أثر شكل توزيع القدرة على ملاءمة المفردات ودقة تقدير معلم الصعوبة في نموذج راش

د. سوميه شكري محمد محمود
قسم علم النفس التربوي
كلية التربية بجامعة المنيا وجامعة الإمام عبد الرحمن بن فيصل

• **المستخلص:**

هدفت الدراسة إلى تحديد أثر شكل توزيع القدرة على ملاءمة المفردات ودقة تقدير مَعلم الصعوبة في نموذج راش لتدريج المفردات ثنائية الاستجابة، ولتحقيق الهدف من الدراسة تم استخدام برنامج (WinGen3) لتوليد القيم الحقيقية لمعلم صعوبة اختبار مكون من ٦٠ مفردة ثنائية الاستجابة مدرجة باستخدام نموذج راش، بحيث تتوزع قيم معلم الصعوبة اعتداليا بمتوسط (صفر) وانحراف معياري (واحد)، كما تم توليد عشرين مجموعة من البيانات تمثل كل منها قيم الدرجات الحقيقية لقدرة أحادية البعد لعينة من المتحنين، بحيث توزعت خمس منها طبقاً للتوزيع الاعتدالي ($\mu=0, \sigma=1$)، وتوزعت الخمس الثانية طبقاً لتوزيع بيتا ($\alpha=2, \beta=5$)؛ لتمثل توزيعاً موجب الالتواء، وتوزعت الخمس الثالثة طبقاً لتوزيع بيتا ($\alpha=5, \beta=2$)؛ لتمثل توزيعاً سالب الالتواء، وتوزعت الخمس الأخيرة طبقاً للتوزيع المنتظم ($a=-3, b=3$)، وذلك بأحجام مختلفة (١٠٠، ٢٠٠، ٣٠٠، ٤٠٠، ٥٠٠) لكل توزيع، ثم تم توليد بيانات استجابة كل عينة من العينات العشرين على الاختبار، مما أنتج عشرين فئة من البيانات، تمثل كل منها استجابات عينة مختلفة التوزيع أو الحجم للاختبار، وتم تحليل بيانات الاستجابات باستخدام برنامج R(3.3) بعد ربطه ببرنامج (SPSS)، وكشفت النتائج عن وجود تباين في ملاءمة المفردات لنموذج راش تبعاً لتباين شكل توزيع القدرة وحجم العينة، وتم استخدام مؤشرَين لتقييم دقة تقدير معلم الصعوبة، الأول هو الخطأ المعياري لتقدير معلم الصعوبة، والثاني هو معامل ارتباط سبيرمان للرتب بين قيم معلم الصعوبة الناتجة عن تحليل بيانات كل مجموعة من الاستجابات والقيم الحقيقية للمعلم، وتم استخدام تحليل التباين الأحادي للقياسات المتكررة؛ للكشف عن دلالة الفروق بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة في كل مجموعة، وأظهرت النتائج وجود فروق دالة إحصائيا ($p<0.001$) بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة في كل مجموعة ترجع لتوزيع القدرة، وكان التوزيع المنتظم هو التوزيع الذي ينتج أدنى مستوى من الخطأ، يليه التوزيع الاعتدالي، وكان المتوسط الأعلى للخطأ في حالة التوزيعات الملتوية، وبيّنت نتائج المؤشر الثاني أن جميع معاملات الارتباط دالة إحصائيا، وتراوحت بين (0.949) و (0.993).

الكلمات المفتاحية: المحاكاة ، نموذج راش ، شكل توزيع القدرة ، دقة التقدير.

The Effect of the Examinee's Ability Distribution Shape on the item fit and the Estimation Accuracy of the Difficulty Parameter in Rasch Model

Dr .Somaya Shokry Muhammad Mahmoud

Abstract

This study aimed at investigating The Effect of the Examinee's Ability Distribution Shape on the item fit and the Estimation Accuracy of the Difficulty Parameter in Rasch Model. To achieve the objective of the study, Wingen3 program was used to generate true item parameters of 60

dichotomous items calibrating on Rasch scale. These parameters were normally distributed (μ=0, σ=1). Twenty groups of examinee's data were generated each group of them represents the real scores of one-dimensional ability of a sample of a group examinees. The first five was normally distributed (μ=0, σ =1), the second five was distributed according to Beta distribution (α =2 , β=5) to present a positive skewness distribution, the third five was distributed according to Beta distribution (α =5 , β=2) to present a negative skewness distribution. The last five was distributed according to the uniform distribution (a= -3, b= 3) with different sample size to each distribution (100,200,300,400,500).The response data of each group to the 60 items were generated, then the response data were analysed using R(3.3) attached to SPSS Program. Results revealed significant differences *in the fit of test items and the estimation accuracy due to the distribution shape of the examinees ability. The most accurate estimation was achieved when the examinees' ability were distributed according to the uniform distribution, while the positive skewness distribution and the negative skewness distribution resulted the least accurate estimation.*

Key words: Simulation, Rasch model, Distribution shape, Estimation accuracy

- **مقدمة**:

يُعد القياس أساس البحث العلمي بشكل عام والبحث النفسي والتربوي بشكل خاص؛ حيث تتوقف دقة نتائج البحث على دقة أدوات القياس المستخدمة، ويمثل الاعتماد على نظرية الاستجابة للمفردة (النظرية الحديثة للقياس النفسي) وسيلة جيدة لتطوير البحث النفسي والتربوي؛ لما تقدمه من مميزاتٍ؛ حيث تعتبر استجابة كل ممتحن لكل مفردةٍ في الاختبار موقفاً متفرداً بذاته، وبذلك تختلف دقة هذه الاستجابة تبعاً لمدى ملاءمة المفردة لمستوى قدرة الممتحن (Xing & Hambleton, 2004)، وتفترض هذه النظرية مجموعة من النماذج توضح العلاقة بين احتمال استجابة محددة بالنسبة لكل فرد على كل مفردة، وليس للاختبار ككل، مما يتيح المرونة عند تعديل الاختبار (Sijtsma & Molenaar, 2002, Linden, 2005)، كما أن تقدير معالم المفردات يكون مستقلاً عن العينة (Alderson, et al., 1995)، وهي بذلك تعالج أوجه القصور في النظرية التقليدية، كعدم خطية الدرجات، وافتراض تساوي تباين خطأ القياس لجميع الممتحنين، واعتماد الخصائص السيكومترية للاختبارات على خصائص عينة الممتحنين (He & Tymms, 2005).

وعلى الرغم من تلكٍ المميزات التي تقدمها النظرية الحديثة في القياس النفسي فإن هناك عزوفاً عن تطبيقها في البحوث النفسية والتربوية، ويرجع ذلك لعدة عوامل، أهمها: الافتراضات الصارمة التي تقوم عليها النظرية، وعدم ألفة الباحثين للبرامج الإحصائية المطلوبة، بالإضافة إلى ضخامة حجم العينة

اللازم للتدريج مقارنة بحجم العينة اللازم لتقدير ثبات أدوات القياس في النظرية التقليدية (الشمراني، ٢٠١٦ ، القيسي، ٢٠١٦)

فقد اعتمدت الدراسات السابقة كلها على عينات تتوزع القدرة فيها اعتدالياً، وأكدت على أن دقة التقدير تزيد بزيادة حجم العينة. ويمكن تفسير ذلك بأن الحاجة إلى زيادة حجم العينة ترجع إلى ضمان توفر عدد كاف من الأفراد متطرفي القدرة، مما يثير التساؤل حول تأثير زيادة عدد الأفراد متطرفي القدرة دون زيادة الحجم الكلي للعينة . أي أن تتوزع القدرة طبقاً للتوزيع المنتظم . على دقة التقدير. لذلك هدفت الدراسة الحالية إلى تقصي أثر شكل توزيع قدرة المتحنين على ملاءمة المفردات ودقة تقدير مَعلم الصعوبة في نموذج راش لتدريج المفردات ثنائية الاستجابة؛ لأنه النموذج الأكثر استخداما، وهو النموذج الوحيد في نظرية الاستجابة للمفردة الذي يحقق خاصية الموضوعية النسبية (عدم التغاير Invariance)، كما أن المفردات ثنائية الاستجابة هي النوع الأكثر استخداماً في الاختبارات التحصيلية واختبارات القدرات.

وبذلك تقدم الدراسة الحالية معلومات عن شكل توزيع القدرة الذي يؤدي للحصول على أدق تقدير بأقل حجم للعينة. وقد تم استخدام برنامج (R3.3) بعد ربطه ببرنامج (SPSS)؛ لإتاحة إجراء التحليل الإحصائي من خلال نوافذ سهلة أعتاد عليها الباحثون، وتم الاعتماد على بيانات محاكاة تم توليدها باستخدام برنامج (WinGen3)؛ حيث أشار (Han & Hambleton, 2014) إلى أهمية دراسات المحاكاة في: تقييم تأثير انتهاك الافتراضات في نظرية الاستجابة للمفردة، ودراسة تأثير اختيار النماذج، وفحص أحجام العينات، وتوزيعات القدرة، وطول الاختبار. كما أن أسلوب المحاكاة يتيح الدراسة المتعمقة لتأثير هذه العوامل في مواقف متعددة مع التحكم في حجم العينة وشكل التوزيع وطول الاختبار، وليس بناءً على موقف تطبيقي واحد (الدرابيع، ٢٠٠١).

- **مشكلة الدراسة**:

على الرغم من مميزات نظرية الاستجابة للمفردة مقارنة بالنظرية التقليدية في القياس، وعلى الرغم من تفرد نموذج راش -مقارنة بالنماذج الأخرى لنظرية الاستجابة للمفردة -بخاصية الموضوعية النسبية، فإن النجاح في الاستفادة من هذه المميزات يعتمد على دقة تقدير معلمي النموذج. وقد تناولت الدراسات في هذا المجال تأثير عدة عوامل في دقة التقدير، كالطريقة الرياضية، والبرنامج الإحصائي المستخدم، وحجم العينة، وطول الاختبار، لكن هناك ندرة في الدراسات التي اهتمت بتأثير شكل توزيع القدرة في نتائج تدريج الاختبار، فمعظم الباحثين يعتمدون على عينات كبيرة تم اختيارها بطريقة عشوائية لتحقق شرط الاعتدالية، متأثرين في ذلك بالنظرية التقليدية.

ومن ناحية أخرى فإن هناك عزوفاً من قِبل الباحثين في التربية وعلم النفس عن تبني نظرية الاستجابة للمفردة عند إجراء البحوث الوصفية أو التجريبية، وقد يرجع ذلك لكبر حجم العينة المطلوبة للتدريج باستخدام نماذج النظرية مقارنة بحجم العينة المطلوب عند استخدام النظرية التقليدية، أو لتوقع عدم ملاءمة المفردات للنموذج، فبينما يتطلب تدريج الاختبارات باستخدام نموذج راش عينة لا يقل حجمها عن ١٠٠ ممتحن، فإن حساب ثبات الاختبار في ضوء النظرية التقليدية يتطلب عينة عشوائية يزيد حجمها عن ٣٠ ممتحناً، مما يمثل تحدياً للباحثين الذين يهدفون إلى تدريج الاختبارات لأغراض تطبيقية.

وقد اهتم عدد من الدراسات السابقة بتأثير كل من حجم العينة وطول الاختبار في جودة ملاءمة المفردات للنموذج ودقة التقدير، وأكدت تلك الدراسات زيادة دقة التقدير بزيادة حجم العينة وطول الاختبار، لذا بات من المفيد تحديد شكل التوزيع الذي يؤدي إلى أعلى مستوى من الدقة بأقل حجم من الممتحنين. كما أن هناك تساؤلاً بديهياً حول تأثير الالتواء في توزيع عينة التدريج، الذي يظهر في العديد من المواقف العملية، حيث تمثل قدرات الأفراد توزيعاً سالب الالتواء عندما يمتلك معظم العينة مستوى من القدرة أعلى من المتوسط، كما هو الحال في مدارس الفائقين، في حين تمثل قدرات الأفراد توزيعاً موجب الالتواء عندما يمتلك معظم أفراد العينة مستوى من القدرة أقل من المتوسط.

ويمكن تلخيص مشكلة الدراسة في التساؤلات التالية:

- هل تختلف ملاءمة المفردات لنموذج راش تبعاً لشكل توزيع القدرة أو حجم العينة ؟.
- هل يختلف الخطأ المعياري لتقدير معلم صعوبة المفردة باختلاف شكل توزيع القدرة عند تساوي أحجام العينات ؟.
- ما هو شكل توزيع القدرة الذي يؤدي إلى أقل خطأ معياري في تقدير معلم الصعوبة بأقل حجم من الممتحنين ؟.
- هل يختلف مدى تحقق خاصية الموضوعية النسبية لمعلم الصعوبة باختلاف شكل التوزيع أو حجم العينة ؟.

- **أهداف البحث**:

- مقارنة نتائج ملاءمة المفردات لنموذج راش للإستجابات الثنائية عندما تتوزع القدرة اعتداليا وعندما يكون توزيعها ملتويا أو منتظماً عند عدة مستويات لحجم العينة.
- مقارنة متوسطات الخطأ المعياري لتقدير معلم الصعوبة عند اختلاف شكل توزيع القدرة وتساوي حجم العينة.

- مقارنة متوسطات الخطأ المعياري لتقدير معلم الصعوبة عند اختلاف شكل توزيع القدرة أو حجم العينة.
- مقارنة معاملات الارتباط بين القيم المقدَّرة والقيم الحقيقة لمعلم الصعوبة عند اختلاف شكل توزيع القدرة أو حجم العينة.

- **أهمية الدراسة**:

- **الأهمية النظرية**:

تقدم الدراسة أدلة واضحة حول تأثير شكل توزيع القدرة لعينة المتحنين في دقة تقدير معلم الصعوبة لنموذج راش، وذلك اعتماداً على تصميم تجريبي وبيانات تم توليدها باستخدام أسلوب المحاكاة.

- **الأهمية التطبيقية**:

تقدم الدراسة معلومات وإرشادات حول الطريقة المناسبة لاختيار عينة التدريج بما يحقق الوصول لشكل توزيع القدرة الذي يؤدي إلى أعلى دقة في التقدير مع أصغر حجم ممكن للعينة.

حدود الدراسة:

تتحدد الدراسة الحالية بكل من:

- البرنامج المستخدم في توليد البيانات (WinGen3).
- البرنامج المستخدم في تحليل البيانات (R3.3).
- أشكال توزيعات القدرة المستخدمة. وهي: التوزيع الاعتدالي ($\mu=0,\ \sigma=1$)، وتوزيع بيتا ($\alpha=2,\ \beta=5$)، وتوزيع بيتا ($\alpha=5,\ \beta=2$)، والتوزيع المنتظم ($a=-3,\ b=3$).

- **مصطلحات الدراسة**:

- شكل توزيع القدرة: هو منحنى تكراري يوضح العلاقة بين مستوى القدرة وتكرارات الأفراد الذين يمتلكونها في العينة.
- نموذج راش للمفردات ثنائية الاستجابة: هو أحد نماذج نظرية الاستجابة للمفردة ثنائية الاستجابة، الذي يحدد العلاقة بين احتمال نجاح المتحن في الإجابة عن المفردة وكل من مستوى قدرته وصعوبة المفردة، ويفترض أن جميع مفردات الاختبار متساوية التمييز، ومعلم التخمين يساوى صفر.
- معلم الصعوبة: هو نقطة على متصل القدرة توضح موقع المنحنى المميز للمفردة، وتقابل احتمال (0.50) للإجابة عن المفردة إجابة صحيحة عندما يكون معلم التخمين مساوياً للصفر، وبمعنى آخر فإن الأفراد الذين يقعون على هذه النقطة من متصل القدرة يستطيع نصفهم الإجابة عن المفردة إجابة صحيحة دون تخمين، في حين يفشل النصف الآخر في الإجابة عن المفردة.

- دقة تقدير معلم الصعوبة: هو مدى اقتراب القيمة المقدَّرة لمعلم الصعوبة من البيانات الملاحظة من القيمة الحقيقية للمعلم كما يفترضها النموذج.
- الخطأ المعياري في التقدير: هو مؤشر لدقة التقدير، فكلما قل الخطأ المعياري زادت الدقة، ويُمثَّل بالجذر التربيعي لمتوسطات مربعات الانحرافات بين القيم المقدَّرة من البيانات الملاحظة والقيم التي يفترضها النموذج.

- **الإطار النظري والدراسات السابقة**:

تتوقف الاستفادة من مميزات نظرية الاستجابة للمفردة على عاملين اساسيين تتمثل في ملاءمة البيانات للنموذج ودقة تقدير معالم المفردات؛ حيث إن دوال معلومات المفردات تكون مفيدة فقط عندما تكون المفردات ملائمة للنموذج، وفي حالة عدم الملاءمة يؤدي الاعتماد على دوال المعلومات إلى نتائج مضللة، كما أن حساب هذه الدوال يعتمد على قيم معالم المفردات، فكلما كان تقدير المعالم دقيقاً زادت دقة دوال المعلومات الناتجة (Hambleton, et al., 1991).

وتمثل دوال معلومات المفردات في نظرية الاستجابة للمفردة مؤشراً لدقة القياس، فهي تعد المقابل لمفهوم الثبات في النظرية التقليدية، حيث تتناول كل من النظرية التقليدية ونظرية الاستجابة للمفردة أخطاء القياس بطريقة مختلفة، فتفترض الأولى أن الأخطاء تتوزع عشوائياً، وأن خطأ القياس ثابت بالنسبة لكل الدرجات، ويعتمد تقدير كلٍ من الدرجة الحقيقية للفرد وفترة الثقة على خواص عينة المتحنين؛ نظراً لاعتماد تقدير ثبات الاختبار على تباين الدرجة الملاحظة، أما في نظرية الاستجابة للمفردة فيتم تقدير خطأ القياس لكل درجة على حدة، أي لكل مستوى من مستويات القدرة (Weiss & Yoes, 1991, Thissen, 2000).

ويرتبط الثبات في النظرية التقليدية بالاختبار ككل، فلا يمكن تجزئته، في حين أن هذه التجزئة ممكنة في نظرية الاستجابة للمفردة، إذ يتم تقييم كل مفردة على حدة في ضوء كمية المعلومات التي تقدمها عند مستوى محدد من القدرة من خلال دالة معلومات المفردة، فيمكن الحصول بسهولة على دالة المعلومات لأي اختبار مكون من أية مجموعة جزئية من هذه المفردات (Linden, 2005, Shultz & Whitney, 2005).

ويعبَّر عن الخطأ المعياري للقياس في النظرية التقليدية Standard Error of Measurement (SE) من خلال المعادلة التالية:

$$SE = \sigma\sqrt{(1-p)}$$

حيث σ هي الانحراف المعياري للدرجات الملاحظة، و p هو معامل ثبات الاختبار (Lord, 1980). ويلاحظ من هذه المعادلة أن الخطأ المعياري للقياس في النظرية التقليدية يقل بزيادة ثبات الاختبار.

بينما يعبَّر عن الخطأ المعياري لتقدير القدرة Standard Error of Estimation (SEE) في نظرية الاستجابة للمفردة بالمعادلة التالية:

$$SEE = 1/\sqrt{I(\theta)}\,,$$

$$I(\theta) = \sum_{i=1}^{n} Ii(\theta) = \frac{2.89ai^2(1-ci)}{[ci + e^{1.7(\theta-bi)}][1 + e^{-1.7(\theta-bi)}]^2}$$

حيث Ii(θ) هي دالة معلومات المفردة i ، و$I(\theta)$ هي دالة معلومات الاختبار، و ai قيمة معلم التمييز للمفردة i ، و ci قيمة معلم التخمين للمفردة i ، و bi قيمة معلم الصعوبة للمفردة i . ويتضح من خلال المعادلة أن دالة المعلومات تزيد كلما زاد معلم التمييز، وكلما اقترب معلم الصعوبة bi من مستوى القدرة θ ، وكلما اقترب معلم التخمين ci من الصفر (Hambleton, et al., 1991). كما أنها تتأثر بالخطأ المعياري لتقدير معالم المفردات، حيث إنها تعتمد على قيمة المعالم، فكلما كانت قيمة المعالم دقيقة وكان الخطأ المعياري في تقديرها صغيراً كانت دالة المعلومات أكثر دقة.

وتعبر دالة معلومات المفردة $Ii(\theta)$ عن كمية المعلومات التي تقدمها المفردة عند مستوى محدد من القدرة θ ، ومن المنطقي أن تتباين هذه المعلومات تبعاً لمستوى القدرة، فعلى سبيل المثال نجد أن المفردة متوسطة الصعوبة تقدم الكثير من المعلومات إذا استخدمت لقياس مجموعة من الطلاب متوسطي القدرة، في حين أنها لا تقدم أية معلومة إذا اختبرنا بها مجموعة من الفائقين؛ لأنهم سينجحون جميعاً في الإجابة عنها، ولن تقدم أية معلومات عن الطلاب منخفضي القدرة؛ لأنهم سيفشلون جميعاً في الإجابة عنها.

كما تستخدم دالة معلومات الاختبار في المفاضلة بين الاختبارات في المواقف المختلفة عند مستوى محدد من القدرة من خلال محك الكفاءة النسبية للاختبار Relative Efficiency (RE)، فالكفاءة النسبية لاختبار ما y بالنسبة لاختبار آخر x هي النسبة بين دالة معلومات الاختبار الأول عند مستوى محدد من القدرة ودالة معلومات الاختبار الثاني عند نفس المستوى، كما هو موضح بالمعادلة التالية :

$$RE = \frac{\{I(\theta, y)\}}{\{I(\theta, x)\}}$$

حيث $\{I(\theta, y)\}$ هي دالة معلومات الاختبار y عند المستوى θ للقدرة، و $\{I(\theta, x)\}$ هي دالة معلومات الاختبار x عند المستوى θ للقدرة (Hambleton, et al., 1996).

ويتضح مما سبق أهمية دالة معلومات المفردة في تقدير دقة القياس في النظرية الحديثة، والدور الذي تلعبه معالم المفردات في حساب هذه الدالة، لذلك اعتمدت الدراسة الحالية على الخطأ المعياري لتقدير معلم الصعوبة كمحك للدقة في نموذج راش؛ إذ أن الخطأ المعياري للتقدير هو معكوس الدقة، فكلما زاد الخطأ المعياري للتقدير قلت الدقة (القيسي، ٢٠١٣).

ومن الدراسات التي اعتمدت على الخطأ المعياري لتقدير المعالم كمحك لدقة التقدير دراسة عبابنة (٢٠٠٤)، التي هدفت إلى فحص أثر حجم العينة وطريقة انتقائها، وعدد الفقرات وطريقة انتقائها، على دقة تقدير معالم الفقرة والقدرة في النموذج ثلاثي المعلم. ومن أهم نتائجها زيادة دقة تقدير معالم الفقرة بزيادة حجم عينة المفحوصين، وعدم استقرار تقديرات معالم الفقرة عند استخدام عينات متغايرة القدرة، وزيادة الدقة في تقدير معلم الصعوبة عندما يلائم مدى قدرة المفحوصين.

وكذلك دراسة الثوابية (٢٠١٠) التي هدفت إلى استقصاء أثر حجم العينة في تقدير معلم صعوبة المفردة والخطأ المعياري في تقديرها باستخدام نموذج راش. وتوصلت الدراسة إلى أنه بزيادة حجم العينة يزداد متوسط معلم صعوبة الفقرة ويتناقص الخطأ المعياري في تقديره.

ودراسة بني عطا والشريفين (٢٠١٢) التي هدفت إلى الكشف عن أثر شكل توزيع القدرة (اعتدالي، ملتوٍ نحو اليمين، ملتوٍ نحو اليسار، ثنائي المنوال) على معالم المفردة ودالة المعلومات في النموذج ثلاثي المعلم، وتوصلت إلى وجود فروق دالة إحصائيا بين كل من متوسطات معالم الصعوبة ومتوسطات معالم التمييز ترجع لاختلاف شكل توزيع القدرة، وكذلك تباين شكل دالة معلومات الاختبار بتباين شكل توزيع القدرة، إلا أن قيم معاملات الارتباط بين معالم المفردات تبعا لشكل التوزيع كانت دالة إحصائيا، مما يدل على استقرار معالم الفقرات عبر التوزيعات المختلفة للقدرة.

ودراسة الشريفين (٢٠١٢) التي اهتمت بتأثير طريقة التقدير وحجم العينة في تقدير معالم النموذج، وتوصلت إلى وجود فروق دالة إحصائيا في متوسطات الأخطاء المعيارية لتقدير معالم المفردات في النموذج ثلاثي المعلم تُعزى إلى التفاعل بين طريقة تقدير معالم المفردة وحجم العينة، رغم عدم دلالة الفروق التي تُعزى لمتغير حجم العينة أو طريقة التقدير، كما كانت هناك فروق دالة

إحصائياً بين متوسطات الأخطاء المعيارية لتقدير معلم القدرة تُعزى لمتغير حجم العينة وللتفاعل بين طريقة التقدير وحجم العينة.

بالإضافة إلى دراسة الزبون (٢٠١٣)، التي هدفت إلى دراسة أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس. وتوصلت إلى أن مقدار تقديرات دالة المعلومات في النموذج ثلاثي المعلم يزداد بزيادة حجم العينة، وأن الخطأ المعياري في تقدير دالة المعلومات يتناقص بزيادة حجم العينة.

وكذلك اعتمدت دراسة شما (٢٠١٣) على قيم الخطأ المعياري كمؤشر على دقة التقدير، حيث هدفت إلى معرفة أثر حجم العينة على دقة تقدير صعوبة المفردات وقدرات الأفراد باستخدام نموذج راش لتدريج اختبار (أوتيس - لينون) للقدرة العقلية (المستوى المتوسط -الصورة j)، وتوصلت إلى أن دقة تقدير صعوبة المفردات تزيد بزيادة حجم العينة، بينما لا يؤثر حجم العينة على دقة تقدير معلم القدرة.

واتفقت مع جميع الدراسات السابقة دراسة الشمراني (٢٠١٦)، التي هدفت إلى الكشف عن أثر كل من حجم العينة وطول الاختبار والتفاعل بينهما في تقدير معالم النموذج ثلاثي المعلم والخطأ المعياري للتقدير، مع جميع الدراسات السابقة ، حيث توصلت إلى وجود فروق دالة إحصائياً في تقدير معلم الصعوبة تُعزى إلى حجم العينة، كما أن الخطأ المعياري يقل كلما زاد حجم العينة.

وقد استخدمت بعض الدراسات مؤشر التحيز (Bias) أو مؤشر الجذر التربيعي لمتوسط مربع الخطأ (RMSA) للتعبير عن دقة القياس في نظرية الاستجابة للمفردة، ويُعرف التحيز على أنه الخطأ المنتظم في التقدير، ويعبر عن مدى استقرار تقدير المعلم عبر مجموعة من العينات، حيث يمثل متوسط انحراف القيم المقدرة للمعلم عن القيم الحقيقية، ويُحسب من خلال المعادلة التالية :

$$bias = \frac{\sum_{j=1}^{n}(\ true\ parameter(j) - estimated\ parameter(j))}{n}$$

بينما يُستخدم مؤشر الجذر التربيعي لمتوسط مربع الخطأ (RMSA) للتعبير عن الدقة المطلقة لتقدير المعلم، ويمثل الجذر التربيعي لمتوسط مربعات الفروق بين القيم الحقيقية والقيم المقدرة للمعلم، ويُحسب من المعادلة التالية:

$$RMSE = \sqrt[2]{\frac{\sum_{j=1}^{n}[true\ parameter(j) - estimated\ parameter(j)]^2}{n}}$$

حيث n مرات تكرار إعادة تقدير المعلم (Khan, 2014).

ومن هذه الدراسات دراسة الدرابيع (٢٠٠١)، حيث هدفت إلى التحقق من فعالية نموذج راش في دقة تقدير قدرات الأفراد ومعلم صعوبة المفردة عند استخدام عينات مختلفة الحجم (٥٠، ١٠٠، ٥٠٠) فرد، واختبارات متباينة الطول (٢٥، ٥٠، ٣٠٠)، حيث تم توليد بيانات ثنائية الاستجابة باستخدام برنامج (IRTDATA)، وتم استخدام برنامج (BILOG3.0) لتقدير قدرة الأفراد ومعلم صعوبة المفردات، وتوصلت الدراسة إلى وجود فروق دالة إحصائيا لكل من متغير طول الاختبار ومتغير التفاعل بين طول الاختبار وحجم العينة على دقة تقدير قدرة الأفراد، ووجود فروق دالة إحصائيا في دقة تقدير معلم صعوبة المفردات لكل من متغير طول الاختبار وحجم العينة ومتغير التفاعل بينهما.

وكذلك اعتمدت كل من دراسة القيسي (٢٠١٣) ودرويش (٢٠١٤) على محك التحيز (Bias) ومحك الجذر التربيعي لمتوسط مربعات الخطأ (RMSE) للحكم على دقة التقدير في النماذج المعلمية واللا معلمية لنظرية الاستجابة للمفردة.

ودراسة Khan(2014) التي هدفت إلى الكشف عن تأثير حجم العينة وطول الاختبار على ملاءمة المفردات واستقرار تقدير معلم الصعوبة، وتم استخدام مؤشر التحيز (Bias) ومتوسط الخطأ المعياري (RMSE)، وتوصلت إلى اختلاف كل من ملاءمة المفردات وتقديرات معلم الصعوبة باختلاف حجم العينة وطول الاختبار، كما بينت هذه الدراسة أنه على الرغم من إمكانية تحقق ملاءمة النموذج في حالة استخدام عينات صغيرة وعدد قليل من المفردات فإن ذلك لايضمن بالضرورة تحقق ثبات أو دقة تقدير المعالم.

ودراسة Karadavut (2017) التي هدفت إلى مقارنة دقة تقدير المعالم في كل من نموذج راش والنموذج ثنائي المعلم والنموذج ثلاثي المعلم عندما تتوزع القدرة اعتداليا وطبقا للتوزيع المنتظم، وتوصلت إلى أن الاعتماد على التوزيع المنتظم يؤدي إلى نتائج أكثر دقة مقارنة بالتوزيع الاعتدالي في النموذجين ثنائي المعلم وثلاثي المعلم، بينما لا توجد فروق بينهما عند استخدام نموذج راش.

ورغم أهمية محكي التحيز والجذر التربيعي لمتوسط مربعات الخطأ، فإنهما لا يناسبان تقدير الدقة في الدراسة الحالية، إذ يعتمدان على مقارنة قيم معلم الصعوبة المقدرة في العينات بنظائرها الحقيقية باعتبارها قيم مطلقة وثابتة، ويكون هذان المؤشران مناسبَين في حالة إعادة تطبيق الاختبار على العينة نفسها

عدة مرات، أو عند تكافؤ العينات المختلفة في مستوى القدرة واختلافها في أحد المتغيرات الديموغرافية، كالنوع أو الخلفية الثقافية، وذلك بهدف كشف تحيز الاختبار.

فلا تدَّعي نظرية الاستجابة للمفردة أن قيم معالم المفردات مطلقة وثابتة عبر العينات مختلفة القدرة، بل إن ما قدمته النظرية هو مفهوم الموضوعية النسبية أو عدم تغاير معالم القدرة والصعوبة Invariance، الذي ينتج عن دمج المعلومات المتعلقة بمعالم المفردة أثناء تقدير قدرات الممتحنين، ودمج المعلومات المتعلقة بقدرات الممتحنين أثناء تقدير معالم المفردة (Hambleton, et al., 1991)، مما يتيح ترتيب صعوبة المفردات موضوعياً (MacNamara, 1996)، وهذا ما جعل نموذج راش أكثر النماذج موضوعية؛ لأنه يحتوي على معلم واحد للمفردة، هو معلم الصعوبة.

ويُعد نموذج راش أكثر نماذج نظرية الاستجابة للمفردة استعمالاً، ويفترض أن المفردات تختلف من حيث معلم الصعوبة فحسب، وأنها متساوية التمييز، حيث يساوي معلم التمييز (١) لجميع المفردات، بينما ينعدم معلم التخمين. وتُعطى المنحنيات المميزة للمفردات في نموذج راش بواسطة المعادلة التالية :

$$Pi(\theta) = \frac{e^{(\theta - bi)}}{1+e^{(\theta - bi)}}, \; i = 1,2,...,n \qquad (1)$$

حيث $Pi(\theta)$ احتمال نجاح الفرد الذي قدرته θ في الإجابة عن المفردة التي معامل صعوبتها bi صوابا (Wright, 1977, Hambleton et al., 1991)، وتتحقق خاصية الموضوعية النسبية لنموذج راش رياضياً كما يلي:

حيث إن مجموع احتمالَي النجاح في الإجابة عن المفردة والفشل في الإجابة عن المفردة يساوي ١، فإن احتمال الفشل في الإجابة عن المفردة يُعطى من العلاقة التالية:

$$Pi(\theta) = \frac{1}{1+e^{(\theta - bi)}} \qquad (2)$$

وبقسمة المعادلة (1) على المعادلة (2)

$$\frac{Pi(\theta)}{1 - Pi(\theta)} = e^{(\theta - bi)}$$

وبأخذ لوغاريتم الطرفين

$$(\theta - bi) = \ln\left(\frac{Pi(\theta)}{1 - Pi(\theta)}\right) \qquad (3)$$

وبتكرار نفس الخطوات بالنسبة لمفردة أخرى معامل صعوبتها bj

$$(\theta - bj) = \ln\left(\frac{Pj(\theta)}{1-Pj(\theta)}\right) \qquad (4)$$

وبطرح المعادلة (٣) من المعادلة (٤)

$$bi - bj = \ln\left(\frac{Pj(\theta)}{1-Pj(\theta)}\right) - \ln\left(\frac{Pi(\theta)}{1-Pj(\theta)}\right) \qquad (5)$$

وما تعنيه المعادلة (٥) أن الفرق بين معلم صعوبة المفردة i ومعلم صعوبة المفردة j – أي المسافة بين موقع المفردة i وموقع المفردة j على متصل القدرة – لا يعتمد على مستوى قدرة عينة التدريج؛ فهذا الفرق يمكن تقديره عند أية قيمة لمتغير القدرة θ ، ولا يعني ذلك أن قيمة معلم صعوبة المفردات تكون مطلقة وثابتة عند استخدام عينات مختلفة القدرة، إنما يعني أن المسافة بين المفردات تكون ثابتة، ولذا فإن المفردات بعد تحليلها باستخدام نموذج راش تسمى مفردات مدرَّجة أي أنها مرتَّبة طبقا لمعلم الصعوبة، وأنها ستحتفظ بهذا الترتيب مهما اختلف مستوى قدرة العينة، وهذا ما فرض ضرورة وجود رابطة من المفردات أو رابطة من الأفراد عند دمج الاختبارات المختلفة في تدريج واحد.

وهكذا تسمح نماذج نظرية الاستجابة للمفردة بإجراء ما يُعرف بالقياس المحرَّر من خصائص الاختبار (Wright & Stone, 1979)، مما يعني إمكانية مقارنة قدرات الأفراد، حتى عندما يأخذون اختبارات مختلفة، بشرط أن تنتمي مفرداتها إلى التدريج نفسه (MacNamara, 1996)، كما تسمح النماذج نفسها بإجراء تدريج المفردة متحررا من خصائص العينة، مما يعني إمكانية مقارنة مجموعات مختلفة من المفردات مباشرة وربطها معا، بشرط وجود رابطة من عينة المتحنين (Wright & Stone, 1979). ومن ثم اعتمدت الدراسة الحالية على معامل ارتباط سبيرمان للرتب بين معالم الصعوبة الناتجة من استخدام عينات مختلفة للتدريج كمحك إضافي لتقدير الدقة.

- **منهج الدراسة**:

اعتمدت الدراسة على المنهج التجريبي باستخدام تصميم القياسات المتكررة، حيث تم قياس الخطأ المعياري لتقدير معلم الصعوبة عدة مرات لعينة واحدة من المفردات عند استخدام عينات متساوية الحجم من المتحنين ومتباينة فقط في شكل توزيع القدرة (اعتدالي، موجب الالتواء، سالب الالتواء، منتظم)، وتم تكرار ذلك عند خمسة مستويات لحجم العينة، هي: (١٠٠، ٢٠٠، ٣٠٠، ٤٠٠، ٥٠٠)

ممتحنٍ، وبذلك تكرر قياس دقة التقدير (٤ × ٥) مرة، مما أنتج ٢٠ قياساً مرتبطا.

- **إجراءات الدراسة**:

- **أولا- توليد البيانات**:

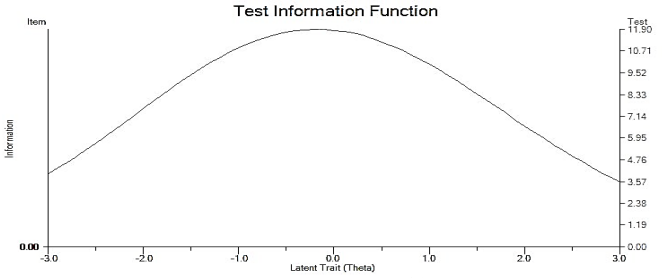
تم توليد البيانات باستخدام برنامج (WinGen3)؛ حيث يتمتع البرنامج بالعديد من الخصائص المفيدة، فهو يدعم النماذج المختلفة لنظرية الاستجابة للمفردة، كما أنه يمكن من توليد مجموعات من معالم المفردات ومجموعات من معالم القدرة بأنواع مختلفة من التوزيعات بغرض توليد بيانات استجابات واقعية للمفردات (Han & Hambleton, 2014)، وقد تمت عملية توليد البيانات في ثلاث مراحل، كالتالي:

- **مرحلة توليد المعالم الحقيقية لمفردات الاختبار**:

تم توليد المعالم الحقيقية لـ ٦٠ مفردة ثنائية الاستجابة مدرجة في ضوء نموذج راش، وتم اختيار هذا الطول لمناسبته للعديد من المواقف العملية، كما أشارت العديد من الدراسات إلى أن أقصى دقة لتقدير معالم المفردات يتحقق عند هذا الطول (الشمراني، ٢٠١٦)، وتم اختيار المفردة ثنائية الاستجابة لأنها الأكثر استخداما في الاختبارات التحصيلية واختبارات القدرات العقلية، وتم اختيار نموذج راش لما يتمتع به من خاصية فريدة هي خاصية الموضوعية النسبية، وتم اختيار التوزيع الاعتدالي بمتوسط صفر وانحراف معياري واحد لتوزيع قيم معالم الصعوبة المولدة، حتى تمتد صعوبة المفردات من -٣ إلى +٣، ويقع معظم المفردات قريبا من المتوسط؛ لأن ذلك التوزيع هو التوزيع المثالي للمفردات الذي يستهدف مطورو الاختبارات الوصول إليه ليلائم قياس المتغيرات النفسية التي تتوزع توزيعا اعتداليا. ويوضح الجدول (١) القيم الحقيقية لمعلم الصعوبة للمفردات المولدة. ويوضح الشكل (١) دالة معلومات الاختبار.

جدول (١) القيم الحقيقية لمعلم صعوبة المفردات المولدة

المفردة	معلم الصعوبة	المفردة	معلم الصعوبة	المفردة	معلم الصعوبة	المفردة	معلم الصعوبة	المفردة	معلم الصعوبة	المفردة	معلم الصعوبة
1	.515	11	2.041	21	.358	31	-1.034	41	.175	51	.227
2	-.639	12	1.156	22	1.813	32	.446	42	-1.313	52	-1.562
3	.031	13	1.076	23	-1.289	33	.311	43	-1.082	53	2.073
4	-.786	14	.981	24	-.627	34	-1.680	44	-.947	54	-.464
5	1.250	15	-1.828	25	-.013	35	-.609	45	-.013	55	.233
6	.281	16	.399	26	-.916	36	2.872	46	-.836	56	-.501
7	-1.294	17	.926	27	-1.153	37	.128	47	.600	57	-1.764
8	.151	18	1.711	28	-.476	38	.208	48	1.201	58	1.307
9	-1.658	19	-.820	29	-1.078	39	-1.181	49	.876	59	.691
10	-1.681	20	-.924	30	.672	40	-.713	50	-1.145	60	.628



شكل (١) دالة معلومات الاختبار

- **مرحلة توليد الدرجات الحقيقية للقدرة**:

تم توليد ٢٠ مجموعة من البيانات، تمثل كل منها الدرجات الحقيقية لمعلم قدرة أحادية البعد لعينات مختلفة التوزيع بخمس مستويات من حجم العينة، هي: (١٠٠، ٢٠٠، ٣٠٠، ٤٠٠، ٥٠٠)، وكانت التوزيعات كالتالي:

- التوزيع الاعتدالي، بمتوسط صفر، وإنحراف معياري ١ .
- توزيع بيتا (α=2, β=5)؛ ليمثل توزيعاً موجب الالتواء.
- توزيع بيتا (α=5, β=2)؛ ليمثل توزيعاً سالب الالتواء.
- التوزيع المنتظم، بمتوسط صفر، وانحراف معياري ١ .

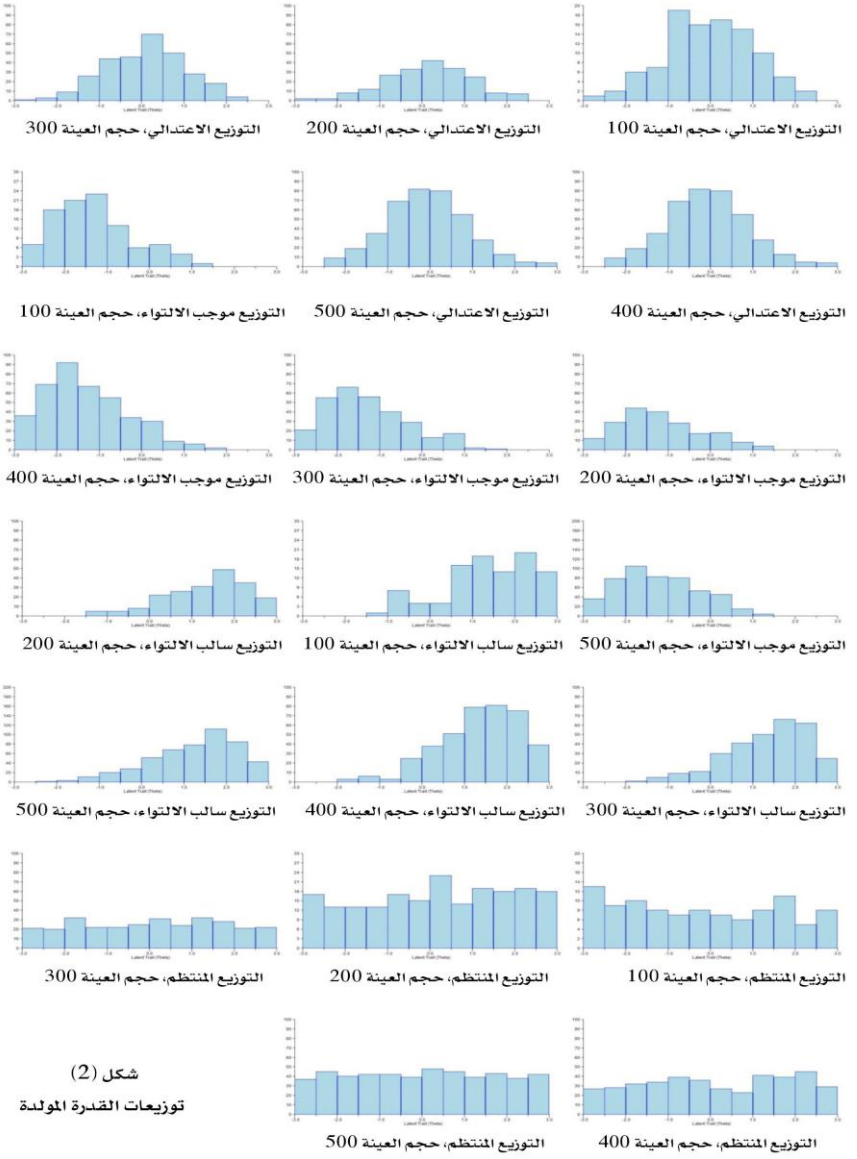
ويوضح الشكل (٢) توزيعات القدرة المولدة.

- **مرحلة توليد بيانات الاستجابة**:

في هذه المرحلة تم فتح ملف الدرجات الحقيقية لمعلم القدرة لعينة الممتحنين، وفتح ملف القيم الحقيقية لمعلم صعوبة مفردات الاختبار، بهدف توليد بيانات استجابة هذه العينة على الاختبار ثم حفظها، وتكررت هذه الخطوة ٢٠ مرة، طبقاً لتوزيعات القدرة وأحجامها التي تم توليدها في المرحلة السابقة.

- **ثانياً - تحليل بيانات الاستجابة في ضوء نموذج راش**:

تم التحقق من افتراض أحادية البعد في البيانات باستخدام التحليل العاملي الاستكشافي لكل مجموعة من الاستجابات، وأكدت النتائج وجود عامل عام في كل مجموعة، حيث كانت نسب التباين المفسر من العامل الأول أكبر من ٢٠ ٪ من التباين الكلي، وكان ناتج قسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني أكبر من ٢، كما هم موضح في الجدول (٢)، ثم تم استخدام برنامج (R3.3) بعد ربطه ببرنامج (SPSS)؛ لإجراء تحليل البيانات في ضوء نموذج راش؛ حيث يتيح ذلك الإجراء تدريج المفردات باستخدام نماذج الاستجابة للمفردة من خلال نافذة سهلة اعتاد الباحثون على التعامل معها. وبذلك نتجت ٢٠ مجموعة من نتائج التحليل لنفس الاختبار المكون من ٦٠ مفردة .



شكل (2)

توزيعات القدرة المولدة

جدول (٢) نتائج التحقق من أحادية البعد في البيانات

شكل التوزيع	حجم العينة		العامل الأول	العامل الثاني	ناتج القسمة
اعتدالي	100	الجذر الكامن	13.55	2.268	5.97
		التباين المفسر	22.58	3.90	
اعتدالي	200	الجذر الكامن	13.55	1.70	7.97
		التباين المفسر	22.59	2.996	
اعتدالي	300	الجذر الكامن	12.10	1.678	7.21
		التباين المفسر	20.17	2.796	
اعتدالي	400	الجذر الكامن	12.83	1.485	8.63
		التباين المفسر	21.39	2.475	
اعتدالي	500	الجذر الكامن	13.00	1.456	8.92
		التباين المفسر	21.68	2.426	
موجب الالتواء	100	الجذر الكامن	13.283	2.362	5.623
		التباين المفسر	22.138	3.937	
موجب الالتواء	200	الجذر الكامن	12.032	1.909	6.30
		التباين المفسر	20.053	3.182	
موجب الالتواء	300	الجذر الكامن	12.617	1.774	7.11
		التباين المفسر	21.029	2.957	
موجب الالتواء	400	الجذر الكامن	12.638	1.633	7.74
		التباين المفسر	21.063	2.721	
موجب الالتواء	500	الجذر الكامن	5.023	1.166	3.025
		التباين المفسر	22.832	5.298	
سالب الالتواء	100	الجذر الكامن	13.19	2.525	5.223
		التباين المفسر	21.98	4.208	
سالب الالتواء	200	الجذر الكامن	13.295	1.884	7.05
		التباين المفسر	22.159	3.140	
سالب الالتواء	300	الجذر الكامن	13.1663	1.882	6.995
		التباين المفسر	21.944	3.137	
سالب الالتواء	400	الجذر الكامن	13.516	1.687	8.01
		التباين المفسر	22.527	2.812	
سالب الالتواء	500	الجذر الكامن	13.314	1.445	9.213
		التباين المفسر	21.827	2.370	
منتظم	100	الجذر الكامن	22.942	2.030	11.3
		التباين المفسر	38.237	3.384	
منتظم	200	الجذر الكامن	21.182	1.923	11.015
		التباين المفسر	35.303	3.205	
منتظم	300	الجذر الكامن	20.625	1.633	12.63
		التباين المفسر	34.375	2.722	
منتظم	400	الجذر الكامن	20.628	1.646	12.53
		التباين المفسر	34.381	2.743	
منتظم	500	الجذر الكامن	21.059	1.628	12.935
		التباين المفسر	35.098	2.713	

- **عرض النتائج ومناقشتها**:

◄◄ للإجابة عن التساؤل الأول: "هل تختلف ملاءمة المفردات لنموذج راش تبعاً لشكل توزيع القدرة أو حجم العينة؟"، تم فحص إحصاءات ملاءمة المفردات للنموذج في كل مجموعة من البيانات، وتبين اختلاف دلالة إحصاءات ملاءمة المفردات بشكل كبير تبعا لتوزيع القدرة وحجم العينة، ويوضح الجدول (٣) إحصاءات المفردات غير الملائمة للنموذج في التوزيعات المختلفة،

جدول(٣) بيانات المفردات غير الملائمة في كل توزيع

شكل التوزيع	حجم العينة	عدد المفردات غير الملائمة	رقم المفردة	كا^2	مستوى الدلالة
اعتدالي	100	2	17	19.974	.018
			18	18.927	.026
اعتدالي	200	2	43	17.714	.039
			57	18.697	.028
اعتدالي	300	zero			
اعتدالي	400	2	7	17.786	.038
			32	20.404	.016
اعتدالي	500	1	55	17.241	.045
موجب الالتواء	100	1	32	22.512	.007
موجب الالتواء	200	1	9	18.465	.030
موجب الالتواء	300	3	21	20.488	.015
			53	28.535	.001
			60	30.990	.000
موجب الالتواء	400	3	23	24.904	.003
			26	21.339	.011
			48	20.361	.016
موجب الالتواء	500	3	17	20.556	.015
			57	22.972	.006
			60	17.609	.040
سالب الالتواء	100	3	2	17.845	.037
			9	30.725	.000
			45	17.391	.043
سالب الالتواء	200	1	25	22.295	.008
سالب الالتواء	300	4	3	20.530	.015
			9	21.851	.009
			13	18.908	.026
			18	18.806	.027
سالب الالتواء	400	4	2	23.179	.006
			18	19.624	.020
			24	19.199	.024
			33	19.879	.019
سالب الالتواء	500	2	3	22.186	.008
			7	18.446	.030
منتظم	100	2	1	18.204	.033
			5	21.402	.011
منتظم	200	1	26	21.610	.010
منتظم	300	1	48	19.038	.025
منتظم	400	1	23	20.741	.014
منتظم	500	2	22	17.745	.038
			29	25.419	.003

فبعض المفردات تكون ملائمة للنموذج في فئة من البيانات بينما تكون غير ملائمة في فئة أخرى، ولذا يجب أن يتم التعامل مع قيم إحصاءات الملاءمة باعتبارها قيماً نسبية تختلف تبعاً لاختلاف عينة التدريج، ويجب إعادة حسابها في كل مرة يستخدم فيها الاختبار المدرج، فكما يتضح من الجدول (٣) أن المفردتين ١٧ و ١٨ - على سبيل المثال - غير ملائمتين للنموذج عندما تتوزع القدرة اعتدالياً في عينة من ١٠٠ ممتحن، حيث كانت قيمة كا٢ دالة إحصائياً عند مستوى (.018) و (.026) على الترتيب، في حين أنهما ملائمتان للنموذج عندما تتوزع القدرة اعتدالياً في عينة من ٢٠٠ ممتحن.

وبذلك تتسق هذه النتائج مع نتائج دراسة عبد الجبار والشافعي (٢٠٠٤) التي توصلت إلى اختلاف دلالة إحصاءات ملاءمة المفردات لنموذج راش باختلاف حجم العينة عند استخدام برنامج (Bilog) أو برنامج (Microscale) لتدريج المفردات. وتتسق كذلك مع نتائج دراسة عبابنة (٢٠٠٤) التي توصلت إلى اختلاف دلالة ملاءمة المفردات للنموذج باختلاف حجم العينة.

- للإجابة عن التساؤل الثاني: "هل يختلف الخطأ المعياري لتقدير معلم صعوبة المفردة باختلاف شكل توزيع القدرة عند تساوي أحجام العينات؟"، تم استخدام برنامج (SPSS) لإجراء تحليل التباين الأحادي للقياسات المتكررة بهدف الكشف عن دلالة الفروق بين متوسطات الخطأ المعياري في تقدير معلم الصعوبة الناتج عن استخدام عينات متساوية الحجم ومختلفة في شكل توزيع القدرة، وتم تعديل درجات الحرية باستخدام إحصاءة (Greenhouse-Geisser)؛ نظراً لعدم تحقق شرط الكروية في البيانات. وفيما يلي نتائج تحليل التباين عند كل مستوى من مستويات حجم العينة.
  - ✓ نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير معلم الصعوبة الناتجة عن استخدام عينات مختلفة التوزيع حجمها ١٠٠ :

كشفت النتائج عن وجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة، ترجع لشكل توزيع القدرة عند حجم العينة ١٠٠، وأشارت نتائج المقارنات البعدية باستخدام اختبار LSD - كما هو موضح في الجدول (٥) - إلى وجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين كل زوج من متوسطات الخطأ المعياري في التوزيعات المختلفة.

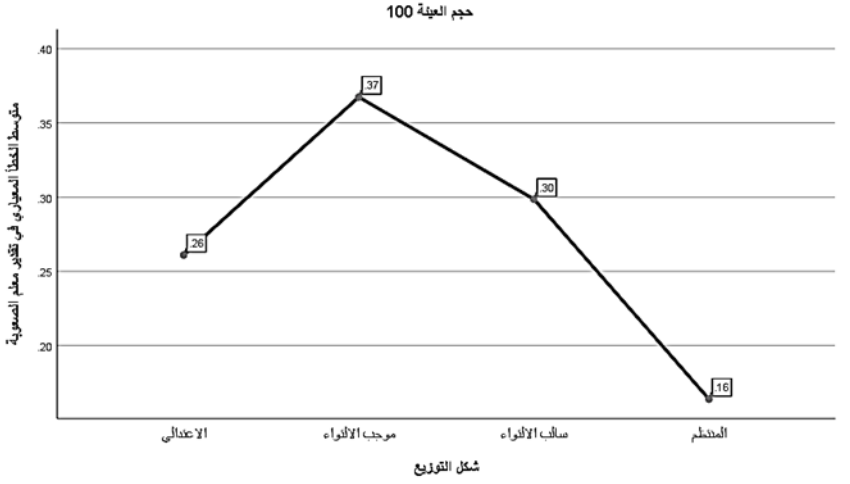
جدول (٤) نتائج تحليل التباين الأحادي للقياسات المتكررة -حجم العينة ١٠٠

درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	حجم التأثير الجزئي
1.260	**1.029**	**67.61**	**P<.001**	**0.534**

جدول(٥) المقارنات الثنائية بين التوزيعات المختلفة - حجم عينة ١٠٠

التوزيع	موجب الالتواء(م=.367)	سالب الالتواء(م=.299)	المنتظم (م=.164)
الاعتدالي (م=.261)	P<.001	P= 0.004	P<.001
موجب الالتواء (م=.367)		P= .005	P<.001
سالب الالتواء (م=.299)			P<.001

ويتضح من الشكل (٣) أن دقة تقدير معلم الصعوبة كانت أفضل في حالة التوزيع المنتظم، حيث كان متوسط الخطأ المعياري في التقدير أقل، ثم تلتها دقة التقدير في حالة التوزيع الاعتدالي، وترتب على استخدام عينات ملتوية التوزيع ارتفاع الخطأ المعياري لتقدير معلم الصعوبة، على الرغم من أن متوسط هذا الخطأ كان أقل بشكل دال في حالة التوزيع سالب الالتواء.



شكل (٣) العلاقة بين شكل التوزيع ومتوسط الخطأ المعياري في تقدير معلم الصعوبة

✓ نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير معلم الصعوبة الناتجة عن استخدام عينات مختلفة التوزيع حجمها ٢٠٠:

كشفت النتائج عن وجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة، ترجع لشكل توزيع القدرة عند حجم العينة ٢٠٠، وأشارت نتائج المقارنات البعدية - كما هو موضح في الجدول (٧) – إلى وجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع الاعتدالي ونظيره في كل من التوزيعات الأخرى، ووجود فروق دالة إحصائيا عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع المنتظم ونظيره في

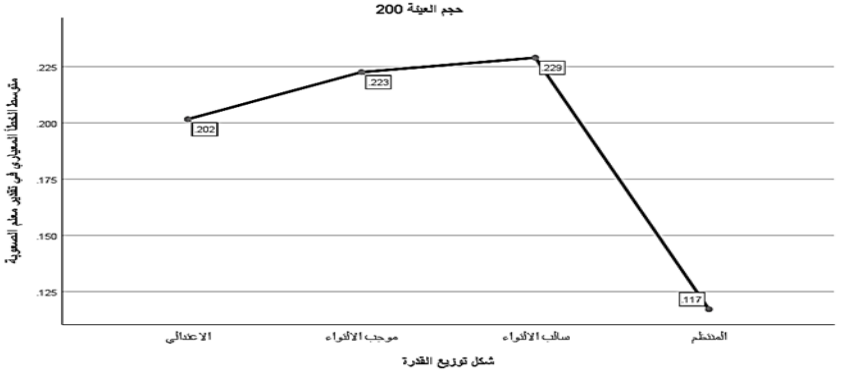
كل من التوزيعات الأخرى، في حين لم تكن هناك فروق دالة إحصائياً بين متوسط الخطأ المعياري لتقدير معلم الصعوبة عند الاعتماد على التوزيع موجب الالتواء، ونظيره عند الاعتماد على التوزيع سالب الالتواء.

جدول (٦) نتائج تحليل التباين الأحادي للقياسات المتكررة - حجم العينة ٢٠٠

درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	حجم التأثير الجزئي
1.144	.420	73.538	P<.001	.555

جدول (٧) المقارنات الثنائية بين التوزيعات المختلفة عند حجم عينة ٢٠٠

التوزيع	موجب الالتواء (م=.223)	سالب الالتواء (م=.229)	المنتظم (م=.117)
الاعتدالي (م =.202)	P=.011	P<.001	P<.001
موجب الالتواء (م=.223)		P=.655	P<.001
سالب الالتواء (م=.229)			P<.001



شكل (٤) العلاقة بين شكل التوزيع ومتوسط الخطأ المعياري في تقدير معلم الصعوبة

ويتضح من الشكل (٤) أن دقة تقدير معلم الصعوبة كانت أفضل في حالة التوزيع المنتظم؛ حيث كان متوسط الخطأ المعياري للتقدير أقل، ثم تلتها دقة التقدير في حالة التوزيع الاعتدالي، وترتب على استخدام عينات ملتوية التوزيع ارتفاع الخطأ المعياري لتقدير معلم الصعوبة.

- ✓ نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير معلم الصعوبة الناتجة عن استخدام عينات مختلفة التوزيع حجمها ٣٠٠ :

اتفقت نتائج الفروق بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة في حالة حجم العينة ٣٠٠ مع النتائج في حالة حجم العينة ٢٠٠، حيث وُجدت فروق

دالة إحصائياً عند مستوى ($p<0.001$) بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة، ترجع لشكل توزيع القدرة عند حجم العينة ٣٠٠، وأشارت نتائج المقارنات البعدية - كما هو موضح في الجدول (٩) - إلى وجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع الاعتدالي ونظيره في كل من التوزيعات الأخرى، ووجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع المنتظم ونظيره في كل من التوزيعات الأخرى، في حين لم تكن هناك فروق دالة إحصائياً بين متوسط الخطأ المعياري لتقدير معلم الصعوبة عند الاعتماد على التوزيع موجب الالتواء، ونظيره في عند الاعتماد على التوزيع سالب الالتواء.

جدول (٨) نتائج تحليل التباين الأحادي للقياسات المتكررة - حجم العينة ٣٠٠

درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	حجم التأثير الجزئي
1.095	.211	46.885	P<.001	.443

جدول (٩) المقارنات الثنائية بين التوزيعات المختلفة - حجم عينة ٣٠٠

التوزيع	موجب الالتواء (م=.195)	سالب الالتواء (م=.188)	المنتظم (م=.116)
الاعتدالي (م =.160)	P<.001	P<.001	P<.001
موجب الالتواء (م=.195)		P= .566	P<.001
سالب الالتواء (م=.188)			P<.001

شكل (٥) العلاقة بين شكل التوزيع ومتوس الخطأ المعياري في تقدير معلم الصعوبة

جدول (١٠) نتائج تحليل التباين الأحادي للقياسات المتكررة -حجم العينة ٤٠٠

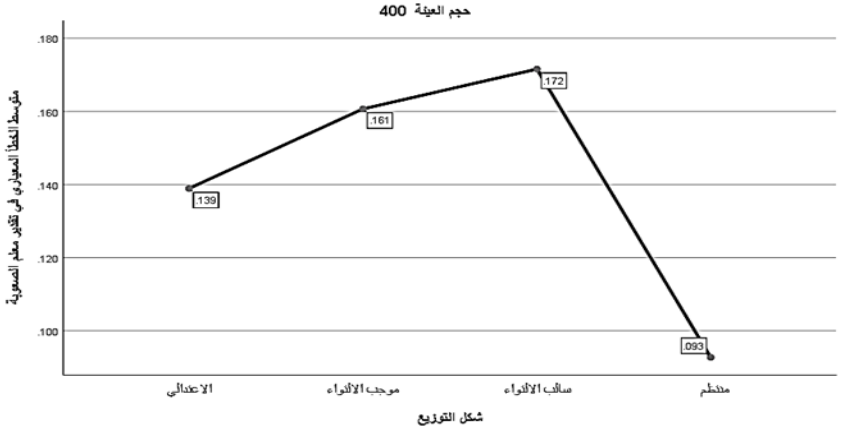
درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	حجم التأثير الجزئي
1.149	.191	57.186	P<.001	.492

✓ نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير معلم الصعوبة الناتجة عن استخدام عينات مختلفة التوزيع حجمها ٤٠٠:

اتفقت نتائج الفروق بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة في حالة حجم العينة ٤٠٠ مع النتائج في حالة حجم العينة ٢٠٠ والنتائج في حالة حجم العينة ٣٠٠، حيث وُجدت فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة، ترجع لشكل توزيع القدرة عند حجم العينة ٤٠٠، وبينت نتائج المقارنات البعدية - كما هو موضح في الجدول (١١) - وجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع الاعتدالي و نظيره في كل من التوزيعات الأخرى، ووجود فروق دالة إحصائياً عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع المنتظم و نظيره في كل من التوزيعات الأخرى، في حين لم تكن هناك فروق دالة إحصائياً بين متوسط الخطأ المعياري لتقدير معلم الصعوبة عند الاعتماد على التوزيع موجب الالتواء، و نظيره عند الاعتماد على التوزيع سالب الالتواء.

جدول (١١) المقارنات الثنائية بين التوزيعات المختلفة - حجم عينة ٤٠٠

التوزيع	موجب الالتواء (م=.161)	سالب الالتواء (م=.172)	المنتظم (م=.093)
الاعتدالي (م=.139)	P<.001	P<.001	P<.001
موجب الالتواء (م=.161)		P=.318	P<.001
سالب الالتواء (م=.172)			P<.001



شكل (٦) العلاقة بين شكل التوزيع ومتوسط الخطأ المعياري في تقدير معلم الصعوبة

✓ نتائج المقارنة بين متوسطات الأخطاء المعيارية لتقدير معلم الصعوبة الناتجة عن استخدام عينات مختلفة التوزيع حجمها ٥٠٠:

اتفقت نتائج الفروق بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة في حالة حجم العينة ٥٠٠ مع النتائج في حالة حجم العينة ٢٠٠ والنتائج في حالة حجم العينة ٣٠٠، والنتائج في حالة حجم العينة ٤٠٠، حيث كشفت النتائج عن وجود فروق دالة إحصائيا عند مستوى ($p<0.001$) بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة، ترجع لشكل توزيع القدرة عند حجم العينة ٥٠٠، وأشارت نتائج المقارنات البعدية - كما هو موضح في الجدول (١٣) - وجود فروق دالة إحصائيا عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع الاعتدالي و نظيره في كل من التوزيعات الأخرى، ووجود فروق دالة إحصائيا عند مستوى ($p<0.001$) بين متوسط الخطأ المعياري في حالة الاعتماد على التوزيع المنتظم و نظيره في كل من التوزيعات الأخرى، في حين لم تكن هناك فروق دالة إحصائيا بين متوسط الخطأ المعياري لتقدير معلم الصعوبة عند الاعتماد على التوزيع موجب الالتواء، و نظيره عند الاعتماد على التوزيع سالب الالتواء.

جدول (١٢) نتائج تحليل التباين الأحادي للقياسات المتكررة - حجم العينة ٥٠٠

درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	حجم التأثير الجزئي
1.124	.079	33.234	P<.001	.360

جدول (١٣) المقارنات الثنائية بين التوزيعات المختلفة -حجم عينة ٥٠٠

التوزيع	موجب الالتواء (م=.148)	سالب الالتواء (م=.146)	منتظم (م=.100)
الاعتدالي (م=.138)	P=.017	P=.155	P<.001
موجب الالتواء (م=.148)		P=.829	P<.001
سالب الالتواء (م=.146)			P<.001

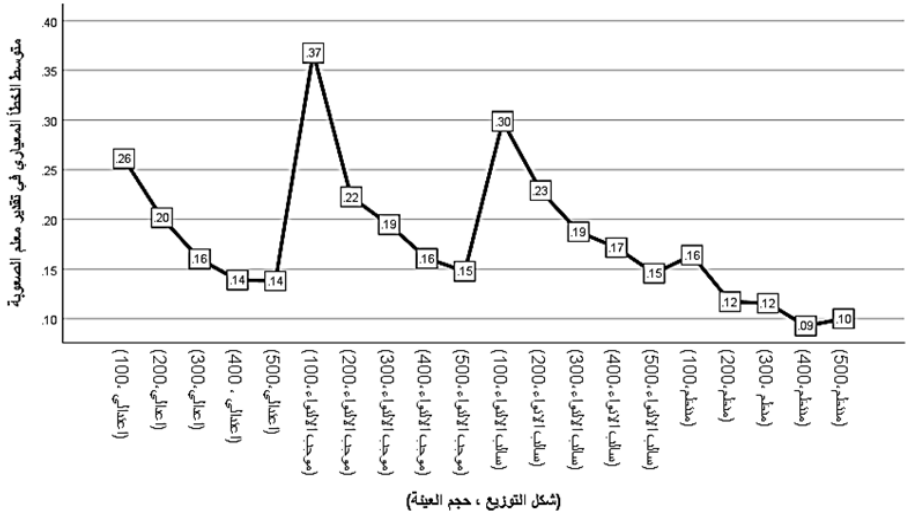
شكل (٧) العلاقة بين شكل التوزيع ومتوسط الخطأ المعياري في تقدير معلم الصعوبة عند حجم العينة ٥٠٠

ويتضح مما سبق أن متغير شكل توزيع القدرة يؤثر في قيمة الخطأ المعياري لتقدير معلم الصعوبة عند جميع مستويات حجم العينة التي تم دراستها ، وتراوح حجم التأثير بين ٠.٣٦ عندما كان حجم العينة ٥٠٠ إلى ٠.٥٥٥ عندما كان حجم العينة ٢٠٠.

- للإجابة عن التساؤل الثالث: "ما هو شكل توزيع القدرة الذي يؤدي إلى أعلى مستوى من الدقة بأقل حجم من الممتحنين؟"، تم إجراء تحليل التباين الأحادي للقياسات المتكررة للمقارنة بين متوسطات الخطأ المعياري لتقدير معلم الصعوبة عند اختلاف شكل توزيع القدرة وحجم العينة، كمتغير واحد عبارة عن زوج مرتب على الصورة (شكل التوزيع، حجم العينة)، وكشفت النتائج عن وجود فروق دالة إحصائيا عند مستوى ($p<0.001$)، وبينت نتائج المقارنات البعدية أن معظم الأزواج بينها فروق دالة إحصائيا عند مستوى ($p<0.001$)، باستثناء بعض الأزواج الموضحة في الجدول (١٥).

جدول (١٤) نتائج تحليل التباين الأحادي للقياسات المتكررة –متغير (نوع التوزيع،حجم العينة)

درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	حجم التأثير الجزئي
1.450	3.706	132.648	P<.001	.692



شكل (٨) العلاقة بين شكل التوزيع ومتوسط الخطأ المعياري في تقدير معلم الصعوبة

جدول (١٥) التوزيعات التي تساوت في متوسط الخطأ المعياري لتقدير معلم الصعوبة

(الاعتدالي، ٢٠٠)	(موجب الالتواء، ٣٠٠)		
(الاعتدالي، ٣٠٠)	(موجب الالتواء، ٤٠٠)	(المنتظم، ١٠٠)	
(الاعتدالي، ٤٠٠)	(الاعتدالي، ٥٠٠)		
(الاعتدالي، ٥٠٠)	(سالب الالتواء، ٥٠٠)		
(موجب الالتواء، ٢٠٠)	(سالب الالتواء، ٢٠٠)		
(موجب الالتواء، ٣٠٠)	(سالب الالتواء، ٣٠٠)	(سالب الالتواء، ٤٠٠)	
(موجب الالتواء، ٤٠٠)	(سالب الالتواء، ٤٠٠)	(سالب الالتواء، ٥٠٠)	(المنتظم، ١٠٠)
(موجب الالتواء، ٥٠٠)	(سالب الالتواء، ٥٠٠)		

يتضح من خلال الشكل (٨) أن التوزيع المنتظم هو التوزيع المقابل لأدنى متوسط للخطأ المعياري في تقدير معلم الصعوبة، كما يتضح من الجدول (١٥) أن الخطأ المعياري في تقدير معلم الصعوبة عند استخدام عينة من ١٠٠ ممتحن تتوزع قدرتها توزيعا منتظما يكافئ الخطأ المعياري في تقدير معلم الصعوبة عند استخدام عينات بحجم مضاعف تتوزع قدرتها اعتداليا، فلم تكن هناك فروق دالة إحصائيا بين متوسط الخطأ المعياري في تقدير معلم الصعوبة عند استخدام عينة حجمها ١٠٠ تتوزع توزيعا منتظما وعينة حجمها ٣٠٠ تتوزع توزيعا اعتداليا أو أخرى حجمها ٤٠٠ أو ٥٠٠ تتوزع توزيعا موجب الالتواء أو سالب الالتواء.

ويتبين من الشكل(٨) أن الخطأ المعياري في تقدير معلم الصعوبة يقل بشكل واضح كلما زاد حجم العينة في نفس التوزيع، وأن شكل التوزيع يؤثر بشكل كبير في هذا الخطأ، فالتوزيع المنتظم هو التوزيع الذي يقابل أدنى مستوى من الخطأ، يليه التوزيع الاعتدالي، بينما يرتفع الخطأ عند الاعتماد على توزيعات ملتوية، فعندما كان حجم العينة ١٠٠ ممتحن وصل متوسط الخطأ ٣٧ ٪ في حالة استخدام التوزيع موجب الالتواء، و ٣٠ ٪ في حالة استخدام التوزيع سالب الالتواء، في حين كان متوسط الخطأ ١٦ ٪ في حالة استخدام التوزيع المنتظم . وهذا يفسر الاختلاف الملاحظ بين نتائج الدراسات السابقة فيما يخص تأثير حجم العينة في دقة تقدير معلم الصعوبة، فقد يرجع ذلك الاختلاف إلى عدم أخذ شكل التوزيع بعين الاعتبار، حيث بينت دراسة الدرابيع (٢٠٠١) أن دقة تقدير معلم الصعوبة عند استخدام عينة حجمها ١٠٠ ممتحن تكون أفضل من دقة تقدير معلم الصعوبة عند استخدام عينة حجمها ٥٠٠ ممتحن، بينما أوصت دراسة الزبون (٢٠١٣) بزيادة حجم العينة قدر المستطاع للتقليل من أخطاء التقدير.

وهكذا تتفق نتيجة الدراسة الحالية مع ما نتائج دراسة العبابنة (٢٠٠٤)، حيث توصلت إلى أن متوسط الخطأ المعياري في تقدير معلم الصعوبة في حالة استخدام عينة عشوائية يكون أقل من متوسط الخطأ المعياري في حالة استخدام

عينة مرتفعة القدرة أو عينة منخفضة القدرة، كما تتفق مع ما توصلت إليه دراسة Karadavut (2017) من حيث زيادة دقة التقدير عندما تتوزع القدرة طبقا للتوزيع المنتظم.

- للإجابة عن السؤال الرابع: "هل يختلف مدى تحقق خاصية الموضوعية النسبية لمعلم الصعوبة باختلاف شكل التوزيع أو حجم العينة؟"، تم استخدام معامل ارتباط سبيرمان للرتب بين تقديرات معلم الصعوبة الناتجة عن كل مجموعة من البيانات والقيم الحقيقية لمعلم الصعوبة، وأظهرت النتائج أن جميع قيم معاملات الارتباط دالة إحصائياً عند مستوى ($p<0.01$)، وتراوحت بين 0.949 و 0.993 - كما هو موضح في الجدول (١٦) - حيث احتفظت المفردات بالمواقع النسبية لها باختلاف حجم عينة التدريج وشكل توزيع القدرة، مما يعد دليلاً عملياً على تحقق خاصية الموضوعية النسبية في نموذج راش. وتتفق هذه النتيجة مع نتائج دراسة بني عطا والشريفين (٢٠١٢) حيث توصلت إلى وجود فروق دالة إحصائيا بين متوسطات معالم الصعوبة، ترجع لاختلاف شكل توزيع القدرة، إلا أن قيم معاملات الارتباط بين معالم الصعوبة الناتجة عن التوزيعات المختلفة كانت دالة إحصائياً.

وتتعارض هذه النتائج مع ما توصلت إليه دراسة (عبد الجبار والشافعي، ٢٠٠٤)؛ نتيجة اعتماد تلك الدراسة على تقييم استقرار معلم الصعوبة من خلال مقارنة الفرق بين القيم المتناظرة للمعلم بمجموع الخطأ المعياري في تقديره، أي بسبب اعتمادها على مقارنة قيم المعلم باعتبارها قيم مطلقة وثابتة، وهو ما قاد الباحثين للتشكيك في نتائج البرامج المستخدمة، لكن نموذج راش في حقيقة الأمر لا يدَّعي أن قيم معلم الصعوبة مطلقة وثابتة، وإنما يدَّعي إمكانية تحقق ثبات المواقع النسبية للمفردات، بمعنى أن ترتيب المفردات بالنسبة لبعضها يظل ثابتاً على متصل القدرة مهما اختلفت عينة التدريج.

جدول (١٦)معامل ارتباط سبيرمان للرتب بين قيم معلم الصعوبة المقدرة في كل مجموعة والقيم الحقيقية له

التوزيع	معامل الارتباط	التوزيع	معامل الارتباط	التوزيع	معامل الارتباط	التوزيع	معامل الارتباط
(اعتدالي،١٠٠)	.971	(موجب الالتواء،١٠٠)	.981	(سالب الالتواء،١٠٠)	.949	(منتظم،١٠٠)	.971
(اعتدالي،٢٠٠)	.978	(موجب الالتواء،٢٠٠)	.969	(سالب الالتواء،٢٠٠)	.965	(منتظم،٢٠٠)	.987
(اعتدالي،٣٠٠)	.990	(موجب الالتواء،٣٠٠)	.985	(سالب الالتواء،٣٠٠)	.978	(منتظم،٣٠٠)	.987
(اعتدالي،٤٠٠)	.994	(موجب الالتواء،٤٠٠)	.991	(سالب الالتواء،٤٠٠)	.987	(منتظم،٤٠٠)	.987
(اعتدالي،٥٠٠)	.990	(موجب الالتواء،٥٠٠)	.987	(سالب الالتواء،٥٠٠)	.990	(منتظم،٥٠٠)	.993

- **التوصيات**:

في ضوء ما سبق من نتائج توصي الدراسة الحالية بضرورة التأكد من ملاءمة المفردات للنموذج في كل مرة نستخدم الاختبار المدرج؛ حيث اختلفت نتائج الملاءمة باختلاف كل من شكل توزيع القدرة وحجم العينة، كما توصي الدراسة بالاعتماد على التوزيع المنتظم للقدرة عند تدريج الاختبارات باستخدام نموذج راش، وبذلك تكون الطريقة الأنسب لاختيار عينة التدريج هي التي تؤدي إلى تمثيل مناسب ومتكافئ لجميع مستويات القدرة، ولذا يمكن اختيار العينة بأسلوب المعاينة العشوائية الطبقية بدلاً من طريقة المعاينة العشوائية، كما يجب تجنب الاعتماد على عينات ملتوية التوزيع. وقد اتفقت نتائج الدراسة الحالية مع نتائج جميع الدراسات السابقة من حيث زيادة دقة التقدير بزيادة حجم عينات التدريج وذلك بالنسبة لجميع التوزيعات المستخدمة.

وللوصول لرؤية متكاملة حول تأثير شكل توزيع القدرة على ملاءمة النموذج للبيانات، ودقة تقدير معالم النماذج في نظرية الاستجابة للمفردة، تقترح الدراسة تكرار البحث عند اختلاف مدى الصعوبة الذي يغطيه الاختبار، فنتائج الدراسة الحالية محددة باختبار مثالي تغطي مفرداته المدى من -٣ إلى +٣ ، فهل تختلف النتائج التي توصلت إليها الدراسة إذا تغير هذا المدى كما هو الحال في الاختبارات متطرفة الصعوبة أو الاختبارات التي بها فجوات على متصل القدرة ؟.

وللتحقق مما توصلت إليه الدراسة الحالية وإمكانية تعميم نتائجها يمكن دراسة نفس المتغيرات عند الاعتماد على برامج أخرى سواء لتوليد البيانات، أوتحليلها في ضوء نموذج راش، وكذلك عند الاعتماد على نماذج أخرى لنظرية الاستجابة للمفردة.

- **قائمة المراجع**:

- بني عطا، زايد صالح والشريفين، نضال (٢٠١٢). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. المجلة الأردنية في العلوم التربوية، ٨(٢)، ١٥١ -١٦٦.
- الثوابية، أحمد محمود (٢٠١٠). أثر حجم العينة على تقدير صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة. مجلة جامعة دمشق للعلوم التربوية -سوريا، ٢٦(١)، ٥٢٥ -٥٥٦.
- الدرابيع، ماهر يونس (٢٠٠١). فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة نموذج راش في دقة تقدير قدرة الفرد ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار. دراسات - العلوم التربوية - الأردن، ٢٨(١)، ١٩٧ -٢٠٨.
- درويش، هيثم فؤاد (٢٠١٤). دقة تقدير معالم الفقرة باستخدام نظرية الاستجابة للفقرة البارامترية المعدلة واللابارامترية باختلاف حجم العينة وطول الاختبار: دراسة مقارنة، (رسالة دكتوراه غير منشورة)، كلية الدراسات العليا، الجامعة الأردنية، الأردن.

- الزبون، حابس سعد موسى (٢٠١٣). أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس. مجلة جامعة النجاح للعلوم الإنسانية - فلسطين، ٢٧(٦)، ١٣١٣ -١٣٣٤.

- الشريفين، نضال كمال (٢٠١٢). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة والخصائص السيكومترية للاختبار في ضوء تغير حجم العينة. المجلة التربوية، الكويت، ٢٦(١٠٤)، ١٧٧ -٢٣٨.

- شما، يمان نزار(٢٠١٣). أثر حجم العينة على دقة تقدير صعوبة المفردات وقدرة الأفراد باستخدام نموذج راش. مجلة كلية الآداب جامعة بغداد - العراق، ١٠٥، ٦٧٣–٦٩٨.

- الشمراني، محمد موسى (٢٠١٦). أثر حجم العينة وطول الاختبار والتفاعل بينهما على تقدير معلم الصعوبة والتمييز والتخمين والخطأ المعياري باستخدام نظرية الاستجابة للفقرة. مجلة كلية التربية، جامعة طنطا - مصر، ٦٤(٤)، ٣٠٦ -٣٦٥.

- عبابنة، عماد غصاب باير(٢٠٠٤). أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة. (رسالة دكتوراه غير منشورة)، كلية الدراسات التربوية العليا، جامعة عمان العربية، الأردن.

- عبدالجبار، عادل بن صلاح عمر، والشافعي، محمد منصور محمد (٢٠٠٤). أثر حجم عينة التحليل على مؤشرات الملاءمة الإحصائية وتقديرات الصعوبة المتضمنة ببرنامجي (مايكروسكال) و (بايلوج) للمفردات باستخدام نموذج (راش): دراسة محاكاة. *رسالة التربية وعلم النفس* - السعودية، ٢٥، ٢٠٥ -٢٦٧.

- القيسي، حسين عبد النبي (٢٠١٣). دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية الاستجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار: دراسة مقارنة. (رسالة دكتوراه غير منشورة). كلية التربية، جامعة اليرموك، الأردن.

- القيسي، حسين عبد النبي (٢٠١٦): أثر حجم العينة وطول الاختبار على دقة تقدير معالم الفقرة والقدرة باستخدام نظرية الاستجابة للمفردة اللامعلمية. مؤتة للبحوث والدراسات -العلوم الإنسانية والاجتماعية - الأردن، ٣١(٥)، ٢٠٣ -٢٤٦.

- Alderson, J., Clapham, C., & Wall, D. (1995). Language test construction and evaluation. Cambridge: Cambridge Uni. Press.

- Hambleton R., Swaminathan H. & Rogers, H. (1991). Fundamentals of item response theory. London: Sage Publications.

- Han, K. T & Hambleton, R. K (2014). User's Manual for WinGen3: Windows Software that Generates IRT Model Parameters and Item Responses. Center for Educational Assessment Research Report No. 642, Amherst, MA: University of Massachusetts, Center for Educational Assessment.

- He, Q. & Tymms, P. (2005). A computer-assisted test design and diagnosis system for use by classroom teachers. Journal of Computer Assisted Learning, 21, 419-429.

- Karadavut, T.(2017). Estimation of item response theory models when ability is uniformly distributed. The Eurasia Proceedings of Educational & Social Sciences(EPESS),7,30-37
- Khan, M. I. (2014). Recovery and stability of item parameter and model fit across varying sample sizes and test lengths in Rasch analysis with small sample. Social Science International, 30(1), 43-60.
- Linden, W. (2005). Statistics for social and behavioral science: Linear models for optimal test design. New York: Spring.
- Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlbaum Associates
- McNamara, T. (1996). Measuring second language performance, London: Longman.
- Shultz, K. & Whitney, D. (2005). Measurement theory in action. London: Sage Publications.
- Sijtsma, K. & Molenaar, I. (2002). Introduction to nonparametric item response theory. London: Sage Publications.
- Thissen, D. (2000). Reliability and measurement precision. In H. Wainer, N. Dorans, R. Flaugher, B. Green, R. Mislevy, L. Steinberg, & D. Thissen (Eds.): Computerized adaptive testing: A primer (2nd ed., pp. 159-184). London: Lawrence Erlbaum Associates.
- Weiss, D. & Yoes, M. (1991). Item response theory. In R. Hambleton & J. Zaal (Eds.), Advances in educational and psychological testing: Theory and applications (69-96). London: Kluwer Academic Publishers.
- Wright, B. (1977). Solving measurement problems with the Rasch Model. Journal of Educational Measurement, 14(2) , 97-116.
- Wright, B. & Stone, M. (1979). Best test design, Chicago: MESA Press.
- Xing, D. & Hambleton, R. (2004). Impact of test design, item quality, and item bank size on the psychometric properties of computer-based credentialing examinations. Educational and Psychological Measurement, 64(1), 5-21.

❋❋❋❋❋❋❋