OPEN ACCESS
تاريخ الاستلام: 29/ 09/ 2024م
تاريخ القبـول: 23/ 11/ 2024م

مجلة الآداب
للدراسات النفسية والتربوية

أثر إحصائي المطابقة وأقصى معلومات والخطأ المعياري في تقدير الصعوبة على الأداء التفاضلي للفقرة باستخدام النموذج الثلاثي المعلم 3PLM

أ .محمد محمد عتيق الخضر*
Email: alkhader2011@gmail.com

أ.د. إسماعيل سلامة البرصان**
Email: ibursan@ksu.edu.sa

الملخص:

هدفت الدراسة إلى تقصي أثر إحصائي المطابقة وأقصى معلومات والخطأ المعياري في تقدير الصعوبة باستخدام النموذج الثلاثي المعلم (3PLM) على الأداء التفاضلي للفقرة، باستخدام طريقة نسبة الأَرْجَحِية ***(Likelihood-Ratio)*** في نظرية الاستجابة للفقرة، باستخدام المنهج الوصفي التحليلي الارتباطي، على عينة تكونت من (2245) طالباً وطالبةً بالمرحلة الثانوية في محافظة ذمار، تم اختيارهم بالطريقة الطبقية العشوائية، على اختبار كاتل للذكاء "المقياس الثالث الصورة (أ)"، تم التحقق من افتراضات نظرية الاستجابة للفقرة، وتم استخراج معالم الفقرات بالنموذج الثلاثي (3PLM) لمتغيرات الدراسة المقاسة ببرنامج **Xcalibre**، وإحصائي الأداء التفاضلي **(DIF Statistic)** بطريقة نسبة الأَرْجَحِية ***(Likelihood-Ratio)*** باستخدام برنامج **(IRTLRDIF)**، وتم استخدام معامل ارتباط إيتا (Eta)، والانحدار اللوجستي الثنائي **(Binary Logistic)** ، باستخدام برنامج **(SPSS.V24)**، وقد توصلت الدراسة إلى أنه توجد علاقة دالة إِحْصَائِيَاً عند مستوى (.05&.01) بين ظهور مؤشر الأداء التفاضلي للفقرة ومتغيرات الدراسة باستخدام النموذج الثلاثي المعلم (3PLM)، إذ بلغت (.201 - .351 - .316) بالترتيب، وهي معاملات ارتباط طردية ضعيفة ومتوسطة، توجد قدرة تنبؤية لمتغير أقصى معلومات للفقرة، فيما لا توجد قدرة تنبؤية لمتغيري إحصائي المطابقة والخطأ المعياري في تقدير صعوبة الفقرة، بالأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)* في نظرية الاستجابة للفقرة.

الكلمات المفتاحية: النموذج الثلاثي المعلم (3PLM) - الأداء التفاضلي للفقرة – إحصائي المطابقة - أقصى معلومات - الخطأ المعياري في تقدير صعوبة الفقرة.

* طالب دكتوراه - مسار القياس والتقويم- قسم علم النفس- كلية التربية- جامعة الملك سعود.
** أستاذ القياس والتقويم - - قسم علم النفس- كلية التربية- جامعة الملك سعود.

للاقتباس: الخضر، محمد محمد عتيق؛ البرصان، إسماعيل سلامة. (2024). **أثر إحصائي المطابقة و أقصى معلومات والخطأ المعياري في تقدير الصعوبة على الأداء التفاضلي للفقرة باستخدام النموذج الثلاثي المعلم 3PLM** ، *مجلة الآداب للدراسات النفسية والتربوية*، 6(4)، 9-39.

© نُشر هذا البحث وفقًا لشروط الرخصة Attribution 4.0 International (CC BY 4.0)، التي تسمح بنسخ البحث وتوزيعه ونقله بأي شكل من الأشكال، كما تسمح بتكييف البحث أو تحويله أو الإضافة إليه لأي غرض كان، بما في ذلك الأغراض التجارية، شريطة نسبة العمل إلى صاحبه مع بيان أي تعديلات أُجريت عليه.

(EISSN): 2708-5775 ISSN: 2707-5788
DOI: https://doi.org/10.53285/artsep.v6i4.2267

OPEN ACCESS
Received: 29-09-2024
Accepted: 23-11-2024

مجلة الآداب
للدراسات النفسية والتربوية

The effect of Fit Statistic, Maximum Information Value and Standard Error of Difficulty Estimation of the Item on the Differential item functioning using Three-parameter model 3PLM

Mohammed Mohammed Ateeq Alkahder *
Email: alkhader2011@gmail.com

Prof. Ismaeel Salamah Al-Bursan **
Email: ibursan@ksu.edu.sa

Abstract:

The study aimed to investigate the effect of fit statistics, maximum information, and standard error (SE) in estimating item difficulty using three-parameter logistic model (3PLM) on differential item functioning (DIF) through the Likelihood-Ratio method within Item Response Theory (IRT). A descriptive-analytical correlational approach was used. A sample of 2,245 secondary school students in Yemen was selected randomly. The sample was assessed through third edition of Cattell Intelligence Test, Form (A). The assumptions of IRT were verified, and item parameters were extracted using the 3PLM model in the Xcalibre software. DIF statistics were calculated using Likelihood-Ratio method in IRTLRDIF software. Eta correlation coefficients and binary logistic regression were analyzed using SPSS V24.The study revealed statistically significant relationships at the .01 and .05 levels between DIF indicators and the study variables using the 3PLM model, with correlation coefficients of .316, .351, and .201, indicating weak to moderate positive correlations. The variable of maximum item information showed predictive ability for DIF, whereas fit statistic and standard error in estimating item difficulty did not prove predictive ability for DIF using the Likelihood-Ratio method within IRT.

Keywords: Three-Parameter Logistic Model (3PLM); Differential Item Functioning ; Fit Statistic; Maximum Information; Standard Error in Estimating Item Difficulty.

* PhD student - Measurement and Evaluation track - Department of Psychology - College of Education - King Saud University.
**Professor of Measurement and Evaluation - Department of Psychology - College of Education - King Saud University.

Cite this article as:: Alkahder, Mohammed Mohammed Ateeq. & Al-Bursan, Ismaeel Salamah. (2024). The effect of Fit Statistic, Maximum Information Value and Standard Error of Difficulty Estimation of the Item on the Differential item functioning using Three-parameter model 3PLM. *Journal of Arts for Psychological & Educational Studies 6*(4) 9-39.

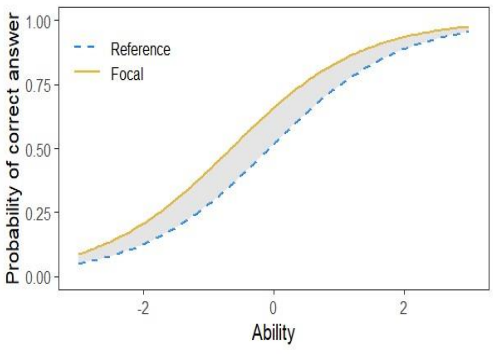
© This material is published under the license of Attribution 4.0 International (CC BY 4.0), which allows the user to copy and redistribute the material in any medium or format. It also allows adapting, transforming or adding to the material for any purpose, even commercially, as long as such modifications are highlighted and the material is credited to its author.

DOI: https://doi.org/10.53285/artsep.v6i4.2267

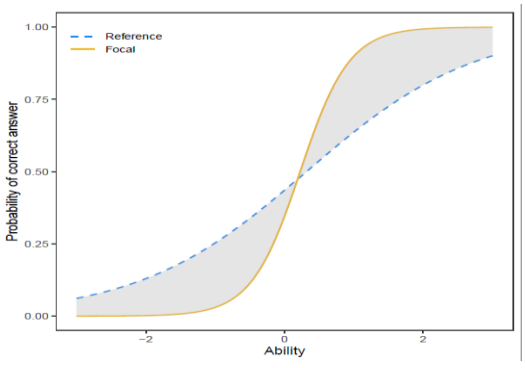
مقدمة:

يُعد تحقيق عدالة القياس الغاية التي يسعى منظرو القياس إلى تحقيقها، وتتمثل في مؤشرات الصدق التي يمكن أن تتحقق في الخصائص السيكو مترية كالصدق (Validity)، والثبات (Reliability) للاختبار ككل، أو لفقراتها كالصعوبة، والتمييز، والتخمين، ودالة المعلومات، وتتحقق عدالة الاختبار عندما يكون هناك تناظر – تساوٍ - بين الدرجات المشاهدة، والقدرات الكامنة، لدعم المساواة في التعامل بين المتقدمين في الاختبار، والاتساق في معاني درجاته للجميع، وإتاحة الفرصة بشكلٍ متساوٍ لكافة المفحوصين بكل متغيراتهم الديموغرافية المختلفة، للأداء بشكلٍ فعلي يُظهر السمة أو الخاصية التي يقيسها الاختبار، أي أن تكون الأفضلية لمجموعة دون أخرى في درجة امتلاكهم للخاصية أو السمة المقاسة فقط، ويتأثر الحكم بعدالة الاختبار بالقرارات التي تبنى على درجات الاختبار، والتي يشعر من خلالها المفحوص بمستوى العدالة التي تحققت (Cains, Bridglall & Chatterji, 2014)، ويتجلى صدق الاختبار في عدة عمليات لجمع البيانات، والأدلة، باعتباره تقييما شاملاً يوفر الدليل النظري والمادي، مُعتمداً على عددٍ من العوامل، وفقاً لظروف استخدام الاختبار؛ لإثبات معنى وملاءمة أيّ تأويل، أو حكم يبنى على نتائج الاختبار(تيغزة، 2009)، فلا يمكن اختزال الصدق في مُؤشر وحيد، ولكنه مجموعة أدلة تتكامل فيما بينها، وتدعم بعضها بعضاً؛ لفحص العلاقة بين الأداء على الاختبار، والحقائق الملموسة، والدرجات الملاحظة للسمة أو الخاصية المراد قياسها، وبذلك يُعد تحقيق الخصائص السيكو مترية للاختبارات والمقاييس النفسية شرطاً أساسياً لضمان تحقيق عدالة القياس (Test-Fairness) في تقدير ما يمتلكه الأفراد من قدرات أو سمات كامنة (عبد الوهاب، 2014). فقضية التحيز في الاختبارات عموما، والاختبارات العقلية خصوصاً، ظهرت عام 1911 على يد ألفرِد بينيه، وتجددت على يد كلٍ من إلس (Eells)، وديفز (Davis)، وهافغهيرست (Havighurst)، هيريك (Herrick)، وتايلر(Tyler)، والتي بدأت في جامعة إلس (Eells's) في شيكاغو، كقضيةٍ لفتت اهتمام المنظمات الحقوقية، فيما وضعت الجمعيات العلمية المتخصصة عالمياً، مثل الجمعية الأمريكية للبحث التربوي (AERA)، والجمعية النفسية الأمريكية (APA)، والمجلس القومي للقياس في التربية (AERA, APA & NCME, 2014) شرطاً بعدم وجود فقرات متحيزة، كأول الخصائص الأساسية للاختبارات العادلة فوجود اختبارات متحيزة، أو بعض من فقراتها، يمثل قصوراً في صدق نتائج الاختبار وعدالة استخدامه، ومن ثم يبتعد عن الهدف الذي وضِع الاختبار لقياسه (Lai & Krishnan, 2013).

لذلك فإن الفروق في الأداء لدى الأفراد متساوي القدرة قد ترجع إلى فروق حقيقية واقعية في السمة المقاسة، وهنا تتحقق الغاية من القياس، أو إلى فروق غير حقيقية ناتجة عن تحيز فقرات الاختبار (Test Bias) إلى فئة معينة، وأحياناً تحيز لبعض فقراته Item Bias فقط، لذا يقوم الأداء التفاضلي للفقرة، بدراسة الخصائص السيكو مترية للفقرة من ناحية الكيفية التي تعمل بها الفقرة في المجموعات بشكل مختلف، ويظهر ذلك بيانياً في منحنى خصائص الفقرة الاختبارية التي بموجبها تعد الفقرة غير متحيزة عندما يكون للفقرة دالة الاستجابة نفسها (Item Response Function) في كل مجموعة فرعية للمفحوصين محل المقارنة المجموعة المرجعية (Reference Group)، والمجموعة المستهدفة (Focal Group)، لذا يعد شرط ضروري للحكم على تحيز الفقرة، ولكنه غير كاف؛ إن لم يكن الدليل الوحيد على ذلك (Hambleton, Swaminithan & Rogers, 1991). وهناك نوعان من الأداء التفاضلي للفقرة: الأول الأداء التفاضلي المنتظم "Uniform Differential Item Functioning (UNDIF)" عندما تكون الأفضلية في الأداء لمجموعة ما على مجموعة أخرى عند جميع مستويات القدرة على مقياس التدريج، ويظهر عند مقارنة منحنى خصائص الفقرة (Item characteristics curve ICC) كممثل للعلاقة بين أداء المفحوص على الفقرة و متصل السمة الكامنة للمجموعتين على فقرة معينة، وهو ما يعني توازي منحنيات خصائص الفقرة، والممثل بالشكل البياني رقم (1)، ولكنهما يختلفان في مستوى صعوبة الفقرة بين المجموعتين (Sideridis, 2013a)، أَمَّا الثاني فهو الأداء التفاضلي غير المنتظم "Non Uniform Differential Item Functioning (Non-UNDIF)" ، ويظهر عندما يحدث تفاعل بين مستوى القدرة وعضوية المجموعة فتكون الأفضلية لأحدهما في الأداء عند مستوى معين من القدرة دون المستويات الأُخرى، وهو ما يعني تقاطع منحنيات خصائص الفقرة (Gruijter & Kamp, 2005). والممثل بالشكل البياني رقم (1-2).



الشكل البياني رقم (1) الأداء التفاضلي المنتظم



الشكل البياني رقم (2) الأداء التفاضلي غير المنتظم

ونتيجة للقصور الذي تعاني منه النظرية التقليدية في القياس وأساليبها المستخدمة في الكشف عن الأداء التفاضلي للفقرة، جاءت نظرية الاستجابة للفقرة؛ لتلافي قصور النظرية التقليدية، بخاصية اللاتغاير (Invariance)، فمعالم الفقرة لا تعتمد على قدرات العينة (Sample Free)، وحساب القدرة الكامنة لا يعتمد على معالم الفقرات (Item Free) (Hambleton et all., 1991). وأتاحت هذه الخاصية الحل لبعض مشكلات القياس التي يصعب حلها في النظرية التقليدية، وبشكل خاص في معادلة الاختبارات، وبنوك الأسئلة، والاختبارات التكيفية، وتحيز الاختبارات (Hambleton, Swaminithan & Rogers, 1991)، فمفهوم التحيز يحتمل تفسيراً إِحْصَائِيَاً؛ إلا أنه ارتبط بدلالات ثقافية واجتماعية سلبية كعدم العدالة والمساواة، لذا توجه الباحثون إلى إيجاد مصطلح يعطي دلالة سيكو مترية، وفي الوقت نفسه تكون مدلولاته الثقافية والاجتماعية مدلولات مهذبة بمسمى الأداء التفاضلي للفقرة باعتباره مؤشراً ضمن مجموعة من الإجراءات التي تستخرج للتحري والاستدلال في الحكم على تحيز الفقرة (دي إيالا، 2017/2009)، وليس دليلاً حتمياً على التحيز، فيما يُعد التحيز حكماً باتاً في أداء الفقرة (Ellis & Raju, 2004). فالفقرة التي تبدي أداءً تفاضلياً لا تتمتع بخاصية اللاتغاير، (أي: أن بيانات الفقرة لا تطابق النموذج)، وعدم المطابقة هو مؤشر للشك بأن معالم الفقرة تتأثر بالعينة، وكذلك قدرة الأفراد تتأثر بالفقرة، وبهذا يتعذر الحصول على إطار مرجعي للحكم على الأداء المقدم على الفقرة في الاختبار، أو أداء المفحوص في الاختبار ككل؛ فمؤشر الأداء التفاضلي للفقرة يعد أُولى خطوات التعرف على وجود التحيز من عدمه، بل شرطاً ضرورياً للحكم على تحيز الفقرة (Camili &Shepard, 1994).

وتعددت طرق الكشف عن الأداء التفاضلي لفقرات الاختبار، واختلفت إجراءاتها وفقاً للافتراضات الأساسية للنظرية التي تنتمي إليها (النظرية التقليدية، أو نظرية الاستجابة للفقرة) (النفيعي، 2008)، ويتم ذلك من خلال التحليل الكمّي للفقرة التي قد تُظِهر أداءً تفاضلياً والذي يعود لطبيعة التفاعل بين مستوى القدرة وانتماء الفرد لمجموعة ما، وهذا في حد ذاته لا يعني حتمية تحيز الفقرة، ولكنه يُعد موجهاً لإخضاع الفقرة للتحليل المنطقي للبحث عن أسبابه (دي إيالا، 2017/2009)، بعرضها على خبراء ومحكمين من ذوي الاختصاص والخبرة والمهارة، والاهتمام بالمفهوم أو السمة محل القياس، وفق محكات مُعدة مُسبقاً تتواءم مع النطاق الشامل للسمة، وخصائص المجتمع الذي سيطبق عليه الاختبار، أو المحتوى الذي تنتمي إليه الفقرة، لتحديد ما إذا كان ظهور مؤشر الأداء التفاضلي للفقرة يثبت تحيز الفقرة أم لا؟ وما مصدر هذا التحيز؟ (Karami, 2011; Camili & Shepard, 1994) فالاستدلال على التحيز يتجاوز البيانات؛ باعتبار الأداء التفاضلي مصطلحاً لتمييز الدليل التجريبي الذي نحصل عليه عند الاستقصاء عن تحيز الفقرة،

بدراسة الخصائص السيكو مترية للفقرة من ناحية الكيفية التي تعمل بها الفقرة في المجموعة بشكل مختلف عن المجموعة الأخرى (Hambleton et al., 1991).

وفي نظرية الاستجابة للفقرة تعددت طرق الكشف عن الأداء التفاضلي ثنائية الاستجابة، باختلاف النموذج المعتمد (1PL- 2PL -3PL- 4PL)، واعتماداً على هذه النماذج طورت عدة إجراءات للكشف عن الأداء التفاضلي، منها: طريقة كاي تربيع للورد (Lord Chi Square)، وطريقة اختبار تحيز الفقرة (SIBTEST Test)، ومقارنة معلمات الفقرة، وطريقة مقارنة مطابقة نماذج استجابة الفقرة للبيانات في المجموعات المختلفة للأفراد وطريقة حساب المساحة بين المنحنيات المميزة للفقرة Area(ICC) difference. (Hambleton & Swaminathan, 2010; Raju, 1988)، واقترح ثيسن (Thissen, 2001)، طريقة نسبة الأَرْجَحِية Likelihood-Ratio IRT-LR إحدى طرق الكشف عن الأداء التفاضلي في نظرية الاستجابة للفقرة باستخدام البرمجية الإحصائية (IRTLRDIF)، وتستخدم هذه الطريقة الإحصائية Likelihood-Ratio Tests (For DIF) من أجل فحص الفرضية الصفرية التي تنص على عدم وجود فروق في معالم الفقرات بين المجموعات، والدلالة الإحصائية لهذه الطريقة تشير إلى وجود الأداء التفاضلي في هذه الفقرة. وتعتمد صياغة الفرضية الصفرية على تعريف لورد Lord الذي ينص على أن وجود فروق في دالة الاستجابة للفقرة بين مجموعة وأخرى دليل على وجود مؤشر الأداء التفاضلي للفقرة، فمجموعة معالم الفقرة تناظر دالة الاستجابة للفقرة، وتستخدم للفقرات ثنائية أو متعددة الاستجابة (Samejima, s (1969,1997). فطريقة نسبة الأَرْجَحِية (Likelihood Ratio) تقوم بإيجاد الفرق بين اثنين من النماذج هما المدمج والمعزز (Compact & Augmented) والمصاغة في المعادلة كالآتي:

$$[\boldsymbol{Gj2 = -2Log\,L(Compact\ Model) + 2Log\,L(Augmented\ Model)}]$$

ويمكن تبسيطها بالمعادلة التالية: $\boldsymbol{[G^2 = -2LL_c + 2LL_A\ ]}$ حيث "$\boldsymbol{2LL_c}$" لوغاريتم نسبة الأَرْجَحِية للنموذج المدمج، و "$\boldsymbol{2LL_A}$" لوغاريتم نسبة الأَرْجَحِية للنموذج المعزز، ويكون توزيعه مماثل لتوزيع مربع كاي X^2، عند درجة حرية تساوي الفرق بين عدد المعالم بين النموذج المدمج والنموذج المعزز، إذ بُني النموذج المدمج (compact Model) على افتراض تساوي المعالم في المجموعة المرجعية، والمجموعة المستهدفة عند مطابقه البيانات مع معالم منحنيات الاستجابة للفقرات التي تضَبّط بحيث تكون متساوية في كلا المجموعتين لكل الفقرات، (Thissen, 2001)، وتتحدد نقطة القطع للقيمة الحرجة بالقيمة (3.84)، فإذا بلغت قيمة $\boldsymbol{G^2}$ أعلى من أو تساوي القيمة الحرجة لمربع كاي؛ فإنه توجد فروق دالة إِحْصَائِيَاً بين النموذجين، وهو دليل على ظهور مؤشر الأداء التفاضلي للفقرة، فيما تم تصنيف مستويات الأداء

التفاضلي للفقرة - وفقاً لمحك كوهين - إلى ثلاثة مستويات، تتراوح قيمة G^2 للمستوى الضعيف (A) بين ($9.4 > G^2 > 3.84$)، فيما تتراوح قيمة G^2 للمستوى المتوسط (B) بين ($41.9 > G^2 > 9.4$)، فيما تكون قيمة G^2 للمستوى المرتفع (C) ($G^2 > 41.9$ Greer, 2004)، واهتمت الدراسات السابقة للأداء التفاضلي بثلاثة مناحٍ مختلفة، شمل المنحى الأول على دراسة أثر بعض المتغيرات على الأداء التفاضلي، للفقرة كحجم العينة، ونماذج الاختبارات، وطول الاختبار، والقيم المتطرفة، ومستوى الصعوبة، ونوع الفقرة، فيما ركزت دراسة واحدة على العلاقة بين اتجاه الأداء التفاضلي للفقرة وبعض معالم الفقرة، وشمل المنحى الثاني مقارنة فاعلية طرق الكشف عن الأداء التفاضلي، وشمل المنحى الثالث الكشف عن الأداء التفاضلي لفقرات الاختبارات أو المقاييس، إذ هدفت دراسة المحروقية (2024)، إلى الكشف عن الأداء التفاضلي في فقرات اختباري الرياضيات في الدراسة الدولية تيمس (TIMSS) والكشف عن العلاقة بين اتجاه الأداء التفاضلي للفقرات (الذكور / الإناث)، وكل من مستوى صعوبة الفقرة (مرتفع / متوسط / منخفض)، ونوع الفقرة (اختيار من متعدد/ مقالي)، ومجال المحتوى، والمجال المعرفيّ للفقرات، باستخدام طريقتي الصعوبة المحوّلة للفقرة، وطريقة مانتل هانزل تبعًا لمتغير الجنس (الإناث الذكور)، في فقرات اختباري الرياضيات في الدراسة الدولية تيمس TIMSS للصفين الرابع والثامن الأساسيين، على عينة من طلبة الصف الرابع والثامن بلغت (13565)، ممثلة من (288) مدرسة حكومية، وخاصة، ودولية، وتوصلت الدراسة - وفق طريقة الصعوبة المحوّلة - إلى أن النسب الإجمالية للفقرات التي أظهرت أداء تَفاضلياً في اختبار الرياضيات للصف الرابع بلغت (7.64%)، بينما بلغت نسبتها للصف الثامن (9.23%)، في المقابل أظهرت طريقة مانتل هانزل (17.2%) للصف الرابع و(26.15%) للصف الثامن. وبلغت نسبة الاتفاق بين نتائج الطريقتين للصف الرابع (87.90%)، وهي درجة توافق متوسطة، وبلغت للصف الثامن (80%)، وهي درجة توافق مقبولة. وأشارت النتائج كذلك إلى وجود علاقة بين اتجاه الأداء التفاضلي للفقرات (الذكور الإناث) وكل من مستوى صعوبة الفقرة (مرتفع / متوسط / منخفض) في الصف الرابع، ومتغير نوع الفقرة (اختيار من متعدد/ مقالي) في الصف الثامن، ومتغير المجال المعرفي (المعرفة / التطبيق / الاستدلال) في الصف الثامن، ولم تظهر وجود علاقة مع متغير مجال المحتوى (الأعداد / الأشكال الهندسيّة والقياس / عرض البيانات/ الجبر) وفقاً لبيانات الصفين (رغم أن الأسلوب لا معلمي ووفق النظرية الكلاسيكية).

وهدفت دراسة ضعضع (2023)، إلى فحص أثر اختيار نماذج الاستجابة للفقرة الثنائية (أحادي المعلم، وثنائي المعلم، وثلاثي المعلم) في الكشف عن الفقرات ذات الأداء التفاضلي باستخدام طريقة المنحنى المميز للفقرة، والمقارنة في نسبة الفقرات ذات الأداء التفاضلي، ومعرفة دلالة الفروق في قيمة معامل الأداء التفاضلي ومتوسطه لاختبار القبول بالمعهد العالي للغات في جامعة حلب المكون من (60) فقرة، على عينة

بلغت (1000) مُتقدماً ومتقدمةً، وأظهرت النتائج: أن (31.481%) من الفقرات لم تظهر أداءً تفاضلياً في كل النماذج، وأن (44.44%) أظهرت أداءً تفاضلياً عند نموذج واحد فقط، وأن (24.074%) أظهرت عند نموذجين، ولم تظهر أي فقرة أداءً تفاضلياً عند كل النماذج، وأظهرت النتائج أن النموذج ثنائي المعلم حقق أعلى نسبة بلغت (68.5%)، يليه النموذج أحادي المعلم بنسبة (22.222%)، وأخيراً النموذج ثلاثي المعلم بنسبة (1.9%).

وهدفت دراسة العبدالله (2022)، إلى دراسة تأثير اختلاف حجم العينة على دقة الكشف عن الأداء التفاضلي للفقرة، بلغت أحجام العينات (1000، 500، 300) على اختبار مكون من عشرين فقرة متعددة التدريج، وتم استخدام أنموذج الاستجابة المتدرجة لتقدير معالم الفقرات والأفراد، باستخدام طريقة مانتل-هانزل، وقد أظهرت نتائج الدراسة العلاقة العكسية بين حجم العينة وعدد الفقرات التي أظهرت أداءً تفاضلياً.

وهدفت دراسة صالح (2022) إلى دراسة أثر حجم العينة في الكشف عن الأداء التفاضلي للمفردة باستخدام طريقة المنحنى المميز للمفردة وفق النموذج ثلاثي المعلم لاختبار الاستدلال غير اللفظي ويتكون من (60) فقرة من نوع الاختيار من متعدد على عينة عشوائية مكونة من (1000) طالبٍ وطالبة، وأظهرت النتائج: أنه عند حجم عينة (1000) أظهرت (24) فقرة أداء تفاضلياً بنسبة (40%) من فقرات الاختبار، وعند حجم عينة (500)، أظهرت النتائج أن (42) فقرة ذات أداء تفاضلي بنسبة (70%) من فقرات الاختبار، يوجد فروق ذات دلالة إحصائية في قيمة الأداء التفاضلي للفقرة وفقاً لمتغير حجم العينة لصالح العينة الأصغر، وأنه يوجد فروق ذات دلالة إحصائية في متوسط قيمة الأداء التفاضلي للفقرة لصالح العينة الأصغر، ويوجد فروق ذات دلالة إحصائية في نسبة الفقرات ذات الأداء التفاضلي لصالح حجم العينة الأصغر.

وهدفت دراسة علي وعيال (2021) إلى معرفة تأثير اختلاف مستوى قدرة الأفراد في الأداء التفاضلي لاختبار القدرة العقلية غير اللفظي، والمكون من (60) فقرة على عينة طبقية عشوائية بلغت (600) طالب وطالبةً، وأظهرت النتائج ملاءمة (55) فقرة من فقرات اختبار ماستر براين للأنموذج ثنائي المعلم، وأن (10.9%) من الفقرات أظهرت أداء تفاضلياً لجميع أفراد العينة، و18.18% للأفراد منخفضي القدرة، و16.36% للأفراد مرتفعي القدرة، وأظهرت عدم وجود فروق ذات دلالة إحصائية في نسبة الفقرات ذات الأداء التفاضلي تبعاً لمستوى القدرة، كذلك عدم وجود فروق ذات دلالة إحصائية في قيمة الأداء التفاضلي ومتوسطها تبعاً لمستوى القدرة.

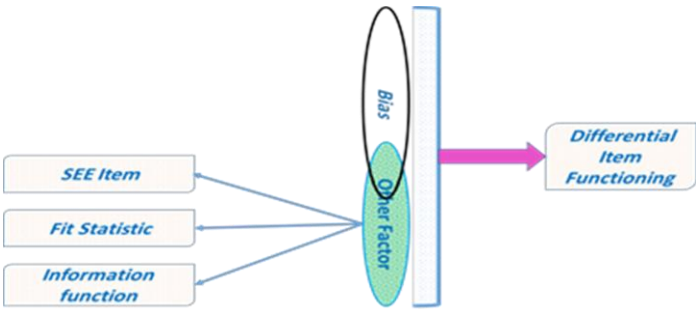
وهدفت دراسة بني عطا (2018)، إلى التحقق من أثر القيم المتطرفة على الأداء التفاضلي لفقرات اختبار الرياضيات في الدراسة الدولية "TEMES" للصف الثامن تبعا لمتغير النوع، على عينة عشوائية بلغت (554) طالباً وطالبةً لعام 2011م، وتكون الاختبار من (26) فقرة، وقد استخدمت طريقة توكي (الرسم الصندوقي)، وطريقة مؤشر الأداء للفقرة والاختبار للكشف عن القيم المتطرفة، والأداء التفاضلي لفقرات الاختبار، وكشفت نتائج الدراسة عن وجود (7) فقرات أبدت أداءً تفاضلياً بنوعيه المنتظم وغير المنتظم بوجود القيم المتطرفة من أصل (26) فقرة بنسبة (27%)، تبعاً لمتغير النوع. وبعد حذف القيم المتطرفة من مجموعة البيانات وإعادة التحليل أظهرت النتائج وجود (4) فقرات قد أبدت أداءً تفاضلياً بنوعيه المنتظم وغير المنتظم بنسبة (15 %)، إذ تناقصت هذه النسبة بمقدار (12 %) عندما استبعدت القيم المتطرفة من التحليل.

ويلاحظ مما سبق اختلاف الدراسات السابقة في دراسة المتغيرات وعلاقتها بالأداء التفاضلي للفقرة كحجم العينة، ونماذج الاختبارات، وطول الاختبار، والقيم المتطرفة، ومستوى الصعوبة، ونوع الفقرة، فيما تعددت طرق الكشف والنماذج اللوجستية المستخدمة، فدراسة المحروقية (2024)، وبني عطا (2018) تناولتا العلاقة بين الأداء التفاضلي للفقرة التفاضلي مع بعض المتغيرات الإحصائية، فيما ركزت دراسة العبدالله (2022)، ودراسة صالح (2022)، على أثر حجم العينة، أما دراسة علي وعيال (2021) فقد هدفت إلى معرفة تأثير اختلاف مستوى قدرة الأفراد، أما دراسة ضعضع (2023)، فقد هدفت إلى معرفة أثر اختيار نماذج الاستجابة للفقرة ثنائية الاستجابة على ظهور مؤشر الأداء التفاضلي للفقرة التفاضلي.

مشكلة الدراسة:

يُعد وجود أخطاء القياس كأحد مكونات الدرجات الملاحظة للأداء على الاختبار مضللاً، ولا يعطي مُؤشراً عن مدى تحيز الفقرات لمجموعة دون أخرى، ومن هنا سعى بناة ومطورو المقاييس لتحقيق موضوعية القياس وعدالته، من خلال تحقيق افتراضات نظرية الاستجابة للفقرة Item Response Theory "(IRT)" ممثلة بخاصية اللاتغاير (Invariance) واعتماد نمط الاستجابة. فمعالم الفقرة لا تعتمد على قدرات العينة Sample Free، وحساب القدرة الكامنة لا يعتمد على معالم الفقرات "Item Free" (Hambleton et all., 1991)، وتمكنت هذه النظرية من الكشف عن التحيز، من خلال مؤشر الأداء التفاضلي للفقرة، الذي يُعد دالة مشتقة إِحْصَائِيَاً يكون للفقرة أداء تفاضلياً لإحدى المجموعتين ذواتي القدرة المتساوية دون الأخرى، ويكون الأداء التفاضلي شرطاً ضرورياً لاعتبار الفقرة متحيزة، إن لم يكن المؤشر الوحيد على التحيز، ولكنه غير كاف للحكم على تحيز الفقرة، بل موجه لإخضاع الفقرة للتحليل المنطقي للبحث عن أسبابه (دي إيالا، 2017/2009)، كون الفقرة التي تبدي أداءً تفاضلياً لا تتمتع بخاصية

اللاتغاير(Invariance) وتصبح غير ملائمة للنموذج، وهنا نجد المتخصصين وفق احتمالين الأول: أن ظهور مؤشر الأداء التفاضلي دليل قاطع لتحيز الفقرة، والثاني: يكون ظهور هذا المؤشر زائفاً وغير كاف للحكم على تحيزها، والشكل رقم (3) يوضح ذلك.



الشكل البياني رقم (3) يبين تصور لمشكلة الدراسة

وبناءً على ذلك يتطلب من منظري القياس توقع فرضيات حول المصادر المحتملة لهذا الأداء التفاضلي للفقرة، ومنها ما توصل إليه روسوس (Roussos &Stout, 1996) بأن الأسباب الموضوعية المحتملة لظهور مؤشر الأداء التفاضلي للفقرة تبقى غير معروفة إلى حد كبير، ويعزوها بيترسون وآخرون (Petersen, at al, 2003) إلى الأخطاء العشوائية وقد تكون بمحض الصدفة، فيما توصلت إليس (Ellis & Ragu, 2003) إلى أن مهمة تطوير طرق موضوعية لتحديد أسباب ومصادر الأداء التفاضلي للفقرة يُعد من أكبر التحديات لهذه المرحلة، ورغم صعوبة تقديم تفسيرات مناسبة لهذه المصادر؛ إلا أننا بصدد دراسة يمكن من خلالها الإسهام في تقديم بعض هذه التفسيرات، فقد يكون ناتجاً عن مؤشرات إحصائية أخرى قد تعود إلى أخطاء القياس المختلفة، التي يسعى الباحثان إلى التحقق من علاقتها بظهور مؤشر الأداء التفاضلي، والممثلة في إحصائي مطابقة الفقرة للنموذج المستخدم لتأثيره بمستويات القدرة المختلفة ومعالم الفقرات، وبذلك نستطيع التعرف على العوامل التي يمكن أن تؤثر على كل منهم (Hambleton & Murray, 1983)، وأقصى معلومات للفقرة، والخطأ المعياري في تقدير صعوبة الفقرة، كون معالم الفقرة والدرجة الكلية للاختبار تتأثر بمؤشر صعوبة الفقرة والخطأ المعياري في تقديرها (كروكر والجينا، 2017/1986) ونظراً لندرة وجود دراسات تتعلق بمتغيرات الدراسة وأهدافها بشكل صريح من خلال مسح الأدب النظري للدراسات التي أُجريت على الأداء التفاضلي، لم توجد دراسة –في حدود علم الباحثين- تناولت علاقة الأداء التفاضلي للفقرة بالمؤشرات الإحصائية التي قد تؤدي إلى ظهور مؤشر للأداء التفاضلي للفقرة؛ إذ تتلخص مشكلة البحث في السؤال الرئيسي الآتي: ما أثر إحصائي المطابقة وأقصى معلومات والخطأ المعياري في تقدير الصعوبة باستخدام النموذج الثلاثي المعلم 3PLM على الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية Likelihood-Ratio؟ ويتفرع عن السؤال الرئيسي السؤالان الفرعيان الآتيان:

1- هل توجد علاقة بين الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، ومؤشر إحصائي المطابقة، وأقصى معلومات للفقرة، والخطأ المعياري في التقدير لصعوبة الفقرة باستخدام النموذج الثلاثي المعلم *(3PLM)* في نظرية الاستجابة للفقرة؟

2- ما أثر كل من إحصائي المطابقة، وأقصى معلومات للفقرة والخطأ المعياري في التقدير لصعوبة الفقرة باستخدام النموذج الثلاثي المعلم *(3PLM)*، على الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)* في نظرية الاستجابة للفقرة؟

أهداف الدراسة:

تهدف الدراسة الحالية إلى الآتي:

1- معرفة العلاقة بين الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، ومؤشر إحصائي المطابقة، وأقصى معلومات للفقرة، والخطأ المعياري في التقدير لصعوبة الفقرة باستخدام النموذج الثلاثي المعلم (3PLM) في نظرية الاستجابة للفقرة.

2- معرفة أثر كل من إحصائي المطابقة، وأقصى معلومات للفقرة والخطأ المعياري في التقدير لصعوبة الفقرة باستخدام النموذج الثلاثي المعلم (3PLM)، على الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)* في نظرية الاستجابة للفقرة؟

أهمية الدراسة:

الأهمية النظرية: تتحدد الأهمية النظرية للدراسة في قدرة النتائج المتوقعة لهذه الدراسة على الإجابة عن تساؤلات الدراسة، والتي قد تسهم في الكشف عن أثر إحصائي المطابقة، وأقصى معلومات للفقرة والخطأ المعياري في التقدير لصعوبة الفقرة، على ظهور الأداء التفاضلي للفقرة، في مواقف اختبارية فعلية تم تطبيقها في مدارس التعليم.

الأهمية التطبيقية: تتمثل الأهمية التطبيقية لهذه الدراسة فيما قد تقدمه من مؤشرات يستدل بها منظرو القياس وبناة ومطورو الاختبارات والمقاييس، في دقة الحكم على تحيز الفقرة من عدمه، رغم ظهور مؤشر الأداء التفاضلي، أو من خلال ما قد تقدمه من نتائج قد تكون مؤشراً داعماً لسلامة حكم الخبراء بعدم تحيز الفقرة، لاستبعاد ما قد يؤثر على صدق القياس، وتجنب التشكيك في تحقيق عدالة الاختبار؛ لكونها تتعلق بصدق الاختبار وعدالته التي تدعم المساواة في التعامل، والاتساق في المعنى والتفسير.

حدود الدراسة:

اقتصرت الدراسة على إحصائي المطابقة للفقرة، متمثلاً في البواقي (Residuals) المعيارية وأقصى معلومات للفقرة (Maximum Information) والخطأ المعياري في التقدير لصعوبة الفقرة (Standard Error of Difficulty Estimation)، على الأداء التفاضلي للفقرة باستخدام نسبة الأَرْجَحِية *(Likelihood-Ratio)*، وفقاً لمتغير النوع، بمدارس المرحلة الثانوية بمدينة ذمار- الجمهورية اليمنية، بالفصل الدراسي الأول للعام الدراسي 2021-2022.

مصطلحات الدراسة:

إحصائي المطابقة للفقرة: (Fit Statistic)

يشير إلى الدرجة المعيارية للفروق بين الصعوبة المتوقعة، والصعوبة الفعلية للفقرة (كروكر والجينا، 2017/1986)، ويعرف إجرائياً بقيمة الإحصائي المستخرج من قسمة البواقي الخام للفقرات على الخطأ المعياري، والذي يعطى بالعلاقة التالية: حيث N_j هي عدد المفحوصين عند مستوى القدرة (j) المقابل لأقصى معلومات للفقرة(Hambleton & Swaminthan, 1985)

$$S = \sqrt{P_{ij}(\theta)\left(1 - p_{ij}(\theta)\right)}/N_j$$

أقصى معلومات للفقرة (Maximum Information)

يعتبر مؤشر دقة القياس في نظرية الاستجابة للفقرة (Edward, 2010)، وتشير إلى أقصى ارتفاع لمنحنى دالة المعلومات للفقرة عند مستوى معين على متصل السمة، (Rose, Bjorner; Becker, Friesc, 2008)، ويعرف إجرائياً بالدالة المشتقة لمستوى القدرة المقدرة للمفحوص وبارامترات الفقرة المطبقة، فالفقرة التي تمتلك أعلى المعلومات عند مستوى معين من القدرة تعطي أدق التقديرات للقدرة المقاسة وبأقل عدد من الفقرات (Siang&Fritz, 2006).

الخطأ المعياري في التقدير لصعوبة الفقرة: (SEE)

يمثل معلم صعوبة الفقرة القاسم المشترك بين جميع نماذج نظرية الاستجابة للفقرة، ويعد الخطأ المعياري في التقدير لصعوبة الفقرة مؤشراً لدقة القياس، "Standard Error of Estimation (SEE)"، ويمثل بالجذر التربيعي لمتوسطات مربعات الانحرافات بين القيم المقدرة من البيانات الملاحظة والقيم التي يفترضها النموذج (دي إيالا، 2017/2009). ويعرف إجرائياً بالانحراف المعياري للخطأ في تقدير معلمة الصعوبة للفقرة، ويمثل معكوس دالة المعلومات،(Hambleton and Jones ,1993) ويستخرج بالمعادلة التالية:

$$SEM = \frac{1}{\sqrt{I(\theta)}}$$ (بيكر، 2001/2010)

الأداء التفاضلي: (DIF)

هو مؤشر إحصائي يستخرج بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، للتعبير عن الفروق في احتمالية الإجابة الصحيحة عن الفقرة باختلاف المجموعة، ممن لهم القدرة نفسها (دي إيالا، 2017/2009). ويعرف إجرائياً بأنه مؤشر لاختلاف الخصائص الإحصائية لفقرة أو مجموعة من الفقرات لدى أفراد لهم القدرة نفسها تعزى لاختلاف انتمائهم لبعض المتغيرات الديموغرافية (Maller, 2003).

النموذج ثلاثي المعالم (3PLM):

أقترحه لورد (Lord Model)، مضيفاً بارامتراً ثالثاً أطلق عليه معلم التخمين (Guwssing Parameter) أو الخط التقاربي الأدنى (Lower Asymptote Line)، ويُمثل احتمال توصلّ الأفراد ذوي القدرة المنخفضة إلى الإجابة الصحيحة عن فقرة اختيار من متعدد عن طريق التخمين عندما تكون (θ=صفر)، أَيّ أن التوصل للإجابة الصحيحة عن طريق التخمين لا يرتبط بارتفاع أو انخفاض مستوى القدرة وأن المدى النظري لـ c ($0 \le c \le 1$)، متفادياً تأثير التخمين العشوائي على الأداء في الاختبار، ويستخرج بالمعادلة الرياضية التالية: حيث Ci : معلم التخمين بالنسبة للفقرة i.

$$pi(\theta) = Ci + (1 - Ci)\frac{e^{Da(\theta - bi)}}{1 + e^{Da(\theta - bi)}}$$

منهج الدِّراسة وإجراءاتها

استخدم الباحثان المنهج الوصفي التحليلي الارتباطي لتحقيق أهداف الدراسة والإجابة عن تساؤلاتها، والذي يعد الأسلوب المناسب للوصول إلى العلاقة أو التنبؤ باستخدام هذه العلاقة بين المتغيرات (أبو علام، 2011)، باعتبار التنبؤ هو الحصاد الأخير للبحث العلمي بعد الوصف والتفسير، وتعتبر نتائج المعرفة التنبؤية ترجيحية (قيلي، 2020).

تكون مجتمع الدراسة من جميع طلاب وطالبات الصف الثاني الثانوي بمدينة ذمار بالجمهورية اليمنية، والبالغ عددهم (4985) طالباً وطالبةً، في الفصل الأول للعام الدراسي 2021- 2022م، بواقع (2069) طالباً، و(2916) طالبةً، حسب آخر إحصائية للفصل الأول من العام الدراسي 2021-2022م (مكتب التربية والتعليم بمحافظة ذمار، 2022م).

تكونت عينة الدراسة من طلاب وطالبات الصف الثاني الثانوي بـ (2245) طالباً وطالبةً، بنسبة تصل إلى (45.04%) تقريباً من المجتمع، تم اختيارهم بالطريقة الطبقية العشوائية، بتقسيم المدينة إلى خمس مناطق تعليمية وفق الجهات الأربع مع مركز المدينة.

استخدم الباحثان اختبار كاتل للذكاء المتحرر من أثر الثقافة "المقياس الثالث الصورة (أ)" يستخدم للأعمار من (13-19) سنة والراشدين المتفوقين، أعد هذا الاختبار عالم النفس " رايموند كاتل " Raymond Cattell كاختبار يصنف من الاختبارات غير المتحيزة ثقافياً، يمكن تصنيفها على أنها اختبارات قوة وسرعة، ويمكن تطبيق الاختبار بالطريقة الفردية أو الجمعية، تكون استجابة المفحوص عليها بعد فهم الأشكال، التي تعرض عليه لملاحظتها، وفهم واكتشاف العلاقة، ثم اختيار الاستجابة المطلوبة، من البدائل المطروحة لكل سؤال (أبو حطب، وأمال صادق، 2005).

ويتألف الاختبار من أربعة اختبارات فرعية يتطلب كل اختبار نمطا مختلفا من الاستجابة، وهي:

1- السلاسل (Series):ويتكون من (14) فقرة، يطلب من المفحوص إكمال المربع الخالي، باختيار الشكل الملائم من ستة بدائل.
2- التصنيف (Classification) :ويتكون من (14) فقرة، يطلب من المفحوص اختيار الشكلين المختلفين من بين سلسلة من الأشكال عددها خمسة.
3- المصفوفات (Matrices) :ويتكون من (13) فقرة، يطلب من المفحوص اختيار شكل من ستة أشكال لاستكمال مصفوفة الاشكال بالفقرة.
4- الشروط (Conditions): ويتكون من (10) فقرات، يتطلب هذا الاختبار من المفحوص إدراك العلاقة في وضع نقطة أو نقطتين في مكان معين، يحدد علاقتهما في مجموعة من الأشكال المتداخلة، ثم يختار المفحوص استجابته من بين سلسلة من الأشكال عددها خمسة بدائل.

إجراءات الدراسة:

تم الحصول على اختبار الذكاء المتحرر من أثر الثقافة "المقياس الثالث الصورة (أ)"، تم الحصول على موافقة لجنة أخلاقيات البحث العلمي برقم (KSU-HE-21-147)، في جلستها رقم (42) بتاريخ 2021/03/02م، تم تطبيق الاختبار على عينة البحث المذكورة، تم التحقق من افتراضات نظرية الاستجابة للفقرة. تم استخراج معالم الفقرات متغيرات الدراسة المقاسة (إحصائي المطابقة للفقرة، وأقصى معلومات للفقرة، والخطأ المعياري في التقدير لصعوبة الفقرة)، وإحصائي الأداء التفاضلي (DIF Statistic) باستخدام النموذج الثلاثي المعلم (3PLM) في نظرية استجابة الفقرة، باستخدام طريقة نسبة الأَرْجَحِية

(Likelihood-Ratio)، وتم تحليل البيانات وفقاً للتساؤلات المحددة في الدراسة. باستخراج معاملات الارتباط باستخدام مربع إيتا والتنبؤ بتحليل الانحدار اللوجستي الثنائي (Binary Logistic)، وأجري تحليل البيانات باستخدام البرامج الإحصائية الآتية: برنامج (SPSS)، وبرنامج (IRTLRDIF)، وبرنامج (Xcalibre).

الخصائص السيكومترية:

لاستخراج مؤشرات الصدق لاختبار الذكاء لكاتل المتحرر من أثر الثقافة "المقياس الثالث الصورة (أ)"، استخراج الاتساق الداخلي، وثبات معامل ألفا-كرونباخ (Cronbach's Alpha)، ومعامل ثبات ماكدونالد أوميجا(McDonald's Omega) ، على عينة استطلاعية مكونة من 260 طالباً وطالبةً.

تم استخدام معاملات الارتباط بطريقة بيرسون (Pearson Correlation Coefficient) لأداء أفراد العينة بين درجات كل فقرة مع الدرجة الكلية للبعد الذي تنتمي إليه والدرجة الكلية للاختبار، وأظهرت النتائج أن جميع قيم معاملات ارتباط الفقرات بالدرجة الكلية للاختبار تراوحت بين (.219 & .569)، وهي ضمن البعد الأول، وجميعها قيم دالة إِحْصَائِيَاً عند مستوى (0.01)، عدا الفقرة السابعة في البعد الأول والفقرة العاشرة في البعد الرابع كانتا دالتين عند (.05)، فيما تراوحت قيم معاملات الارتباط بين الفقرة والبعد الذي تنتمي إليه بالبعد الأول "السلاسل Series" بين (.202 & .660)، وجميعها قيم دالة إِحْصَائِيَاً عند مستوى (0.01)، عدا الفقرة السابعة كانت دالة عند (.05)، وفي البعد الثاني " التصنيف Classification" تراوحت معاملات الارتباط بين (.360 & .695)، فيما تراوحت قيم معاملات الارتباط في البعد الثالث " المصفوفات Matrices" بين (.395 & .771)، وجميعها قيم دالة إِحْصَائِيَاً عند مستوى (0.01)، فيما تراوحت معاملات الارتباط في البعد الرابع " الشروط Conditions" بين (.236 & .485)، وجميعها قيم دالة إِحْصَائِيَاً عند مستوى (0.01)، فيما تراوحت معاملات ارتباط وجميعها قيم دالة إِحْصَائِيَاً عند مستوى (0.01)، عدا الفقرة العاشرة كانت دالة عند (.05)، فيما تراوحت قيم معاملات ارتباط الأبعاد بالدرجة الكلية للاختبار بين (.454&.826)، و تراوحت قيم معاملات ارتباط الأبعاد ببعضها بين (.304&.847)، وجميعها قيم دالة إِحْصَائِيَاً عند مستوى (0.01)، وهي مؤشر على تحقق الدليل الثالث من أدلة الصدق الثمانية الداعمة لتفسيرات درجة الاختبار وفق ما أورده (نتكو &بركهارت، 2018) ، القائم على البنية الداخلية، ويرتكز على العلاقة بين مكونات الأداة أو الاختبار، والمسمى بالاتساق الداخلي (Validity of internal consistency)، وتم ذلك من خلال إيجاد معاملات الارتباط بين الفقرات والدرجة الكلية للبعد الذي تنتمي إليه والدرجة الكلية للاختبار (تيغزة، 2011)، تم استخراج مؤشر الثبات

(Reliability) لاختبار الذكاء لكاتل المتحرر من أثر الثقافة "المقياس الثالث، الصورة (أ)"، باستخدام طريقة معامل ألفالكرونباخ (Cronbach's Alpha)، والتجزئة النصفية (Split half Reliability) ، وثبات ماكدونالد أوميجا (McDonald's Omega) الذي اقترحه (McDonald, 1999, 1978)، استنادًا إلى التحليل العاملي للفقرات. إذ بلغ معامل ثبات ألفا لكرونباخ للدرجة الكلية للاختبار (.880)، فيما بلغ معامل ثبات التجزئة النصفية (.850)، فيما بلغ معامل ثبات ماكدونالد اوميجا (.903)، وهي معاملات ثبات مرتفعة، وفيما يتعلق بثبات الأبعاد فقد تراوحت قيم ثبات ألفا لكرونباخ بين (.75-.86)، وتراوحت قيم معامل ثبات التجزئة النصفية بين (.76-.87)، فيما تراوحت قيم معامل ثبات ماكدونالد اوميجا بين (85-.88)، وهي معاملات ثبات جيدة ومرتفعة.

التحقق من افتراضات نظرية الاستجابة للفقرة.

1- أحادية البعد:

للتحقق من أحادية البعد باستخدام أسلوب التحليل العاملي (Factor Analysis)، لاختبار الذكاء لكاتل المتحرر من أثر الثقافة "المقياس الثالث الصورة (أ)"، بطريقة المكونات الأساسية Principal (Component Analysis)، بعد تدوير العوامل بطريقة (Vairimax)، تم اختبار كفاية العينة ككل عن طريق اختبار Kaiser-Meyer-Olkin (KMO)، إذ بلغت قيمته (KMO=0.926)، وهي أعلى من الحد الأدنى المطلوب الذي يجب ألا يقل عن (0.6) حسب محك Kaiser. وتم اختبار افتراض الكروية بدلالة قيمةكا2 (X^2) ، لاختبار (Bartlett's - Sphericity) إذ كانت $(\chi^2(1245) = 16755{,}324\ p = .000)$، وهي دالة إِحْصَائِيَاً وفقاً لمحك Guilford، أي أن مصفوفة معاملات الارتباط في كل اختبار ليست مصفوفة الوحدة، مما يؤيد رفضنا للفرضية الصفرية التي تقول: أنه لا يوجد ارتباط بين فقرات الاختبار. وتم إيجاد قيم الجذور الكامنة (Eigenvalues)، ونسب التباين المفسر لكل عامل من العوامل، كما تم تدوير المحاور باستخدام طريقة التدوير المتعامد (Varimax Rotation) للعوامل التي قيم جذرها الكامن (≤1)، وذلك اعتماداً على محك (Kaiser's Criterion)، وحُددت جوهرية العامل باحتوائه على ثلاثة تشبعات جوهرية على الأقل، وقد استخدمت القيمة (0.30) كحد أدنى لقبول تشبعات الفقرات بالعوامل (Factor Loading)، ووفقاً لهذه الإجراءات فسرت ما نسبته (44.920%) من التباين على اختبار الذكاء لكاتل المتحرر من أثر الثقافة "المقياس الثالث الصورة (أ)"، كما هو مبين بالجدول رقم (1).

جدول (1):

قيم الجذر الكامن (Eigenvalues) ونسبة التباين المفسر للتحليل العاملي الاستكشافي

العامل	الجذور الكامنة الابتدائية			مجموع مربعات التشبعات المستخلصة (الجذور الكامنة النهائية)		
	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمي
1	5.700	11.400	11.400	5.700	11.400	11.400
2	1.755	3.510	14.910	1.755	3.510	14.910
3	1.332	2.664	17.574	1.332	2.664	17.574
4	1.239	2.479	20.053	1.239	2.479	20.053
5	1.220	2.440	22.492	1.220	2.440	22.492
6	1.200	2.399	24.891	1.200	2.399	24.891
7	1.168	2.335	27.226	1.168	2.335	27.226
8	1.132	2.264	29.490	1.132	2.264	29.490
9	1.116	2.232	31.723	1.116	2.232	31.723
10	1.103	2.206	33.929	1.103	2.206	33.929
11	1.076	2.151	36.080	1.076	2.151	36.080
12	1.048	2.096	38.176	1.048	2.096	38.176
13	1.039	2.079	40.255	1.039	2.079	40.255
14	1.011	2.023	42.278	1.011	2.023	42.278
15	1.006	2.013	44.290	1.006	2.013	44.290
16	0.994	1.988	46.279			

وقد تبين أن الفقرات التي تشبعت تشبعاً دالاً على العامل الأول بلغت (24) فقرة، وهذا يؤكد تمركز فقرات الاختبار حول عامل عام تشترك فيه فقرات الاختبار، وهو المُؤشر الأول لأحادية البعد

للاختبار، وقد تراوحت الجذور الكامنة لهذه العوامل بين (5.7) للعامل الأول و(1.755) للعامل الثاني، وهذا يُعد المُؤشر الثاني على أحادية البعد حسب مُؤشر لورد (Lord, 1980) إذ بلغت نسبة قيمة الجذر الكامن للعامل الأول إلى العامل الثاني حوالي (3:1) تقريباً، مما يعني أن هناك عامل واحد مهيمن، واعتماداً على نسبة التباين المفسر للعامل الأول، والذي بلغت (11.400) ويمثل ما نسبته (25.7%) من التباين الكلي، وظهر المُؤشر الثالث من أنَّهُ إذا استطاع العامل الأول أن يفسر ما نسبته (20%)على الأقل من التباين الكلي فإن ذلك يُعد أحد مؤشرات أحادية البعد (Wiberg, 2007)، وهذا ما يبدو واضحاً في المُؤشر الرابع من اختبار الفرز لـ (Scree Plot)، الموضح في الشكل البياني رقم (4)، والذي يظهر عاملاً واحداً فقط قبل نقطة القطع الأولى بانحدار شديد بين قيم الجذر الكامن الأول والثاني، بينما يقل الانحدار بشكلٍ واضح لبقية الجذور الكامنة للعوامل الأخرى، مُؤيداً لأحادية البعد للاختبار وخاصة لحجم العينات الكبيرة (Stevens, 2009)

الشكل البياني رقم (4) اختبار الفرز لـ Scree Plot

فيما أظهرت النتائج التشبعات العاملية (Loadings)، وقيم الشيوع أو الاشتراكيات (Communalities) كمحك خامس، أن أقل تشبع بلغ (416.) تقريبًا للمفردة (13) على العامل الأول، وأعلى تشبع بلغ (674.) للمفردة (10) على العامل الأول أيضًا وفقاً لمحك Guilford، ويلاحظ أن جميع تشبعات الفقرات على العوامل أعلى من (40.)، وهي تشبعات مرتفعة وفق المحك المفضل، أما قيم الاشتراكيات فقد تراوحت بين (306.) و(632.) للمفردتين (40)، و(10) على الترتيب بمتوسط بلغ (56.). ويفضل أن يكون متوسط قيم الشيوع أكبر من أو يساوي (60.)، وهي مناسبة وجيدة وتدل على أحادية البعد وفق هذا المحك.

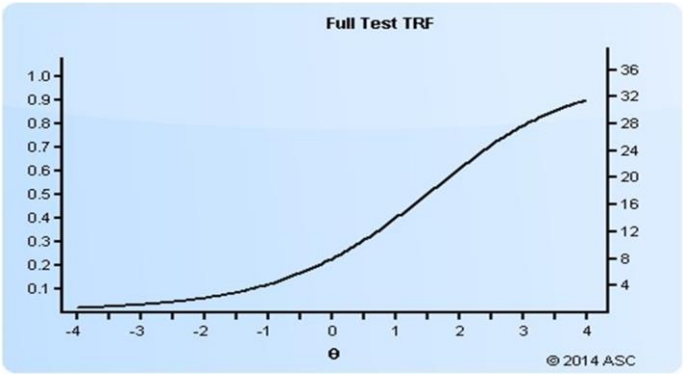
2- **الاستقلال الموضعي:**

يقصد بالاستقلال الموضعي الاستقلال الإحصائي لأداء المفحوصين على فقرات الاختبار، أي: أن أداء المفحوص على فقرة لا يتأثر بأدائه على فقرة أخرى (Brailovsky & Lawson,2006) ويرى هامبلتون

وسـوامنيثان (Hambleton&Swaminthan,1985) أنَّ مفهـوم الاسـتقلال الموضـعي وأحاديـة البعـد متكافئان؛ وبالتالي، فتحقق أحدها يضمن تحقق الآخر، وهو ما يؤكده لورد (lord,1980) (كروكر والجينا، 2017/1986) بأنَّ الاستقلال الموضعي يمكن الاستدلال عليه من خلال أحادية البعد، فتحقيقه يعني تحقق افتراض الاسـتقلال الموضـعي، وهـو مـا أوصـى بـه هـاملتون وسـوامينثان ووورم (Hambleton & Swaminathan, 2010; Warm, 1978). إذ تم استخراج مؤشر (Q3) الذي اقترحه ين (Yen, 1984)، من خـلال حسـاب معامـل ارتبـاط بيرسـون للبـواقي بـين الفقـرات مـن خـلال المعادلـة التاليـة:$Q3_{jj} = r(d_j, d_j)$ وأظهرت النتائج أن أعلى قيمة كانت بين الفقرة رقم (40) والفقرة رقم (32) إذ بلغت القيمة (372.)، بينما نجد باقي الارتباطات تقل عن هذه القيمة، وهي تقل عن القيمة (50.) كمحك، وهو مُؤشر يدل على أَنَّهُ لا يوجد فقرات تتأثر الإجابة عليها بفقرةٌ أخرى من فقرات الاختبار (Onder, 2007).

3- المنحنى المميز للفقرات (ICC):

تفترض النظريـة وجـود دالـة مميزة يتخـذ كل منهـا شـكل منحنى الترجيح اللوغـاريتمي الاحتمـالي، وتتوفر للاختبار ككل ولكل فقرة على حدة؛ والذي يسمى بمنحنى خصائص الاختبار والفقرة، يمثل المنحنى المميز للاختبار وصف العلاقة بين كل من القدرة واحتمالية الاستجابة الصحيحة ممثلة بالدرجة المتوقعة للمفحوصـــــين، ويوضـــح الشـــكل التالي تمثيل منحنى خصـائص الاختبـار بيانيـــاً باســـتخدام برنـــامج اكســـكليبر (xcalibre).



الشكل البياني (5) المنحنى المميز للاختبار

4- التحرر من السرعة:

تم التحقق من هذا الافتراض وفقاً لما ذكره هامبلتون وسوامينثان وروجرز (1991) "Hambleton & Swaminthan and Rogers" أنه إذا كان (75%) من المفحوصين أكملوا الإجابة عن الاختبار أو تم الإجابة على (80%) من فقرات الاختبار فإن السرعة لا تُعد في هذه الحالة عاملاً، وهذا ما تم تحقيقه بإجابة (95.5

%) من المفحوصين أكملوا الإجابة عن الاختبار بعد حذف الاستجابات الناقصة لعدد (105) مفحوص، فيما كانت إجابة العينة على جميع فقرات الاختبار، بالإضافة إلى ما أكدته نتائج التحقق من أحادية البعد، كون وجود تأثير للسرعة سيظهر عامل آخر ولا يتحقق شرط أحادية البعد، وللتحقق من هذا الافتراض يتم حساب الزمن المحدد للاختبار وإتاحة الوقت الكافي الذي يُمكّن جميع أفراد العينة من الإجابة عن جميع فقرات الاختبار، مما يؤكد تحقق افتراض التحرر من السرعة.

نتائج الإجابة عن السؤال الأول:

والذي ينص على "هل توجد علاقة بين الأداء التفاضلي للفقرة، باستخدام طريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، ومؤشر إحصائي المطابقة للفقرة وأقصى معلومات للفقرة والخطأ المعياري في التقدير لصعوبة الفقرة باستخدام النموذج الثلاثي المعلم في نظرية الاستجابة للفقرة (3PLM)؟

للإجابة عن هذا السؤال تم استخراج العلاقة باستخدام معامل ارتباط إيتا (Eta correlation coefficient)، وأظهرت النتائج أنه توجد علاقة دالة إِحْصَائِيَاً بين الأداء التفاضلي للفقرة باستخدام طريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، ومؤشر إحصائي المطابقة للفقرة، للنموذج الثلاثي المعلم (3PLM) بمعامل ارتباط بلغ (201.) وهو معامل ارتباط طردي ضعيف، كون البواقي المعيارية (Residuals) مؤشراً على مطابقة النموذج للبيانات، وهي نتيجة تؤيد ما تفترضه نظرية الاستجابة للفقرة بأن ضعف مطابقة الفقرة مؤشر للشك في معالم الفقرة المقدرة في النموذج الثلاثي المعلم (3PLM)، ولا يستدل بها على نمط استجابة المفحوص على فقرات الاختبار، فيما يُعد ضعف مطابقة الفرد كذلك مؤشراً على الشك في مستوى القدرة المقدرة، ومؤشراً ضعيفاً على موقع الفرد على متصل القدرة للسمة المقاسة (Reise, 1990)، فيما توجد علاقة دالة إِحْصَائِيَاً بين الأداء التفاضلي للفقرة باستخدام طريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، وأقصى معلومات، للنموذج الثلاثي المعلم (3PLM) بمعامل ارتباط بلغ (351.)، وهو معامل ارتباط طردي متوسط أيضاً، وهي نتيجة تؤيد ما تفترضه نظرية الاستجابة للفقرة باعتبار أن أقصى معلومات للفقرة تمثل نقطة واحدة على مقياس التدريج، وأن العلاقة فعلاً طردية بين كمية المعلومات التي تقدمها الفقرة والقدرة التمييزية لها، وتكون عندها كمية المعلومات التي تشارك بها الفقرة في تقدير قدرة المفحوص $\boldsymbol{\theta}$ كبيرة، وقد تكون انعكاسًا لحجم القيم اللاصفرية لمعلم التخمين $\boldsymbol{\chi_j}$، وأن التقدير الضعيف لواحد أو أكثر من معالم الفقرة يؤثر في تقدير المعالم الأخرى، وتتناسب عكسياً كلما كان معلم التخمين $\boldsymbol{\chi_j}$ أكبر انخفضت أقصى معلومات للفقرة، أي أن فاعلية تمييز الفقرة تتأثر بحجم معلم التخمين $\boldsymbol{\chi_j}$ بافتراض ثبات أية

عوامل أخرى، فكلما زادت معلومات الفقرة زادت قدرتها التمييزية بين الأفراد الأمر الذي يؤدي إلى ظهور العلاقة بين أقصى معلومات وظهور مؤشر للأداء المختلف حتى بين المجموعات لذوي القدرة المتساوية.

وفيما توجد علاقة دالة إِحْصَائِيَاً بين الأداء التفاضلي للفقرة باستخدام طريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، والخطأ المعياري في التقدير لصعوبة الفقرة، للنموذج الثلاثي المعلم (3PLM) بمعامل ارتباط بلغ (316.) وهو معامل ارتباط طردي متوسط أيضاً، فكلما زاد الخطأ المعياري زادت احتمالية ظهور مؤشر الأداء التفاضلي للفقرة، وهي نتيجة تتوافق مع تنظير منطق القياس على أنه كلما زاد الخطأ المعياري لتقدير صعوبة الفقرة قل الثبات، وهو مؤشر للشك في تقدير موقع الفقرة الأمر الذي ينعكس على تقدير معلم القدرة عند الأفراد إذ يكون موضعاً للشك في صحة وسلامة قدرة وموقع الأفراد، أي أنه قد يتوقع أن قدرات الأفراد باختلاف مجموعاتهم غير متساوية على مقياس التدريج، وبهذا فإن الفقرة لا تتمتع بخاصية اللاتغاير ويكون ظهور مؤشر الأداء التفاضلي للفقرة نتيجة حتمية.

نتائج الإجابة عن السؤال الثاني:

والذي ينص على "ما أثر كل من إحصائي المطابقة، وأقصى معلومات للفقرة والخطأ المعياري في تقدير لصعوبة الفقرة، على الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية باستخدام النموذج الثلاثي المعلم (3PLM) في نظرية الاستجابة للفقرة"؟

وللإجابة عن هذا التساؤل تم استخدم تحليل الانحدار اللوجستي الثنائي (Binary Logistic) كون المتغير التابع هنا: الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية متغير تصنيفي ثنائي (يوجد& لا يوجد)، وأظهرت النتائج أن مؤشر الأداء التفاضلي للفقرة بطريقة نسبة الأَرْجَحِية *(Likelihood-Ratio)*، (يوجد - لا يوجد) في المرحلة الأولية التي تعتمد على الثابت فقط (38) يوجد مقابل (13) لا يوجد كما ظهرت في الحالات المشاهدة. فيما تبين أن محاولات التقدير للإمكان لثابت النموذج (Log likelihood) 2-بلغ (57.901) والذي استقر عند الخطوة الرابعة بسبب أن الفرق في التغير في تقدير المعالم أصبح لا يذكر وأقل من (0.001)، وبلغت قيمة ثابت النموذج ودلالته الإحصائية التي تعتمد على الثابت فقط (1.073) وبلغ إحصائي والد (11.45) وهي دالة إِحْصَائِيَاً عند مستوى (001.)، مما يدل على تأثير الثابت في النموذج تأثيراً يختلف عن الصفر، وبذلك ترفض الفرضية الصفرية.

فيما تبين من نتائج الخطوة الثانية التي تتمثل في النموذج الثاني المتضمن لجميع المتغيرات المستقلة (إحصائي المطابقة، وأقصى معلومات للفقرة والخطأ المعياري في تقدير لصعوبة الفقرة)، أن نتائج تاريخ محاولات التقدير للإمكان للنموذج (Log likelihood) 2- والذي استقر في الخطوة السادسة كمحاولة

أخيرة في تقدير الإمكان الأقصى والذي بلغ (45.846)، وأن معلماتها تمثل أفضل نتيجة لأن سالب ضعف دالة الإمكان في نهايتها الصغرى، وهي قيمة أقل من تاريخ محاولات التقدير للإمكان للثابت، مما يفيد بجودة توافق النموذج.

وبلغ اختبار امينيبوس (Omnibus) الخاص بالمراحل الثلاث للكشف عن دلالة النموذج، إذ بلغت قيمة مربع كاي للنموذج (12.055) وهي دالة إِحْصَائِيَاً، أي أن النموذج يمثل البيانات تمثيلاً جيداً، مما يؤكد تأثير المتغيرات المستقلة وقدرتها التنبؤية في المتغير التابع.

وتظهر نتائج ملخص نموذج الانحدار الدال على جودة توافق النموذج مع البيانات أن قيمة نقلكرك (Nagelkerke R Square) التي تناضر معامل التحديد في الانحدار الخطي بلغت (.310) تقريباً وتشير إلى نسبة التباين المفسر في المتغير التابع والذي يمكن أن يعزى الى المتغيرات المستقلة المتضمنة في النموذج، أي أن هناك نسبة تقدر بـ (31%) من تصنيف المتغير التابع تعزى للمتغيرات المستقلة.

أما قيمة ودلالة اختبار هوسمر وليميشو (Hosmer and Lemeshow Test)، والذي بلغ قيمة مربع كاي (2.862) وهي قيمة غير دالة إِحْصَائِيَاً إذ بلغت قيمة sig(.943) مما يشير إلى أن الحالات المشاهدة لا تختلف عن الحالات المتوقعة المتنبأ بها كدليل على توافق نموذج الانحدار بصورة جيدة مع واقع البيانات المستخدمة.

وقد أظهرت نتائج التكرارات الملاحظة توافق بالتكرارات المتنبأ بها لاختبار (Hosmer and Lemeshow Test) إذ استقرت على الخطوة العاشرة حيث تبين تساوي الفرق بين التكرارات الملاحظة والمتنبأ بها إذ بلغت (.099) كفارق بين لا يوجد أداء تفاضلي (0.00-0.099)، وهو نفس الفرق بين يوجد أداء تفاضلي (5.901-6.00). والجدول (2) يوضح ذلك.

جدول (2)

تكرارات تصنيف ظهور مؤشر الأداء التفاضلي للفقرة (يوجد – لا يوجد) في ضوء البيانات المشاهدة والمتنبأ بها

التنبؤ					
المشاهدة		الأداء التفاضلي		المجموع	النسب الصحيحة
		لا يوجد	يوجد		
الأداء التفاضلي	لا يوجد	3	10	13	32.1
	يوجد	2	36	38	74.7
النسبة الكلية					76.5

فيما يظهر من الجدول (2) تكرارات تصنيف الفقرات في ضوء البيانات المشاهدة والمتنبأ بها من قبل النموذج إذ بلغت نسبة التصنيف الصحيح الكفاءة نسبة التصنيف الصحيح "(Et) Efficiency Specificity" للنموذج لظهور مؤشر الأداء التفاضلي (76.5%) مما يشير إلى جودة النموذج في تصنيف ظهور مؤشر الأداء اعتماداً على هذه المتغيرات المستقلة وقيمة أفضل من النموذج الثابت، إذ بلغت دقة التصنيف للنموذج أو النوعية (Specificity) لظهور مؤشر الأداء التفاضلي للفقرة (94.74%)، وهو مؤشر مميز للقدرة التنبؤية للمتغيرات المستقلة. والجدول (3) يوضح ذلك.

جدول (3)

قيمة ثابت النموذج والمتغيرات المستقلة ودلالتها الإحصائية

Exp(B)	Sig	df	Wald	S.E.	B	المتغيرات
0.317	0.189	1	1.726	0.873	-1.148	إحصائي المطابقة للفقرة
8.274	0.048	1	3.899	1.920	3.790	أقصى معلومات للفقرة
0.007	0.709	1	0.139	13.494	-5.027	الخطأ المعياري في تقدير صعوبة الفقرة
3.408	0.412	1	0.672	1.495	1.226	الثابت

يظهر من الجدول (3) معاملات الانحدار (B)، للمتغيرات المستقلة والثابت من خلال العمود (B)، وهي القيم التي تستخدم في معادلة التنبؤ، والتي تتضمن كل المتغيرات المستقلة الدالة مع قيمة الثابت. يتضح من النتائج أن قيمة معامل الانحدار لمتغير إحصائي المطابقة للفقرة (1.148-)، وبلغت قيمة إحصائي والد Wald(1.726)، وهي قيمة غير دالة إِحْصَائِيَاً عند مستوى دلالة (.189)، وبلغت نسبة الأَرْجَحِية *(Likelihood-Ratio)* (0.317) وهذا يُعد مؤشراً دالاً على صحة الفرض الصفري، وأن التأثير لهذا المتغير لا يختلف عن الصفر، فيما أظهرت النتائج أن متغير أقصى معلومات للفقرة هو المتغير الوحيد الذي له قدرة تنبؤية بتصنيف ظهور مؤشر الأداء التفاضلي للفقرة (يوجد – لا يوجد) إذ بلغت قيمة معامل الانحدار (3.790)، وبلغت قيمة إحصائي Wald (3.899)، وهي قيمة دالة إِحْصَائِيَاً عند مستوى دلالة (.045)، وهذا مؤشر على التأثير الدال لهذا المتغير، وأنه كلما ارتفعت كمية المعلومات التي تقدمها الفقرة والتي تظهر في مؤشر أقصى دالة معلومات عند أي نقطة صعوبة على مقياس التدريج لمواقع الفقرات وتقدير قدرة الأفراد بوحدة واحدة؛ زادت نسبة أرجحية ظهور مؤشر الأداء التفاضلي للفقرة بمقدار (8.274)، فيما ظهرت قيمة معامل الانحدار لمتغير الخطأ المعياري في تقدير صعوبة الفقرة (5.027-)، وبلغت قيمة إحصائي والد Wald

(139.)، وهي قيمة غير دالة إِحْصَائِيَاً عند مستوى دلالة (709.)، و بلغت نسبة الأَرْجَحِية *(Likelihood-Ratio* (0.007)، وهذا يعد مؤشراً دالاً على صحة الفرض الصفري، وأن التأثير لهذا المتغير لا يختلف عن الصفر، ويمكن صياغة معادلة التنبؤ لهذه المتغيرات واحتمالية ظهور مؤشر الأداء التفاضلي للفقرة التفاضلي والتي تظهر بالمعادلة التنبؤية التالية:

$$log(p/1-p) = b0 + b1*x_1 + b2*x_2 + b3*x_3$$

log(p/1-p) = -1.226-1.148* إحصائي المطابقة *3.790 + أقصى معلومات للفقرة *3.790 + الخطأ المعياري في تقدير صعوبة الفقرة

المراجع العربية والانجليزية:

أولاً: المراجع العربية:

بني عطا، زايد صالح إبراهيم (2018). أثر القيم المتطرفة على الأداء التفاضلي لفقرات اختبار الرياضيات في الدراسة الدولية تيمس تبعاً لمتغير الجنس، *العلوم التربوية، 45*، 157-176.

بيكر، فرانك (2010). *أسس نظرية الاستجابة للفقرة*. (الطريري، عبد الرحمن؛ أبو هاشم، السيد؛ شلبي، سوسن، مترجمون) دار جامعة الملك سعود للنشر (الكتاب الأصلي منشور 2001 Baker, B).

تيغزة، أمحمد بوزيان (2009). نظرية الصدق الحديثة ومتضمناتها التطويرية لواقع القياس، *ندوة علم النفس والتنمية الفردية والمجتمعية*، جامعة الملك سعود 1430هـ.

أبو حطب، فؤاد عبد اللطيف؛ صادق، آمال مختار (2005). *اختبار كاتل للعامل العام، مقياس الذكاء المتحرر من أثر الثقافة "المقياس الثالث الصورة (أ)*. مكتبة الأنجلو المصرية.

دي إيالا، أر جي (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة*. (الكيلاني، عبد الله؛ البرصان، إسماعيل، المترجمان). دار جامعة الملك سعود للنشر (الكتاب الأصلي منشور. *(Ayala, 2009*

صالح، عامر مهدي. (2022). أثر حجم العينة في الأداء التفاضلي للمفردة على وفق أنموذج ثلاثي المعلم. *مجلة الجامعة العراقية، 55* (3)، 564-580.

ضعضع، هبة عبد اللطيف. (2023). أثر اختلاف نموذج الاستجابة للفقرة (1PL,2PL,3PL) في الإداء التفاضلي للفقرة. *مجلة الاستاذ للعلوم الانسانية والاجتماعية. 62*(1)، 23-45.

العبد الله، زياد أحمد. (2022). أثر حجم العينة في الأداء التفاضلي للفقرة وفقا لنظرية الاستجابة للفقرة .*مجلة البحوث التربوية والنفسية، 19* (72)، 119-143.

عبد الوهاب، محمد محمود (2014). الكشف عن الأداء التفاضلي على فقرات مقياس أساليب المعاملة الوالدية بين الآباء والأمهات وبين الأبناء من الجنسين. *دراسات عربية في التربية وعلم النفس، 45*(4)، 155-178.

أبو علام، رجاء محمود. (2011). *مناهج البحث في العلوم النفسية والتربوية*، دار النشر للجامعات.

علي، سعدي أحمد وعيال، ياسين حميد. (2021). تأثير اختلاف مستوى القدرة في الأداء التفاضلي وفقاً للأنموذج ثنائي المعلم لنظرية الاستجابة للفقرة. *مجلة الآداب*،*1* (138)، 137-162.

قيلي ح. (2020). التنبؤ العلمي. مجلة الحكمة للدراسات الفلسفية، 8(2)، 70-87.

كروكر، والجينا (2017). *مدخل إلى نظرية القياس التقليدية والمعاصرة* (الحموري، هند؛ دعنا، زينات، المترجمتان). دار الفكر ناشرون وموزعون، ط1، (الكتاب الأصلي منشور (Crocker, L., & Algina, J, 1986) .

المحروقية، زينب بنت محمد بن حمود، عمارة، إيهاب محمد نجيب، وابن كاظم علي بن مهدي. (2024) *الأداء التفاضلي لمفردات اختبار الرياضيات للصفين الرابع والثامن في الدراسة الدولية TIMSS 2019 تبعاً المتغير الجنس في سلطنة عمان* (رسالة ماجستير غير منشورة). جامعة السلطان قابوس

مكتب التربية والتعليم بمحافظة ذمار (2019). إحصائيات طلبة المرحلة الثانوية، إدارة الإحصاء.

نتكو، أنثوني؛ بروكهارت، سوزان. (2018). *التقويم التربوي للطلبة* (ترجمة علي القرني وآخرين)، مكتب التربية العربي لدول الخليج (الكتاب الأصلي منشور (Anthony J.at all, 2007).

النفيعي، عبد الرحمن عبد الله (2008). *مقارنة فاعلية عدد من الطرق الإحصائية للكشف عن الأداء التفاضلي لفقرات الاختبارات وأثر ذلك على الخطأ المعياري من النوع الأول وقوة الاختبار*. [رسالة دكتوراه غير منشورة] . جامعة أم القرى.

Arabic references

Banī ʿAṭā, Zāyid Ṣāliḥ Ibrāhīm (2018). Athar al-Qayyim al-mutaṭarrifah ʿalá al-adāʾ altfāḍly lfqrāt ikhtibār al-riyāḍīyāt fī al-dirāsah al-Dawlīyah tyms tbʿan lmtghyr al-jins, al-*ʿUlūm al-Tarbawīyah, 45*, 157-176.

Bīkar, Frānk (2010). *Usus Naẓarīyat al-istijābah llfqrh*. (al-Ṭurayrī, ʿAbd al-Raḥmān ; Abū Hāshim, al-Sayyid ; Shalabī, Sawsan, mtrjmwn) Dār Jāmiʿat al-Malik Saʿūd lil-Nashr (al-Kitāb al-aṣlī manshūr Baker, B 2001).

Tyghzh, Amḥammad Būziyān (2009). Naẓarīyat al-ṣidq al-ḥadīthah wmtḍmnāthā al-taṭwīrīyah li-wāqiʿ al-qiyās, *Nadwat ʿilm al-nafs wa-al-tanmiyah al-fardīyah wa-al-mujtamaʿīyah,* Jāmiʿat al-Malik Saʿūd, al-Riyāḍ..

Abū Ḥaṭab, Fuʾād ʿAbd al-Laṭīf ; Ṣādiq, Āmāl Mukhtār (2005). *ikhtibār kātl lil-ʿāmil al-ʿāmm, miqyās al-dhakāʾ almtḥrr min Athar al-Thaqāfah "al-miqyās al-thālith al-Ṣūrah (U)" "*, Maktabat al-Anjlū al-Miṣrīyah.

Dī iyālā, ar Jī (2017). *al-naẓarīyah wa-al-taṭbīq fī Naẓarīyat al-istijābah llfqrh.* (al-Kīlānī, ʿAbd Allāh ; al-Burṣān, Ismāʿīl, almtrjmān). Dār Jāmiʿat al-Malik Saʿūd lil-Nashr (al-Kitāb al-aṣlī manshūr. Ayala, 2009)

Ṣāliḥ, ʿĀmir Mahdī. (2022). Athar ḥajm al-ʿayyinah fī al-adāʾ altfāḍly llmfrdh ʿalá wafqa Unmūdhaj thulāthī al-Muʿallim. *Majallat al-Jāmiʿah al-ʿIrāqīyah, 55* (3), 564-580.

Ḍʿḍʿ, Hibat ʿAbd al-Laṭīf. (2023). Athar ikhtilāf namūdhaj al-istijābah li-L Faqrah (1PL, 2PL, 3PL) fī alʾdāʾ altfāḍly li-L Faqrah. *Majallat al-Ustādh lil-ʿUlūm al-Insānīyah wa-al-Ijtimāʿīyah. 62* (1), 23-45.

Al-ʿAbd Allāh, Ziyād Aḥmad. (2022). Athar ḥajm al-ʿayyinah fī al-adāʾ altfāḍly li-L Faqrah wafqan li-naẓarīyat al-istijābah llfqrh. *Majallat al-Buḥūth al-Tarbawīyah wa-al-nafsīyah. 19* (72), 119-143.

ʿAbd al-Wahhāb, Muḥammad Maḥmūd (2014). al-kashf ʿan al-adāʾ altfāḍly ʿalá fqrāt miqyās Asālīb al-Muʿāmalah al-wālidīyah bayna al-Ābāʾ wa-al-ummahāt wa-bayna al-abnāʾ min al-jinsayn. *Dirāsāt ʿArabīyah fī al-Tarbiyah wa-ʿilm al-nafs, 45* (4), 155-178.

Abū ʿAllām, Rajāʾ Maḥmūd. (2011). *Manāhij al-Baḥth fī al-ʿUlūm al-nafsīyah wa-al-tarbawīyah*, Dār al-Nashr lil-Jāmiʿāt.

ʿAlī, Saʿdī Aḥmad wa-ʿiyāl, Yāsīn Ḥamīd. (2021). Taʾthīr ikhtilāf mustawá al-qudrah fī al-adāʾ altfāḍly wafqan llʾnmwdhj thunāʾī al-Muʿallim li-naẓarīyat al-istijābah llfqrh. *Majallat al-Ādāb, 1 (138),* 137-162.

Krwkr, wāljynā (2017). *madkhal ilá Naẓarīyat al-qiyās al-taqlīdīyah wa-al-muʿāṣarah* (al-Ḥammūrī, Hind ; Daʿnā, Zīnāt, almtrjmtān). Dār al-Fikr Nāshirūn wa-Muwazziʿūn, Ṭ1, (al-Kitāb al-aṣlī manshūr Crocker, L., & Algina, J, 1986).

Almḥrwqyh, Zaynab bint Muḥammad ibn Ḥammūd, ʿImārah, Īhāb Muḥammad Najīb, wa-Ibn Kāẓim ʿAlī ibn Mahdī. (2024) al-adāʾ altfāḍly li-mufradāt ikhtibār al-riyāḍīyāt llṣfyn al-rābiʿ wa-al-thāmin fī al-dirāsah al-Dawlīyah TIMSS2019 tbʿan al-mutaghayyir al-jins fī Salṭanat ʿAmmān (Risālat mājistīr ghayr manshūrah). Jāmiʿat al-Sulṭān Qābūs

Maktab al-Tarbiyah wa-al-taʿlīm bi-Muḥāfaẓat Dhamār (2019). iḥṣāʾīyāt ṭalabat al-marḥalah al-thānawīyah, Idārat al-Iḥṣāʾ.

Ntkw, anthwny ; brwkhārt, Sūzān. (2012). *al-Taqwīm al-tarbawī lil-Ṭalabah* (tarjamat ʿAlī al-Quranī wa-ākharīn), Maktab al-Tarbiyah al-ʿArabī li-Duwal al-Khalīj (al-Kitāb al-aṣlī manshūr (Anthony J. at all, 2007).

Al-Nufayʿī, ʿAbd al-Raḥmān ʿAbd Allāh (2008). *muqāranah fāʿilīyat ʿadad min al-ṭuruq al-iḥṣāʾīyah lil-kashf ʿan al-adāʾ Altfāḍly lfqrāt al-ikhtibārāt wa-athar dhālika ʿalá al-khaṭaʾ al-miʿyārī min al-nawʿ al-Awwal wa-qūwat al-ikhtibār*. [Risālat duktūrāh ghayr manshūrah]. Jāmiʿat Umm al-Qurá.

ثانياً: المراجع الأجنبية:

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing. American Psychological Association.*

Cains, J.;Bridglall, B.;& Chatterji,M. (2014).Understanding validity and fairness issues in high-stakes individual testing situations. *Quality Assurance in Education* ,22(1),5-18.

Camili, G & Shepard, L.A.(1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publication

Edward, H. (2010). Interpretation of the Three-Parameter Testlet Response Model and Information Function. *Applied Psychological Measurement, 34* (7): pp467-482. https://doi.org/10.1177/0146621610364975

Ellis, B. B., & Raju, N. S. (2004). Test and item bias: what they are, what aren't, and how to detect them: measuring up. In J. Wall and G. Walz (Eds.), Measuring Up*: Assessment Issues for Teachers, Counselors, and Administrators (pp. 89-98).* Greensboro: *CAPS Press.*

Greer, T. G. (2004). *Detection of diff erential item functioning (DIF) on the SATV: Acomparison of four methods: Mantel-Haenszel, logistic regression, simultaneous item bias and likelihood ratio test*. Unpublished doctoral dissertation, University of Houston.

Gruijter, D. and Kamp, L. (2005). *Statistical Test Theory for Education and Psychology*, Retrieved December 30, 2005.

Hambleton R., Jones, R. (1993), Comparison of classical test theory and item response theory and their applications to test development, NCME items: *The Instructional topics in educational measurement series, Educational Measurement: Issues and Practice*, Module 16.

Hambleton, R. & Murray. L. (1983). Some goodness of fit investigations for response models. In R.K. Hambleton (ED.), *Applications of item response theory* (PP. 71-94). Vancouver, British Columbia.

Hambleton, R., & Swaminathan, H. (1985). *Item response theory, principles and applications. Kluwer* -Nijhoff, USA.

Hambleton, R., Swaminathan, H. (2010). *Item Response Theory. Principles and Application. Boston*: Kluwer-Nigh off Publishing.

Hambleton, R., Swaminthan, H., & Rogers, J. (1991). *Fundamentals of item response theory. Newbury park*: Sage publications.

Karami, H. (2011). *Detecting Gender Bias in A Language Proficiency Test. International Journal of Language Studies,* 5:2,27-38.

Lai, H., & Krishnan, V. (2013). *Applications of Differential Item Functioning (DIF) and Natural Language Processing (NLP) on Alberta's Early Child Development Instrument (EDI) data, Early Childhood Mapping (ECMap) Project, Community-University Partnership (CUP), Faculty of Extension,* University of Alberta, Edmonton, Alberta, Canada.

Lord, F. (1980). *Application of Item Response Theory to Practical Testing Problems*. NJ: Lawrence Erlbaum. Google Scholar

Maller, S. J. (2003). Best practices in detecting bias in nonverbal tests. *In Handbook of nonverbal assessment* (pp. 23-47). Boston, MA: Springer US.

McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement, 38*, 75–79.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Petersen,M.A Groenvold, M, Bjorner, J.B, Aaronson, N, Conroy, T, Cull, A, Fayers, P, Hjermstad, M, Sprangers, M and Sullivan, M, (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Quality of Life Research 12: 373–385.*

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, *53*(4), 495-502. https://doi.org/10.1007/BF02294403

Reise, S. P., & Waller (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*, 45—58

Rose, B. Bjorner, J. Becker, F. Friesc, E. (2008). Evaluation of A Preliminary Physical Function Item Bank Supported the Expected Advantages of the Patient-Reported Outcomes Measurement Information System (Promise), *Journal of clinical Epidemiology, 61*, 17–33. https://doi.org/10.1016/j.jclinepi.2006.06.025

Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I error performance. **Journal of Educational Measurement, 33**(2), 215–230. https://doi.org/10.1111/j.1745-3984.1996.tb00490.x

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph*, 17

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*. 85–100. New York: Springer.

Siang C. & Fritz D.(2006). How Big Is Big Enough? Sample Size Requirements for CAST Item Parameter Estimation*, Applied Psychological Measurement, 19(3):pp 241- 255.*

Sideridis, g and Tsaousis, I, (2013a). DIF Analysis for Item and Test on the NCA Tests The General Ability Test (GAT) Art Major. *National Center for Assessment in Higher Education. TR035-2013.*

Stevens, J.P. (2009*). Applied multivariate statistics for the social sciences*. New York: Taylor & Francis.

Thissen, D. (2001). *IRTLRDIF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. University of North Carolina at Chapel Hill.

Warm, A. (1978). *A primer of Item Response Theory*: US. Coast Guard Institute Oklahoma, 73/69.

Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test*: A theoretic comparison of methods. EM No 60. Retreived August 20, 2017. from.www.edusci.umuse/digitalAssets/159/59534-emno -60.

الملاحق:

جدول (1):

معاملات الارتباط باستخدام مربع إيتا بين الأداء التفاضلي للفقرة، ومؤشر إحصائي المطابقة للفقرة وأقصى معلومات للفقرة والخطأ المعياري في التقدير لصعوبة الفقرة بطريقة نسبة الأرجحية

	إحصائي المطابقة	أقصى معلومات للفقرة	الخطأ المعياري في التقدير لصعوبة الفقرة
النموذج الثلاثي المعلم	**.201**	**.351**	**.316**

جدول (3)

تصنيف ظهور مؤشر الأداء التفاضلي للفقرة (يوجد – لا يوجد) في النموذج الثابت

النموذج			Predicted	
			DIF	Percentage Correct
			لا يوجد	يوجد
3PLM	**DIF**	لا يوجد	**0**	13	0
		يوجد	**0**	38	100.00
	النسبة الكلية				74.5

جدول (4)

محاولات تقدير الإمكان الأكبر لثابت النموذج (Step 0)

	Iteration	-2 Log likelihood	Coefficients Constant
3PLM	4	57.901	1.073

جدول (5)

قيمة ثابت النموذج ودلالته الإحصائية

Step		B	S.E.	Wald	df	Sig.	Exp(B)
Constant	**3PLM**	1.073	0.321	**11.145**	1	**0.001**	2.923

جدول (8)

ملخص النموذج للانحدار (اختبار جودة وإحصاءات توافق النموذج مع البيانات)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
3PLM	45.846[a]	0.211	0.310

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

جدول (7)

اختبار امينيبوس Omnibus للكشف عن دلالة النموذج

Sig.	df	Chi-square		
0.007	3	12.055	Step	Step 1
0.007	3	12.055	Block	
0.007	3	12.055	Model	

جدول (6)

إعادة تقديرات معالم محاولات تقدير الإمكان الأكبر للنموذج بالمتغيرات المستقلة (Step 1)

النموذج			-2 Log likelihood	Coefficients			
				Constant	إحصائي المطابقة للفقرة	أقصى معلومات للفقرة	الخطأ المعياري في التقدير لصعوبة الفقرة
الأداء التفاضلي للفقرة	Step 1	1	48.940	1.445	-0.772	1.358	-5.745
		2	46.306	1.437	-1.077	2.691	-5.705
		3	45.869	1.297	-1.150	3.526	-5.278
		4	45.846	1.231	-1.148	3.773	-5.045
		5	45.846	1.226	-1.148	3.790	-5.027
		6	**45.846**	1.226	-1.148	3.790	-5.027