

2020

## The Effect of Ability Estimation Method and Sample Size on the Percentages of Model Misfitting Response Patterns and Distributions Characteristics of Person-fit Indices on Mixed-Format Tests

Ali Abed Alrheem Tawalbeh

*Yarmouk University/Jordan*, alitawalbeh19773010@gmail.com

Nedal Kamal Al-Shraifin

*Yarmouk University/Jordan*, shraifin@yu.edu.jo

Follow this and additional works at: [https://digitalcommons.aaru.edu.jo/jaou\\_edpsych](https://digitalcommons.aaru.edu.jo/jaou_edpsych)

---

### Recommended Citation

Tawalbeh, Ali Abed Alrheem and Al-Shraifin, Nedal Kamal (2020) "The Effect of Ability Estimation Method and Sample Size on the Percentages of Model Misfitting Response Patterns and Distributions Characteristics of Person-fit Indices on Mixed-Format Tests," *Journal of Al-Quds Open University for Educational & Psychological Research & Studies*: Vol. 11 : No. 30 , Article 9.

Available at: [https://digitalcommons.aaru.edu.jo/jaou\\_edpsych/vol11/iss30/9](https://digitalcommons.aaru.edu.jo/jaou_edpsych/vol11/iss30/9)

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Journal of Al-Quds Open University for Educational & Psychological Research & Studies by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact [rakan@aar.edu.jo](mailto:rakan@aar.edu.jo), [marah@aar.edu.jo](mailto:marah@aar.edu.jo), [dr\\_ahmad@aar.edu.jo](mailto:dr_ahmad@aar.edu.jo).

# أثر طريقة تقدير القدرة وحجم العينة في نسب أنماط الاستجابة غير المطابقة للنموذج وخصائص التوزيعات لمؤشرات مطابقة الفرد على الاختبارات مختلطة الشكل

## The Effect of Ability Estimation Method and Sample Size on the Percentages of Model Misfitting Response Patterns and Distributions Characteristics of Person-fit Indices on Mixed-Format Tests

**Ali Abed Alrheem Tawalbeh**

Ph.D. Student/Yarmouk University/ Jordan

alitawalbeh19773010@gmail.com

**علي عبد الرحيم أحمد طوالة**

طالب دكتوراه/ جامعة اليرموك/ الأردن

**Nedal Kamal al-shraifin**

Prof./Yarmouk University/Jordan

Shraifin@Yu.edu.jo

**نضال كمال الشرايفين**

أستاذ دكتور / جامعة اليرموك/ الأردن

Received: 6/ 4/ 2019, Accepted: 2/ 6/ 2019.

DOI: 10.33977/1182-011-030-004

https://journals.qou.edu/index.php/nafsia

تاريخ الاستلام: 6/ 4/ 2019م، تاريخ القبول: 2/ 6/ 2019م.

E-ISSN: 2307-4655

P-ISSN: 2307-4647

the two indices increase and the distribution of the two indices is close to a normal distribution by increasing the sample size in different ability estimation methods. The percentages of misfitting patterns detected by the index  $l_{z(MIX)}^*$  was higher than those detected by the  $l_{z(MIX)}$  index, and that the distribution of the  $l_{z(MIX)}^*$  index was closer to the normal distribution than the  $l_{z(MIX)}$  index in different methods for estimating ability and in all samples.

In light of the findings of the study, the researchers recommended using the index and the WLE method for ability estimation to detect misfitting response patterns in mixed-format tests.

**Keywords:** Person-Fit Indices, Mixed-Format Tests, Ability Estimation Method.

## المقدمة

أدى ظهور نظرية الاستجابة للفقرة (Item Response Theory) (IRT) وتطورها إلى إيجاد حلول لمعظم المشاكل التي واجهت الاختبارات المصممة وفق النظرية التقليدية في القياس، والتي تؤثر في موضوعية ودقة القياس، حيث تقوم هذه النظرية على مجموعة من المبادئ من خلال محاولتها نمذجة العلاقة بين مستوى سمة معينة لدى الفرد واحتمال إجابته على فقرات اختبار تقيس هذه السمة (علام، 2005). وتعتمد نظرية الاستجابة للفقرة على مجموعة من الافتراضات الأساسية هي: افتراض أحادية البعد، وافتراض منحنى خصائص الفقرة وافتراض الاستقلال الموضوعي، وافتراض التحرر من السرعة (Hambleton & Swaminathan, 1985).

وقد ظهرت العديد من نماذج الاستجابة للفقرة التي تصف بشكل رياضي العلاقة بين مستوى قدرة المفحوص واحتمالية استجابته على الفقرة استجابة صحيحة، وهناك تصنيف شائع لهذه النماذج يعتمد على الطريقة التي يجري فيها تدرج الاستجابة على الفقرة هو: نماذج الاستجابة ثنائية التدرج (Dichotomous Re-sponse Models) وتستخدم مع الفقرات التي يكون فيها تدرجان فقط للاستجابة (0,1)، مثل النموذج أحادي المعلمة (1PL) والنموذج ثنائي المعلمة (2PL) والنموذج ثلاثي المعلمة (3PL)، ونماذج الاستجابة متعددة التدرج (Polytomous Response Models) وتستخدم مع الفقرات التي يكون فيها أكثر من تدرجان للاستجابة على الفقرة، مثل نموذج الاستجابة المتدرجة (GRM) ونموذج الاستجابة الأسمية (NRM) ونموذج التقدير الجزئي العام (GPCM).

ومن القضايا المرتبطة بنظرية الاستجابة للفقرة والتي تؤثر بشكل كبير في صدق نتائج القياس ودقة القرارات التي تبني على هذه النتائج، هو مدى مطابقة نمط استجابة الفرد لنموذج الاستجابة للفقرة (IRT model) المستخدم، والذي يعرف عادة باسم مطابقة الفرد (Meijer and Sijtsma, 2001)، ويعد تحليل مطابقة الفرد أحد الأنواع الخمسة للتحليلات الأكثر شيوعاً للكشف عن التزييف في درجات الاختبار (Olson and Fremer, 2013).

## الملخص:

هدفت هذه الدراسة إلى الكشف عن أثر طريقة تقدير القدرة وحجم العينة في خصائص توزيعات مؤشري مطابقة الفرد  $l_{z(MIX)}$  و  $l_{z(MIX)}^*$  ونسب أنماط الاستجابة غير المطابقة التي يكشف عنها هذان المؤشران على الاختبارات مختلطة الشكل، ولتحقيق أهداف الدراسة قام الباحثان باختيار عينات عشوائية بأعداد (1000، 2000، 5000) طالباً من طلبة الصف الثامن الذين قاموا بالإجابة على فقرات الرياضيات في الكراسة الاختبارية الأولى من اختبار (TIMSS) الدولي لعام 2015، واستخدم الباحثان ثلاث طرق لتقدير قدرات الطلاب هي: (MAP, WLE, MLE).

وقد أظهرت نتائج الدراسة أن نسب الكشف عن أنماط الاستجابة غير المطابقة للمؤشرين تزداد، كما أن توزيع المؤشرين يزداد قرباً من التوزيع الطبيعي عند استخدام طريقة (WLE) لتقدير القدرة في عينات الدراسة المختلفة، وأظهرت النتائج أيضاً أن نسب الكشف عن أنماط الاستجابة غير المطابقة للمؤشرين تزداد، كما أن توزيع المؤشرين يزداد قرباً من التوزيع الطبيعي بزيادة حجم العينة في الطرق المختلفة لتقدير القدرة، وكانت نسب الأنماط غير المطابقة التي كشف عنها مؤشر  $l_{z(MIX)}^*$  أعلى من النسب التي كشف عنها مؤشر  $l_{z(MIX)}$  وكان توزيع مؤشر  $l_{z(MIX)}^*$  أقرب للتوزيع الطبيعي من مؤشر  $l_{z(MIX)}$  في الطرق المختلفة لتقدير القدرة وفي جميع العينات.

وفي ضوء ما توصلت إليه الدراسة من نتائج يوصي الباحثان باستخدام مؤشر  $l_{z(MIX)}^*$  وطريقة (WLE) لتقدير القدرة، للكشف عن أنماط الاستجابة غير المطابقة في الاختبارات مختلطة الشكل.

الكلمات المفتاحية: مؤشرات مطابقة الفرد، الاختبارات مختلطة الشكل، طرق تقدير القدرة.

## Abstract

This study aimed to detect the effect of the ability estimation method and sample size on the distributions characteristics of person-fit indices and  $l_{z(MIX)}$ , and  $l_{z(MIX)}^*$  the percentages of the misfitting response patterns detected by these two indices on mixed-form tests. To achieve the purpose of the study, the researchers choose random samples (1000, 2000, 5000) of eighth grade students who responded to the mathematics items in the first booklet of the international TIMSS test 2015. The researchers also used three methods for estimating student's ability (MAP, WLE, MLE).

The results of the study showed that the percentages of misfitting response patterns for the two indices increase and the distribution of the two indices is close to a normal distribution when using the WLE method for ability estimation in the different study samples. The results of the study also showed that the percentages of misfitting response patterns for

ونتيجة النقص الحاد في البحث في تقييم مطابقة الفرد لنماذج الاستجابة للفقرة في الاختبارات مختلطة الشكل، بسبب عدم وجود مؤشرات مطابقة تناسب هذا الشكل من الاختبارات؛ قدم سنهاري (Sinharay, 2016) مؤشري،  $l_z(MIX)$  و  $l_z^*(MIX)$  للاختبارات مختلطة الشكل كتطوير لمؤشري  $(l_z^*, l_z)$ ، اللذين يستخدمان في الاختبارات التي تتكون من فقرات ثنائية التدرج فقط، ويتم حساب مؤشر  $l_z(MIX)$  الذي طوره سنهاري (Sinharay, 2016) وفق المعادلة التالية:

$$l_z(MIX) = \frac{l - E(l)}{\sqrt{Var(l)}} \dots \dots \dots (1)$$

إذ تمثل  $l$  لوغاريتم الأرجحية لدرجة المفحوص، وتمثل  $E(l)$  القيمة المتوقعة للوغاريتم الأرجحية لدرجة المفحوص، ويمثل المقام  $Var(l)$  تباين المؤشر  $l_z(MIX)$  (تباين لوغاريتم الأرجحية لدرجة المفحوص) الذي يعتمد على دالة الوزن الحقيقي للقدرة  $w_{ij}(\hat{\theta})$ .

ونتيجة انحراف توزيع مؤشر  $l_z(MIX)$  عن التوزيع الطبيعي عندما يتم استبدال القدرة الحقيقية  $(\theta)$  بالقدرة المقدرة  $(\hat{\theta})$ ؛ قدم سنهاري (Sinharay, 2016) تصحيح لهذا المؤشر بهدف الحصول على التوزيع الطبيعي المعياري، من خلال الاعتماد على الأوزان المعدلة للقدرة  $\tilde{w}_{ij}(\theta)$  في معادلة حساب مؤشر المطابقة

$$l_z^*(MIX) = \frac{l - E(l) + \hat{c}_n(\hat{\theta})s_0(\hat{\theta})}{\sqrt{Var(l)}} \dots \dots \dots (2)$$

ويجري حساب  $\hat{c}_n(\hat{\theta})$  وفق المعادلة التالية:

$$\hat{c}_n(\hat{\theta}) = \frac{\sum_{i=1}^n \sum_{j=0}^{m_i} \hat{p}_{ij}(\hat{\theta})w_{ij}(\hat{\theta})}{\sum_{i=1}^n \sum_{j=0}^{m_i} \hat{p}_{ij}(\hat{\theta})s_{ij}(\hat{\theta})}$$

ويجري حساب  $s_0(\hat{\theta})$  وفق المعادلة التالية:

$$s_0(\hat{\theta}) = \begin{cases} 0 & \text{if } \hat{\theta} = MLE \\ \frac{I(\hat{\theta})}{2I(\hat{\theta})} & \text{if } \hat{\theta} = WLE \\ \frac{d \log f(\hat{\theta})}{d(\hat{\theta})} & \text{if } \hat{\theta} = MAP \end{cases}$$

كما يتم حساب  $s_{ij}(\hat{\theta})$  وفق المعادلة التالية:

$$s_{ij}(\hat{\theta}) = \frac{\hat{p}_{ij}(\hat{\theta})}{p_{ij}(\hat{\theta})}$$

وتمثل  $p_{ij}(\hat{\theta})$  احتمال حصول الفرد ذو القدرة  $(\theta)$  على

ويهتم تحليل مطابقة الفرد (person-fit analysis) بالكشف عن الأداء غير النمطي على الاختبار كما ينعكس في نمط الاستجابة (response pattern) على فقرات الاختبار (Meijer and Sijtsma, 2001)، ويساعد تحليل مطابقة الفرد الباحث في الحصول على معلومات عن سلوك استجابة الأفراد على فقرات الاختبار (Glas and Meijer, 2003)، ويجب مؤشر مطابقة الفرد على أسئلة مثل "هل القدرة المقدرة هي المقياس المناسب أو التمثيل المناسب للقدرة الحقيقية للمفحوص؟" (De Ayala, 2009).

وتظهر أنماط الاستجابة غير المطابقة التي يقدمها الفرد غير المطابق لنموذج الاستجابة للفقرة المستخدم، عندما يجيب الأفراد ذوو القدرة المتدنية على العديد من الفقرات الصعبة إجابة صحيحة أو يجيب الأفراد ذوي القدرة العالية بشكل خاطئ على الفقرات السهلة بما يتجاوز ما يمكن توقعه عن طريق الصدفة (Steinkamp, 2017). وتؤدي العديد من العوامل لظهور الاستجابات غير المطابقة (misfitting responses) منها: التخمين والغش والتباطؤ في تأدية الاختبار، والتكاسل والاستجابة الإبداعية، والأخطاء في تعبئة ورقة الإجابة، ووجود أكثر من قدرة متضمنة في الاختبار، والمرض، وقلق الاختبار ونقص الدافعية، والخصائص الديموغرافية كالجنس، والعرق، والتحيز الثقافي (Meijer, 1996; Tatsuoka, 1996; Birenbaum, 1986; De la Torre and Deng, 2008).

ويعرف نمط الاستجابة غير المطابق وفق نظرية الاستجابة للفقرة على أنه نمط استجابة ملاحظ ينحرف عن نمط الاستجابة المتوقع، استناداً إلى قدرة الفرد وإلى خصائص نموذج الاستجابة للفقرة المستخدم (Meijer and Sijtsma, 2001; Reise, 1990). ومن بين المؤشرات المعتمدة على نظرية الاستجابة للفقرة (IRT) عدد المؤشرات المبنية على الأرجحية (likelihood-based indi-ces) كمؤشرات  $(l_z^*, l_z)$  الأكثر شيوعاً وقوة في الكشف عن أنماط الاستجابة غير المطابقة (Armstrong, Stoumbos, Kung and Shi, 2007)، حيث تمتاز هذه المؤشرات بسهولة الحساب والفهم وتقيس بشكل مباشر المطابقة بين نموذج الاستجابة للفقرة ونمط استجابة المفحوص (De La Torre and Deng, 2008). وتفترض هذه المؤشرات معيارية توزيع المؤشر حتى يكون مفيداً إلى أقصى حد مدى مطابقة نمط استجابة المفحوص لنمط الاستجابة الذي يتنبأ به النموذج (Karabatsos, 2003). إلا أن استخدام هذه المؤشرات اقتصر في الغالب على الفقرات ثنائية التدرج (Sinharay, 2015).

وبعد أن أصبحت الاختبارات مختلطة الشكل شائعة الاستخدام في الاختبارات المدرسية والوطنية، نتيجةً لسماح التنوع في أشكال الفقرات بقياس مجموعة أوسع من المهارات مقارنة باستخدام شكل واحد، إذ يسمح دمج الأشكال المختلفة للفقرات في نفس الاختبار بالاستفادة من نقاط قوتها مع التعويض عن نقاط ضعفها، ويمكن أن تتكامل ميزات التمثيل العيني الواسع للمحتوى والثبات المرتفع والنتائج الموضوعية التي توفرها فقرات الإجابة المنتقاة، مع ميزات صدق المحتوى وصدق البناء والمنعة ضد التخمين والوصول المتكامل للإدراك المعقد التي تقدمها فقرات الاستجابة المصاغة (Cao, 2008).

(Posterior):

تعد هذه الطريقة أحد أساليب بيز (Bayes) في تقدير القدرة، حيث تعتمد أساليب بيز (Bayes) في تقدير القدرة على دمج المعلومات عن أرجحية (likelihood) الحصول على الإجابة الصحيحة أو غير الصحيحة، مع الافتراضات عن الشكل والوسط الحسابي والتباين لتوزيع القدرة في المجتمع المستهدف، وتفترض أساليب بيز (Bayes) أن التوزيع القبلي (prior) لقدرات الأفراد، هو توزيع طبيعي بوسط حسابي (0) وانحراف معياري (1)، وبعد الحصول على استجابات الأفراد على الفقرة الأولى، يجري حساب دالة الأرجحية للاستجابات ودمجها مع المعلومات عن التوزيع القبلي لقدرات الأفراد، للحصول على التوزيع المعدل للقدرة والذي يسمى التوزيع البعدي (posterior)، ثم يعمل هذا التوزيع البعدي للقدرة كتوزيع قبلي ليتم دمجها مع دالة الأرجحية لاستجابات الأفراد على الفقرة الثانية، وتستمر هذه العملية بالتتابع (فقرة فقرة) عن طريق استخدام التوزيع البعدي للقدرة السابقة كتوزيع قبلي للفقرة التي تليها، ودمج معلومات هذا التوزيع مع دالة الأرجحية لاستجابة الأفراد على الفقرة حتى تنتهي جميع فقرات الاختبار (Wang & Vispoel, 1998).

وتستخدم طريقة (MAP) القيمة المنوالية (mode) للتوزيع البعدي كتقدير للقدرة ( $\theta$ )، ويجري الحصول على تقدير القدرة بطريقة (MAP) لنمط الاستجابة المعطى (u) عن طريق حساب المشتقة الأولى للوغاريتم التوزيع البعدي للقدرة ومساواتها بالصفء، حسب المعادلة التالية:

$$\frac{\partial \log[P(\theta|u)]}{\partial \theta} = \frac{\partial \log[L(u|\theta)]}{\partial \theta} + \frac{\partial \log[P(\theta)]}{\partial \theta} = 0 \dots (6)$$

وتمثل  $P(\theta|u)$  التوزيع البعدي للقدرة ( $\theta$ ) لنمط الاستجابة (u)، وتمثل التوزيع القبلي للقدرة ( $\theta$ ).

ويجري الحكم على فاعلية مؤشرات المطابقة المبنية على الأرجحية من خلال محكين: المحك الأول، يتمثل بخصائص توزيع مؤشر المطابقة ومدى قربه من التوزيع الطبيعي المعياري، فكلما اقترب توزيع المؤشر من التوزيع الطبيعي تحرر من مستويات القدرة المختلفة، وأصبح بالإمكان تحديد قيمة حرجة واحدة لإصدار الأحكام على قيم المؤشر، ويجري الاستدلال على هذا المحك بدراسة إحصائيات (الوسط الحسابي، والانحراف المعياري، والالتواء، والتفلطح) للتوزيع ومقارنتها بخصائص التوزيع الطبيعي أو من خلال استخدام اختبار كولموجوروف-سميرنوف (Kolmogorov-Smirnov)، والمحك الثاني، يتمثل بقوة المؤشر وقدرته على الكشف عن أنماط الاستجابة غير المطابقة من خلال نسبة أنماط الاستجابة غير المطابقة التي يكشف عنها المؤشر (De La Torre & Deng, 2008; Birenbaum, 1986).

وتختبر مؤشرات مطابقة الفرد  $l_z(MIX)$  و  $l_z^*(MIX)$  وفرضية صفرية تفترض أن أنماط الاستجابة متوافقة مع نموذج الاستجابة للفقرة المستخدم، مقابل فرضية بديلة تفترض أن أنماط الاستجابة غير متوافقة مع النموذج المستخدم، ويتم رفض الفرضية الصفرية إذا كانت إحصائية الاختبار المحسوبة أقل من القيمة الحرجة المحددة، بحيث يحقق الاختبار مستوى الدلالة الإحصائية، ويجري

الدرجة  $j$  في الفقرة  $i$ ، وتمثل  $w_{ij}(\hat{\theta})$  دالة الوزن الحقيقي للقدرة التي يعبر عنها بالمعادلة التالية:

$$w_{ij}(\hat{\theta}) = \log p_{ij}(\hat{\theta})$$

كما تمثل  $\hat{p}_{ij}(\hat{\theta})$  المشتقة الأولى لدالة احتمال حصول المفحوص ذو القدرة ( $\theta$ ) على الدرجة  $j$  في الفقرة  $i$ ، ويمثل المقام  $Var l(\hat{\theta})$  التباين المعدل لمؤشر  $l_z^*(MIX)$  (تباين لوغاريتم الأرجحية لدرجة المفحوص) الذي يعتمد على دالة الوزن المعدل للقدرة  $\tilde{w}_{ij}(\hat{\theta})$  والتي يعبر عنها بالمعادلة التالية:

$$\tilde{w}_{ij}(\hat{\theta}) = w_{ij}(\hat{\theta}) - \hat{c}_n(\hat{\theta})s_{ij}(\hat{\theta})$$

ويتأثر حساب مؤشرات مطابقة الفرد المبنية على الأرجح، بقيمة قدرة الفرد ومن الطرق الشائعة لتقدير القدرة:

1. طريقة الأرجحية العظمى (Maximum Likelihood) (MLE) (Estimation)

يعتمد تقدير قيمة القدرة ( $\theta$ ) في هذه الطريقة على نمط استجابة المفحوص على مجموعة الفقرات، ويتم الحصول على تقديرات القدرة للأفراد ( $\theta$ ) بطريقة (MLE) والتي تعظم دالة الأرجح، عن طريق إيجاد المشتقة الأولى للوغاريتم دالة الأرجحية مساواتها بالصفء حسب المعادلة التالية (Hambelton & Swami, 1985):

$$\frac{\partial}{\partial \theta} \ln l = 0 \dots (3)$$

ويعبر عن لوغاريتم دالة الأرجحية في الاختبارات مختلطة الشكل بالمعادلة التالية (Sinharay, 2016):

$$l = \sum_{i=1}^n \sum_{j=0}^{m_i} d_j(x_i) \log [p_{ij}(\hat{\theta})] \dots (4)$$

وتمثل عدد فئات الاستجابة - 1، وتمثل  $n$  عدد فقرات الاختبار وتعبّر  $p_{ij}(\hat{\theta})$  عن احتمال حصول المفحوص على الدرجة  $j$  في الفقرة  $i$ .

2. طريقة الأرجحية الموزونة (Weighted Likelihood) (WLE) (hood Estimation)

قدم وارم (Warm, 1989) هذه الطريقة التي تعتمد في تقدير القدرة على توزيع دالة الأرجحية ببعض الدوال المناسبة، للتقليل من التحيز في تقدير القدرة بطريقة الأرجحية العظمى (MLE)، ويمكن الحصول على تقدير الأرجحية الموزونة للقدرة من خلال حل المعادلة التالية:

$$\frac{\partial \ln L((u|\theta))}{\partial \theta} + \frac{\partial \ln w(\theta)}{\partial \theta} = 0 \dots (5)$$

وتمثل  $L(\theta|u)$  دالة الأرجحية لنمط الاستجابة، وتمثل  $w(\theta)$  دالة وزن الأرجحية عند مستوى القدرة ( $\theta$ ).

3. طريقة التقدير البعدي الأعظم (Maximum A Posterior) (MAP)



مولدة وفق النموذج ثنائي المعلمة (2PLM)، وباستخدام ثلاث طرق لتقدير القدرة هي: طريقة الأرجحية العظمى (MLE) وطريقة بيز (EAP) وطريقة التقدير الموزونة (BIW)، أظهرت النتائج أن الوسط الحسابي والانحراف المعياري لتوزيع مؤشر المطابقة كانت قريبة من القيم النظرية (0,1) عند استخدام القدرة الحقيقية في معظم الظروف، وأن التباين لتوزيع مؤشر المطابقة كانت أقل من القيمة النظرية (1) عند استخدام القدرة المقدرة في جميع ظروف الدراسة، كما أظهرت النتائج أن قوة المؤشر تتأثر بطريقة تقدير القدرة، إذ كانت قوة المؤشر في الكشف عن أنماط الاستجابة غير المطابقة أكبر عند استخدام طريقة (BIW) مقارنة بالطرق الأخرى المستخدمة في الدراسة، في حين كانت قوة المؤشر  $I_Z$  في الكشف متقاربة عند استخدام طريقتي (EAP) و (MLE) لتقدير القدرة.

وأجرى ميجر ونيرنج (Meijer & Nering,1997) دراسة هدفت إلى تقصي أثر طريقة تقدير القدرة (MLE) و (EAP) و (BIW) على قوة المؤشر  $I_Z$  في الكشف عن أنماط الاستجابة غير المطابقة، باستخدام بيانات مولدة وفق النموذج الثنائي والثلاثي المعلمة، وأظهرت نتائج الدراسة أن طريقة (BIW) تقلل من التحيز في تقدير القدرة، وتحسن من قوة الكشف عن أنماط الاستجابة غير المطابقة، مقارنة بطريقتي (MLE) و (EAP).

وقام سنجرز (Snijders,2001) بدراسة هدفت إلى إجراء تصحيح على توزيع مؤشر  $I_Z$  عند استخدام القدرة المقدرة لحسابه من خلال التعديل الإحصائي لتباين المؤشر، ولغايات الدراسة استخدم الباحث بيانات مولدة وفق النموذج اللوجستي ثنائي المعلمة، في ظل ظروف متنوعة من طول الاختبار ومستوى القدرة ومستوى الدلالة الإحصائية، وقد أظهرت نتائج الدراسة أن معدلات الخطأ من النوع الأول للمؤشر المصحح  $I_Z^*$  كانت أقرب إلى مستويات الدلالة الاسمية من مؤشر  $I_Z$  وفي جميع مستويات القدرة، وأنه بزيادة طول الاختبار تقترب معدلات الخطأ من النوع الأول من المستويات الاسمية.

وهدفت دراسة لوبيز ومونتيسينوس (Lopez & Montesi- nos , 2005) إلى تعرف خصائص التوزيع وقوة الكشف عن أنماط الاستجابة غير المطابقة لمؤشرات ( $ECL2_Z, ECL4_Z, I_Z$ ) باستخدام بيانات مولدة وفق نموذج راش، وفي ظل ظروف متنوعة من حجم العينة ومدى الصعوبة وشكل توزيع القدرة وطول الاختبار، وقد أظهرت نتائج الدراسة أن المتوسطات الحسابية والانحرافات المعيارية لقيم مؤشر  $I_Z$  تزداد بزيادة حجم العينة، وفي جميع ظروف الدراسة، كما تزداد قوة الكشف عن أنماط الاستجابة غير المطابقة بزيادة حجم العينة، وتقل قوة الكشف عن أنماط الاستجابة للفقره بشكل كبير عندما يقل حجم العينة عن (500) فرد.

وأجرى حمادنه (2011) دراسة هدفت إلى التحقق من فاعلية مؤشر  $I_Z$  من خلال قدرته على ضبط الخطأ من النوع الأول، وقوته في الكشف عن أنماط استجابة الفرد غير المطابقة، باستخدام بيانات مولدة لاختبار مكون من (50) فقره وفق النموذج ثنائي وثلاثي المعلمة، واستخدم الباحث طريقة بيز (EAP) لتقدير قدرات الأفراد، وقد أظهرت نتائج الدراسة أن معدلات الخطأ من النوع الأول كانت أكثر قرباً من المستوى الاسمي للدلالة بزيادة حجم العينة، في كل من النموذجين ثنائي المعلمة وثلاثي المعلمة، وأن قوة الإحصائي

الاعتماد على اختبار الذيل الأيسر أحادي الاتجاه لاكتشاف أنماط الاستجابة غير المطابقة (Armstrong et al.,2017).

وعندما يطابق نمط الاستجابة النموذج، فإن قيمة مؤشر المطابقة تقترب من الصفر، بينما تعتبر القيم السالبة لمؤشر المطابقة دلالة على عدم ملاءمة نمط الاستجابة للنموذج، وتشير القيم الموجبة للمؤشر إلى أن أنماط استجابة الفرد مطابقة بدرجة أكبر مما يتنبأ به النموذج (Reise, 1990; De Ayala, 2009).

وفي هذه الدراسة يعدّ الفرد غير مطابق للنموذج إذا كانت قيمة مؤشر المطابقة المقابلة لنمط استجابته أقل من الدرجة المعيارية (-1.645) حيث تمثل هذه الدرجة القيمة الحرجة عند مستوى الدلالة ( $\alpha=0.05$ ). لاختبار الذيل الواحد باعتبار توزيع المؤشر توزيعاً طبيعياً، وقد استخدم هذا المعيار في دراسات (Sinharay, 2016; Meijer & Tenderio,2012 ; Meijer & Nering,1997; Seo & Weiss,2013; Nering,1995).

ونتيجةً للاعتماد المتزايد على مؤشرات مطابقة الفرد المبنية على الأرجحية في الكشف عن أنماط الاستجابة غير المطابقة، اتجهت العديد من الدراسات والأبحاث نحو الكشف عن الخصائص الإحصائية لهذه المؤشرات، ومدى دقتها وفعاليتها في تحديد أنماط الاستجابة غير المطابقة، فقد أجرى نونان وبوس وجيسارولي (Noonan , Boss & Gessaroli ,1992) دراسة هدفت لتقصي أثر طول الاختبار (80,40) فقره.

ونموذج الاستجابة للفقره (2PL,3PL) على خصائص توزيع مؤشر  $I_Z$  باستخدام بيانات مولدة، وقد أظهرت نتائج الدراسة أن زيادة طول الاختبار في كلا النموذجين قابلها انخفاض في قيم الالتواء والتفلطح، وكانت هذه القيم في النموذج ثلاثي المعلمة أقل منها في النموذج ثنائي المعلمة، وكان شكل توزيع مؤشر  $I_Z$  بشكل عام وفي جميع الظروف ملتو نحو اليسار وله تفلطح موجب.

وقام نيرنج (Nering,1995) بدراسة هدفت إلى الكشف عن خصائص توزيع مؤشر  $I_Z$  عند استخدام القدرة الحقيقية والقدرة للمقدرة، في ضوء متغيرات طريقة تقدير القدرة (Maximum Like-lihood, Bayesian) ومستوى التخمين ومدى صعوبة الفقرات، باستخدام بيانات مولدة ضمن النموذج ثلاثي المعلمة (3PL)، وأظهرت نتائج الدراسة أن الوسط الحسابي والانحراف المعياري لتوزيع قيم مؤشر المطابقة كانت قريبة من القيم النظرية للتوزيع الطبيعي (0,1)، عند استخدام القدرة الحقيقية، في حين انحراف توزيع مؤشر المطابقة عن التوزيع الطبيعي عند استخدام القدرة المقدرة، وأن هناك أثراً لطريقة تقدير القدرة على خصائص توزيع المؤشر؛ إذ كانت قيم الوسط الحسابي والانحراف المعياري والالتواء والتفلطح لتوزيع المؤشر أقرب إلى القيم النظرية لخصائص التوزيع الطبيعي عند استخدام طريقة (MLE) لتقدير القدرة مقارنة بطريقة بيز (method Bayesian)، وكان شكل توزيع المؤشر أقرب إلى التوزيع الطبيعي عند استخدام طريقة (MLE) اعتماداً على قيمة اختبار (Kolmogrov-Smirnov).

وفي دراسة ريس (Reise,1995) التي هدفت إلى تقصي فاعلية مؤشر  $I_Z$  في الكشف عن أنماط الاستجابة غير المطابقة عند استخدام القدرة الحقيقية والقدرة المقدرة، باستخدام بيانات

تزداد بزيادة حجم العينة في كل من النموذج ثنائي المعلمة وثلاثي المعلمة.

وهدفت دراسة ماجيس وريش وبيلاندي (Magis, Raïche, & Béland, 2012) إلى المقارنة بين شكلي التوزيع لمؤشري  $l_z$  و  $l_z^*$  عند استخدام النموذج ثنائي المعلمة (2PL) واستخدام طريقة الأرجحية العظمى الموزونة لتقدير القدرة (WLE)، باستخدام بيانات حقيقية، وأظهرت نتائج الدراسة عدم مطابقة توزيع المؤشرين للتوزيع الطبيعي بشكل تام، وأن توزيع مؤشر  $l_z^*$  كان أقل تفلطحاً من توزيع مؤشر  $l_z$ ، وأن مؤشر  $l_z^*$  لا يطابق التوزيع الطبيعي بشكل تام في حالة البيانات الحقيقية.

وأجرى ميجر وتندريو (Meijer & Tenderio, 2012) دراسة هدفت إلى الكشف عن خصائص توزيع مؤشري المطابقة  $l_z$  و  $l_z^*$  ودرجة مطابقتها للتوزيع الطبيعي المعياري باستخدام بيانات حقيقية، ولغايات تقدير المعالم استخدم الباحثان النموذج ثنائي المعلمة (2PL)، وأظهرت نتائج الدراسة أن الذيل الأيسر لتوزيع قيم مؤشر  $l_z^*$  أكثر قرباً من التوزيع الطبيعي مقارنة بمؤشر  $l_z$ ، وأن شكل توزيع المؤشر  $l_z^*$  بشكل عام أقرب للتوزيع الطبيعي المعياري من مؤشر  $l_z$ .

وقام سيو وويس (Seo & Weiss, 2013) بدراسة هدفت إلى الكشف عن الأفراد غير المطابقين لنموذج الاستجابة للفقرة ثلاثي المعلمة (3PL) باستخدام مؤشر  $l_z$ ، وباستخدام بيانات الاستجابة الحقيقية للطلبة الجامعيين على 20 اختبار تحصيلي كل منها مكون من فقرات اختيار من متعدد، استخدم فيها الباحث ثلاث طرق لتقدير القدرة (MAP, EAP), MLE. وقد أظهرت نتائج الدراسة أن توزيع المؤشر بشكل عام ملتو نحو اليسار في طرق تقدير القدرة المختلفة، وكان شكل التوزيع لقيم مؤشر المطابقة عند تقدير القدرة بطريقتي (MAP, EAP)، أكثر التواءاً نحو اليسار من شكل التوزيع عند تقدير القدرة بطريقة (MLE)، وأن نسب الأفراد غير المطابقين الذين كشف عنهم مؤشر المطابقة عند تقدير القدرة بطريقتي (MAP, EAP) أقل بشكل عام من نسب الأفراد غير المطابقين الذين كشف عنهم مؤشر المطابقة عند تقدير القدرة بطريقة (MLE)، في حين كانت النسب متقاربة لكل من طريقتي (MAP, EAP).

وأجرى كوي و موسافي (Cui and Mousavi, 2015) دراسة هدفت إلى تقصي خصائص توزيع مؤشر  $l_z$  وفاعليته في الكشف عن أنماط الاستجابة غير المطابقة، في اختبار تحصيلي عام في الرياضيات لطلبة الصف التاسع في كندا، وقد أظهرت النتائج أن توزيع مؤشر  $l_z$  ليس معيارياً بشكل تام، وأن شكل التوزيع بشكل عام ملتو نحو اليسار وله تفلطح موجب.

وهدفت دراسة سنهاري (Sinharay, 2015) إلى تقصي فاعلية مؤشري  $l_z(MIX)$  و  $\zeta_2$  في الكشف عن أنماط الاستجابة غير المطابقة في الاختبارات مختلطة الشكل، بدلالة الخطأ من النوع الأول وقوة الكشف، باستخدام بيانات مولدة وفق النموذج ثلاثي المعلمة (3PL) ونموذج التقدير الجزئي العام (GPCM)، في ظل ظروف متغيرة من طول الاختبار (60,30,12) فقرة، وقد أظهرت نتائج الدراسة أن معدلات الخطأ من النوع الأول لمؤشر  $l_z(MIX)$  أقل من مستويات الدلالة الإسمية، وأن مؤشر  $\zeta_2$  بشكل عام أكثر قوة من مؤشر  $l_z(MIX)$ ، وأظهرت النتائج أيضاً أن قوة الكشف عن

أنماط الاستجابة غير المطابقة تزداد بزيادة طول الاختبار.

وقام سنهاري (Sinharay, 2016) بدراسة هدفت إلى الكشف عن فاعلية مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الكشف عن أنماط الاستجابة غير المطابقة في الاختبارات مختلطة الشكل، باستخدام بيانات مولدة وفق نموذج التقدير الجزئي العام (GPCM) ونموذج الاستجابة للفقرة ثلاثي المعلمة (3PL)، في ضوء متغيرات طول الاختبار ومستوى الدلالة الإحصائية، وقد استخدم الباحث طريقة (MLE) لتقدير القدرة، وأظهرت النتائج أن معدلات الخطأ من النوع الأول لمؤشر  $l_z^*(MIX)$  تقترب من المستويات الإسمية لهذا الخطأ أكثر من مؤشر  $l_z(MIX)$  في جميع ظروف الدراسة، كما أظهرت نتائج الدراسة أن المؤشر  $l_z^*(MIX)$  أكثر قوة في الكشف عن الأفراد غير المطابقين للنموذج من المؤشر  $l_z(MIX)$ ، وفي جميع ظروف الدراسة، وبالنسبة لشكل توزيع المؤشران أظهرت النتائج أن توزيع المؤشران بشكل عام ملتو نحو اليسار وله تفلطح موجب، وأن توزيع مؤشر  $l_z^*(MIX)$  بشكل عام أقرب إلى التوزيع الطبيعي من توزيع مؤشر  $l_z(MIX)$ .

وأجرى سنهاري (Sinharay, 2018) دراسة هدفت إلى المقارنة بين فاعلية مؤشرات  $l_z^*(MIX)$ ،  $\zeta_1^*$ ،  $\zeta_2^*$ ،  $\zeta_1$ ،  $\zeta_2$  في الكشف عن أنماط الاستجابة غير المطابقة في الاختبارات مختلطة الشكل، بدلالة الخطأ من النوع الأول وقوة الكشف، باستخدام بيانات مولدة وفق نموذج التقدير الجزئي العام (GPCM) ونموذج الاستجابة للفقرة ثلاثي المعلمة (3PL)، في ظل ظروف متغيرة من طول الاختبار (60,30,12) فقرة ومستوى القدرة، وقد أظهرت نتائج الدراسة أن معدلات الخطأ من النوع الأول لمؤشر  $l_z^*(MIX)$  كانت الأكثر قرباً لمستويات الدلالة الإسمية عند القيم السالبة لمستويات القدرة، في حين كانت معدلات الخطأ من النوع الأول لمؤشر  $\zeta_2^*$  الأقرب لمستويات الدلالة الإسمية عند القيم الموجبة لمستويات القدرة، وأن معدلات الخطأ من النوع الأول لجميع المؤشرات تقترب من المستويات الإسمية بزيادة طول الاختبار، وأظهرت تحليلات القوة أن مؤشر  $l_z^*(MIX)$  كان أكثر قوة في الكشف عن أنماط الاستجابة غير المطابقة من مؤشرات  $\zeta_1^*$ ،  $\zeta_2^*$ ،  $\zeta_1$ ،  $\zeta_2$  عند مستويات القدرة أقل من (-1)، وأقل قوة منها حول مستوى القدرة (0) في جميع ظروف الدراسة.

وبالرغم من تنوع أهداف الدراسات السابقة واختلاف طبيعة البيانات التي تتعامل معها، إضافة إلى تنوع المتغيرات التي تناولتها، إلا أن غالبيتها اقتصر على تقصي فاعلية مؤشري  $l_z$  و  $l_z^*$  وخصائص توزيعهما في الفقرات ثنائية التدرج، وأن هناك ندرة في الأبحاث التي اهتمت بدراسة فاعليتهما في الاختبارات مختلطة الشكل، ولم تهتم سوى دراسة سنهاري (Sinharay, 2016) بالمقارنة بين فاعلية مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الاختبارات مختلطة الشكل، غير أن هذه الدراسة اقتصر على استخدام البيانات المولدة، ولم تهتم بدراسة أثر حجم العينة وطريقة تقدير القدرة، على خصائص توزيع المؤشرين، ونسب أنماط الاستجابة غير المطابقة التي يكشفان عنها في الاختبارات مختلطة الشكل بشكل عام، وعند استخدام البيانات الحقيقية بشكل خاص.

ولما أظهرت نتائج الدراسات السابقة أن خصائص توزيع مؤشري  $l_z$  و  $l_z^*$  وفاعليتهما في الكشف عن أنماط الاستجابة

خصائص توزيع المؤشر ونسب أنماط الاستجابة غير المطابقة التي يكشف عنها دليل على فاعليته، ودقته في تصنيف أنماط الاستجابة إلى مطابقة وغير مطابقة.

### أهمية الدراسة

تبرز أهمية الدراسة من خلال جانبين:

#### الأهمية النظرية :

تكتسب هذه الدراسة أهميتها النظرية، من خلال محاولتها الكشف عن أثر طريقة تقدير القدرة وحجم العينة على خصائص توزيع مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$ ، ونسب أنماط الاستجابة غير المطابقة التي يكشف عنها المؤشران في الاختبارات مختلطة الشكل؛ إذ يُعدُّ التحقق من التوزيع النظري لمؤشرات المطابقة المبنية على الأرجحية قبل البدء بإجراءات اختبار الفرضيات الخاصة بمطابقة الأفراد لنموذج الاستجابة للفقره من الأمور الهامة، فكلما اقترب توزيع المؤشر من التوزيع الطبيعي، أصبحت نتائج اختبار الفرضيات المبنية على هذا التوزيع أكثر دقة، كما أن زيادة نسب أنماط الاستجابة غير المطابقة التي يكشف عنها المؤشر تعد دليل على قوته وفاعليته، بالنظر إلى الآثار السلبية المترتبة على عدم استبعاد الأنماط غير المطابقة على صدق نتائج الاختبار، ودقة القرارات المبنية على هذه النتائج.

#### الأهمية العملية:

تبرز الأهمية العملية لهذه الدراسة، من خلال دراستها التجريبية المعتمدة على البيانات الفعلية، والتي قد تزود مطبقي الاختبارات والعاملين في مجال قياس السمات المعرفية والنفسية، على المستوى الفردي كالمعلمين وأساتذة الجامعات، أو على المستوى الوطني والمؤسسي كالعاملين في مجال الاختبارات الوطنية واختبارات انتقاء الموظفين، بمعلومات عن الظروف المتعلقة بطريقة تقدير القدرة وحجم العينة، التي تزيد من فاعلية ودقة مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الكشف عن أنماط الاستجابة غير المطابقة في الاختبارات مختلطة الشكل، بعد أن أصبحت هذه الاختبارات شائعة في التقييم المدرسي والوطني، إذ تتيح هذه المؤشرات إمكانية تعرف الأفراد الذين تظهر استجابتهم أنماطاً شاذة غير متوافقة مع ما هو متوقع منهم في ضوء قدراتهم؛ بسبب الغش أو التخمين أو اللامبالاة، التي تهدد صدق نتائج هذه الاختبارات ودقة القرارات المبنية على هذه النتائج، وبخاصة المتعلقة بالأفراد، مثل اتخاذ القرار بشأن القبول الجامعي، واختيار المتقدمين للوظيفة، وتشخيص الحالة النفسية.

#### التعريفات الاصطلاحية والإجرائية

مؤشر مطابقة الفرد (Person-Fit Index): هو مؤشر إحصائي يستخدم للحكم على درجة مطابقة نمط استجابة الفرد لنموذج الاستجابة للفقره المستخدم، وفي هذه الدراسة جرى استخدام مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  للكشف عن أنماط الاستجابة غير المطابقة، وهما من المؤشرات المبنية على الأرجحية (-likelihood based Indices).

الاختبارات مختلطة الشكل (mixed-format tests):

يرير المطابقة، تتأثر بحجم العينة كدراسة (Lopez & Montesi- nos,2005) ودراسة حمادنه (2011)، وبطريقة تقدير القدرة كدراسات (Seo & Weiss, 2013; Reise, 1995; Meijer & Nering, 1995; Nering, 1997). لذلك من المحتمل أن تؤثر هذه العوامل على خصائص توزيعات مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  ونسب أنماط الاستجابة غير المطابقة التي يكشفان عنها في الاختبارات مختلطة الشكل، باعتبارهما صور مطورة من مؤشري  $l_z$  و  $l_z^*$  للاختبارات مختلطة الشكل، وهذا ما تسعى هذه الدراسة لاستكشافه.

### مشكلة الدراسة وأسئلتها

تعد مطابقة نمط استجابة الفرد لنموذج الاستجابة للفقره المستخدم من القضايا الهامة لتحديد فيما إذا كان الفرد يستجيب للفقره وفقاً للبنية الأساسية التي يجري قياسها، أو ما إذا كانت هناك عوامل أخرى تؤثر على سلوك الاستجابة، إذ يتعرض الأفراد ذو أنماط الاستجابة غير المطابقة لخطر تلقي نتائج اختبارات غير دقيقة، قد تؤدي إلى اتخاذ قرارات غير عادلة أو غير ملائمة بناءً على درجاتهم على تلك الاختبارات.

كما تؤدي الاستجابات غير المطابقة في مصفوفة البيانات إلى تأثير سلبي على معايرة الفقرات، ويمكن أن تؤثر على تقديرات القدرة، وعندما لا يطابق الفرد النموذج، فإن تقدير السمة لا يمكن تفسيره بسهولة، ويصبح أي استخدام لاحق للدرجات المبنية على النموذج غير صادق (Meijer, 1996).

وبسبب زيادة الاعتماد على الاختبارات ذات أشكال الفقرات المختلطة في القياس التربوي، وقلّة الدراسات التي اهتمت بدراسة فاعلية مؤشري مطابقة الفرد  $l_z(MIX)$  و  $l_z^*(MIX)$  في الاختبارات مختلطة الشكل بشكل عام، وفي حالة البيانات الحقيقية بشكل خاص، واستجابة لتوصيات دراسات (Conijn, Emons, & Sijtsma, 2014; Meijer, Niessen, & Tendeiro, 2015; Meijer & Tendeiro, 2012) بضرورة إجراء مزيد من الدراسات حول سلوك مؤشرات المطابقة في البيانات الحقيقية، إذ من المحتمل أن تؤثر خصائص الاختبار وخصائص المفحوصين على توزيع مؤشر المطابقة عند استخدام البيانات الحقيقية. جاءت هذه الدراسة للكشف عن أثر طريقة تقدير القدرة وحجم العينة على فاعلية مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الاختبارات مختلطة الشكل من خلال الإجابة عن الأسئلة التالية:

1. هل تختلف خصائص توزيعات مؤشري المطابقة وتبعاً لحجم العينة (5000، 2000، 1000) وطريقة  $l_z(MIX)$  و  $l_z^*(MIX)$  تقدير القدرة (WLE، MLE، MAP)؟

2. هل تختلف نسبة أنماط الاستجابة غير المطابقة التي يكشف عنها مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  تبعاً لحجم العينة (5000، 2000، 1000) وطريقة تقدير القدرة (WLE، MLE، MAP)،؟

### أهداف الدراسة

تهدف هذه الدراسة إلى الكشف عن أثر طريقة تقدير القدرة وحجم العينة، على خصائص توزيع مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  ونسب أنماط الاستجابة غير المطابقة التي يكشفان عنها في الاختبارات مختلطة الشكل، باستخدام بيانات فعلية، باعتبار



الثامن في هذه الدول، والذين قاموا بالاستجابة على فقرات الرياضيات في الكراسة الاختبارية رقم (1)، والبالغ عددهم (17166) طالباً وطالبة.

### أداة الدراسة وبياناتها

اعتمدت هذه الدراسة على البيانات التي توفرها استجابات طلبة الصف الثامن على فقرات الرياضيات في اختبار (TIMSS) دورة العام 2015، واختبار (TIMSS) هو دراسة دولية في العلوم والرياضيات، تنفذ على طلبة الصفين الرابع والثامن بشكل دوري كل أربع سنوات منذ عام 1995، تحت إشراف الرابطة الدولية لتقييم التحصيل التربوي (IEA)، واختبار العام 2015 هو الاختبار السادس في هذه السلسلة.

ويهدف اختبار (TIMSS) إلى المقارنة بين مستويات تحصيل طلبة مجموعة من دول العالم في الرياضيات والعلوم، ودراسة العوامل التي تؤثر في تحصيلهم في هاتين المادتين وتحسين التعلم والتعليم فيهما، من خلال توفير بيانات عن تحصيل الطلبة في أنظمة تربوية متباينة، وفي ظل ممارسات تدريسية وبيئات مدرسية مختلفة.

وتوزعت أسئلة الرياضيات والعلوم في اختبار (TIMSS) لعام 2015 على (14) كراسة اختبارية، تتكون كل كراسة من عنقودين في الرياضيات وعنقودين في العلوم، ويتكون العنقود من مجموعة صغيرة من الأسئلة، حيث يظهر كل سؤال من أسئلة الاختبار في عنقود واحد فقط، ويظهر كل عنقود في كراستين، ويعطى الطالب في جلسة الاختبار كراسة واحدة بصورة عشوائية، يجب عن أسئلتها في فترة زمنية تبلغ 90 دقيقة موزعة على جلستين تفصلهما استراحة لمدة 10 دقائق، ويبين الجدول (1) توزيع عناقيد الأسئلة على كراسات الاختبار.

الجدول (1)

توزيع عناقيد الأسئلة على كراسات الاختبار (الصف الثامن)

رقم الكراسة	الجزء الأول	الجزء الثاني	الجزء الثالث	الجزء الرابع
1	M01	M02	S01	S02
2	S02	S03	M02	M03
3	M03	M04	S03	S04
4	S04	S05	M04	M05
5	M05	M06	S05	S06
6	S06	S07	M06	M07
7	M07	M08	S07	S08
8	S08	S09	M08	M09
9	M09	M10	S09	S10
10	S10	S11	M10	M11
11	M11	M12	S11	S12
12	S12	S13	M12	M13
13	M13	M14	S13	S14

هي الاختبارات التي تتنوع فيها أشكال الفقرات تبعاً لطريقة الاستجابة عليها وطريقة تدرج هذه الاستجابة، وتشمل: الفقرات ثنائية التدرج (dichotomous items) التي يكون فيها تدرجان فقط للاستجابة (0,1)، مثل فقرات الاختيار من متعدد وفقرات الإجابة (بنعم أو لا) وفقرات المزاج والمطابقة، والفقرات متعددة التدرج (polytomous items)، التي يكون فيها أكثر من تدرجان للاستجابة على الفقرة، وتتنوع المهام في هذا الشكل من الفقرات بين ملء الفراغات، وإكمال الجمل، والإجابات القصيرة، إلى كتابة مقالات متعددة الصفحات، أو حل متعدد الخطوات لمشكلة أو مسألة رياضية، وفقرات مقاييس الاتجاهات والشخصية والمويل التي تتضمن تدرج يقابل كل فقرة، وقد تكونت أداة الدراسة من مزيج من فقرات الاختيار من متعدد (ثنائية التدرج)، وفقرات الإجابة المصاغة التي تتطلب خطوات حل مرتبة (متعددة التدرج).

نمط الاستجابة غير المطابق (misfitting response pattern): هو نمط الاستجابة الذي ينحرف عن النمط المتوقع من الشخص اعتماداً على قدرته، وعلى خصائص وافتراضات نموذج الاستجابة للفقرة المستخدم، وفي هذه الدراسة جرى استخدام نموذج الاستجابة للفقرة ثلاثي المعلمة (3PL) ونموذج التقدير الجزئي العام (GPCM)، ويعتبر الشخص غير مطابق إذا كانت قيمة مؤشر المطابقة لنمط استجابته أقل من القيمة المعيارية (-1.645).

### حدود الدراسة ومحدداتها

اقتصرت هذه الدراسة على:

1. استخدام بيانات استجابات طلبة الصف الثامن على فقرات الرياضيات في الكراسة الاختبارية الأولى في اختبار (TIMSS 2015)، للإجابة عن أسئلة الدراسة الحالية.
2. استخدام ثلاث طرق لتقدير القدرة الأرجحية العظمى (MLE)، وطريقة الأرجحية الموزونة (WLE)، وطريقة التقدير البعدي الأعظم (MAP).
3. استخدام مؤشري  $I_z(MIX)$  و  $I_z^*(MIX)$  مطابقة الفرد وللكشف عن أنماط الاستجابة غير المطابقة.

### الطريقة والإجراءات

#### منهجية الدراسة

استخدمت الدراسة المنهج الوصفي التحليلي، إذ جرى استخدام استجابات الطلبة على فقرات الرياضيات في الكراسة الاختبارية الأولى في اختبار (TIMSS 2015) للإجابة على أسئلة الدراسة، والمتمثلة بخصائص توزيع مؤشري المطابقة، ونسب أنماط الاستجابة غير المطابقة.

#### مجتمع الدراسة وعينتها

تكون مجتمع الدراسة من جميع طلبة الصف الثامن في الدول المشاركة في اختبار (TIMSS) لعام 2015، والبالغ عددها (39) دولة منها (10) دول عربية هي: الأردن، ومصر، ولبنان، والكويت، والمغرب، وقطر، والبحرين، وعمان، والسعودية، والإمارات.

وللإجابة عن أسئلة الدراسة، اختار الباحثان عينات حجمها (5000,2000,1000) طالباً، بطريقة عشوائية من طلبة الصف

### الجدول (3)

توزيع فقرات الإجابة المصاغة (CR) في الكراسة الأولى حسب موضوع الفقرة والمجال الذي تنتمي إليه

رمز الفقرة	موضوع الفقرة	مجال الفقرة
M062111A	الأعداد	التطبيق
M062111B	الأعداد	التطبيق
M062237	الجبر	التطبيق
M062314	الجبر	المعرفة
M042093	الجبر	التطبيق
M042159	البيانات والاحتمالات	التطبيق
M042164	البيانات والاحتمالات	التعليل
M042167	البيانات والاحتمالات	التعليل
M062208	الأعداد	المعرفة
M042302A	الأعداد	التطبيق
M042302B	الأعداد	التطبيق
M042302C	الأعداد	التعليل
M042081	الأعداد	المعرفة
M062183	الهندسة	التطبيق
M062286	الهندسة	التطبيق

### الجدول (4)

توزيع فقرات الاختيار من متعدد (MC) في الكراسة الأولى حسب موضوع الفقرة والمجال الذي تنتمي إليه

رمز الفقرة	موضوع الفقرة	مجال الفقرة
M042182	الأعداد	التطبيق
M042052	الأعداد	المعرفة
M062208A	الأعداد	المعرفة
M062208B	الأعداد	المعرفة
M062208C	الأعداد	المعرفة
M062208D	الأعداد	المعرفة
M062153	الأعداد	التطبيق
M042049	الجبر	المعرفة
M042076	الجبر	المعرفة
M062074	الجبر	التطبيق
M042100	الجبر	المعرفة
M042202	الجبر	التطبيق
M042240	الجبر	التطبيق
M062325	البيانات والاحتمالات	المعرفة

رقم الكراسة	الجزء الأول	الجزء الثاني
14	S01 S14	M01 M14

M : عنقود رياضيات S : عنقود علوم

وبلغ عدد أسئلة الرياضيات في اختبار (TIMSS) لعام 2015 للصف الثامن (212) سؤالاً توزعت على موضوعات: الأعداد والجبر والهندسة والبيانات، وغطت مجالات: المعرفة والتطبيق والتعليل، وتنوعت فيها أشكال الأسئلة بين الاختيار من متعدد (MC) بنسبة (51%) من فقرات الاختبار، والأسئلة ذات الإجابة المصاغة (CR) التي تتطلب من المفحوص أن يصوغ الإجابة بنسبة بلغت (49%) من أسئلة الاختبار ويبين الجدول (2) توزيع أسئلة الرياضيات بحسب موضوع ومجال ونوع السؤال.

### الجدول (2)

توزيع أسئلة اختبار الرياضيات حسب موضوع ومجال ونوع السؤال

البعد	المحتوى	عدد الأسئلة من نوع الاختيار من متعدد	عدد الأسئلة من نوع بناء الإجابة (المقالي)	مجموع الأسئلة	العلامات %
البيانات	الأعداد	29(29)	35(41)	(70)64	30.6%
	الجبر	35(35)	27(30)	(65)62	28.4%
	الهندسة	22(22)	21(25)	(47)43	20.5%
	البيانات	29(31)	14(16)	(47)43	20.5%
	المجموع	115(117)	97(112)	212(229)	100%
معرفة	النسبة المئوية للعلامات	51%	49%		
	معرفة	50(50)	19(20)	(70)69	30.5%
	تطبيق	48(48)	47(55)	(103)95	45%
التعليل	التعليل	17(19)	31(37)	(56)48	24.5%
	المجموع	115(117)	97(112)	212(229)	100%
	النسبة المئوية للعلامات	51%	49%		

مجموع العلامات يظهر بين قوسين

وقد اقتصر هذه الدراسة على بيانات استجابة طلبة الصف الثامن على فقرات الرياضيات في الكراسة الاختبارية الأولى والتي تكونت من (35) فقرة، منها (20) فقرة اختيار من متعدد (MC) بنسبة (57%)، و(15) فقرة تتطلب من المفحوص أن يصوغ الإجابة (CR) بنسبة (43%) من مجموع فقرات الرياضيات في الكراسة الأولى، منها (11) فقرة يتطلب حلها خطوة واحدة فقط و(4) فقرات يتطلب حلها خطوتين متابعتين، ويبين الجدولان (4.3) توزيع فقرات الرياضيات في الكراسة الأولى على الموضوعات والمجالات التي يقيسها الاختبار.

العظمى (MLE)، الأرجحية الموزونة (WLE)، التقدير البعدي الأعظم (MAP)

2. حجم العينة وله ثلاث فئات هي: 5000 طالب، 2000 طالب، 1000 طالب.

### المتغيرات التابعة

1. نسب أنماط الاستجابة غير المطابقة لمؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$ .

2. خصائص توزيع مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  ممثلة بقيم الوسط الحسابي والانحراف المعياري والالتواء والتفلطح.

### إجراءات الدراسة

للإجابة عن أسئلة الدراسة، جر تنفيذ الخطوات التالية:

1. الحصول على ملف استجابات طلبة الصف الثامن على فقرات الرياضيات لاختبار (TIMSS 2015) من موقع الاختبار على الإنترنت (TIMSS & PIRLS International Study Center, 2015)، ثم حصر استجابات الطلبة على الفقرات التي تنتمي للكراسة الاختبارية الأولى، وتخزينها في ملف على برنامج (SPSS V23).

2. اختيار ثلاث عينات عشوائية بحجم (5000، 2000، 1000) طالباً من الطلبة الذين استجابوا على فقرات الرياضيات في الكراسة الأولى، باستخدام أمر الاختيار العشوائي الموجود في برنامج (SPSS V23).

3. تحويل الاستجابات على الفقرات ثنائية التدرج (MC) إلى الترميز (0) في حالة الإجابة الخاطئة والترميز (1) في حالة الإجابة الصحيحة، وتحويل الاستجابات على الفقرات ذات الإجابة المصاغة (CR) التي تتطلب من الطالب صياغة الاستجابة إلى الترميز (0) في حالة الإجابة الخاطئة، والترميز (1) في حالة الإجابة على الخطوة الأولى إجابة صحيحة، والترميز (2) في حالة الإجابة على الخطوة الثانية إجابة صحيحة، استناداً إلى مفتاح التصحيح الخاص بالكراسة الاختبارية الأولى، باستخدام أوامر البرنامج الإحصائي (SPSS V23)

4. دلالات صدق وثبات أداة الدراسة:

جرى التحقق من ثبات أداة الدراسة من خلال حساب معامل كرونباخ ألفا، كمؤشر على ثبات الاتساق الداخلي لفقرات أداة الدراسة (فقرات الرياضيات في الكراسة الاختبارية الأولى)، في عينات الدراسة الثلاث، ويبين الجدول (5) قيم معاملات ثبات كرونباخ ألفا في عينات الدراسة الثلاث.

الجدول (5)

قيم معامل ثبات كرونباخ ألفا في عينات الدراسة الثلاث

حجم العينة	معامل كرونباخ ألفا
1000	.890
2000	.882
5000	.892

يتضح من الجدول (5) أن أداة الدراسة تتمتع بدرجة مرتفعة

رمز الفقرة	موضوع الفقرة	مجال الفقرة
M062106	البيانات والاحتمالات	المعرفة
M062124	البيانات والاحتمالات	المعرفة
M062202	الهندسة	التعليل
M062246	الهندسة	التعليل
M042271	الهندسة	التطبيق
M042268	الهندسة	التعليل

ومن الأمثلة على فقرات الرياضيات في الكراسة الاختبارية الأولى (المركز الوطني لتنمية الموارد البشرية، 2017):

مثال(1): فقرة رياضيات ثنائية التدرج (MC) لطلبة الصف الثامن في اختبار TIMSS 2015 وصفت جمانة الرسم البياني لاقتران:

- الرسم البياني هو خط مستقيم.
  - يقطع الرسم البياني المحور ص في ٣.
- ما الاقتران الذي يمثله الرسم البياني لجمانة؟

- ① ص = س + ٢ + ٣
- ② ص = س + ٣ + ١
- ③ ص = س + ٣ - ١
- ص = س + ٣

مثال(2): فقرة رياضيات متعددة التدرج (CR) لطلبة الصف الثامن في اختبار TIMSS 2015

مجموع أطوال أضلاع هذا المثلث يساوي ٣٠ سم.

أ. لكتب معادلة يمكنك من إيجاد قيمة (س).

المعادلة:  $30 = 10 + 20$

ب. ما طول الضلع الأطول في هذا المثلث بالسنتيمترات ؟

الإجابة: ٦٠ سم

### متغيرات الدراسة

#### المتغيرات المستقلة

1. طريقة تقدير القدرة ولها ثلاث فئات هي: الأرجحية

الأول في عينات الدراسة المختلفة كانت أعلى من (20%)، كما كان ناتج قسمة قيمة الجذر الكامن للمكون الأول على قيمة الجذر الكامن للمكون الثاني في عينات الدراسة المختلفة أعلى من (2)، مما يشير إلى وجود عامل واحد سائد في فقرات الرياضيات في الكراسة الأولى، وبالتالي تتحقق أحادية البعد استناداً إلى معيار هاتي(Hattie,1985) في الحكم على أحادية البعد

#### 6. التحقق من افتراض الاستقلال الموضوعي:

جرى التحقق من افتراض الاستقلال الموضوعي لفقرات الرياضيات في كراسة الاختبار الأولى، باستخدام البرنامج للإحصائي (IRTPRO V.4.2) وباستخدام اختبار (local depen- dence) ، ويعتبر تقاطع زوج الفقرات ذو تبعية موضعية إذا تخطت قيمة القيمة (5) أو ساوتها، وإذا كانت قيمتها أقل من تلك القيمة، فمعنى ذلك تحقيق تقاطع زوج الفقرات للاستقلال الموضوعي، ويبين الجدول (8) نتائج اختبار الاستقلال الموضوعي لفقرات الرياضيات في الكراسة الأولى.

الجدول (8)

نتائج اختبار الاستقلال الموضوعي لفقرات الكراسة الاختبارية الأولى في عينات الدراسة الثلاث

حجم العينة	حالة الاستقلال الموضوعي					
	تكرار تقاطع أزواج الفقرات	النسبة المئوية %	تكرار تقاطع أزواج الفقرات	النسبة المئوية %	تبعية موضعية	
					تكرار تقاطع أزواج الفقرات	النسبة المئوية %
1000	14	2.352	581	97.648	595	100
2000	21	3.529	574	96.471	595	100
5000	18	3.025	577	96.975	595	100

يلاحظ من الجدول (8) أن جميع قيم النسب المئوية لتكرارات تقاطع أزواج الفقرات المحققة للتبعية الموضوعية لم تزد عن (3.529%) في عينات الدراسة المختلفة، ما يعني أن جميع قيم النسب المئوية لتكرارات تقاطع أزواج الفقرات المحققة للاستقلال الموضوعي قد زادت عن (96.471%)، وبهذا يكون الاستقلال الموضوعي متحققاً في فقرات الكراسة الاختبارية الأولى وفي عينات الدراسة المختلفة.

#### 7. التحقق من افتراض (منحنى خصائص الفقرة ICC):

للتحقق من افتراض اطراد السمة (منحنى خصائص الفقرة ICC) وفقاً لنظرية استجابة الفقرة في النموذج ثلاثي المعلمة ونموذج التقدير الجزئي العام، فقد جرى رسم منحنى خصائص الفقرة (ICC) لكل فقرة من فقرات الاختبار في عينات الدراسة الثلاث، وتوضح الأشكال (6,5,4,3,2,1) منحنيات خصائص بعض الفقرات في عينات الدراسة الثلاث.

من ثبات الاتساق الداخلي، في العينات المختلفة للدراسة.

كما جرى التحقق من صدق أداة الدراسة من خلال حساب معاملات الارتباط بين الدرجة والفقرة والدرجة الكلية على أداة الدراسة، ويبين الجدول (6) قيم معاملات ارتباط الدرجة على الفقرة مع الدرجة الكلية على أداة الدراسة في عينات الدراسة الثلاث.

الجدول (6)

معاملات ارتباط الدرجة على الفقرة مع الدرجة الكلية على أداة الدراسة في عينات الدراسة الثلاث

حجم العينة	أقل قيمة لمعامل ارتباط الدرجة على الفقرة مع الدرجة الكلية	أعلى قيمة لمعامل ارتباط الدرجة على الفقرة مع الدرجة الكلية	المتوسط الحسابي لقيم معامل ارتباط الدرجة على الفقرة مع الدرجة الكلية
1000	*.394	.764	*.582
2000	*.435	.751	*.554
5000	*.469	.793	*.581

\*دالة إحصائية عند مستوى الدلالة (0.05).

يلاحظ من الجدول (6) أن قيم معاملات الارتباط كانت ذات دلالة إحصائية عند مستوى الدلالة (0.05)، ويعد ذلك دليلاً على صدق أداة الدراسة وتجانس فقراتها، ويعني ذلك وجود انسجام كبير بين ما تقيسه الفقرة وما تقيسه الأداة بشكل عام.

#### 5. التحقق من افتراض أحادية البعد:

جرى التحقق من افتراض أحادية البعد لفقرات الرياضيات في الكراسة الأولى، باستخدام طريقة المكونات الأساسية للتحليل العاملي الاستكشافي في برنامج (SPSS V23)، ويبين الجدول (7) قيم الجذور الكامنة ونسبة التباين المفسر للعامل الأول والثاني، وناتج قسمة قيمة الجذر الكامن للعامل الأول على قيمة الجذر الكامن للعامل الثاني لفقرات الرياضيات في الكراسة الاختبارية الأولى لعينات الدراسة الثلاث، التي استجابت على فقرات الرياضيات في الكراسة الاختبارية الأولى.

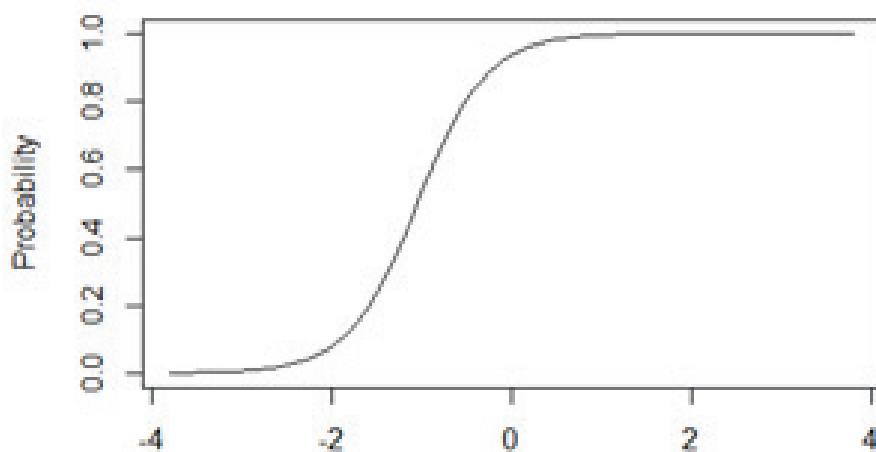
الجدول (7)

نتائج التحليل العاملي الاستكشافي لفقرات الرياضيات في الكراسة الأولى لعينات الدراسة الثلاث

حجم العينة	المكون	قيمة الجذر الكامن	التباين المفسر %	ناتج قسمة قيمة الجذر الكامن للمكون الأول على قيمة الجذر الكامن للمكون الثاني
1000	الأول	7.956	22.732	2.785
	الثاني	2.856	8.160	
2000	الأول	8.094	23.125	2.923
	الثاني	2.769	7.910	
5000	الأول	8.669	24.768	3.292
	الثاني	2.633	7.524	

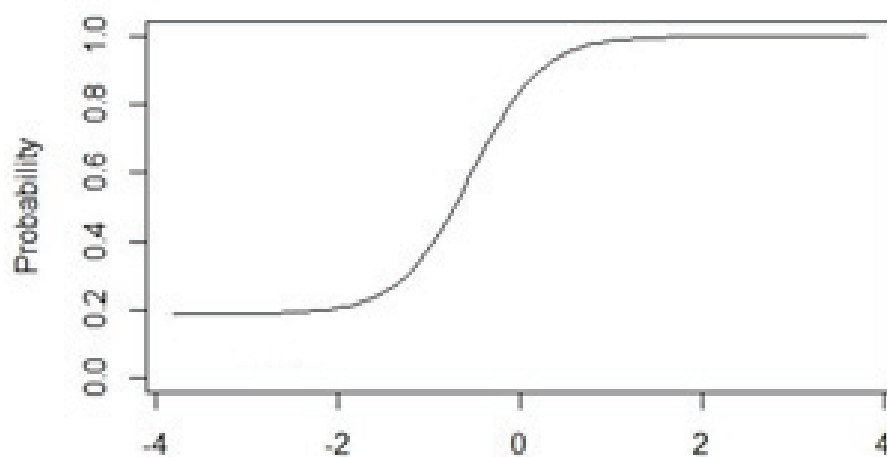
يلاحظ من الجدول (7) أن نسبة التباين التي فسرها المكون





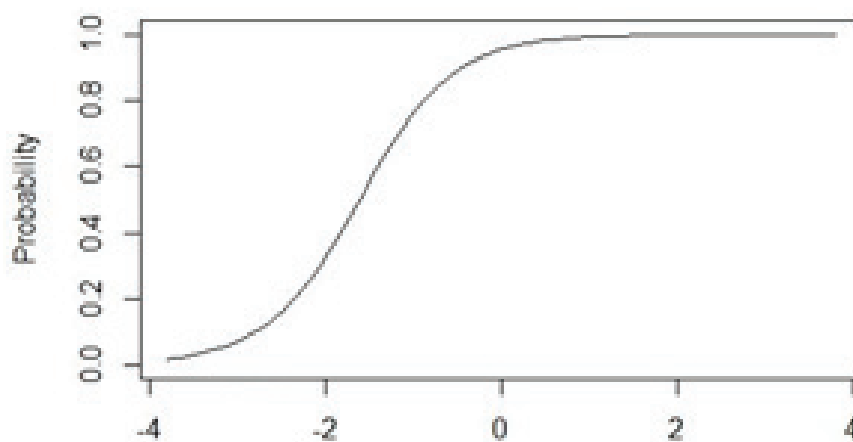
(الشكل 1)

منحنى خصائص فقرة ثنائية التدرج في العينة الأولى (1000) طالب



(الشكل 2)

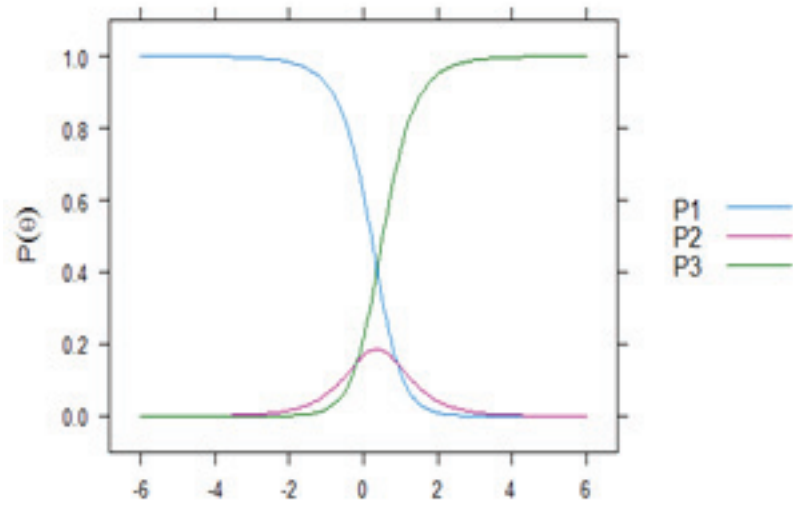
منحنى خصائص فقرة ثنائية التدرج في العينة الثانية (2000) طالب



(الشكل 3)

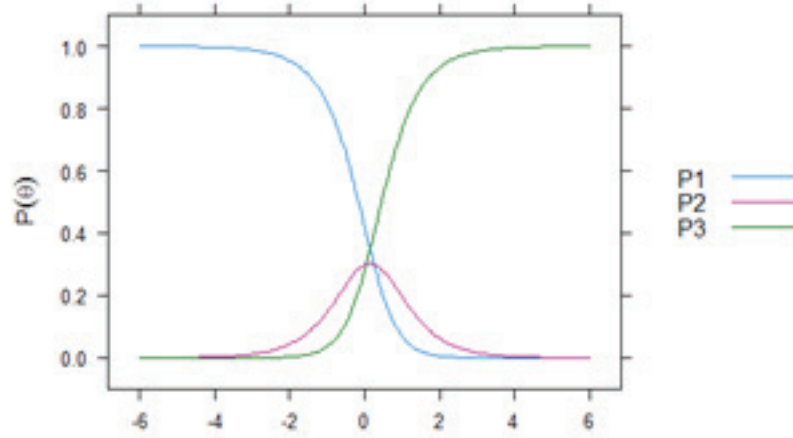
منحنى خصائص فقرة ثنائية التدرج في العينة الثالثة (5000) طالب

يلاحظ من الأشكال (1,2,3) أن جميع الفقرات تحقق افتراض اطراد السمة المقاسة، على مستوى كل فقرة من الفقرات ثنائية التدرج، في عينات الدراسة الثلاث.



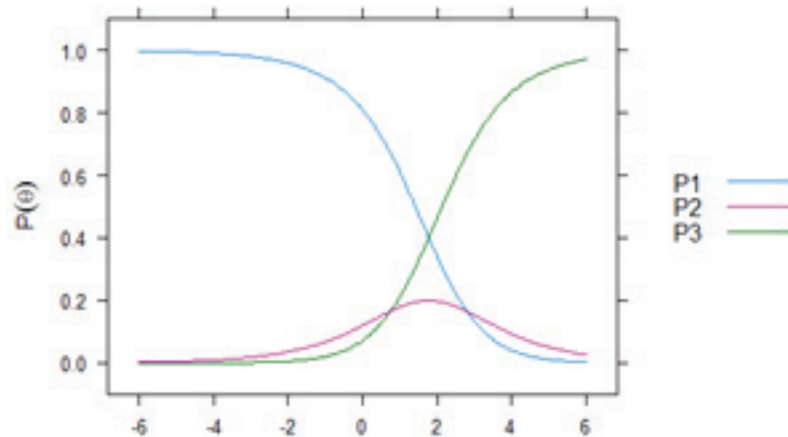
الشكل (4)

منحنى خصائص فقرة متعددة التدرج في العينة الأولى (1000) طالب .



الشكل (5)

منحنى خصائص فقرة متعددة التدرج في العينة الثانية (2000) طالب



الشكل (6)

منحنى خصائص فقرة متعددة التدرج في العينة الثالثة (5000) طالب

يلاحظ من الأشكال (6,5,4) والتي تمثل عينات من الفقرات متعددة التدرج أنها تحقق افتراض اطراد السمة المقاسة.

على حساب قيمة واحدة لمؤشر مطابقة الفرد، عند استجابته على فقرات اختبار مختلط الشكل، يتضمن فقرات ثنائية التدرج وفقرات متعددة التدرج، في ضوء إجابته على هذه الفقرات، اعتماداً على المعادلات (1) و(2) المتضمنة في هذه الحزمة، والتي جاءت بصيغة عامة لتناسب الفقرات المبنية وفق نموذج الاستجابة للفقرة ثلاثي المعلمة (3PL) ونموذج التقدير الجزئي العام (GPCM)، كما تمتاز هذه الحزمة بإمكانية استخدام طرق متعددة لتقدير قدرات الأفراد (MLE, MAP, WLE)، وقد جرى تنفيذ الخطوات التالية للحصول على إحصائيات مؤشري المطابقة:

- تشغيل برمجية (R).
- استدعاء الأوامر الخاصة بحزمة (lzstarMix v1).
- تزويد البرمجية بملف معالم الفقرات التي جرى تقديرها من استجابات العينة الأولى المكونة من (1000) طالب، باستخدام حزمة (ltm)، وملف استجابات الطلبة على الفقرات في هذه العينة، ثم إصدار الأوامر لتقدير إحصائيات مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  باستخدام طريقة (MLE) لتقدير قدرات المفحوصين، ثم إعادة الخطوة باستخدام طريقتي (WLE) و (MAP) لتقدير القدرة.

- تزويد البرمجية بملف معالم الفقرات التي جرى تقديرها من استجابات العينة الثانية المكونة من (2000) طالب، وملف استجابات الطلبة على الفقرات في هذه العينة، ثم إصدار الأوامر لتقدير إحصائيات مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  باستخدام طريقة (MLE) لتقدير قدرات المفحوصين، ثم إعادة الخطوة باستخدام طريقتي (WLE) و (MAP) لتقدير القدرة.

- تزويد البرمجية بملف معالم الفقرات التي جرى تقديرها من استجابات العينة الثالثة المكونة من (5000) طالب، وملف استجابات الطلبة على الفقرات في هذه العينة، ثم إصدار الأوامر لتقدير إحصائيات مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  باستخدام طريقة (MLE) لتقدير قدرات المفحوصين، ثم إعادة الخطوة باستخدام طريقتي (WLE) و (MAP) لتقدير القدرة.

12. إدخال ملفات إحصائيات مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  والخاصة بكل طرف من ظروف الدراسة (حجم العينة وطريقة تقدير القدرة) إلى البرنامج الإحصائي (SPSS V23) ودمجها في ملف واحد، ثم جرى تصنيف أفراد الدراسة إلى مطابقين وغير مطابقين استناداً إلى درجة القطع المستخدمة في الدراسة، حيث يعتبر الفرد غير مطابق إذا كانت قيمة مؤشر المطابقة المقابل لنمط استجابته أقل من (-1645).

## المعالجات الإحصائية

1. للإجابة عن سؤال الدراسة الأول، جرى حساب الأوساط الحسابية والانحرافات المعيارية والالتواء والتفلطح لقيم مؤشري المطابقة، التي أفرزتها برمجية (lzstarMix)، كما استخدم اختبار (Kolmogorov-Smirnov) للحكم على مدى مطابقة شكل توزيع قيم مؤشري المطابقة للتوزيع الطبيعي المعياري، في ضوء ظروف الدراسة المختلفة، باستخدام برنامج (SPSS V23)..

2. للإجابة عن سؤال الدراسة الثاني، جرى حساب التكرارات

## 8. التحقق من افتراض التحرر من السرعة:

تم التحقق من افتراض التحرر من السرعة، من خلال تحليل استجابات أفراد عينات الدراسة على فقرات أداة الدراسة، حيث أجاب غالبية الطلبة على جميع الفقرات، ما يعني أن عامل السرعة لم يلعب دوراً في الإجابة عن فقرات أداة الدراسة، أي أن إخفاق المفحوصين في الإجابة عن فقرات أداة الدراسة يرجع إلى انخفاض قدرتهم وليس إلى عدم تمكنهم من الوصول إلى الفقرات، بسبب عدم كفاية الوقت لمحاولة الإجابة عنها.

## 9. التحقق من مطابقة البيانات للنموذج:

للتحقق من مطابقة البيانات للنموذج جرى استخدام الحزمة الإحصائية (ltm v.1.1-1 package) التي تعمل في بيئة (R) لتحليل البيانات، وجرى الاعتماد على المؤشر الإحصائي مربع كاي ( $\chi^2$ ) الذي تعتمده الحزمة الإحصائية في الحكم على مطابقة الفقرات للنموذج، إذ الفقرة غير مطابقة للنموذج إذا كانت قيمة مربع كاي ( $\chi^2$ ) دالة إحصائية، ويوضح الجدول (9) نتائج اختبار مطابقة الفقرات لنموذجي الاستجابة للفقرة المستخدمان في الدراسة، ثلاثي المعلمة (3PL) ونموذج التقدير الجزئي العام (GPCM)، في عينات الدراسة الثلاث.

الجدول (9)

نتائج اختبار مطابقة الفقرات لنموذج الاستجابة للفقرة ثلاثي المعلمة (3PL) ونموذج التقدير الجزئي العام (GPCM) في عينات الدراسة الثلاث

حجم العينة	النموذج ثلاثي المعلمة (3PL)		نموذج التقدير الجزئي العام (GPCM)	
	عدد الفقرات المطابقة	عدد الفقرات غير المطابقة	عدد الفقرات المطابقة	عدد الفقرات غير المطابقة
1000	20	-	15	-
2000	20	-	15	-
5000	20	-	15	-

يتضح من الجدول (9) مطابقة جميع فقرات الرياضيات ثنائية التدرج (الاختبار من متعدد) في الكراسة الاختبارية الأولى من اختبار (TIMMS 2015) والبالغ عددها (20) فقرة لنموذج الاستجابة للفقرة ثلاثي المعلمة (3PL) في عينات الدراسة الثلاث، كما يتضح أيضاً من الجدول مطابقة جميع الفقرات متعددة التدرج (الإجابة المصاغة) والبالغ عددها (15) فقرة لنموذج التقدير الجزئي العام (GPCM)، في عينات الدراسة الثلاث.

10. تقدير معالم فقرات الرياضيات في الكراسة الاختبارية الأولى من استجابات الطلبة عليها في عينات الدراسة الثلاث (2000، 5000) باستخدام الحزمة الإحصائية (ltm v.1.1-1) التي طورها رايزوبولس (Rizopoulos, 2006) وفق النموذج ثلاثي المعلمة (3PL) للفقرات ثنائية التدرج (الاختبار من متعدد) ونموذج التقدير الجزئي العام (GPCM) للفقرات متعددة التدرج (الإجابة المصاغة)

11. حساب قيم مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  باستخدام الحزمة الإحصائية (lzstarMix v.1) التي طورها سنهاري (Sanharay, 2016)، وتعمل هذه الحزمة في بيئة (R) وتتمايز بقدرتها

تبعاً لطريقة تقدير القدرة (MAP,WLE ,MLE) وحجم العينة (5000,2000,1000)؛

للإجابة عن هذا السؤال؛ تم حساب الإحصاءات الوصفية لقيم مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  وتبعاً لعينات الدراسة الثلاث (5000,2000,1000) ولطريقة تقدير القدرة (MAP,WLE,MLE)، كما جرى استخدام اختبار كولمجروف - سميرنوف (Kolmogorov Smirnov) للتحقق من مدى مطابقة توزيعات المؤشرين للتوزيع الطبيعي المعياري، كما هو مبين في الجدول (10).

والنسب المئوية لأنماط الاستجابة المطابقة وغير المطابقة، كما تم استخدام اختبار مربع كاي  $\chi^2$  واختبار كوكران (Cochran's Q test) للتحقق من دلالة الفروق الظاهرية بين النسب، باستخدام برنامج (SPSS V23).

## نتائج الدراسة

النتائج الخاصة بسؤال الدراسة الأول الذي نصّ على: «هل تختلف خصائص توزيعات مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$ »

الجدول (10)

الأوساط الحسابية والانحرافات المعيارية والانحرافات والتفطحات لقيم مؤشري المطابقة وقيم اختبار (K-S) (Kolmogorov - Smirnov) ومستوى دلالة في عينات الدراسة الثلاث

حجم العينة	مؤشر المطابقة	طريقة تقدير القدرة	الوسط الحسابي	الانحراف المعياري	الالتواء	التفطح	قيمة اختبار K-S	الدلالة الاحصائية لاختبار K-S
1000	$l_z(MIX)$	WLE	.0193	1.3398	-1.2832	2.3889	4.9504	.0000
		MLE	.0765	1.3425	-1.3343	2.5127	5.2962	.0000
		MAP	.1359	1.3460	-1.3723	2.5977	5.7471	.0000
2000	$l_z^*(MIX)$	WLE	.0186	1.5004	-1.2258	1.9687	4.8832	.0000
		MLE	.0188	1.5031	-1.2334	2.0142	4.9774	.0000
		MAP	.1782	1.5085	-1.2904	2.1875	5.3491	.0000
5000	$l_z(MIX)$	WLE	.0351	1.2167	-1.1015	1.9963	2.9931	.0000
		MLE	.0885	1.2176	-1.1475	2.1162	3.0851	.0000
		MAP	.1575	1.2198	-1.2033	2.2349	3.3712	.0000
5000	$l_z^*(MIX)$	WLE	.0368	1.3567	-1.0187	1.4440	2.8259	.0000
		MLE	.0368	1.3580	-1.0223	1.4791	2.8611	.0000
		MAP	.2016	1.3657	-1.0754	1.6683	2.9211	.0000
5000	$l_z(MIX)$	WLE	.0560	1.2091	9750.-	1.6219	2.7996	.0000
		MLE	.1146	1.2110	-1.0186	1.7111	3.0492	.0000
		MAP	.1824	1.2150	-1.0674	1.7954	3.1714	.0000
5000	$l_z^*(MIX)$	WLE	.0618	1.3582	-9168	1.2040	2.7024	.0000
		MLE	.0616	1.3630	-9185	1.2331	2.8258	.0000
		MAP	.2308	1.3677	-9619	1.3709	2.8680	.0000

يتضح من الجدول (10) أن قيم الأوساط الحسابية والانحرافات المعيارية والالتواء والتفطح لمؤشر أكثر قرباً من القيم النظرية لخصائص التوزيع الطبيعي المعياري، عند استخدام طريقة (WLE) لتقدير القدرة، تليها طريقة (MLE)، ثم طريقة (MAP) في عينات الدراسة الثلاث، حيث تكون قيم كل من الوسط الحسابي والالتواء والتفطح (0) في التوزيع الطبيعي المعياري، بينما تكون قيمة الانحراف المعياري (1).

كما أظهرت النتائج في الجدول (10) أن توزيع مؤشر  $l_z(MIX)$  أكثر قرباً من التوزيع الطبيعي عند استخدام طريقة (WLE) لتقدير القدرة، تليها طريقة (MLE) ثم طريقة (MAP).

يتضح من الجدول (10) أن قيم الأوساط الحسابية والانحرافات المعيارية والالتواء والتفطح لمؤشر أكثر قرباً من القيم النظرية لخصائص التوزيع الطبيعي المعياري، عند استخدام طريقة (WLE) لتقدير القدرة، تليها طريقة (MLE)، ثم طريقة (MAP) في عينات الدراسة الثلاث، حيث تكون قيم كل من الوسط الحسابي والالتواء والتفطح (0) في التوزيع الطبيعي المعياري، بينما تكون قيمة الانحراف المعياري (1).

كما أظهرت النتائج في الجدول (10) أن توزيع مؤشر  $l_z(MIX)$  أكثر قرباً من التوزيع الطبيعي عند استخدام طريقة



النتائج الخاصة بالسؤال الثاني الذي نصّ على: "هل تختلف نسبة أنماط الاستجابة غير المطابقة التي يكشف عنها مؤشري و تبعاً لحجم العينة (5000,2000,1000) ولطريقة تقدير القدرة (MAP,WLE,MLE)؟"

للإجابة عن سؤال الدراسة الثاني: جرى حساب التكرارات والنسب المئوية لأنماط الاستجابة غير المطابقة التي كشف عنها مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$ ، تبعاً لعينات الدراسة الثلاث (5000,2000,1000)، ولطريقة تقدير القدرة (MAP, MLE, WLE)، كما هو مبين في الجداول (11).

وأظهرت النتائج في الجدول (10) أن توزيع مؤشر  $l_z^*(MIX)$  أكثر قرباً من التوزيع الطبيعي عند استخدام طريقة (WLE) لتقدير القدرة تليها طريقة (MLE) ثم طريقة (MAP) في عينات الدراسة الثلاث، اعتماداً على قيمة اختبار (Kolmogorov- Smirnov).

كما أظهرت النتائج في الجدول (10) أن توزيع مؤشر  $l_z(MIX)$  أكثر قرباً من التوزيع الطبيعي المعياري، مقارنة بمؤشر  $l_z(MIX)$  في طرق تقدير القدرة المختلفة، وفي عينات الدراسة الثلاث، وأن توزيعات المؤشرين بشكل عام ملتوية نحو اليسار ولها تقلطح موجب.

#### الجدول (11)

نسب أنماط الاستجابة المطابقة وغير المطابقة التي كشف عنها مؤشري المطابقة و في عينات الدراسة الثلاث (5000,2000,1000) تبعاً لطريقة تقدير القدرة (MAP,MLE,WLE)

حجم العينة	مؤشر المطابقة	طريقة تقدير القدرة	تكرار أنماط الاستجابة غير المطابقة	تكرار أنماط الاستجابة المطابقة	نسبة أنماط الاستجابة المطابقة	نسبة أنماط الاستجابة غير المطابقة
1000	$l_z(MIX)$	WLE	88	912	8.80%	91.20%
		MLE	85	915	8.50%	91.50%
		MAP	82	918	8.20%	91.80%
	$l_z^*(MIX)$	WLE	101	899	10.10%	89.90%
		MLE	98	902	9.80%	90.20%
		MAP	94	906	9.40%	90.60%
2000	$l_z(MIX)$	WLE	182	1818	9.10%	90.90%
		MLE	179	1821	8.95%	91.05%
		MAP	172	1828	8.60%	91.40%
	$l_z^*(MIX)$	WLE	223	1777	11.15%	88.85%
		MLE	221	1779	11.05%	88.95%
		MAP	195	1805	9.75%	90.25%
5000	$l_z(MIX)$	WLE	545	5455	10.90%	89.10%
		MLE	520	4480	10.40%	89.60%
		MAP	505	4495	10.10%	89.90%
	$l_z^*(MIX)$	WLE	670	4330	13.40%	86.60%
		MLE	655	4345	13.10%	86.90%
		MAP	575	4425	11.50%	88.50%

يتضح من الجدول (11) أن نسب الكشف عن أنماط الاستجابة غير المطابقة لمؤشر  $l_z(MIX)$  في عينته حجمها (1000) طالب، تراوحت بين (8.20%) و(8.80%)، وتراوحت هذه النسب في عينته حجمها (2000) طالب بين (9.40%) و(10.10%)، وتراوحت هذه النسب حين تراوحت نسب الكشف في عينته حجمها (5000) طالب بين (11.50%) و(13.40%)، كانت أعلى هذه النسب عند استخدام طريقة (WLE) لتقدير القدرة، تليها طريقة (MLE) ثم طريقة (MAP).

ويتضح من الجدول (11) أيضاً أن نسب الكشف لمؤشر  $l_z^*(MIX)$  أعلى من نسب الكشف لمؤشر  $l_z(MIX)$  في طرق تقدير القدرة المختلفة، وفي جميع العينات، وأن نسب الكشف عن أنماط

يتضح من الجدول (11) أن نسب الكشف عن أنماط الاستجابة غير المطابقة لمؤشر  $l_z(MIX)$  في عينته حجمها (1000) طالب، تراوحت بين (8.20%) و(8.80%)، وتراوحت نسب الكشف في عينته حجمها (2000) طالب بين (8.60%) و(9.10%)، في حين تراوحت نسب الكشف في عينته حجمها (5000) طالب بين (10.10%) و(10.90%)، وكانت أعلى هذه النسب عند استخدام طريقة (WLE) لتقدير القدرة، تليها طريقة (MLE) ثم طريقة (MAP)، في عينات الدراسة الثلاث.

كما يتضح من الجدول (11) أن نسب الكشف عن أنماط

طريقة تقدير القدرة (MAP,MLE,WLE) ضمن حجم العينة المختارة (5000,2000,1000) كانت ذات دلالة إحصائية عند مستوى  $\alpha$  (0.05)، في عينات الدراسة الثلاث.

ويتضح من الجدول (13) أيضاً أن الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشر  $l_z^*(MIX)$  عند تغيير طريقة تقدير القدرة (MAP,MLE,WLE)، ضمن حجم العينة المختارة (5000,2000,1000) كانت ذات دلالة إحصائية  $\alpha$  (0.05) في عينات الدراسة الثلاث.

الجدول (14)

نتائج اختبار كوكران (Cochran's Q test) لفحص دلالة الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشر  $l_z^*(MIX)$  مقابل النسب التي كشفها مؤشر  $l_z(MIX)$  تبعاً لطريقة تقدير القدرة (MAP,MLE,WLE) ضمن العينة المختارة (5000,2000,1000)

حجم العينة	طريقة تقدير القدرة	قيمة الإحصائي Q	مستوى الدلالة
1000	WLE	12.0000	.0021
	MLE	12.0000	.0021
	MAP	11.0000	.0023
2000	WLE	41.0000	.0000
	MLE	42.0000	.0000
	MAP	23.0000	.0000
5000	WLE	103.0000	.0000
	MLE	116.0000	.0000
	MAP	58.0000	.0000

يتضح من الجدول (14) أن الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشر  $l_z(MIX)$  والنسب التي كشفها مؤشر  $l_z^*(MIX)$  كانت ذات دلالة إحصائية عند مستوى  $\alpha$  (0.05). لطرق تقدير القدرة المختلفة، وفي عينات الدراسة الثلاث.

## مناقشة النتائج

### مناقشة النتائج الخاصة بسؤال الدراسة الأول

أظهرت نتائج هذا السؤال، انحراف توزيع مؤشري المطابقة  $l_z(MIX)$  و  $l_z^*(MIX)$  عن التوزيع الطبيعي، في عينات الدراسة الثلاث، وفي طرق تقدير القدرة المختلفة، وأن توزيع المؤشرين بشكل عام ملتو نحو اليسار وله تفلطح موجب، ويعزى ذلك إلى أن هذه الدراسة اعتمدت على قيم القدرة المقدر من البيانات الفعلية، حيث تم بناء افتراض التوزيع الطبيعي للمؤشر اعتماداً على تقديرات القدرة الحقيقية من البيانات المولدة، وقد اتفقت هذه النتيجة مع نتيجة دراسة سنهاري (Sinharay, 2016) التي أظهرت أن شكل توزيع مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  ملتو نحو اليسار وله تفلطح موجب، كما تشابهت مع نتائج دراسة نونان وبوس وجيسارولي (Noonan, Boss & Gessaroli, 1992) ودراسة كوي وموسافي (Cui & Mousafi, 2015) ودراسة سيو وويس (Seo & Weiss, 2013) التي أظهرت كل منها أن توزيع مؤشر  $l_z$  بشكل عام ملتو نحو اليسار وله تفلطح موجب عند استخدام القدرة المقدر.

الاستجابة غير المطابقة لمؤشري و تتزايد بزيادة حجم العينة، في الطرق المختلفة لتقدير القدرة.

وللتحقق من دلالة الفروق الظاهرية بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$ ، تبعاً لطريقة تقدير القدرة وحجم العينة، تم استخدام اختبار مربع كاي، واختبار كوكران (Cochran's Q test)، والجدول (14,13,12) توضح نتائج هذه الاختبارات.

الجدول (12)

نتائج اختبار مربع كاي لفحص دلالة الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشري المطابقة و تبعاً لطريقة القدرة (MAP,MLE,WLE) عبر الحجم المختلفة للعينة (5000,2000,1000).

طريقة تقدير القدرة	WLE	MLE	MAP
قيمة اختبار	مستوى الدلالة	قيمة اختبار	مستوى الدلالة
$l_z(MIX)$	7.463	6.300	6.321
$l_z^*(MIX)$	12.324	11.643	6.867

يتضح من الجدول (12) أن الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشر  $l_z(MIX)$  عبر الحجم المختلفة للعينة (5000,2000,1000)، كانت دالة إحصائية عند مستوى  $\alpha$  (0.05) عند استخدام طريقة (WLE) لتقدير القدرة، كما كانت هذه الفروق دالة إحصائية عند استخدام طريقة (MLE)، وكانت هذه الفروق أيضاً دالة إحصائية عند استخدام طريقة (MAP).

كما يتضح من الجدول (12) أن الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشر  $l_z^*(MIX)$  عبر الحجم المختلفة للعينة (5000,2000,1000)، عند استخدام طريقة (WLE) لتقدير القدرة، كانت دالة إحصائية عند مستوى الدلالة  $\alpha$  (0.05)، كما كانت هذه الفروق دالة إحصائية عند استخدام طريقة (MLE)، وكانت هذه الفروق أيضاً دالة إحصائية عند استخدام طريقة (MAP).

الجدول (13)

نتائج اختبار كوكران (Cochran's Q test) لفحص دلالة الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشري وباختلاف طريقة تقدير القدرة (MAP,MLE,WLE) ضمن العينة المختارة (5000,2000,1000).

مؤشر المطابقة	حجم العينة	قيمة الإحصائي Q	مستوى الدلالة
$l_z(MIX)$	1000	12.2500	.0022
	2000	14.3636	.0007
	5000	48.0625	.0000
$l_z^*(MIX)$	1000	13.9474	.0012
	2000	31.2857	.0000
	5000	98.5384	.0000

يتضح من الجدول (13) أن الفروق بين نسب أنماط الاستجابة غير المطابقة التي كشفها مؤشر  $l_z(MIX)$  عند تغيير

طريقة (WLE) مقارنة بطريقتي (MLE) و (MAP)، وباقترب توزيع المؤشر من التوزيع الطبيعي، تزداد المساحة تحت القيمة المعيارية الحرجة (1.645-) المستخدمة في الحكم على نمط الاستجابة، وبالتالي يزداد عدد المفحوصين الذين تقع قيمة مؤشر المطابقة لنمط استجاباتهم تحت هذه القيمة، وقد تشابهت هذه النتيجة مع نتيجة دراسة ريس (Reise, 1995) ودراسة ميجر ونيرنج (Meijer & Nering, 1997) اللتان أظهرتا أن طرق تقدير القدرة الموزونة (BIW) تزيد من قوة مؤشر  $I_Z^*$  في الكشف عن أنماط الاستجابة غير المطابقة.

كما أظهرت النتائج أن نسب الكشف عن أنماط الاستجابة غير المطابقة لمؤشري  $I_Z(MIX)$  و  $I_Z^*(MIX)$  تتزايد بزيادة حجم العينة، في طرق تقدير القدرة المختلفة، ويعزى ذلك إلى أن شكل توزيع المؤشرين يصبح أكثر قرباً من التوزيع الطبيعي بزيادة حجم العينة، كما أنه بزيادة حجم العينة تزداد الفرصة لظهور أنماط الاستجابة غير المطابقة نظراً لزيادة عدد الأفراد.

وأظهرت النتائج أيضاً أن نسب الكشف عن أنماط الاستجابة غير المطابقة لمؤشر  $I_Z^*(MIX)$  أعلى من النسب التي يكشف عنها مؤشر  $I_Z(MIX)$  في عينات الدراسة المختلفة، وفي طرق تقدير القدرة المختلفة، ويعزى ذلك إلى أن توزيع مؤشر  $I_Z^*(MIX)$  أكثر قرباً من التوزيع الطبيعي المعياري من توزيع مؤشر  $I_Z(MIX)$  وبالتالي تزداد الدقة في تصنيف الأفراد إلى مطابقين أو غير مطابقين، وتصبح المساحة تحت درجة القطع في توزيع المؤشر  $I_Z^*(MIX)$  أكبر من المساحة في توزيع المؤشر  $I_Z(MIX)$ .

### استنتاجات الدراسة

خلصت الدراسة إلى مجموعة من الاستنتاجات، أبرزها:

- كان مؤشر  $I_Z^*(MIX)$  أكثر فاعلية من مؤشر  $I_Z(MIX)$  في الظروف المختلفة لطريقة تقدير القدرة وحجم العينة، نظراً لقرب توزيعه من التوزيع الطبيعي مما يزيد من دقة القرارات الخاصة بالحكم على أنماط الاستجابة، وزيادة نسب أنماط الاستجابة غير المطابقة التي يكشف عنها، والتي يهدد بقاءها صدق نتائج الاختبار ودقة القرارات المتعلقة بالمفحوصين.

- ازدادت فاعلية مؤشري  $I_Z(MIX)$  و  $I_Z^*(MIX)$  عند استخدام طريقة (WLE) لتقدير القدرة مقارنة بطرق التقدير الأخرى، إذ اقترب توزيع المؤشرين من التوزيع الطبيعي، وازدادت نسب الكشف عند استخدام هذه الطريقة.

- ازدادت فاعلية مؤشري  $I_Z(MIX)$  و  $I_Z^*(MIX)$  عند زيادة حجم العينة، إذ ازداد قرب توزيع المؤشرين من التوزيع الطبيعي، وارتفعت نسب أنماط الاستجابة غير المطابقة التي كشف المؤشران عنها.

### التوصيات

في ضوء نتائج الدراسة يوصي الباحثان بما يلي:

- استخدام مؤشر  $I_Z^*(MIX)$  للكشف عن أنماط الاستجابة غير المطابقة في الاختبارات مختلطة الشكل، نظراً لقرب توزيعه من التوزيع الطبيعي، وارتفاع نسب أنماط الاستجابة غير المطابقة

وأظهرت النتائج أيضاً أن توزيعات مؤشري المطابقة  $I_Z(MIX)$  و  $I_Z^*(MIX)$  كانت أكثر قرباً من التوزيع الطبيعي عند استخدام طريقة (WLE) لتقدير القدرة تليها طريقة (MLE) ثم طريقة (MAP)، في عينات الدراسة الثلاث، ويعزى ذلك إلى أن طريقة (WLE) أكثر دقة في تقدير القدرة من طريقتي (MLE) و (MAP)، نظراً لاعتماد طريقة (WLE) على توزيع دالة الأرجحية ببعض الدوال المناسبة، للتقليل من التحيز في تقدير القدرة بطريقة (MLE) حسب المعادلة (5)، كما تتأثر طريقة (MAP) بالتوزيع القبلي الفرضي للقدرة في مجتمع الدراسة؛ مما يزيد من عدم الدقة في تقدير القدرة وزيادة الأخطاء المعيارية، ويؤدي التقدير غير الدقيق لقدرة الفرد إلى أخطاء في حساب قيمة مؤشر المطابقة، وبالتالي التأثير على شكل وخصائص توزيع مؤشر المطابقة، وقد تشابهت هذه النتيجة مع نتيجة دراسة سيو وويس (Seo & Weiss, 2013) التي أظهرت أن توزيع مؤشر  $I_Z$  كان أكثر قرباً للتوزيع الطبيعي، عند استخدام طريقة (MLE) لتقدير القدرة مقارنة بطريقة (MAP)، كما تشابهت مع نتائج دراسة نيرنج (Nering, 1995) التي أظهرت أن توزيع مؤشر  $I_Z$  كان أقرب إلى التوزيع الطبيعي عند استخدام طريقة (MLE) مقارنة بطريقة بيز (Bayesian Method).

كما أظهرت النتائج أن توزيعات مؤشري المطابقة  $I_Z(MIX)$  و  $I_Z^*(MIX)$  تقترب من التوزيع الطبيعي بزيادة حجم العينة، ويعزى ذلك إلى زيادة قيم مؤشري المطابقة التي تتمركز في وسط التوزيع مقارنة بالقيم عند أطرافه عند زيادة حجم العينة، وقد تشابهت هذه النتيجة مع نتيجة دراسة (Lopez & Montesinos, 2005) التي أظهرت أن توزيع مؤشر  $I_Z$  يقترب من التوزيع الطبيعي بزيادة حجم العينة.

وأظهرت النتائج أيضاً أن توزيع مؤشر  $I_Z^*(MIX)$  أكثر قرباً من التوزيع الطبيعي المعياري، مقارنة بتوزيع مؤشر  $I_Z(MIX)$  في الطرق المختلفة لتقدير القدرة، وفي عينات الدراسة الثلاث، ويعزى ذلك إلى الأساس الرياضي لمؤشر  $I_Z^*(MIX)$  القائم على تصحيح توزيع المؤشر، من خلال الاعتماد على الأوزان المعدلة للقدرة في حساب المؤشر بدلاً من الأوزان العادية في حساب مؤشر  $I_Z(MIX)$  كما في المعادلات (2،1)، وقد اتفقت هذه النتيجة مع نتائج دراسة سنهاري (Sinharay, 2016) التي أظهرت أن توزيع مؤشر  $I_Z^*(MIX)$  أكثر قرباً من التوزيع الطبيعي المعياري من توزيع مؤشر، وتشابهت مع نتائج دراسة ماجيس وريش وبييلاند (Magis, Raïche, & Bé-land, 2012) التي أظهرت أن توزيع مؤشر  $I_Z^*$  أقل تفلطحاً من توزيع مؤشر  $I_Z$  ونتائج دراسة ميجر وتندريو (Meijer & Tende-rio, 2012) التي أظهرت أن شكل توزيع المؤشر  $I_Z^*$  بشكل عام أقرب للتوزيع الطبيعي المعياري من مؤشر  $I_Z$ .

### مناقشة النتائج الخاصة بسؤال الدراسة الثاني

أظهرت نتائج هذا السؤال، أن أعلى نسب الكشف عن أنماط الاستجابة غير المطابقة لمؤشري  $I_Z(MIX)$  و  $I_Z^*(MIX)$  كانت عند استخدام طريقة (WLE) لتقدير القدرة تليها طريقة (MLE) ثم طريقة (MAP)، في عينات الدراسة الثلاث، ويعزى ذلك إلى أن شكل التوزيع للمؤشرين عند استخدام طريقة (WLE) كان أكثر قرباً للتوزيع الطبيعي، مقارنة بطريقتي (MLE) و (MAP)، وبالتالي تزداد الدقة في تصنيف الأفراد إلى مطابقين أو غير مطابقين عند استخدام

- Journal of Testing*, 15(1): 23-49.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford.
  - De La Torre, J., & Deng, W. (2008). *Improving Person-Fit Assessment by Correcting the Ability Estimate and Its Reference Distribution*. *Journal of Educational Measurement*, 45(2): 159-177.
  - Drasgow, F., Levine, M. V., & Williams, E. A. (1985). *Appropriateness measurement with polychotomous item response models and standardized indices*. *British Journal of Mathematical and Statistical Psychology*, 38: 67-86.
  - Glas, C. A., & Meijer, R. R. (2003). *A Bayesian approach to person fit analysis in item response theory models*. *Applied Psychological Measurement*, 27(3): 217-233.
  - Hambleton, R.K., & Swaminathan .H. (1985). *Item Response Theory Principles and applications*, Boston: Kluwer: Nijhoff Publishing.
  - Hattie, J. (1985). *Methodology review: assessing unidimensionality of tests and items*. *Applied psychological measurement*, 9(2): 139-164.
  - Karabatsos, G. (2003). *Comparing the aberrant response detection performance of thirty-six person-fit statistics*. *Applied Measurement in Education*, 16(4): 277-298.
  - Lopez, A.J & Montesinos, H.D. (2005). *Fitting rasch model using appropriateness Measure Statistics*. *The Spanish journal of psychology*, 8: 100 – 110 .
  - Magis, D., Raiche, G., & Béland, S. (2012). *A didactic presentation of Snijders's lz\* index of person fit with emphasis on response model selection and ability estimation*. *Journal of Educational and Behavioral Statistics*, 37(1): 57-81.
  - Meijer, R. R. (1996). *Person-fit research: An introduction*. *Applied Measurement in Education*, 9: 3-8.
  - Meijer, R. R., & Nering, M. L. (1997). *Trait level estimation for nonfitting response vectors*. *Applied Psychological Measurement*, 21(4): 321-336.
  - Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2015). *A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics: Examples and a computer program*. *Assessment*, 23(1): 52-62.
  - Meijer, R. R., & Sijtsma, K. (2001). *Methodology review: Evaluating person fit*. *Applied psychological measurement*, 25(2): 107-135
  - Meijer, R. R., & Tendeiro, J. N. (2012). *The use of the lz and lz\* person-fit statistics and problems derived from model misspecification*. *Journal of Educational and Behavioral Statistics*, 37(6): 758-766.
  - Nering, M.L.(1995). *The distribution of person fit using true and estimated person*. *Applied psychological Measurement*, 19:121-129.
  - Noonan, B. W., Boss, M. W., & Gessaroli, M. E. (1992). *The effect of test length and IRT model on the distribution and stability of three appropriateness indexes*. *Applied Psychological Measurement*, 16(4): 345-352.
  - Olson, J. F., & Fremer, J. (2013). *TILSA test security guidebook: Preventing, detecting and investigating test securities irregularities*. Washington DC : Council of Chief State School Officers.
  - Reise, S. P. (1990). *A comparison of item-and person-fit methods of assessing model-data fit in IRT*. *Applied Psychological Measurement*, 14(2): 127-137.
  - Reise, S.P. (1995) . *Scoring method and the detection of person misfit in a personality assessment context* . *Applied psychological Measurement*, 19:213 – 229.
  - Rizopoulos, D. (2006) .*ltm: An R package for latent variable modeling and item response theory analyses* .*Journal of Statistical Software*, 17(5): 1-25.
  - Seo, D. G., & Weiss, D. J. (2013). *lz Person-fit index to identify misfit students with achievement test data*. *Educational and*

التي يكشف عنها مقارنة بمؤشر  $l_z(MIX)$  .

- استخدام طريقة (WLE) لتقدير القدرة عند استخدام مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الاختبارات مختلطة الشكل، نظراً لدقتها في تقدير القدرة، التي تنعكس على خصائص توزيع المؤشرين ونسب أنماط الاستجابة غير المطابقة التي يكشف عنها المؤشران.
- إجراء دراسات حول أثر نسبة الفقرات ذات الإجابة المصاغة في الاختبارات مختلطة الشكل، على فاعلية مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الكشف عن أنماط الاستجابة غير المطابقة.
- إجراء دراسات حول أثر عدد الخطوات في الفقرات ذات الإجابة المصاغة، على فاعلية مؤشري  $l_z(MIX)$  و  $l_z^*(MIX)$  في الكشف عن أنماط الاستجابة غير المطابقة.

## قائمة المصادر والمراجع

### المراجع العربية

- حمادنه، مروان عبدالله.(2011). فاعلية أسلوب تحسين مطابقة الفرد القائم على تصحيح تقدير القدرة وتوزيعها المرجعي عند الاختلاف في حجم العينة والنموذج اللوجستي (أطروحة دكتوراه غير منشورة). جامعة اليرموك. الأردن.
- علام، صلاح الدين.(2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة، مصر: دار الفكر العربي.
- المركز الوطني لتنمية الموارد البشرية.(2017). التقرير الوطني الأردني عن الدراسة الدولية للرياضيات والعلوم لعام 2015، عمان، الأردن.

### المراجع المترجمة

- Hamadneh, M. (2011). *The effectiveness of an individual matching improvement method to correct capacity estimate and reference distribution when differences in sample size and logistical model (unpublished doctorate)*. Yarmouk University. Jordan.
- Allam, S. (2005). *Single-dimensional and multi-dimensional experimental response models and their applications in psychological and educational measurement* . Cairo, Egypt: Dar Al-Fikr Al-Arabi.
- National Center for Human Resources Development. (2017). *Jordanian National Report on the International Study of Mathematics and Science 2015*, Amman, Jordan
- Armstrong, R. D., Stoumbos, Z. G., Kung, M. T., & Shi, M. (2007). *On the performance of the lz person-fit statistic*. *Practical Assessment, Research & Evaluation*, 12(16): 1-15.
- Birenbaum, M. (1986). *Effect of dissimulation motivation and anxiety on response pattern appropriateness measures* . *Applied Psychological Measurement*, 10(2): 167-174.
- Cao, Y. (2008). *Mixed-format test equating: Effects of test dimensionality and common-item sets (Unpublished Doctoral Dissertation)*. University of Maryland . America.
- Conijn, J. M., Emons, W. H., & Sijtsma, K. (2014). *Statistic lz-based person-fit methods for noncognitive multiscale measures*. *Applied Psychological Measurement*, 38(2): 122-136
- Cui, Y., & Mousavi, A. (2015). *Explore the Usefulness of Person-Fit Analysis on Large-Scale Assessment*. *International*



- Psychological Measurement*, 73(6): 994-1016.
- Sinharay, S. (2015). Assessment of person fit for mixed-format tests. *Journal of Educational and Behavioral Statistics*, 40(4): 343-365
  - Sinharay, S. (2016). Asymptotic corrections of standardized extended caution indices. *Applied psychological measurement*, 40(6): 418-433.
  - Sinharay, S. (2018). Extension of caution indices to mixed-format tests. *British Journal of Mathematical and Statistical Psychology*, 71(2): 363-386.
  - Snijders, T. A. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika*, 66(3): 331-342.
  - Steinkamp, S. (2017). *Identifying Aberrant Responding: Use Of Multiple Measures ( Unpublished Doctoral Dissertation)*. University of Minnesota.America.
  - Tatsouka, K.K. (1996). Use of generalized person-fit indexes, zetas for statistical pattern classification . *Applied Measurement In Education* .9:65-75.
  - TIMSS & PIRLS International Study Center .(2015). *TIMSS international database*. Retrieved 14-1-2019 from <https://timssandpirls.bc.edu/timss2015/international-database/>
  - Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2): 109-135.
  - Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3): 427-450.