبح، 2014	1.115	0.1624	-6.866	<.001	-1.433	-.797	10.901	1.3
73	الصبح، 2014	-.784	0.0935	-8.384	0	-.967	-.601	13.494	1.6
77	الشريفين، 2012	-.230	0.0888	-2.589	0.01	-.404	-.056	13.651	1.7
78	الشريفين، 2012	-.300	0.0552	-5.433	<.001	-.408	-.192	14.618	1.8
79	الشريفين، 2012	-.140	0.0236	-5.941	<.001	-.186	-.094	15.171	1.8
80	بني عطا، 2017	-.159	0.004	-39.750	0	-.167	-.151	15.296	1.9
81	بني عطا، 2017	-.123	0.0017	-74.334	0	-.126	-.120	15.299	1.9
82	بني عطا، 2017	-.099	0.0009	-113.786	0	-.101	-.097	15.299	1.9
83	بني عطا، 2017	-.107	0.0007	-155.965	0	-.108	-.106	15.299	1.9
84	بني عطا، 2017	-.109	0.0019	-57.416	0	-.113	-.105	15.299	1.9
85	بني عطا، 2017	-.074	0.0009	-86.023	0	-.076	-.072	15.299	1.9
86	بني عطا، 2017	-.077	0.0005	-164.916	0	-.078	-.076	15.299	1.9
87	بني عطا، 2017	-.052	0.0003	-160.731	0	-.053	-.051	15.299	1.9
88	بني عطا، 2017	-.059	0.002	-29.784	0	-.063	-.055	15.299	1.9
89	بني عطا، 2017	-.094	0.0051	-18.393	0	-.104	-.084	15.293	1.9
90	بني عطا، 2017	-.082	0.0006	-143.838	0	-.083	-.081	15.299	1.9

91	بني عطا، 2017	-.077	0.0004	-214.109	0	-.078	-.076	15.299	1.9
92	بني عطا، 2017	-.045	0.0065	-6.963	<.001	-.058	-.032	15.29	1.9
93	بني عطا، 2017	-.043	0.0007	-58.087	0	-.044	-.042	15.299	1.9
94	بني عطا، 2017	-.043	0.0004	-116.600	0	-.044	-.042	15.299	1.9
95	بني عطا، 2017	-.031	0.0003	-121.905	0	-.031	-.031	15.299	1.9
97	الرحيل والدرابسة، 2014	-.051	0.0018	-28.724	0	-.055	-.048	15.299	1.9

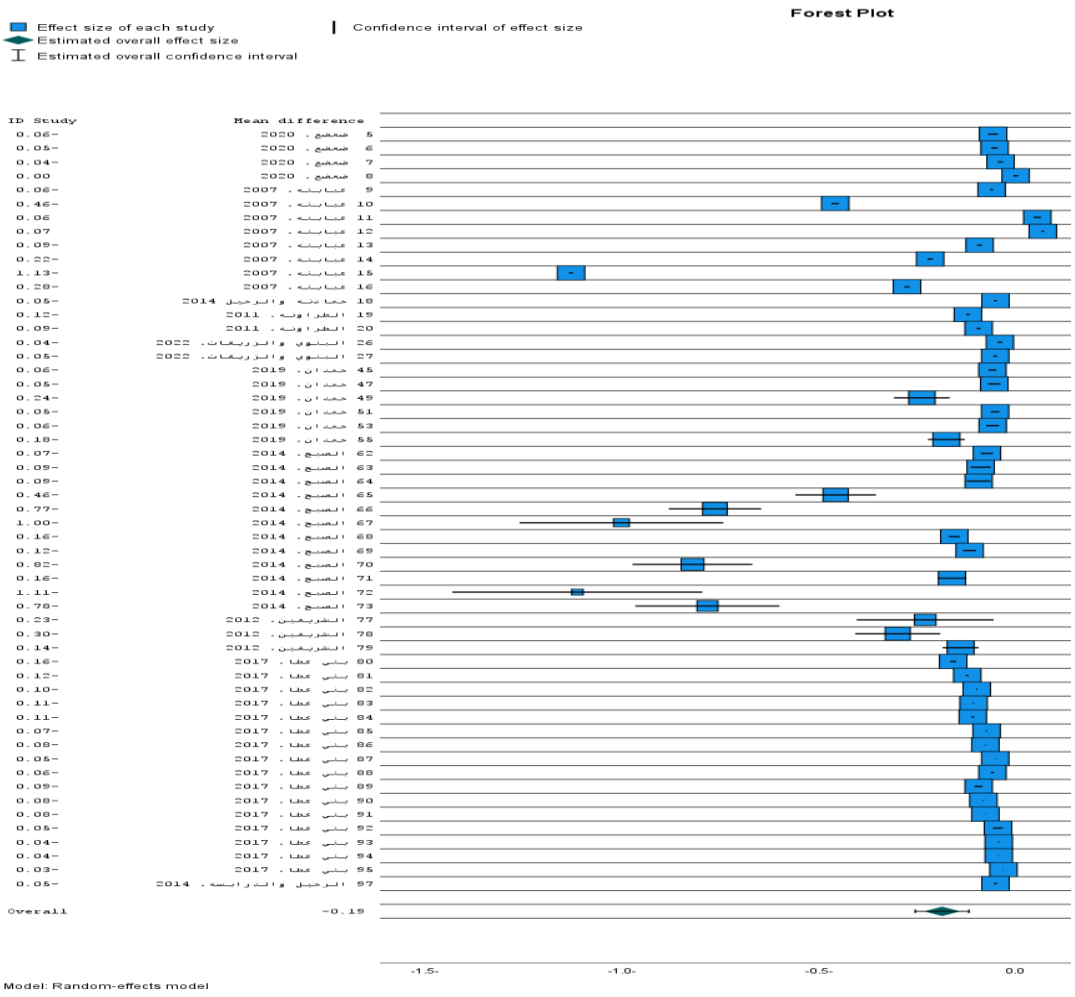
يتبين من الجدول (2) أن حجم الأثر الكلي بالاعتماد على أحجام الأثر وأوزان كل من الدراسات المضمنة في عينة الدراسة قد بلغت قيمته (0.187) بخطأ معياري (0.0348) حيث كانت قيمته ذات دلالة إحصائية كما هو مبين في الجدول (3).

جدول (3): حجم الأثر الكلي لطريقة التقدير على دقة تقدير القدرة التي تعزى للدراسات المضمنة في عينة الدراسة

حجم الأثر	الخطأ المعياري	Z	القيمة الاحتمالية**	فترة الثقة		فترة التنبؤ *	
				الحد الأدنى	الحد الأعلى	الحد الأدنى	الحد الأعلى
−.187	.0348	−5.372	<.001	−.255	−.119	−.187	.0348

*: بالاعتماد على توزيع t. **: مستوى الدلالة 0.05

ويعد حجم الأثر البالغ (0.187) صغير حسب تصنيف كوهين لمستويات حجم الأثر (Cohen, 1977)، ويشير لان الفروق في تقدير القدرة التي تعزى لطريقة التقدير كانت بشكل عام لصالح طريقة الأرجحية العظمى ولكنها في مستوى صغير عمليا. وهذا ما يوضحه التمثيل البياني لأحجام الأثر للدراسات وحجم الأثر الكلي باستخدام التمثيل البياني (Forest) كما هو مبين في الشكل (1).



شكل (1): التمثيل البياني لأحجام أثر طريقة التقدير على دقة تقدير القدرة لكل من الدراسات وحجم الأثر الكلي

يتضح من التمثيل البياني في الشكل (1) أن بعض الدراسات تظهر حجم أثر مرتفع للفروق في دقة تقدير القدرة باختلاف طريقة التقدير مثل الدراسة (عبابنة، 2007) الظرف (15) ولصالح طريقة الأرجحية العظمى كما أنها كانت بوزن مرتفع حيث كانت مساحة الصندوق الممثل لها كبيرة بالمقارنة مع

دراسة (الصبح، 2014) الظرف (72) التي كان وزنها منخفض نسبياً وتتسم بالدقة المنخفضة والتي تظهر من خلال فترة الثقة العالية حولها.

ويمكن تفسير ذلك من خلال الفروق الطفيفة في دقة التقدير بين الطريقتين عند وجود اختبارات طويلة ومتوسطة الطول، حيث أن طريقة بييز تكون أدق عندما تكون الاختبارات قصيرة جدا وهنالك احتمالية كبيرة لظهور الأنماط المتطرفة من الاستجابات.

إذ تتفق نتيجة هذا السؤال مع نتائج جميع الدراسات التي تم تضمينها في هذه الدراسة وبكل تفرعاتها باستثناء دراسة (عبابنة، 2007) الدراستين الفرعيتين (11، 12) حسب ترميزها من قبل الباحث في هذه الدراسة في الملحق (أ) والتي اختلفت مع هذه الدراسة وكشفت عن أفضلية طريقة بييز في تقدير القدرة بمقابل طريقة الأرجحية العظمى.

للإجابة عن السؤال الأول للدراسة والذي نص على:" هل يختلف حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات التي أجريت في الفترة الممتدة من (2000–2022) والتي تعزى لكل من طول الاختبار، حجم العينة، النموذج اللوجستي ونوع البيانات؟"، تم إجراء تحليل ما وراء التحليل للدراسات المضمنة في عينة الدراسة بعد تجميعها حسب المتغير الوسيط كما يلي:

أولا: طول الاختبار حيث تم تجميع البيانات في مجموعتين كما هو مبين في الجدول (4).

جدول (4): الدراسات التي تناولت أثر طريقة التقدير على دقة تقدير القدرة حسب طول الاختبار

طول الاختبار	عدد الدراسات	النسبة المئوية (%)
=<33	29	52.7
>33	26	47.3
المجموع	55	100.0

حيث تم اعتبار عدد الفقرات (33) هو الحد الفاصل لطول الاختبار بين المجموعتين لان القيمة التي تمثل الوسيط لجميع الظروف في الدراسات المضمنة في عينة الدراسة لأجل الحصول على مجموعات متساوية وتماشياً مع الدراسات التي كانت تبحث في أثر طول الاختبار حيث كانت تعتمد على أطوال مثل (25، 40) في اغلب الأحيان وتعتبر القيمة(33) قيمة متوسطة.

ولتقدير حجم الأثر الكلي لكل من المجموعتين تم إجراء تحليل ما بعد التحليل على المجموعتين كما هو مبين في الجدول (5).

جدول (5): حجم الأثر الكلي لطريقة التقدير على دقة تقدير القدرة للدراسات المجمعة بحسب طول الاختبار

طول الاختبار	حجم الأثر	الخطأ المعياري	Z	القيمة الاحتمالية**	فترة الثقة		فترة التنبؤ *	
					الحد الأدنى	الحد الأعلى	الحد الأدنى	الحد الأعلى
=<33	-.221	.0540	-4.089	<.001	-.327	-.115	-.823	.381
>33	-.147	.0414	-3.559	<.001	-.229	-.066	-.587	.292
الكلي	-.187	.0348	-5.372	<.001	-.255	-.119	-.704	.331

*: بالاعتماد على توزيع t. **: مستوى الدلالة 0.05

يتبين من خلال الجدول (5) أن حجم الأثر الكلي قد بلغ (0.221) للاختبارات القصيرة ولصالح طريقة الأرجحية العظمى، وهي قيمة مقبولة الى حد ما. في حين بلغ حجم الأثر الكلي (0.147) للاختبارات الطويلة ولصالح طريقة الأرجحية العظمى وعلى الرغم من دلالته الإحصائية الا انه يعتبر في الحدود المنخفضة.

ولفحص اختلاف معدل حجم الأثر لطريقة التقدير على دقة تقدير القدرة لمجموعات الدراسات المصنفة حسب

طول الاختبار تم استخدام اختبار كوكران (Q) لفحص تجانس حجوم الأثر كما هو مبين في الجدول (6).

الجدول (6): فحص تجانس معدل حجم الأثر الكلي لطريقة التقدير على دقة تقدير القدرة باختلاف طول الاختبار.

قيمة اختبار كوكران	درجة الحرية	الدلالة
1.166	1	.280

من نتائج اختبار كوكران يتبين تحقق تجانس حجوم الأثر وعدم اختلافها باختلاف طول الاختبار. أي أن حجم أثر طريقة التقدير على دقة تقدير القدرة لم يختلف لمجموعتي الدراسات ذات الاختبارات القصيرة (أقل أو تساوي 33 فقرة) والاختبارات الطويلة (أكثر من 33 فقرة).

ثانياً: حجم العينة حيث تم تجميع البيانات في مجموعتين كما هو مبين في الجدول (7).

جدول (7): الدراسات التي تناولت أثر طريقة التقدير على دقة تقدير القدرة حسب حجم العينة

حجم العينة	عدد الدراسات	النسبة المئوية (%)
=<990	26	47.3
>990	29	52.7
المجموع	55	100.0

حيث تم اعتبار حجم العينة (990) هو الحد الفاصل لحجم العينة بين المجموعتين لأنها القيمة التي تمثل الوسيط تقريباً لأجل الحصول على مجموعات متساوية وتماشياً مع الدراسات التي كانت تبحث في أثر حجم العينة حيث كانت تعتمد على أطوال حول العدد (1000) في اغلبها.

ولتقدير حجم الأثر الكلي لكل من المجموعتين تم إجراء تحليل ما بعد التحليل على المجموعتين كما هو مبين في الجدول (8).

جدول (8): حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات المجمعة بحسب حجم العينة

حجم العينة	حجم الأثر	الخطأ المعياري	Z	القيمة الاحتمالية**	فترة الثقة		فترة التنبؤ *	
					الحد الأدنى	الحد الأعلى	الحد الأدنى	الحد الأعلى
=<990	-.178	.0400	-4.437	<.001	-.256	-.099	-.600	.245
>990	-.193	.0556	-3.479	<.001	-.302	-.084	-.815	.428
الكلي	-.187	.0348	-5.372	<.001	-.255	-.119	-.70	.331

*: بالاعتماد على توزيع t. **: مستوى الدلالة 0.05

يتبين من خلال الجدول (8) أن حجم الأثر الكلي قد بلغ (0.178) للعينات الصغيرة ولصالح طريقة الأرجحية العظمى، كما أن حجم الأثر الكلي قد بلغ (0.193) للعينات الكبيرة ولصالح طريقة الأرجحية العظمى ويلاحظ أنها قيم صغيرة بالإضافة إلى أن الفروق بين حجمي الأثر فروق طفيفة.

ولفحص اختلاف معدل حجم الأثر لطريقة التقدير على دقة تقدير القدرة لمجموعات الدراسات المصنفة حسب حجم العينة تم استخدام اختبار كوكران (Q) لفحص تجانس حجوم الأثر كما هو مبين في الجدول (9).

الجدول (9): فحص تجانس معدل حجم الأثر الكلي لطريقة التقدير على دقة تقدير القدرة باختلاف حجم العينة.

قيمة اختبار كوكران	درجة الحرية	الدلالة
.054	1	.817

من نتائج اختبار كوكران يتبين تحقق تجانس حجوم الأثر وعدم اختلافها باختلاف حجم العينة. أي ان حجم أثر طريقة التقدير على دقة تقدير القدرة لم يختلف لمجموعتي الدراسات ذات الاحجام الصغيرة (الأصغر من 990 فرد) والاحجام الكبيرة (الأكبر من 990 فرد).

ثالثاً: النموذج المستخدم حيث تم تجميع البيانات في ثلاث مجموعات كما هو مبين في الجدول (10).

جدول (10): الدراسات التي تناولت أثر طريقة التقدير على دقة تقدير القدرة حسب النموذج اللوجستي المستخدم

النموذج	عدد الدراسات	النسبة المئوية (%)
1plm	8	14.5
2plm	2	3.6
3plm	45	81.8
المجموع	55	100.0

يلاحظ من الجدول (10) أن توزيع الدراسات حسب النموذج اللوجستي المستخدم غير منتظم ويلاحظ أن النموذج الأكثر استخداما هو النموذج الثلاثي ويعزى ذلك لعدة أسباب منها مراعاة النموذج الثلاثي لظروف الاختبار فمن جهة يحتمل وجود تخمين ومن جهة أخرى يفترض اختلاف تمييز الفقرات وكذلك يعتبر النموذج الثلاثي من النماذج الأكثر رواجا لقدرته على التعامل مع الأنواع المختلفة من الاختبارات.

ولتقدير حجم الأثر الكلي لكل من المجموعات الثلاثة تم إجراء تحليل ما بعد التحليل على كل منها كما هو مبين في الجدول (11).

جدول (11): حجم الأثر الكلي للفروق في دقة تقدير القدرة في ضوء طريقتي التقدير (بييز والأرجحية العظمى) للدراسات المجمعة بحسب النموذج اللوجستي المستخدم

النموذج	حجم الأثر	الخطأ المعياري	Z	القيمة الاحتمالية**	فترة الثقة		فترة التنبؤ *	
					الحد الأدنى	الحد الأعلى	الحد الأدنى	الحد الأعلى
1plm	-.044	.0069	-6.330	<.001	-.057	-.030	-.094	.006
2plm	-.056	.0058	-9.776	.000	-.068	-.045	.	.
3plm	-.220	.0419	-5.264	<.001	-.303	-.138	-.788	.347
الكلي	-.187	.0348	-5.372	<.001	-.255	-.119	-.704	.331

*: بالاعتماد على توزيع t. **: مستوى الدلالة 0.05

يتبين من خلال الجدول (11) أن حجم الأثر الكلي الأعلى قد بلغ (0.220) للنموذج الثلاثي المعلمة ولصالح طريقة الأرجحية العظمى، وهو في الحدود المقبولة. إلا أن حجم الأثر الكلي الأقل قد بلغ (0.044) للنموذج أحادي المعلمة ولصالح طريقة الأرجحية العظمى.

ولفحص اختلاف معدل حجم الأثر لطريقة التقدير على دقة تقدير القدرة لمجموعات الدراسات المصنفة حسب النموذج اللوجستي المستخدم تم استخدام اختبار كوكران (Q) لفحص تجانس حجوم الأثر كما هو مبين في الجدول (12).

الجدول (12): فحص تجانس معدل حجم الأثر الكلي لطريقة التقدير على دقة تقدير القدرة باختلاف النموذج اللوجستي.

قيمة اختبار كوكران	درجة الحرية	الدلالة
18.081	2	<.001

من نتائج اختبار كوكران يتبين عدم تحقق تجانس حجوم الأثر واختلافها باختلاف النموذج اللوجستي المستخدم. أي ان حجم أثر طريقة التقدير على دقة تقدير القدرة يختلف لمجموعات الدراسات ذات النماذج اللوجستية المختلفة.

رابعاً: نوع البيانات حيث تم تجميع البيانات في مجموعتين كما هو مبين في الجدول (13).

جدول (13): الدراسات التي تناولت أثر طريقة التقدير على دقة تقدير القدرة حسب نوع البيانات

نوع البيانات	عدد الدراسات	النسبة المئوية (%)
محاكاة	41	74.5
حقيقية	14	25.5
المجموع	55	100.0

حيث يلاحظ ان النسبة الأعلى من الدراسات تستخدم البيانات الافتراضية التي تحاكي الظروف التجريبية للمقارنة بين طريقتي التقدير.

ولتقدير حجم الأثر الكلي لكل من المجموعتين تم إجراء تحليل ما بعد التحليل على المجموعتين كما هو مبين في الجدول (14).

جدول (14): حجم الأثر الكلي للدراسات المجمعة بحسب نوع البيانات

نوع البيانات	حجم الأثر	الخطأ المعياري	Z	القيمة الاحتمالية**	فترة الثقة		فترة التنبؤ *	
					الحد الأدنى	الحد الأعلى	الحد الأدنى	الحد الأعلى
محاكاة	-.174	.0401	-4.348	<.001	-.253	-.096	-.696	.347
حقيقية	-.225	.0723	-3.113	.002	-.367	-.083	-.829	.378
الكلي	-.187	.0348	-5.372	<.001	-.255	-.119	-.704	.331

*: بالاعتماد على توزيع t. **: مستوى الدلالة 0.05

يتبين من خلال الجدول (14) أن حجم الأثر الكلي قد بلغ (0.174) للدراسات التي تستخدم البيانات الافتراضية ولصالح طريقة الأرجحية العظمى وهي قيمة صغيرة، إلا ان حجم الأثر الكلي البالغ (0.225) للدراسات التي تستخدم البيانات الحقيقية في الحدود المقبولة وهو لصالح طريقة الأرجحية العظمى.

ولفحص اختلاف معدل حجم الأثر لطريقة التقدير على دقة تقدير القدرة لمجموعات الدراسات المصنفة حسب نوع البيانات المستخدمة في الدراسة تم استخدام اختبار كوكران (Q) لفحص تجانس حجوم الأثر كما هو مبين في الجدول (15).

الجدول (15): فحص تجانس معدل حجم الأثر الكلي لطريقة التقدير على دقة تقدير القدرة باختلاف نوع البيانات.

قيمة اختبار كوكران	درجة الحرية	الدلالة
.377	1	.539

من نتائج اختبار كوكران يتبين تحقق تجانس حجوم الأثر وعدم اختلافها باختلاف نوع البيانات المستخدمة في الدراسة. أي ان حجم أثر طريقة التقدير على دقة تقدير القدرة لم يختلف لمجموعتي الدراسات ذات البيانات الافتراضية (المولدة) والبيانات الحقيقية.

ويمكن تفسير ذلك من خلال الصيغ الرياضية التي يتم استخدامها في تقدير القدرة باستخدام طريقتي التقدير

حيث أن تلك الصيغ ثابته باختلاف حجم العينة وطول الاختبار وحتى نوع البيانات، إلا أن تلك الصيغ تختلف باختلاف النموذج اللوجستي المستخدم في التقدير ولا ننسى أن تقدير القدرة باستخدام النموذج اللوجستي الثلاثي يفترض وجود نسبة من التخمين وهو ما يعطي طريقة الأرجحية أفضلية أكبر من طريقة بييز في حال النموذجين الأحادي والثنائي، فالتقدير باستخدام طريقة الأرجحية العظمى يفترض أن عدد المعالم المقدرة بمثابة مؤشرات لدقة التقدير ومدى الوصول للقيمة الأدق لقدرة الفرد، وتعتمد طريقة بييز على جزء من معادلة الأرجحية العظمى الا أنها تستخدم توزيعات قبلية مفترضة لتعديل احتمالية الاستجابة على فقرات المقياس قد تزود المحلل بتقديرات مظللة وبخاصة في ضوء وجود معالم غير معلمة الصعوبة ذات التوزيع الاعتدالي المشابه لتوزيع الأفراد وبخاصة معلمة التخمين ذات التوزيع الملتوي غالبا.

التوصيات:

في ضوء النتائج التي تم التوصل لها فان الباحث يوصي بما يلي:

توصيات نظرية:

1. إعادة إجراء الدراسة على الدراسات والأبحاث التي تم إجراؤها على البيانات في ضوء نماذج نظرية الاستجابة للفقرة متعددة التدريج.

2. استخدام حجوم الأثر المعدلة التي تم اقتراحها بإضافة ثابت لتعديل صيغة كوهين.

3. دراسة أثر التفاعل بين العوامل المختلفة على دقة تقدير القدرة باستخدام طريقتي التقدير.

4. دراسة أثر توزيع معلمة القدرة على دقة تقديرها باستخدام طريقتي التقدير.

توصيات عملية:

1. استخدام طريقة الأرجحية العظمى لتقدير القدرة في جميع الظروف باستثناء التقدير وفق النموذج الأحادي للاختبارات القصيرة جدًا.

المصادر والمراجع:

البنوي، علاء وزريقات، هيثم. (2022). أثر نسبة القيم المفقودة على دقة التقدير وفق نموذج راش في ضوء طريقتي التقدير بييز والأرجحية العظمى. *مجلة العلوم الإنسانية والطبيعية*، 3(2)، 559-570.

بني عطا، زايد. (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*، 5(2)، 581-606.

حمدان، غسان. (2019). *دراسة مقارنة لطرائق تقدير المعالم في نظرية الاستجابة للمفردة*. أطروحة دكتوراه غير منشورة، جامعة حلب، كلية العلوم، قسم الإحصاء الرياضي، سوريا.

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

سكران، السيد. (2006). ما وراء التحليل كمنهج وصفي تحليلي لتجميع نتائج البحوث وتكاملها في مجال التربية وعلم النفس. *مجلة كلية التربية بالزقازيق*، 53، 1-38.

الشرايدة، سمية والصمادي، عبدالله. (2018). ما وراء التحليل لأثر برامج تنمية مهارات التفكير التباعدي الواردة في أطروحات الماجستير والدكتوراه بجامعة الخليج العربي للفترة (1994-2016). *دراسات، العلوم التربوية*، 45(3)، 49-69.

الشريفين، نضال. (2012). أثر طريقة تقدير معالم الفقرات وقدرات الأفراد على قيم معالم الفقرة والخصائص السيكومترية للاختبار في ضوء تغير حجم العينة. *المجلة التربوية*، 104(1)، 177-238.

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

ضعضع، عبد اللطيف وطومان، منار وطيفور، مصطفى. (2020). أثر حجم العينة وطرائق التقدير في دقة تقدير معالم نموذج راش. *مجلة جرش للبحوث والدراسات*، 21(1)، 131-170.

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

عبابنة، عماد. (2007). مقارنة فاعلية طريقة الأرجحية العظمى وطريقة بييز في تقدير معلمة القدرة عند استخدام النموذج اللوجستي الثلاثي. *مجلة الأكاديمية العربية المفتوحة في الدنمارك*، 2007، 3، 5-22.

عبابنة، عماد. (2009). الاختبارات محكية المرجع: فلسلفتها وأسس تطويرها، ط2، عمان: دار المسيرة للنشر والتوزيع والطباعة.

عبد الحميد، محمد جمال. (2016). *أسلوب ما وراء التحليل لنتائج البحوث والدراسات السابقة*. ط1، عالم الكتب للنشر والتوزيع.

علام، صلاح الدين. (2005). *نماذج الاستجابة للفقرة الاختيارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. ط1، دار الفكر للنشر والتوزيع.

Bangert, R. & Rndner, L. (1991). *Meta-analysis in education research*. Eric Digest, Retrieved from http://www.eric.ed.Gov.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*(4), 431–444. https://doi.org/10.1177/014662168200600405.

Burns, M. & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37*(1), 126–136.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education. (6th ed.)*, Routledge.

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach*. (5th ed), Thousand Oaks, CA: Sage.

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.

Garre, F. & Vermunt, J. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*, 33(1), 43-59.

Gay. L. (2000). *Educational research Competencies for Analysis and Applications*. NJ: prentice Hall.

Hambleton, R. & Swaminatham, H. (1985). *Item Respose theory: principles and application***.** Kluwek NIJ puplishing.

Lewy, B. & Nielsen, A. (2002). Comparison of the frequentist properties of Bayes and the maximum likelihood estimators in an age-structured fish stock assessment model. *Canadian Journal of Fisheries and Aquatic Sciences*, 59, 136–143.

Lord, F. (1986). Maximum Likelihood and Bayesian Parameter Estimation in Item Response Theory. *Journal of Educational Measurement*, 23, 157-162.

Rakkapo, S. & Arayathanitkul, K. (2016). Analysis test of understanding of vectors with the three-parameter logistic model of item response theory and item response curves technique. *Physical Review Physics Education Research*, 12(2), 1-10.