

The Effectiveness of the Likelihood Ratio Test and Multiple Indicators Multiple Causes Method in Detecting Differential Item Functioning According to Sample Size, Percentage of Items with DIF, and Ability Distribution

فاعلية طريقتي نسبة الأرجحية العظمى والمؤشرات المتعددة والأسباب المتعددة في الكشف عن الاداء التفاضلي تبعاً لنسبة الفقرات ذات الاداء التفاضلي وشكل توزيع القدرة

Ala'a Qawaqneh^{1*}, Zaid Bani Ata¹.

¹Yarmouk University, Irbid, Jordan.

ARTICLE INFO

Article history:

Received 02 Jun 2022

Accepted 26 Jul 2022

Published 01 Jul 2024

*Corresponding author:

Yarmouk University, Irbid, Jordan.

Email: alm_qawaqneh@yahoo.com.

Abstract

This study aimed to verify the effectiveness of the Likelihood Ratio Test (IRTLR) and the Multiple Indicators Multiple Causes (MIMIC) method in detecting differential item functioning (DIF) according to the percentage of items with DIF and ability distribution. To achieve the study's aim, nine ability distributions and three percentages of items with differential functioning were generated using the three-parameter model with the WINGEN program on 50 dichotomous items. The generated data were analyzed using the R program and the MULTILOG program. The results revealed statistically significant differences at $\alpha = 0.05$ between Type I error and power means due to the percentage of items with DIF, DIF methods, and their interaction. Type I error increased with a higher percentage of items with DIF for both methods. The results also indicated that the power of the test decreased as the percentage of items with DIF increased. Furthermore, statistically significant differences at $\alpha = 0.05$ were found due to ability distributions, DIF methods, and their interaction, with the MIMIC method being better at detecting DIF from ability distribution.

Keywords: Ability Distribution, Three- Parameter Logistic Model, Differential Item Functioning, Likelihood Ratio Test, Multiple Indicators and Causes.

المخلص

هدفت الدراسة الحالية إلى التحقق من فاعلية طريقتي نسبة الأرجحية العظمى (IRTLR) والمؤشرات المتعددة والأسباب المتعددة (MIMIC) في الكشف عن الاداء التفاضلي للفقرات تبعاً لنسبة الفقرات ذات الاداء التفاضلي وشكل توزيع القدرة. ولتحقيق الهدف من الدراسة تم توليد استجابات لتسعة أشكال من توزيعات القدرة وثلاثة نسب من الفقرات ذات الاداء التفاضلي وفق النموذج اللوجستي ثلاثي المعلمة من خلال استخدام برنامج WINGEN على 50 فقرة ثنائية الاستجابة. تم تحليل البيانات المولدة باستخدام برنامج R وبرنامج MULTILOG. كشفت نتائج الدراسة عن وجود فروق ذات دلالة

إحصائية عند ($\alpha = 0.05$) بين الأوساط الحسابية لمعدل الخطأ من النوع الأول وقوة الاختبار لكل نسبة من نسب الفقرات ذات الأداء التفاضلي وطريقة الكشف عن الأداء التفاضلي والتفاعل بينهم، وكانت تزداد معدلات الخطأ من النوع الأول بزيادة نسبة الفقرات ذات الأداء التفاضلي لكلا الطريقتين. كما بينت النتائج أن قوة الاختبار تقل بزيادة نسبة الفقرات ذات الأداء التفاضلي لكلا الطريقتين. وبينت النتائج كذلك وجود فروق ذات دلالة إحصائية ($\alpha = 0.05$) لكل من شكل توزيع القدرة وطريقة الكشف عن الأداء التفاضلي والتفاعل بينهم، وتفوقت طريقة (MIMIC) على طريقة (IRTLLR) تبعاً لكل شكل من أشكال توزيع القدرة.

الكلمات المفتاحية: شكل توزيع القدرة، النموذج اللوجستي ثلاثي المعلمة، الأداء التفاضلي للفقرة، اختبار نسبة الأرجحية، المؤشرات المتعددة والأسباب المتعددة.

١. المقدمة

١.١ الإطار النظري والدراسات السابقة

تلعب الاختبارات دوراً أساسياً في عملية التقويم التربوي، حيث تستخدم نتائجها في اتخاذ القرارات المتنوعة، وكلما ازدادت أهمية القرارات المتوقع اتخاذها كلما ازدادت الحاجة إلى أن تكون المعلومات التي نحصل عليها من الاختبار دقيقة، وذات صلة وثيقة بالغرض الذي أعدت من أجله (عوده، ٢٠٠٥). فالاختبارات المعدة بشكل جيد، والتي تتمتع بخصائص الصدق والثبات والموضوعية تساعد التربويين في الحصول على نتائج دقيقة وموثوقة.

كما أولى المختصون في مجال القياس النفسي والتربوي قضية عدالة الاختبار بالغ الاهتمام والتقدير، نظراً لأهمية القرارات التي يجري اتخاذها استناداً إلى نتائج الاختبارات، سواء في اختيار الأفراد للوظائف المختلفة، أو تصنيفهم تبعاً لقدراتهم، أو تحصيلهم، أو تحديد الصفوف المناسبة لهم. ويتصف الاختبار بالعدالة عندما يكون صادقاً بالنسبة لجميع الأفراد والمجموعات، ويعطي فرصاً متساوية لجميع المفحوصين لإبداء المعرفة والمهارات المكتسبة المتعلقة بغرض الاختبار (Roever, 2005).

ويرى كل من بير وبارك (Pare & Park, 2006) أن الاختبار يكون متحيزاً إذا انتهك مبدأ العدالة والمساواة بين المفحوصين، ولذلك يجب أن يتمتع الاختبار بدرجة عالية من الصدق لجميع المفحوصين والمجموعات، وذلك بأن يعطي فرصاً متساوية لهم جميعاً، فإذا ميز الاختبار المصمم لقياس سمة ما بين مجموعة ومجموعة أخرى من نفس المستوى، فإن هذا يعد أحد مهددات صدق المقياس وفقراته، لذلك يعد التحقق من خلو الاختبار وفقراته من التحيز خطوة مهمة من خطوات بناء الاختبار.

وبعد فحص الأداء التفاضلي لفقرات الاختبار (DIF) من أهم الطرق للتحقق من عدالة الاختبار وصدقه والتعرف على الفقرات المتحيزة. ويعرف الأداء التفاضلي للفقرة (DIF) بأنه "دالة مشتقة إحصائياً لتعبر عن الفروق في احتمال استجابة الفرد الصحيحة للفقرة بين مجموعتين ممن هم في نفس المستوى من القدرة (Teresi, 2001). وتجدر الإشارة هنا إلى أن ثمة فرق بين مفهوم التحيز والأداء التفاضلي للفقرة، فيرى دورانز وهولاند (Dorans & Holland, 1993) أن (DIF) يهتم بالخصائص السيكومترية للفقرة من حيث الكيفية التي تعمل بها في المجموعات المختلفة، بينما يهتم تحيز الفقرة بالمعنى الاجتماعي بشكل كبير.

ووفقاً لونغ ولان (Wang & Lane, 1996) يمكن تصنيف (DIF) تبعاً لطبيعة التفاعل بين مستوى القدرة وعضوية المجموعة إلى نوعان:

- الأداء التفاضلي المنتظم: يظهر عندما لا يكون هناك تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد، أي أن احتمال الإجابة الصحيحة للفقرة يكون أكبر دائماً لإحدى المجموعات عند جميع مستويات القدرة.
- الأداء التفاضلي غير المنتظم: يظهر عند حدوث تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد، أي أن فروق احتمال الإجابة الصحيحة ليست ثابتة عند جميع مستويات القدرة، فمرة يظهر الأداء التفاضلي لصالح المجموعة المرجعية، ومرة أخرى يظهر لصالح المجموعة المستهدفة عند مستويات قدرة مختلفة.

ظهرت عدة طرق إحصائية للكشف عن الأداء التفاضلي لفقرات الاختبار في بداية الستينات نتيجة للجهود الكبيرة التي بذلها علماء القياس، والاهتمام المتزايد بصدق نتائج الاختبارات، تستند بعض هذه الطرق الى النظرية الكلاسيكية في القياس (CTT)، مثل طريقة مانتل-هانزل (Mantel-Haenszel Method) وطريقة الانحدار اللوجستي (Logistic Regression Method) ومنها ما يستند إلى نظرية الاستجابة للفقرة (IRT): كطريقة كاي تربيع للورد (Lord's Chi-Square)، وطريقة اختبار نسبة الأرجحية (IRT-LR). ومنها ما يستند إلى نمذجة المعادلة البنائية (SEM): كطريقة التحليل العاملي التوكيدي متعدد المجموعات (MG-CFA)، وطريقة المؤشرات المتعددة والأسباب المتعددة (MIMIC).

وبعد اختبار نسبة الأرجحية (IRT-LR). طريقة من طرق الكشف عن (DIF)، وإحدى الطرق المعلمية القائمة على نظرية (IRT)، وهي إحدى الطرق التي سيعتمدها الباحثان في إجراء التحليلات الإحصائية المتعلقة بالدراسة الحالية. لاحظ ثيسن وآخرون (Thissen et al., 1986) أن طريقي المساحة وكاي تربيع للورد صعبة من الناحية الحسابية، إذ أنها تتطلب الحصول على تقديرات التباين والتغاير لمعالم الفقرة، إضافة إلى حاجتها لمعادلة الدرجات، لذا اقترحوا طريقة بديلة لطرق نظرية (IRT)، وهي طريقة (IRT-LR) والتي تتميز بسهولة حسابها مقارنة بالطرق السابقة، إضافة إلى عدم حاجتها لإجراء عملية معادلة الدرجات، إذ يتم تقدير المعالم تزامنياً، وتم توسيعه عام ١٩٩٣ للبيانات الثنائية ومتعددة الاستجابة.

تقوم هذه الطريقة على استخدام الاحصائي نسبة الأرجحية (Likelihood Ratio) من أجل فحص الفرضية الصفرية والتي تنص على عدم وجود فروق في معالم الفقرات بين المجموعتين المرجعية والمستهدفة، وتعتمد صياغة الفرضية الصفرية على تعريف لورد لمفهوم التحيز الذي ينص على أن وجود فروق في دالة استجابة الفقرة بين مجموعتين مؤشر على وجود التحيز، ذلك أن مجموعة معالم الفقرة تناظر دالة استجابة الفقرة أو منحني خصائص الفقرة، وهي تصلح للكشف عن (DIF) في الفقرات ثنائية الاستجابة أو المتعددة الاستجابة، وتتماشى هذه الطريقة مع نموذج استجابة الفقرة الثلاثي ونموذج استجابة الفقرة المتدرج (اللبيدي، ٢٠٠٨).

كما وتقوم طريقة (IRT-LR) على إيجاد الفرق بين اثنين من النماذج، الأول يحتوي على جميع الفقرات باستثناء الفقرة المراد الكشف عن أدائها التفاضلي، ويسمى هذا النموذج بالنموذج المدمج (Compact Model)، والثاني يتكون من الفقرة المراد الكشف عن أدائها التفاضلي بالإضافة إلى الفقرة التي تم استثنائها وفقاً للنموذج المدمج ويسمى هذا النموذج بالنموذج الموسع (Augment Model)، ويتم حساب الفرق من خلال حساب المؤشر الإحصائي لاختبار نسبة الأرجحية (G^2) والذي يتم إيجاده من خلال المعادلة التالية:

$$G^2 = -2LLC - (-2LLA) \dots\dots\dots (1)$$

حيث إن:

LLC: لوغاريتم نسبة الأرجحية للنموذج المدمج

LLA: لوغاريتم نسبة الأرجحية للنموذج الموسع

كما وتعد (MIMIC) طريقة من طرق الكشف عن (DIF)، وهي الطريقة الأخرى التي سيعتمدها الباحثان في إجراء التحليلات الإحصائية المتعلقة بالدراسة الحالية. تم وصف طريقة (MIMIC) لأول مرة من قبل جورسكوج وجولد بيرغر (Jöreskog & Goldberger, 1975) كطريقة خاصة لـ (SEM)، تعد هذه الطريقة بالغة الأهمية للتحقق من الصدق لأنها تسمح بالتحقيق في الفروق بين المجموعات المتعددة في البنية الكامنة (Hancock, 2001)، ويمكن استخدامها في الكشف عن (DIF) في المؤشرات الملاحظة للمتغيرات الكامنة، وهي فعالة مع المجموعات المتباينة غير المتجانسة.

إذا كانت قيمة المؤشر الإحصائي لاختبار نسبة الأرجحية (G^2) تزيد عن القيمة الحرجة لمربع كاي بثلاث درجات، فإن النموذج المدمج والنموذج الموسع توجد بينهما فروق دالة إحصائية، ويعد ذلك مؤشراً على وجود أداء تفاضلي للفقرة (Greer, 2004).

طور موثن واسباروف (Muthen & Asparouhov, 2002) معادلة يمكن استخدامها للكشف عن الأداء التفاضلي للفقرات ثنائية الاستجابة:

$$Y_i = \lambda_i \theta + \beta_i z_k + \epsilon_i \dots\dots\dots (2)$$

متغير الاستجابة الكامنة: γ_i^*

متغير وهمي يشير الى عضوية المجموعة k : ζ_k

الميل الذي يربط متغير المجموعة بالاستجابة: β_i

قدرة الفرد θ :

معلمة تمييز الفقرة: λ_i

الخطأ العشوائي: ϵ_i

وقام وودز وجريم (Woods & Grimm, 2011) بتطوير طريقة (MIMIC-interaction) للكشف عن الأداء التفاضلي للفقرة (المنتظم، وغير المنتظم) في الفقرات ثنائية الاستجابة في وقت واحد، وذلك وفق المعادلة التالية:

$$Y_i^* = \lambda_i \theta + \beta_i \zeta_k + \omega_i \theta \zeta_k + \epsilon_i \dots \dots \dots (3)$$

مقارنة بالمعادلة (٢)، تمت إضافة مصطلح (ω_i) ليشير إلى التفاعل المحتمل بين المتغير الكامن وعضوية المجموعة، مما يشير إلى الأداء التفاضلي غير المنتظم.

تختلف طريقة (MIMIC) عن طرق الكشف عن الأداء التفاضلي الأخرى في أنها تستطيع الكشف عن الأداء التفاضلي للفقرة عن طريق إضافة أو حذف المسارات المباشرة للفقرات التي تشير إلى الأداء التفاضلي (المنتظم، وغير المنتظم)، ويتم حساب التأثير من خلال المسارات من مجموعة المتغيرات إلى العامل المشترك، والذي يمكننا من خلاله معرفة كيفية تأثير عضوية المجموعة على أوساط المتغير الكامن (Gao, 2019).

وقد تناولت العديد من الدراسات موضوع الأداء التفاضلي للفقرات في ضوء المتغيرات التي تؤثر على طريقتي الكشف عن الأداء التفاضلي للفقرة (IRT-LR, MIMIC). فقد أجرى وودز (Woods, 2009) دراسة هدفت إلى المقارنة بين طريقة (MIMIC) وطريقة (IRT-LR) للكشف عن (DIF) تحت ظروف تجريبية مختلفة، وذلك من خلال فحص معدلات الخطأ من النوع الأول وقوة الاختبار، ركزت الدراسة على التعرف على أثر كل طريقة وحجم المجموعة المستهدفة الصغير (٢٥، ٥٠، ١٠٠، ٢٠٠، ٤٠٠) على معدلات الخطأ من النوع الأول وقوة الاختبار، أظهرت النتائج أن أداء طريقة (MIMIC) أفضل من طريقة (IRT-LR) في الكشف عن الأداء التفاضلي المنتظم عندما تكون المجموعة المستهدفة صغيرة.

ومن جهة أخرى أجرى لوبيز (Lopezm, 2012) دراسة هدفت إلى دراسة فعالية ثلاثة طرق للكشف عن (DIF) للفقرة، وهي طريقة (IRT-LR) وطريقة تحيز الفقرة المتزامن (SIBTEST) وطريقة الانحدار اللوجستي، تم استخدام ظروف تجريبية مختلفة هي طول الاختبار (١٥، ٣٠)، وحجم عينات للمجموعتين المرجعية والمستهدفة (٢٥٠/٥٠٠، ٥٠٠/١٠٠٠)، وتوزيع قدرات مختلفة، كانت توزيعات القدرة للمجموعتين المرجعية والمستهدفة توزيعات طبيعية $N(0, 1)$. أظهرت النتائج أن طريقة الانحدار اللوجستي كانت أفضل الطرق في قوة الكشف عن الأداء التفاضلي، كما أشارت النتائج إلى أن طريقة (IRT-LR) كانت قوية، ولكنها تأثرت بتضخم معدل الخطأ من النوع الأول.

أجرى جاملي وآخرون (Jamali, et al., 2017) دراسة هدفت إلى تقييم طريقة (MIMIC) للكشف عن الاداء التفاضلي المنتظم للفقرة عندما يكون توزيع السمات الكامنة غير طبيعي وحجم العينة للمجموعة المستهدفة صغير، في هذه الدراسة القائمة على المحاكاة، تم فحص معدلات الخطأ من النوع الاول وقوة طريقة (MIMIC) للكشف عن الاداء التفاضلي للفقرة في مجموعات مختلفة بالإشارة الى نسبة حجم عينة المجموعة المستهدفة، وحجم تأثير (DIF) المنتظم وطول المقياس، وعدد فئات الاستجابة، وتوزيع السمات الكامنة، أدى الانحراف المتوسط والعالي في توزيع السمات الكامنة الى انخفاض قوة طريقة (MIMIC)، وأشارت النتائج الى أنه من خلال زيادة طول المقياس وعدد فئات الاستجابة وحجم (DIF) أدى الى تحسين قوة طريقة (MIMIC) بنسب ٣,٤٧% و ٤,٨٣% و ٢٠,٣٥% على التوالي، كما أنه قلل من الخطأ من النوع الاول، وكشفت هذه الدراسة أن قوة (MIMIC) كانت عند مستوى مقبول عندما تكون توزيعات السمات الكامنة ملتوية.

وعلى صعيد آخر أجرى زيمبرا (Zimbra, 2018) دراسة هدفت إلى تقييم طريقة (MIMIC) في الكشف عن (DIF) المنتظم وغير المنتظم للفقرة ومقارنة أدائه بطريقة (IRT-LR) واختبار والد (Wald Test)، تم إجراء دراسة محاكاة تحت ظروف تجريبية مختلفة، طول الاختبار (٢٥) فقرة وحجم العينة للمجموعتين المرجعية والمستهدفة (١٠٠٠/١٠٠٠).

٥٠٠/٥٠٠، ١٥٠٠/٥٠٠، ٧٥٠/٢٥٠، وحجم (DIF) (٣، ٥، ٧، ١٠)، تم توليد استجابات ثنائية التدرج وفق النموذج اللوجستي ثلاثي المعلمة، وتوزيع القدرة توزيع طبيعي بوسط حسابي (٠) وانحراف معياري (١). تشير النتائج إلى أن طريقة (MIMIC) تفوقت على طريقة (IRT-LR) واختبار والد (Wald Test) بناءً على معدلات الخطأ من النوع الأول وقوة الاختبار عندما كان حجم الاختبار كبير.

قام جو (٢٠١٩) Gao بإجراء دراسة هدفت لمقارنة ستة طرق للكشف عن الأداء التفاضلي المنتظم وغير المنتظم (مانتل-هانزل، الانحدار اللوجستي (IRT-LR، MIMIC) اختبار والد، اختبار العشوائية)، تم إجراء دراسة محاكاة وفق ظروف تجريبية مختلفة، نسبة الفقرات (١٠٪، ٢٠٪)، وحجم العينة للمجموعتين المرجعية والمستهدفة (٢٥٠/٢٥٠، ٢٥٠/٥٠٠، ٥٠٠/٥٠٠، ٥٠٠/١٠٠٠)، وطول الاختبار (٢٠، ٤٠)، ومتوسط القدرة (١٠، ١٠٠). أظهرت النتائج أن طريقة (MIMIC) كان لها قوة عالية في الكشف عن (DIF) المنتظم وغير المنتظم للفقرات عندما تكون الاختبارات أطول، وتعتبر طريقة IRT-LR وطريقة التوزيع العشوائي فعالة تماماً في الكشف عن الأداء التفاضلي المنتظم، لكن طريقة التوزيع العشوائي لا تطبق إلا عندما تكون مجموعتا الأفراد لهما نفس متوسط القدرة. يتضح من خلال الدراسات السابقة بأن هناك توجهاً عاماً نحو البحث عن كفاءة وفاعلية طريقتي (IRT-LR، MIMIC) في الكشف عن (DIF) للفقرات، حيث أجمعت الدراسات على وجود عوامل كثيرة تؤثر على فاعلية هذه الطرق في الكشف عن (DIF) للفقرات من أهمها طول الاختبار، وحجم العينة، ونسبة الفقرات ذات الأداء التفاضلي. وعلى الرغم من كثرة الدراسات التي اهتمت بدراسة كفاءة وفاعلية طريقتي (IRT-LR، MIMIC) في الكشف عن (DIF) للفقرات، ودراسة العوامل المؤثرة عليها، إلا أن المتصفح لها يجد غالبيتها اهتم بدراسة أثر حجم العينة وطول الاختبار على كفاءة هذه الطرق، وقليل منها اهتم بدراسة أثر شكل توزيع القدرة ونسبة الفقرات ذات الأداء التفاضلي معاً وبشكل متزامن على كفاءة وفاعلية طريقتي الكشف عن الأداء التفاضلي للفقرات. لذا جاءت هذه الدراسة للكشف عن فاعلية طريقتي (IRT-LR، MIMIC) في الكشف عن الأداء التفاضلي للفقرات في ضوء تغير نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة، كمحاولة لتقييم دقة وكفاءة هذه الطرق في الكشف عن الأداء التفاضلي للفقرات من خلال استخدام أسلوب المحاكاة للواقع التطبيقي لهذه الطرق. وتجدر الإشارة كذلك بأن هذه الدراسة ستختلف عن الدراسات السابقة بأخذها نسب متباينة في نسب الفقرات ذات الأداء التفاضلي للفقرات، من النسب الصغيرة وحتى النسب الكبيرة، ومستويات متباينة من شكل توزيع القدرة (طبيعي، ملتو نحو اليمين، ملتو نحو اليسار). وفي الوقت نفسه يلاحظ ندرة الدراسات التربوية العربية على المستوى العربي والمحلي التي اهتمت بتقييم فاعلية طريقتي (IRT-LR، MIMIC) والمقارنة بينهما لتحديد أيهما أفضل في الكشف عن الأداء التفاضلي للفقرات، فلم تجد الباحثة أي دراسة تناولت موضوع الدراسة الحالية.

٢. مشكلة الدراسة وأسئلتها

يسعى مطورو ومستخدمو الاختبارات إلى توفير فرص متساوية أمام المفحوصين لإبداء المعرفة والمهارات ذات الصلة بغرض الاختبار، لذلك تم تطوير عدد كبير من الطرق الإحصائية للكشف عن (DIF) لفقرات الاختبار. اهتمت بعض الدراسات بالكشف عن الأداء التفاضلي في اختبارات معينة (مبارك، ٢٠١٣). بينما اهتمت دراسات أخرى بالمقارنة بين طرق الكشف عن الأداء التفاضلي في ضوء معايير وظروف محددة كحجم العينة ونسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة (Jamali et al., 2017؛ Finch, 2005). وقد تم تناول هذه الظروف بصورة منفصلة في الدراسات سابقة للأداء التفاضلي للفقرات، وهذا التناول المنفرد لم يوفر معلومات كافية عن كفاءة الطرق المستخدمة ودقتها في الكشف عن الأداء التفاضلي للفقرات، ونظراً لأن بعض طرق الكشف عن الأداء التفاضلي للفقرات تستند إلى الإحصاء التحليلي، فمن الممكن أن تؤثر نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة على الخطأ من النوع الأول والقوة الإحصائية، لذا سيتم تناول هذه الظروف معاً بصورة متزامنة من خلال استخدام طريقتي (MIMIC) و (IRT-LR).

ونظراً لزيادة الاهتمام بطريقة (MIMIC) في الكشف عن الأداء التفاضلي للفقرات والنقص النسبي في الدراسات التي تقيم فاعليتها وكفاءتها، أشارت بعض الدراسات السابقة (Woods, 2009؛ Gao, 2019؛ Finch, 2005) إلى الحاجة لدراسة هذه الطريقة ومقارنتها مع الطرق الأخرى في الكشف عن (DIF)، وقد تم اختيار (IRT-LR) بقصد التركيز على طريقة جديدة نسبياً، وهي طريقة (MIMIC) ومقارنتها مع طريقة معروفة تنطلق من أساس نظري مختلف عن طريقة (MIMIC).

وبمراجعة الدراسات التربوية الأجنبية في هذا المجال- حسب اطلاع الباحثان- يظهر ندرة تناولها للأداء التفاضلي باستخدام طريقتي (MIMIC) و (IRT-LR) وفق هذه الظروف، حيث لم تتناول الدراسات السابقة المقارنة بين طريقتي (IRT-LR، MIMIC) ضمن نسب الفقرات ذات الأداء التفاضلي (الصغيرة، والمتوسطة، والكبيرة)، وشكل توزيع القدرة (طبيعي، ملتو نحو اليمين، ملتو نحو اليسار)، كما يلاحظ انعدامها على صعيد الدراسات التربوية العربية. لذا جاءت هذه

الدراسة للتمييز عن سابقتها لتقييم فاعلية طريقتي (IRT-LR، MIMIC) في الكشف عن الأداء التفاضلي للفقرة في ضوء تغير نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة، وتحديد أيهما أفضل في الكشف عن الأداء التفاضلي للفقرة من خلال استخدام أسلوب المحاكاة للواقع التطبيقي لهذه الطرق.

٢,١ أسئلة الدراسة

لذا حاولت هذه الدراسة على وجه التحديد الإجابة عن السؤالين الآتيين:

١. ما أثر نسبة الفقرات ذات الاداء التفاضلي، وشكل توزيع القدرة على الخطأ من النوع الاول في الكشف عن الاداء التفاضلي باستخدام طريقتي (IRT-LR، MIMIC)؟

٢. ما أثر نسبة الفقرات ذات الاداء التفاضلي، وشكل توزيع القدرة على قوة الاختبار في الكشف عن الاداء التفاضلي باستخدام طريقتي (IRT-LR، MIMIC)؟

٣. أهداف وأهمية الدراسة

٣,١ أهداف الدراسة

هدفت هذه الدراسة إلى تقييم فاعلية طريقتين من طرق الكشف عن (DIF) تستند أحد هذه الطرق إلى (SEM) وهي طريقة (MIMIC) والطريقة الأخرى تستند إلى (IRT) وهي طريقة (IRT-LR)، وتم تقييم فاعلية هذه الطرق من خلال استخدام أسلوب المحاكاة في توليد البيانات وفقاً للنموذج الثلاثي المعلمة، وكذلك تقييم الفاعلية وفقاً لمعايير الخطأ من النوع الاول وقوة الاختبار الإحصائي. كما هدفت الدراسة إلى معرفة أثر نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة على فاعلية طريقتي الكشف عن الاداء التفاضلي للفقرة، وأثر ذلك على الخطأ من النوع الاول وقوة الاختبار.

٣,٢ أهمية الدراسة

تكتسب الدراسة الحالية أهميتها النظرية من خلال محاولتها الكشف عن الطريقة المثلى للكشف عن الاداء التفاضلي عندما تكون نسب الفقرات مختلفة، وشكل توزيع القدرة (طبيعي، ملتو نحو اليمين، ملتو نحو اليسار)، وذلك بمقارنة طريقة (IRT-LR) وطريقة (MIMIC) دون ادعاء أفضلية طريقة على أخرى، بحيث حددت النتائج التي تم التوصل إليها في الدراسة مزايا كل طريقة وعيوبها.

أما من الناحية العملية، فتتمثل في فتح الأبواب للدراسات العربية التي تستخدم المحاكاة في الكشف عن الأداء التفاضلي للفقرة باستخدام طريقة (MIMIC). كما سعت الدراسة إلى توفير مصدر موثوق للباحثين في اختيار الطرق المناسبة للكشف عن الاداء التفاضلي لفقرات الاختبار الذي سيتم استخدامه وخصائص مجموعة الاختبار، وبالتالي المساعدة في تحديد طرق الكشف عن الاداء التفاضلي الأكثر موثوقية، وتجعل الاختبارات أكثر عدلاً.

٤. مصطلحات الدراسة

الخطأ من النوع الأول

هو النسبة المئوية لعدد مرات الكشف عن الاداء التفاضلي للفقرة بشكل خاطئ وبشكل إجرائي، وندعي وجود أداء تفاضلي وهو غير موجود عند مستوى دلالة ($\alpha = 0.05$)، (إجباره، ٢٠١٥).

قوة الاختبار

تمثل النسبة التي يتم فيها الكشف عن الأداء التفاضلي للفقرة بين المجموعتين المرجعية والمستهدفة في الفقرة بشكل صحيح عند مستوى الدلالة ($\alpha = 0.05$)، (أبو علام والمرايحة، ٢٠١٥).

IRT-LR

إحدى الطرق الإحصائية القائمة على (IRT)، تستخدم للكشف عن الأداء التفاضلي للفقرة، وتقوم على مقارنة نموذجين الأول يسمى القاعدي، والثاني يسمى المعزز، من خلال إيجاد المؤشر الإحصائي لاختبار نسبة الأرجحية (G^2)، (أبو علام والمرايحة، ٢٠١٤).

إحدى الطرق الإحصائية القائمة على نماذج المعادلات البنائية (SEM)، تم اقتراحها لأول مرة من قبل جورسكوج وجولد بيرغر (Jöreskog & Goldberger, 1975) تختبر هذه الطريقة العلاقة بين المتغيرات الكامنة والمتغيرات المشاهدة، كما تهدف إلى اختبار العلاقة الإحصائية بين متغير كامن غير قابل للقياس مرتبط بمؤشرات ومتغيرات مشاهدة قابلة للقياس ترتبط بدورها بمتغيرات سببية مشاهدة وقابلة للقياس (Finch, 2005).

الاداء التفاضلي للفقرة

دالة مشتقة إحصائياً لتعبر عن الفروق في احتمال استجابة الفرد الصحيحة للفقرة بين مجموعتين ممن هم في نفس المستوى من القدرة (Teresi, 2001).

نسبة الفقرات ذات الاداء التفاضلي

وهي عدد الفقرات ذات الأداء التفاضلي الموجودة في الاختبار مقسومة على عدد الفقرات الكلية للاختبار (ابوعلام، والمرابحة، ٢٠١٥).

شكل توزيع القدرة

هي الخصائص المتعلقة بالتوزيع التكراري للقدرة من حيث شكل التوزيع ووسطه الحسابي وانحرافه المعياري والتوائه وتفلطحه (بني عطا والشريفين، ٢٠١٢).

النموذج ثلاثي المعلمة

هو أحد نماذج نظرية الاستجابة للفقرة ثنائية التدرج، حيث يمكن لهذا النموذج تقدير أربع معلمات هي: قدرة الفرد، ومعامل صعوبة الفقرة، ومعامل التمييز، ومعامل تخمين الفقرة (بني عطا، ٢٠١٤).

٥. محددات الدراسة

اقتصرت الدراسة على استخدام طريقة (MIMIC) القائمة على المعادلة البنائية وطريقة (IRT-LR) القائمة على (IRT) للكشف عن (DIF).

- تحددت نسب الفقرات ذات الاداء التفاضلي للفقرات بثلاث نسب (٠٪، ٢٥٪، ٥٠٪) وشكل توزيع القدرة (طبيعي، ملتو نحو اليمين، ملتو نحو اليسار).
- استخدام النموذج ثلاثي المعلمة.
- استخدام طريقة المحاكاة في توليد البيانات.
- استخدام برنامج (R) وبرنامج (MULTILOG) في تحليل البيانات.

٦. الطريقة

٦,١ تصميم الدراسة

لتحقيق أهداف الدراسة الحالية في الكشف عن فاعلية طريقتي (MIMIC، IRTLR) للكشف عن (DIF) المنتظم باختلاف نسب الفقرات ذات الأداء التفاضلي وشكل توزيع القدرة، اتبعت الدراسة التصميم التجريبي العاملي (٣×٩) بالاعتماد على المحاكاة، حيث كانت المتغيرات المستقلة في التصميم نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة، وقد استخدمت ثلاثة نسب للأداء التفاضلي (٠٪، ٢٥٪، ٥٠٪) لتغطية المتصل المتعلق بنسب الأداء التفاضلي، وكذلك هي النسب الأكثر استخداماً في الدراسات السابقة مثل دراسة (Finch, 2005). وفي الوقت نفسه تم اختيار المتناظرات المختلفة بين ثلاث أنواع من توزيع القدرة (طبيعي، ملتو نحو اليمين، ملتو نحو اليسار) للمجموعتين المرجعية والمستهدفة وهي التوزيعات الغالبة على توزيع قدرة الأفراد.

٦,٢ توليد البيانات

ولدت البيانات باستخدام طريقة المحاكاة؛ إذ أشار ويلكوس (Wilcox, 1988) إلى أهمية دراسات المحاكاة؛ لأنها تتيح المجال للحصول على بيانات من مواقف مختلفة وتحت شروط مختلفة لإيجاد حلول لمشكلات إحصائية لا يمكن الوصول إليها بالمواقف العملية. وقد ولدت البيانات باستخدام برنامج WINGEN وهو من تصميم وإنتاج هان وهامبيلتون (Han & Hambleton, 2007)، حيث يمكن من خلال هذا البرنامج توليد استجابات لاختبارات أحادية البعد سواء كانت ثنائية التدرج أم متعددة التدرج، بالإضافة إلى استجابات متعددة الأبعاد، وتجدر الإشارة إلى أن البرنامج يؤهل المستخدم لتوليد استجابات لأكثر من مجموعة من المفحوصين بأنواع مختلفة من التوزيعات، ويمكن من خلاله تحليل الاستجابات الثنائية المولدة لإيجاد المعالم المقدرة لكل من الفرد والفقرة من خلال استخدام أحد البرامج التي تستند على (IRT). وقد ولدت عملية البيانات وفق المراحل الآتية:

المرحلة الأولى: توليد اختبار طوله ٥٠ فقرة وفق النموذج اللوجستي ثلاثي المعلمة، حيث يعد هذا الطول مناسب للاختبارات المستخدمة في الواقع العملي التي تعد في المدارس والجامعات والاختبارات الموضوعية المقننة. وقد تم توليد معالم الفقرات وفق الشروط التالية:

- ولدت معلمة الصعوبة للفقرات وفق التوزيع الطبيعي (١, ٠) $\sim$ Normal، حيث كان الوسط الحسابي والانحراف المعياري لقيم معلمة الصعوبة ٠,٥٣، ٠,٩٢ على الترتيب.
- ولدت معلمة التمييز للفقرات وفق توزيع منتظم (٢, ٠) $\sim$ Uniform وهذا المدى يعطي معاملات تمييز جيدة، وتعد قيم معلمة التمييز الحقيقية التي ولدت مماثلة للقيم الحقيقية التي استخدمها هامبيلتون وسوامينثان (Hambleton & Swaminthan, 1985) حيث أكدا على أن تكون قيم معلمة التمييز تتراوح ما بين ٠ و ٢ لوجت، وكان الوسط الحسابي والانحراف المعياري لقيم معلمة التمييز ١,١١، ٠,٤٠ على الترتيب.
- ولدت معلمة التخمين للفقرات وفق توزيع بيتا (٣٢, ٨) $\sim$ Beta وهذا التوزيع ينتج قيم لمعلمة التخمين تماثل قيم التخمين للاختبار الموضوعي (ثنائي الاستجابة) المؤلف من خمسة بدائل، حيث كان الوسط الحسابي والانحراف المعياري لقيم معلمة التخمين ٠,١٥، ٠,٠٥ على الترتيب. ويلخص الجدول (١) قيم معالم الفقرات التمييز والصعوبة والتخمين التي تم توليدها.

جدول (١): قيم معالم التمييز والصعوبة والتخمين

أرقام الفقرات	معلم التمييز	معلم الصعوبة	معلم التخمين	أرقام الفقرات	معلم التمييز	معلم الصعوبة	معلم التخمين
١	١,١	٠,٧-	٠,٢٠	٢٧	٠,٧	٠,٥	٠,٢٠
٢	٠,٧	٠,٦-	٠,٢٠	٢٨	٠,٤	٠,٤-	٠,٢٠
٣	١,٤	٠,١	٠,٢٠	٢٩	٠,٤	٠,٤-	٠,٢٠
٤	٠,٩	٠,٩	٠,١٦	٣٠	١,٢	٠,٥-	٠,٢٠
٥	١,٢	٠,٧	٠,١٢	٣١	٠,٧	١,٠-	٠,٢٠
٦	١,٦	١,١	٠,٠٦	٣٢	٠,٧	٠,٢-	٠,٢٠
٧	١,٦	١,١	٠,٠٦	٣٣	٠,٧	٠,٢-	٠,٢٠
٨	١,٦	٠,١-	٠,١٦	٣٤	٠,٥	٠,٠	٠,٢٠
٩	١,٢	٠,٥	٠,٢٠	٣٥	٠,٩	٠,٥	٠,١٤٠
١٠	٢,٠	١,٦	٠,١٦	٣٦	١,١	١,٤	٠,٠٤
١١	١,٠	١,٦	٠,١٣	٣٧	١,٢	٠,٦-	٠,٢٠
١٢	١,٥	١,٧	٠,٠٩	٣٨	١,٢	٠,٦-	٠,٢٠
١٣	١,٠	٠,٧	٠,١٥	٣٩	٠,٦	٠,٥-	٠,٢٠
١٤	١,١	٢,٠	٠,٠٦	٤٠	١,٦	٠,٣	٠,١٨
١٥	١,١	٢,٤	٠,٠٩	٤١	١,١	٠,٠	٠,٢٠
١٦	٢,٠	١,٤	٠,١١	٤٢	١,٥	٢,٠	٠,٠٦
١٧	١,٧	١,٣	٠,١٧	٤٣	١,٩	١,٩	٠,١١

١٨	٠,٥	٠,٦-	٠,٢٠	٤٤	٠,٩	٠,٥-	٠,٢٠
١٩	٠,٩	١,٦	٠,١١	٤٥	٠,٧	٠,٥-	٠,٢٠
٢٠	١,٣	٠,٤	٠,١٨	٤٦	١,٤	١,٦	٠,١١
٢١	١,١	١,٢	٠,٠٥	٤٧	١,٤	١,٦	٠,١١
٢٢	١,٢	١,١	٠,٠٥	٤٨	١,٠	١,٧	٠,٠٨
٢٣	١,٣	٠,٢	٠,٢٠	٤٩	١,٢	١,١	٠,١٥
٢٤	١,٣	٠,٢	٠,٢٠	٥٠	١,٢	١,١	٠,١٥
٢٥	٠,٥	٠,٨-	٠,٢٠				
٢٦	٠,٧	٠,٥	٠,٢٠				

المرحلة الثانية: اختيار الفقرات ذات الأداء التفاضلي للنسبة (٠٪) حيث لا يوجد فقرات ذات أداء تفاضلي، واختيار الفقرات ذات الأداء التفاضلي للنسبة (٢٥٪) بمعدل ١٣ فقرة، و (٥٠٪) من الفقرات ذات الأداء التفاضلي بما يعادل ٢٥ فقرة.

المرحلة الثالثة: ولدت استجابات ١٠٠٠ مفحوص، ويعد مثل هذا الحجم مناسباً للظروف التطبيقية للاختبارات في الواقع العملي، كما أوصى لورد (Lord, 2012) بأن يكون عدد الأفراد ١٠٠٠ للحصول على أفضل التقديرات. تم توليد الاستجابات لكل نسبة من نسب الفقرات ذات الأداء التفاضلي، ولكل توزيع من توزيعات القدرة، وبين الجدول (٢) توزيع أفراد المجموعتين (المرجعية، المستهدفة) لكل من متناظرات التوزيعات المختلفة.

جدول (٢): شكل توزيع القدرة للمجموعتين المرجعية والمستهدفة

شكل التوزيع للمجموعتين			
المرجعية	المستهدفة	المرجعية	المستهدفة
طبيعي	طبيعي	NORMAL (0,1, skewness=0)	NORMAL (0,1, skewness=0)
طبيعي	ملتو نحو اليمين	NORMAL (0,1, skewness=0)	NORMAL (0,1, skewness=2)
طبيعي	ملتو نحو اليسار	NORMAL (0,1, skewness=0)	NORMAL (0,1, skewness=-2)
ملتو نحو اليمين	طبيعي	NORMAL (0,1, skewness=2)	NORMAL (0,1, skewness=0)
ملتو نحو اليمين	ملتو نحو اليمين	NORMAL (0,1, skewness=2)	NORMAL (0,1, skewness=2)
ملتو نحو اليمين	ملتو نحو اليسار	NORMAL (0,1, skewness=2)	NORMAL (0,1, skewness=-2)
ملتو نحو اليسار	طبيعي	NORMAL (0,1, skewness=-2)	NORMAL (0,1, skewness=0)
ملتو نحو اليسار	ملتو نحو اليمين	NORMAL (0,1, skewness=-2)	NORMAL (0,1, skewness=2)
ملتو نحو اليسار	ملتو نحو اليسار	NORMAL (0,1, skewness=-2)	NORMAL (0,1, skewness=-2)

وللتحقق من شكل توزيع الاستجابات التي تم توليدها بأنها تأخذ شكل التوزيع الطبيعي أم لا. فقد قام الباحثان باختبار الأشكال السابقة لتوزيع القدرة من خلال استخدام اختبار كولموغوروف-سمرنوف (Kolmogorove-Smirnov)، وبينت نتائج اختبار كولموغوروف-سمرنوف بأن قيمه دالة إحصائية عند مستوى دلالة (٠,٠٥=α) لكل من التوزيع الملتوي نحو اليمين واليسار، مما يعني ابتعادها عن التوزيع الطبيعي في حين لم تكن قيمه دالة إحصائية للتوزيع الطبيعي.

المرحلة الرابعة: ولدت استجابات المفحوصين بواقع ١٠٠ مرة باستخدام القيم نفسها للمعالم الحقيقية التي ولدت في المرحلة الأولى، تم إجراء تغيير على قيم معلمة الصعوبة لتبدي الفقرات أداءً تفاضلياً، وذلك بإضافة ٠,٤ لقيمة معلمة الصعوبة لفقرات المجموعة المستهدفة وهي القيمة المستخدمة في أغلب الدراسات. وبذلك يكون عدد البيانات التي ولدت يساوي (٣×٩×١٠٠). وقد اختيرت عدد مرات التوليد ١٠٠ مرة؛ لأن عدد مرات التوليد التي استخدمت في دراسات المحاكاة كانت تتراوح ما بين ٣ إلى ١٠٠ مرة، وأن التقديرات كانت تستقر عندما يكون عدد مرات التوليد ٥٠ مرة فما فوق (Kamata, 1998). تم تحليل البيانات باستخدام برنامج R للكشف عن (DIF) باستخدام طريقتي (IRT LR, MIMIC)، كما وتم استخدام برنامج MULTILOG من خلال ربطه ببرنامج R وذلك للتغلب على مشكلة حاجة الحزم الإحصائية (DIFR) إلى وقت طويل لكي تعطي النتائج.

٦,٣ المعالجات الإحصائية

من أجل الإجابة عن أسئلة الدراسة، تم استخدام المعالجات الإحصائية الآتية:

- استخدام تحليل التباين الثلاثي للكشف عن أثر الظروف التجريبية على فاعلية كل من طريقي الكشف عن الأداء التفاضلي للفقرة.
- استخدام تحليل التباين الثنائي للمقارنة بين طريقي الكشف عن الأداء التفاضلي للفقرة في ضوء كل ظرف من الظروف التجريبية.

٧. نتائج الدراسة ومناقشتها

النتائج المتعلقة بالسؤال الأول الذي ينص على: ما أثر نسبة الفقرات ذات الاداء التفاضلي، وشكل توزيع القدرة على الخطأ من النوع الاول في الكشف عن الاداء التفاضلي باستخدام طريقي (MIMIC، IRTLRL)؟ تم استخدام تحليل التباين الثلاثي للكشف عن دلالة الفروق بين الأوساط الحسابية لنسب الفقرات وتوزيع القدرة وطريقي الكشف عن الأداء التفاضلي في الخطأ من النوع الأول. ويبين الجدول (٣) نتائج هذا التحليل.

جدول (٣): نتائج تحليل التباين الثلاثي للكشف عن دلالة الفروق بين متوسطات نسب الفقرات وتوزيع القدرة وطريقي الكشف عن الأداء التفاضلي في الخطأ من النوع الأول.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	الدلالة العملية 2η
نسب الفقرات	٥٠,٩٤	٢	٢٥,٤٧	٥٠٩,٤	٠,٠٠٠	٠,٠٦
توزيع القدرة	٨٤,١٦	٨	١٠,٢٥	٢٠٤,٩٠	٠	٠,٠٩
الطريقة	٢٠,٢٦	١	٢٠,٢٦	٤٠٥,٣٢	٠,٠٠٠	٠,٠١
نسب الفقرات x توزيع القدرة	١٢٢,٢٤	١٦	٧,٦٤	١٥٢,٨٤	٠,٠٠٠	٠,١٣
الطريقة x نسب الفقرات	١٨٠,٢٢	٢	٩٠,١١	١٨٠٢,١٦	٠,٠٠٠	٠,١٠
الطريقة x توزيع القدرة	٣٥١,٧٦	٨	٤٣,٩٧	٨٧٨,٣٨	٠,٠٠٠	٠,١٨
الطريقة x نسب الفقرات x توزيع القدرة	١٠٦,٣٠٤	١٦	٦,٦٤٤	١٣٢,٨٩	٠,٠٠٠	٠,٠٦
الخطأ	٤٧,٣	٩٤٦	٠,٠٥			
الكل	٩٦٣,١٨٤	٩٩٩	٠,٩٦			

تشير النتائج الواردة في الجدول (٣) بأن هناك فروقاً ذات دلالة إحصائية عند مستوى دلالة إحصائية ($\alpha=0,05$) بين المتوسطات الحسابية لمعدل الخطأ من النوع الأول تعزى لكل من نسب الفقرات ذات الاداء التفاضلي وشكل توزيع القدرة والتفاعل بينهم، وتشير النتائج كذلك بأن التفاعل بين الطريقة وتوزيع القدرة كان الأكثر تأثيراً على معدل الخطأ من النوع الأول، حيث بلغت قيمته (١٨٪)، وبينت النتائج بأن هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعدل الخطأ من النوع الأول يعزى للطريقة ولكنها كانت الأقل تأثيراً مقارنة بالشرط الأخرى حيث بلغت قيمتها (١٪).

جدول (٤): الوسط الحسابي للخطأ من النوع الأول لطريقي (MIMIC) و (IRTLRL) تبعاً لمتغيرات الدراسة

المرجعية	توزيع القدرة	النسبة %	IRTLRL	MIMIC
ملتو نحو اليسار	المستهدفة	٠	٠,٠٦٠	٠,٠٥٣
ملتو نحو اليسار	ملتو نحو اليسار	٢٥	٠,٠٧٤	٠,١٢٩
ملتو نحو اليسار	طبيعي	٥٠	٠,٣٣٠	٠,٣٤١
ملتو نحو اليسار	طبيعي	٠	٠,٠٤٠	٠,٠٤٨
ملتو نحو اليسار	طبيعي	٢٥	٠,٠٨١	٠,٢١٩
ملتو نحو اليسار	طبيعي	٥٠	٠,٣٧٣	٠,٤١٠
ملتو نحو اليسار	ملتو نحو اليمين	٠	٠,٠٦٤	٠,٠٤٠

٢٥	٠,١٣٨	٠,٣٢٢	طبيعي	طبيعي
٥٠	٠,٣٧٧	٠,٤٨٦		
٠	٠,٠٠١	٠,٠٢٣		
٢٥	٠,١١١	٠,٣٠٢		
٥٠	٠,٣٦١	٠,٣٥٠		
٠	٠,٠١٨	٠,٠٤٦	ملتو نحو اليمين	طبيعي
٢٥	٠,١٢٢	٠,٢٨٩		
٥٠	٠,٤٠٨	٠,٣٩٦		
٠	٠,٠١٦	٠,٠٥٨		
٢٥	٠,١٢٣	٠,٣٧٢		
٥٠	٠,٣١٢	٠,٣٠٥	ملتو نحو اليسار	طبيعي
٠	٠,١٦٩	٠,٠٥٥		
٢٥	٠,١٠٧	٠,٣١١		
٥٠	٠,٢٩٣	٠,٣٠٢		
٠	٠,١٧٨	٠,٠٦٢		
٢٥	٠,١٢٨	٠,١٧٩	ملتو نحو اليمين	طبيعي
٥٠	٠,٤١٥	٠,٣٠٨		
٠	٠,١٦٤	٠,٠٤٤		
٢٥	٠,١٤٩	٠,١٣٣		
٥٠	٠,٥٢٧	٠,٣٤٩		

يتضح من النتائج الواردة في الجدول (٤) تبايناً ملحوظاً لقيم معدلات الخطأ من النوع الأول لطريقة (MIMIC) لنسبة الفقرات ذات الأداء التفاضلي وشكل توزيع القدرة، حيث كان أعلى معدل للخطأ من النوع الأول عندما كانت نسبة الفقرات ذات الأداء التفاضلي تساوي ٥٠٪ وتوزيع القدرة ملتو نحو اليسار للمجموعة المرجعية وملتو نحو اليمين للمجموعة المستهدفة وبلغت قيمته (٠,٤٨٦)، بينما كان أقل معدل للخطأ من النوع الأول عندما كانت نسبة الفقرات ذات الأداء التفاضلي تساوي ٠٪، وتوزيع القدرة طبيعي للمجموعتين المرجعية المستهدفة وبلغت قيمته (٠,٠٢٣). أما بالنسبة لطريقة (IRTLR) فقد أظهرت النتائج الواردة في الجدول نفسه تبايناً في معدلات الخطأ من النوع الأول، حيث كان أعلى معدل للخطأ من النوع الأول عندما كانت نسبة الفقرات ذات الأداء التفاضلي تساوي ٥٠٪ وتوزيع القدرة ملتو نحو اليمين للمجموعة المرجعية وملتو نحو اليمين للمجموعة المستهدفة وبلغت قيمته (٠,٥٢٧)، بينما كان أقل معدل للخطأ من النوع الأول عندما كانت نسبة الفقرات ذات الأداء التفاضلي تساوي ٠٪ وتوزيع القدرة طبيعي للمجموعتين المرجعية والمستهدفة وبلغت قيمته (٠,٠٠١).

ولتحديد تأثير كل عامل من عوامل الدراسة على الخطأ من النوع الأول لطريقتي (MIMIC، IRTL) قام الباحثان بدراسة كل عامل على حده.

نسبة الفقرات ذات الأداء التفاضلي

لدراسة أثر نسبة الفقرات ذات الأداء التفاضلي (صفر%، ٢٥٪، ٥٠٪)، وطريقتي الكشف عن الأداء التفاضلي (MIMIC، IRTL) والتفاعل بينهما في الخطأ من النوع الأول، استخدم الباحثان تحليل التباين الثنائي، وجاءت النتائج كما يوضحها جدول (٥).

جدول (٥): نتائج تحليل التباين الثنائي للكشف عن دلالة الفروق بين متوسطات الخطأ من النوع الأول حسب متغير نسبة الفقرات ذات الأداء التفاضلي

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	الدلالة العملية 2η
نسبة الفقرات ذات الأداء التفاضلي	١٨٨,١٣٤	٢	٩٤,٠٦٧	٧٢٣٥,٩٢	٠,٠٠٠	٠,٨٣

طريقة الكشف عن الأداء التفاضلي	٢,٤٥٤	١	٢,٤٥٤	١٨٨,٧٣	٠,٠٠٠	٠,٠٠
الطريقة x نسبة الفقرات الخطأ	٢١,٨١٨	٢	١٠,٩٠٩	٨٣٩,١٣	٠,٠٠٠	٠,٠٤
الخطأ	١٢,٩٢٢	٩٩٤	٠,٠١٣			
الكلي	٢٢٥,٣٢٨	٩٩٩	٠,٢٢٦			

ولتحديد دلالة الفروق بين الطريقتين قام الباحثان بمقارنة الفروق بين الطريقتين عند كل نسبة من نسب الفقرات ذات الأداء التفاضلي، وذلك باستخدام اختبارات لعينتين مرتبطتين، وجاءت النتائج كما يوضحها جدول (٦).

جدول (٦): اختبارات لعينتين مرتبطتين للفروق بين طريقتي الكشف عن الأداء التفاضلي المنتظم عند كل نسبة من نسب الفقرات ذات الأداء التفاضلي

نسب الفقرات	طريقة الكشف عن الفقرات ذات الأداء التفاضلي	المتوسط	درجة الحرية	قيمة ت	مستوى الدلالة
صفر %	نسبة الأرجحية MIMIC	٠,٠٨ ٠,٠٥	٣٥٩٩	١١,٧٩	٠,٠٠٠
٢٥ %	نسبة الأرجحية MIMIC	٠,١٨ ٠,٢٩	٣٥٩٩	٣٥,٩	٠,٠٠٠
٥٠ %	نسبة الأرجحية MIMIC	٠,٤٤ ٠,٤٢	٣٥٩٩	٧,٥٨	٠,٠٠٠

يتبين من جدول (٥) بأن هناك فروق ذات دلالة إحصائية عند مستوى دلالة إحصائية ($\alpha=0.05$) بين الأوساط الحسابية لمعدل الخطأ من النوع الأول تعزى لنسبة الفقرات ذات الأداء التفاضلي وطريقة الكشف عن الأداء التفاضلي والتفاعل بينهما.

وأظهرت نتائج الجدول (٦) بأن هناك فروقاً دالة إحصائية بين الطريقتين عند كل نسبة من نسب الفقرات ذات الأداء التفاضلي لصالح طريقة (MIMIC) عندما كانت نسبة الفقرات ذات الأداء التفاضلي (٠,٠ %، ٥٠ %)، بمعنى أن أداء طريقة (MIMIC) كان أفضل من أداء طريقة (IRTLR) عندما كانت نسبة الفقرات ذات الأداء التفاضلي (٠,٠ %، ٥٠ %)، بينما كان أداء طريقة (IRTLR) أفضل مقارنة بأداء طريقة (MIMIC) عندما كانت نسبة الفقرات ذات الأداء التفاضلي (٢٥ %). وكما يتضح من جدول ٦ أن معدلات الخطأ من النوع الأول تزداد بزيادة نسبة الفقرات ذات الأداء التفاضلي وذلك لكلا الطريقتين. وربما يعود ذلك إلى تأثير معدل الخطأ من النوع الأول بدرجة كبيرة بنسبة الفقرات ذات الأداء التفاضلي، ووجود فقرات ذات أداء تفاضلي في الاختبار يفترض ضمناً أن بنية السمة المقاسة تختلف من مجموعة لأخرى، وهو ما يؤثر بالتالي على دقة الكشف عن الأداء التفاضلي للفقرة لكلا الطريقتين، مما يعني ضرورة فحص الاختبار وحذف الفقرات ذات الأداء التفاضلي قبل استخدام أي من الطريقتين. ومثل هذه النتيجة جاءت متفقة إلى حد كبير مع نتائج الدراسات (Finch, 2005؛ woods, 2009).

شكل توزيع القدرة

ولمعرفة أثر شكل توزيع القدرة للمجموعتين المرجعية والمستهدفة باختلاف مستوياته وطريقتي الكشف عن الأداء التفاضلي والتفاعل بينهما، قام الباحثان باستخدام تحليل التباين الثنائي، وجاءت النتائج كما يوضحها جدول (٧).

جدول (٧): تحليل التباين الثنائي للكشف عن دلالة الفروق بين الأوساط الحسابية للخطأ من النوع الأول حسب متغير شكل توزيع القدرة

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	الدلالة العملية 2η
شكل توزيع القدرة	١٥,٩٦	٨	١,٩٩٥	٣٣,٢٥	٠,٠٠٠	٠,٠٥
طريقة الكشف عن الأداء التفاضلي	١٣,٢٧٧	١	١٣,٢٧٧	٢٢١,٢٩	٠,٠٠٠	٠,٠٤
الطريقة x شكل توزيع القدرة	٢٣٠,١٩٢	٨	٢٨,٧٧٤	٤٧٩,٥٦	٠,٠٠٠	٠,٧٢
الخطأ	٥٨,٩٢	٩٨٢	٠,٠٦			
الكلي	٣١٨,٣٤٩	٩٩٩	٠,٣١			

ولتحديد دلالة الفروق بين الطريقتين قام الباحثان بمقارنة الفروق بين الطريقتين عند كل توزيع من توزيعات القدرة للمجموعتين المرجعية والمستهدفة، وذلك باستخدام اختبارات لعينتين مرتبطتين، وجاءت النتائج كما يوضحها جدول (٨).
جدول (٨): اختبارات لعينتين مرتبطتين للفروق بين طريقتي الكشف عن الأداء التفاضلي المنتظم لكل توزيع من توزيعات القدرة.

شكل توزيع القدرة	طريقة الكشف عن الفقرات ذات الأداء التفاضلي	المتوسط	درجة الحرية	قيمة ت	مستوى الدلالة
ملتو نحو اليسار / ملتو نحو اليسار	نسبة الأرجحية	٠,١٧	١١٩٩	٤,١٢	٠,٠٠٠
	MIMIC	٠,١٨			
ملتو نحو اليسار / طبيعي	نسبة الأرجحية	٠,١٨	١١٩٩	٢٨,٩١	٠,٠٠٠
	MIMIC	٠,٢٧			
ملتو نحو اليسار / ملتو نحو اليمين	نسبة الأرجحية	٠,٢٢	١١٩٩	٢٤,٥١	٠,٠٠٠
	MIMIC	٠,٣٢			
طبيعي / طبيعي	نسبة الأرجحية	٠,١١	١١٩٩	٢١,٣١	٠,٠٠٠
	MIMIC	٠,١٧			
طبيعي / ملتو نحو اليمين	نسبة الأرجحية	٠,٢٤	١١٩٩	١١,٧٤	٠,٠٠٠
	MIMIC	٠,٢٩			
طبيعي / ملتو نحو اليسار	نسبة الأرجحية	٠,٢٧	١١٩٩	٢٢,٠٦	٠,٠٠٠
	MIMIC	٠,٢٨			
ملتو نحو اليمين / ملتو نحو اليسار	نسبة الأرجحية	٠,٢٤	١١٩٩	٦,٠٦	٠,٠٠٠
	MIMIC	٠,٢٧			
ملتو نحو اليمين / طبيعي	نسبة الأرجحية	٠,٢٩	١١٩٩	١٧,٢٩	٠,٠٠٠
	MIMIC	٠,٢٢			
ملتو نحو اليمين / ملتو نحو اليمين	نسبة الأرجحية	٠,٢١	١١٩٩	٣٢,١٤	٠,٠٠٠
	MIMIC	٠,١٩			

تشير النتائج الواردة في جدول (٧) بأن هناك فروقاً ذات دلالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين المتوسطات الحسابية لمعدل الخطأ من النوع الأول تعزى لشكل توزيع القدرة وطريقة الكشف عن الأداء التفاضلي والتفاعل بينهما. وتشير النتائج كذلك بأن التفاعل بين توزيع القدرة وطريقة الكشف عن الأداء التفاضلي قد أسهم في تباين قيم معدلات الخطأ من النوع الأول أكثر من شكل توزيع القدرة، وطريقة الكشف عن الأداء التفاضلي.

تشير النتائج الواردة في جدول (٨) بأن هناك فروقاً ذات دلالة إحصائية عند مستوى دلالة ($\alpha=0.05$) بين المتوسطات الحسابية لمعدل الخطأ من النوع الأول لطريقتي الكشف عن الأداء التفاضلي تبعاً لكل توزيع من توزيعات القدرة للمجموعتين المرجعية والمستهدفة. وبينت النتائج أن الوسط الحسابي لمعدلات الخطأ من النوع الأول لطريقة (IRTLR) كان أقل من الوسط الحسابي لمعدلات الخطأ من النوع الأول لطريقة (MIMIC) تبعاً لكل توزيع من توزيعات القدرة باستثناء عندما كان توزيع القدرة ملتو نحو اليمين للمجموعة المرجعية و (ملتو نحو اليمين و طبيعي) للمجموعة المستهدفة. وبينت النتائج أن الخطأ من النوع الأول كان أقل من ($\alpha=0.05$) لطريقتي الكشف عن الأداء التفاضلي عندما كان توزيع القدرة طبيعي للمجموعتين المرجعية والمستهدفة، وتضخم الخطأ من النوع الأول بشكل أكبر عندما كان توزيع القدرة ملتوياً لإحدى المجموعتين، وربما يعود السبب في ذلك إلى أن توزيع القدرة الغير متساوي في المجموعتين يؤدي إلى التضخم في الخطأ من النوع الأول للفرضية الصفرية المتعلقة بعدم وجود أداء تفاضلي للفقرة، مما يدل على أن تغيير شكل توزيع القدرة يؤثر على طريقتي الكشف عن الأداء التفاضلي للفقرة. وجاءت هذه النتيجة متفقة مع نتائج دراسات (Jamali, Woods, 2008)؛ (2017).

نتائج الإجابة عن السؤال الثاني: للإجابة عن السؤال الثاني الذي نصه "ما أثر نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة على قوة الاختبار في الكشف عن الأداء التفاضلي المنتظم باستخدام طريقتي (MIMIC، و IRTLRL)؟ استخدم الباحثان تحليل التباين الثلاثي للكشف عن دلالة الفروق بين متوسطات نسب الفقرات وتوزيع القدرة، وطريقتي الكشف عن الأداء التفاضلي في قوة الاختبار وقد جاءت النتائج كما يوضحها جدول (٩).

جدول (٩): نتائج تحليل التباين الثلاثي للكشف عن دلالة الفروق بين متوسطات نسب الفقرات وتوزيع القدرة وطريقتي الكشف عن الأداء التفاضلي في قوة الاختبار

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	الدلالة العملية 2η
نسب الفقرات ١	٧٠,٩٢	١	٧٠,٩٢	٢٧٢,٧٧	٠,٠٠٠	٠,٠٤
توزيع القدرة	١٦٧,٥٤٤	٨	٢٠,٩٤	٨٠,٥٥	٠,٠٠٠	٠,١١
الطريقة	٧٧٧,٤١	١	٧٧٧,٤١	٢٩٩٠,٠٤	٠,٠٠٠	٠,٥٣
نسب الفقرات x توزيع القدرة	٢٤,٨	٨	٣,١٠	١١,٩١	٠,٠٠٠	٠,٠١
الطريقة x نسب الفقرات	٢٧,٨٨	١	٢٧,٨٨	١٠٧,٢٤	٠,٠٠٠	٠,٠١
الطريقة x توزيع القدرة	٧٨,٨٦	٨	٩,٨٥	٣٧,٨٧	٠,٠٠٠	٠,٠٥
الطريقة x نسب الفقرات x توزيع القدرة	٦٠,١٦	٨	٧,٥٢	٢٨,٩٣	٠,٠٠٠	٠,٠٤
الخطأ	٢٥٠,٦٤	٩٦٤	٠,٢٦			
الكل	١,٤٥٨,٢١٤	٩٩٩	١,٤٦			

كشفت النتائج الواردة (٩) بأن هناك فروقاً ذات دلالة إحصائية عند مستوى ($\alpha = 0.05$) بين المتوسطات الحسابية لقوة الاختبار تعزى لكل من نسبة الفقرات ذات الأداء التفاضلي، وشكل توزيع القدرة والتفاعل بينهم، وتشير النتائج بأن الطريقة كانت أكثر تأثيراً في قوة الاختبار، حيث أثرت بـ (٥٣٪) في قوة الاختبار. ويوضح جدول (١٠) الأوساط الحسابية لقوة الاختبار لطريقتي (MIMIC، IRTLRL) تبعاً لجميع متغيرات الدراسة.

جدول (١٠): الوسط الحسابي لقوة الاختبار لطريقتي (MIMIC، IRTLRL) تبعاً لجميع متغيرات الدراسة

MIMIC	IRTLRL	النسبة %	توزيع القدرة	
			المستهدفة	المرجعية
٠,٧٠٩	٠,٦٠٠	٢٥	ملتو نحو اليسار	ملتو نحو اليسار
٠,٦٩٠	٠,٥٧٠	٥٠		
٠,٦٦٦	٠,٥٦٥	٢٥	طبيعي	ملتو نحو اليسار
٠,٦٢٧	٠,٥٨٧	٥٠		
٠,٦٣٨	٠,٥٣٢	٢٥	ملتو نحو اليمين	ملتو نحو اليسار
٠,٦٢٤	٠,٥٥٤	٥٠		
٠,٨٨٨	٠,٨٩٣	٢٥	طبيعي	طبيعي
٠,٦٥٠	٠,٥٩٧	٥٠		
٠,٦٤٣	٠,٥٥٩	٢٥	ملتو نحو اليمين	طبيعي
٠,٦٢٢	٠,٥٧٢	٥٠		
٠,٧١٥	٠,٦١٢	٢٥	ملتو نحو اليسار	طبيعي
٠,٧٢٤	٠,٥٧١	٥٠		
٠,٧٢٨	٠,٥٩٠	٢٥	ملتو نحو اليسار	ملتو نحو اليمين
٠,٧٢٩	٠,٥٦١	٥٠		
٠,٧١٥	٠,٥٨٨	٢٥	طبيعي	ملتو نحو اليمين
٠,٦٧٨	٠,٥٩٩	٥٠	ملتو نحو اليمين	ملتو نحو اليمين
٠,٦٤٤	٠,٦١١	٢٥		

٠,٦٧٨	٠,٦١٧	٥٠
-------	-------	----

يتضح من النتائج الواردة في جدول (١٠) بأن هناك اختلافات قليلة لقيم المتوسطات الحسابية لطريقة (MIMIC) تبعاً لنسبة الفقرات ذات الأداء التفاضلي وشكل توزيع القدرة، حيث كان أعلى متوسط حسابي لقوة الاختبار عندما كانت نسبة الفقرات ذات الأداء التفاضلي تساوي ٢٥٪ وتوزيع القدرة طبيعي للمجموعتين المرجعية والمستهدفة وبلغت قيمته (٠,٨٨٨). أما بالنسبة لطريقة (IRTLR) كان أعلى متوسط حسابي لقوة الاختبار عندما كانت نسبة الفقرات ذات الأداء التفاضلي تساوي ٢٥٪ وتوزيع القدرة طبيعي للمجموعتين المرجعية والمستهدفة وبلغت قيمته (٠,٨٩٣).

ولتحديد تأثير كل عامل من عوامل الدراسة على قوة الاختبار لطريقتي (IRTLR، MIMIC) قام الباحثان بدراسة كل عامل على حده.

نسبة الفقرات ذات الأداء التفاضلي

جدول (١١): تحليل التباين الثنائي للكشف عن دلالة الفروق بين متوسطات نسب الفقرات ذات الأداء التفاضلي وطريقة الكشف عن الأداء التفاضلي في قوة الاختبار

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الدلالة الإحصائية	الدلالة العملية 2η
نسب الفقرات ذات الأداء التفاضلي	١,٢٠١	١	١,٢٠١	٤٠,٠٤	٠,٠٠٠	٠,٠١
طريقة الكشف عن الأداء التفاضلي	٧٨,٨٩٩	١	٧٨,٨٩٩	٢٦٢٩,٩٦	٠,٠٠٠	٠,٦٩
الطريقة x نسب الفقرات	٢,٨٣٠	١	٢,٨٣٠	٩٤,٣٣	٠,٠٠٠	٠,٠٢
الخطأ	٢٩,٨٨	٩٩٦	٠,٠٣			
الكل	١١٢,٨١	٩٩٩	٠,١١٢			

لمعرفة دلالة الفروق بين الطريقتين قام الباحثان بمقارنة الفروق بين الطريقتين عند كل نسبة من نسب الفقرات ذات الأداء التفاضلي وذلك باستخدام اختبار ت لعينتين مرتبطتين، وجاءت النتائج كما يوضحها جدول (١٢).

جدول (١٢): اختبار ت لعينتين مرتبطتين للفروق بين طريقتي الكشف عن الأداء التفاضلي المنتظم لكل نسبة من نسب الفقرات ذات الأداء التفاضلي

نسب الفقرات ذات الأداء التفاضلي	طريقة الكشف عن الفقرات ذات الأداء التفاضلي	المتوسط	درجة الحرية	قيمة ت	مستوى الدلالة
٢٥٪	نسبة الأرجحية	٠,٦٦	٣٥٩٩	٤٤,٥٨	٠,٠٠٠
	MIMIC	٠,٧٤			
٥٠٪	نسبة الأرجحية	٠,٦٥	٣٥٩٩	٢٨,٥	٠,٠٠٠
	MIMIC	٠,٧٠			

بينت النتائج الواردة في جدول (١١) بأن هناك فروق ذات دلالة إحصائية عند مستوى $(\alpha=0,05)$ بين الاوساط الحسابية لقوة الاختبار تعزى لنسبة الفقرات ذات الأداء التفاضلي وطريقة الكشف عن الأداء التفاضلي والتفاعل بينهما، وتشير النتائج كذلك بأن طريقة الكشف عن الأداء التفاضلي قد أثرت في قيم قوة الاختبار أكثر من نسبة الفقرات ذات الأداء التفاضلي والتفاعل بينهما.

يتضح من جدول (١٢) أن هناك فروقاً ذات دلالة إحصائية عند مستوى دلالة $(\alpha=0,05)$ بين الاوساط الحسابية لقوة اختبار طريقتي (IRTLR، MIMIC) تعزى لنسب الفقرات ذات الأداء التفاضلي، مما يظهر تأثير قوة الاختبار بنسب الفقرات ذات الأداء التفاضلي، وقد يعزى ذلك إلى زيادة نسبة الفقرات ذات الأداء التفاضلي في الاختبار والتي وصلت إلى (٥٠٪) من الفقرات، حيث أن زيادة نسبة الفقرات ذات الأداء التفاضلي أدت إلى انخفاض قوة الاختبار لكلا طريقتي الكشف عن الأداء التفاضلي، كما أن وجود فقرات ذات أداء تفاضلي في الاختبار سيحد من صدقه الداخلي وثباته، وينتهك خاصيتي أحادية البعد

واللاتغابر (invariance)، لذلك في حالة التخلص منها سيميل الاختبار بفقراته إلى أحادية البعد، وبالتالي ضمان عدالته. وجاءت نتائج هذه الدراسة متفقة مع نتائج دراسة فينش (Finch, 2005).

توزيع القدرة

لدراسة أثر توزيع القدرة للمجموعتين المرجعية والمستهدفة وطريقتي الكشف عن (MIMIC، IRTL) والتفاعل بينهما على قوة الاختبار، استخدم الباحثان تحليل التباين الثنائي. ويظهر الجدول (١٣) نتائج هذا التحليل.

جدول (١٣): تحليل التباين الثنائي للكشف عن دلالة الفروق بين متوسطات شكل توزيع القدرة وطريقة الكشف عن الأداء التفاضلي في قوة الاختبار

الدلالة العملية 2η	الدلالة الإحصائية	قيمة F	متوسط المربعات	درجة الحرية	مجموع المربعات	مصدر التباين
٠,٠٨	٠,٠٠٠	١١,٩٠	٠,٧١٤	٨	٥,٧١٢	توزيع القدرة
٠,٦٥	٠,٠٠٠	٦٩١,٨٠	٤١,٥٠٨	١	٤١,٥٠٨	طريقة الكشف عن الأداء التفاضلي
٠,٢	٠,٠٠٠	٣٤,٠٩	٢,٠٤٥	٨	١٦,٣٦٠	الطريقة x توزيع القدرة
			٠,٠٦	٩٨٢	٠,٠٥٨	الخطأ
			٠,٠٦	٩٩٩	٦٣,٦٣٨	الكل

لتحديد دلالة الفروق بين الطريقتين، قام الباحثان بمقارنة الفروق بين الطريقتين عند كل توزيع من توزيعات القدرة للمجموعتين المرجعية والمستهدفة، وذلك باستخدام اختبار ت لعينتين مرتبطتين، وجاءت النتائج كما يوضحها جدول (١٤).

جدول (١٤): اختبار ت لعينتين مرتبطتين للفروق بين طريقتي الكشف عن الأداء التفاضلي المنتظم لكل توزيع من توزيعات القدرة للمجموعتين المرجعية والمستهدفة

مستوى الدلالة	قيمة ت	درجة الحرية	المتوسط	طريقة الكشف عن الفقرات ذات الأداء التفاضلي	توزيع القدرة
٠,٠٠٠	٢٥,٤١	٧٩٩	٠,٦٥	نسبة الأرجحية	ملتو نحو اليسار/ملتو نحو اليسار
			٠,٧٤	MIMIC	
٠,٠٠٠	١٠,٧	٧٩٩	٠,٦٤	نسبة الأرجحية	ملتو نحو اليسار /طبيعي
			٠,٦٨	MIMIC	
٠,٠٠٠	١٧,٥٥	٧٩٩	٠,٦٢	نسبة الأرجحية	ملتو نحو اليسار / ملتو نحو اليمين
			٠,٦٨	MIMIC	
٠,٠٠٠	١٣,٩٥	٧٩٩	٠,٨١	نسبة الأرجحية	طبيعي/طبيعي
			٠,٧٩	MIMIC	
٠,٠٠٠	١٣,٨٩	٧٩٩	٠,٦٣	نسبة الأرجحية	طبيعي/موجب
			٠,٦٩	MIMIC	
٠,٠٠٠	٢٦,٧٦	٧٩٩	٠,٦٦	نسبة الأرجحية	طبيعي/ ملتو نحو اليسار
			٠,٧٦	MIMIC	
٠,٠٠٠	٢٨,٧٩	٧٩٩	٠,٦٦	نسبة الأرجحية	ملتو نحو اليمين / ملتو نحو اليسار
			٠,٧٧	MIMIC	
٠,٠٠٠	١٦,٧٦	٧٩٩	٠,٦٧	نسبة الأرجحية	ملتو نحو اليمين /طبيعي
			٠,٧٤	MIMIC	
٠,٠٠٠	٨,٢٣	٧٩٩	٠,٦٨	نسبة الأرجحية	ملتو نحو اليمين/ ملتو نحو اليمين
			٠,٧٢	MIMIC	

بينت النتائج الواردة في جدول (١٣) بأن هناك فروق ذات دلالة إحصائية عند مستوى $(\alpha=0.05)$ بين الاوساط الحسابية لقوة الاختبار تعزى لتوزيع القدرة، وطريقة الكشف عن الأداء التفاضلي والتفاعل بينهما، وتشير النتائج كذلك بأن طريقة الكشف عن الأداء التفاضلي قد أسهمت في تباين قيم قوة الاختبار أكثر من شكل توزيع القدرة والتفاعل بينهما.

تشير النتائج الواردة في جدول (١٤) بأن هناك فروقاً ذات دلالة إحصائية عند مستوى دلالة $(\alpha=0.05)$ بين المتوسطات الحسابية لقوة اختبار طريقتي الكشف عن الأداء التفاضلي تبعاً لكل توزيع من توزيعات القدرة للمجموعتين المرجعية والمستهدفة. وبينت النتائج أن الوسط الحسابي لقوة اختبار طريقة (IRTLR) كانت أقل من الوسط الحسابي لقوة اختبار طريقة (MIMIC) لكل توزيع من توزيعات القدرة، مما يعني أداء طريقة (MIMIC) أفضل من أداء طريقة (IRTLR) تبعاً لكل توزيع من توزيعات القدرة وفقاً لمحك قوة الاختبار.

٨. التوصيات

في ضوء ما توصلت إليه الدراسة من نتائج يوصي الباحثان بما يلي:

١. إجراء دراسات عن الأداء التفاضلي لفقرات من النوع المقالي - متعدد الاستجابة - باستخدام طريقتي الكشف عن الأداء التفاضلي للفقرات (IRTLR، MIMIC).
٢. استخدام طريقتي الكشف عن الأداء التفاضلي للفقرات لقدرتهما على ضبط معدلات الخطأ من النوع الأول، عندما تكون نسبة الفقرات ذات الأداء التفاضلي $(\%)$ ، وشكل توزيع القدرة طبيعي للمجموعتين المرجعية والمستهدفة.
٣. إجراء المزيد من الدراسات من خلال بيانات حقيقية باستخدام طرق الكشف عن الأداء التفاضلي المستخدمة في الدراسة للتوصل إلى مزيد من النتائج.
٤. إجراء دراسات مماثلة تأخذ بعين الاعتبار عوامل أخرى قد تلعب دوراً في تأثيرها على فاعلية طريقتي الكشف المستخدمة في الدراسة مثل: حجم الأثر، حجم الأداء التفاضلي.
٥. إعادة إجراء الدراسة على النماذج الثنائية الأخرى (أحادي المعلمة، وثنائي المعلمة) وتحت نفس الشروط.

بيان تضارب المصالح

يقر جميع المؤلفين أنه ليس لديهم أي تضارب في المصالح.

المراجع

- إجباره، محمد. (٢٠١٥). *أثر نسبة الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول وقوة الاختبار في الكشف عنه باستخدام طريقتي الانحدار اللوجستي ونسبة الأرجحية العظمى* (رقم المنشور ٧٥٩٤٧٠) [أطروحة دكتوراه منشورة، جامعة اليرموك]. قاعدة بيانات دار المنظومة، الرسائل الجامعية.
- بني عطا، زايد. (٢٠١٤). *تقصي دقة تقدير النموذج اللوجستي ثلاثي المعلمة لمعالم الفقرة وقدرة الأفراد في ضوء تغير طول الاختبار وحجم العينة: دراسة محاكاة*. مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية، ١١ (٢)، ٣٧-١.
- بني عطا، زايد، والشريفين، نضال. (٢٠١٢). *أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار*. المجلة الأردنية في العلوم التربوية، ١٨ (٢)، ١٥٥-١٦٦.
- علام، محمود، والمراوحة، عامر. (٢٠١٤). *أثر طول الاختبار وحجم العينة على تضخم نسبة الخطأ من النوع الأول لإحصائيات طرق الكشف عن دالة المفردة التفاضلية: دراسة محاكاة*. المؤسسة العربية للاستشارات العلمية وتنمية الموارد البشرية، ١٥ (٤٧)، ١-١٢٤.
- علام، محمود، والمراوحة، عامر. (٢٠١٥). *أثر نسبة المفردات ذات الأداء التفاضلي وحجم العينة على قوة وفاعلية اختبار تحيز المفردة المتزامن: دراسة محاكاة*. المجلة العربية للعلوم الاجتماعية، ٢ (٧)، ١٣٩-١٦٥.
- عودة، أحمد. (٢٠٠٥). *القياس والتقويم في العملية التدريسية*. دار الأمل.

اللبيدي، نزار. (٢٠٠٨). مقارنة أربع طرائق للكشف عن الدالة المفردة التفاضلية: دراسة محاكاة [أطروحة دكتوراة غير منشورة]. الجامعة الأردنية.

مبارك، مأمون. (٢٠١٣). الأداء التفاضلي لاختبار الرياضيات في الدراسة الدولية بيزا ٢٠٠٦. *مجلة العلوم التربوية*، ٢٥ (٢)، ٢٤١-٢٦١.

References

- Dorans, N., & Holland, P. (1993). DIF detection and description: Mantel-Haenszel and standardization. *ETS Research Report Series, 1992*(1), i-40.
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement, 29*(4), 278-295.
- Gao, X. (2019). *A comparison of six DIF detection methods* [Unpublished master's thesis]. University of Connecticut.
- Greer, T. G. (2004). *Detection of differential item functioning (DIF) on the SATV: A comparison of four methods: Mantel-Haenszel, logistic regression, simultaneous item bias, and likelihood ratio test* [Unpublished doctoral dissertation]. University of Houston.
- Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer Science & Business Media.
- Hancock, G. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*(3), 373-388.
- Han, K. T., & Hambleton, R. K. (2007). *User's manual for WinGen: Windows software that generates IRT model parameters and item responses*. Center for Educational Assessment Research Report, 642, 516-524.
- Jamali, J., Ayatollahi, S., & Jafari, P. (2017). The effect of small sample size on measurement equivalence of psychometric questionnaires in MIMIC model: A simulation study. *Bio Med Research International*.
- Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351), 631-639.
- Kamata, A. (1998). *Some generalization of the Rasch model: An application of the hierarchical generalized linear model* [Unpublished doctoral dissertation]. Michigan State University, Ann Arbor.
- Lopez, G. E. (2012). *Detection and classification of DIF types using parametric and nonparametric methods: A comparison of the IRT-likelihood ratio test, crossing-SIBTEST, and logistic regression procedures* (Publication No. 3506081) [Doctoral dissertation]. University of South Florida, ProQuest Dissertations & Theses Global.
- Lord, M. (2012). *Applications of item response theory to practical testing problems*. Routledge.
- Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes, 4*(5), 1-22.

- Pare, T., & Park, P. (2006). Examining the relationship between differential item functioning and differential test functioning. *Language Testing*, 23(4), 475–496.
- Roever, C. (2005). That's not fair! Fairness, bias and differential item functioning in language testing. *SLS Brownbag*, 9(15), 1-14.
- Teresi, J. (2001). Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical, and mental health. *Journal of Mental Health and Aging*, 7(1), 31-40.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, 99(1), 118.
- Wang, N., & Lane, S. (1996). Detection of gender-related differential item functioning in a mathematics performance assessment. *Applied Measurement in Education*, 9(2), 175-199.
- Wilcox, R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology*, 41(1), 109-117.
- Woods, C. (2008). IRT-LR-DIF with estimation of the focal-group density as an empirical histogram. *Educational and Psychological Measurement*, 68(4), 571-586.
- Woods, C. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two group analysis. *Multivariate Behavioral Research*, 44(1), 1-27.
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5), 339-361.
- Zimbra, D. J. (2018). *An examination of the MIMIC method for detecting DIF and comparison to the IRT likelihood ratio and Wald tests* [Doctoral dissertation]. University of Hawaii at Mānoa.