



جامعة المنصورة
كلية التربية

—



الفروق بين طريقة دلتا المعلمية لتقدير الدرجات وطريقة كيرنل الالمعلمية في دقة تقدير معلمة القدرة

إعداد:

أ/ هند عبد الله سليمان الجاسر

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

د/ إقبال زين العابدين درندري

أستاذ مشارك القياس والتقويم والإحصاء، قسم علم النفس، كلية التربية،
جامعة الملك سعود، المملكة العربية السعودية

مجلة كلية التربية – جامعة المنصورة
العدد 129 – يناير 2025م

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

أ/ هند عبد الله سليمان الجاسر¹

د/ إقبال زين العابدين درندري²

المستخلص باللغة العربية:

هدفت الدراسة الحالية إلى الكشف عن الفروق في دقة تقدير معلمة القدرة تبعاً لطريقة تقدير الدرجات (طريقة دلتا بالإطار الكامن لتقدير الدرجات **Delta Scoring Method- Latent Framework -DSM-L** وطريقة تهذيب النواة كيرنل **KS- Kernel Smoothing** اللامعلمية)، وحجم العينة (100-250-500-1000)، وتوزيع البيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، والتفاعل بينها؛ من خلال استخدام بيانات مُولدة (بأسلوب المحاكاة)، وفق طريقة مونت كارلو (**MCM**) باستخدام (50) فقرة ثنائية التدرج (0,1). وقد تم الحكم على دقة التقدير باستخدام مؤشر الخطأ المعياري للتقدير (**SEE**). وتوصلت نتائج الدراسة إلى ما يلي: وجود فروق دالة إحصائية ($P < .05$) في دقة تقدير معلمة القدرة المقدرة تبعاً لحجم العينة وتوزيع البيانات، وعدم وجود فروق دالة إحصائية ($P > .05$) في دقة تقدير معلمة القدرة تبعاً لطريقة تقدير الدرجات (دلتا **DSM-L** المعلمية وكيرنل **KS** اللامعلمية)، ووجود فروق دالة إحصائية ($P < .05$) في التفاعل بين كل من طريقة التقدير وحجم العينة وتوزيع البيانات في دقة تقدير معلمة القدرة المقدرة. وتوضح النتائج أنه لا توجد فروق دالة إحصائية ($P > .05$) في دقة التقدير تبعاً للطريقة في التوزيع الاعتدالي، ولكن ظهرت فروق دالة إحصائية ($P < .05$) في التوزيعات الملتوية وكان لحجم العينة تأثير في معظم التوزيعات. وبناء على النتائج، توصي الدراسة باختيار الطريقة المناسبة حسب حجم العينة وتوزيع البيانات عند تقدير القدرة للوصول إلى أدق التقديرات. فمع العينات الصغيرة والبيانات الملتوية، وخاصة الموجبة والسالبة الشديدة، يفضل استخدام طريقة كيرنل (**KS**)، ومع العينات الكبيرة والبيانات الاعتدالية، يوصى باستخدام طريقة دلتا (**DSM-L**)، نظراً لدقة تقديراتها، وعدم حاجتها للتحقق من العديد من الافتراضات مثل نماذج نظرية الاستجابة للفقرة. الكلمات المفتاحية: دقة تقدير معلمة القدرة، طريقة دلتا المعلمية، طريقة كيرنل اللامعلمية، مؤشر الخطأ المعياري للتقدير (**SEE**).

¹ باحثة دكتوراة في القياس والتقويم، قسم علم النفس، كلية التربية جامعة الملك سعود

*البريد الإلكتروني: H-M-R44@hotmail.com

² أستاذ مشارك القياس والتقويم والإحصاء، قسم علم النفس، كلية التربية، جامعة الملك سعود،

المملكة العربية السعودية *البريد الإلكتروني: eqbal@ksu.edu.sa

Differences Between the Parametric Delta Scoring Estimation Method and the Non-Parametric Kernel Method in the Accuracy of Ability Parameter Estimation

Ms.Hind Abdullah ALjaser

Doctoral Researcher of Measurement and Evaluation, Department of Psychology,
College of Education, King Saud University, Saudi Arabia.

H-M-R44@hotmail.com

Dr. Eqbal Z. Darandari

Associate Professor of Measurement, Evaluation and Statistics, Department of
Psychology, College of Education, King Saud University, Saudi Arabia.

eqbal@ksu.edu.sa

Abstract:

The current study aimed to detect differences in the accuracy of estimating the ability parameter according to the method of score estimation (parametric Delta Scoring method- Latent Framework -DSM-L, and nonparametric Kernel Smoothing-KS method), the sample size (100- 250- 500- 1000), and the data distribution (normal - slightly positive skewness - strong positive skewness - slightly negative skewness - strong negative skewness), and the interaction between them; using data generated (by simulation), according to the Monte Carlo method (MCM) using (50) binary items (0,1). The accuracy of the estimation was judged using the standard error of estimation index (SEE). The results of the study showed the following: There are statistically significant differences ($P < .05$) in the accuracy of estimating the ability parameter according to the sample size and data distribution; and there are no statistically significant differences ($P > .05$) in the accuracy of the ability parameter according to the method of estimating (the parametric DSM-L delta and the nonparametric kernel -KS); and there are statistically significant differences in the three-way interaction ($P < .05$) in the accuracy of estimating the ability parameter among each of the estimation methods, sample sizes, and data distributions. The results show that there are no statistically significant differences ($P > .05$) in the accuracy of the estimate according to the method in the normal distribution, but there are statistically significant differences ($P < .05$) in the skewed distributions, and the sample size have an effect in most of the distributions. Based on the results, the study recommends selecting the appropriate method, according to the sample size and data distribution, when estimating the ability to reach the most accurate estimates. With small samples and skewed data, especially the strongly positive and negative ones, it is preferable to use the Kernel (KS) method, and with large samples and normal data, it is recommended to use the Delta (DSM-L) method, due to the accuracy of its estimates and the lack of need to verify many assumptions such as item response theory models.

Keywords: Accuracy of Ability Parameter Estimation, Parametric Delta Method, Non-Parametric Kernel Method, Standard Error of Estimation (SEE).

المقدمة:

تُعَدّ الاختبارات المعرفية والنفسية من أهم أدوات القياس والتقييم الشائعة، التي تستند في بنائها وتحليلها على نماذج نظريات القياس المختلفة، وتُقَدَّم 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 تتمتع بخصائص سيكومترية لا تتأثر بهما؛ وهذا ما أدى إلى ظهور نظرية الاستجابة للفقرة (Item Response Theory-IRT)؛ إذ وضع رُود هذه النظرية عددًا من النماذج الرياضية (Hambleton & Swaminthan, 1991).

وتنقسم نماذج نظرية الاستجابة للفقرة (IRT) إلى نوعين رئيسيين، 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 تشددًا. وتُعَدّ 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 التدرج وذات مستويات قياس اسمية (Nozawa, 2008).

واستكمالًا للجهود المبذولة، وتجنبًا للتعقيدات الحسابية والأخطاء في التقدير، والمشاكل الفنية في نماذج نظرية الاستجابة للفقرة؛ فقد طُوّر دومينغو وديميتروف (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 التدرج، ويعتمد على نمذجة العلاقة القائمة بين استجابات الأفراد على فقرات المقياس والقدرة الكامنة، كما في نماذج نظرية استجابة الفقرة. وفي عام (2015) طُوّر ديميتروف (Dimitrov) طريقة تقدير الدرجات لتقدير مغلّمة القدرة للفرد ومعلمات 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 المتوقعة للفقرة الاختبارية، إضافة إلى إمكانية تدرج 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 تقادي 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 تعددت نظريات القياس والنماذج والأساليب المختلفة لتقدير قدرات الأفراد ومعلمات الفقرات، وأصبحت محط اهتمام الباحثين لتقديم تقييم دقيق؛ لاتخاذ قرارات موضوعية دقيقة. ولأن دقة تقديرات معلمات الأفراد والفقرات المستخلصة من نماذج نظرية القياس الحديثة تتأثر بعدد من العوامل؛ فقد استهدفتها الكثير من الدراسات باستخدام بيانات حقيقية أو مولدة بالحاكاة، ودرست مدى تأثير عدد من العوامل مثل: طريقة التقدير، وحجم العينة، وطول أداة القياس، وشكل التوزيع في دقة تقديرات هذه المعلمات (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 تُمكن المختص من الإجابة عن أي تساؤل يُثار حول الفقرة أو الاختبار أو المفحوص (الشريفي، 2006).

وهذا ما شجع على استخدام نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية في هذه الدراسة، توافقا مع توصيات سبق أن أشار إليها هامبلتون وسلاتر (Hambleton & Slater, 2011). إذ أوصيا بضرورة إجراء تغييرات أساسية على الاختبارات، كالانتقال من الاعتماد على القياس الكلاسيكي إلى تبني النماذج الحديثة في القياس؛ للتغلب على عيوب الاختبارات التقليدية. وبالرغم من المميزات التي تمتاز بها نظرية الاستجابة للفقرة (IRT)؛ لكن هناك مشكلة في توفير أحجام العينات وأطوال الاختبارات المطلوبة حسب النموذج المعلمي المُستخدم، وفي صعوبة توفير الافتراضات التي تتطلبها نماذج نظرية الاستجابة للفقرة (IRT) المعلمية؛ لذا يلجأ بعض الباحثين إلى النماذج المعلمية المبسطة 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 أحجام عينات وأطوال اختبار كبيرة مقارنة بنظيرتها المعلمية. وبين رامسي (Ramsay, 1991) أن عدداً قليلاً من الأفراد لا يزيد عن (100)، وعدد فقرات الاختبار (20)؛ هما المطلوبان لتقدير المنحنى المميز للفقرة (Item Characteristic Curve - ICC)، الذي يُطلق عليه دالة الاستجابة للفقرة (Item Response Function - IRF).

وبالرغم من أن هناك عدداً من العوامل التي قد تؤثر في دقة تقديرات معلمات هذه النماذج، مثل: طرق تقدير الدرجات المُستخدمة، 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 -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: الشريفي ومناصرة (2017)، والشاورة (2017)، والقيسي (2013)، ومورديرليك 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: دومينكو وديميتروف (Dimitrov & Domingue, 2015)، وديميتروف وأتانسوف (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 التدرج، فكل النموذجين مبني على معالجة مشكلات في نماذج الاستجابة للفقرة المعلمية. ويكثر استخدامهما في الفقرات ثنائية التدرج، مع أهمية معلمة 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 استخدم في هذه الدراسة أسلوب المحاكاة وفق طريقة مونت كارلو (Monte Carlo Methods-MCM)، التي تُساعد على الوصول إلى النتائج المثالية، حيث يتم ضبط جميع المتغيرات، كما أنه يُوفّر الوقت والجهد على الباحث؛ ومن هنا تبرز الحاجة إلى استكشاف أثر طرق تقدير الدرجات المُستخدمة، وحجم العينة، وتوزيع البيانات باستخدام هذه النماذج في دقة تقدير المعلمات، خاصة معلمة 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 التالي:

س1: هل تختلف دقة تقدير معلمة القدرة المُقدّرة باستخدام فقرات ثنائية التدرج تبعاً لطريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، وحجم العينة (100-250-500-1000)، وتوزيع البيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، والتفاعل بينها؟

أهداف الدراسة:

هدفت الدراسة الحالية إلى المقارنة بين دقة تقدير معلمة القدرة المُقدّرة باستخدام فقرات ثنائية التدرج تبعاً لطريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، وحجم العينة (100-250-500-1000)، وتوزيع البيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، والتفاعل بينها.

أهمية الدراسة:

الأهمية النظرية:

1. تتكامل هذه الدراسة مع الدراسات وجهود الباحثين في المقارنة بين نماذج نظريات القياس؛ للوصول إلى بناء معرفي متكامل في البحوث.

2. 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.
3. تحاول الدراسة الكشف عن أثر حجم العينة وشكل التوزيع في دقة تقدير مَعْلَمَة القدرة باستخدام نماذج مختلفة، حيث هناك قلة في الدراسات -التي درست تأثير توزيع البيانات في دقة التقدير.
4. تُوجّه نتائج هذه الدراسة اهتمام الباحثين في مجال القياس والتقويم النفسي والتربوي نحو تناول أثر عدد من المتغيرات المختلفة في الخصائص السيكومترية، باستخدام نماذج نظرية الاستجابة للفقرة (IRT)، وطرق تقدير الدرجات المعلمية واللامعلمية، ونماذج دلتا لتقدير الدرجات (DSM)، والمقارنة بينها.

الأهمية التطبيقية:

1. من المتوقع أن تُوفّر نتائج هذه الدراسة معلومات تساعد الباحثين على توظيفها لتقدير السمات الكامنة للأفراد بأعلى كفاءة ممكنة باستخدام النماذج المستعملة في الدراسة الحالية.
2. ستُساعد 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 مرجعيًا.
3. تُقدم الدراسة معلومات وإرشادات تساعد مُطوري الاختبارات النفسية والتربوية في الحصول على تقديرات دقيقة لمعلومات الفقرة باختلاف طرق التقدير، وأحجام العينة، وشكل توزيع القدرة.

حدود الدراسة:

1. اقتصرَت الدراسة على استخدام بيانات مُولّدة (بأسلوب المحاكاة) (Simulation)، وفق طريقة مونت كارلو (MCM)، وهذه الطريقة تُوفّر بيانات ذات ضبط عالٍ للمتغيرات؛ لذا استخدمت الدراسة أسلوب المحاكاة في توليد البيانات لـ (50) فقرة ثنائية التدرّج (0،1)؛ حيث أوصى لورد (Lord, 1980) أن يكون طول الاختبار (50) فقرة، وأن يكون عدد الأفراد (1000)؛ للحصول على أفضل التقديرات.
2. استخدمت الدراسة الأحجام (100- 250 - 500 - 1000)؛ لدراسة أثر أحجام العينات في دقة التقدير باستخدام نموذج الاستجابة للفقرة ثنائي المعلمة (Two-Parameter Logistic Model- 2PLM). وقد أوضحت دراسة دي إيالا (2017/2009) ارتباط حجم العينة وعدد الفقرات بالنموذج المُستخدَم، حيث إن النموذج المُستخدَم في الدراسة يتناسب مع أحجام عينات أكبر من (200)، وأقل أو يساوي (1000).
3. اعتمدت الدراسة شكل توزيعات محددة للبيانات (اعتدالي- التواء موجب بسيط- التواء موجب شديد- التواء سالب بسيط- التواء سالب شديد)، باستخدام برنامج (WinGen v1.4) (Han, 2007).

4. اعتمدت الدراسة 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 المُقدَّرة؛ وبالتالي فإن مؤشر دقة التقدير هو درجة الاتفاق بين قيم التقديرات الملاحظة والمُتوقَّعة (Guo, 2006).

التعريف الإجرائي: هو مقدار الفرق بين قيم معلمات القدرة المُقدَّرة حسب طريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، والقيمة الفعلية (المولَّدة) 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 (Dimitrov, Domingue & Dimitrov, 2015; 2018).

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

الدراسات السابقة:

سوف يتم عرض الدراسات السابقة ضمن ثلاث محاور كالتالي:

المحور الأول: الدراسات التي أُجريت حول دقة تقدير معلمات الفقرات والأفراد في نماذج نظرية الاستجابة للفقرة (IRT) - النماذج المعلمية (PIRT) والنماذج اللامعلمية (NIRT) - وفق عدد من المتغيرات:

قام بني عطا والشريفين (2012) بدراسة هدفت إلى التحقق من أثر اختلاف شكل توزيع القدرة على معلمات الفقرة ودالة المعلومات للاختبار. ولتحقيق الغرض من الدراسة وُلدت استجابات أربعة أشكال من توزيعات القدرة وفق النموذج اللوجستي ثلاثي المعلمة، و(60) فقرة ثنائية الاستجابة. وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات معلمات الفقرات. وعلى

الرغم من تباين معلمات الفقرات؛ فقد أظهرت النتائج أن قيم معاملات الارتباط بين تقديرات معلمات الفقرات المعيارية تبعاً لشكل توزيع القدرة؛ كانت دالة إحصائياً بوصفها مؤشراً على استقرار معلمات الفقرات عبر التوزيعات المختلفة للقدرة.

كما هدفت دراسة الشريفيين (2012) إلى الكشف عن أثر طريقة تقدير معلمات الفقرة وقدرات الأفراد في قيم معلمات الفقرة والخصائص السيكومترية للاختبار، في ضوء متغير حجم العينة. ولتحقيق هدف الدراسة بُني اختبار تحصيلي في الفيزياء من نوع الاختيار من أربعة بدائل، تكوّن من (33) فقرة، وطُبق الاختبار على عينة مكونة من (1000) طالبا وطالبة من طلبة الصف الثاني الثانوي العلمي. وأظهرت النتائج وجود فروق ذات دلالة إحصائية في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تُعزى إلى مُتغير حجم العينة، وللفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تُعزى إلى طريقة التقدير. كما أشارت النتائج إلى أن دقة تقديرات معلّمة 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 (Maximum Likelihood Estimation – MLE)؛ بغض النظر عن حجم العينة.

أما دراسة شما (2013) فهدفت إلى التّعرف على تأثير حجم العينة في دقة تقدير صعوبة الفقرات وقدرة الأفراد باستخدام نموذج راش. ولتحقيق الهدف من الدراسة طُبق اختبار أوتيس - لينون للقدرة العقلية على ثلاث عينات مختلفة من حيث الحجم تكوّنت من: (200-800-1500) 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.

وهدف دراسة بني عطا (2014) إلى الكشف عن دقة تقدير النموذج اللوجستي الثلاثي المعلم لمعلمات الفقرة والقدرة، من خلال تغير طول الاختبار وحجم العينة. وقد اعتمدت الدراسة على توليد بيانات محاكاة ثنائية الاستجابة لستة من أطوال الاختبار (10-25-50-75-100) فقرة، ووُلدت بيانات لستة أحجام مختلفة من العينة (100-250-500-1000). وأظهرت نتائج الدراسة تأثيراً دالاً إحصائياً لمُتغير طول الاختبار ومُتغير حجم العينة والفاعل بينهما في دقة تقديرات معلمات الفقرات والأفراد.

وهدف دراسة الحواري (2015) إلى الكشف عن أثر طول الاختبار وشكل توزيع القدرة في تقديرات القدرة للأفراد، وتقديرات معلمات الصعوبة للفقرات، واقتران المعلومات للفقرات، وفق نموذج راش في نظرية استجابة الفقرة. ومن ضمن إجراءات الدراسة: تم ولید البيانات، ومجموعة الفقرات، وقدرات الأفراد؛ إذ وُلدت بيانات بواقع (1000) مفحوص، بطول الاختبار (30,60) فقرة، وشكل توزيعات القدرة (ملتوٍ التواء سالباً- طبيعي، ملتوٍ التواء موجباً)، وفق النموذج الأحادي المعلمة. وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية في تقديرات معلمات قدرة الأفراد تُعزى إلى شكل توزيع القدرة وطول الاختبار، حيث كانت تقديرات قدرة الأفراد أكثر دقة لصالح شكل الالتواء السالب والموجب.

بينما هدفت دراسة الشريفيين ومناصرة (2017) إلى تقدير خصائص توزيع قدرات الأفراد ومعلمات فقرات اختبار وفق نماذج نظرية الاستجابة للفقرة المعلمية (PIRT) واللامعلمية (NIRT)، وإجراء مقارنة بين النماذج المُستخدمة: 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 (25) فقرة من نوع الاختيار من متعدد، وتكوّنت عينة الدراسة من (1010) طالبا وطالبة من طلبة الصف الرابع الأساسي، ممن طُبّق عليهم الاختبار في المدارس الحكومية والخاصة الأردنية للعام الدراسي (2012-2013). وأظهرت النتائج وجود فروق ذات دلالة إحصائية 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.

كما هدفت دراسة الخرشة (2018) إلى الكشف عن أثر حجم العينة، واختلاف شكل توزيع القدرة (طبيعي ملتوٍ نحو اليمين - ملتوٍ نحو اليسار) في معلمات الفقرة. ولتحقيق الغرض من الدراسة وُلدت استجابات تسع مجموعات من الأفراد وفق النموذج اللوجستي ثلاثي المعلمة و(50) فقرة ثنائية الاستجابة. وكشفت نتائج تحليل التباين الثنائي عن وجود فروق ذات دلالة إحصائية عن مستوى دلالة بين متوسطات معلمات الفقرة لل فقرات تبعاً لحجم العينة، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين متوسطات معلمات الفقرة تبعاً لشكل التوزيع.

وهدف دراسة مورديرليك وكوتش (Mor Dirlik & Koç, 2019) إلى مقارنة معلمات القدرة المقدرة في النماذج المعلمية واللامعلمية (NIRT)، وفقاً لعوامل مختلفة منها: طول الاختبار، وحجم العينة، والصفات السيكومترية لل فقرات. وأُستخدم اختبار الرياضيات للصف الثامن في الدراسة الدولية تيمس (TIMSS 2011) لعينة مكوّنة من (7254) طالبا وطالبة من دول مختلفة. وبالنسبة لحجم العينة فكان (500-1000-3000)، وكان طول الاختبار (5-15-25) فقرة. وأظهرت النتائج أن جميع معلمات الفقرة المقدرة وفقاً للنموذجين المعلمي (PIRT) واللامعلمي (NIRT) مرتبطة ارتباطاً عالياً. أما بالنسبة لمعلمات القدرة؛ فقد ظهر في النماذج المعلمية (PIRT) عدم الاتساق في التقديرات - خاصة عند (5 و15) فقرة- بينما القدرات المقدرة في حال (25) فقرة؛ فكانت مرتبطة بشكل أكبر. وفي حال النماذج اللامعلمية (NIRT)، أظهرت جميع أحجام العينات وأطوال الاختبار ارتباطاً عالياً، كما أظهرت النتائج الارتباط المرتفع بين النماذج المعلمية (PIRT) واللامعلمية (NIRT) فقط في حال حجم العينة (1000) والطول (25) فقرة، وكلما زاد طول الاختبار وحجم العينة؛ كانت التقديرات أكثر اتساقاً.

في حين هدفت دراسة ضعضع وآخرين (2020) إلى الكشف عن أثر طرق تقدير الأرجحية العظمى، وتقدير بيز، وطريقة بروكس، باستخدام عينات ذات أحجام مختلفة (500-1000-1500-2000) على دقة تقدير معلمات الفقرة والأفراد باستخدام نموذج راش، بالاعتماد على الخطأ المعياري للتقدير (SEE). ولتحقيق هدف الدراسة وُلدت استجابات لـ(2000) مفحوص لاختبار مكوّن من (40) مفردة ثنائية الاستجابة، بتوزيع طبيعي معياري لمعلم للقدرة والصعوبة،

وأُخذت عينات عشوائية جزئية منها (500-1000-1500). ومن بين النتائج التي توصلت إليها الدراسة: أنه توجد فروق في دقة تقدير معلمة الصعوبة وقدرة الأفراد تُعزى إلى حجم العينة وطريقة التقدير والتفاعل بينهما، وتزايد دقة تقدير معلمة الصعوبة وقدرة الأفراد بازدياد حجم العينة.

أما دراسة البادية وآخرين (2018) 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 (8484) طالبًا وطالبة. ومن هذه العينة أُستخدمت (12) عينة بأحجام مختلفة سُحبت بطريقة عشوائية، حيث سُحبت أربع عينات مختلفة (200-500-1000-1500)، وفي كل عينة من العينات الأربع سُحبت ثلاث عينات مختلفة بالحجم نفسه؛ ليصبح مجموع العينات (12)، وأُستخدمت طريقة الأرجحية العظمى الهامشية (Marginal Maximum Likelihood - MML) في التقدير. وللحكم على دقة التقدير أُعتمد على الجذر التربيعي لمتوسط مربعات الخطأ (Root Mean Squared Error - RMSE) بين المعلمات المُقدّرة للعينة الكلية، والمعلمات المُقدّرة من العينات (12) 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 (500) فردا كافٍ ليعطي تقديرات للقدرة وصعوبة الفقرة بشكل دقيق باستخدام نموذج راش.

وهدف دراسة علوان وجاسم (Alwan & Jasim, 2022) إلى الكشف عن أثر الاختلاف في توزيع مستوى القدرة الملنوي موجبًا لمعاملات فقرات اختبار القدرة العقلية وفقًا لنظرية الاستجابة للفقرة، وطُبّق الاختبار على عينة مُكوّنة من (1000) طالبًا وطالبة، بالاعتماد على نموذج المعلمة الثلاثي. وتألّف الاختبار في صورته النهائية من (78) فقرة. ولتحقيق أهداف البحث وُلدت البيانات باستخدام (المحاكاة)؛ للحصول على شكل التوزيع الملنوي الموجب، وأظهرت النتائج وجود دلالة إحصائية في الصعوبة والتمييز والتقدير لل فقرات وفقًا للشكل المُتغيّر للتوزيع الملنوي الموجب.

أما دراسة الشافعي (2024) فهدفت إلى الكشف عن تأثير حجم العينة في تقديرات صعوبات الفقرات وقدرات الأفراد باستخدام تحليلات النموذج اللوجستي ثلاثي المعلمة، وقد بلغ حجم عينة الدراسة (1800) فردا، وطُبّق الباحث اختبارًا لمادة الجبر للصف الثاني الثانوي مكوّن من (40) مفردة اختبارية من نوع الاختيار من متعدد، على أربعة عينات ذات أحجام مختلفة (300-600-900-1800) فرد. وأُستخدمت أخطاء القياس مؤشرًا لدقة القياس، كما استخدمت أسلوب تحليل تباين القياس المتكرر؛ 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 باختلاف حجم العينة، كما تبين أن دقة تقديرات قدرات الأفراد تزداد بزيادة حجم العينة.

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

هدفت دراسة القيسي (2013) إلى الكشف عن دقة تقدير معلمات 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 (100 - 250 - 500 - 1000) فرداً، وعدد فقرات الاختبار (20 - 40 - 60). ولتقدير معلمات الفقرة والقدرة أستخدمت طريقة الأرجحية العظمى الهامشية (MML) المعلمية، وأستخدمت برمجية تستجراف (TESTGRAF) لتقدير معلمات الفقرة والقدرة باستخدام طريقة تهذيب النواة اللامعلمية (كيرنل KS). وأظهرت النتائج أن المعلمات المقدرة بطريقة التقدير المعلمية (PIRT)؛ كانت أفضل من المعلمات المقدرة وفق طريقة التقدير اللامعلمية (NIRT).

وأجرى القيسي (2016) دراسة هدفت إلى الكشف عن أثر حجم العينة وطول الاختبار في دقة تقدير معلمات الفقرة والقدرة باستخدام نماذج نظرية الاستجابة للفقرة اللامعلمية (NIRT)، اعتماداً على مؤشري دقة القياس: (التحيز BIAS)، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE)، وباستخدام بيانات مولدة، واستخدام طريقة تهذيب النواة (كيرنل KS) اللامعلمية (NIRT) لتقدير معلمات الفقرة والقدرة. وأظهرت نتائج الدراسة وجود فروق دالة إحصائية في متوسطات مؤشر دقة التقدير في معلمة التمييز (a) تُعزى إلى متغير حجم العينة، وفي معلمة التخمين (c) تُعزى إلى متغيري الدراسة (حجم العينة، وطول الاختبار) والتفاعل بينهما، ودقة تقدير معلمة القدرة (θ) تُعزى لمتغيري الدراسة (حجم العينة، وطول الاختبار)، في حين لم تظهر فروق دالة إحصائية في متوسطات مؤشر دقة التقدير في معلمة الصعوبة (b).

كما أجرى الشاورة (2017) دراسة هدفت إلى التعرف على الفاعلية النسبية لطريقة بيز (Bayes) المعلمية، وطريقة تهذيب النواة كيرنل (KS) اللامعلمية (NIRT) في تقدير معلمة القدرة وفق النموذج اللوجستي الثنائي باختلاف حجم العينة، واستخدمت الدراسة البيانات المولدة. وأظهرت نتائج الدراسة وجود فرق دال إحصائية بين قيمتي معامل الارتباط بين معلمة القدرة الحقيقية ومعلمة القدرة المقدرة وفق طريقة تقدير بيز المعلمية من جهة، ومعامل الارتباط بين معلمة القدرة الحقيقية ومعلمة القدرة المقدرة وفق طريقة التقدير كيرنل (KS) اللامعلمية (NIRT) من جهة أخرى؛ لصالح معامل الارتباط بين معلمة القدرة الحقيقية ومعلمة القدرة المقدرة وفق طريقة التقدير بيز المعلمية، عندما كان حجم العينة (250 و 500)، بينما لم يثبت وجود فرق دال إحصائية عند حجم العينة (1500) فرداً.

وهدف دراسة أجراها عفتبانه وبغائي (Effatpanah & Baghaei, 2022) إلى الكشف عن جودة تقييم الفاحصين باستخدام منحني خصائص الفقرة اللامعلمي، واستخدمت الدراسة طريقة تقدير منحني خاصية الفقرة اللامعلمي لنمذجة استجابات الأفراد وفحصها، وطُبقت على عينة مكونة من (217) طالباً في الجامعة الإسلامية بإيران، منهم (87) ذكوراً و(130) إناثاً في اختبار الكتابة

(250) كلمة باللغة الإنجليزية كلغة ثانية عن موضوع: دور التقدم التكنولوجي الحديث في البطالة. وأستخدم معيار تقييم تحليلي يقيس أربعة معايير للكتابة على مقياس مُتدرّج مُكوّن من خمس نقاط، وأستخدمت طريقة تهذيب النواة كيرنل (KS) اللامعلمية (NIRT)؛ بهدف الكشف عن مدى استقرار تقديرات المنحنى في حالة تمثيلها بيانيًا. وأظهرت نتائج المنحنيات أن التقدير اللامعلمي لمنحنيات الخصائص المميزة للفقرة باستخدام طريقة كيرنل (KS)؛ يمكن أن يحدّد تأثيرات التقييم المختلفة، وتوفّر معلومات تشخيصية قيّمة لفحص جودة التصنيف واستكشاف أنماط التصنيف.

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

هدفت دراسة دومينغو وديميتروف (Domingue & Dimitrov, 2015) إلى تناول مدى اتساق قيم معلمات قدرة الأفراد المُقدّرة بأسلوب الدلتا لتقدير الدرجات، والمُتمثلة في (D-Scores) مع مُسلّمات القياس الموحد في المركز الوطني للقياس (The National Center for Assessment - NCA). وقد استندت تحليلات دراستهما على استجابات (45612) من المعلمين السعوديين في اختبار المعلمين (Teacher Test)، الذي يُقدّمه المركز الوطني للقياس والمُكوّن من (79) فقرة. وبرهنت النتائج أن قيم معلمة قدرة الأفراد المُقدّرة بأسلوب الدلتا (D-Scores)؛ تنتج انتهاكات أقل لمُسلّمات القياس الموحد من نظيرتها ثيتا (θ Scores) المُقدّرة في ضوء النموذج اللوجستي الثلاثي المعلمة (Three-Parameter Logistic Model-3PLM)؛ على الرغم من وجود ارتباط عالٍ بين قيم معلمات القدرة المُقدّرة، حيث بلغت قيمة معامل الارتباط بين قيم (θ) و (D) (.91).

وبحث ديميتروف وأتاناسوف (Dimitrov & Atanasov, 2021) دقة تقديرات معلمات 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 أستخدمت بيانات حقيقية وبيانات محاكاة، وتكوّنت عينة البيانات الحقيقية من استجابات (1288) فردًا في اختبار كفايات اللغة الإنجليزية ستيب (Student Training in Engineering Program - STEP) المُقدّم من المركز الوطني للقياس في السعودية، والمُكوّن من (100) فقرة ثنائية التدرّج. وأسفرت نتائج تحليل البيانات عن وجود ارتباط عالٍ بين معلمات الفقرات والأفراد المُقدّرة بنموذج الدالة النسبة ثنائي المعلمة (RFM2) وفق الإطار الكامن (DSM-L)، والمعلمة 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. وُدّرس أثر مُتغيّري: حجم العينة وطول الاختبار في دقة تقديرات المعلمة، حيث وُلدت استجابات لثلاث عينات بحجم (500-1000-3000)، وأربعة اختبارات بأطوال (20-40-60-80) فقرة. وأظهرت نتائج تحليل البيانات أن تقدير معلمات الفقرة والفرد وفق الإطار الكامن (DSM-L)؛ أكثر

دقة وارتباطاً بمثيلاتها المُقدَّرة وفق نظرية الاستجابة للفقرة من التقديرات وفق الإطار الكلاسيكي (DSM-C) في جميع الظروف الاختبارية المُحدَّدة.

وفي دراسة أجراها روبيتزش (Robitzsch, 2021) هدفت إلى بحث مدى التكافؤ بين النموذج الكامن لتقدير الدرجات دلًا (DSM-L) المُقترح من ديميتروف، ونموذج الاستجابة للفقرة اللوجستي ثنائي المعلمة (2PLM) ويُعادلُه إحصائيًا، أُستخدمت بيانات برنامج التقييم الدولي للطلاب بيزا في ألمانيا (PISA -2006 -Program for International Student Assessment)، المُكوّن من (28) فقرة، مع عينة أُخذت من (26) دولة مختلفة. وأظهرت النتائج أنه بعد إجراء عدد كبير من التحويلات الخطية في النموذجين (DSM-L) و (2PLM)؛ فإنهما يعطيان نتائج إحصائية متشابهة؛ لكن يظل نموذج ديميتروف الأنسب في تفسير النتائج.

التعقيب على الدراسات السابقة:

من خلال العرض السابق للدراسات السابقة؛ لوحظ كيف برزت جهود الباحثين في الكشف عن تباينات القيم المُقدَّرة للقدرة ومعلومات الفقرة حسب النماذج المُستخدمة وعدد من المُتغيّرات. وقد تناولت أغلبية الدراسات بشكل مركز مُتغيّر حجم العينة، بالإضافة إلى مُتغيّر طول الاختبار، واستخدمت بعض الدراسات أثر مُتغيّر شكل توزيع البيانات في دقة تقدير معلومات الفرد وال فقرات، سواء كانت بيانات مولَّدة (محاكاة) أو حقيقية. وتتفق الدراسة الحالية مع دراسات: بني عطا (2014)، وبني عطا والشرفين (2012)، والحواري (2015)، والخرشة (2018)، والشافعي (2024)، وشما (2013)، وعلوان وجاسم (Alwan & Jasim, 2022)، في المقارنة بين النماذج المعلمية (PIRT) والنماذج اللامعلمية (NIRT). وخصّصت هذه الدراسة طريقة تهذيب النواة (كيرنل KS) من النماذج اللامعلمية (NIRT)، حيث اتفقت مع دراسات: الشاورة (2017)، وعفتبانه وبغائي (Effatpanah & Baghaei, 2022)، والقيسي (2016، 2013)، كما تفقت الدراسة الحالية مع الدراسات التي استخدمت طريقة دلًا لتقدير الدرجات، مثل دراسات: ودومينكو وديميتروف (Dimitrov & Atanasov, 2015)، وديميتروف وأتانسوف (Dimitrov & Atanasov, 2021)، وروبيتزش (Robitzsch, 2021).

كما اتفقت الدراسة الحالية في الكشف عن أثر حجم العينة في دقة تقرير معلومات الفقرة والقدرة مع دراسات: البادية وآخرون (2020)، وبني عطا (2014)، وديميتروف وأتانسوف (Dimitrov & Atanasov, 2021)، والشافعي (2024)، وشما (2013)، والشاورة (2017)، والقيسي (2013). كذلك اتفقت مع بعض الدراسات التي جمعت بين مُتغيّري: حجم العينة وتوزيع البيانات مثل دراستي: الخرشة (2018)، واتفقت مع الدراسات التي جمعت بين أثر حجم العينة وطرق التقدير، مثل دراستي: الشرفين (2012)، وضعضع وآخرون (2020).

وقد خلصت الدراسات السابقة مثل: البادية وآخرون (2020)، وبني عطا (2014)، والشافعي (2024)، وديميتروف وأتانسوف (Dimitrov & Atanasov, 2021)، والقيسي (2016) إلى وجود أثر دال لحجم العينة في دقة التقدير، ودراسة الشاورة (2017)، التي توصّلت إلى أن اختلاف الحجم كان له أثر؛ باستثناء دراسة شما (2013)، التي توصّلت إلى أن اختلاف

حجم العينة ليس له تأثير في دقة تقدير قدرات الأفراد. كما خلصت دراسة ضعضع (2020) إلى وجود فروق دالة إحصائية تبعاً لحجم العينة وطريقة التقدير والتفاعل بينهما في دقة تقدير معلمات الأفراد، وخلصت دراسة الشريفيين (2012) إلى وجود فروق دالة إحصائية تبعاً لحجم العينة، ولا توجد فروق دالة إحصائية تبعاً لطريقة التقدير، بينما توجد فروق دالة إحصائية تبعاً للتفاعل بين حجم العينة وطريقة التقدير في دقة تقدير معلمات الأفراد. كما خلصت دراسات: الحواري (2015)، والشريفيين ومناصرة (2017)، وبني عطا والشريفيين (2012)، وعلوان وجاسم، Alwan & Jasim, (2022) إلى وجود فروق دالة إحصائية تبعاً لتوزيع البيانات في دقة تقدير معلمات الأفراد.

ومما سبق؛ يتبين أن الدراسة الحالية تميزت في استخدامها لنماذج نظرية الاستجابة للفقرة اللامعلمية (NIRT)، التي تفترض أن البيانات تقع على مستوى القياس الرتبى، وتتطلب افتراضات أقل تشدداً من نظيرتها المعلمية (PIRT)؛ لذلك فإنها تُتيح المجال لتوظيف هذه النظرية في تقدير خصائص الفقرات وقدرات الأفراد، واستخدام الأساليب والطرق الواردة في الدراسة. كما تتميز الدراسة الحالية بمحاولة التوصل إلى الفروق بين الإطار الكامن لأسلوب دلتا المعلمي لتقدير الدرجات (DSM-L)، وطريقة كيرنل لتقدير الدرجات (KS) اللامعلمية (NIRT) في دقة تقدير معلمة القدرة، بينما بحثت معظم الدراسات السابقة عن الفروق بين نماذج نظرية الاستجابة للفقرة (IRT) المعلمية (PIRT) واللامعلمية (NIRT).

كما تتميز الدراسة الحالية باستخدام بيانات مولدة (المحاكاة)، وفق طريقة مونت كارلو (MCM)، التي تساعد على الوصول إلى النتائج المثالية، حيث تضبط جميع المتغيرات، وتوفر الوقت والجهد على الباحث، وقد تباينت الدراسات السابقة بين استخدامها لبيانات حقيقية ومولدة (محاكاة) ضمن أحجام وتوزيعات للبيانات مختلفة. وقد أُستفيد من الإجراءات المُتبعة بالدراسات السابقة في أثناء تطبيق الدراسة الحالية بما يتناسب مع طبيعة المتغيرات المُستخدمة والأساليب الإحصائية، كما فُسرت نتائج هذه الدراسة في ضوء النتائج التي توصلت إليها الدراسات السابقة.

منهج الدراسة:

استخدمت الدراسة منهج المحاكاة التجريبي المقارن لملاءمته لأهداف الدراسة وأسئلتها؛ حيث أن هذا المنهج يُعد أسلوباً فعالاً 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 (50) فقرة ثنائية التدرج. وأحجام العينات (100، 250، 500، 1000) وكان لكل 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 ($\mu=0, -1, -2, 1, 2$) 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 التعرف على تأثير المتغيرات المستقلة: طريقة التقدير (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية) وأحجام العينات (100، 250، 500، 1000) 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 (2×4×5) (تصميم بين- داخل المجموعات Between – Within Subject Design) المستخدم في الدراسة.

جدول 1. التصميم العاملي للدراسة

العامل (A) (طريقة التقدير) (داخل المجموعات)		العامل (C) (مستويات توزيع البيانات) (بين المجموعات)		العامل (B) (حجم العينة بين المجموعات)
(DSM-L) A ₂	(KS) A ₁			
A ₂ B ₁ C ₁	A ₁ B ₁ C ₁	C1 (توزيع اعتدالي) $\mu=0,\sigma = 1$		
A ₂ B ₁ C ₂	A ₁ B ₁ C ₂	C2 (التواء موجب بسيط) $\mu=-1,\sigma = 1$		
A ₂ B ₁ C ₃	A ₁ B ₁ C ₃	C3 (التواء موجب شديد) $\mu=-2,\sigma = 1$		
A ₂ B ₁ C ₄	A ₁ B ₁ C ₄	C4 (التواء سالب بسيط) $\mu=1,\sigma = 1$		
A ₂ B ₁ C ₅	A ₁ B ₁ C ₅	C5 (التواء سالب شديد) $\mu=2,\sigma = 1$		
A ₂ B ₂ C ₁	A ₁ B ₂ C ₁	C1 (توزيع اعتدالي) $\mu=0,\sigma = 1$		
A ₂ B ₂ C ₂	A ₁ B ₂ C ₂	C2 (التواء موجب بسيط) $\mu=-1,\sigma = 1$		
A ₂ B ₂ C ₃	A ₁ B ₂ C ₃	C3 (التواء موجب شديد) $\mu=-2,\sigma = 1$		
A ₂ B ₂ C ₄	A ₁ B ₂ C ₄	C4 (التواء سالب بسيط) $\mu=1,\sigma = 1$		
A ₂ B ₂ C ₅	A ₁ B ₂ C ₅	C5 (التواء سالب شديد) $\mu=2,\sigma = 1$		
A ₂ B ₃ C ₁	A ₁ B ₃ C ₁	C1 (توزيع اعتدالي) $\mu=0,\sigma = 1$		
A ₂ B ₃ C ₂	A ₁ B ₃ C ₂	C2 (التواء موجب بسيط) $\mu=-1,\sigma = 1$		
A ₂ B ₃ C ₃	A ₁ B ₃ C ₃	C3 (التواء موجب شديد) $\mu=-2,\sigma = 1$		
A ₂ B ₃ C ₄	A ₁ B ₃ C ₄	C4 (التواء سالب بسيط) $\mu=1,\sigma = 1$		
A ₂ B ₃ C ₅	A ₁ B ₃ C ₅	C5 (التواء سالب شديد) $\mu=2,\sigma = 1$		
A ₂ B ₄ C ₁	A ₁ B ₄ C ₁	C1 (توزيع اعتدالي) $\mu=0,\sigma = 1$		
A ₂ B ₄ C ₂	A ₁ B ₄ C ₂	C2 (التواء موجب بسيط) $\mu=-1,\sigma = 1$		
A ₂ B ₄ C ₃	A ₁ B ₄ C ₃	C3 (التواء موجب شديد) $\mu=-2,\sigma = 1$		
A ₂ B ₄ C ₄	A ₁ B ₄ C ₄	C4 (التواء سالب بسيط) $\mu=1,\sigma = 1$		
A ₂ B ₄ C ₅	A ₁ B ₄ C ₅	C5 (التواء سالب شديد) $\mu=2,\sigma = 1$		

إجراءات الدراسة:

المرحلة الأولى (توليد البيانات): تم توليد البيانات وذلك لـ (50) فقرة ثنائية التدرج (0,1)، وهي عبارة عن القدرة الفعلية (المولدة) للأفراد وفق أحجام عينات مختلفة (100، 250، 500، 1000) 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, 2007)، حيث كانت معلمات الصعوبة وفقاً للتوزيع الاعتدالي (Normal) بوسط حسابي (0) وانحراف معياري (1)، ومعلمات التمييز وفقاً للتوزيع المنتظم (Uniform) بقيمة صغرى (0.3) وقيمة عظمى (1.9)، وتعد قيم التمييز التي تم توليدها قريبة للقيم التي استخدمها هامبلتون وسواميناثان (Hambleton & Swaminathan, 2010) حيث أكد على أن تكون قيم معلمة التمييز متراوحة ما بين (0 و 2) لوجت، والجدول (2) يوضح الإحصاءات الوصفية الخاصة بمعلمة القدرة ومعلمات الفقرات المولدة (الصعوبة، والتمييز).

جدول 2. الإحصاءات الوصفية الخاصة بمعلمة القدرة ومعلمات الفقرات المولدة (الصعوبة، والتمييز)

حجم العينة	مستويات توزيع القدرة	خصائص توزيع المعلمات	
		الصعوبة	التمييز
(100)	(توزيع اعتدالي)	$\mu=0, \sigma=1$	(توزيع منتظم) القيمة الصغرى = 0.3 القيمة العظمى = 1.9
	(التواء موجب بسيط)	$\mu=-1, \sigma=1$	
	(التواء موجب شديد)	$\mu=-2, \sigma=1$	
	(التواء سالب بسيط)	$\mu=1, \sigma=1$	
	(التواء سالب شديد)	$\mu=2, \sigma=1$	
(250)	(توزيع اعتدالي)	$\mu=0, \sigma=1$	(توزيع اعتدالي) القيمة الصغرى = 0.3 القيمة العظمى = 1.9
	(التواء موجب بسيط)	$\mu=-1, \sigma=1$	
	(التواء موجب شديد)	$\mu=-2, \sigma=1$	
	(التواء سالب بسيط)	$\mu=1, \sigma=1$	
	(التواء سالب شديد)	$\mu=2, \sigma=1$	
(500)	(توزيع اعتدالي)	$\mu=0, \sigma=1$	(توزيع اعتدالي) القيمة الصغرى = 0.3 القيمة العظمى = 1.9
	(التواء موجب بسيط)	$\mu=-1, \sigma=1$	
	(التواء موجب شديد)	$\mu=-2, \sigma=1$	
	(التواء سالب بسيط)	$\mu=1, \sigma=1$	
	(التواء سالب شديد)	$\mu=2, \sigma=1$	
(1000)	(توزيع اعتدالي)	$\mu=0, \sigma=1$	(توزيع منتظم)

حجم العينة	مستويات توزيع القدرة	خصائص توزيع المعلمات	
		القدرة	الصعوبة
	(التواء موجب بسيط)	$\mu = -1, \sigma = 1$	اعتدالي
	(التواء موجب شديد)	$\mu = -2, \sigma = 1$	$\mu = 0$
	(التواء سالب بسيط)	$\mu = 1, \sigma = 1$	$\sigma = 1$
	(التواء سالب شديد)	$\mu = 2, \sigma = 1$	
			القيمة الصغرى = 0.3 القيمة العظمى = 1.9

ويوضح الجدول (3) المتوسطات الحسابية والانحرافات المعيارية والقيمة العظمى والصغرى لمعلمة القدرة الفعلية (المولدة) لكل حالة من الحالات المشمولة بالدراسة وفقاً لمتغيري (حجم العينة، توزيع البيانات).

جدول 3. المتوسطات الحسابية والانحرافات المعيارية لمعلمة القدرة وفقاً لمتغيري الدراسة (حجم العينة، توزيع البيانات)

1. العينة	الإحصائي	مستويات توزيع القدرة			
		توزيع اعتدالي	التواء موجب بسيط	التواء موجب شديد	التواء سالب بسيط
(100)	المتوسط الحسابي	0.242	-0.879	-1.801	1.162
	الانحراف المعياري	0.968	1.097	0.899	1.042
	القيمة العظمى	2.864	1.931	0.120	3.171
	القيمة الصغرى	-1.492	-4.939	-3.407	-1.981
(250)	المتوسط الحسابي	0.131	-0.921	-1.877	1.122
	الانحراف المعياري	0.967	1.016	0.992	1.010
	القيمة العظمى	2.719	1.773	1.112	4.173
	القيمة الصغرى	-2.455	-3.598	-4.870	-1.731
(500)	المتوسط الحسابي	0.133	-0.888	-1.979	1.119
	الانحراف المعياري	1.028	1.003	0.997	0.959
	القيمة العظمى	3.076	2.191	1.752	4.584
	القيمة الصغرى	-2.720	-3.905	-4.747	-2.029
(1000)	المتوسط الحسابي	0.054	-0.972	-1.982	1.005
	الانحراف المعياري	1.000	1.035	1.011	1.033
	القيمة العظمى	3.348	2.127	1.878	4.589
	القيمة الصغرى	-3.089	-5.463	-5.249	-2.502

ويوضح الجدول (4) المتوسطات الحسابية والانحرافات المعيارية لمعاملات الفقرات المولدة من الصعوبة، والتميز التي تم توليدها وفقاً لمتغيري الدراسة (حجم العينة، وتوزيع البيانات).
جدول 4. المتوسطات الحسابية والانحرافات المعيارية لمعاملات الفقرات المولدة (الصعوبة، والتميز) وفقاً لمتغيري الدراسة (حجم العينة، وتوزيع البيانات)

حجم العينة	الإحصائي	توزيع اعتدالي		التواء موجب بسيط		التواء موجب شديد		التواء سالب بسيط		التواء سالب شديد	
		تميز	صعوبة	تميز	صعوبة	تميز	صعوبة	تميز	صعوبة	تميز	صعوبة
(100)	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
(250)	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
(500)	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

المتوسط الحسابي	الانحراف المعياري	القيمة العظمى	القيمة الصغرى	توزيع اعتدالي		التواء موجب بسيط		التواء موجب شديد		التواء سالب بسيط		التواء سالب شديد	
				تميز	صعوبة	تميز	صعوبة	تميز	صعوبة	تميز	صعوبة	تميز	صعوبة
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 سؤال الدراسة، تم تحليل 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, 2007).

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 ومن ثم تدوير العوامل باستخدام طريقة فاريماكس. والجدول (5) يوضح قيم الجذر الكامن ونسب التباين المفسر للعاملين الأول والثاني وناتج قسمة قيمة الجذر للعامل الأول على جذر العامل الثاني في التوزيعات والأحجام المختلفة. والجدول (6) يوضح الرسم البياني (Scree Plot) لقيم الجذور الكامنة لجميع العوامل لبيانات الدراسة باختلاف حجم العينة وتوزيع البيانات.

جدول 5. 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
		النسبة	الجذر الكامن	النسبة	الجذر الكامن	النسبة	الجذر الكامن	الجذر الكامن للعامل الثاني - الجذر الكامن للعامل الثالث
(100)	(توزيع اعتدالي)	12.792	25.585	2.283	4.565	2.128	4.256	5.605
	(التواء موجب بسيط)	13.631	27.262	2.206	4.412	1.943	3.885	6.179
	(التواء موجب شديد)	11.239	22.477	2.564	5.128	2.150	4.299	4.383
	(التواء سالب بسيط)	13.183	26.365	2.487	4.973	2.185	4.369	5.301
	(التواء سالب شديد)	12.265	24.529	3.292	6.583	2.608	5.216	3.726
(250)	(توزيع اعتدالي)	12.272	24.543	2.026	4.052	1.537	3.075	6.057
	(التواء موجب بسيط)	13.180	26.359	1.972	3.944	1.673	3.345	6.683
	(التواء موجب شديد)	11.093	22.185	3.086	6.171	1.851	3.702	3.595
	(التواء سالب بسيط)	11.833	23.666	2.195	4.390	1.721	3.443	5.391
	(التواء سالب شديد)	10.799	21.597	2.334	4.668	2.276	4.551	4.627
(500)	(توزيع اعتدالي)	13.108	26.215	2.387	4.774	1.239	2.478	5.491
	(التواء موجب بسيط)	11.970	23.940	1.981	3.963	1.382	2.763	6.041
	(التواء موجب شديد)	9.271	18.541	2.677	5.354	2.006	4.011	3.463
	(التواء سالب بسيط)	9.374	18.748	2.712	5.424	1.868	3.737	3.456
	(التواء سالب شديد)	10.568	21.137	2.590	5.180	1.661	3.322	4.080

								شديد)	(1000)
10.788	6.046	2.222	1.111	4.173	2.087	25.232	12.616	(توزيع اعتدالي)	
13.619	5.676	3.018	1.509	4.596	2.298	26.091	13.045	(التواء موجب بسيط)	
14.220	4.691	3.179	1.590	4.294	2.147	20.146	10.073	(التواء موجب شديد)	
14.471	5.908	2.520	1.260	3.815	1.907	22.540	11.270	(التواء سالب بسيط)	
16.945	4.753	3.486	1.743	4.478	2.239	21.287	10.644	(التواء سالب شديد)	

جدول 6. 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
100					
250					
500					
1000					

يتضح من النتائج الواردة في الجدول (5) بأن جميع نسب التباين المفسر من العامل الأول كانت أعلى من (20%) في الأحجام والتوزيعات المختلفة فيما عدا (التواء موجب شديد والتواء سالب بسيط لحجم العينة 500) كانت القيم مرتفعة وقريبة جداً من المحك المطلوب (الشافعي، 2014). 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، 2015؛ الشرفاوي، 2022). وهذه النتيجة تُعتبر مؤشراً أول على تحقق أحادية البعد، وأن ناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني أكبر من (2) في الأحجام والتوزيعات المختلفة بما فيها (التواء موجب شديد والتواء سالب بسيط) لحجم العينة (500)؛ 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 (6) 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. ومن الواضح تحقق أحادية البعد لجميع الظروف من الأحجام والتوزيعات المختلفة بناءً على المؤشرات السابقة.

الافتراض الثاني: الاستقلال الموضعي (Local 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 (0.30)، وتبين أنه في جميع ظروف الدراسة التي يكون فيها التوزيع اعتدالياً يكون الاستقلال الموضعي متحققاً، وهذا يتفق مع دراسة (الشرفاوي، 2022). أما في حالة التوزيع الملتوي فيظهر لنا وجود أزواج للفقرات بينها ارتباط موضعي وهي بين (2%-9%) حيث تمثل نسبة بسيطة جداً مقارنة بالعدد الكلي للأزواج التي حققت الاستقلال الموضعي وهي بين (91%-100%). وبفحص البيانات تبين ارتفاع نسبة تحقق الاستقلال الموضعي كلما كان حجم العينة كبيراً وايضاً كلما كان توزيع البيانات اعتدالياً أو قريباً من الاعتدالية، لذا فإن عدم تحقق الافتراض الاستقلال الموضعي لعدد من الأزواج قد لا يشير بالضرورة إلى تأثر الإجابة عن فقرة بالفقرة الأخرى، ولا بد من الأخذ بالاعتبار عوامل أخرى تؤثر على الاستقلال الموضعي للفقرات. كما لاحظنا 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 ذكر في الدراسات (DeMars, 2010; Hambleton et al., 1991; Hulin et al., 1983; Raykov & Marcoulides, 2016; Reise & Revicki, 2015) من أن هذا الشرط يتحقق ضمناً بتحقيق شرط أحادية البعد؛ حيث أن هناك ارتباطاً وثيقاً

بين تحقق افتراض أحادية البعد وتحقق افتراض الاستقلال الموضوعي؛ وفقا لما سبق فإن الاستجابات عن الفقرات حققت افتراض الاستقلال الموضوعي وتم الإبقاء على عدد الفقرات 50 فقرة ولم يتم حذف أي فقرة.

الافتراض الثالث: الاضطرابية (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 (Bilog MG 3.0) وذلك عن طريق منحنيات خصائص الفقرة (ICC) لكل فقرة من الفقرات باختلاف حجم العينة (100، 250، 500، 1000) وتوزيع البيانات (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد).

– ثانياً: مطابقة البيانات المولدة للنموذج:

للتحقق من مطابقة الفقرات والأفراد للنموذج الثنائي المعلمة (2PLM)، تم استخدام لغة البرمجة (R) بواسطة حزمة (A package for R - mirt) وحساب إحصائيات جودة الملاءمة (GOF - Goodness of fit). وقد تم الحكم على مطابقة الفقرات وفق المؤشر الإحصائي ($S-X^2$) الذي قدمه كل من ثيسن وأولاندو في (Orlando & Thissen, 2000) للمفردات ثنائي التدرج، كما أثبت هذا المؤشر فاعليته في دراسة (Orlando & Thissen, 2003) في الكشف عن مطابقة الفقرات في اختبارات مختلفة مكونة من (10-40-80) فقرة وعينة (1000) فرد وعند مستويات الدلالة ($P < 0.05$) لضبط الخطأ من النوع الأول، وقد كانت النتائج متقاربة باختلاف مستوى الدلالة، حيث تعد الفقرة غير مطابقة للنموذج إذا قلت قيمة الدلالة الإحصائية عن (0.05)، ويمكن التجاوز عن عدم مطابقة الفقرة إذا كانت قيمة المؤشر غير دالة عند مستوى الدلالة (0.01). لذلك اختير هذا المؤشر للحكم على مطابقة الفقرات لمناسبتها لبيانات وظروف الدراسة.

وقد كانت نسبة وعدد الفقرات غير المطابقة للنموذج الثنائي المعلمة (2PLM) من خلال المؤشر ($S-X^2$) والتي تراوحت بين (2%-4%) وهي نسبة بسيطة جداً مقارنة بنسبة الفقرات المطابقة. كما لوحظ أن المطابقة تكون جيدة عندما يكون التوزيع اعتدالياً في جميع ظروف الدراسة، ويظهر عدم مطابقة بسيط كلما كان الالتواء شديداً في توزيع البيانات. كما لوحظ جودة المطابقة كلما زاد حجم العينة، وهذه النتيجة جاءت متوافقة مع ما توصلت إليه دراسة بني عطا والشريفين (2012)، لذا تم الاحتفاظ بجميع الفقرات في هذه الدراسة بغرض إجراء عملية المقارنة بين معلمات القدرة الفعلية (المولدة) والقدرة المقدرة بالطرق المستخدمة في هذه الدراسة حيث 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 (2- و 2+) فإن مطابقة الأفراد جيدة، وتكون أفضل مطابقة كلما اقتربت القيم من الصفر، وإذا كانت القيم أعلى أو أقل من (2- و 2) تكون غير مطابقة. وتشير بعض الدراسات إلى أنه إذا كانت نسبة عدم المطابقة أكثر من (5%) فقد يشير ذلك إلى وجود مشكلة في النموذج أو البيانات

(Linacre, 2002)، ويعتبر الشخص غير مناسب إذا كان اثنان على الأقل من المؤشرات z - ($zh.outfit$, $z-infit$) خارج النطاق (2- و 2+). وقد توصلت الدراسة إلى أن نسبة الأفراد غير المطابقين تراوحت بين (1% - 3%) وهذا يعني أنا النسبة في حدود النسبة المسموح بها حسب ما ورد في الدراسات (Linacre, 2002).

ومما 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 مناسباً بشكل جيد إذا كانت جميع المؤشرات تشير إلى ملاءمة جيدة، وأخيراً تم اعتبار النموذج مناسباً بشكل مقبول عندما تشير بعض المؤشرات إلى ملاءمة جيدة بينما تشير المؤشرات الأخرى إلى عدم ملاءمة. وفي هذه الدراسة تم الاعتماد على المؤشرات:

- مؤشر مربع كاي $M2$: وهو مؤشر لتقييم ملاءمة النموذج، وتشير القيمة الأقل لـ $M2$ عموماً إلى مطابقة أفضل بين النموذج والبيانات (Maydeu & Joe, 2005).
- p - value: قيمة p المرتبطة بالمؤشر $M2$. وتشير القيمة p العالية ($p > 0.05$) إلى أن النموذج يناسب البيانات بشكل جيد. وتشير قيمة p المنخفضة ($p < 0.05$) إلى أن النموذج لا يتناسب تماماً مع البيانات، وهو ما يحدث غالباً مع أحجام العينات الأكبر، حيث يصبح الاختبار أكثر حساسية حتى للانحرافات الطفيفة عن الملاءمة المثالية. وتتأثر قيمة p بحجم العينة، وحجم العينة الكبير، وحتى العينات 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): وتشير القيم المنخفضة للمؤشرات إلى ملاءمة أفضل، حيث تشير القيم الأقل من (0.06) بشكل عام إلى ملاءمة جيدة (Hu & Bentler, 2009)، وهو مؤشر جيد يُعالج حساسية المؤشر ($M2$) لحجم العينة وعدد المعلمات (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 (0.08) جيدة عادةً. وفي حالة كون (SRMSR) مرتفعاً نسبياً، يشير ذلك 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 (0.90) 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). وأشارت نتائج المؤشرات 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 إليه كثير من الدراسات (منها: الصيخان، 2024؛ عامر، 2004؛ Dogan, 2015)، أما باقي المؤشرات فأعطت نتائج جيدة تشير إلى مطابقة البيانات بشكل عام للنموذج في هذه الدراسة، وهذه النتيجة تتفق مع دراسة كارون وآخرون (Karon et al., 2009) 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 التدرج (scalability coefficients) والتي تستخدم للتحقق من التجانس الاطرادي في فقرات الاختبار أو المقياس وهي: معاملات تدرج أزواج الفقرات (H_{ij}): وتم حساب معاملات تدرج أزواج الفقرات، وأظهرت النتائج مطابقة جميع الفقرات لنموذج موكن حيث كانت قيم معاملات التدرج جميعها موجبة لكل ظرف من ظروف الدراسة، ومعامل (H_i) لكل فقرة في التدرج بالنسبة لباقي الفقرات. وأظهرت النتائج تمتع بعض الفقرات بقيم تدرج عالية جداً، ونسبة بسيطة من الفقرات كانت قيم التدرج منخفضة فيها لكنها قريبة جداً وتكاد تصل إلى القيمة (0.3). وبشكل عام كانت المطابقة جيدة وكانت نسبة الفقرات غير المطابقة بسيطة جداً بالنسبة للعدد الكلي للفقرات، كما أن وجود نسبة بسيطة جداً من الفقرات غير المطابقة يمكن التجاوز عنه، حيث كانت الفقرات غير المطابقة في حال وجود التواء في توزيع البيانات، وهذا يتفق مع ما توصلت إليه الدراسة التي أجراها زيكار (Zickar, 1997) حيث طابقت معظم الفقرات مقياس هوغان للشخصية وليس كلها، ودراسة مؤمني (2015) التي خلصت إلى عدم مطابقة خمس فقرات من أصل (28) فقرة من الفقرات لنموذج موكن الاطرادي. وبالنسبة لمعامل التدرج الكلي للفقرات جميعها (H): تم حساب معامل التدرج الكلي للحكم على جودة التدرج، حيث ذكر كل من سيجتسما ومولينار (Sijtsma & Molenaar, 2002) أن موكن حدد معايير لمعاملات التدرج لضمان جودة التدرج، واقترح حداً أدنى لمعامل تدرج الفقرة ($H_i = 0.3$)، كما اقترح بعض المعايير للحكم على جودة التدرج الكلي، وهي كالتالي: مقياس غير قابل للتدرج ($H < 0.3$)، مقياس ضعيف ($0.3 \leq H \leq 0.4$)، مقياس متوسط ($0.4 \leq H \leq 0.5$)، مقياس قوي ($H \geq 0.5$). وقد أوضحت النتائج مستوى جودة تدرج متوسطة بشكل عام لمعظم ظروف الدراسة، وهذا يتفق مع دراسة جيرلينج وآخرون (Geerlings et al., 2014)، كما لوحظ أن المستوى ضعيف في حال الالتواءات الشديدة، وهذا ينسجم مع ما توصلت إليه كل من دراسة الحواري (2015)، وريس (Reis, 1986)، وهيو وبينتلي (Hu & Bentler, 2009).

(Bentler, 2009) وغيرها من الدراسات من أن طبيعية توزيع البيانات يؤثر في عدة أمور منها تحقق الافتراضات ومطابقة البيانات وتقدير معلمات الفقرات وقدرات الأفراد.

تبين مما سبق تحقيق جميع فقرات الاختبار المكونة من (50) فقرة لشروط نموذج موكن للتجانس الاطرادي فيما يخص معاملات التدرج (H) بأنواعها الثلاثة، الأمر الذي يجيز تسمية التدرج الناتج بتدرج موكن، فقد كانت جميع معاملات تدرج أزواج الفقرات موجبة، وجميع معاملات تدرج الفقرات موجبة، إضافة إلى أن معامل التدرج 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 (50)، 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 الدراسة على مؤشر متوسط الاختلاف المطلق (Mean Absolute Difference-MAD) المشار إليه من قبل ديمتروف وأتانسوف (Dimitrov, 2023; Dimitrov & Luo, 2017) في التعرف على مدى مطابقة الفقرات، وللحكم على مستوى المطابقة تم الاعتماد على المحكات التالية:

- إذا كانت قيمة المؤشر $(MAD \leq 0.07)$ دل ذلك على المطابقة الجيدة للفقرة.
- إذا كانت قيمة المؤشر $(0.07 < MAD < 0.10)$ دل على مطابقة مقبولة للفقرة.
- إذا كانت قيمة المؤشر $(MAD \geq 0.10)$ دل على المطابقة الضعيفة للفقرة.

وتوصلت الدراسة إلى 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 تم اختيار أحد مؤشرات مطابقة الأفراد المستخرجة باستخدام برنامج دلتا والذي يتمثل في إحصائي (Ud Statistic) والذي يعتمد على المؤشر الإحصائي (U3) (Person-Fit Statistic)، وتتحصر قيمة المؤشر (Ud) بين (0-1) 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 للنموذج الثنائي (2PLM) حيث أن مؤشر المطابقة المستخدم في الدراسة الحالية يعتمد على المؤشر الإحصائي (U3) الذي وضعه فان دير فليير (van der Flier, 1980). واتضح أن معظم القيم كانت محصورة بين (0 - 0.6) وتقرب من (0) عندما يكون توزيع البيانات اعتدالياً 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 دلنا المعلمية اتضح أن نسبة الفقرات التي كانت مطابقتها ضعيفة لنماذج دلنا لتقدير الدرجات لم تتجاوز (6%)، ولم تصل قيم مؤشر المطابقة (Ud) في مطابقة الأفراد إلى حدود عدم المطابقة التامة وهي الواحد الصحيح لجميع ظروف الدراسة؛ لذا تم الإبقاء على جميع الفقرات والأفراد في الدراسة .

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

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

1. 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 فارماكس (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 الاطردية من خلال برنامج (Bilog MG 3.0) وذلك عن طريق منحنيات خصائص الفقرة (ICC) لكل فقرة من الفقرات.
2. تم حساب المتوسطات الحسابية والانحرافات المعيارية والخطأ المعياري للتقدير (S.E.E) لكل ظرف اختباري (تصميم عاملي 5X4X2) .
3. تم تقدير معلمة القدرة وفق النماذج المعلمية (PIRT) بطريقة دلنا عن طريق برنامج الدلتا (DELTA 3.1)، باستخدام الإطار الكامن لأسلوب دلنا لتقدير الدرجات (DSM-L) والتي تستخدم أحد أسلوبَي الأرجحية العظمى المشتركة (JML) أو الهامشية (MML). وتم تقدير معلمة القدرة وفق نماذج (NIRT) بطريقة كيرنل (KS) اللامعلمية باستخدام برنامج R بواسطة حزمة (KernSmoothIRT) التي طورها ماز وآخرون (Mazza et al., 2014, 2020) لتقدير القدرات والتي تستخدم طريقة تهذيب النواة في التقدير .
4. 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، وايضاً تم حساب حجم الأثر (Effect Size) للإجابة عن سؤال الدراسة. وتم استخدام شيفيه (*Scheffe*) للمقارنات البعدية. وقد 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. وفي حالة انتهاك افتراض الكروية، تم تطبيق تصحيح جرينهاوس-جايسر (*Greenhouse-Geisser*) لجميع العوامل داخل الموضوعات (*Kassambara, 2023*). وفي 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 التباين، خاصة عند التعامل مع التصميمات المعقدة، نظراً 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*)، حيث تم اعتبار القيمة كبيرة إذا كانت (≥ 0.14)، ومتوسطة إذا كانت (0.06 to 0.13)، ومنخفضة إذا كانت (< 0.06).

نتائج الدراسة وتفسيرها:

النتائج المتعلقة بالإجابة عن سؤال الدراسة: "هل تختلف دقة تقدير معلمة القدرة المُقدّرة باستخدام فقرات ثنائية التدرّج تبعاً لطريقة تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، وحجم العينة (100-250-500-1000)، وتوزيع البيانات (اعتدالي-التواء موجب بسيط-التواء موجب شديد-التواء سالب بسيط-التواء سالب شديد)، والتفاعل بينها؟" للإجابة عن هذا السؤال تم تقدير القدرة بطريقتي تقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية)، والجدول (7) يوضح المتوسط الحسابي والانحراف المعياري لقيم الخطأ المعياري (*SEE*) لتقدير القدرة وفقاً لطريقة التقدير باختلاف حجم العينة وتوزيع البيانات.

جدول 7. 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 (7) أنه بشكل عام، تُظهر طريقة دلتا تبايناً أعلى في الخطأ المعياري للتقدير (SEE) مقارنة بطريقة كيرنل عبر جميع أحجام العينات والتوزيعات. وبالنسبة لطريقة دلتا، يميل الخطأ المعياري للتقدير (SEE) إلى الزيادة في التوزيعات الملتوية، وخاصة في التوزيعات الملتوية الموجبة والسالبة الشديدة، مع ملاحظة انحرافات معيارية أكبر. وعلى وجه التحديد، بالنسبة لحجم عينة (100)، تُظهر طريقة دلتا متوسطات وانحرافات معيارية عالية نسبياً، وخاصة للتوزيعات الملتوية الشديدة، مثل (المتوسط = 0.125 والانحراف المعياري = 0.115) عند الالتواء السالب الشديد. ومع زيادة حجم العينة، ينخفض الخطأ المعياري للتقدير لطريقة دلتا للتوزيعات الاعتدالية ولكنه يظل مرتفعاً 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 نسبياً، مثل (المتوسط الحسابي = 0.09 والانحراف المعياري = 0.021) عند الالتواء السالب الشديد بحجم عينة 100. وتُظهر طريقة كيرنل أيضاً نمطاً أكثر اتساقاً عبر أحجام العينات المختلفة، مما يشير إلى أنها تعطي أفضل التقديرات عبر أنواع التوزيعات. ومع زيادة حجم العينة، تُظهر كلتا الطريقتين اتجاهاً عاماً لتناقص الأخطاء المعيارية للتوزيعات الاعتدالية، وبالتالي تزيد دقة التقدير، لكن طريقة كيرنل تظل أكثر استقراراً وأقل حساسية للتغيرات في شكل التوزيع وحجم العينة مقارنة بطريقة دلتا.

وفي ضوء ما تقدم تم إجراء تحليل التباين المختلط ثلاثي الاتجاه وكذلك ثنائي الاتجاه لاختبار التفاعلات بين طرق التقدير وحجم العينة وتوزيع البيانات متبوعاً بحساب حجم التأثير

لجميع البيانات. وتم الكشف عن الفروق بين متوسطات الخطأ المعياري للتقدير باختلاف طريقة التقدير الدرجات (دلتا بالإطار الكامن DSM-L المعلمية، وكيرنل KS اللامعلمية) وحجم العينة (100، 250، 500، 1000) وتوزيع البيانات (اعتدالي، التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط، التواء سالب شديد). ويوضح الجدول (8) نتائج تحليل التباين المختلط الثلاثي الاتجاه لتقديرات القدرة باختلاف طريقة التقدير وحجم العينة وتوزيع البيانات التي تم التوصل إليها.

جدول 8. نتائج تحليل التباين المختلط الثلاثي الاتجاه لتقديرات القدرة باختلاف طريقة التقدير وحجم العينة وتوزيع البيانات (ن=9250)

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

** قيمة ف دالة إحصائية عند مستوى 0.01.

يوضح الجدول (8) نتائج تحليل التباين المختلط الثلاثي الاتجاهات وتأثيرات ثلاث متغيرات - توزيع البيانات وحجم العينة وطريقة التقدير - إلى جانب تفاعلاتها على المتغير التابع (دقة تقدير القدرة)؛ حيث يوجد تأثير دال إحصائيًا لتوزيع البيانات على دقة تقدير معلمة القدرة بطريقتي دلتا المعلمية وكيرنل اللامعلمية لتقدير الدرجات، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < .001$)؛ مما يشير إلى أن المستويات المختلفة للتوزيع لها تأثير كبير على دقة التقدير. وكانت الفروق لصالح العينات الاعتدالية حسب نتائج شيفيه. كما تشير قيمة مربع إيتا الجزئي (η^2) البالغة (0.188) إلى أنه يمكن إرجاع حوالي (19%) من التباين الإجمالي في دقة التقدير إلى تأثير توزيع البيانات، والذي يُعتبر حجم تأثير كبير تبعًا للمحكات المحددة في كوهين (Cohen, 1988).

وبالمثل، يُظهر حجم العينة أيضًا تأثيرًا مهمًا، حيث يوجد تأثير دال إحصائيًا لحجم العينة على دقة تقدير معلمة القدرة بطريقتي دلتا المعلمية وكيرنل اللامعلمية لتقدير الدرجات، حيث بلغت

قيمة "ف" والقيمة الاحتمالية ($F(3, 18460) = 21.131, P < .001$). وكانت الفروق لصالح العينات المتوسطة والكبيرة. 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 (2.9%) من التباين في دقة التقدير.

أما بالنسبة لطريقة التقدير بلغت قيمة "ف" والقيمة الاحتمالية ($F(1, 18460) = 102.255, P < .001$)، وكانت الفروق لصالح كيرنل في العينات الصغيرة، حسب نتائج شيفيه. كما تشير قيمة مربع إيتا الجزئي (η^2) لطريقة التقدير البالغة (0.047) إلى أن التأثير صغير نسبيًا، 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، حيث بلغت قيمة "ف" والقيمة الاحتمالية للتفاعل بين توزيع البيانات وحجم العينة ($F(12, 18460) = 53.684, P < .001$). وتشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.297) إلى أنه يفسر (29.7%) من التباين في دقة التقدير، وهو تأثير كبير تبعًا للمحكات المحددة في كوهين (Cohen, 1988)، حيث لوحظ أن توزيع البيانات كان له تأثير كبير ومستقل على دقة التقدير، وعند تفاعله مع حجم العينة كان لهم تأثير كبير. وهذا يعني أن شكل التوزيع يتأثر بحجم العينة وكل نوع من التواء البيانات يكون أكثر دقة في التقدير عند حجم معين. وكان التفاعل بين توزيع البيانات وطريقة التقدير مؤثرًا بشكل مماثل، حيث بلغت قيمة "ف" لهذا التفاعل والقيمة الاحتمالية ($F(4, 18460) = 53.917, P < .001$)، وتشير قيمة مربع إيتا الجزئي (η^2) لهذا التفاعل البالغة (0.099) إلى أنه يفسر (9.9%) من التباين في دقة التقدير مما يشير إلى أن التأثير المشترك لهذه المتغيرات يؤثر بشكل متوسط على دقة التقدير. وكان التفاعل بين الحجم والطريقة دالًا إحصائيًا أيضًا، وإن كان بحجم تأثير صغير حيث بلغت قيمة "ف" لهذا التفاعل والقيمة الاحتمالية ($F(1, 18460) = 18.903, P < .001$)، 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 كان دالًا إحصائيًا، حيث بلغت قيمة "ف" لهذا التفاعل الثلاثي والقيمة الاحتمالية ($F(12, 18460) = 56.531, P < .001$)، 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.

كما تم إجراء تحليل التباين الثنائي الاتجاه لمقارنة تأثير أحجام العينات والطرق المختلفة على تقدير القدرة عبر أنواع التوزيع للبيانات، ويوضح الجدول (9) النتائج المهمة فيما يتعلق بالتأثيرات الرئيسية وتفاعلها.

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

توزيع اعتدالي (ن=1850)						
مصدر التباين	مجموع المربعات	متوسط مجموع المربعات	درجات الحرية	قيمة "ف"	الدلالة الإحصائية	مربع إيتا الجزئي
حجم العينة	0.101	0.034	3	**93.805	P < .001	.0314
طريقة التقدير	0.001	0.001	1	1.607	.205	.0012
الحجم×طريقة التقدير	0.076	0.025	3	**71.084	P < .001	.0238
التواء موجب بسيط (ن=1850)						
المتغيرات	مجموع المربعات	متوسط مجموع المربعات	درجات الحرية	قيمة "ف"	الدلالة الإحصائية	مربع إيتا الجزئي
حجم العينة	0.778	0.259	3	**107.641	P < .001	.0393
طريقة التقدير	0.042	0.042	1	**17.437	P < .001	.0021
الحجم×طريقة التقدير	0.753	0.251	3	**104.269	P < .001	.0381
التواء موجب شديد (ن=1850)						
المتغيرات	مجموع المربعات	متوسط مجموع المربعات	درجات الحرية	قيمة "ف"	الدلالة الإحصائية	مربع إيتا الجزئي
حجم العينة	0.777	0.259	3	**72.182	P < .001	.0273
طريقة التقدير	0.071	0.071	1	**19.763	P < .001	.0025
الحجم×طريقة التقدير	1.141	0.38	3	**105.971	P < .001	.04
التواء سالب بسيط (ن=1850)						
المتغيرات	مجموع المربعات	متوسط مجموع المربعات	درجات الحرية	قيمة "ف"	الدلالة الإحصائية	مربع إيتا الجزئي
حجم العينة	0.259	0.086	3	**25.167	P < .001	.0095

طريقة التقدير	0.274	0.274	1	**79.715	P < .001	.01
الحجم×طريقة التقدير	0.57	0.19	3	**55.339	P < .001	.0208
التواء سالب شديد (ن=1850)						
المتغيرات	مجموع المربعات	متوسط مجموع المربعات	درجات الحرية	قيمة "ف"	الدالة الإحصائية	مربع إيتا الجزئي
حجم العينة	0.038	0.013	3	2.378	.068	.002
طريقة التقدير	0.717	0.717	1	**133.833	P < .001	.03
الحجم×طريقة التقدير	0.014	0.005	3	0.871	.455	.001

** قيمة ف دالة إحصائية عند مستوى 0.01.

ويتضح من الجدول (9) أنه في حالة التوزيع الاعتدالي، فإن تأثير حجم العينة كان دالاً إحصائياً، كما هو موضح من خلال قيمة "ف" والقيمة الاحتمالية ($P < .001$, $F(3,1846) = 93.805$)، ويعكس متوسط مجموع مربعات الحجم البالغ (0.034) قدرًا من التباين الذي يفسره هذا المتغير، وتدعم قيمة "ف" أن الحجم له تأثير دال إحصائياً على دقة تقدير القدرة، وكان لصالح العينات الكبيرة، حسب نتائج شيفيه. وتشير قيمة مربع إيتا الجزئي التي بلغت (0.0314) إلى أنه يمكن إرجاع ما يقارب من (3.14%) من التباين في دقة التقدير إلى الاختلافات بين حجم العينة. على الرغم من أن هذا قد يبدو صغيراً تبعاً للمحكات المحددة في كوهين (Cohen, 1988)، إلا أنه يمثل تأثيراً ذا مغزى في سياق الدراسة. على النقيض من ذلك، فإن تأثير طريقة التقدير ليس دالاً إحصائياً، حيث تشير قيمة "ف" والقيمة الاحتمالية ($P = 0.205$, $F(1,1846) = 1.607$) إلى عدم وجود تأثير دال إحصائياً لطريقة التقدير على دقة تقدير القدرة. كما أن مربع إيتا الجزئي للطريقة صغير أيضاً وفسر نسبة صغيرة جداً (0.02%) من التباين بالطرق المختلفة للتقدير، مما يجعلها غير ذات أهمية عملياً. وأخيراً، فإن التفاعل بين الحجم وطريقة التقدير كان تأثيره دالاً إحصائياً، حيث كانت قيمة "ف" والقيمة الاحتمالية ($P < .001$, $F(3,1846) = 71.084$). وهذا يشير إلى أن تأثير حجم العينة على دقة التقدير يختلف وفقاً للطريقة المستخدمة، وأن هذا التفاعل يفسر قدرًا من التباين. كما أن مربع إيتا الجزئي للتفاعل بلغ (0.0238)، مما يعني أنه يمكن عزو (2.38%) من التباين في دقة التقدير إلى التفاعل بين الحجم وطريقة التقدير، وهي نسبة صغيرة. وبشكل عام، في حين أن حجم العينة كان له تأثير كبير على النتيجة، فإن طريقة التقدير وحدها لم يكن لها تأثير بنفس القدر. ومع ذلك، فإن الجمع بين حجم العينة وطريقة التقدير نتج عنه تأثير تفاعلي كبير، مما يظهر أن تأثير حجم العينة يختلف اعتماداً على طريقة التقدير المستخدمة.

وفي حالة الالتواء الموجب البسيط، تشير نتائج الجدول (9) لتحليل التباين ثنائي الاتجاه إلى تأثيرات دالة إحصائية لحجم العينة وطريقة التقدير وتفاعلها على دقة تقدير القدرة. ويظهر متغير حجم العينة تأثيراً على دقة التقدير، والذي يعكس من خلال قيمة "ف" والقيمة الاحتمالية () ،

$P < 0.001$ ($F(3,1846) = 107.641$). كما أن متوسط مجموع مربعات الحجم (0.259) كبير وهذا يؤكد أن حجم العينة يفسر جزءاً كبيراً من التباين. ويشير مربع إيتا الجزئي البالغ (0.039) إلى أن (3.93%) من التباين يمكن تفسيره من خلال الاختلافات في حجم العينة، وهي نسبة صغيرة تبعاً للمحكات المحددة في كوهين (Cohen, 1988). لذا فإن حجم العينة يلعب دوراً رئيسياً في التأثير على دقة تقدير القدرة داخل هذا التوزيع. أما بالنسبة لطريقة التقدير فكان تأثيرها دال إحصائياً أيضاً، من خلال قيمة "ف" والقيمة الاحتمالية ($P < 0.001$, $F(1,1846) = 17.437$). وعلى الرغم من أن متوسط مجموع مربعات طريقة التقدير (0.042) أصغر مقارنة بتأثير حجم العينة، إلا أنه لا يزال له تأثير. كما أن مربع إيتا الجزئي للطريقة كان صغيراً (0.0021)، مما يعني أن (0.21%) من التباين يمكن تفسيره من خلال الاختلافات في طريقة التقدير. وفي حين أن هذا التأثير صغير، إلا أنه ذا دلالة إحصائية ويشير إلى أن طريقة التقدير المستخدمة لها بعض التأثير على دقة تقدير القدرة. كما نجد أن التفاعل بين حجم العينة وطريقة التقدير كان ذا دلالة إحصائية أيضاً، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$, $F(3,1846) = 104.269$). كما يُشير متوسط مجموع مربعات التفاعل (0.251) إلى أن تأثير التفاعل هذا يفسر جزءاً ملحوظاً من التباين. ويوضح مربع إيتا الجزئي البالغ (0.0381) أن (3.81%) من التباين يمكن أن يُعزى إلى التفاعل بين حجم العينة وطريقة التقدير، وهي نسبة بسيطة، مما يشير إلى أن تأثير الحجم يتأثر بطريقة التقدير المستخدمة. وباختصار، تُظهر النتائج أن كلاً من حجم العينة وطريقة التقدير يؤثران بشكل دال إحصائياً وكبير على النتيجة في حالة الالتواء الموجب البسيط، حيث يكون للحجم تأثير أكبر على دقة تقدير القدرة. بالإضافة إلى ذلك، يشير التفاعل الكبير بين الحجم وطريقة التقدير إلى أن تأثير حجم العينة يختلف حسب طريقة التقدير المطبقة.

وفي حالة الالتواء الموجب الشديد، تكشف نتائج تحليل التباين ثنائي الاتجاه في الجدول (9) عن تأثيرات كبيرة لحجم العينة وطريقة التقدير وتفاعلها على دقة تقدير القدرة. ويؤثر حجم العينة بشكل كبير على دقة التقدير، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$, $F(3,3692) = 72.182$). كما بلغت قيمة متوسط مجموع المربعات (0.259)، مما يُظهر تأثيراً قوياً لحجم العينة على تفسير التباين. كما يشير مربع إيتا الجزئي البالغ (0.0273) إلى أن (2.73%) من التباين يمكن تفسيره باختلافات في حجم العينة، وهي نسبة صغيرة. هذا التأثير، على الرغم من كونه معتدلاً، إلا أنه ملحوظ ويسلط الضوء على أهمية حجم العينة في التأثير على دقة التقدير. وبالنسبة لتأثير طريقة التقدير كان دالاً إحصائياً أيضاً، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$, $F(1,3692) = 19.763$). وتشير قيمة متوسط مجموع المربعات البالغة (0.071) إلى أنها تساهم في تفسير التباين. كما أن قيمة مربع إيتا الجزئي بلغ (0.0025)، مما يعني أن (0.25%) من التباين يمكن تفسيره بطريقة التقدير. هذا التأثير صغير، مما يشير إلى أن طريقة التقدير لها تأثير بسيط ولكن يمكن ملاحظته على النتيجة. أما التفاعل بين حجم العينة وطريقة التقدير فكان أيضاً مهماً حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$, $F(3,3692) = 105.971$). كما بلغت قيمة متوسط مجموع المربعات للتفاعل هو (0.38)، وتوضح قيمة اختبار ف المرتفعة تأثير تفاعل قوي. كما يوضح مربع إيتا الجزئي البالغ (0.04) أن (4%) من

التباين يمكن تفسيره بالتفاعل بين حجم العينة وطريقة التقدير، وهذا تأثير صغير. ويشير ذلك إلى أن تأثير حجم العينة يتأثر بشكل كبير بطريقة التقدير. وباختصار، تشير النتائج إلى أن كلا من حجم العينة وطريقة التقدير يؤثران بشكل كبير على النتيجة في حالة الالتواء الموجب الشديد، حيث يكون لحجم العينة تأثير كبير إلى حد ما في دقة تقدير القدرة. والتفاعل بين حجم العينة وطريقة التقدير قوي بشكل خاص، مما يشير إلى أن تأثير حجم العينة يعتمد بشكل كبير على طريقة التقدير المستخدمة.

وفي حالة الالتواء السالب البسيط، تشير نتائج تحليل التباين ثنائي الاتجاه في جدول (9) لتقييم تأثيرات الحجم وطريقة التقدير وتفاعلها على دقة تقدير القدرة، أن لحجم العينة تأثير مهم ودال إحصائياً، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$)، $F(3, 1846) = 25.167$ ، كما بلغت قيمة متوسط مجموع مربعات الحجم هو (0.086)، مما يشير إلى أن الاختلافات في حجم العينة تؤثر على دقة التقدير. وبلغت قيمة مربع إيتا الجزئي للحجم هو (0.0095)، مما يشير إلى أن (0.95%) من التباين في دقة التقدير يتم تفسيره بواسطة حجم العينة. في حين أن هذا تأثير صغير، وهو يشير إلى أن الحجم لا يزال يلعب دوراً في التأثير على النتائج، على الرغم من أن مساهمته في التباين الكلي طفيفة نسبياً. أما بالنسبة لطريقة التقدير فإن لها تأثير دال إحصائياً، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$)، $F(1, 1846) = 79.715$ ، كما بلغت قيمة متوسط مجموع المربعات (0.274)، مما يشير إلى تأثير كبير لطريقة التقدير على تفسير التباين. وبلغت قيمة مربع إيتا الجزئي للطريقة (0.01)، مما يشير إلى أن (1%) من التباين يتم تفسيره بواسطة طريقة التقدير. وهذا يمثل تأثيراً صغيراً ولكنه ملحوظ، مما يدل على أن طريقة التقدير تؤثر بشكل هادف على دقة التقدير في هذا التوزيع. وكذلك التفاعل بين حجم العينة وطريقة التقدير دال إحصائياً أيضاً، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P < 0.001$)، $F(3, 1846) = 55.339$ ، كما بلغت قيمة متوسط مجموع المربعات للتفاعل هو (0.19). وبلغت قيمة مربع إيتا الجزئي للتفاعل (0.0208)، مما يشير إلى أن (2.08%) من التباين في دقة التقدير يمكن أن يعزى إلى التفاعل بين حجم العينة وطريقة التقدير. وهذا يشير إلى أن تأثير الحجم يعتمد على طريقة التقدير المستخدمة، ويلعب هذا التفاعل دوراً مهماً إلى حد ما في تفسير التباين في دقة تقدير القدرة. وباختصار، تشير نتائج الالتواء السالب البسيط إلى أن كلا من حجم العينة وطريقة التقدير يؤثران بشكل كبير على دقة تقدير القدرة، حيث يكون لطريقة التقدير تأثير أقوى قليلاً من حجم العينة. كما يلعب التفاعل بين الحجم وطريقة التقدير دوراً حاسماً، حيث يفسر جزءاً معتدلاً من التباين. وتسلط هذه النتائج الضوء على أن تأثير حجم العينة على دقة تقدير القدرة يتأثر بطريقة التقدير المستخدمة، مما يؤكد على أهمية النظر في كلا العاملين معاً.

وفي حالة الالتواء السالب الشديد، تشير نتائج تحليل تباين ثنائي الاتجاه لفحص تأثيرات حجم العينة وطريقة التقدير وتفاعلها على دقة تقدير القدرة في جدول (9)، إلى عدم تأثير حجم العينة بشكل دال إحصائياً وإن كان شديد الاقتراب من ذلك، حيث بلغت قيمة "ف" والقيمة الاحتمالية ($P = 0.068$)، $F(3, 1846) = 2.378$ ، وهي أكبر قليلاً من ($\alpha = 0.05$). كما بلغت

قيمة متوسط مجموع المربعات (0.013). وبلغت قيمة مربع إيتا الجزئي للحجم (0.002)، مما يشير إلى أن حجم العينة يمثل (0.2%) من التباين في دقة التقدير وهذا تأثير صغير جداً تبعاً 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 مساهماً رئيسياً في هذا التوزيع، والأدلة ليست قوية بما يكفي للقول بأن له تأثير قوي. من ناحية أخرى، فإن طريقة التقدير لها تأثير دال إحصائياً، حيث بلغت قيمة "ف" والقيمة الاحتمالية (0.001) $P < 133.833$, $F(1, 1846)$ (=). كما بلغت قيمة متوسط مجموع المربعات (0.717)، مما يشير إلى تأثير قوي على دقة التقدير. وبلغت قيمة مربع إيتا الجزئي للطريقة (0.03)، مما يعني أن (3%) من التباين في دقة التقدير يتم تفسيره بواسطة متغير طريقة التقدير. ويمكن اعتباره تأثيراً كبيراً وأكثر وضوحاً مقارنة بتأثير حجم العينة ويظهر أن طريقة التقدير هي متبني مهم في هذا التوزيع. ولم يكن التفاعل بين حجم العينة وطريقة التقدير دالاً إحصائياً، حيث بلغت قيمة "ف" والقيمة الاحتمالية (0.455) $P = 0.871$, $F(3, 1846)$ ، وهي أكبر من ($\alpha = 0.05$). كما بلغت قيمة متوسط مجموع المربعات (0.005). وبلغت قيمة مربع إيتا الجزئي للتفاعل (0.001)، مما يشير إلى أن التفاعل يفسر نسبة بسيطة جداً ولا تُذكر من التباين في دقة تقدير القدرة. وتشير هذه النتيجة إلى أن تأثير حجم العينة 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 دال إحصائياً ومهم على دقة التقدير للقدرة، في حين يظهر تأثير حجم العينة صغيراً وغير دال. وكان التفاعل بين الحجم وطريقة التقدير ليس دالاً إحصائياً ويفسر القليل جداً من التباين، مما يشير إلى أن هذين العاملين لا يتفاعلا بطريقة ذات مغزى في هذه المجموعة من البيانات. والنتيجة الرئيسية هي التأثير المهيمن للطريقة على النتيجة.

مناقشة النتائج والخلاصة والتوصيات:

أوضحت النتائج أن طريقة دلتا (DSM-L) المعملية تظهر تبايناً أعلى في الخطأ المعياري للتقدير (SEE) مقارنة بطريقة كيرنل (KS) اللامعملية عبر جميع أحجام العينات والتوزيعات. وبالنسبة لطريقة دلتا، يميل الخطأ المعياري (SEE) لتقدير القدرة إلى الزيادة في التوزيعات الملتوية، وخاصة في التوزيعات الملتوية الموجبة والسالبة الشديدة، مع ملاحظة انحرافات معيارية أكبر. وعلى وجه التحديد، بالنسبة لحجم عينة (100)، تُظهر طريقة دلتا متوسطات وانحرافات معيارية عالية نسبياً، وخاصة للتوزيعات الملتوية الشديدة. ومع زيادة حجم العينة، ينخفض الخطأ المعياري للتقدير لطريقة دلتا للتوزيعات الاعتدالية ولكنه يظل مرتفعاً للتوزيعات الملتوية. في المقابل، تعطي طريقة كيرنل غالباً متوسطات أقل للخطأ المعياري للتقدير، بغض النظر عن التوزيع أو حجم العينة. حتى في التوزيعات الملتوية، تحافظ طريقة كيرنل على أخطاء معيارية مستقرة ومنخفضة نسبياً. ومع زيادة حجم العينة، تُظهر كلتا الطريقتين اتجاهاً عاماً لتناقص الأخطاء المعيارية للتقدير في التوزيعات الاعتدالية، وبالتالي تزيد دقة التقدير، لكن طريقة كيرنل تظل أكثر استقراراً وأقل حساسية للتغيرات في شكل التوزيع وحجم العينة مقارنة بطريقة دلتا. وهذا يتفق مع ما

توصلت اليه عدة دراسات تناولت موضوع دقة التقدير، منها: دراسة بني عطا والشريفين (2012)، والحواري (2015)، وعلوان وجاسم (Alwan & Jasim, 2022)، بأن الانحراف عن التوزيع الاعتدالي يزيد من الخطأ المعياري وبالتالي تقل دقة التقدير، ودراسة عفتبانه وبغائي (Effatpanah & Baghaei, 2022) لدقة تقديرات كيرنل (KS).

كما أوضحت نتائج تحليل التباين المختلط ثلاثي الاتجاه لتأثيرات توزيع البيانات والحجم وطريقة التقدير إلى جانب تفاعلاتها على المتغير التابع (دقة التقدير للقدرة)؛ وجود تأثير دال احصائيًا لتوزيع البيانات على دقة تقدير معلمة القدرة بطريقتي دلتا المعلمية وكيرنل اللامعلمية لتقدير الدرجات. وهذه النتيجة تتفق مع ما توصلت اليه كثير من الدراسات، منها: بني عطا والشريفين (2012) بأنه يوجد أثر لاختلاف شكل توزيع القدرة على معلمات الفقرات، حيث تباينت عبر التوزيعات المختلفة لقدرات الأفراد، ودراسة الحواري (2015) والتي أشارت إلى وجود فروق ذات دلالة إحصائية بين تقديرات القدرة تُعزى لشكل التوزيع، حيث كانت أكثر دقة في التوزيعات الملتوية الموجبة والسالبة، ودراسة علوان وجاسم (Alwan & Jasim, 2022) التي توصلت لاختلاف تقديرات القدرة تبعًا لتوزيع البيانات.

كما أوضحت النتائج وجود تأثير دال احصائيًا لحجم العينة على دقة تقدير معلمة القدرة بطريقتي دلتا المعلمية وكيرنل اللامعلمية لتقدير الدرجات. وهذه النتيجة أتت متوافقة مع كثير من الدراسات التي أكدت تأثير حجم العينة على دقة التقدير، ومنها دراسة الشافعي (2024) التي أظهرت أن زيادة حجم العينة يساعد على الوصول إلى دقة تقدير قدرة الأفراد، ودراسة ضعضع وآخرين (2020) والتي توصلت إلى أن دقة التقدير تزداد بزيادة حجم العينة، ودراسة البادية وآخرين (2018) والتي توصلت إلى أن زيادة حجم العينة يزيد من دقة تقدير القدرة، كما تتفق مع بني عطا (2014) والشريفين (2012) في أن دقة تقدير القدرة تتأثر بحجم العينة وطريقة التقدير، والتفاعل بينهما. وتختلف مع دراسة شما (2013) في أن حجم العينة ليس له تأثير في دقة تقدير قدرات الأفراد.

كما أظهرت النتائج عدم وجود تأثير دال احصائيًا لطريقة التقدير على دقة تقدير معلمة القدرة، وهذا يعني أن طريقة التقدير قد لا يكون لها تأثير مستقل على دقة التقدير. وهذه النتيجة أتت متوافقة مع ما توصلت اليه دراسة وروبينزش (Robitzsch, 2021)، والشريفين (2012) بأنه لا يوجد فروق دالة احصائية تُعزى لطريقة التقدير المستخدمة، بينما اختلفت مع كل من دراسة ضعضع وآخرين (2020)، وبني عطا (2017) حيث أكدت على وجود فروق ذات دلالة إحصائية في دقة تقدير معلمة القدرة تُعزى لطريقة التقدير المستخدمة.

كما كان للتفاعل بين توزيع البيانات وحجم العينة تأثير كبير على دقة التقدير. وهذه النتيجة اتفقت مع دراسة الشرفاوي (2022)، التي توصلت إلى أن دقة تقدير القدرة تتأثر بمستويات حجم العينة عند التوزيعات الاعتدالية والملتوية التواء سالب، ودراسة الخرشة (2018)، التي توصلت إلى وجود أثر دال احصائيًا لكل من حجم العينة واختلاف شكل توزيع القدرة على دقة التقدير. أيضًا بالنسبة للتفاعل بين توزيع البيانات وطريقة التقدير كان له تأثير دال على دقة التقدير. أما بالنسبة

للتفاعل بين حجم العينة وطريقة التقدير كان التأثير محدودا على التباين في دقة التقدير، وهذه النتيجة تتفق مع دراسة كل من ضعضع وآخرين(2020) التي توصلت إلى وجود دلالة إحصائية عند تقدير قدرة الأفراد وفق حجم العينة وطريقة التقدير والتفاعل بينهما، بينما تختلف النتيجة مع دراسة الشرفاوي(2022) التي توصلت إلى أن دقة تقدير قدرة الأفراد لا تتأثر بتفاعل حجم العينة وطرق التقدير التي أستخدمت. وأخيراً، فإن التفاعل الثلاثي بين توزيع البيانات وحجم العينة وطريقة التقدير يشير إلى حجم تأثير كبير. وبشكل عام، تُظهر النتائج تفاعلات معقدة بين المتغيرات، مع بعض أحجام التأثير المعتدلة التي تفسر التباين في دقة التقدير عبر ظروف مختلفة.

كما أظهرت نتائج تحليل التباين الثنائي لمقارنة تأثير متغيري حجم العينة وطريقة التقدير وتفاعلهما على دقة التقدير للقدرة ضمن توزيعات البيانات المختلفة، أنه في التوزيع اعتدالي تبين أن حجم العينة له تأثير قوي على دقة التقدير. بينما كان تأثير طريقة التقدير غير دال إحصائياً. كما أن التفاعل بين حجم العينة وطريقة التقدير دال إحصائياً وله تأثير صغير على دقة التقدير. أما نتائج التحليل في حالة توزيع البيانات (التواء موجب بسيط، التواء موجب شديد، التواء سالب بسيط) فكانت ذات دلالة إحصائية لوجود تأثير لكل من حجم العينة وطريقة التقدير والتفاعل بينهما على دقة تقدير القدرة. وفي حالة الالتواء السالب الشديد تشير النتائج إلى أن الطريقة لها تأثير دال إحصائياً على دقة التقدير، في حين يظهر الحجم تأثيراً صغيراً وغير مهم. والتفاعل بين الحجم والطريقة ليس مهماً ويفسر القليل جداً من التباين في دقة التقدير، مما يشير إلى أن هذين المتغيرين لا يتفاعلا بطريقتهم ذات مغزى في هذه المجموعة من البيانات.

والنتيجة الرئيسية هي التأثير المهيمن للطريقة على النتيجة، وأنه في حالة الالتواءات الشديدة السلبية نحتاج إلى اختيار الطريقة الأفضل من طرق التقدير المستخدمة للوصول إلى أدق التقديرات. وكانت الفروق لصالح طريقة كيرنل تارة ولصالح طريقة دلتا تارة أخرى باختلاف حجم العينة والتوزيع ومالت لصالح كيرنل في العينات الصغيرة الملتوية التوزيع. وتتفق النتيجة مع دراسة مورديريك وكوتش (Mor Dirlik & Koç, 2019) التي أوضحت اتفاقاً بين النماذج المعلمية (PIRT) واللامعلمية (NIRT) في دقة التقدير حال حجم العينات الكبير، وكلما زاد حجم العينة؛ كانت التقديرات أكثر اتساقاً. كما تتفق جزئياً مع الشاورة (2017) في أن تقدير معلمة القدرة كان أفضل بالطريقة المعلمية مقارنة بطريقة كيرنل (KS) اللامعلمية، عندما كان حجم العينة (250 و500)، بينما لم توجد فرق دالة إحصائية عند حجم العينة (1500) فرداً. وتتفق مع دراسة القيسي (2016) في أن هناك فروق في دقة تقدير معلمة القدرة تعزى لحجم العينة عند استخدام طريقة كيرنل (KS). كما لا تتفق مع دراسة القيسي (2013) في أن دقة المعالم المقدرة بطريقة التقدير المعلمية أفضل من المعالم المقدرة وفق طريقة التقدير اللامعلمية.

كذلك تدعم النتائج دراسة ديمتروف وأتانسوف (Dimitrov & Atanasov, 2021) التي أشارت لدقة تقديرات القدرة وفق لطريقة دلتا (DSM-L)؛ في الظروف الاختبارية وأحجام العينات المختلفة (500-1000). ويؤيد ذلك ما أشار له دومينكو وديميتروف (Dimitrov, 2015) & Domingue) و روبيتزش (Robitzsch, 2021) من أن قيم معلمة قدرة الأفراد المقدرة بأسلوب

الدلتا (DSM-L) تعطي نتائج إحصائية متشابهة مع طرق نظرية الاستجابة للفقرة التقليدية؛ وتنتج انتهاكات أقل لمُسلّمات القياس الموحد منها.

ويمكن أن تعزى نتائج الدراسة للأثر الواضح لكل من حجم العينة وتوزيع البيانات وخاصة عند الابتعاد عن الاعتدالية، حيث أن دقة التقدير تتأثر بعوامل مختلفة، بما في ذلك توزيع القدرة، وحجم العينة، فعندما ينحرف توزيع القدرة بشكل كبير عما هو مفترض في النموذج اللوجستي، فقد يؤدي ذلك إلى انتهاكات لافتراضات النموذج ويؤثر على دقة التقدير (Hambleton & Swaminathan, 2010).

بالرغم من أهمية النتائج التي توصل إليها البحث الحالي في الكشف عن دقة تقدير قدرة الأفراد باستخدام نظرية الاستجابة للفقرة وفقاً لطريقة دلتا بالإطار الكامن (DSM-L) المعلمية، وكيرنل (KS) اللامعلمية، إلا أن هناك بعض العوامل المؤثرة الأخرى التي قد تلعب دوراً مهماً في دقة تقدير قدرات الأفراد ولم يتم ضبطها في البحث الحالي مثل تقصي دقة التقدير في أطوال الاختبارات ونسبة البيانات المفقودة، مما قد يسهم في إثراء النتائج.

التوصيات:

بناءً على نتائج الدراسة تم تقديم التوصيات التالية:

1. بشكل عام، يمكن استخدام طريقة دلتا (DSM-L) المعلمية أو طريقة كيرنل (KS) اللامعلمية عند تقدير القدرة حيث هناك اتفاق كبير بينها في دقة التقدير، وكلما زاد حجم العينة؛ كانت التقديرات أكثر انساقاً.
2. يوصى باستخدام طريقة دلتا (DSM-L) المعلمية وفق الإطار الكامن في تحليل فقرات الاختبارات ثنائية التدرج، خاصة في العينات الكبيرة (500 فأكثر) والاعتدالية، نظراً لدقة تقديراتها، حيث أشارت نتائج الدراسة الحالية إلى صغر قيم الأخطاء المعيارية لتقدير قدرات الأفراد الناتجة عنها وبالتالي زيادة دقة التقدير، وعدم وجود فروق دالة إحصائية في التقديرات في ضوء هذه الطريقة وطريقة كيرنل (KS) اللامعلمية، وتتفوق طريقة دلتا بأنها لا تحتاج إلى التحقق من العديد من الافتراضات مثل نماذج نظرية الاستجابة للفقرة.
3. يفضل اختيار الطريقة المناسبة حسب حجم العينة والتوزيع عند تقدير القدرة للوصول إلى أدق التقديرات، فمع العينات الصغيرة (أقل من 500) والملتوية وخاصة الموجبة والسالبة الشديدة، يفضل استخدام طريقة كيرنل (KS) اللامعلمية، حيث تعطي غالباً متوسطات أقل للخطأ المعياري للتقدير، بغض النظر عن التوزيع أو حجم العينة، مقارنة بطريقة دلتا (DSM-L) المعلمية.
4. أفضلية زيادة حجم العينة عند استخدام طريقة دلتا (DSM-L) المعلمية وطريقة كيرنل (KS) اللامعلمية، للحصول على تقديرات دقيقة للقدرة حيث يمكن اعتبار زيادة حجم العينة عاملاً إيجابياً في دقة التقدير وخفض الخطأ المعياري للتقدير.

-
5. إمكانية استخدام عينات ذات أحجام متفاوتة، مشابهة للأحجام المستخدمة في الدراسة الحالية، عند تطبيق طريقة دلتا (DSM-L) المعلمية وطريقة كيرنل (KS) اللامعلمية، لتقدير الدرجات في الدراسات المستقبلية لتحليل مفردات المقاييس. فقد أظهرت نتائج الدراسة الحالية جودة مطابقة البيانات بشكل عام سواء للأفراد أو الفقرات لهذين النموذجين باختلاف حجم العينة.
6. أهمية استخدام حجم التأثير لدعم النتائج والقرارات الناتجة من الدلالة الإحصائية، لما له من أهمية عملية في إيضاح أهمية النتائج التي تم التوصل إليها وتفسير معناها بشكل شامل للوصول لفهم أعمق للظواهر موضع الدراسة.
7. التوسع في استخدام البيانات المولدة بالمحاكاة في الدراسات لاختيار طرق التقدير تبعا للمتغيرات المختلفة، حيث تساهم في ضبط المتغيرات المطلوب دراستها والتخلص من تأثير بعض العوامل على استجابات الأفراد مثل الغش والتخمين، واللامبالاة.

المقترحات والدراسات المستقبلية:

يمكن تقديم المقترحات التالية:

1. دراسة أثر طرق التقدير باستخدام النماذج اللوجستية الثنائية أو المتعددة الاستجابة الأخرى على دقة تقدير قدرة الأفراد ومعلمات الفقرة.
2. دراسة أثر طرق التقدير على دقة تقدير قدرة الأفراد ومعلمات الفقرة تحت ظروف مختلفة من أشكال التوزيع، كالتوزيع المنتظم والتوزيع بيتا والتوزيع الأسّي وغيرها.
3. إجراء دراسات مشابهة باستخدام متغيرات أخرى مثل أطوال مختلفة من الاختبارات، وأساليب قياس دقة تقدير أخرى.
4. إجراء المزيد من الدراسات المماثلة عبر مستويات قدرات الأفراد المختلفة، حيث قد تؤثر على مطابقة الفقرات ودقة التقدير.
5. إجراء المزيد من الدراسات التجريبية على نماذج دلتا لتقدير الدرجات باستخدام عينات صغيرة الحجم للكشف عن كيفية عمل هذه النماذج.

المراجع:

المراجع العربية:

- البادية، فاطمة حمد، ابن كاظم، علي بن مهدي، والمحززي، راشد بن سيف (2018). أثر حجم العينة على دقة تقدير خصائص المفردة والقدرة في اختبار التنمية المعرفية في مادة العلوم لطلبة الصف السابع بسلطنة عمان. *دراسات*، 73، 106-125.
<http://search.mandumah.com/Record/947730>
- بني عطا، زايد، والشريفين، نضال (2012). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. *المجلة الأردنية في العلوم والتربية*، 8(2)، 151-166.
<http://search.mandumah.com/Record/444491>
- بني عطا، زايد صالح (2014). 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*، 11(2)، 1-37.
<http://search.mandumah.com/Record/810211>
- الحواري، أروى. (2015). أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وفق نموذج راش في نظرية استجابة الفقرة. *مجلة جامعة النجاح للأبحاث - العلوم الإنسانية*، 29(8)، 1463-1488.
<http://search.mandumah.com/Record/931120>
- الخرشة، طه عقلة (2018). أثر حجم العينة وشكل توزيع قدرة المفحوصين على معالم الفقرة عند استخدام النموذج اللوغارتمي الثلاثي. *مجلة العلوم التربوية والنفسية*، 19(4)، 397-424.
<http://search.mandumah.com/Record/948502>
- دي إيلا، آر. جي (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة* (عبد الله الكيلاني وإسماعيل البرصان، مترجمان). دار جامعة الملك سعود للنشر. (نشر العمل الأصلي في 2009).
- الشافعي، محمد محمود (2024). تأثير حجم العينة على دقة تقديرات: معامل صعوبة فقرات الاختبار، ومعامل القدرة للأفراد في عينة التحليل باستخدام النموذج اللوجستي ثلاثي المعامل. *مجلة الإسكندرية التربوية*، 34(2)، 191-210.
<https://dx.doi.org/10.21608/jealex.2024.371247>
- الشافعي، محمد منصور (2014). تأثير انتهاك بعض افتراضات نظرية الاستجابة للمفردة على تدرج بنك الأسئلة ودقة معادلة درجات الاختبارات البنكية المسحوبة. *مجلة كلية التربية بجامعة بنها*، 25(98)، 431-498.
- الشرفاوي، عبد الكريم (2022). طرق تقدير معالم الفقرة والقدرة وأثرها في دقة التقدير باستخدام نظرية الاستجابة للمفردة راش أنموذجًا [رسالة دكتوراة غير منشورة]. جامعة الجزائر.

- الشريفيين، نضال كمال (2006). الخصائص السيكومترية لاختبار محكي المرجع في القياس والتقويم التربوي وفق النظرية الحديثة في القياس التربوي والنفسي. *مجلة العلوم التربوية والنفسية*، 7 (4)، 79-109. <http://search.mandumah.com/Record/2445>
- الشريفيين، نضال كمال (2012). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. *المجلة التربوية*، 26 (104)، 177-238. <http://search.mandumah.com/Record/415514>
- الشريفيين، نضال كمال، ومناصرة، سوسن عاطف (2017). خصائص توزيع قدرات الأفراد ومعالم فقرات الاختبار وفق نماذج نظرية الاستجابة للاستجابة للفقرة العملية واللامعملية: دراسة مقارنة. *المجلة التربوية*، 31 (124)، 265-320. <http://search.mandumah.com/Record/832513>
- شما، يمان نزار (2013). أثر حجم العينة على دقة تقدير صعوبة المفردات وقدرة الأفراد باستخدام نموذج راش. *مجلة الآداب*، 105، 673-698. <http://search.mandumah.com/Record/666339>
- الشواورة، شادي يوسف (2017). الفاعلية النسبية لطريقة بيزز البارامترية وطريقة تهذيب النواة اللابارامترية في تقدير معلمة القدرة وفق النموذج اللوجستي الثنائي باختلاف حجم العينة. *العلوم التربوية*، 25 (4)، 338-361. <http://search.mandumah.com/Record/918216>
- ضعضع، هبة عبد اللطيف، طومان، منار أحمد، وطيفور، مصطفى أحمد (2020). أثر حجم العينة وطرائق التقدير في دقة تقدير معالم نموذج راش. *جروش للبحوث والدراسات*، 21 (1)، 131-170. <http://demo.mandumah.com/Record/1060646>
- القيسي، حسين عبد النبي (2013). 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 (2016). 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، 31 (5)، 203-246. <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 Skewed 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](https://doi.org/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 Q₃: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied psychological measurement*, 41(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- 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](https://doi.org/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](https://doi.org/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*. 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>
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, 16(2), 878. <https://www.rasch.org/rmt/rmt162f.htm>
- Maydeu-Olivares, A., & Joe, H. (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2ⁿ Contingency Tables: A Unified Framework. *Journal of the American Statistical Association*, 100(471), 1009–1020. <https://doi.org/10.1198/016214504000002069>
- Mazza, A., Punzo, A., & McGuire, B. (2014). KernSmoothIRT: An R package for kernel smoothing in item response theory. *Journal of Statistical Software*, 58(6), 1-34. <http://www.jstatsoft.org/v58/06/>
- Mazza, A., Punzo, A., & McGuire, B. (2020). KernelSmoothIRT: Nonparametric Item Response Theory [Computer software] R package version 6.4. <https://cran.r-project.org/web/packages/KernSmoothIRT/index.html>
- 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* [Unpublished Doctoral dissertation]. 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](https://doi.org/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](https://doi.org/10.1177/0013164415576958)
- Reise, S. P., & Revicki, D. A. (Eds.). (2015). *Handbook of item response theory modeling: Applications to typical performance assessment*. Routledge/Taylor & Francis Group.
- 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](https://doi.org/10.1037/0022-3514.50.6.1123)
- 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.
- Van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties* [Comparability of individual test performance]. Swets & Zeitlinger.
- Zickar, M. (1997). *Identifying Untraited Individuals Using Model based Measurement* [Unpublished Doctoral dissertation]. University of Illinois at Urbana-Champaign <https://hdl.handle.net/2142/82216>
-