جامعة المنصورة
كلية التربية

الفروق بين طريقة دلتا المعلمية لتقدير الدرجات وطريقة كيرنل اللامعلمية في دقة تقدير مَعْلَمَة القدرة

إعداد

أ/ هند عبد اللّه سليمان الجاسر
درجة الدكتوراه، (مسار القياس والتقويم)، قسم علم النفس
كلية التربية جامعة الملك سعود

د/ إقبال زين العابدين درندري
أستاذ مشارك، قسم علم النفس
كلية التربية جامعة الملك سعود

مجلة كلية التربية – جامعة المنصورة
العدد ١٢٨ – أكتوبر ٢٠٢٤

الفروق بين طريقة دلتا المعلمية لتقدير الدرجات وطريقة كيرنل اللامعلمية في دقة تقدير مَعْلَمَة القدرة

أ / هند عبد الله سليمان الجاسر
درجة الدكتوراه، (مسار القياس والتقويم)، قسم علم النفس
كلية التربية جامعة الملك سعود

د / إقبال زين العابدين درندري
أستاذ مشارك، قسم علم النفس
كلية التربية جامعة الملك سعود

المستخلص:

*هدفت الدراسة الحالية إلى الكشف عن الفروق بين طريقة دلتا المعلمية لتقدير الدرجات وطريقة كيرنل اللامعلمية في دقة تقدير مَعْلَمَة القدرة من خلال استخدام بيانات مُولّدة (بأسلوب المحاكاة)، وفق طريقة مونت كارلو **(MCM)** على (٥٠) فقرة ثنائية التدريج (٠،١). وقد تم الحكم على دقة التقدير باستخدام مؤشر الخطأ المعياري للتقدير **(S.E.E)** . وتوصلت نتائج الدراسة إلى ما يلي: وجود فروق دالة احصائياً عند مستوى الدلالة (α=٠,٠٥) في دقة تقدير معلمة القدرة المقدرة تبعا لحجم العينة وتوزيع البيانات، وعدم وجود فروق دالة احصائياً عند مستوى الدلالة (α=.٠٥) في دقة تقدير معلمة القدرة تبعاً لطريقة تقدير الدرجات (دلتا DSM-L المعلمية وكيرنل KS اللامعلمية)"، وجود فروق دالة احصائياً في التفاعل الثلاثي عند مستوى الدلالة (α=.٠٥) في دقة تقدير معلمة القدرة المقدرة بين كل من طريقة التقدير وحجم العينة وتوزيع البيانات، كما توضح النتائج أن طريقة دلتا توفر باستمرار تقديرات للقدرة أكثر استقرارًا وموثوقية عبر أنواع توزيع مختلفة وأحجام عينات مقارنة بطريقة كيرنل، والتي تميل إلى إنتاج تقديرات ذات تباين أعلى ودقة أقل، وخاصة في التوزيعات الملتوية. وبناء على ما توصلت اليه الرسالة توصي الباحثة باستخدام طريقة (دلتا DSM-L) في تحليل المفردات الاختباري ثنائية الاستجابة وفي التقدير، حيث تبين من نتائج هذه الدراسة أن كلا الطريقتين تعطيان نتائج إحصائية متشابهة، إلا أن طريقة دلتا تظل الأنسب في تفسير النتائج.*

الكلمات المفتاحية: *دقة تقدير مَعْلَمَة القدرة، طريقة دلتا المعلمية، طريقة كيرنل اللامعلمية، مؤشر الخطأ المعياري للتقدير (S.E.E) .*

Abstract:

This study aimed to Differences Between the Parametric Delta Scoring Estimation Method and the Non-Parametric Kernel Method on the Accuracy of Ability Parameter Estimation. The analysis was conducted using simulated data generated through the Monte Carlo method (MCM), consisting of 50 binary items (0,1) with sample sizes of 100, 250, 500, and 1000, and different data distributions (normal, simple positive skew, severe positive skew, simple negative skew, and severe negative skew). To answer the research questions, the DELTA program was used to estimate item and person parameters according to the parametric Delta model, and the R program was employed to estimate parameters using the non-parametric Kernel method (KS). The accuracy of the estimation was assessed using the Standard Error of Estimation (S.E.E). A three-way mixed analysis of variance (ANOVA) was conducted, and the results revealed the following: statistically significant differences were found at the α =.05 level in the accuracy of ability parameter estimation based on sample size and data distribution; no statistically

significant differences were found at the α =.05 level in the accuracy of ability parameter estimation based on the scoring estimation method (Delta DSM-L parametric and Kernel KS non-parametric); statistically significant differences were observed in the three-way interaction (estimation method, sample size, and data distribution) at the α =.05 level in the accuracy of ability parameter estimation,The results showed that the Delta method consistently provided more stable and reliable ability estimates across different data distributions and sample sizes compared to the Kernel method, which tended to produce estimates with higher variance and lower accuracy, especially in skewed distributions. Based on the findings, the researcher recommends using the Delta DSM-L method for analyzing binary-response test items and for estimation purposes, as the results of this study indicate that both methods yield similar statistical outcomes. However, the Delta method remains the more suitable choice for interpreting results.
Keywords: Accuracy of Ability Parameter Estimation, Parametric Delta Method, Non-Parametric Kernel Method, Standard Error of Estimation (S.E.E).

المقدمة:

تُعدّ الاختبارات المعرفية والنفسية من أهم أدوات القياس والتقويم الشائعة، التي تستند في بنائها وتحليلها على نماذج نظريات القياس المختلفة، وتُقدّم أساسًا لعمل تنبؤات حول السمات والقدرات المُقاسة. وقد بُني معظم المقاييس النفسية باستخدام النظرية الكلاسيكية للقياس (Classical Test Theory-CTT)؛ لكن نظرًا لوجود بعض أوجه القصور فيها - حيث إن خصائص الاختبار والفقرات تتأثر بخصائص عينة الأفراد والفقرات- فقد اهتم الباحثون في القياس بتطوير مقاييس تتمتّع بخصائص سيكومترية لا تتأثر بهما؛ وهذا ما أدّى إلى ظهور نظرية الاستجابة للفقرة (Item Response Theory-IRT)؛ إذ وضع رُوّاد هذه النظرية عددًا من النماذج الرياضية (Hambleton & Swaminthan, 1991).

وتنقسم نماذج نظرية الاستجابة للفقرة (IRT) إلى نوعين رئيسين، وهما: النماذج المعلمية (Parametric)، التي يكون فيها شكل الدالة مُحددًا، والنماذج اللامعلمية (NIRT-Nonparametric) ، التي لا تُحدّد شكلًا للدالة على أن تكون غير متناقصة. والنماذج اللامعلمية أقرب إلى دوال الاستجابة الحقيقية من تلك التي تعطيها النماذج المعلمية؛ لأنها تعتمد على افتراضات أقل تشدّدًا. وتُعدّ طريقة تهذيب النواة كيرنل (Kernel Smoothing -KS) اللامعلمية من أفضل طرق تقدير كمية المعلومات، من خلال المنحنيات اللامعلمية مقارنة بغيرها من الطرق المعلمية؛ لبساطة تقديرها وتوافر البرامج الحاسوبية لها، إضافة إلى إمكانية استخدامها لبيانات ثنائية ومتعددة التدريج وذات مستويات قياس اسمية (Nozawa, 2008).

واستكمالًا للجهود المبذولة، وتجنّبًا للتعقيدات الحسابية والأخطاء في التقدير، والمشاكل الفنية في نماذج نظرية الاستجابة للفقرة؛ فقد طوّر دومينقو وديمتروف (Domingue & Dimitrov, 2015) عددًا من النماذج الرياضية الحديثة المستندة على أسلوب دلتا لتقدير الدرجات (Delta Scoring Method-DSM)، وهو أسلوب يُستخدم في تحليل الفقرات ثنائية ومتعددة التدريج، ويعتمد على نمذجة العلاقة القائمة بين استجابات الأفراد على فقرات المقياس والقدرة الكامنة، كما في نماذج نظرية استجابة الفقرة. وفي عام (٢٠١٥) طوّر ديمتروف (Dimitrov) طريقة تقدير الدرجات لتقدير مَعْلَمَة القدرة للفرد ومعالم الفقرات بداية بالاعتماد على الإطار الكلاسيكي لأسلوب دلتا لتقدير الدرجات (Delta Scoring Method- Classical Framework -DSM-C) (Dimitrov, 2016).

وتميّزت الطريقة بالبساطة والوضوح كما في النظرية الكلاسيكية، وأنها تأخذ في الحسبان الصعوبة المتوقّعة للفقرة الاختبارية، إضافة إلى إمكانية تدريج درجات الأفراد وصعوبة الفقرات على مقياس فئوي واحد يُعبّر عن السمة المُقاسة (Dimitrov, 2018; Domingue & Dimitrov, 2015). ثم طُوّرت الطريقة مؤخرًا بالاعتماد على الإطار الكامن لأسلوب دلتا لتقدير الدرجات (Delta Scoring Method- Latent Framework -DSM-L) ليكون أقرب إلى نظرية الاستجابة للفقرة؛ لتوفير خصائصها الجيدة مع تفادي مشكلات التعقيد. حيث تستخدم درجة الفرد (D_w) المقدرة في ضوء الإطار الكلاسيكي لأسلوب دلتا لتقدير الدرجات (DSM-C) كقيمة مبدئية لتقدير معلم قدرة الفرد في الإطار الكامن لأسلوب دلتا لتقدير الدرجات (DSM-L) (Dimitrov & Atanasov, 2021).

وقد تعدّدت نظريات القياس والنماذج والأساليب المختلفة لتقدير قدرات الأفراد ومعالم الفقرات، وأصبحت محط اهتمام الباحثين لتقديم تقييم دقيق؛ لاتخاذ قرارات موضوعية دقيقة. ولأن دقة تقديرات معالم الأفراد والفقرات المُستخلصة من نماذج نظرية القياس الحديثة تتأثر بعدد من العوامل؛ فقد استهدفتها الكثير من الدراسات باستخدام بيانات حقيقية أو مولّدة بالمحاكاة، ودرست مدى تأثير عدد من العوامل مثل: طريقة التقدير، وحجم العينة، وطول أداة القياس، وشكل التوزيع في دقة تقديرات هذه المعالم (Dimitrov & Atanasov, 2021).

مشكلة الدراسة:

تُعدّ نظرية الاستجابة للفقرة (IRT) نظرية بديلة عن النظرية الكلاسيكية (CTT)، التي يُفترض فيها تلافي عيوب النظرية الكلاسيكية، كما أن نماذج هذه النظرية تُمكّن المختص من الإجابة عن أي تساؤل يُثار حول الفقرة أو الاختبار أو المفحوص (الشريفين، ٢٠٠٦).

وهذا ما شجع الباحثة على استخدام نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية، توافقًا مع توصيات سبق أن أشار إليها هامبلتون وسلاتر (Hambleton & Slater, 2011). إذ أوصيا بضرورة إجراء تغيرات أساسية على الاختبارات، كالانتقال من الاعتماد على القياس الكلاسيكي إلى تبني النماذج الحديثة في القياس؛ للتغلّب على عيوب الاختبارات التقليدية. وبالرغم من المميزات التي تمتاز بها نظرية الاستجابة للفقرة (IRT)؛ لكن هناك مشكلة في توفير حجوم العينات وأطوال الاختبارات المطلوبة حسب النموذج المعلمي المُستخدَم، وفي صعوبة توافر الافتراضات التي تتطلّبها نماذج نظرية الاستجابة للفقرة (IRT) المعلمية؛ لذا يلجأ بعض الباحثين إلى النماذج المعلمية المبسّطة والجيدة في الوقت نفسه، مثل: أسلوب دلتا لتقدير الدرجات (DSM) (Dimitrov, 2023)، أو النماذج اللامعلمية البديلة، مثل: طريقة تهذيب النواة كيرنل (KS)، التي لا تحتاج إلى حجوم عينات وأطوال اختبار كبيرة مقارنة بنظيرتها المعلمية. وبيّن رامسي (Ramsay, 1991) أن عددًا قليلًا من الأفراد لا يزيد عن (١٠٠)، وعدد فقرات الاختبار (٢٠)؛ هما المطلوبان لتقدير المنحنى المميز للفقرة (Item Characteristic Curve -ICC)، الذي يُطلق علية دالة الاستجابة للفقرة (Item Response Function-IRF).

وبالرغم من أن هناك عددًا من العوامل التي قد تؤثر في دقة تقديرات معالم هذه النماذج، مثل: طرق تقدير الدرجات المُستخدَمة، وحجم العينة، وطول الاختبار، وتوزيع البيانات؛ لكن كان هناك اختلاف في نتائج الدراسات حول هذه العوامل المؤثرة، كما أنه لم تُجرَ عليها دراسات عربية أو أجنبية – في حدود علم الباحثة- كافية لمعرفة كفاءتها تحت الظروف المختلفة؛ رغم اهتمام الباحثين بدراسة أثر هذه العوامل في ضوء النماذج المختلفة الأخرى، وعدم تأثرها بالثقافة. ومن الدراسات التي اهتمت بذلك دراسات: الشريفين ومناصرة (٢٠١٧)، والشواورة (٢٠١٧)، والقيسي (٢٠١٣)، وموردیرلیك وكوتش (Mor Dirlik & Koç, 2019)، التي قارنت بين نماذج نظرية الاستجابة للفقرة (IRT) المعلمية (Parametric) واللامعلمية (Nonparametric). واستخدم بعض الباحثين أسلوب دلتا لتقدير الدرجات (DSM) المعلمية - سواء الإطار الكلاسيكي (DSM-C)، أو الإطار الكامن (DSM-L)- مثل: ودومينقو وديمتروف (Domingue & Dimitrov, 2015)، وديمتروف

وأتانسوف (Dimitrov & Atanasov, 2021)، وروبيتزش (Robitzsch, 2021)، كما لا توجد دراسة للمقارنة بين طريقة الإطار الكامن لأسلوب دلتا المعلمية لتقدير الدرجات (DSM-L)، وطريقة تهذيب النواة كيرنل (KS) اللامعلمية؛ من حيث أثر تلك العوامل في دقة تقدير المعلمات للمفردات ثنائية التدريج، فكلا النموذجين مبني على معالجة مشكلات في نماذج الاستجابة للفقرة المعلمية. ويكثر استخدامهما في الفقرات ثنائية التدريج، مع أهمية مَعْلَمَة القدرة، حيث يمكن من خلالها التنبؤ بأداء الفرد على الفقرة الاختبارية، كما أنها تُقدّم مؤشرات مهمة تكون لبنة أساسية في عملية صنع القرار وإصدار الأحكام حول الأفراد، هذا بالإضافة إلى قلة الدراسات التي أُجريت على توزيع البيانات على تأثير دقة التقدير في هذه النماذج.

وقد استخدمت الباحثة أسلوب المحاكاة وفق طريقة مونت كارلو (Monte Carlo Methods-MCM)، التي تُساعد على الوصول إلى النتائج المثالية، حيث يتم ضبط جميع المُتغيّرات، كما أناه تُوفر الوقت والجهد على الباحث؛ ومن هنا تبرز الحاجة إلى استكشاف أثر طرق تقدير الدرجات المُستخدَمة، وحجم العينة، وتوزيع البيانات باستخدام هذه النماذج في دقة تقدير المعلمات، خاصة مَعْلَمَة القدرة؛ لأهميتها لبقية التقديرات، باستخدام هذه النماذج مع استخدام بيانات المحاكاة؛ لتوفير المعلومات التي تساعد الباحثين على توظيفها لتقدير السمات الكامنة للأفراد بأعلى كفاءة. وتتلخّص مشكلة البحث في السؤال الأتي:

١. هل تختلف دقة تقدير مَعْلَمَة القدرة المُقدّرة باستخدام فقرات ثنائية التدريج تبعًا لطريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، وحجم العينة (١٠٠- ٢٥٠- ٥٠٠- ١٠٠٠)، وتوزيع البيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، والتفاعل بينها؟

أهداف الدراسة: هدفت الدراسة الحالية إلى ما يأتي:

١. المقارنة بين دقة تقدير مَعْلَمَة القدرة المُقدّرة باستخدام فقرات ثنائية التدريج تبعًا لطريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، وحجم العينة (١٠٠- ٢٥٠- ٥٠٠- ١٠٠٠)، وتوزيع البيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، والتفاعل بينها.

أهمية الدراسة:

الأهمية النظرية:

١. تتكامل هذه الدراسة مع الدراسات وجهود الباحثين في المقارنة بين نماذج نظريات القياس؛ للوصول إلى بناء معرفي مُتكامل في البحوث.
٢. تُسهم هذه الدراسة في تدعيم جوانب التوافق والاختلاف بين أسلوبين بديلين في نظرية الاستجابة للفقرة (IRT)، وهما: أسلوب دلتا المعلمي لتقدير الدرجات حسب الإطار الكامن (DSM-L)، وطريقة تهذيب النواة كيرنل (KS) اللامعلمية.
٣. تحاول الدراسة الكشف عن أثر حجم العينة وشكل التوزيع في دقة تقدير مَعْلَمَة القدرة باستخدام نماذج مختلفة، حيث هناك قلة في الدراسات – في حدود علم الباحثة- التي درست تأثير توزيع البيانات في دقة التقدير.
٤. تُوجّه نتائج هذه الدراسة اهتمام الباحثين في مجال القياس والتقويم النفسي والتربوي نحو تناول أثر عدد من المُتغيّرات المختلفة في الخصائص السيكومترية، باستخدام نماذج نظرية الاستجابة للفقرة (IRT)، وطرق تقدير الدرجات المعلمية واللامعلمية، ونماذج دلتا لتقدير الدرجات (DSM)، والمقارنة بينها.

الأهمية التطبيقية:

١. تُوفّر نتائج هذه الدراسة معلومات تساعد الباحثين على توظيفها لتقدير السمات الكامنة للأفراد بأعلى كفاءة ممكنة باستخدام النماذج المستعملة في الدراسة الحالية.

٢. تُساعد نتائج هذه الدراسة المهتمين والباحثين على اختيار الطريقة المناسبة، حيث تشمل استعراض طريقتين مفصّلتين - حسب توزيع البيانات وحجم العينة- وهما: طريقة دلتا المعلمية لتقدير الدرجات حسب الإطار الكامن (DSM-L)، وطريقة تهذيب النواة كيرنل (KS) اللامعلمية؛ بهدف أن تكون إطارًا مرجعيًّا.
٣. تُقدم الدراسة معلومات وإرشادات تساعد مُطوري الاختبارات النفسية والتربوية في الحصول على تقديرات دقيقة لمعالم الفقرة باختلاف طرق التقدير، وأحجام العينة، وشكل توزيع القدرة.

حدود الدراسة:

١. اقتصرت الدراسة على استخدام بيانات مُولّدة (بأسلوب المحاكاة) (Simulation)، وفق طريقة مونت كارلو (MCM) ، وهذه الطريقة تُوفّر بيانات ذات ضبط عالٍ للمُتغيّرات؛ لذا استخدمت الباحثة أسلوب المحاكاة في توليد البيانات لـ(٥٠) فقرة ثنائية التدريج (٠،١)؛ حيث أوصى لورد (Lord, 1980) أن يكون طول الاختبار (٥٠) فقرة، ولأن يكون عدد الأفراد (١٠٠٠)؛ للحصول على أفضل التقديرات.
٢. استخدمت الباحثة الأحجام (١٠٠- ٢٥٠- ٥٠٠- ١٠٠٠)؛ لدراسة أثر أحجام العينات في دقة التقدير باستخدام نموذج الاستجابة للفقرة ثنائي المعلمة (2PLM). وقد أوضحت دراسة دي إيالا (٢٠٠٩/٢٠١٧) ارتباط حجم العينة وعدد الفقرات بالنموذج المُستخدَم، حيث إن النموذج المُستخدَم في الدراسة يتناسب مع أحجام عينات أكبر من (٢٠٠)، وأقل أو يساوي (١٠٠٠).
٣. اعتمدت الباحثة شكل توزيعات محددة للبيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، باستخدام برنامج (WinGen v1.4) (Han, 2007).
٤. اعتمدت الباحثة طريقتين للتقدير، وهما: طريقة دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية.

مصطلحات الدراسة:

دقة التقدير (Accuracy of Estimation):

التعريف العلمي: الدقة هي الدرجة التي يتفق فيها القياس الفعلي مع القيمة المُقدّرة؛ وبالتالي فإن مؤشر دقة التقدير هو درجة الاتفاق بين قيم التقديرات الملاحظة والمُتوقّعة (Guo,2006).

التعريف الإجرائي: هو مقدار الفرق بين قيم معالم القدرة المُقدّرة حسب طريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، والقيمة الفعلية (المولّدة) للقدرة. وفي الدراسة الحالية حُسب المتوسط الحسابي، والانحراف المعياري، والخطأ المعياري للتقدير(Standard Error of Estimate-S.E.E) .

طريقة تهذيب النواة كيرنل (Kernel Smoothing -KS):

التعريف العلمي: هي طريقة تقدير لامعلمية تقوم على تقدير كمية المعلومات باستخدام الانحدار اللامعلمي من خلال المنحنيات والرسوم البيانية (Nozawa, 2008).

التعريف الإجرائي: هي الأسلوب المُستخدَم لتقدير القدرة باستخدام نماذج استجابة الفقرة اللامعلمية في هذه الدراسة، باستخدام برنامج تستجراف (Test Graf) وبرنامج (R).

أسلوب دلتا لتقدير الدرجات (Delta Scoring Models-DSM) :

التعريف العلمي: هي عدد من النماذج الاحتمالية وضعها ديمتروف (Dimitrov)، بحيث يُمكن استخدامها لتحليل البيانات الثنائية التدريج، أو البيانات الترتيبية متعددة التدريج (2018; Dimitrov, Domingue & Dimitrov,2015) .

التعريف الإجرائي: هو نموذج الدالة النسبية ثنائي المعلم، بحيث تُقدّر معالم النموذج وفق الإطار الكامن لأسلوب دلتا لتقدير الدرجات (DSM-L).

الدراسات السابقة:سوف يتم عرض الدراسات السابقة ضمن ثلاث محاول كالتالي:

المحور الأول: الدراسات التي أجريت حول دقة تقدير معالم الفقرات والأفراد في نماذج نظرية الإستجابة للفقرة (IRT)- النماذج المعلمية(PIRT) والنماذج اللامعل (NIRT)- وفق عدد من المُتغيّرات:

دراسة بني عطا والشريفين (٢٠١٢)، التي هدفت إلى التحقق من أثر اختلاف شكل توزيع القدرة في معالم الفقرة ودالة المعلومات للاختبار. ولتحقيق الغرض من الدراسة وُلّدت استجابات أربعة أشكال من توزيعات القدرة وفق النموذج اللوجستي ثلاثي المعلمة، و(٦٠) فقرة ثنائية الاستجابة. وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات معالم الفقرات، وعلى الرغم من تباين معالم الفقرات؛ فقد أظهرت النتائج أن قيم مُعامِلات الارتباط بين تقديرات معالم الفقرات المعايرة تبعًا لشكل توزيع القدرة؛ كانت دالة إحصائيًّا بوصفها مؤشرًا على استقرار معالم الفقرات عبر التوزيعات المختلفة للقدرة.

وهدفت دراسة الشريفين (٢٠١٢) إلى الكشف عن أثر طريقة تقدير معالم الفقرة وقدرات الأفراد في قيم معالم الفقرة والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. ولتحقيق هدف الدراسة بُني اختبار تحصيلي في الفيزياء من نوع الاختيار من أربعة بدائل، تكوّن من (٣٣) فقرة، وطُبّق الاختبار على عينة مكوّنة من (١٠٠٠) طالب وطالبة من طلبة الصف الثاني الثانوي العلمي. وأظهرت النتائج وجود فروق ذات دلالة إحصائية في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تُعزى إلى مُتغيّر حجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تُعزى إلى طريقة التقدير. كما أشارت النتائج إلى أن دقة تقديرات مَعْلَمَة القدرة تزداد في حالة عينة الأفراد ذوي القدرة العالية، وعينة الأفراد ذوي القدرة المُتدنية عند استخدام طريقة توقع التوزيع البعدي (EAP)، في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسطة باستخدام طريقة الأرجحية العظمى (MLE)؛ بغض النظر عن حجم العينة.

أما دراسة شما (٢٠١٣) فهدفت إلى التّعرّف على تأثير حجم العينة في دقة تقدير صعوبة الفقرات وقدرة الأفراد باستخدام نموذج راش. ولتحقيق الهدف من الدراسة طُبّق اختبار أوتيس - لينون للقدرة العقلية على ثلاث عينات مختلفة من حيث الحجم تكوّنت من: (٢٠٠- ٨٠٠- ١٥٠٠) من طلبة المرحلة الأساسية. وأظهرت نتائج الدراسة أن دقة تقديرات صعوبة الفقرات تزداد مع زيادة حجم العينة، كما أن اختلاف حجم العينة ليس له تأثير في دقة تقدير قدرات الأفراد.

وهدفت دراسة بني عطا (٢٠١٤) إلى الكشف عن دقة تقدير النموذج اللوجستي الثلاثي المعلم لمعالم الفقرة والقدرة، من خلال تغير طول الاختبار وحجم العينة. وقد اعتمدت الدراسة على توليد بيانات محاكاة ثنائية الاستجابة لستة من أطوال الاختبار (١٠-٢٥-٥٠-٧٥-١٠٠) فقرة، ووُلّدت بيانات لستة أحجام مختلفة من العينة (١٠٠-٢٥٠-٥٠٠-١٠٠٠). وأظهرت نتائج الدراسة تأثيرًا دالًّا إحصائيًّا لمُتغيّر طول الاختبار ومُتغيّر حجم العينة والتفاعل بينهما في دقة تقديرات معالم الفقرات والأفراد.

وهدفت دراسة الحواري (٢٠١٥) إلى الكشف عن أثر طول الاختبار وشكل توزيع القدرة في تقديرات القدرة للأفراد، وتقديرات معالم الصعوبة للفقرات، واقتران المعلومات للفقرات، وفق نموذج راش في نظرية استجابة الفقرة. ومن ضمن إجراءات الدراسة: تم وليد البيانات، ومجموعة الفقرات، وقدرات الأفراد؛ إذ وُلّدت بيانات بواقع (١٠٠٠) مفحوص، بطول الاختبار (٣٠،٦٠) فقرة، وشكل توزيعات القدرة (ملتوٍ التواء سالبًا- طبيعي، ملتوٍ التواء موجبًا)، وفق النموذج الأحادي المعلمة. وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقديرات معالم قدرة الأفراد تُعزى إلى شكل توزيع القدرة وطول الاختبار، حيث كانت تقديرات قدرة الأفراد أكثر دقة لصالح شكل الالتواء السالب والموجب.

بينما هدفت دراسة الشريفين ومناصرة (٢٠١٧) إلى تقدير خصائص توزيع قدرات الأفراد ومعالم فقرات اختبار وفق نماذج نظرية الاستجابة للفقرة المعلمية (PIRT) واللامعلمية (NIRT)، وإجراء مقارنة بين النماذج المُستخدَمة: النموذج الثلاثي المعلمة، والنموذج الرباعي المعلمة، ونموذج موكن اللامعلمي (Mokken). ولتحقيق هدف الدراسة حُصل على بيانات اختبار ضبط نوعية التعليم في مادة العلوم العامة للصف الرابع الأساسي، الذي تألف من (٢٥) فقرة من نوع الاختيار من متعدد، وتكوّنت عينة الدراسة من (١٠١٠) طلاب وطالبات من طلبة الصف الرابع الأساسي، ممن طُبّق عليهم الاختبار في المدارس الحكومية والخاصة الأردنية للعام الدراسي (٢٠١٢-٢٠١٣). وأظهرت النتائج وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (a=.05) بين تقديرات الخطأ المعياري في تقديرات القدرة بين النموذج المعلمي الرباعي والنموذج المعلمي الثلاثي؛ لصالح النموذج المعلمي الرباعي، وبين النموذج اللامعلمي الثلاثي والنموذج المعلمي الثلاثي؛ لصالح النموذج اللامعلمي الثلاثي.

وهدفت دراسة الخرشة (٢٠١٨) إلى الكشف عن أثر حجم العينة، واختلاف شكل توزيع القدرة (طبيعي ملتوٍ نحو اليمين- ملتوٍ نحو اليسار) في معالم الفقرة. ولتحقيق الغرض من الدراسة وُلّدت استجابات تسع مجموعات من الأفراد وفق النموذج اللوجستي ثلاثي المعلمة و(٥٠) فقرة ثنائية الاستجابة. وكشفت نتائج تحليل التباين الثنائي عن وجود فروق ذات دلالة إحصائية عن مستوى دلالة بين متوسطات معالم الفقرة للفقرات تبعًا لحجم العينة، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات معالم الفقرة تبعًا لشكل التوزيع.

وهدفت دراسة موردیرلیك وكوتش (Mor Dirlik & Koç, 2019) إلى مقارنة معلمات القدرة المقدرة في النماذج المعلمية واللامعلمية (NIRT)، وفقًا لعوامل مختلفة منها: طول الاختبار، وحجم العينة، والصفات السيكومترية للفقرات. واُستخدم اختبار الرياضيات للصف الثامن في الدراسة الدولية تيمس (TIMSS 2011) لعينة مكوّنة من (٧٢٥٤) طالبًا وطالبة من دول مختلفة. وبالنسبة لحجم العينة فكان (٥٠٠- ١٠٠٠- ٣٠٠٠)، وكان طول الاختبار (٥- ١٥- ٢٥) فقرة. وأظهرت النتائج أن جميع معلمات الفقرة المُقدّرة وفقًا للنموذجين المعلمي (PIRT) واللامعلمي(NIRT) مرتبطة ارتباطًا عاليًا. أما بالنسبة لمعلمات القدرة؛ فقد ظهر في النماذج المعلمية (PIRT) عدم الاتساق في التقديرات - خاصة عند (٥ و١٥) فقرة- بينما القدرات المُقدّرة في حال (٢٥) فقرة؛ فكانت مرتبطة بشكل أكبر. وفي حال النماذج اللامعلمية (NIRT)، أظهرت جميع أحجام العينات وأطوال الاختبار ارتباطًا عاليًا، كما أظهرت النتائج الارتباط المرتفع بين النماذج المعلمية (PIRT) واللامعلمية (NIRT) فقط في حال حجم العينة (١٠٠٠) والطول (٢٥) فقرة، وكلما زاد طول الاختبار وحجم العينة؛ كانت التقديرات أكثر اتساقًا.

في حين هدفت دراسة ضعضع وآخرين (٢٠٢٠) إلى الكشف عن أثر طرائق التقدير الأرجحية العظمى، وتقدير بييز، وطريقة بروكس، باستخدام عينات ذات أحجام مختلفة (٥٠٠- ١٠٠٠- ١٥٠٠- ٢٠٠٠) على دقة تقدير معالم الفقرة والأفراد باستخدام نموذج راش، بالاعتماد على الخطأ المعياري للتقدير(SEE). ولتحقيق هدف الدراسة وُلّدت استجابات لـ(٢٠٠٠) مفحوص لاختبار مكوّن من (٤٠) مفردة ثنائية الاستجابة، بتوزيع طبيعي معياري لمعلم للقدرة والصعوبة، وأُخذت عينات عشوائية جزئية منها (٥٠٠- ١٠٠٠- ١٥٠٠). ومن بين النتائج التي توصّلت إليها الدراسة: أنه توجد فروق في دقة تقدير معلم الصعوبة وقدرة الأفراد تُعزى إلى حجم العينة وطريقة التقدير والتفاعل بينهما، وتزايد دقة تقدير معلم الصعوبة وقدرة الأفراد بازدياد حجم العينة.

أما دراسة البادية وآخرين(٢٠١٨) فهدفت إلى الكشف عن أثر حجم العينة في دقة تقدير خصائص الفقرة والقدرة في اختبار التنمية المعرفية باستخدام نموذج راش، وتكوّنت عينة الدراسة من طلبة الصف السابع بمحافظة شمال الباطنة، الذين خضعوا لاختبار التنمية المعرفية حيث بلغ عددهم (٨٤٨٤) طالبًا وطالبة. ومن هذه العينة اُستخدمت (١٢) عينة بأحجام مختلفة سُحبت بطريقة عشوائية، حيث سُحبت أربع عينات مختلفة (٢٠٠- ٥٠٠- ١٠٠٠- ١٥٠٠)، وفي كل عينة من

العينات الأربع سُحبت ثلاث مختلفة بالحجم نفسه؛ ليصبح مجموع العينات (١٢)، واُستخدمت طريقة الأرجحية العظمى الهامشية (MML) في التقدير. وللحكم على دقة التقدير اُعتمد على الجذر التربيعي لمتوسط مربعات الخطأ (RMSE) بين المعالم المُقدّرة للعينة الكلية، والمعالم المُقدّرة من العينات (١٢) المسحوبة المختلفة. وتوصّلت الدراسة إلى وجود أثر لحجم العينة، حيث تزداد دقة تقديرات معلم الصعوبة، وقدرة الأفراد، ودالة معلومات الاختبار بزيادة حجم العينة. وأن حجم عينة (٥٠٠) فرد كافٍ ليعطي تقديرات للقدرة وصعوبة الفقرة بشكل دقيق باستخدام نموذج راش.

وهدفت دراسة علوان وجاسم (Alwan & Jasim, 2022) إلى الكشف عن أثر الاختلاف في توزيع مستوى القدرة الملتوي موجبًا لمعالم فقرات اختبار القدرة العقلية وفقًا لنظرية الاستجابة للفقرة، وطُبّق الاختبار على عينة مُكوّنة من (١٠٠٠) طالب وطالبة، بالاعتماد على نموذج المعلم ثلاثي الاتجاهات، وتألف الاختبار في صورته النهائية من (٧٨) فقرة. ولتحقيق أهداف البحث وُلّدت البيانات باستخدام (المحاكاة)؛ للحصول على شكل التوزيع الملتوي الموجب، وأظهرت النتائج وجود دلالة إحصائية في الصعوبة والتمييز والتقدير للفقرات وفقًا للشكل المُتغيّر للتوزيع الملتوي الموجب.

أما دراسة الشافعي (٢٠٢٤) فهدفت إلى الكشف عن تأثير حجم العينة في تقديرات صعوبات الفقرات وقدرات الأفراد باستخدام تحليلات النموذج اللوجيستي ثلاثي المعلم، وقد بلغ حجم عينة الدراسة (١٨٠٠) فرد، وطبّق الباحث اختبارًا لمادة الجبر للصف الثاني الثانوي مكوّن من (٤٠) مفردة اختبارية من نوع الاختيار من متعدد، على أربعة عينات ذات أحجام مختلفة (٣٠٠- ٦٠٠-٩٠٠- ١٨٠٠) فرد، واُستخدمت أخطاء القياس مؤشرًا لدقة القياس، كما استخدمت أسلوب تحليل تباين القياس المتعدد؛ للكشف عن الفروق الدالة بين تقديرات القدرة المتناظرة الناتجة عن تحليل استجابات أحجام مختلفة من العينة. وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية عند تقدير صعوبات الفقرات الاختبارية لاختلاف حجم العينة، كما تبيّن أن دقة تقديرات قدرات الأفراد تزداد بزيادة حجم العينة.

المحور الثاني: الدراسات التي قارنت بين طريقة التقدير اللامعلمية (NIRT) ***تهذيب النواة (كيرنل KS) وطرق أخرى:***

وهدفت دراسة القيسي (٢٠١٣) إلى الكشف عن دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة المعلمية واللامعلمية باختلاف حجم العينة وطول الاختبار، وذلك عن طريق مقارنة طريقة الأرجحية العظمى الهامشية المعلمية (MML) وطريقة تهذيب النواة اللامعلمية (كيرنل KS)، باختلاف حجم العينة (١٠٠- ٢٥٠- ٥٠٠- ١٠٠٠) فرد، وعدد فقرات الاختبار (٢٠- ٤٠- ٦٠). ولتقدير معالم الفقرة والقدرة اُستخدمت طريقة الأرجحية العظمى الهامشية(MML) المعلمية، واُستخدمت برمجية تستجراف (TESTGRAF) لتقدير معالم الفقرة والقدرة باستخدام طريقة تهذيب النواة اللامعلمية (كيرنل KS). وأظهرت النتائج أن المعالم المقدرة بطريقة التقدير المعلمية (PIRT)؛ كانت أفضل من المعالم المُقدرة وفق طريقة التقدير اللامعلمية (NIRT).

وأجرى القيسي (٢٠١٦) دراسة هدفت إلى الكشف عن أثر حجم العينة وطول الاختبار في دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية الاستجابة للفقرة اللامعلمية (NIRT)، اعتمادًا على مؤشري دقة القياس: (التحيز BIAS)، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE)، وباستخدام بيانات مولّدة، واستخدام طريقة تهذيب النواة (كيرنل KS) اللامعلمية (NIRT) لتقدير معالم الفقرة والقدرة. وأظهرت نتائج الدراسة وجود فروق دالة إحصائيًّا (α=.٠٥) في متوسطات مؤشر دقة التقدير في مَعْلَمَة التمييز (a) تُعزى إلى مُتغيّر حجم العينة، وفي مَعْلَمَة التخمين (c) تُعزى إلى متغيّري الدراسة (حجم العينة، وطول الاختبار) والتفاعل بينهما، ودقة تقدير مَعْلَمَة القدرة (θ) تُعزى لمتغيري الدراسة (حجم العينة، وطول الاختبار)، في حين لم تظهر فروق دالة إحصائيًّا (α=.٠٥) في متوسطات مؤشر دقة التقدير في مَعْلَمَة الصعوبة (b).

كما أجرى الشواورة (٢٠١٧) دراسة هدفت إلى التّعرُّف على الفاعلية النسبية لطريقة بييز(Bayes) المعلمية، وطريقة تهذيب النواة كيرنل (KS) اللامعلمية (NIRT) في تقدير مَعْلَمَة القدرة وفق النموذج اللوجستي الثنائي باختلاف حجم العينة، واستخدمت الدراسة البيانات المولدة. وأظهرت نتائج الدراسة وجود فرق دال إحصائيًا عند مستوى الدلالة (α=.٠٥) بين قيمتي مُعامِل الارتباط بين مَعْلَمَة القدرة الحقيقية ومَعْلَمَة القدرة المُقدّرة وفق طريقة التقدير بييز المعلمية من جهة، ومُعامِل الارتباط بين مَعْلَمَة القدرة الحقيقية ومَعْلَمَة القدرة المُقدّرة وفق طريقة التقدير كيرنل (KS) اللامعلمية (NIRT) من جهة أخرى؛ لصالح معامل الارتباط بين مَعْلَمَة القدرة الحقيقية ومَعْلَمَة القدرة المُقدّرة وفق طريقة التقدير بييز المعلمية، عندما كان حجم العينة (٢٥٠ و٥٠٠)، بينما لم يثبت وجود فرق دال إحصائيًّا عند حجم العينة (١٥٠٠) فرد.

وهدفت دراسة أجراها عفتبانة وبغائي (Effatpanah & Baghaei, 2022) إلى الكشف عن جودة تقييم الفاحصين باستخدام منحنى خصائص الفقرة اللامعلمي، واستخدمت الدراسة طريقة تقدير منحنى خاصية الفقرة اللامعلمي لنمذجة استجابات الأفراد وفحصها، وطُبّقت على عينة مُكوّنة من (٢١٧) طالبًا في الجامعة الإسلامية بإيران، منهم (٨٧) ذكورًا و(١٣٠) إناثًا في اختبار الكتابة (٢٥٠) كلمة باللغة الإنجليزية كلغة ثانية عن موضوع: دور التقدم التكنولوجي الحديث في البطالة. واُستخدم معيار تقييم تحليلي يقيس أربعة معايير للكتابة على مقياس مُتدرّج مُكوّن من خمس نقاط، واُستخدمت طريقة تهذيب النواة كيرنل (KS) اللامعلمية (NIRT)؛ بهدف الكشف عن مدى استقرار تقديرات المنحنى في حالة تمثيلها بيانيًّا. وأظهرت نتائج المنحنيات أن التقدير اللامعلمي لمنحنيات الخصائص المميزة للفقرة باستخدام طريقة كيرنل (KS)؛ يمكن أن تُحدّد تأثيرات التقييم المختلفة، وتُوفّر معلومات تشخيصية قيّمة لفحص جودة التصنيف واستكشاف أنماط التصنيف.

المحور الثالث: الدراسات التي تناولت أسلوب دلتا لتقدير الدرجات في تقدير معالم الفقرة والفرد ومقارنته بنماذج أخرى:

هدفت دراسة دومينقو وديمتروف (Domingue &Dimitrov, 2015) إلى تناول مدى اتساق قيم معالم قدرة الأفراد المُقدّرة بأسلوب الدلتا لتقدير الدرجات، والمُتمثّلة في (D-Scores) مع مُسلّمات القياس الموحّد في المركز الوطني للقياس (The National Center for Assessment -NCA). وقد استندت تحليلات دراستهما على استجابات (٤٥٦١٢) من المعلمين السعوديين في اختبار المعلمين (Teacher Test)، الذي يُقدّمه المركز الوطني للقياس والمُكوّن من (٧٩) فقرة. وبرهنت النتائج أن قيم معلم قدرة الأفراد المُقدّرة بأسلوب الدلتا (D-Scores)؛ تنتج انتهاكات أقل لمُسلّمات القياس الموحّد من نظيرتها ثيتا (Scores θ) المُقدّرة في ضوء النموذج اللوجستي الثلاثي المعلم (Three-Parameter Logistic Model-3PLM)؛ الرغم من وجود ارتباط عالٍ بين قيم معالم القدرة المُقدّرة، حيث بلغت قيمة معامل الارتباط بين قيم (θ) و (D) (٩١.).

وبحث ديمتروف وأتانسوف (Dimitrov & Atanasov, 2021) دقة تقديرات معالم الفقرات والأفراد في ضوء نموذج الدالة النسبة ثنائي المعلم (Two-Parameter -RFM2 Rational Function Model) وفق الإطار الكلاسيكي لأسلوب دلتا لتقدير الدرجات (-DSM C)، والإطار الكامن لأسلوب دلتا لتقدير الدرجات (DSM-L). ولاستخلاص النتائج اُستخدمت بيانات حقيقية وبيانات محاكاة، وتكوّنت عينة البيانات الحقيقية من استجابات (١٢٨٨) فردًا في اختبار كفايات اللغة الإنجليزية ستيب (- Student Training in Engineering Program STEP) المُقدّم من المركز الوطني للقياس في السعودية، والمُكوّن من (١٠٠) فقرة ثنائية التدريج. وأسفرت نتائج تحليل البيانات عن وجود ارتباط عالٍ بين معالم الفقرات والأفراد المُقدّرة بنموذج الدالة النسبة ثنائي المعلمة (RFM2) وفق الإطار الكامن (DSM-L)، والمعالم المُقدّرة بنموذجي الدالة النسبة ثنائي المعلمة (RFM2) وفق الإطار الكلاسيكي لأسلوب دلتا لتقدير الدرجات (-DSM C)، والنموذج اللوجستي ثنائي المعلمة(2PLM) وفق نظرية الاستجابة للفقرة. ودُرس أثر مُتغيّري: حجم العينة وطول الاختبار في دقة تقديرات المعالم، حيث وُلّدت استجابات لثلاث عينات

بحجم (٥٠٠- ١٠٠٠- ٣٠٠٠)، وأربعة اختبارات بأطوال (٢٠- ٤٠- ٦٠- ٨٠) فقرة. وأظهرت نتائج تحليل البيانات أن تقدير معالم الفقرة والفرد وفق الإطار الكامن (DSM-L)؛ أكثر دقة وارتباطًا بمثيلاتها المُقدّرة وفق نظرية الاستجابة للفقرة من التقديرات وفق الإطار الكلاسيكي (DSM-C) في جميع الظروف الاختبارية المُحدّدة.

وفي دراسة أجراها روبيتزش (Robitzsch, 2021) هدفت إلى بحث مدى التكافؤ بين النموذج الكامن لتقدير الدرجات دلتا (DSM-L) المُقترح من ديميتروف، ونموذج الاستجابة للفقرة اللوجستي ثنائي المعلمة (2PLM) ويُعادله إحصائيًّا. واُستخدمت بيانات برنامج التقييم الدولي للطلاب بيزا في ألمانيا (Program for International Student -PISA 2006 Assessment)، المُكوّن من (٢٨) فقرة، مع عينة أُخذت من (٢٦) دولة مختلفة. وأظهرت النتائج أنه بعد إجراء عدد كبير من التحويلات الخطية في النموذجين (DSM-L) وَ (2PLM)؛ فإنهما يعطيان نتائج إحصائية متشابهة؛ لكن يظل نموذج ديميتروف الأنسب في تفسير النتائج.

التعقيب على الدراسات السابقة:

من خلال العرض السابق للدراسات السابقة؛ لاحظت الباحثة كيف برزت جهود الباحثين في الكشف عن تباينات القيم المُقدّرة للقدرة ومعالم الفقرة حسب النماذج المُستخدَمة وعدد من المُتغيّرات. وقد تناولت أغلبية الدراسات بشكل مركّز مُتغيّر حجم العينة، بالإضافة إلى مُتغيّر طول الاختبار، واستخدمت بعض الدراسات أثر مُتغيّر شكل توزيع البيانات في دقة تقدير معالم الفرد والفقرات، سواء كانت بيانات مولّدة (محاكاة) أو حقيقية. تتفق الدراسة الحالية مع دراسات: بني عطا (٢٠١٤)، وبني عطا والشريفين (٢٠١٢)، والحواري (٢٠١٥)، والخرشة (٢٠١٨)، والشافعي (٢٠٢٤)، وشما (٢٠١٣)، وعلوان وجاسم (Alwan & Jasim, 2022)، في المقارنة بين النماذج المعلمية (PIRT) والنماذج اللامعلمية (NIRT). وخصّت هذه الدراسة طريقة تهذيب النواة (كيرنل KS) من النماذج اللامعلمية (NIRT)، حيث اتفقت مع دراسات: الشواورة (٢٠١٧)، وعفتبانة وبغائي (Effatpanah & Baghaei, 2022)، والقيسي (٢٠١٣،٢٠١٦)، اتفقت الدراسة الحالية مع الدراسات التي استخدمت طريقة دلتا لتقدير الدرجات، مثل دراسات: ودومينقو وديمتروف (Domingue & Dimitrov, 2015)، وديمتروف وأتانسوف (Dimitrov & Atanasov, 2021)، وروبيتزش(Robitzsch, 2021).

اتفقت الدراسة الحالية في الكشف عن أثر حجم العينة في دقة تقرير معالم الفقرة والقدرة مع دراسات: البادية وآخرون (٢٠٢٠)، وبني عطا (٢٠١٤)، وديمتروف وأتانسوف (Dimitrov & Atanasov, 2021)، والشافعي (٢٠٢٤)، وشما (٢٠١٣)، والشواورة (٢٠١٧)، والقيسي (٢٠١٣).

- كما اتفقت مع بعض الدراسات التي جمعت بين مُتغيّري: حجم العينة وتوزيع البيانات مثل دراستي: الخرشة (٢٠١٨)، واتفقت مع الدراسات التي جمعت بين أثر حجم العينة وطرق التقدير، مثل دراستي: الشريفين (٢٠١٢)، وضعضع وآخرون (٢٠٢٠).

خلصت الدراسات السابقة مثل: البادية وآخرون (٢٠٢٠)، وبني عطا (٢٠١٤)، والشافعي (٢٠٢٤)، وديمتروف وأتانسوف (Dimitrov & Atanasov, 2021)، والقيسي (٢٠١٦) إلى وجود أثر دال لحجم العينة في دقة التقدير، ودراسة الشواورة (٢٠١٧)، التي توصّلت إلى أن اختلاف الحجم كان له أثر. باستثناء دراسة شما (٢٠١٣)، التي توصّلت إلى أن اختلاف حجم العينة ليس له تأثير في دقة تقدير قدرات الأفراد، كما خلصت دراسة ضعضع (٢٠٢٠) إلى وجود فروق دالة إحصائيًّا لحجم العينة وطريقة التقدير والتفاعل بينهما على دقة تقدير معالم الأفراد، وخلصت دراسة الشريفين (٢٠١٢) إلى وجود فروق دالة إحصائيًّا لحجم العينة، ولا توجد دلالة إحصائية لطريقة التقدير، بينما توجد دلالة إحصائية للتفاعل بين حجم العينة وطريقة التقدير على دقة تقدير معالم الأفراد. كما خلصت دراسات: الحواري (٢٠١٥)، والشريفين ومناصرة (٢٠١٧)، وبني عطا

والشريفين (٢٠١٢)، وعلوان وجاسم (Alwan & Jasim, 2022) إلى وجود دلالة إحصائية لتوزيع البيانات على دقة تقدير معالم الأفراد.

ومما سبق؛ يتبيّن أن الدراسة الحالية تميّزت في استخدامها لنماذج نظرية الاستجابة للفقرة اللامعلمية (NIRT)، التي تفترض أن البيانات تقع على مستوى القياس الرتبي، وتتطلّب افتراضات أقل تشددًا من نظيرتها المعلمية (PIRT)؛ لذلك فإنها تُتيح المجال لتوظيف هذه النظرية في تقدير خصائص الفقرات وقدرات الأفراد، واستخدام الأساليب والطرق الواردة في الدراسة.

كما تتميّز الدراسة الحالية بمحاولة التوصّل إلى الفروق بين الإطار الكامن لأسلوب دلتا المعلمي لتقدير الدرجات (DSM-L)، وطريقة كيرنل لتقدير الدرجات (KS) اللامعلمية (NIRT) في دقة تقدير مَعْلَمَة القدرة والكفاءة النسبية لهما، بينما بحثت معظم الدراسات السابقة عن الفروق بين نماذج نظرية الاستجابة للفقرة (IRT) المعلمية (PIRT) واللامعلمية (NIRT).

كما تتميّز الدراسة الحالية باستخدام بيانات مولّدة (المحاكاة)، وفق طريقة مونت كارلو (MCM)، التي تساعد على الوصول إلى النتائج المثالية، حيث تضبط جميع المُتغيّرات، وتُوفّر الوقت والجهد على الباحث، وقد تباينت الدراسات السابقة بين استخدامها لبيانات حقيقية ومولّدة (محاكاة) ضمن أحجام وتوزيعات للبيانات مختلفة.

وقد استفيد من الإجراءات المُتبعة بالدراسات السابقة في أثناء تطبيق الدراسة الحالية بما يتناسب مع طبيعة المُتغيّرات المُستخدَمة والأساليب الإحصائية، كما فُسّرت نتائج هذه الدراسة في ضوء النتائج التي توصّلت إليها الدراسات السابقة.

فرضيات الدراسة:

لا توجد فروق دالة إحصائيًّا عند مستوى الدلالة (α=٠,٠٥)، في دقة تقدير مَعْلَمَة القدرة المُقدّرة باستخدام فقرات ثنائية التدريج، تبعًا لطريقة تقدير الدرجات (دلتا DSM-L المعلمية، وكيرنل KS اللامعلمية)، وحجم العينة (١٠٠- ٢٥٠- ٥٠٠- ١٠٠٠)، وتوزيع البيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، والتفاعل بينها.

منهج الدراسة: استخدمت الباحثة منهج المحاكاة التجريبي المقارن لملاءمته لأهداف الدراسة وأسئلتها. حيث أن هذا المنهج يُعد أسلوب فعال لاختبار الفروق والتأثيرات بين مجموعات مختلفة في ظروف متحكم فيها، مما يساهم في تحسين الفهم حول كيفية تأثير العوامل المختلفة على النتائج في العديد من المجالات البحثية.

بيانات الدراسة: تم توليد بيانات الدراسة باستخدام المحاكاة وفق طريقة مونت كارلو (MCM) وذلك لـ (٥٠) فقرة ثنائية التدريج. وأحجام العينات (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠) وحاكت كل من هذه العينات خمس توزيعات للبيانات في مجتمعات الدراسة المماثلة (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد) بمتوسطات حسابية (٢،١،٠-١،-٢،-μ) على الترتيب وانحراف معياري يساوي $(\sigma = 1)$. واعتمدت طريقة مونت كارلو على الاختيار العشوائي لقيمة من التوزيع المحدد لكل افتراض في كل عينة، وتمثل العينة مجموعة غير متحيزة من المجتمع الذي يمثل البيانات بأكملها.

تصميم الدراسة: استهدفت الدراسة الحالية التعرُّف على تأثير المتغيرات المستقلة طريقة التقدير (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية) وأحجام العينات (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠) وتوزيع البيانات (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد) على المتغير التابع دقة تقدير معلمة القدرة. ويوضح الجدول التالي التصميم العاملي (٢×٤×٥) (تصميم بين- داخل المجموعات Between – Within Subject Design) المستخدم في الدراسة.

جدول ١. التصميم العاملي للدراسة

		العامل (C) (مستويات توزيع البيانات) (بين المجموعات)	العامل (A) (طريقة التقدير) (داخل المجموعات)	
			(KS) A_1	DSM-) A_2 (L
العامل (B) (حجم العينة) (بين المجموعات)	(١٠٠) B_1	C1 (توزيع اعتدالي) $\mu=0,\sigma=1$	$A_1B_1C_1$	$A_2B_1C_1$
		C2 (التواء موجب بسيط) $\mu=-١,\sigma=1$	$A_1B_1C_2$	$A_2B_1C_2$
		C٣ (التواء موجب شديد) $\mu=-2,\sigma=1$	$A_1B_1C_3$	$A_2B_1C_3$
		C٤ (التواء سالب بسيط) $\mu=1,\sigma=1$	$A_1B_1C_4$	$A_2B_1C_4$
		C٥ (التواء سالب شديد) $\mu=2,\sigma=1$	$A_1B_1C_5$	$A_2B_1C_5$
	(250) B2	C1 (توزيع اعتدالي) $\mu=0,\sigma=1$	$A_1B_2C_1$	$A_2B_2C_1$
		C2 (التواء موجب بسيط) $\mu=-١,\sigma=1$	$A_1B_2C_2$	$A_2B_2C_2$
		C٣ (التواء موجب شديد) $\mu=-2,\sigma=1$	$A_1B_2C_3$	$A_2B_2C_3$
		C٤ (التواء سالب بسيط) $\mu=1,\sigma=1$	$A_1B_2C_4$	$A_2B_2C_4$
		C٥ (التواء سالب شديد) $\mu=2,\sigma=1$	$A_1B_2C_5$	$A_2B_2C_5$
	(500) B_3	C1 (توزيع اعتدالي) $\mu=0,\sigma=1$	$A_1B_3C_1$	$A_2B_3C_1$
		C2 (التواء موجب بسيط) $\mu=-١,\sigma=1$	$A_1B_3C_2$	$A_2B_3C_2$
		C٣ (التواء موجب شديد) $\mu=-2,\sigma=1$	$A_1B_3C_3$	$A_2B_3C_3$
		C٤ (التواء سالب بسيط) $\mu=1,\sigma=1$	$A_1B_3C_4$	$A_2B_3C_4$
		C٥ (التواء سالب شديد) $\mu=2,\sigma=1$	$A_1B_3C_5$	$A_2B_3C_5$
	(1000) B_4	C1 (توزيع اعتدالي) $\mu=0,\sigma=1$	$A_1B_4C_1$	$A_2B_4C_1$
		C2 (التواء موجب بسيط) $\mu=-١,\sigma=1$	$A_1B_4C_2$	$A_2B_4C_2$
		C٣ (التواء موجب شديد) $\mu=-2,\sigma=1$	$A_1B_4C_3$	$A_2B_4C_3$
		C٤ (التواء سالب بسيط) $\mu=1,\sigma=1$	$A_1B_4C_4$	$A_2B_4C_4$
		C٥ (التواء سالب شديد) $\mu=2,\sigma=1$	$A_1B_4C_5$	$A_2B_4C_5$

إجراءات الدراسة:

المرحلة الأولى (توليد البيانات): تم توليد البيانات وذلك لـ (٥٠) فقرة ثنائية التدريج (٠،١)، وهي عبارة عن القدرة الفعلية (المولدة) للأفراد وفق أحجام عينات مختلفة (١٠٠، 250، 500، ١٠٠٠) وتوزيعات بيانات مختلفة لكل منها (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد). كما تم توليد الفقرات وفق النموذج اللوجستي ثنائي المعلمة (2PLM) باستخدام برنامج (WinGen v1.4) (Han، ٢٠٠٧)، حيث كانت معالم الصعوبة وفقاً للتوزيع الاعتدالي (Normal) بوسط حسابي (٠) وانحراف معياري (١)، ومعالم التمييز وفقاً للتوزيع المنتظم (Uniform) بقيمة صغرى (0.3) وقيمة عظمى (1.9)، وتعد قيم التمييز التي تم توليدها قريبة للقيم التي استخدمها هامبلتون وسوامينيثان (Hambleton & ٢٠١٠ Swaminathan,) حيث أكد على أن تكون قيم معلمة التمييز تتراوح مابين (٠ و ٢) لوجت، والجدول (٣) يوضح الإحصاءات الوصفية الخاصة بمعلمة القدرة ومعالم الفقرات المولدة (الصعوبة، والتمييز).

جدول ٢.

الإحصاءات الوصفية الخاصة بمعلمة القدرة ومعالم الفقرات المولدة (الصعوبة، والتمييز)

حجم العينة	مستويات توزيع القدرة	خصائص توزيع المعلمات		
		القدرة	الصعوبة	التمييز
(١٠٠)	(توزيع اعتدالى)	$\mu=0,\sigma=1$	(توزيع اعتدالي) $\mu=0$ $\sigma=1$	(توزيع منتظم) القيمة الصغرى= 0.3 القيمة العظمى=1.9
	(التواء موجب بسيط)	$\mu=-١,\sigma=1$		
	(التواء موجب شديد)	$\mu=-2,\sigma=1$		
	(التواء سالب بسيط)	$\mu=1,\sigma=1$		
	(التواء سالب شديد)	$\mu=2,\sigma=1$		
(٢٥٠)	(توزيع اعتدالى)	$\mu=0,\sigma=1$	(توزيع اعتدالي) $\mu=0$ $\sigma=1$	توزيع منتظم) القيمة الصغرى= 0.3 القيمة العظمى=1.9
	(التواء موجب بسيط)	$\mu=-١,\sigma=1$		
	(التواء موجب شديد)	$\mu=-2,\sigma=1$		
	(التواء سالب بسيط)	$\mu=1,\sigma=1$		
	(التواء سالب شديد)	$\mu=2,\sigma=1$		
(٥٠٠)	(توزيع اعتدالى)	$\mu=0,\sigma=1$	(توزيع اعتدالي) $\mu=0$ $\sigma=1$	(توزيع منتظم) القيمة الصغرى= 0.3 القيمة العظمى=1.9
	(التواء موجب بسيط)	$\mu=-١,\sigma=1$		
	(التواء موجب شديد)	$\mu=-2,\sigma=1$		
	(التواء سالب بسيط)	$\mu=1,\sigma=1$		
	(التواء سالب شديد)	$\mu=2,\sigma=1$		
(١٠٠٠)	(توزيع اعتدالى)	$\mu=0,\sigma=1$	(توزيع اعتدالي) $\mu=0$ $\sigma=1$	(توزيع منتظم) القيمة الصغرى= 0.3 القيمة العظمى=1.9
	(التواء موجب بسيط)	$\mu=-١,\sigma=1$		
	(التواء موجب شديد)	$\mu=-2,\sigma=1$		
	(التواء سالب بسيط)	$\mu=1,\sigma=1$		
	(التواء سالب شديد)	$\mu=2,\sigma=1$		

يوضح الجدول (٢) المتوسطات الحسابية والانحرافات المعيارية والقيمة العظمى والصغرى لمعلمة القدرة الفعلية (المولدة) لكل حالة من الحالات المشمولة بالدراسة وفقاً لمتغيري (حجم العينة، توزيع البيانات).

جدول٣.المتوسطات الحسابية والانحرافات المعيارية لمعلمة القدرة وفقاً لمتغيري الدراسة (حجم العينة، توزيع البيانات)

حجم العينة	الإحصائي	مستويات توزيع القدرة				
		توزيع اعتدالي	التواء موجب بسيط	التواء موجب شديد	التواء سالب بسيط	التواء سالب شديد
(١٠٠)	المتوسط الحسابي	0.242	-0.879	-1.801	1.162	2.032
	الانحراف المعياري	0.968	1.097	0.899	1.042	0.855
	القيمة العظمى	2.864	1.931	0.120	3.171	4.066
	القيمة الصغرى	-1.492	-4.939	-3.407	-1.981	-0.253
(٢٥٠)	المتوسط الحسابي	0.131	-0.921	-1.877	1.122	2.038
	الانحراف المعياري	0.967	1.016	0.992	1.010	0.955
	القيمة العظمى	2.719	1.773	1.112	4.173	4.271
	القيمة الصغرى	-2.455	-3.598	-4.870	-1.731	-0.109
(٥٠٠)	المتوسط الحسابي	0.133	-0.888	-1.979	1.119	2.030
	الانحراف المعياري	1.028	1.003	0.997	0.959	0.984
	القيمة العظمى	3.076	2.191	1.752	4.584	4.807
	القيمة الصغرى	-2.720	-3.905	-4.747	-2.029	-0.997
(1000)	المتوسط الحسابى	0.054	-0.972	-1.982	1.005	2.038
	الانحراف المعياري	1.000	1.035	1.011	1.033	1.000
	القيمة العظمى	3.348	2.127	1.878	4.589	5.061
	القيمة الصغرى	-3.089	-5.463	-5.249	-2.502	-1.659

يوضح الجدول (٣) المتوسطات الحسابية والانحرافات المعيارية لمعالم الفقرات المولدة من الصعوبة، والتمييز التي تم توليدها وفقاً لمتغيري الدراسة (حجم العينة، وتوزيع البيانات).

جدول ٤. المتوسطات الحسابية والانحرافات المعيارية لمعالم الفقرات المولدة (الصعوبة، والتمييز) وفقاً لمتغيري الدراسة (حجم العينة، وتوزيع البيانات)

حجم العينة	الإحصائي	توزيع اعتدالي		التواء موجب بسيط		التواء موجب شديد		التواء سالب بسيط		التواء سالب شديد	
		تمييز	صعوبة	تمييز	صعوبة	تمييز	صعوبة	تمييز	صعوبة	تمييز	صعوبة
(١٠٠)	المتوسط الحسابي	1.279	0.234	1.121	0.204	1.194	-0.127	1.087	0.276	1.301	0.442
	الانحراف المعياري	0.480	0.979	0.425	0.986	0.405	0.988	0.499	1.050	0.345	1.175
	القيمة العظمى	1.900	2.073	1.892	2.099	1.852	1.692	1.869	2.716	1.867	2.757
	القيمة الصغرى	0.378	-1.791	0.353	-1.789	0.352	-2.165	0.337	-1.787	0.642	-1.672
(٢٥٠)	المتوسط الحسابي	1.481	0.256	1.532	0.035	1.528	-0.247	1.482	0.191	1.356	0.283
	الانحراف المعياري	0.302	1.103	0.235	1.052	0.280	1.167	0.309	1.078	0.262	0.899
	القيمة العظمى	1.888	2.428	1.886	2.063	1.888	2.447	1.867	2.378	1.753	1.902
	القيمة الصغرى	0.656	-2.587	0.941	-2.184	0.866	-2.769	0.333	-2.111	0.767	-1.565
(٥٠٠)	المتوسط الحسابي	1.534	-0.208	1.483	-0.112	1.437	-0.014	1.477	-0.109	1.497	0.364
	الانحراف المعياري	0.245	1.169	0.321	1.085	0.351	1.128	0.335	1.485	0.309	1.046
	القيمة العظمى	1.889	1.636	1.900	2.282	1.880	1.942	1.880	2.888	1.881	2.333
	القيمة الصغرى	0.939	-2.526	0.425	-2.095	0.316	-2.259	0.331	-2.892	0.620	-1.622
(1000)	المتوسط الحسابي	1.524	0.153	1.579	0.037	1.500	-0.156	1.514	-0.234	1.564	0.238
	الانحراف المعياري	0.313	1.216	0.293	1.147	0.257	1.132	0.349	1.170	0.233	1.141
	القيمة العظمى	1.899	3.196	1.894	2.804	1.899	1.776	1.895	1.98	1.88	2.049
	القيمة الصغرى	0.558	-2.373	0.435	-1.969	0.804	-2.17	0.308	-2.647	1.023	-2.474

المرحلة الثانية (تحليل البيانات): من أجل تحقيق الهدف من الدراسة والاجابة على سؤالي الدراسة، قامت الباحثة بتحليل البيانات التي تم توليدها وفق الخطوات التالية:

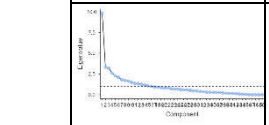
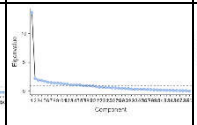
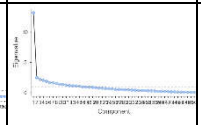
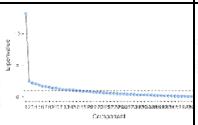
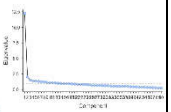
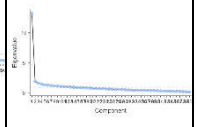
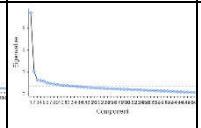
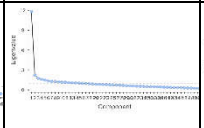
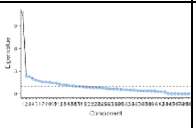
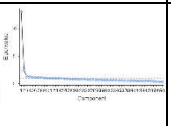
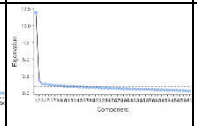
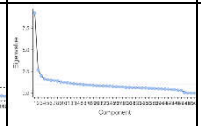
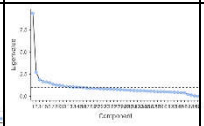
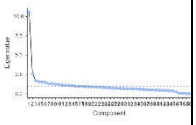
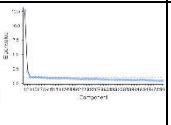
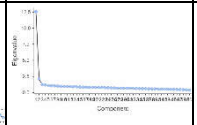
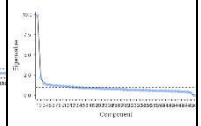
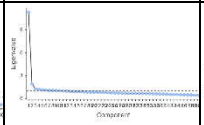
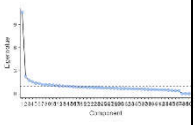
-أولاً: التحقق من افتراضات النموذج الثنائي المعلمة (2PLM): وهو النموذج الذي تم توليد البيانات على أساسه باستخدام برنامج (WinGen v1.4) (Han، ٢٠٠٧).

- الافتراض الأول: أحادية البعد (Unidimensionality): تم التحقق من أحادية البعد باستخدام كل من برنامج Spss و Jamovi 2.3.28 عن طريق التحليل العاملي الاستكشافي لكل حجم من أحجام العينات الأربعة ولكل توزيع من التوزيعات الخمسة باستخدام طريقة المكونات الرئيسية (Principal Components Analysis) ومن ثم تدوير العوامل باستخدام طريقة (Varimax). والجدول (٥) يوضح قيم الجذر الكامن ونسب التباين المفسر للعاملين الأول والثاني وناتج قسمة قيمة الجذر للعامل الأول على جذر العامل الثاني في التوزيعات والأحجام المختلفة.

جدول ٥. العوامل الكامنة المستخلصة من التحليل العاملي الاستكشافي لبيانات الدراسة باختلاف حجم العينة وتوزيع البيانات

حجم العينة	(مستويات توزيع البيانات)	العامل الأول		العامل الثاني		العامل الثالث		الجذر الكامن للعامل الأول/الجذر الكامن للعامل الثاني	الجذر الكامن للعامل الثاني-الجذر الكامن للعامل الثالث
		الجذر الكامن	نسبة التباين المفسر	الجذر الكامن	نسبة التباين المفسر	الجذر الكامن	نسبة التباين المفسر		
(١٠٠)	توزيع اعتدالي))	١٢,٧٩٢	٢٥,٥٨٥	٢,٢٨٣	٤,٥٦٥	٢,١٢٨	٤,٢٥٦	٥,٦٠٥	٦٧,٨٩٢
	(التواء موجب بسيط)	١٣,٦٣١	٢٧,٢٦٢	٢,٢٠٦	٤,٤١٢	١,٩٤٣	٣,٨٨٥	٦,١٧٩	٤٣,٣٥٩
	التواء موجب شديد))	١١,٢٣٩	٢٢,٤٧٧	٢,٥٦٤	٥,١٢٨	٢,١٥٠	٤,٢٩٩	٤,٣٨٣	٢٠,٩٤٤
	(التواء سالب بسيط)	١٣,١٨٣	٢٦,٣٦٥	٢,٤٨٧	٤,٩٧٣	٢,١٨٥	٤,٣٦٩	٥,٣٠١	٣٥,٤٠٥
	(التواء سالب شديد)	١٢,٢٦٥	٢٤,٥٢٩	٣,٢٩٢	٦,٥٨٣	٢,٦٠٨	٥,٢١٦	٣,٧٢٦	١٣,١١٨
(٢٥٠)	توزيع اعتدالي))	١٢,٢٧٢	٢٤,٥٤٣	٢,٠٢٦	٤,٠٥٢	١,٥٣٧	٣,٠٧٥	٦,٠٥٧	٢٠,٩٥٣
	(التواء موجب بسيط)	١٣,١٨٠	٢٦,٣٥٩	١,٩٧٢	٣,٩٤٤	١,٦٧٣	٣,٣٤٥	٦,٦٨٣	٣٧,٤٨٥
	التواء موجب شديد))	١١,٠٩٣	٢٢,١٨٥	٣,٠٨٦	٦,١٧١	١,٨٥١	٣,٧٠٢	٣,٥٩٥	٦,٤٨٣
	(التواء سالب بسيط)	١١,٨٣٣	٢٣,٦٦٦	٢,١٩٥	٤,٣٩٠	١,٧٢١	٣,٤٤٣	٥,٣٩١	٢٠,٣٣٣
	(التواء سالب شديد)	١٠,٧٩٩	٢١,٥٩٧	٢,٣٣٤	٤,٦٦٨	٢,٢٧٦	٤,٥٥١	٤,٦٢٧	١٤٥,٩٤٨
(٥٠٠)	توزيع اعتدالي))	١٣,١٠٨	٢٦,٢١٥	٢,٣٨٧	٤,٧٧٤	١,٢٣٩	٢,٤٧٨	٥,٤٩١	٩,٣٣٩
	(التواء موجب بسيط)	١١,٩٧٠	٢٣,٩٤٠	١,٩٨١	٣,٩٦٣	١,٣٨٢	٢,٧٦٣	٦,٠٤١	١٦,٦٧٦
	التواء موجب شديد))	٩,٢٧١	١٨,٥٤١	٢,٦٧٧	٥,٣٥٤	٢,٠٠٦	٤,٠١١	٣,٤٦٣	٩,٨١٩
	(التواء سالب بسيط)	٩,٣٧٤	١٨,٧٤٨	٢,٧١٢	٥,٤٢٤	١,٨٦٨	٣,٧٣٧	٣,٤٥٦	٧,٨٩٥
	(التواء سالب شديد)	١٠,٥٦٨	٢١,١٣٧	٢,٥٩٠	٥,١٨٠	١,٦٦١	٣,٣٢٢	٤,٠٨٠	٨,٥٨٨
(١٠٠٠)	توزيع اعتدالي))	١٢,٦١٦	٢٥,٢٣٢	٢,٠٨٧	٤,١٧٣	١,١١١	٢,٢٢٢	٦,٠٤٦	١٠,٧٨٨
	(التواء موجب بسيط)	١٣,٠٤٥	٢٦,٠٩١	٢,٢٩٨	٤,٥٩٦	١,٥٠٩	٣,٠١٨	٥,٦٧٦	١٣,٦١٩
	التواء موجب شديد))	١٠,٠٧٣	٢٠,١٤٦	٢,١٤٧	٤,٢٩٤	١,٥٩٠	٣,١٧٩	٤,٦٩١	١٤,٢٢٠
	(التواء سالب بسيط)	١١,٢٧٠	٢٢,٥٤٠	١,٩٠٧	٣,٨١٥	١,٢٦٠	٢,٥٢٠	٥,٩٠٨	١٤,٤٧١
	(التواء سالب شديد)	١٠,٦٤٤	٢١,٢٨٧	٢,٢٣٩	٤,٤٧٨	١,٧٤٣	٣,٤٨٦	٤,٧٥٣	١٦,٩٤٥

جدول ٦. الرسم البياني (Scree Plot) لقيم الجذور الكامنة لجميع العوامل لبيانات الدراسة باختلاف حجم العينة وتوزيع البيانات

التوزيع / حجم العينة	اعتدالي	موجب بسيط	موجب شديد	سالب بسيط	سالب شديد
١٠٠					
٢٥٠					
٥٠٠					
١٠٠٠					

يتضح من النتائج الواردة في الجدول (٦) بأن جميع نسب التباين المفسر من العامل الأول كانت أعلى من (٢٠%) في الأحجام والتوزيعات المختلفة فيما عدا (التواء موجب شديد والتواء سالب بسيط لحجم العينة ٥٠٠) كانت القيم مرتفعة وقريبة جداً من المحك المطلوب، وكما هو واضح في جميع ظروف الدراسة الارتباك في القيم في حال الالتواءات بين المرتفع والمنخفض عنه في التوزيع الاعتدالي، وتتفق هذه النتيجة مع نتيجة دراسة (الحواري، ٢٠١٥؛ الشرفاوي، ٢٠٢٢)، وهذه النتيجة تُعتبر مؤشر أول على تحقق أحادية البعد، وأن ناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني أكبر من (٢) في الأحجام والتوزيعات المختلفة بما فيها (التواء موجب شديد والتواء سالب بسيط) لحجم العينة (٥٠٠)؛ والتي كانت قريبة من المحك المطلوب في نسب التباين المفسر من العامل الأول كمؤشر ثانٍ على تحقق أحادية البعد، وكذلك أظهرت نتائج عملية قسمة حاصل طرح الجذر الكامن للعامل الثاني من الجذر الكامن للعامل الأول على حاصل طرح الجذر الكامن الثالث من الجذر الكامن للعامل الثاني قيماً ضخمة في أغلب التوزيعات مما يشير إلى تحقق افتراض أحادية البعد كمؤشر ثالث، والجدول(٦) يوضح نتائج التحليل العاملي للبيانات المولدة عن طريق الرسم البياني (Scree Plot) لقيم الجذور الكامنة لكل عامل على المحور الصادي ورقم المكون على المحور السيني وفقاً لمتغيرات الدراسة(حجم العينة، مستويات توزيع

البيانات) كمؤشر رابع على تحقق أحادية البعد. من الواضح تحقق أحادية البعد لجميع الظروف من الاحجام والتوزيعات المختلفة بناء على المؤشرات السابقة.

- الافتراض الثاني: الاستقلال الموضعي (Loca Independence): تم التحقق من افتراض الاستقلال الموضعي من خلال مؤشر ين (Yen's Index) المعروف بـ (Q_3)، وهو معامل الارتباط بين البواقي لأزواج الفقرات بعد ضبط القدرة أو السمة، حيث تم تحليل البيانات باستخدام حزمة (SnowIRT) من برنامج Jamovi 2.3.28 ، وإيجاد معامل الارتباط بين البواقي لأزواج فقرات إحصائياً، وقد كانت غالبية قيم معامل الارتباط بين البواقي لجميع أزواج فقرات الاختبار أقل من أو تساوي (٠,٣٠)، وتبين أنه في جميع ظروف الدراسة التي يكون فيها التوزيع اعتدالي يكون الاستقلال الموضعي متحقق، وهذا يتفق مع دراسة(الشرفاوي،٢٠٢٢)، أما في حالة التوزيع الملتوي فيظهر لنا وجود أزواج للفقرات بينها ارتباط موضعي وهي بين (٢%-٩%) حيث تمثل نسبة بسيطة جداً مقارنة بالعدد الكلي للأزواج التي حققت الاستقلال الموضعي وهي بين (٩١%-١٠٠%)، وبفحص البيانات تبين ارتفاع نسبة تحقق الاستقلال الموضعي كلما كان حجم العينة كبير وايضًا كلما كان توزيع البيانات اعتداليًا أو قريب من الاعتدالية، لذا ترا الباحثة أن عدم تحقق الافتراض الاستقلال الموضعي لعدد من الأزواج قد لا يشير بالضرورة إلى تأثر الإجابة عن فقرة بالفقرة الأخرى، ولابد من الأخذ بالاعتبار عوامل أخرى تؤثر على الاستقلال الموضعي للفقرات كما لحظنا التفاوت في نسبة الأزواج المستقلة موضعياً تبعاً لحجم العينة ومستويات توزيع البيانات والتي قد تؤثر على دقة درجة القطع التي افترضها ين (Yen) للحكم على (Q_3)(Christensen et al., 2017).

وبالاستناد إلى ما ذكره (DeMars, 2010; Hambleton et al., 1991; Hulin et al., 1983; Raykov & Marcoulides, 2016; Reise & Revicki, 2015) أن هذا الشرط يتحقق ضمنياً بتحقق شرط أحادية البعد؛ حيث أن هناك ارتباطاً وثيقاً بين تحقق افتراض أحادية البعد وتحقق افتراض الاستقلال الموضعي؛ وفقا لما سبق ترى الباحثة أن الاستجابات عن الفقرات حققت افتراض الاستقلال الموضعي والإبقاء على عدد الفقرات ٥٠ فقرة ولن يتم حذف أي فقرة.

- الافتراض الثالث: الاطرادية (Monotonicity): ويُقصد بهذا الافتراض أنه بزيادة قدرة الفرد تزداد احتمالية إجابته الصحيحة عن الفقرة أو تبقى ثابتة ضمن مستويات القدرة المختلفة (Stochl, 2006)، وتم التحقق من هذا الافتراض باستخدام برنامج (Bailog MG 3.0) وذلك عن طريق منحنيات خصائص الفقرة (ICC) لكل فقرة من الفقرات باختلاف حجم العينة (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠) وتوزيع البيانات (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد)، ويوضح الشكل التالي تحقق الاطرادية لأحد ظروف الدراسة حيث حجم العينة ١٠٠، وتوزيع البيانات اعتدالي، عدد الفقرات ٥٠ فقرة.

- ثانياً: مطابقة البيانات المولدة للنموذج:

- للتحقق من مطابقة الفقرات والأفراد للنموذج الثنائي المعلمة (2PLM)، تم استخدام لغة البرمجة (R) بواسطة حزمة (mirt -A package for R) وحساب إحصائيات جودة الملاءمة (Goodness of fit -GOF).

وقد تم الحكم على مطابقة الفقرات وفق المؤشر الاحصائي ($S\text{-}X^2$) الذي قدمه كل من ثيسن وأولاندو في (Orlando & Thissen, 2000) للمفردات ثنائي التدريج، كما أثبت هذا المؤشر فاعليته في دراسة (Orlando & Thissen, 2003) في الكشف عن مطابقة الفقرات في

اختبارات مختلفة مكونة من (١٠-٤٠-٨٠) فقرة وعينة(١٠٠٠) فرد وعند مستويات الدلالة (P<.05) لضبط الخطأ من النوع الأول، وقد كانت النتائج متقاربة باختلاف مستوى الدلالة، حيث تعد الفقرة غير مطابقة للنموذج اذا قلت قيمة الدلالة الإحصائية عن(.05)، ويمكن التجاوز عن عدم مطابقة الفقرة اذا كانت قيمة المؤشر غير دالة عند مستوى الدلالة(.01). لذلك اختير هذا المؤشر للحكم على مطابقة الفقرات لمناسبته لبيانات وظروف الدراسة.

وقد كانت نسبة وعدد الفقرات الغير مطابقة للنموذج الثنائي المعلمة (2PLM) من خلال المؤشر ($S\text{-}X^2$) والتي تراوحت بين (٢%-٤%) وهي نسبة بسيطة جداً مقارنة بنسبة الفقرات المطابقة، كما لاحظنا المطابقة تكون جيدة عندما يكون التوزيع اعتدالي في جميع ظروف الدراسة ويظهر عدم مطابقة بسيطة كلما كان الالتواء شديد في توزيع البيانات، كما لاحظنا جودة المطابقة كلما زاد حجم العينة وهذه النتيجة جاءت متوافقة مع ما توصلت اليه دراسة (بني عطا والشريفين، ٢٠١٢)، لذا ارتأت الباحثة الاحتفاظ بجميع الفقرات بغرض اجراء عملية المقارنة بين معلمات القدرة الفعلية (المولدة) والقدرة المقدرة بالطرق المستخدمة في هذه الدراسة وحيث أن النسبة البسيطة لا تؤثر على نتائج الدراسة .

أما بالنسبة لمطابقة الأفراد للنموذج الثنائي المعلمة (2PLM)، فقد تم الحكم على مطابقة الأفراد وفق المؤشرات الإحصائية (zh، z-infit،z-outfit) ، إذا كانت قيم المؤشرات الثلاثة بين (٢- و٢+) نقول أن مطابقة الأفراد جيدة، وتكون أفضل مطابقة كلما اقتربت القيم من الصفر، وإذا كانت القيم أعلى أو أقل من (٢- و٢) تكون غير مطابقة. تشير بعض الدراسات إلى أنه إذا كانت نسبة عدم المطابقة أكثر من (٥%) فقد يشير ذلك إلى وجود مشكلة في النموذج أو البيانات (Linacre, 2002)، يعتبر الشخص غير مناسب إذا كان اثنان على الأقل من المؤشرات ,z-outfit (zh،z-infit) خارج النطاق (٢- وَ٢).

وقد توصلنا الى نسبة الأفراد الغير مطابقين والتي تراوحت بين (١%- ٣%) وهذا يعني أنا النسبة في حدود النسبة المسموح بها حسب ما ورد في دراسة (Linacre, 2002).

مما سبق يتضح جودة مطابقة الفقرات والأفراد للنموذج الثنائي المعلمة (2PLM).

وللكشف عن مطابقة البيانات بشكل عام للنموذج الثنائي المعلمة (2PLM)، تم استخدام العديد من المؤشرات للحكم على جودة النموذج بشكل عام. وفي هذه الدراسة تم اعتبار النموذج غير مناسب إذا كانت جميع المؤشرات تشير إلى عدم ملاءمته وتم اعتبار النموذج مناسبًا جيدًا إذا كانت جميع المؤشرات تشير إلى ملاءمة جيدة وأخيرًا تم اعتبار النموذج مناسبًا بشكل معتدل عندما تشير بعض المؤشرات إلى ملاءمة جيدة بينما تشير المؤشرات الأخرى إلى عدم ملاءمة.

في هذه الدراسة اعتمدت الباحثة على المؤشرات:

- مؤشر مربع كاي M2: وهو مؤشر لتقييم ملاءمة النموذج، تشير القيمة الأقل لـ M2 عمومًا إلى مطابقة أفضل بين النموذج والبيانات (Maydeu & Joe, 2005).

•p - value: قيمة p المرتبطة بالمؤشر M2. تشير القيمة p العالية (p>٠,٠٥) إلى أن النموذج يناسب البيانات جيدًا. وتشير قيمة p المنخفضة (p<٠,٠٥) إلى أن النموذج لا يتناسب تمامًا مع البيانات، وهو ما يحدث غالبًا مع أحجام العينات الأكبر، حيث يصبح الاختبار أكثر حساسية حتى للانحرافات الطفيفة عن الملاءمة المثالية. تتأثر قيمة p بحجم العينة، ومع عينة أكبر، حتى عدم المطابقة الصغيرة يمكن أن تؤدي إلى قيمة p كبيرة، وهذا هو السبب في أنها قد تشير إلى ضعف المطابقة حتى عندما تشير المؤشرات الأخرى إلى مطابقة معقولة.

- مؤشر جذر متوسط مربع الخطأ التقريبي (Root Mean Square Error of Approximation-RMSEA):

تشير القيم المنخفضة لـلمؤشرات (RMSEA, RMSEA _95, RMSEA _5) إلى ملاءمة أفضل، حيث تشير القيم الأقل من (0.06) بشكل عام إلى ملاءمة جيد (Hu & Bentler, 2009)، وهو مؤشر جيد يُعالج حساسية المؤشر (M2) لحجم العينة وعدد المعالم (Browne & Cudeck, 1993).

- مؤشر متوسط الجذر التربيعي المعياري المتبقي (Standardized Root Mean Square Residual-SRMSR) والذي يقيس الفرق المعياري بين الارتباطات الملاحظة والمتوقعة. تشير القيمة الأقل إلى ملاءمة أفضل، حيث تعتبر القيم التي تقل عن (٠,٠٨) جيدة عادةً. في هذه الحالة، يكون (SRMSR) مرتفعًا نسبيًا، مما يشير إلى أنه قد تكون هناك اختلافات ملحوظة بين تنبؤات النموذج والبيانات المرصودة. وهذا يشير إلى مجالات النموذج التي يمكن تحسينها.

- مؤشر تاكر لويس (Tucker-Lewis Index -TLI) ومؤشر الملاءمة المقارن(CFI -Comparative Fit Index)، حيث تشير القيم الأكبر من (٠,٩٠) إلى ملاءمة جيدة. هذه المؤشرات أقل حساسية لحجم العينة وتميل إلى تقديم نتائج أفضل لملاءمة النموذج، خاصة عندما يكون النموذج معقدًا أو يكون حجم العينة كبيرًا(Hu & Bentler, 2009) .

وتشير النتائج إلى مطابقة البيانات بشكل عام للنموذج، مع ملاحظة حساسية مؤشر مربع كاي لحجم العينة حيث كان يعطي نتائج عدم مطابقة مع زيادة حجم العينة وهذه النتيجة تدعم ما توصلت اليه كثير من الدراسات منها:(الصيخان، ٢٠٢٤؛ عامر، ٢٠٠٤؛ Dogan,2015) ، أما باقي المؤشرات كانت تعطي نتائج جيدة تشير إلى مطابقة البيانات بشكل عام للنموذج في هذه الدراسة، وهذه النتيجة تتفق مع دراسة كارون وآخرون(Karon et al., ٢٠٠٩) حيث أظهرت النتائج فاعلية مؤشر تاكر لويس (TLI) في حالة أحجام العينات الكبيرة التوزيع الملتوي للبيانات.

ثالثاً: مطابقة البيانات للنماذج المستخدمة للتقدير في الدراسة:

- مطابقة البيانات لنموذج موكن للتجانس الاطرادي اللامعلمي

للتحقق من مطابقة الفقرات لنموذج موكن للتجانس الاطرادي اللامعلمي تم استخدام لغة البرمجة (R) بواسطة حزمة (Mokken-A package for R)، لاستخراج معاملات التدريج (scalability coefficients) والتي تستخدم للتحقق من التجانس الاطرادي في فقرات الاختبار أو المقياس وهي: معاملات تدريج أزواج الفقرات (H_{ij}): تم حساب معاملات تدريج أزواج الفقرات، وأظهرت النتائج مطابقة جميع الفقرات لنموذج موكن حيث كانت قيم معاملات التدريج جميعها موجبة لكل ظرف من ظروف الدراسة. معامل (H_i) لكل فقرة في التدريج بالنسبة لباقي الفقرات، وأظهرت النتائج تمتع بعض الفقرات بقيم تدريج عالية جداً، ونسبة بسيطة من الفقرات كانت قيم التدريج منخفضة لكنها قريبة جداً وتكاد تصل إلى القيمة (0.3)، بشكل عام المطابقة جيدة ونسبة الفقرات الغير مطابقة بسيطة بالنسبة للعدد الكلي للفقرات، كما أن وجود نسبة بسيطة جداً من الفقرات الغير مطابقة يمكن التجاوز عنها، حيث كانت الفقرات الغير مطابقة في حال وجود التواء في توزيع البيانات وهذه النتيجة تتفق مع ما توصلت اليه الدراسة التي أجراها زيكا (Zickar,1997) حيث طابقت معظم الفقرات مقياس هوغان للشخصية، ودراسة (مؤمني، ٢٠١٥) والتي خلصت إلى عدم مطابقة خمس فقرات من أصل (٢٨) فقرة من الفقرات لنموذج موكن الاطرادي.

معامل التدريج الكلي للفقرات جميعها(H): تم حساب معامل التدريج الكلي للحكم على جودة التدريج، حيث ذكر كل من سيجتسما ومولينار (Sijtsma & Molenaar، ٢٠٠٢) أن موكن حدد معايير لمعاملات التدريج لضمان جودة التدريج، واقترح حداً أدنى لمعامل تدريج الفقرة (٠,٣ = Hi)، كما اقترح بعض المعايير للحكم على جودة التدريج الكلي، وهي كالتالي: مقياس غير قابل للتدريج (H < ٠,٣)، مقياس ضعيف (٠,٤ ≤ H ≤ ٠,٣)، مقياس متوسط (٠,٥ ≤ H ≤ ٠,٤)، مقياس قوي (H ≥ ٠,٥). وقد تبين مستوى جودة تدريج متوسطة بشكل عام لمعظم ظروف الدراسة، وهذا يتفق مع دراسة جيرلينج وآخرون (Geerlings et al., 2014)، كما لاحظنا أن المستوى ضعيف في حال الالتواءات الشديدة، وهذا ينسجم مع ما توصّلت اليه كل من دراسة (الحواري، ٢٠١٥)، وريس (Reis، ١٩٨٦)، وكارن وآخرون (Karon et al., 2008)، وهيو وبينتلير (Hu & Bentler ,2009) وغيرها من الدراسات من أن طبيعية توزيع البيانات يؤثر في عدة أمور منها تحقق الافتراضات ومطابقة البيانات وتقدير معالم الفقرات وقدرات الأفراد.

تبين مما سبق تحقيق جميع فقرات الاختبار المكونة من (٥٠) فقرة لشروط نموذج موكن للتجانس الاطرادي فيما يخص معاملات التدريج (H) بأنواعها الثلاثة، الأمر الذي يجيز تسمية التدريج الناتج بتدريج موكن، فقد كانت جميع معاملات تدريج أزواج الفقرات موجبة، وجميع معاملات تدريج الفقرات موجبة، إضافةً إلى أن معامل التدريج الكلي للاختبار بلغ قيمة متوسطة في الغالب، مما يشير إلى المطابقة بشكل عام.

وللتحقق من مطابقة الأفراد لنموذج موكن للتجانس الاطرادي اللامعلمي: لم يحصل أي مفحوص على العلامة الكاملة على الاختبار، وهي (٥٠)، وكذلك لم يحصل أي مفحوص على العلامة الصغرى على الاختبار، وهي (الصفر)، وبهذا يكون جميع لأفراد عينة الدراسة مطابقين لنموذج التجانس الاطرادي لموكن.

مطابقة البيانات لنماذج دلتا المعلمية: تم استخدام برنامج دلتا (DELTA) حيث اعتمدت الباحثة على مؤشر متوسط الاختلاف المطلق (Mean Absolute Difference-MAD) المشار اليه من قبل ديمتروف وأتانسوف (Dimitrov,2023 ;Dimitrov & Luo, 2017) في التعرف على مدى مطابقة الفقرات، وللحكم على مستوى المطابقة لابد من الاعتماد على المحكات التالية:

- إذا كانت قيمة المؤشر (MAD≤ 0.07) دل ذلك على المطابقة الجيدة للفقرة.
- إذا كانت قيمة المؤشر (0.07<MAD< 0.10) تدل على مطابقة مقبولة للفقرة.
- إذا كانت قيمة المؤشر (MAD≥ 0.10) تدل على المطابقة الضعيفة للفقرة

توصلت الدراسة الى جودة مطابقة معظم الفقرات لنموذج دلتا لتقدير الدرجات وخاصة في ظروف الدراسة التي يكون التوزيع فيها اعتدالياً، بينما يلاحظ وجود نسبة بسيطة من عدم المطابقة في ظروف الدراسة التي يكون التوزيع فيها ملتوياً سواء التواء سلبي أو ايجابي.

وللكشف عن مطابقة الأفراد اختارت الباحثة أحد مؤشرات مطابقة الأفراد المستخرجة باستخدام برنامج دلتا والذي يتمثل في إحصائي (Ud Statistic) والذي يعتمد على المؤشر الإحصائي (Person-Fit Statistic(U3، وتنحصر قيمة المؤشر (Ud) بين (٠-١) حيث يشير الصفر إلى مطابقة استجابة الفرد للنموذج بينما يشير الواحد إلى عدم المطابقة التامة (Dimitrov, 2023; Dimitrov et al., 2020). وقد تم اعتماد هذا المؤشر في هذه الدراسة باعتباره المؤشر الأنسب لمقارنة مدى مطابقة استجابات الأفراد لنموذج دلتا لتقدير الدرجات مع مدى مطابقة استجابات هؤلاء الأفراد لنموذج الثنائي(2PLM) حيث أن مؤشر المطابقة المستخدم في الدراسة الحالية يعتمد على المؤشر الإحصائي (U٣) الذي وضعه فان دير فلير (van der Flier, 1980)

.واتضح أن معظم القيم كانت محصورة بين (٠ - ٠.6) وتقترب من (٠) عندما يكون توزيع البيانات اعتداليًا وتبدأ بالابتعاد عن الصفر كلما زادت درجة الالتواء، كما يتضح عدم تجاوز قيم مؤشر المطابقة لـ (0.7) لمعظم أفراد الدراسة، مما يدل على وجود قدر بسيط جداً من عدم جودة مطابقة استجابات الأفراد للنموذج.

مما سبق من مطابقة الفقرات والأفراد لنماذج دلتا المعلمية وجدنا أن نسبة الفقرات التي كانت مطابقتها ضعيفة لنماذج دلتا لتقدير الدرجات لم تتجاوز (٦%)، وعدم وصول قيم مؤشر المطابقة (Ud) في مطابقة الأفراد إلى حدود عدم المطابقة التامة وهي الواحد الصحيح لجميع ظروف الدراسة؛ ارتأت الباحثة الإبقاء على جميع الفقرات والأفراد في الدراسة .

المرحلة الثالثة (تقدير معلمة القدرة): اعتمدت هذه الدراسة في تقدير معلمة القدرة وفق النماذج المعلمية (PIRT) بطريقة دلتا على برنامج الدلتا (DELTA 3.1)، باستخدام الإطار الكامن لأسلوب دلتا لتقدير الدرجات (DSM-L) والتي تستخدم أحد أسلوبي الأرجحية العظمى (المشتركة (JML) أو الهامشية (MML)). أما تقدير معلمة القدرة وفق نماذج (NIRT) بطريقة كيرنل (KS) اللامعلمية تم استخدام برنامج R بوسطة حزمة (KernSmoothIRT)التي طورها ماز وآخرون (Mazza et al., 2014, 2020) لتقدير القدرات والتي تستخدم طريقة تهذيب النواة في التقدير. القدرات المقدرة بطريقة دلتا المعلمية (DSM-L) لتقدير الدرجات والقدرات المقدرة بطريقة كيرنل(KS) اللامعلمية لتقدير الدرجات.

الأساليب الإحصائية: لتحليل البيانات والتوصل إلى نتائج الدراسة تم استخدام الأساليب الإحصائية التالية:

١. التحليل العاملي الاستكشافي (Exploratory Factor Analysis-EFA) للتحقق من أحادية البعد للبيانات المولدة لكل ظرف من الظروف الاختبارية بالدراسة باستخدام طريقة المكونات الرئيسية (Principal Component Analysis)وتدوير العوامل باستخدام طريقة فاريماكس (Varimax) باستخدام برنامج SPSS وحساب ما يلزم من اشتراطات الأساليب الإحصائية المستخدمة للدراسة. وتم التحقق من افتراض الاستقلال الموضعي باستخدام حزمة (SnowIRT) من برنامج (Jamovi 2.3.28). وكذلك تم التحقق من افتراض الاطرادية من خلال برنامج (Bailog MG 3.0) وذلك عن طريق منحنيات خصائص الفقرة (ICC) لكل فقرة من الفقرات.

٢. تم حساب المتوسطات الحسابية والانحرافات المعيارية والخطأ المعياري للتقدير(S.E.E) لكل ظرف اختباري (تصميم عاملي5X4X2) .

٣. تم تقدير معلمة القدرة وفق النماذج المعلمية (PIRT) بطريقة دلتا عن طريق برنامج الدلتا (DELTA 3.1)، باستخدام الإطار الكامن لأسلوب دلتا لتقدير الدرجات (DSM-L) والتي تستخدم أحد أسلوبي الأرجحية العظمى (المشتركة (JML) أو الهامشية (MML)). وتم تقدير معلمة القدرة وفق نماذج (NIRT) بطريقة كيرنل (KS) اللامعلمية باستخدام برنامج R بوسطة حزمة (KernSmoothIRT)التي طورها ماز وآخرون (Mazza et al., 2014,2020) لتقدير القدرات والتي تستخدم طريقة تهذيب النواة في التقدير.

٤. تم تطبيق تحليل التباين المختلط الثلاثي (Three-Way Mixed- Design ANOVA)، وكذلك ثنائي الاتجاه (Two-Way Mixed- Design ANOVA) (Gliner et al., 2016) باستخدام برنامج R، لدلالة الفروق بين متوسطات مؤشرات دقة التقدير تبعًا لكل من متغيرات الدراسة والتفاعل بينها، وايضًا تم حساب حجم الأثر (Effect Size) للإجابة عن السؤال الأول واختبار فرضية الدراسة الأولى.

نتائج الدراسة ومناقشتها وتفسيرها:

النتائج المتعلقة بالإجابة عن السؤال الأول: "هل تختلف دقة تقدير معلمة القدرة المقدرة باستخدام فقرات ثنائية التدريج تبعاً لطريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية) وحجم العينة (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠) وتوزيع البيانات (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد) والتفاعل بينهم؟"

للإجابة على هذا السؤال تم تقدير القدرة بطريقتي تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، والجدول (٧) يوضح المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري (SEE) لتقدير القدرة وفقاً لطريقة التقدير باختلاف حجم العينة وتوزيع البيانات.

جدول ٧. المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري (SEE) لتقدير القدرة وفقاً لطريقة التقدير باختلاف الحجم وتوزيع البيانات

الحجم	الطريقة	توزيع البيانات									
		اعتدالي		التواء موجب بسيط		التواء موجب شديد		التواء سالب بسيط		التواء سالب شديد	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
100	دلتا	0.102	0.114	0.049	0.015	0.107	0.125	0.096	0.099	0.125	0.115
	كيرنل	0.071	0.022	0.069	0.02	0.08	0.022	0.079	0.023	0.09	0.021
250	دلتا	0.051	0.014	0.097	0.114	0.035	0.015	0.108	0.106	0.115	0.107
	كيرنل	0.065	0.018	0.065	0.018	0.077	0.019	0.065	0.022	0.079	0.021
500	دلتا	0.053	0.015	0.107	0.114	0.027	0.005	0.06	0.014	0.115	0.109
	كيرنل	0.058	0.016	0.063	0.017	0.081	0.02	0.07	0.017	0.076	0.019
1000	دلتا	0.049	0.014	0.04	0.015	0.093	0.105	0.112	0.105	0.115	0.11
	كيرنل	0.056	0.014	0.058	0.014	0.073	0.017	0.065	0.016	0.069	0.017

يتضح من الجدول (٧) أنه بشكل عام، تُظهر طريقة دلتا تبايُنًا أعلى في الخطأ المعياري للتقدير (SEE) مقارنة بطريقة كيرنل عبر جميع أحجام العينات والتوزيعات. بالنسبة لطريقة دلتا، يميل الخطأ المعياري للتقدير (SEE) إلى الزيادة في التوزيعات الملتوية، وخاصة في التوزيعات الملتوية الموجبة والسالبة الشديدة، مع ملاحظة انحرافات معيارية أكبر. على وجه التحديد، بالنسبة لحجم عينة (١٠٠)، تُظهر طريقة دلتا متوسطات وانحرافات معيارية عالية نسبيًا، وخاصة للتوزيعات الملتوية الشديدة، مثل (المتوسط =٠,١٢٥ والانحراف المعياري= ٠,١١٥) عند الالتواء السالب الشديد. مع زيادة حجم العينة، ينخفض الخطأ المعياري للتقدير لطريقة دلتا للتوزيعات الاعتدالية ولكنه يظل مرتفعًا للتوزيعات الملتوية. في المقابل، تعطي طريقة كيرنل غالباً متوسطات وانحرافات معيارية أقل، بغض النظر عن التوزيع أو حجم العينة. حتى في التوزيعات الملتوية، تحافظ طريقة كيرنل على أخطاء معيارية مستقرة ومنخفضة نسبيًا، مثل (المتوسط الحسابي=٠,٠٩ والانحراف المعياري = ٠,٠٢١) عند الالتواء السالب الشديد بحجم عينة ١٠٠. تُظهر طريقة كيرنل أيضًا نمطًا أكثر اتساقًا عبر أحجام العينات المختلفة، مما يشير إلى أنها تعطي أفضل التقديرات عبر أنواع التوزيعات. مع زيادة حجم العينة، تُظهر كلتا الطريقتين اتجاهًا عامًا لتناقص الأخطاء المعيارية للتوزيعات الاعتدالية، وبالتالي تزيد دقة التقدير، لكن طريقة كيرنل تظل أكثر استقرارًا وأقل حساسية للتغيرات في شكل التوزيع وحجم العينة مقارنة بطريقة دلتا.

١. وفي ضوء ما تقدم تم اجراء تحليل التباين المختلط ثلاثي الاتجاه (Three-Way Mixed- Design ANOVA) (Gliner et al., 2016) باستخدام برنامج R، وكذلك ثنائي الاتجاه (Two-Way Mixed- Design ANOVA) لاختبار التفاعلات بين

طرق التقدير وحجم العينة وتوزيع البيانات متبوعًا بحساب حجم التأثير. حيث تم الكشف عن الفروق بين متوسطات الخطأ المعياري للتقدير باختلاف طريقة التقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية) وحجم العينة (١٠٠، ٢٥٠، ٥٠٠، ١٠٠٠) وتوزيع البيانات (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد)، كما تم التحقق من افتراضات تحليل التباين ثلاثي الاتجاه وثنائي الاتجاه باستخدام اختبار ليفين لتجانس التباين (Leven's Test)، واختبار ماوكلي للكروية (Mauchly Test)، واختبار بوكس أم (Box's M Test) لتجانس التباين. في حالة انتهاك افتراض الكروية، تم تطبيق تصحيح جرينهاوس-جايسر(Greenhouse-Geisser) لجميع العوامل داخل الموضوعات(Kassambara, 2023). في حالة انتهاك افتراضات تجانس التباين، تم استخدام نموذج التأثيرات المختلطة الخطي لأنه أكثر مرونة ولديه القدرة على التعامل مع المواقف التي لا يتم فيها تلبية افتراضات معينة حول البيانات بدقة، دون التسبب في أخطاء كبيرة في النتائج مثل انتهاكات تجانس التباين أو تجانس التغاير، خاصة عند التعامل مع التصميمات المعقدة. نظرًا لأن نماذج التأثيرات المختلطة يمكنها أن تأخذ في الاعتبار التباين العشوائي وتتأثر بشكل أقل بانتهاكات الافتراضات (Cooper et al., 2013). واتباعاً لما أوصت به كل من الجمعية الأمريكية لعلم النفس (American Psychological Association, 2020) بأهمية استخدام حجم التأثير(Effect Size-ES) لدعم النتائج والقرارات الناتجة من الدلالة الإحصائية، تم حساب حجم الأثر من خلال مربع إيتا الجزئي المعمم(η^2)، وتم تفسيره حسب المحكات المحددة وتم تفسيره حسب المحكات المحددة في كوهين (Cohen, 1988) ، ويوضح الجدول (٨) نتائج تحليل التباين المختلط الثلاثي الاتجاه لتقديرات القدرة باختلاف طريقة التقدير وحجم العينة وتوزيع البيانات التي تم التوصل اليها.

جدول8. نتائج تحليل التباين المختلط الثلاثي الاتجاه لتقديرات القدرة باختلاف طريقة التقدير وحجم العينة وتوزيع البيانات

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	قيمة اختبار ف "المحسوبة F"	الدلالة الإحصائية p.value	مربع ايتا الجزئي " η^2"
توزيع البيانات	1.419	4	0.355	102.084	P < .001	0.188
الحجم	0.220	3	0.073	21.131	P < .001	0.029
طريقة التقدير	0.355	1	0.355	102.255	P < .001	0.047
توزيع البيانات × الحجم	2.238	12	0.187	53.684	P < .001	0.297
توزيع البيانات× طريقة التقدير	0.749	4	0.187	53.917	P < .001	0.099
الطريقة × الحجم	0.197	3	0.066	18.903	P < .001	0.026
توزيع البيانات × الحجم× طريقة التقدير	2.357	12	0.196	56.531	P < .001	0.313

يوضح الجدول (8) نتائج تحليل التباين المختلط ثلاثي الاتجاهات وتأثيرات ثلاث متغيرات - توزيع البيانات والحجم وطريقة التقدير- إلى جانب تفاعلاتها على المتغير التابع (دقة التقدير). حيث يوجد تأثير دلال احصائيًا لتوزيع البيانات على دقة تقدير معلمة القدرة بطريقتي دلتا المعلمية وكيرنل اللامعلمية لتقدير الدرجات. حيث بلغت قيمة اختبار ف والقيمة الاحتمالية ($P <$ .001, $F\ (4, 18460) =$ ١٠٢,٠٨٤)، مما يشير إلى أن المستويات المختلفة للتوزيع لها تأثير كبير على دقة التقدير. كما تشير قيمة مربع إيتا الجزئي (η^2) البالغة (٠,١٨٨) إلى أنه يمكن إرجاع

حوالي (١٩٪) من التباين الإجمالي في دقة التقدير إلى تأثير توزيع البيانات، والذي يُعتبر حجم تأثير كبير جدًا تبعًا للمحكات المحددة في كوهين (Cohen, 1988).

وبالمثل، يُظهر حجم العينة أيضًا تأثيرًا مهمًا، حيث يوجد تأثير دال احصائيًا لحجم العينة على دقة تقدير معلمة القدرة بطريقتي دلتا المعلمية وكيرنل اللامعلمية لتقدير الدرجات، حيث بلغت قيمة اختبار ف والقيمة الاحتمالية (F (3, 18460) = ٢١,١٣١,P < .001). ومع ذلك، تشير قيمة مربع إيتا الجزئي (η^2) لحجم العينة البالغة (0.029) إلى أن التأثير صغير نسبيًا تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، مما يفسر حوالي (٢,٩٪) من التباين في دقة التقدير.

أما بالنسبة لطريقة التقدير بلغت قيمة اختبار ف والقيمة الاحتمالية (١٠٢,٢٥٥,P < .001 = F (1, 18460)، كما تشير قيمة مربع إيتا الجزئي (η^2) لطريقة التقدير البالغة (0.047) إلى أن التأثير صغير إلى متوسط نسبيًا تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، مما يفسر حوالي(4.7٪) من التباين في دقة التقدير، وهذا يعني أن طريقة التقدير المستخدمة لها تأثير على دقة التقدير.

كما أن تأثيرات التفاعل جديرة بالملاحظة. التفاعل بين توزيع البيانات وحجم العينة مهم جداً، حيث أن قيمة اختبار ف والقيمة الاحتمالية (F (12, 18460) = ٥٣,٦٨٤,P < .001). وتشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (٠,٢٩٧) إلى أنه يفسر (٢٩,٧٪) من التباين في دقة التقدير، وهو تأثير كبير تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، حيث لاحظنا أن توزيع البيانات كان له تأثير كبير ومستقل على دقة التقدير، وعند تفاعله مع حجم العينة كان لهم تأثير كبير جدًا وهذا يعني أن شكل التوزيع يتأثر بحجم العينة وكل نوع من التواء البيانات يكون أكثر دقة في التقدير عند حجم معين. التفاعل بين توزيع البيانات وطريقة التقدير مهم بشكل مماثل حيث بلغت قيمة اختبار ف لهذا التفاعل والدلالة الإحصائية (F (4, 18460) =٥٣,٩١٧,P < .001)، وتشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.099) إلى أنه يفسر (٩,٩٪) من التباين في دقة التقدير مما يشير إلى أن التأثير المشترك لهذه المتغيرات يؤثر بشكل متوسط إلى كبير نسبيًا على دقة التقدير تبعًا للمحكات المحددة في كوهين (Cohen, 1988). التفاعل بين الحجم والطريقة مهم أيضًا، وإن كان بحجم تأثير صغير حيث بلغت قيمة اختبار ف لهذا التفاعل والدلالة الإحصائية (F (1, 18460) = 18.903 ,P < .001)، وتشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.026) إلى أنه يفسر (2.6٪) من التباين في دقة التقدير مما يشير إلى أن التأثير المشترك لهذه المتغيرات يؤثر بشكل صغير نسبيًا على دقة التقدير تبعًا للمحكات المحددة في كوهين (Cohen, 1988).

وأخيرًا، فإن التفاعل الثلاثي بين توزيع البيانات وحجم العينة وطريقة التقدير مهم جداً، حيث بلغت قيمة اختبار ف لهذا التفاعل الثلاثي والدلالة الإحصائية (F (12, 18460)= ,P < .001 56.531)، وتشير القيمة المرتفعة جدًا لمربع إيتا الجزئي (η^2) لهذا التفاعل الثلاثي والبالغة (0.313)، إلى أنه يفسر(31.3٪) من التباين في دقة التقدير مما يشير إلى أن التأثير المشترك لهذه المتغيرات كبير جدًا على دقة التقدير تبعًا للمحكات المحددة في كوهين (Cohen, 1988) . كما أن التأثير المشترك لجميع المتغيرات الثلاثة يلعب دورًا مهمًا في تحديد النتيجة. وبشكل عام، تُظهر النتائج تفاعلات معقدة بين المتغيرات، مع بعض أحجام التأثير المعتدلة التي تفسر التباين في دقة التقدير عبر ظروف مختلفة.

وتم اجراء تحليل التباين الثنائي لمقارنة الأحجام والطرق المختلفة ضمن التوزيع الاعتدالي، ويوضح الجدول (9) العديد من النتائج المهمة فيما يتعلق بالتأثيرات الرئيسية وتفاعلها.

جدول 9. نتائج تحليل التباين الثنائي الاتجاه لتقديرات القدرة باختلاف طريقة التقدير وحجم العينة

توزيع اعتدالي						
مصدر التباين	مجموع المربعات	متوسط مجموع المربعات	درجات الحرية	قيمة اختبار ف المحسوبة" F"	الدلالة الإحصائية	مربع ايتا الجزئي
الحجم	0.101	0.034	3	93.805	<0.001	0.0314
طريقة التقدير	0.001	0.001	1	1.607	0.205	P < .0012
الحجم×طريقة التقدير	0.076	0.025	3	71.084	<0.001	0.0238
التواء موجب بسيط						
المتغيرات	**مجموع المربعات**	**متوسط مجموع المربعات**	**درجات الحرية**	**قيمة اختبار ف المحسوبة" F"**	**الدلالة الإحصائية**	**مربع ايتا الجزئي**
الحجم	0.778	0.259	3	107.641	<0.001	0.0393
طريقة التقدير	0.042	0.042	1	17.437	<0.001	0.0021
الحجم×طريقة التقدير	0.753	0.251	3	104.269	<0.001	0.0381
التواء موجب شديد						
المتغيرات	**مجموع المربعات**	**متوسط مجموع المربعات**	**درجات الحرية**	**قيمة اختبار ف المحسوبة" F**	**الدلالة الإحصائية**	**مربع ايتا الجزئي**
الحجم	0.777	0.259	3	72.182	<0.001	0.0273
طريقة التقدير	0.071	0.071	1	19.763	<0.001	0.0025
الحجم×طريقة التقدير	1.141	0.38	3	105.971	<0.001	0.04
التواء سالب بسيط						
المتغيرات	**مجموع المربعات**	**متوسط مجموع المربعات**	**درجات الحرية**	**قيمة اختبار ف المحسوبة" F**	**الدلالة الإحصائية**	**مربع ايتا الجزئي**
الحجم	0.259	0.086	3	25.167	<0.001	0.0095
طريقة التقدير	0.274	0.274	1	79.715	<0.001	0.01
الحجم×طريقة التقدير	0.57	0.19	3	55.339	<0.001	0.0208
التواء سالب شديد						
المتغيرات	**مجموع المربعات**	**متوسط مجموع المربعات**	**درجات الحرية**	**قيمة اختبار ف المحسوبة" F**	**الدلالة الإحصائية**	**مربع ايتا الجزئي**
الحجم	0.038	0.013	3	2.378	0.068	0.002
طريقة التقدير	0.717	0.717	1	133.833	<0.001	0.03
الحجم×طريقة التقدير	0.014	0.005	3	0.871	0.455	P < .001611

من الجدول (٩) وفي حالة التوزيع الاعتدالي، تبين تأثير الحجم مهم جداً، كما هو موضح من خلال قيمة اختبار ف ومستوى الدلالة (P< 0.001, ٩٣,٨٠٥= (3,1846) F)، يعكس متوسط مجموع مربعات الحجم البالغ (٠,٠٣٤) قدرًا كبيرًا من التباين الذي يفسره هذا المتغير، وتدعم قيمة اختبار ف أيضًا أن الحجم له تأثير قوي على دقة التقدير. وتشير قيمة مربع إيتا الجزئي التي بلغت

(٠,٠٣١٤) إلى أنه يمكن إرجاع ما يقارب من (٣,١٤٪) من التباين في دقة التقدير إلى الاختلافات بين حجم العينة. على الرغم من أن هذا قد يبدو صغيرًا تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، إلا أنه يمثل تأثيرًا ذا مغزى في سياق الدراسة.

على النقيض من ذلك، فإن تأثير طريقة التقدير ليس مهمًا إحصائيًا. تشير قيمة اختبار ف والدلالة الإحصائية (F (1,1846) =١,٦٠٧ ,P= 0.205) إلى عدم وجود دلاله على أن طريقة التقدير لها تأثير على دقة التقدير. كما أن مربع إيتا الجزئي للطريقة صغير ايضاً حيث بلغ (P < 2.001)، مما يعني أن نسبة صغيرة جدًا (٠,٠٢%) من التباين يمكن تفسيرها بالطرق المختلفة للتقدير تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، مما يجعلها غير ذات أهمية عمليًا.

وأخيرًا، فإن التفاعل بين الحجم وطريقة التقدير مهم جداً، قيمة اختبار ف ومستوى الدلالة (F (3,1846) =٧١,٠٨٤ ,P< 0.001). وهذا يشير إلى أن تأثير الحجم على دقة التقدير يختلف وفقًا للطريقة المستخدمة، وأن هذا التفاعل يفسر قدرًا من التباين. كما أن مربع إيتا الجزئي للتفاعل بلغ (٠,٠٢٣٨)، مما يعني أنه يمكن عزو(٢,٣٨%) من التباين في دقة التقدير إلى التفاعل بين الحجم وطريقة التقدير، وهي نسبة صغيرة تبعًا للمحكات المحددة في كوهين (Cohen, 1988).

وفي الختام، في حين أن حجم العينة له تأثير كبير على النتيجة، فإن طريقة التقدير وحدها لا يكون لها تأثير بنفس القدر. ومع ذلك، فإن الجمع بين حجم العينة وطريقة التقدير يُنتج تأثير تفاعلي كبير، مما يُظهر أن تأثير حجم العينة يختلف اعتمادًا على طريقة التقدير المستخدمة.

وفي حالة الالتواء الموجب البسيط، تشير نتائج الجدول (١٧) لتحليل التباين ثنائي الاتجاه إلى تأثيرات كبيرة لحجم العينة وطريقة التقدير وتفاعلهما. يظهر متغير الحجم تأثيرًا كبيرًا للغاية على دقة التقدير، والذي ينعكس من خلال قيمة اختبار ف ومستوى الدلالة (١٠٧,٦٤١ ,P< 0.001 = (3,1846) F). كما أن متوسط مجموع مربعات الحجم (٠,٢٥٩) كبير وهذا يؤكد أن الحجم يفسر جزءًا كبيرًا من التباين. ويشير مربع إيتا الجزئي البالغ (٠,٠٣٩٣) إلى أن(٣,٩٣٪) من التباين يمكن تفسيره من خلال الاختلافات في حجم العينة، وهي نسبة صغيرة إلى متوسطة تبعًا للمحكات المحددة في كوهين (Cohen, 1988). إذن حجم العينة يلعب دورًا رئيسيًا في التأثير على دقة التقدير داخل هذا التوزيع.

بالنسبة لطريقة التقدير مهمة إحصائيًا أيضًا، من خلال قيمة اختبار ف ومستوى الدلالة (F (1,1846) = ١٧,٤٣٧ ,P< 0.001). على الرغم من أن متوسط مجموع مربعات طريقة التقدير (٠,٠٤٢) أصغر مقارنة بحجم العينة، إلا أنه لا يزال له تأثير كبير. كما أن مربع إيتا الجزئي للطريقة هو (٠,٠٠٢١)، مما يعني أن (٠,٢١%)من التباين يمكن تفسيره من خلال الاختلافات في طريقة التقدير. وفي حين أن هذا التأثير صغير تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، إلا أنه ذو دلالة إحصائية ويشير إلى أن طريقة التقدير المستخدمة لها بعض التأثير على دقة التقدير.

أيضاً نجد أن التفاعل بين حجم العينة وطريقة التقدير ذو دلالة إحصائية عالية أيضًا، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,1846) = ١٠٤,٢٦٩ ,P< 0.001). كما يُشير متوسط مجموع مربعات التفاعل (٠,٢٥١) إلى أن تأثير التفاعل هذا يفسر جزءًا ملحوظًا من التباين. يوضح مربع إيتا الجزئي البالغ (٠,٠٣٨١) أن(٣,٨١%) من التباين يمكن أن يُعزى إلى التفاعل بين حجم العينة وطريقة التقدير، وهي نسبة بسيطة إلى متوسطة تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، مما يشير إلى أن تأثير الحجم يتأثر بطريقة التقدير المستخدمة.

باختصار، تُظهر النتائج أن كلًا من حجم العينة وطريقة التقدير يؤثران بشكل كبير على النتيجة في حالة الالتواء الموجب البسيط، حيث يكون للحجم تأثير أكبر. بالإضافة إلى ذلك، يشير

التفاعل الكبير بين الحجم وطريقة التقدير إلى أن تأثير حجم العينة يختلف حسب طريقة التقدير المطبقة.

في حالة الالتواء الموجب الشديد، تكشف نتائج تحليل التباين ثنائي الاتجاه عن تأثيرات كبيرة لحجم العينة وطريقة التقدير وتفاعلهما. يؤثر حجم العينة بشكل كبير على دقة التقدير، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,3692) = ٧٢,١٨٢ ,P< 0.001). كما بلغت قيمة متوسط مجموع المربعات (٠,٢٥٩)، والذي يُظهر تأثيرًا قويًا لحجم العينة على تفسير التباين. كما يشير مربع إيتا الجزئي البالغ (٠,٠٢٧٣) إلى أن(٢,٧٣%)من التباين يمكن تفسيره باختلافات في حجم العينة، وهي نسبة صغيرة إلى متوسطة تبعًا للمحكات المحددة في كوهين (Cohen, 1988). هذا التأثير، على الرغم من كونه معتدلاً، إلا أنه ملحوظ ويسلط الضوء على أهمية حجم العينة في التأثير على دقة التقدير.

بالنسبة لطريقة التقدير مهمة إحصائيًا أيضًا، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (1,3692) = ١٩,٧٦٣ ,P< 0.001). وتشير قيمة متوسط مجموع المربعات البالغة (٠,٠٧١) إلى أنها تساهم في تفسير التباين. كما أن قيمة مربع إيتا الجزئي بلغ (٠,٠٠٢٥)، مما يعني أن (٠,٢٥%) من التباين يمكن تفسيره بطريقة التقدير. هذا التأثير صغير ولكنه مهم تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، مما يشير إلى أن طريقة التقدير لها تأثير بسيط ولكن يمكن ملاحظته على النتيجة.

التفاعل بين حجم العينة وطريقة التقدير مهم جداً حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,3692) = ١٠٥,٩٧١ ,P< 0.001). كما بلغت قيمة متوسط مجموع المربعات للتفاعل هو (٠,٣٨)، وتوضح قيمة اختبار ف المرتفعة تأثير تفاعل قوي. كما يوضح مربع إيتا الجزئي البالغ (٠,٠٤) أن (٤%) من التباين يمكن تفسيره بالتفاعل بين حجم العينة وطريقة التقدير، وهذا تأثير صغير إلى متوسط تبعًا للمحكات المحددة في كوهين (Cohen, 1988). يشير هذا إلى أن تأثير حجم العينة يتأثر بشكل كبير بطريقة التقدير.

باختصار، تشير النتائج إلى أن كل من الحجم وطريقة التقدير يؤثران بشكل كبير على النتيجة في حالة الالتواء الموجب الشديد، حيث يكون لحجم العينة تأثير كبير إلى حد ما. إن التفاعل بين حجم العينة وطريقة التقدير قوي بشكل خاص، مما يشير إلى أن تأثير حجم العينة يعتمد بشكل كبير على طريقة التقدير المستخدمة.

في حالة الالتواء السالب البسيط، تم استخدام تحليل التباين ثنائي الاتجاه لتقييم تأثيرات الحجم وطريقة التقدير وتفاعلهما على دقة التقدير. يُظهر حجم العينة تأثيرًا مهمًا، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,1846) = ٢٥,١٦٧ ,P< 0.001). كما بلغت قيمة متوسط مجموع مربعات الحجم هو (٠,٠٨٦)، مما يشير إلى أن الاختلافات في حجم العينة يؤثر على دقة التقدير. وبلغت قيمة مربع إيتا الجزئي للحجم هو (٠,٠٠٩٥)، مما يشير إلى أن (٠,٩٥٪) من التباين في دقة التقدير يتم تفسيره بواسطة حجم العينة. في حين أن هذا تأثير صغير تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، فإنه يشير إلى أن الحجم لا يزال يلعب دورًا في التأثير على النتائج، على الرغم من أن مساهمته في التباين الكلي طفيفة نسبيًا.

أما بالنسبة لطريقة التقدير فإن لها تأثير مهم للغاية أيضًا، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (1,1846) = ٧٩,٧١٥ ,P< 0.001). كما بلغت قيمة متوسط مجموع المربعات (٠,٢٧٤)، مما يشير إلى تأثير كبير لطريقة التقدير على تفسير التباين. وبلغت قيمة مربع إيتا الجزئي للطريقة (٠,٠١)، مما يشير إلى أن (١%) من التباين يتم تفسيره بواسطة طريقة

التقدير. وهذا يمثل تأثيرًا صغيرًا ولكنه ملحوظ، مما يدل على أن طريقة التقدير تؤثر بشكل هادف على دقة التقدير في هذا التوزيع.

التفاعل بين حجم العينة وطريقة التقدير مهم للغاية أيضًا، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,1846) = ٥٥,٣٣٩ ,P< 0.001). كما بلغت قيمة متوسط مجموع المربعات التفاعل هو (٠,١٩). وبلغت قيمة مربع إيتا الجزئي للتفاعل (٠,٠٢٠٨)، مما يشير إلى أن(٢,٠٨%) من التباين في دقة التقدير يمكن أن يعزى إلى التفاعل بين حجم العينة وطريقة التقدير، وهذا تأثير صغير تبعًا للمحكات المحددة في كوهين (Cohen, 1988). وهذا يشير إلى أن تأثير الحجم يعتمد على طريقة التقدير المستخدمة، ويلعب هذا التفاعل دورًا مهمًا إلى حد ما في تفسير التباين في دقة التقدير.

باختصار، تشير نتائج الالتواء السالب البسيط إلى أن كلًا من حجم العينة وطريقة التقدير يؤثران بشكل كبير على دقة التقدير، حيث يكون لطريقة التقدير تأثير أقوى قليلاً من حجم العينة. كما يلعب التفاعل بين الحجم وطريقة التقدير دورًا حاسمًا، حيث يفسر جزءًا معتدلاً من التباين. تسلط هذه النتائج الضوء على أن تأثير حجم العينة على دقة التقدير يتأثر بطريقة التقدير المستخدمة، مما يؤكد على أهمية النظر في كلا العاملين معًا.

في حالة الالتواء السالب الشديد، تم استخدام تحليل تباين ثنائي الاتجاه لفحص تأثيرات الحجم وطريقة التقدير وتفاعلهما على دقة التقدير. اقترب متغير حجم العينة من الأهمية حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,1846) =٢,٣٧٨ ,P= 0.068)، وهي أكبر قليلاً من (α =٠٥.). كما بلغت قيمة متوسط مجموع المربعات (٠,٠١٣). وبلغت قيمة مربع ايتا الجزئي للحجم (٠,٠٠٢)، مما يشير إلى أن حجم العينة يمثل (٠,٢%) من التباين في دقة التقدير وهذا تأثير صغير جدًا تبعًا للمحكات المحددة في كوهين (Cohen, 1988). كما يشير حجم التأثير الصغير هذا إلى أنه على الرغم من أن الحجم قد يكون له بعض التأثير على دقة التقدير، إلا أنه ليس مساهمًا رئيسيًا في هذا التوزيع، والأدلة ليست قوية بما يكفي للقول بأن له تأثير كبير عند مستوى (٠,٠٥).

من ناحية أخرى، فإن طريقة التقدير لها تأثير كبير جداً، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (1,1846) =١٣٣,٨٣٣ ,P< 0.001). كما بلغت قيمة متوسط مجموع المربعات (٠,٧١٧)، مما يشير إلى تأثير قوي على دقة التقدير. وبلغت قيمة مربع ايتا الجزئي للطريقة (٠,٠٣)، مما يعني أن(٣%) من التباين في دقة التقدير يتم تفسيره بواسطة متغير طريقة التقدير. ويمكن اعتباره تأثيرًا كبيراً وأكثر وضوحًا مقارنة بتأثير حجم العينة ويُظهر أن طريقة التقدير هي متنبئ مهم في هذا التوزيع. لم يكن التفاعل بين حجم العينة وطريقة التقدير مهمًا، حيث بلغت قيمة اختبار ف ومستوى الدلالة (F (3,1846) =٠,٨٧١ ,P= 0.455)، وهي أكبر من (α =٠,٠٥). كما بلغت قيمة متوسط مجموع المربعات (٠,٠٠٥). وبلغت قيمة مربع ايتا الجزئي للتفاعل (٦١١P < 001.)، مما يشير إلى أن التفاعل يفسر نسبة بسيطة جداً ولا تُذكر من التباين في دقة التقدير تبعًا للمحكات المحددة في كوهين (Cohen, 1988). تشير هذه النتيجة إلى أن تأثير الحجم لا يعتمد على طريقة التقدير المستخدمة، حيث لا يساهم التفاعل بين هذين العاملين بشكل كبير في تفسير التباين في دقة التقدير. وباختصار، يشير تحليل توزيع الالتواء السالب الشديد إلى أن طريقة التقدير لها تأثير قوي ومهم على دقة التقدير، في حين يظهر الحجم تأثيرًا صغيرًا غير مهم. والتفاعل بين الحجم وطريقة التقدير ليس مهمًا ويفسر القليل جدًا من التباين، مما يشير إلى أن هذين العاملين لا يتفاعلان بطريقة ذات مغزى في هذه المجموعة من البيانات. والنتيجة الرئيسية هي التأثير المهيمن للطريقة على النتيجة.

التوصيات: بالرغم من أهمية النتائج التي توصل إليها البحث الحالي في الكشف عن دقة تقدير قدرة الأفراد باستخدام نظرية الاستجابة للمفردة وفق لطريقة دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية ، إلا ان هناك بعض العوامل المؤثرة الأخرى قد تلعب دورا مهما في دقة تقدير قدرات الأفراد ولم يتم ضبطها في البحث الحالي مثل تقصي دقة التقدير في اطوال اختبارات أكثر تباين ونسبة البيانات المفقودة مما قد يسهم في اثراء النظرية الحديثة في القياس، وبشكل عام توصي الباحثة بما يلي:

١. استخدام عينات ذات أحجام مشابهة للأحجام المستخدمة في الدراسة الحالية عند طريقة كيرنل ونموذج دلتا لتقدير الدرجات في الدراسات المستقبلية لتحليل مفردات المقاييس. فقد أظهرت نتائج الدراسة الحالية جودة مطابقة البيانات بشكل عام سواء للأفراد أو المفردات لهذين النموذجين باختلاف حجم العينة.
٢. زيادة حجم العينة يؤدي إلى تقديرات دقيقة لمعالم الفقرة والفرد لذلك يمكن اعتبار زيادة حجم العينة عامل إيجابي في دقة التقدير وخفض الخطأ المعياري للتقدير.
٣. استخدام نماذج نظرية الاستجابة للمفردة يحتاج إلى التحقق من العديد من الافتراضات على عكس استخدام أي نموذج من نماذج دلتا لذا توصي الباحثة باستخدام نموذج دلتا لتقدير الدرجات (نموذج الدالة النسبية ثنائي المعلم وفق الإطار الكامن) في تحليل المفردات الاختبارية. حيث أشارت نتائج الدراسة الحالية إلى عدم وجود فروق جوهرية بين المعالم المقدرة في ضوء طريقة كيرنل اللامعلمية ونموذج دلتا لتقدير الدرجات.
٤. واتباعاً لما أوصت به كل من الجمعية الأمريكية لعلم النفس (APA) بأهمية استخدام حجم التأثير (Effect Size-ES) لدعم النتائج والقرارات الناتجة من الدلالة الإحصائية، تم حساب حجم الأثر من خلال مربع إيتا الجزئي المعمم(η^2)، توصي الباحثة بضرورة اهتمام الباحثين بفحص وإدراج حجم التأثير (ES) ضمن نتائج الدراسات، لما له من أهمية عملية في إيضاح أهمية النتائج التي تم التوصل إليها وتفسير معناها بشكل شامل للوصول لفهم أعمق للظواهر موضع الدراسة.
٥. استخدام البيانات المولدة بالمحاكاة يساهم في التخلص من تأثير بعض العوامل على استجابات الأفراد مثل الغش والتخمين، كما تساعد في ضبط الموقف الاختباري من خلال معالجة مشكلة عدم الجدية واللامبالاة، وترك بعض الفقرات وغيرها. كما يمكنها تقديم نتائج نظرية يمكن الاعتماد عليها ميدانياً وتسهل جمع البيانات والمعلومات، وهي عملية واقتصادية في التكلفة والوقت.

المقترحات والدراسات المستقبلية: نتيجة لما وجدته الباحثة أثناء إجراء الدراسة الحالية من جوانب تحتاج إلى لفت النظر إليها والاهتمام بها، يمكن تقديم المقترحات التالية:

١. دراسة أثر طرق التقدير باستخدام النماذج اللوجستية الثنائية أو المتعددة الاستجابة الأخرى على دقة تقدير قدرة الأفراد ومعالم الفقرة.
٢. دراسة أثر طرق التقدير على دقة تقدير قدرة الأفراد ومعالم الفقرة تحت ظروف مختلفة من أشكال التوزيع، كالتوزيع المنتظم والتوزيع بيتا والتوزيع الأسي وغيرها.
٣. اجراء دراسات مشابهة ولكن باستخدام متغيرات أخرى مثل أطوال مختلفة من الاختبارات.
٤. إجراء المزيد من الدراسات تركز على توزيع البيانات ومقارنة التوزيع الاعتدالي بالتوزيعات الملتوية التي تعبر عن مستويات الأفراد المنخفضة والمرتفعة، لما كان لها من أثر على نتائج الدراسة من مطابقة الفقرات ودقة التقدير.

٥. إجراء المزيد من الدراسات التجريبية على نماذج دلتا لتقدير الدرجات باستخدام عينات صغيرة الحجم للكشف عن كيفية عمل هذه النماذج.

المراجع:

البادية، فاطمة حمد، ابن كاظم، علي بن مهدي، والمحرزي، راشد بن سيف (٢٠١٨). أثر حجم العينة على دقة تقدير خصائص المفردة والقدرة في اختبار التنمية المعرفية في مادة العلوم لطلبة الصف السابع بسلطنة عمان. *دراسات، ٧٣*، ١٠٦-١٢٥. http://search.mandumah.com/Record/947730

بني عطا، زايد، والشريفين، نضال (٢٠١٢). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. *المجلة الأردنية في العلوم والتربية، ٨*(٢)،١٥١-١٦٦. http://search.mandumah.com/Record/444491

بني عطا، زايد صالح (٢٠١٤). تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة: دراسة محاكاة. *مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية، ١١* (٢)، ١-٣٧. http://search.mandumah.com/Record/810211

الحواري، أروى. (٢٠١٥). أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وفق نموذج راش في نظرية استجابة الفقرة. *مجلة جامعة النجاح للأبحاث - العلوم الإنسانية، ٢٩* (٨)، ١٤٦٣- ١٤٨٨. http://search.mandumah.com/Record/931120

الخرشة، طه عقلة (٢٠١٨). أثر حجم العينة وشكل توزيع قدرة المفحوصين على معالم الفقرة عند استخدام النموذج اللوغارتيمي الثلاثي. *مجلة العلوم التربوية والنفسية، ١٩* (٤)، ٣٩٧-٤٢٤. http://search.mandumah.com/Record/948502

دي إيالا، آر. جي (٢٠١٧). *النظرية والتطبيق في نظرية الاستجابة* للفقرة (عبد الله الكيلاني وإسماعيل البرصان، مترجمان). دار جامعة الملك سعود للنشر. (نُشر العمل الأصلي في ٢٠٠٩).

الشافعي، محمد محمود (٢٠٢٤). تأثير حجم العينة على دقة تقديرات: معامل صعوبة فقرات الاختبار، ومعامل القدرة للأفراد في عينة التحليل باستخدام النموذج اللوجستي ثلاثي المعامل. *مجلة الإسكندرية التربوية، ٣٤* (٢)، ١٩١-٢١٠. https://dx.doi.org/10.21608/jealex.2024.371247

الشرفاوي، عبد الكريم (٢٠٢٢). *طرق تقدير معالم الفقرة والقدرة وأثرها في دقة التقدير باستخدام نظرية الاستجابة للمفردة راش أنموذجًا* [رسالة دكتوراة غير منشورة]. جامعة الجزائر.

الشريفين، نضال كمال (٢٠٠٦). الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي. *مجلة العلوم التربوية والنفسية، ٧* (٤)، ٧٩-١٠٩. http://search.mandumah.com/Record/2445

الشريفين، نضال كمال (٢٠١٢). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. *المجلة التربوية، ٢٦* (١٠٤)، ١٧٧ -٢٣٨-. http://search.mandumah.com/Record/415514

الشريفين، نضال كمال، ومناصرة، سوسن عاطف. (٢٠١٧). خصائص توزيع قدرات الأفراد ومعالم فقرات الإختبارالاختبار وفق نماذج نظرية الإستجابةالاستجابة للفقرة المعملية

واللامعلمية: دراسة مقارنة. *المجلة التربوية*،٣١ (١٢٤)، ٢٦٥-٣٢٠.
http://search.mandumah.com/Record/832513
شما، يمان نزار (٢٠١٣). أثر حجم العينة على دقة تقدير صعوبة المفردات وقدرة الأفراد باستخدام نموذج راش. *مجلة الآداب*، ١٠٥، ٦٧٣-٦٩٨.
http://search.mandumah.com/Record/666339
الشواورة، شادي يوسف. (٢٠١٧). الفاعلية النسبية لطريقة بييز البارامترية وطريقة تهذيب النواة اللابارامترية في تقدير معلمة القدرة وفق النموذج اللوجستي الثنائي باختلاف حجم العينة. *العلوم التربوية*، ٢٥ (٤)، ٣٣٨-٣٦١.
http://search.mandumah.com/Record/918216
ضعضع، هبة عبد اللطيف، طومان، منار أحمد، وطيفور، مصطفى أحمد (٢٠٢٠). أثر حجم العينة وطرائق التقدير في دقة تقدير معالم نموذج راش. *جرش للبحوث والدراسات*،٢١ (١)، ١٣١ -١٧٠http://demo.mandumah.com/Record/1060646
القيسي، حسين عبد النبي. (٢٠١٣). *دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار: دراسة مقارنة* [رسالة دكتوراة، جامعة اليرموك]. دار المنظومة.
http://search.mandumah.com/Record/729749
القيسي، حسين عبد النبي (٢٠١٦). أثر حجم العينة وطول الاختبار على دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية الاستجابة للفقرة اللامعلمية. مجلة مؤتة للبحوث والدراسات، سلسلة العلوم الإنسانية والاجتماعية، ٣١(٥)، ٢٠٣-٢٤٦.
http://search.mandumah.com/Record/785056
Alwan, A. M., & Jasim, K. J. (2022). The Effect of the Difference in the Distribution of the Level of Ability that is Skeweded Positive for the Parameters of the Items of the Mental Ability Test According to the Item Response Theory. *International Journal of Early Childhood Special Education, 14*(1), 1150-1160. DOI:10.9756/INT-JECSE/V14I1.221131
American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.
Browne, M. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & Long, J. S. (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park.
Christensen, K., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q_3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied psychological measurement*, *41*(3), 178–194. https://doi.org/10.1177/0146621616677520
Cooper, J. N., Sulkowski, J. P., Deans, K. J., & Minneci, P. C. (2013). Approaches to estimate between- and within-subject correlation

coefficients in longitudinal repeated-measures studies. *JSM Proceedings, Section on Statistics in Epidemiology, 3323–3333*.

DeMars, C. (2010). *Item Response Theory: Understanding Statistics Measurement*. Oxford University Press.

Dimitrov, D. (2016). An approach to scoring and equating tests with binary items Piloting with large-scale assessments. *Educational and Psychological Measurement*, *76*(6), 954-975. DOI:10.1177/0013164416631100

Dimitrov, D (2018). The Delta Scoring method of tests with binary items: A note on true score estimation and equating. *Educational and Psychological Measurement, 78*(5), 805-825. DOI:10.1177/0013164417724187

Dimitrov, D (2023). *D-scoring method of measurement: Classical and latent frameworks* (1st ed.). Routledge.

Dimitrov, D. M., & Atanasov, D. V. (2021). Latent D-Scoring modeling: Estimation of item and person parameters. *Educational and Psychological Measurement,81*(2), 388–404. https://doi.org/10.1177/0013164420941147

Dimitrov, D. M., Atanasov, D. V., & Luo, Y. (2020). Person-Fit Assessment under the D-scoring Method. *Measurement: Interdisciplinary Research and Perspectives*, *18*(3), 111–123. https://doi.org/10.1080/15366367.2020.1725733

Domingue, B., & Dimitrov, D. (2015). A comparison of IRT theta estimates and delta scores from the perspective of additive conjoint measurement. *National Centre for Assessment in Higher Education.* Saudi Arabia.

Effatpanah, F., & Baghaei, P. (2022). Exploring rater quality in rater-mediated assessment using the non-parametric item characteristic curve estimation. *Psychological Test and Assessment Modeling, 64*(3), 216-252.

Geerlings, H., Laros, J., Tellegen, P., & Glas, C. (2014). Testing the difficulty theory of the SON-R 5(1/2)-17, a non-verbal test of intelligence. *The British journal of mathematical and statistical psychology, 67* (2), 248-65 .

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2016). *Research methods in applied settings: An integrated approach to design and analysis* (3th edi.). Routledge

Guo, F. (2006). Expected classification accuracy using the latent distribution. *Practical Assessment, Research & Evaluation, 11*(6), 1-9. https://doi.org/10.7275/bxba-7466

Hambleton, R. K., & Slater, S. C. (2011). Item response theory models and testing practices: Current international status and future directions. *European Journal of Psychological Assessment, 13*(1), 21–28. https://doi.org/10.1027/1015-5759.13.1.21

Hambleton, R., & Swaminathan, H. (2010). *Item response theory: Principles and applications*. Kluwer-Nijhoff.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). *Fundamentals of item response theory*. Sage Publications.

Han, K. T. (2007). WinGen: Windows software that generates item response theory parameters and item responses. *Applied Psychological Measurement, 31*(5), 457–459. https://doi.org/10.1177/0146621607299271

Hu, L., & Bentler, P. M. (2009). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hulin, C., Drasgow, F. & Parsons, C. (1983). *Item Response Theory: Application to Psychological Measurement*. Homewood IL: Dow Jones Irwin.

Kassambara, A. (2023). rstatix: Pipe-Friendly Framework for Basic Statistical Tests (Version 0.7.2) [R package]. https://rpkgs.datanovia.com/rstatix/

Lei, P.-W., Dunbar, S. B., & Kolen, M. J. (2004). A Comparison of Parametric and Nonparametric Approaches to Item Analysis for Multiple-Choice Tests. *Educational and Psychological Measurement, 64*(4), 565–587. https://doi.org/10.1177/0013164403261760

Maydeu-Olivares, A., & Joe, H. (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2^n Contingency Tables: A Unified Framework. *Journal of the American Statistical Association*, *100*(471), 1009–1020. https://doi.org/10.1198/016214504000002069

Mor Dirlik, E. & Koç, N. (2019). The comparison of item and ability estimations calculated from the parametric and non-parametric item response theory according to the several factors. *Elementary*

Education Online, 18 (3),1016-1035. https://ilkogretim-online.org/index.php/pub/article/view/1342

Nozawa, Y. (2008). *Comparison of parametric and nonparametric IRT equating methods under the common-item non-equivalent groups design*. [Doctoral dissertation, University of Iowa]. University of Iowa.

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27*(4),289-298. DOI:10.1177/0146621603027004004

Ramsay, J. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56*(4), 611-630. https://doi.org/10.1007/BF02294494

Raykov, T., Marcoulides, G. (2016). On the Relationship between Classical Test Theory and Item Response Theory: From One to the Other and Back. *Educational and Psychological Measurement*, *76* (2), 325-338. DOI:10.1177/0013164415576958

Reisenzein, R. (1986). A structural equation analysis of Weiner's attribution-affect model of helping behavior. *Journal of Personality and Social Psychology*, *50*(6), 1123-1133. DOI:10.1037/0022-3514.50.6.1123

Reise, S. P., & Revicki, D. A. (Eds.). (2015). *Handbook of item response theory modeling: Applications to typical performance assessment.* Routledge/Taylor & Francis Group.

Robitzsch, A. (2021). About the equivalence of the latent D-scoring model and the two-parameter logistic item response theory model. *Mathematics ,9*(13). https://doi.org/10.3390/math9131465

Sijtsma, K. & Molenaar, I. (2002). *Introduction to nonparametric item response modeling*. Sage Publications, Inc.

Téllez, A., García, C., & Corral-Verdugo, V. (2015). Effect size, confidence intervals and statistical power in psychological research. *Psychology in Russia: State of the Art, 8*(3), 27-46. DOI:10.11621/pir.2015.0303

Van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties* [Comparability of individual test performance]. Lisse, the Netherlands: Swets & Zeitlinger.

Zickar, M. (1997). *Identifying Untraited Individuals Using Model based Measurement* [PhD thesis] .University of Illinois at Urbana-Champaign https://hdl.handle.net/2142/82216