

Impact of the Type and Percentage of Differential Item Functioning on the Ability Parameter of Individuals According to Parametric and Nonparametric Models of IRT

أثر نوع ونسبة الفقرات ذات الأداء التفاضلي على معلمة القدرة للأفراد وفق النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي لنظرية الاستجابة للفقرة

Issra Al-Khatib^{1*}, Hassan Al-Omari¹.

¹The University of Jordan, Amman, Jordan.

ARTICLE INFO

Article history:

Received 10 Feb 2022

Accepted 18 Apr 2022

Published 01 Jan 2024

*Corresponding author:

The University of Jordan, Amman, Jordan.

Email: asra32915@gmail.com.

Abstract

This study aimed to investigate the impact of the type and percentage of Differential Item Functioning (10%, 20%, and 30%) on estimating individuals' abilities according to the three-parameter model and Mokken's non-parametric model of item response theory. The experimental method was used to address the study's questions by applying hypothetical tests, each consisting of 60 double-response items, generated using the WinGen software, to a sample of 2,000 hypothetical individuals for each experimental condition. The study results showed that differences in the ability parameter between the three-parameter logistic model and Mokken's non-parametric model were not statistically significant across all experimental conditions related to the type and percentage of Differential Item Functioning. There were also no differences in the ability parameter due to the type of model according to the experimental conditions related to the type and percentage of differential performance. The study recommended using Mokken's non-parametric model when seeking the highest degree of reliability in the test.

Keywords: Differential Items Functioning, Abilities of Individuals, Three-Parameter Logistic Model, Moken Model.

المخلص

هدفت هذه الدراسة إلى معرفة أثر نوع ونسبة الأداء التفاضلي (١٠٪، ٢٠٪، ٣٠٪) على تقدير قدرات الافراد وفق النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي لنظرية الاستجابة للفقرة، استخدم فيها المنهج التجريبي للإجابة عن تساؤلات الدراسة بتطبيق اختبارات افتراضية طول كل منها (٦٠) فقرة ثنائية الاستجابة تم توليدها باستخدام برمجية (WinGen)، على عينة مكونة من (٢٠٠٠) فرد افتراضي لكل ظرف تجريبي. اشارت نتائج الدراسة الى فروق في معلمة القدرة العائدة للنموذج الثلاثي ونموذج موكن اللامعلمي غير دالة إحصائياً ولكل الظروف التجريبية المتعلقة بنوع ونسبة

الأداء التفاضلي وعدم وجود فروق في معلمة القدرة تعزى لنوع النموذج وفقاً للظروف التجريبية المتعلقة بنوع ونسبة الأداء التفاضلي، وقد أوصت الدراسة باستخدام نموذج موكن اللامعلمي عند توخي أعلى درجات الثبات في الاختبار.

الكلمات المفتاحية: الأداء التفاضلي للفقرة، قدرات الافراد، النموذج الثلاثي المعلمة، نموذج موكن.

١. المقدمة

١,١ الإطار النظري والدراسات السابقة

ركزت جهود الباحثين ومطوري الاختبارات في مجال بناء وتطوير الاختبارات على القيام بتحديد الفقرات التي تؤدي الغرض من الاختبار في ضوء الصعوبة والتمييز والتخمين، وعلى الرغم من أهمية هذه الدلالات إلا أنها ليست كافية للحكم على صلاحية فقرات الاختبار ليحقق الغرض الذي أعد من أجله، إذ أن الفقرات يمكن أن تتأثر بعوامل أخرى كالجنس واللغة والدين والمستوى الاقتصادي أو الاجتماعي، إضافة لقدرة المفحوصين، مما قد يؤثر في النتائج فيظهر نمطاً متحيزاً نحو مجموعة ضد أخرى ويؤثر في القرارات المتخذة بناء عليه، وعليه توصف فقرات الاختبار أو المقياس بأنها متحيزة (Schmidt & Hunter, 1998).

وتعود بداية الدراسات التي بحثت في تحيز الفقرات الى بينيه (Binet) عندما لاحظ أثناء تطبيقه لاختبارات الذكاء أن أداء المفحوصين من البيض أفضل من السود، وأن أداء المفحوصين من الطبقات الاقتصادية والاجتماعية العليا أفضل من أداء المفحوصين من الطبقات الاقتصادية والاجتماعية الأدنى (البستجي، ٢٠٠٤)، ومنذ ذلك الوقت حظي مفهوم التحيز للمجموعات على اهتمام الباحثين في مجال القياس والتقويم النفسي والتربوي، والذي أطلق عليه فيما بعد الأداء التفاضلي (DIF)، أو مصطلح (DIP) بدلاً عن (DIF) (Ryan & Chiu, 2001)، فقد استخدم مفهوم الأداء التفاضلي (DIF) في الاختبارات النفسية والتربوية للكشف عن الفقرات المتحيزة وفق المتغيرات الديموغرافية المختلفة من مثل الجنس أو العرق حيث يمكن أن تفضل مجموعة عن أخرى من ذات مستوى القدرة، (Stark, et al., 2004)، وعليه فإن الفقرات ذات الأداء التفاضلي توصف بأنها متحيزة لمجموعة ما لعوامل لا تتعلق بالقدرة (Camilli, et al., 1994; Gruijter & Kamp, 2005)، لأن الأفراد الذين لديهم نفس القدرة لديهم احتمالات مختلفة للإجابة بشكل صحيح على الفقرة باختلاف المجموعة (Reynolds & Lowe, 2009; Warn, et al., 2014).

لقد أجريت دراسات عديدة للكشف عن الفقرات ذات الأداء التفاضلي من خلال طرق عدة، منها ما يستند على النظرية الكلاسيكية كطريقة مانتل هانزل وطريقة الصعوبة المحولة وطريقة الانحدار اللوجستي، ومنها ما يستند على نظرية الاستجابة للفقرة كطريقة مؤشر الفرق في المساحة بين منحنيات خصائص الفقرة، وطريقة نسبة الأرجحية وغيرها. وتصنف طرق الكشف عن الأداء التفاضلي حسب نظرية الاستجابة للفقرة (IRT) الى نوعين من الطرق، الطرق المعلمية (Parametric)، والطرق اللامعلمية (Nonparametric) (الشريفين، ٢٠١٨؛ Raju & Ellis, 2002)، حيث يشير مصطلح معلم الى أن البيانات تحقق افتراضات وشروط النموذج من أحادية البعد والاستقلال الموضوعي ومنحنى خصائص الفقرة والتحرر من السرعة، وعند انتهاك أحد هذه الافتراضات يتم اللجوء إلى الطرق اللامعلمية، ومن الطرق المستخدمة في الكشف عن الأداء التفاضلي في نظرية الاستجابة للفقرة (IRT) طريقة مؤشر الفرق في المساحة بين منحنيات خصائص الفقرة وطريقة نسبة الأرجحية وطريقة راجو وغيرها (Park, 2010). حيث يصنف الأداء التفاضلي للفقرة حسب مكان منحنى خصائص الفقرة الذي يربط بين احتمال الإجابة الصحيحة للمجموعتين المرجعية والمستهدفة حيث تشتمل المجموعة المستهدفة الفقرات ذات الأداء التفاضلي (المنتظم أو غير المنتظم)، أما المجموعة التي نقارن فيها فتسمى بالمجموعة المرجعية. ومن خلال المساحة العائدة للفروق العامودية بين المنحنيين يمكننا الحكم على مدى وجود أداء تفاضلي.

لقد اشارت الدراسات والأبحاث مثل (Hambleton & Rogers, 1986; Swaminathan & Rogers, 1990; pae, 2004; Karami & Salmani Nodoushan, 2011; Raquel, 2019) الى نوعين من الأداء التفاضلي هما الأداء التفاضلي المنتظم (UDIF) والذي يظهر عندما لا يحدث أي تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد؛ أي أن احتمال الإجابة الصحيحة يكون أكبر دائماً لإحدى المجموعتين عند جميع مستويات القدرة، وتكون معالم التمييز، والتخمين متساوية بين المجموعات، ولكن معالم الصعوبة لمنحنى خصائص الفقرة غير متساوية بين المجموعات، بينما يظهر النوع الثاني ويسمى الأداء التفاضلي غير المنتظم (NUDIF)، حينما يوجد تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها

الفرد؛ أي أن الفرق في احتمالية الاستجابة الصحيحة ليست متشابهة في مستويات القدرة، فقد يكون لصالح إحدى المجموعات عند مستوى قدرة معين، ولصالح مجموعة أخرى عند مستوى قدرة آخر، مع اختلاف في معالم التمييز مختلفة عبر المجموعات وتساوي في معالم الصعوبة والتخمين (Hanson, 1998).

النماذج اللامعلمية للاستجابة للفقرة (Non Parametric Item Response Theory Models)

وهي تلك النماذج التي لا تحدد شكل محدد لدالة استجابة الفقرة باستثناء ألا تكون متناقضة، كما أنها لا تفترض أي شكل سابق؛ كما يمكن الافتراض أن دوال الاستجابات التي يتم الحصول عليها من خلال النماذج اللامعلمية أقرب إلى الاستجابات الحقيقية التي يتم الحصول عليها من النماذج المعلمية؛ لأنها تعتمد على افتراضات أقل حول النموذج اللوجستي (Van der Linden & Hambleton, 1997).

ويشير سيجتسا ومولينار (Sijtsma & Molenaar, 2002) إلى أن النماذج اللامعلمية تقوم على مجموعة من الافتراضات على غرار النماذج المعلمية إلا أنها أقل تشدداً، وهي أحادية البعد (One-dimensionality) وتعني أن استجابات المفحوصين على الفقرات تتبع متغير كامن أحادي البعد يرمز له بالرمز θ ، والاستقلال الموضعي (Local Independence) وتتمثل بأن احتمالية الإجابة من قبل المفحوص على أي فقرة غير مرتبطة بالاستجابة على أي فقرة أخرى في الاختبار، والإطارية (Monotonicity) وتعني أنه كلما زادت قدرة المفحوص θ ، ازدادت احتمالية الإجابة الصحيحة على الفقرة أو تبقى ثابتة ضمن مستويات القدرة المختلفة، وأخيراً الإطارية المضاعفة (Double Monotonicity) والمتعلق بامتلاك دوال استجابة غير متقاطعة لفقرات الاختبار التي تشكل التدرج. وعلى هذا النحو تنقسم النماذج اللامعلمية إلى قسمين رئيسيين هما:

– نموذج التجانس الإطاري (Monotone Homogeneous Model): وهو النموذج المعروف بنموذج موكن (Mokken)، والذي يستخدم لتحليل التدرج في الاستجابات الثنائية، إذ وصفه سيجتسا (Sijtsma, 1998) بأنه نسخة معدلة لتحليل التدرج وفقاً لأسلوب جوتمان، حيث يختلف بشكل أساسي عن النموذجين الثنائي والثلاثي المعلمة في أن دالة استجابة الفقرة ليس من الضرورة أن تأخذ شكلاً لوجستياً بل من الممكن أن تكون على شكل معادلة خطية أو أسية أو غيرها، وعندما تتوافق دالة الاستجابة للفقرة مع نموذج التجانس الإطاري فإنها تعد مثالية إذا كان الهدف ترتيب الأفراد وفقاً لقدرتهم على متصل القدرة θ باستخدام درجتهم الكلية على الاختبار (Sijtsma & Molenaar, 2002).

– نموذج الاطراد المضاعف (Double Monotonicity Model): وهو النموذج الثاني لموكن، والذي يفترض جميع افتراضات نموذج التجانس الإطاري، بالإضافة إلى افتراض عدم تقاطع دوال الاستجابة لفقرات الاختبار مع السماح لها بالتماس في المناطق المتطرفة، مما يجعل منه نموذجاً صعب التحقيق (Sijtsma, 1998)، وبالتالي يعتبر هذا النموذج حالة خاصة من نموذج التجانس الإطاري حيث أن مجموعة البيانات التي يمكن تفسيرها عن طريق نموذج التجانس الإطاري يمكن تفسيرها عن طريق نموذج الإطارية المضاعفة والعكس غير صحيح (Sijtsma & Molenaar, 2002)، وتبعاً لهذا النموذج يتم ترتيب الأفراد تبعاً للقدرة θ ، مع إمكانية ترتيب الفقرات على المتصل كما في نظرية الاستجابة المعلمية، إلا أنه لا يسمح بتقدير رقمي لمعلمة الصعوبة (Sijtsma & Verweij, 1992).

تقديرات المعالم في نظرية الاستجابة للفقرة

إن الهدف الأول والأساسي في ضوء نظرية الاستجابة للفقرة هو تحديد موقع المفحوص على مقياس القدرة، ومن ثم معرفة احتمال الإجابة الصحيحة لكل مفحوص، وتقييمه في حدود القدرة التي يمتلكها، أما الهدف الثاني فهو إجراء المقارنات بين المفحوصين بهدف إعطائهم العلامات واتخاذ القرارات المختلفة بحقهم (Fitzpatrick & Wendy, 2001)، وعليه هناك عدة طرق تستخدم لتقدير قدرة المفحوصين والمعالم للفقرات؛ وهذه الطرق هي:

– تقدير الأرجحية العظمى (Maximum Likelihood Estimation Methods): يتم في هذه الطريقة تقدير المعالم من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها، عند توفر معلومات كافية عن عينة المفحوصين، حيث تعتمد هذه الطريقة على نمط استجابة المفحوصين على مجموعة من الفقرات من النوع (١)،

٠)، إلا أن أحد محددات هذه الطريقة تظهر عند إجابة المفحوص إجابة صحيحة أو خاطئة على كافة فقرات الاختبار، وبالتالي تصبح قدرته إما $(-\infty, \infty)$ ، ومن محدداتها أيضًا عدم توفر القدرة في تقدير معالم الفقرات التي يجيب عليها جميع المفحوصين إجابة خاطئة أو صحيحة، لذا يتم حذف المفحوصين الذين يجيبون إجابات صحيحة أو خاطئة على جميع الفقرات، أو حذف الفقرات التي أجاب عليها كافة المفحوصين والتي لم يجيب عليها أي مفحوص، للتغلب على هذه المحددات.

– طريقة تقدير بيز (Bayesian Estimation Method): يتم اللجوء إلى هذه الطريقة لتلافي المحددات السابقة الذكر، والتي تمنع استخدام طرق الأرجحية العظمى، وتمتاز باستخدامها لمعلومات أولية متوفرة من خبرات سابقة بالإضافة إلى استخدام معلومات عينة المفحوصين، وبهذا تساعد طريقة بيز في الحصول على تقديرات جيدة لمعالم الفقرات وقدرات المفحوصين، في حال توافرت معلومات عن توزيع القدرة لدى مجموعة من الأفراد المفحوصين (Garre & Vermunt, 2006).

التحيز والأداء التفاضلي للفقرة (bias & Differential Item Functioning)

يُنظر لمفهوم التحيز على أنه مفهوم يخفي وراءه معني سلبياً يرتبط بفكرة عدم العدالة، والاهتمام بالمفاهيم المشحونة بالجانب السياسي والاجتماعي أكثر من ربطه بالمفهوم السيكمومتري، فالتحيز يحمل معني اجتماعياً يشير إلى عدم العدالة والمساواة، في حين يشير الأداء التفاضلي إلى الخصائص السيكمومترية للفقرة من حيث الكيفية التي تعمل بها الفقرة في المجموعة بشكل مختلف عن مجموعة أخرى (Dorans & Holland, 1993)، فقد أشار وليمز (Williams, 1997) إلى أن الفقرة تعد متحيزة إذا قاست غير ما هو متوقع من المحتوى، وأن الفقرة تبدي أداءً تفاضلياً إذا كانت احتمالات استجابة الفرد الصحيحة عنها مختلفة للمفحوصين المتساوين في السمة التي يتم قياسها في المجموعات المختلفة. فالأداء التفاضلي بحيث تتم عملية المقارنة بين منحنيات خصائص الفقرة (ICCS) في المجموعتين الفرعيتين لمجتمع إحصائي معين، لكشف الأداء التفاضلي للفقرة بينهما، والتي يتم تمثيلها لاحقاً وفق النموذج اللوجستي المستخدم (Gierl, et al. 2000).

طرق الكشف عن الأداء التفاضلي

اقترح علماء القياس العديد من الطرق التي يمكن من خلالها الكشف عن الأداء التفاضلي للفقرات في كافة أشكال الاختبارات، إذ أن الهدف من هذه الطرق والأساليب الإحصائية تنقية الاختبارات من التحيز، وتحقيق مبدأ العدالة والمساواة بين مجموعات المفحوصين المختلفة (Camili & Shepard, 1994) وهناك ثلاثة طرق للكشف وفقاً لنظرية الاستجابة للفقرة وهي:

– طريقة مقارنة منحنيات خصائص الفقرة: وتقوم على افتراض أن الفقرات التي تتشابه في معالمها من حيث الصعوبة والتخمين والتمييز في مجموعات المفحوصين النوعية المختلفة، عند مستوى قدرة واحد باختلاف النماذج اللوجستية لا تبدي أداءً تفاضلياً، في حين أن الفقرات التي تختلف معالمها عند مستويات واحدة من القدرة للفقرة نفسها تبدي أداءً تفاضلياً، حيث يمكن حساب فرق المساحة بين منحنيات خصائص الفقرة باختلاف النموذج اللوجستي المستخدم (Chung & Huisu, 2004).

– طريقة مقارنة المعالم المقدرة للفقرة: وتستخدم فيها المجموعات المرجعية والمستهدفة من خلال معلم قدرة الأفراد (θ) ، ومعلم $(B_{uncorrected})$ ، الذي يتم تقديره من خلال التحيز (BIAS) (Rebecca, et al., 1997).

– طريقة منحني نسبة الأرجحية (G^2) : وتستخدم في الفقرات ثنائية التدرج ومتعددة التدرج بأسلوب (Likelihood – Ratio Test)، في فحص الفرضية الصفريّة التي تنص على عدم وجود فروق ذات دلالة إحصائية في دالة استجابة الفقرة بين المجموعات، أو عدم وجود فروق ذات دلالة إحصائية في معالم الفقرات بين المجموعات الاختبارية، فوجود الدلالة الإحصائية يدل على وجود أداء تفاضلي للفقرة (البشاشة، ٢٠١٦).

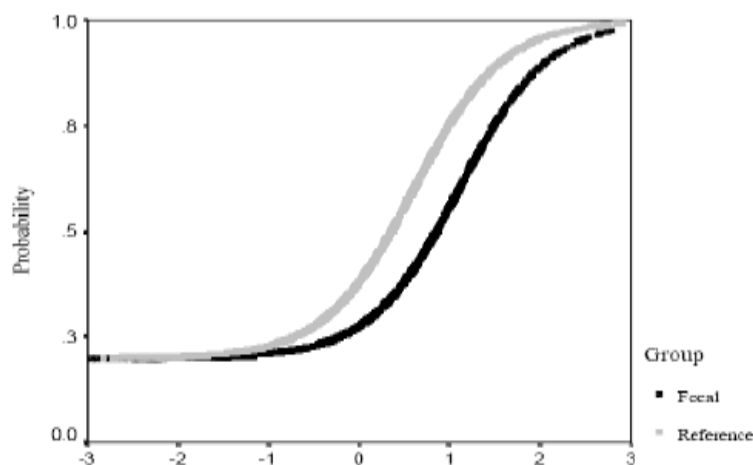
خطوات الكشف عن الأداء التفاضلي وأنواعه

إن أية دراسة تهدف إلى الكشف عن الأداء التفاضلي للفقرات تستوجب إجراء مجموعة من الخطوات الرئيسية، ثمثّل في استخدام الإجراءات الإحصائية المحددة للكشف عن الأداء التفاضلي ومعرفة العناصر المسؤولة عن الأداء التفاضلي

واتخاذ القرارات حول مصادر الأداء التفاضلي فيما إذا كانت ذات علاقة بالسمة المقيسة وبنية الاختبار ككل (Uiterwijk & Vallen, 2005). من الجدير بالذكر أن هناك قسمين رئيسيين للأداء التفاضلي وهما:

أولاً: الأداء التفاضلي المنتظم (Uniform Differential Item Functioning)

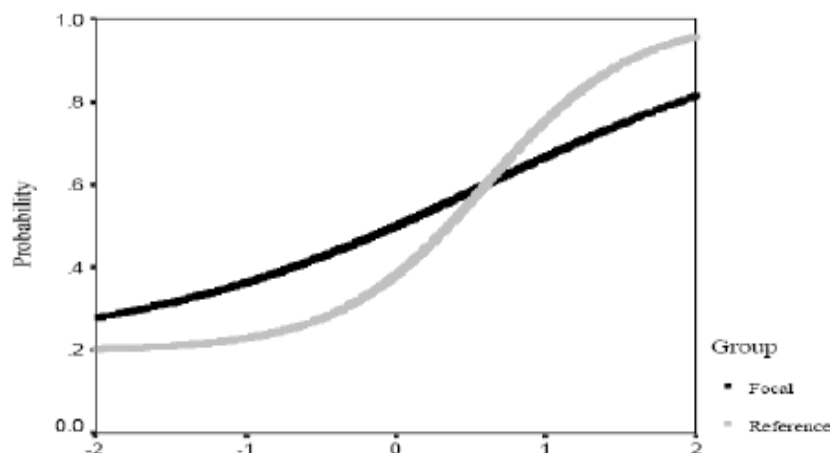
وهو ذلك النوع الذي ينتج عند عدم حدوث تفاعل بين مستوى القدرة (θ) والمجموعة التي ينتمي إليها المفحوص، أي أن احتمال الإجابة الصحيحة من قبل المفحوصين يكون أكبر دائماً لإحدى المجموعتين الاختباريتين عند كافة مستويات القدرة، إذ تكون كل من معلمة التخمين والتمييز متساوية بين المجموعات الاختبارية، ولكن معلم الصعوبة لمنحنى خصائص الفقرة غير متساوية بين المجموعات، والشكل (١) يوضح ذلك (Chung & Huisu, 2004).



شكل (١): منحنى خصائص فقرة ذات أداء تفاضلي بين المجموعتين

ثانياً: الأداء التفاضلي غير منتظم (Non-Uniform Differential Item Functioning)

وهو ذلك النوع من الأداء التفاضلي الذي يظهر عند وجود تفاعل بين مستوى القدرة (θ) والمجموعة التي ينتمي إليها المفحوص؛ أي أن نمط الفروق في احتمالية الاستجابة الصحيحة ليست متشابهة في كافة مستويات القدرة، فقد تكون لصالح إحدى المجموعات الاختبارية عند مستوى قدرة معين (θ)، ولصالح مجموعة أخرى عند مستوى قدرة آخر مختلف، وبذلك فإن معالم التمييز مختلفة عبر المجموعات، في حين أن كل من معلم التخمين والصعوبة متساوية عبر المجموعات، والشكل (٢) يوضح ذلك (Uiterwijk & Vallen, 2005).



شكل (٢): منحنى خصائص فقرة ذات أداء تفاضلي غير منتظم بين المجموعتين

قام العديد من الباحثين والمختصين بالقياس النفسي والتربوي بأجراء العديد من الدراسات والابحاث المتعلقة بالأداء التفاضلي للفقرات والاختبارات، فمنهم من أجرى دراسات حول الكشف من مدى وجود الاداء التفاضلي، ومنهم من قام بمقارنة طرق الكشف عن الاداء التفاضلي سواء من خلال النظرية الكلاسيكية أو نظرية الاستجابة للفقرة، أو من خلال كلاهما.

فقد قام هيو جبروك (Haughbrook, 2020) بدراسة هدفت الى الكشف عن الاداء التفاضلي للفقرات من خلال عدة اختبارات هي مقياس (Woodcock-Johnson III Picture Vocabulary scale)، للتحيزات العرقية باستخدام تحليل (DIF)، وفحص تقرير المعلم لإنجاز الطالب، ومقياس فرعي للكفاءة الأكاديمية (SSRS-T)، باستخدام تحليل (DIF) ايضاً، وقد تم اختيار المشاركين في هذه الدراسة من دراسة أكبر (مشروع KIDS)، والتي اجريت في شمال فلوريدا بين عامي (٢٠٠٥-٢٠٠٦) و (٢٠١٢-٢٠١٣)، وقد تكونت عينة الدراسة من (٣٦٧٨) مفحوص، وأشارت النتائج إلى أن (١١) من (١٧) فقرة من (WJPV) أظهرت (DIF) وستة من تسعة فقرات (SSRS-T) أظهرت (DIF) فيما يتعلق بتأثيرات الاختبار، وتشير النتائج إلى أن كلا من طرق قياس تحصيل الطلاب، والتقييمات الموحدة وتقارير المعلم، تظهر تحيزاً من حيث الاداء التفاضلي وأداء الاختبار حسب العرق، أيضاً في حالة المتغيرات المدرجة في هذه الدراسة، احتوى الاختبار القياسي (WJPV) على المزيد من العناصر باستخدام (DIF) التي كانت كبيرة وكبيرة الحجم.

كما قام زكري (٢٠٢٠) بدراسة هدفت الكشف عن الأداء التفاضلي لفقرات اختبار "إمبو" لأساليب المعاملة الوالدية لدى طلبة المرحلة الثانوية وفقاً لمتغير الجنس، باستخدام طريقة مانتل هانزل، وتكونت عينة الدراسة من (٢٧٤) طالباً وطالبة من الصف الثاني ثانوي بإدارة تعليم محافظة صبيا للعام الدراسي (٢٠١٨/٢٠١٩)، منهم (١٣٤) طالباً و (١٤٠) طالبة تم اختيارهم بالطريقة العشوائية العنقودية، وتم استخدام اختبار "إمبو" لأساليب المعاملة الوالدية، وأظهرت النتائج وجود (١٤) فقرة أظهرت أداء تفاضلياً من فقرات الاختبار (صورة الأب) وفقاً لمتغير الجنس، منها (٨) فقرات لصالح الطلاب، و (٦) فقرات لصالح الطالبات، كما تبين عدم وجود أثر دال إحصائي لمؤشرات صدق البناء الداخلي (RMSEA, NCP, AIC, SRMR, CFI) للاختبار نتيجة لحذف الفقرات ذات الاداء التفاضلي من الأداة. كما أظهرت النتائج وجود (١٧) فقرة أظهرت أداء تفاضلياً من فقرات الاختبار (صورة الأم) وفقاً لمتغير الجنس، منها (٩) فقرات لصالح الطالبات، و (٨) فقرات لصالح الطلاب، كما تبين عدم وجود أثر دال إحصائي لمؤشرات صدق البناء الداخلي (RMSEA, NCP, AIC, SRMR, CFI) للاختبار (صورة الأم) نتيجة لحذف الفقرات ذات الاداء التفاضلي من الأداة.

وأجرى الشريفين (٢٠١٨) دراسة هدفت الكشف عن أثر نوع الأداء التفاضلي للفقرات (منتظم، غير منتظم) على الخصائص السيكمترية للفقرات، والاختبار وفق النموذج اللوجستي ثلاثي المعلمة، والنموذج اللوجستي الثلاثي اللامعلمي لنظرية الاستجابة للفقرة. ولتحقيق هدف الدراسة تم استخدام بيانات مولدة لاختبار من نوع الاختيار من متعدد، تكون كل منهما من (٥٠) فقرة، ل (١٠٠٠) مفحوص، حيث أظهرت النتائج وجود فرق دال إحصائي بين الوسطين الحسابيين لكل من معالم التميز ومعالم التخمين يعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج المعلمي مقارنة بالنموذج اللامعلمي، وأظهرت النتائج ايضاً وجود فرق دال إحصائي بين الوسطين الحسابيين لمعالم الصعوبة يعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، وأشارت النتائج الى وجود فروق دالة إحصائية في الوسط الحسابي لكل من الصعوبة والتخمين تعزى للتفاعل بين النموذج المستخدم ونوع الاداء التفاضلي، كما أشارت النتائج إلى وجود فرق دال إحصائي بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى للنموذج، لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، وبينت النتائج أن قيم معاملات الثبات المقدرة وفق النموذجين كانت عالية، وأن قيم معاملات الثبات المقدرة وفق النموذج المعلمي كانت الأعلى، سواء كان الاداء التفاضلي للفقرات منتظماً أو غير منتظم.

كما قام القرعان والكويتي (AlQuraan & ALKuwaiti, 2017) بدراسة هدفت إلى تقصي أثر حقل المعرفة (التخصص) للطلاب كمصدر للأداء التفاضلي في فقرات تقييم فعالية وجودة التعليم العالي من وجهة نظر الطالب، ولتحقيق هدف الدراسة، فقد تم الاستفادة من البيانات التي قامت جامعة الإمام عبد الرحمن بن فيصل بجمعها لأغراض الحصول على الاعتماد الأكاديمي العام، فقد تم تحليل (٣٩٤٥٩) استبانة في ثلاث كليات (التربية، الصحة، الهندسة) باستخدام نموذج راش المعدل، وأظهرت النتائج وجود أربع فقرات أبدت أداءً تفاضلياً حسب الكلية، وأكد محتوى الفقرات احتمالية أن تكون تلك الفقرات متحيزة لكلية دون أخرى. كما بينت الدراسة أن حذف الفقرات ذات الاداء التفاضلي يسهم وبدلالة إحصائية في تحسين صدق البناء للأداة.

وقام العمري وشطناوي (٢٠١٦) بدراسة هدفت إلى الكشف عن الأداء التفاضلي لفقرات الاختبار الوطني الأردني لمادة الرياضيات للصف العاشر تبعاً لمتغير الجنس، وبلغت عينة الدراسة (٢٤٠٠) طالباً وطالبة بواقع (١٢٠٠) طالب و (١٢٠٠) طالبة ممن تقدموا للاختبار للعام الدراسي ٢٠١٣/٢٠١٤ ولتحقيق هدف الدراسة تم الحصول على البيانات المتعلقة باستجابات الطلبة على الاختبار المكوّن من (٣٩) فقرة، وتم استخدام (مؤشر الأداء التفاضلي اللاتعويضي (NCDIF) في الكشف عن الأداء التفاضلي للفقرات حيث بلغ عدد الفقرات في صورتها النهائية بعد مطابقة الفقرات للنماذج المستخدمة (٢٥) فقرة مطابقة للنموذج الأحادي المعلمة و (٣٤) فقرة مطابقة للنموذج الثنائي المعلمة وكشفت النتائج عن وجود أداء تفاضلي منتظم لـ (٩) فقرات من أصل (١٦) فقرة هي عدد الفقرات المشتركة بين الطلاب والطالبات وعن وجود أداء تفاضلي غير منتظم لـ (١٥) فقرة من أصل (٢٦) فقرة هي عدد الفقرات المشتركة بين الطلاب والطالبات وخلصت هذه الدراسة إلى جملة من التوصيات تدعو إلى إجراء دراسات مختلفة تتناول الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم وخصوصاً في مباحث أخرى غير الرياضيات، والكشف عن الأداء التفاضلي للاختبارات الوطنية الأخرى التي تم تطبيقها إلكترونياً على الطلبة.

وفي دراسة استخدم فيها أسلوب المحاكاة في توليد البيانات تحت ظروف اختبارية متباينة، قام (النوافلة، ٢٠١٣) بمعرفة أثر نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي في تقدير معالم الفقرات وقدرات الأفراد وفق النموذج اللوجستي ثلاثي المعلمة، اشارت النتائج فيها الى وجود دلالة إحصائية بين متوسطات تقديرات معالم الصعوبة تعزى إلى نسب الفقرات ذات الأداء التفاضلي لصالح المجموعة المستهدفة، وفروقات مختلفة بين متوسطات معالم التمييز والتخمين لصالح المجموعة المرجعية، كما تبين عدم وجود فروق ذات دلالة إحصائية في تقديرات القدرة تعزى إلى متغيري الدراسة، أما كمية المعلومات فإنها تنخفض كلما ازدادت نسبة الفقرات ذات الأداء التفاضلي.

كما أجرى القيسي (٢٠١٣) دراسة هدفت إلى مقارنة دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة المعلمة واللامعلمية باختلاف حجم العينة وطول الاختبار اعتماداً على مؤشري التحيز، والجذر التربيعي لمتوسط مربعات الخط، وذلك باستخدام بيانات مولدة (محاكاة)، وتوصلت الدراسة إلى أن تقديرات معالم الفقرة والقدرة بالطريقة المعلمة أقل منها في الطريقة اللامعلمية، كما أن ما يتعلق بمؤشر الدقة في القياس فقد دلت النتائج إلى أن قيم المتوسطات لطريقة التقدير المعلمة كانت أقل من المتوسطات لطريقة التقدير اللامعلمية في كل حالات الدراسة.

وبهدف مقارنة فعالية ثلاث طرق في الكشف عن الأداء التفاضلي من خلال استخدام نماذج معلمية ونماذج لامعلمية أجرى جبريال (Gabriel, 2012) دراسة ، استخدم فيها طريقة اختبار تحيز الفقرة المتزامن والمتقاطع مع الأداء، وطريقة اختبار الأرجحية العظمى في نظرية الاستجابة للفقرة، وطريقة الانحدار اللوجستي في ظروف تجريبية مختلفة من حيث طول الاختبار وحجم العينة ومقدار الأداء التفاضلي للفقرة والاختبار وتوزيع قدرات مختلفة للمجموعات المرجعية والمستهدفة، وتم توليد البيانات وفق النموذج اللوجستي ثلاثي المعلمة، وحساب الخطأ من النوع الأول ومن النوع الثاني، وقوة الاختبار الإحصائي، وبيّنت النتائج أن طريقة الانحدار اللوجستي كانت أفضل الطرق في قوة الكشف عن الأداء التفاضلي، ومعدل الخطأ من النوع الأول، كما اظهرت النتائج عدم وجود طريقة واحدة فعالة تميز بين نوعي الأداء التفاضلي.

١,١,١ التعقيب على الدراسات السابقة

في ضوء ما تم عرضه دراسات سابقة يمكن القول ان معظمها ركزت على الأداء التفاضلي والكشف عن الفقرات ذات الأداء التفاضلي باستخدام طرق مختلفة ومتنوعة واستناداً لكلا النظريتين، إلا أن ما توصلت له هذه الدراسات عدم وجود طريقة محددة فعالة رغم تعدد تلك الطرق؛ ويعزى ذلك إلى أن هذه الطرق كانت تعتمد على نوع الأداء التفاضلي، ومصدر الأداء التفاضلي، ومستوى تمييز الفقرة وصعوبتها، وطول الاختبار، وحجم العينة. كما يتضح ان هناك ندرة في الدراسات التي تناولت أثر نسبة الفقرات التي تظهر اداء تفاضلياً على قدرات الافراد وفق النموذجين المعلمي واللامعلمي، وعليه تعد هذه الدراسة استكمالاً للجهود البحثية السابقة في هذا المجال، لإضافة المزيد من المعلومات كمعرفة تراكمية لما تم البحث فيه ودراسته في هذا الموضوع.

٢. مشكلة الدراسة وأسئلتها

يعد الأداء التفاضلي قضية رئيسية من قضايا القياس النفسي والتربوي، فهو يهدد صدق بناء وثبات الاختبار، كما أنه يهدد دقة تقدير المعالم للفقرات والقدرة للأفراد، مما يؤدي الى الوصول الى نتائج غير دقيقة وغير صادقة، وبالتالي وجود

مخاوف من اتخاذ قرارات تربوية غير صائبة، وقد نصت المعايير الخاصة بالاختبارات التربوية والنفسية على أنه يجب فحص جميع التقييمات بحثاً عن الفقرات التي قد تظهر أداءً تفضلياً من أجل ضمان العدالة والإنصاف في الاختبارات (AERA, 1999 APA & NCME)، كما أن المصدر الأساسي للأداء التفاضلي هو الفقرة، والذي يحدث عندما تختلف الفقرات الموجودة في وظيفة التقييم لمجموعات مختلفة من الأفراد (Croudace & Brown, 2012; Van de Vijver & Tanzer, 2004)، والأهم من ذلك، أن وجود الأداء التفاضلي في الاختبار يمكن أن يؤدي إلى آثار ضارة على الأفراد، مثل عدم تكافؤ الفرص، والتشخيص الخاطئ، وعدم المساواة في الفرص التعليمية، والاستنتاج غير الدقيق في البحث (Croudace & Brown, 2012).

لقد اهتم الباحثون بهذه القضية وأثروها بالدراسة والبحث، حيث أجريت العديد من الدراسات والابحاث التي تسعى للمقارنة بين الطرق المستخدمة للكشف عن الأداء التفاضلي، وقد أظهرت بعضها عدم وجود فروق ذات دلالة إحصائية بين أداء طريقتي التحليل العاملي المقيد من جهة ومانتل هانزل وطريقة مؤشر المساحة بين منحنى خصائص المفردة في جميع اختبارات الدراسة من جهة أخرى (الناجي، ٢٠١١)، بينما تبين وجود اختلاف بتحديد الفقرات ذات الأداء التفاضلي تبعاً للطريقة المتبعة (البيسنجي، ٢٠٠٤)، كما تبين عدم وجود اتفاق بين طرائق الكشف عن الفقرات ذات الأداء التفاضلي (أبو حماد، ٢٠٠٨)، وأن هناك تفاوتاً في نتائج الحكم على الفقرة بأنها ذات أداء تفضلي تبعاً للطريقة المستخدمة في الكشف (Acar & Kelecioğlu, 2010)، كما أنه لا يوجد طريقة واحدة فعالة تميز بين نوعي الأداء التفاضلي (Gabriel, 2012). لذا جاءت هذه الدراسة استكمالاً لتلك الجهود من أجل معرفة أثر نسبة الفقرات ذات الأداء التفاضلي ونوعه أو التفاعل بينها وفق النموذج اللوجستي ثلاثي المعلمة وفق نموذج موكن اللامعلمي على تقدير قدرات الافراد وعليه تتلخص مشكلة الدراسة في الإجابة عن الأسئلة التالية.

٢,١ أسئلة الدراسة

١. هل توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha = 0.05$) بين متوسطات تقديرات قدرات الافراد تعزى لنسبة الفقرات ذات الأداء التفاضلي ولنوع الأداء التفاضلي او للتفاعل بينهم وفق النموذج اللوجستي ثلاثي المعلمة؟
٢. هل توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha = 0.05$) بين متوسطات تقديرات قدرات الافراد تعزى لنسبة الفقرات ذات الأداء التفاضلي ولنوع الأداء التفاضلي او للتفاعل بينهم وفق نموذج موكن اللامعلمي؟
٣. هل توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha = 0.05$) بين متوسطات تقديرات قدرات الافراد تعزى للنموذج المستخدم (ثلاثي المعلمة، موكن اللامعلمي) في ضوء نسبة ونوع الفقرات ذات الأداء التفاضلي؟

٣. أهمية الدراسة

تبرز أهمية الدراسة في استكمال جهود الباحثين في هذا الموضوع والذي يعد احد المواضيع التي نالت ولا تزال تنال اهتمام المختصين في القياس النفسي والتربوي منذ عقود مضت والذي لا يزال بحاجة الى المزيد من البحث والدراسة والتقصي، والذي نسعي من خلاله الى الكشف عن أثر نوع ونسبة الفقرات ذات الأداء التفاضلي على معلمة القدرة للأفراد في ضوء النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي، مع اجراء مقارنات وفق نوع الأداء التفاضلي في ضوء النماذج المعلمية واللامعلمية في نظرية الاستجابة للفقرة، والذي يعد بدوره محاولة من سلسلة المحاولات العلمية والاكاديمية لمواكبة الدراسات المتواصلة والمستندة الى نظرية الاستجابة للفقرة، ولمسايرة التطورات التي تحصل في مجال القياس النفسي والتربوي بشكل عام. وكذلك في استخدام البرمجيات الإحصائية المختلفة في توليد البيانات والاستجابات المختلفة عليها والتحقق من خصائصها الإحصائية وذلك مثل برمجية (Wingen) لتوليد بيانات الاختبارات، برمجية البايولوج (Bilog-Mg) لاستخراج تقديرات القدرة من خلال النموذج اللوجستي الثلاثي. وبرنامج (TestGraf) لتحليل واستخراج تقديرات القدرة من خلال النموذج اللامعلمي (نموذج موكن)، وبرمجية الرزم الإحصائية (SPSS) لاستخراج الاحصاءات الوصفية والفروق الإحصائية؛ والتي يتم من خلالها محاكاة البيانات الفعلية افتراضياً بهدف التعرف على المنهج المناسب لعملية تطوير الاختبارات والتحقق من خصائصها السيكمترية وفق النماذج المناسبة.

٤. التعريفات الإجرائية

الأداء التفاضلي

خطأ منتظم يجعل الاداء على الاختبار لمجموعة أفضل من اخرى، أو هو تقدير متدني أو مرتفع بانتظام لمعلمة المجتمع اعتماداً على إحصائي العينة.

نوع الأداء التفاضلي

نوع الأداء الذي تظهر فيه الفقرة دلالة مشتقة إحصائياً للتعبير عن الفروق في احتمالات الإجابة، والذي قد يكون منتظم أو غير منتظم.

المجموعة المستهدفة (Group Focal)

هي المجموعة التي تشتمل على فقرات ذات أداء تفاضلي (منتظم أو غير منتظم)، ويتم مقارنتها مع المجموعة المرجعية.

المجموعة المرجعية (Group Referenced)

هي المجموعة التي لا تشتمل على أي فقرة ذات أداء تفاضلي، وهي المجموعة التي تقارن المجموعة المستهدفة معها.

٥. محددات الدراسة

اقتصرت هذه الدراسة على استخدام ما يلي:

- نموذج نظرية الاستجابة للفقرة ثلاثي المعلم ونموذج موكن اللامعلمي.
- ثلاث قيم لنسب الفقرات ذات الأداء التفاضلي في الاختبار، وهي (١٠٪، ٢٠٪، ٤٠٪).
- فقرات ثنائية التدرج (٠، ١).
- بيانات مولدة (Simulated Data) من خلال برنامج (WinGen)، والذي تم تطويره كبرنامج حاسوبي لتوليد بيانات استجابة ثنائية ومتعددة التدرج للعديد من نماذج نظرية الاستجابة للفقرة (Han & Hambleton, 2007).

٦. الطريقة

٦,١ منهج الدراسة

اعتمدت الدراسة على المنهج التجريبي حيث تم ضبط الظروف التجريبية وذلك من خلال محاكاتها لبيانات افتراضية تم توليدها باستخدام البرمجيات التي تم الإشارة إليها.

٦,٢ عينة الدراسة

تكونت عينة الدراسة من (٢٠٠) فرد افتراضي لكل ظرف تجريبي ويحاكي توزيع قدراتهم الظروف التجريبية المبينة في الجدول (١).

جدول (١): توزيع قدرة الأفراد الافتراضيين حسب ظروف الدراسة التجريبية

غير منتظم		منتظم		النسبة
المستهدفة	المرجعية	المستهدفة	المرجعية	
$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	١٠٪
$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	٢٠٪
$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	$1000 \sim N(0,1)$	٣٠٪

٦,٣ أداة الدراسة

تكونت أداة الدراسة من اختبارات افتراضية طول كل منها (٦٠) فقرة ثنائية الاستجابة تم توليدها باستخدام برمجية (WinGen)، إذ تم ولكل ظرف اختباري توليد مجموعتين من المستجيبين بواقع (١٠٠٠) مفحوص في كل مجموعة تخضع قدراتهم للتوزيع الطبيعي بمتوسط حسابي (٠) وانحراف معياري (١) وذلك انسجاماً مع الدراسات السابقة التي تفترض التوزيع الطبيعي المعياري للمستجيبين، ولتحقيق شروط الأداء التفاضلي فقد تم توليد مجموعتين متكافئتين من حيث توزيع القدرة. وذلك باستخدام برمجية (WinGen) المتخصصة في توليد القدرة والفقرات والاستجابات.

٦,٤ إجراءات الدراسة

للإجابة عن أسئلة الدراسة تم اتباع الإجراءات التالية:

- توليد المستجيبين الافتراضيين، والمشار لخصائصهم في عينة الدراسة.
- توليد الاختبارات الافتراضية والمشار لها في أداة الدراسة.
- توليد استجابات افتراضية، حيث تم استخدام برمجية (WinGen) لتوليد استجابات افتراضية تحاكي قدرات عينة الدراسة الافتراضية حيث نتج (٦) مصفوفات من الاستجابات.

٦,٥ التحليل الإحصائي

تحليل الاستجابات المولدة في الخطوة السابقة من خلال النموذجين (الثلاثي المعلمي، وموكن اللامعلمي) على النحو التالي:

- تحليل الاستجابات في النموذج ثلاثي المعلمة: تم التحقق من مدى مطابقة البيانات لافتراضات النموذج ثلاثي المعلمة والمتمثلة في افتراضات نظرية الاستجابة للفقرة حيث تم فحص مدى تحقق أحادية البعد من خلال التحليل العاملي باستخدام برمجية (SPSS) ومن خلال نسبة التباين المفسر التي يفسرها الجذر الكامن الأول التي يفترض أن تتجاوز قيمتها الـ (٢٠٪) وكذلك من خلال نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني كما هو مبين في الجدول (٢).

جدول (٢): عدد العوامل وقيم الجذر الكامن ونسب التباين المفسر ونسبة الجذر الكامن الأول إلى الجذر الكامن الثاني

العامل	الجذر الكامن	نسبة التباين المفسر	نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني
١	٧,٠٢٣	١١,٧٠٥	٥,٣١٩
٢	١,٣٢٠	٢,٢٠١	
٣	١,٢٠٨	٢,٠١٤	
٤	١,١٧٨	١,٩٦٣	
١	٦,٩٦٦	١١,٦١٠	٥,٣٧٢
٢	١,٢٩٧	٢,١٦١	
٣	١,٢٠٠	٢,٠٠٠	
٤	١,١٨٩	١,٩٨٢	
١	٧,٠٦٩	١١,٧٨١	٥,٥٣٦
٢	١,٢٧٧	٢,١٢٨	
٣	١,١٩٤	١,٩٩٠	
٤	١,١٦٨	١,٩٤٧	
١	٧,٠٧٠	١١,٧٨٤	٥,٥٣٦
٢	١,٢٧٧	٢,١٢٩	
٣	١,٢٠٩	٢,٠١٥	
٤	١,١٦٩	١,٩٤٨	

١	٦,٨٣٩	١١,٣٩٩	٥,١٧٢
٢	١,٣٢٢	٢,٢٠٤	
٣	١,٢٠٨	٢,٠١٤	
٤	١,١٨٦	١,٩٧٦	
١	٦,٨٠٨	١١,٣٤٧	٥,١٢١
٢	١,٣٢٩	٢,٢١٦	
٣	١,١٩٩	١,٩٩٨	
٤	١,١٨٩	١,٩٨١	

غير منتظم ٢٠٪

غير منتظم ٣٠٪

يتبين من خلال الجدول (٢) أن أحادية البعد متحققة حيث إنه وبالاعتماد على مؤشر (Tanaka) يتضح ان نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني أكبر من (٢) ولكل الظروف التجريبية. كما تم فحص مدى تحقق الاستقلال الموضوعي (Local-Dependent) باستخدام برمجية (LDID) المتخصصة في التحقق من الاستقلال الموضوعي من خلال مؤشر (Q3) والمبينة في الجدول (٣). أما بالنسبة للتحرر من السرعة فلم يتم التحقق منه بسبب أن الاختبار والأفراد والاستجابات افتراضية.

جدول (٣): عدد الأزواج الكلي وعدد الأزواج غير المحققة للاستقلال الموضوعي ونسبتها

N18	N12	N6	U18	U12	U6	
١٧٧٠	١٧٧٠	١٧٧٠	١٧٧٠	١٧٧٠	١٧٧٠	عدد الأزواج الكلي
٥٩	٦٠	٧٢	٧٤	٥٤	٦٥	عدد الأزواج غير محققة
٣,٣٪	٣,٤٪	٤,١٪	٤,٢٪	٣,١٪	٣,٧٪	نسبة عدم التحقيق

يتبين من خلال الجدول (٣) أن نسبة عدم التحقيق قد بلغت في حدها الأعلى (٤,٢٪) والتي تعد مؤشر لنسبة الأزواج المرتبطة من الفقرات وهي نسبة منخفضة، مما يشير إلى تحقق الاستقلال الموضوعي.

وللتحقق من مدى مطابقة الفقرات والأفراد للنموذج ثلاثي المعلمة تم استخدام برمجية بايلوج (BILOG-MG) المتخصصة في نظرية الاستجابة للفقرة (Item Response Theory) حيث تم الكشف عن الأفراد المطابقين في كل مصفوفة استجابات من خلال مؤشر مربع كاي لمطابقة الفقرات والتي تراوحت ما بين (١٤,٧-٢,١) كما كان كافة الأفراد مطابقين حسب مؤشر الاحتمال الهامشي.

– تحليل الاستجابات في نموذج موكن اللامعلمي: تم استخدام برمجية (Test-Graf) في تحليل كل من مصفوفات الاستجابات بهدف تقدير قدرات الأفراد.

٧. نتائج الدراسة ومناقشتها

للإجابة عن السؤال الأول "هل توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha = 0.05$) بين متوسطات تقديرات قدرات الأفراد تعزى لنسبة الفقرات ذات الأداء التفاضلي ولنوع الأداء التفاضلي او للتفاعل بينهم وفق النموذج اللوجستي ثلاثي المعلمة؟ تم تقدير القدرة باختلاف الظروف التجريبية (نسبة الأداء التفاضلي ونوعه) حسب النموذج ثلاثي المعلمة باستخدام برمجية بايلوج (BILOG-MG) واستخراج المتوسط الحسابي لمعلمة القدرة لكل من ظروف الاختبار كما يبينها الجدول (٤).

جدول (٤): المتوسط الحسابي والانحراف المعياري للقدرة في ضوء نوع ونسبة الأداء / النموذج الثلاثي

نوع الأداء	نسبة الأداء	المتوسط الحسابي	الانحراف المعياري
منتظم	١٠٪	٠,٠٠٠١	٠,٩٣٥٨
	٢٠٪	-٠,٠٠١٤	٠,٩٤٢٠
	٣٠٪	٠,٠٠٠٨	٠,٩٣٦٣
غير منتظم	١٠٪	٠,٠٠٠٥	٠,٩٤١١
	٢٠٪	-٠,٠٠٢٠	٠,٩٤١٤

٠,٩٣٩٥	٠,٠٠٠١	٣٠٪
--------	--------	-----

* حجم العينة (٢٠٠٠) مفحوص وطول الاختبار (٦٠) فقرة

يلاحظ من المتوسطات الحسابية في الجدول (٤) وجود فروق ملاحظة بين متوسطات القدرة باختلاف الظروف العائد لنوع ونسبة الأداء التفاضلي وللكشف عن دلالة تلك الفروق في المتوسطات الحسابية لمعلمة القدرة تم استخدام تحليل التباين الثنائي كما هو مبين في الجدول (٥).

جدول (٥): تحليل التباين الثنائي لفحص دلالة الفروق بين متوسطات القدرة العائدة لنوع ونسبة الأداء التفاضلي / النموذج الثلاثي

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الاحتمالية	مربع ايتا
نوع الأداء	٠,٠٠٠	١	٠,٠٠٠	٠,٠٠٠	٠,٩٨٧	٠,٠٠٠
نسبة الأداء	٠,٠١١	٢	٠,٠٠٦	٠,٠٠٦	٠,٩٩٤	٠,٠٠٠
التفاعل	٠,٠٠١	٢	٠,٠٠٠	٠,٠٠٠	١,٠٠٠	٠,٠٠٠
الخطأ	١٠٥٨٣,١٧٨	١١٩٩٤	٠,٨٨٢			
الكل	١٠٥٨٣,١٩٠	١١٩٩٩				

يلاحظ من نتائج تحليل التباين الثنائي أن الفروق في معلمة القدرة ($p > ٠,٠٥$) العائدة للتفاعل بين نسبة ونوع الأداء التفاضلي غير دالة إحصائياً عند مستوى الدلالة ($\alpha = ٠,٠٥$) مما يشير لان قيم القدرة لا تختلف باختلاف الظروف التجريبية المتمثلة في نوع ونسبة الأداء التفاضلي.

للإجابة عن السؤال الثاني والذي نص على "هل توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha = ٠,٠٥$) بين متوسطات تقديرات قدرات الافراد تعزى لنسبة الفقرات ذات الأداء التفاضلي ولنوع الأداء التفاضلي او للتفاعل بينهم وفق نموذج موكن اللامعلمي؟ تم تقدير القدرة باختلاف الظروف التجريبية (نسبة الأداء التفاضلي ونوعه) حسب نموذج موكن باستخدام برمجية (Test-Graf) واستخراج المتوسط الحسابي لمعلمة القدرة لكل من ظروف الاختبار كما يبينها الجدول (٦).

جدول (٦): المتوسط الحسابي والانحراف المعياري لمعلمة القدرة في ضوء نوع ونسبة الأداء / نموذج موكن

نوع الأداء	نسبة الأداء	المتوسط الحسابي	الانحراف المعياري
منتظم	١٠٪	-٠,٠٠٢٦	١,٠٨٦٠
	٢٠٪	٠,٠٠٦٨	١,٠٨٣٦
	٣٠٪	٠,٠٠٠٨	١,٠٨٨٥
غير منتظم	١٠٪	-٠,٠٠١٠	١,٠٨٥٨
	٢٠٪	٠,٠٠٠٦	١,٠٨٤٣
	٣٠٪	٠,٠٠١٥	١,٠٨٦٠

* حجم العينة (٢٠٠٠) مفحوص وطول الاختبار (٦٠) فقرة

يلاحظ من المتوسطات الحسابية في الجدول (٦) وجود فروق ملاحظة بين متوسطات القدرة باختلاف الظروف العائد لنوع ونسبة الأداء التفاضلي وللكشف عن دلالة تلك الفروق في المتوسطات الحسابية لمعلمة القدرة تم استخدام تحليل التباين الثنائي كما هو مبين في الجدول (٧).

جدول (٧): تحليل التباين الثنائي لفحص دلالة الفروق بين متوسطات معلمة القدرة العائدة لنوع ونسبة الأداء التفاضلي / نموذج موكن

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	الاحتمالية	مربع ايتا
نوع الأداء	٠,٠٠٥	١	٠,٠٠٥	٠,٠٠٤	٠,٩٤٩	٠,٠٠٠
نسبة الأداء	٠,٠٦٠	٢	٠,٠٣٠	٠,٠٢٥	٠,٩٧٥	٠,٠٠٠

التفاعل	٠,٣٦	٢	٠,١٨	٠,١٥	٠,٩٨٥	٠,٠٠٠
الخطأ	١٤١٣٧,٦٦٢	١١٩٩٤	١,١٧٩			
الكل	١٤١٣٧,٧٦٤	١١٩٩٩				

يلاحظ من نتائج تحليل التباين الثنائي أن الفروق في معلمة القدرة ($p > 0,05$) العائدة للتفاعل بين نسبة ونوع الأداء التفاضلي غير دالة إحصائياً عند مستوى الدلالة ($\alpha = 0,05$) مما يشير لان قيم القدرة لا تختلف باختلاف الظروف التجريبية المتمثلة في نوع ونسبة الأداء التفاضلي.

للإجابة عن السؤال الثالث والذي نص على "هل توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين متوسطات تقديرات قدرات الافراد تعزى للنموذج المستخدم (ثلاثي المعلمة، موكن اللامعلمي) في ضوء نسبة ونوع الفقرات ذات الأداء التفاضلي؟ تم استخدام اختبار (t) لفحص الفروق في تقدير معلمة القدرة باستخدام النموذجين في ضوء الظروف التجريبية المتمثلة في (نوع ونسبة الأداء التفاضلي) كما هو مبين في الجدول (٨).

جدول (٨): اختبار (t) لفحص دلالة الفروق بين متوسطات معلمة القدرة العائدة للنموذج في ضوء نوع ونسبة الأداء التفاضلي*

نوع الأداء	نسبة الأداء	المعلمة	قيمة اختبار t	درجة الحرية	الاحتمالية
منتظم	١٠٪	القدرة	٠,٠٨٣	٣٩٩٨	٠,٩٣٤
	٢٠٪	القدرة	-٠,٢٥٦	٣٩٩٨	٠,٧٩٨
	٣٠٪	القدرة	٠,٠٠١	٣٩٩٨	١,٠٠٠
غير منتظم	١٠٪	القدرة	٠,٠٤٤	٣٩٩٨	٠,٩٦٥
	٢٠٪	القدرة	-٠,٠٨٠	٣٩٩٨	٠,٩٣٦
	٣٠٪	القدرة	-٠,٠٤٣	٣٩٩٨	٠,٩٦٥

* مستوى الدلالة ($\alpha = 0,05$)

يلاحظ من نتائج اختبار (t) أن الفروق في معلمة القدرة العائدة للنموذج غير دالة إحصائياً ($p > 0,05$) ولكل الظروف التجريبية المتعلقة بنوع ونسبة الأداء التفاضلي مما يشير لان قيم معلمة القدرة لا تختلف باختلاف النموذج المستخدم.

٧,١ مناقشة النتائج

توصلت الدراسة إلى أن الفروق في معلمة القدرة ($p > 0,05$) العائدة للتفاعل بين نسبة ونوع الأداء التفاضلي غير دالة إحصائياً عند مستوى الدلالة ($\alpha = 0,05$) لكل من النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي مما يشير الى ان قيم القدرة لا تختلف باختلاف الظروف التجريبية المتمثلة في نوع ونسبة الأداء التفاضلي. كما توصلت إلى أن الفروق في معلمة القدرة العائدة للنموذج غير دالة إحصائياً ($p > 0,05$) ولكل الظروف التجريبية المتعلقة بنوع ونسبة الأداء التفاضلي مما يشير لان قيم معلمة القدرة لا تختلف باختلاف النموذج المستخدم (النموذج ثلاثي المعلمة ونموذج موكن اللامعلمي)، وقد يعزى ذلك الى ان التقدير يكون ثابت بتغير الفقرات والظروف وهو ما يطلق عليه اللاتغير (Invariance) والمتمثل في عدم تأثر تقدير قدرة الأفراد بمعالم فقرات الاختبار. كما ويمكن تفسير عدم وجود فروق في تقدير قدرة الأفراد باختلاف كلا النموذجين بسبب اعتماد البرمجيات على نمط استجابة الفرد في تقدير القدرة وليس على مجموع علاماته على الاختبار وهذا ما تتفق على تقديره كل النماذج حيث إن الأداء التفاضلي في الفقرات لا يؤثر على تقدير قدرة الأفراد حيث إن آلية توليد الأداء التفاضلي تفترض خلق الأداء التفاضلي في معالم الفقرات. هذه النتيجة مختلفة مع ما تم التوصل إليه في دراسة (النوافلة، ٢٠١٣) والتي بينت وجود دلالة إحصائية بين متوسطات تقديرات معالم الصعوبة تعزى إلى نسب الفقرات ذات الأداء التفاضلي لصالح المجموعة المستهدفة، وفروقات مختلفة بين متوسطات معالم التمييز والتخمين لصالح المجموعة المرجعية، ودراسة (القيسي، ٢٠١٣) التي بينت أن تقديرات معالم القدرة بالطريقة المعلمية أقل منها في الطريقة اللامعلمية.

٨. التوصيات

في ضوء النتائج التي تم التوصل لها يمكن التوصية بما يلي:

١. تفضيل استخدام نموذج موكن اللامعلمي عند توخي أعلى درجات الثبات في الاختبار.
٢. ضرورة التحقق من افتراضات النظرية الحديثة وفي ضوء ذلك يتم اختيار النموذج الأنسب (معلمي / لا معلمي) عند تقدير معالم الفقرات.
٣. يمكن استخدام أي من النموذجين (المعلمي / اللامعلمي) دون شروط وافتراضات إذا كان البحث يهدف إلى تقدير قدرات الأفراد.
٤. إمكانية إجراء دراسات أخرى للمقارنة بين النماذج المعلمية واللامعلمية لمعلمة واحدة واستخدام بعض المتغيرات التجريبية التي لم تتناولها هذه الدراسة مثل طول الاختبار.

بيان تضارب المصالح

يقر جميع المؤلفين أنه ليس لديهم أي تضارب في المصالح.

المراجع

- أبو حماد، كوثر. (٢٠٠٨). *تقصي فاعلية منحني خصائص المفردة في الكشف عن الأداء التفاضلي لمتغير الجنس في عينة مختارة من اختبار في الرياضيات* [رسالة ماجستير غير منشورة]. جامعة اليرموك، إربد، الأردن.
- البستنجي، محمود. (٢٠٠٤). *مقارنة أربع طرق للكشف عن الأداء التفاضلي لمتغير الجنس في فقرات اختبار قدرات عقلية خاصة للفئة العمرية (١٥-١٦) سنة في الأردن* [أطروحة دكتوراه غير منشورة]. جامعة عمان العربية للدراسات العليا، عمان، الأردن.
- البشاشة، خالد. (٢٠١٦). *الكشف عن الأداء التفاضلي لمتغير الجنس لفقرات الاختبار الوطني الأردني لضبط نوعية التعلم لمادة الرياضيات للصف العاشر. مجلة كلية التربية بالزقازيق، ٢٨ (٧٩)، ١-٣٥.*
- زكري، علي. (٢٠٢٠). *الكشف عن الأداء التفاضلي لفقرات اختبار "إمبو" لأساليب المعاملة الوالدية لدى طلبة المرحلة الثانوية. مجلة كلية التربية، جامعة الأزهر، ٣ (١٨٦).*
- الشرقاوي، وآخرون. (١٩٩٦). *اتجاهات معاصرة في القياس والتقويم النفسي والتربوي*. القاهرة: الانجلو المصرية.
- الشريفين، نضال. (٢٠١٨). *أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرات. مجلة دراسات العلوم التربوية، ٤٥، ٦٣٢-٦٥٠.*
- القيسي، حسين. (٢٠١٣). *دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار* [رسالة دكتوراه غير منشورة]. جامعة اليرموك، إربد، الأردن.
- العمرى، حسان، والشطناوي، عبد الله. (٢٠١٦). *الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف العاشر تبعاً لمتغير الجنس. مجلة جامعة النجاح للبحوث الإنسانية، ٣٠ (٨)، ١٥٣٠-١٥٥٤.*
- الناجي، عادل. (٢٠١١). *فاعلية طريقة التحليل العاملي المقيد مقارنة بعدة طرق للكشف عن الأداء التفاضلي على المفردة* [أطروحة دكتوراه غير منشورة]. جامعة اليرموك، إربد، الأردن.
- النوافلة، علي. (٢٠١٣). *أثر نسبة الفقرات ذات الأداء التفاضلي ومستواه في تقدير معالم فقرات الاختبار وقدرات الأفراد وفق نموذج الاستجابة ثلاثي المعلمة* [أطروحة دكتوراه غير منشورة]. جامعة اليرموك، إربد، الأردن.

References

- Acar, T., & Kelecioğlu, H. (2010). Comparison of Differential Item Functioning Determination Techniques: HGLM, LR and IRT-LR. *Educational Sciences: Theory and Practice*, 10(2), 639-649.
- AERA, APA, & NCME. (1999). *The standards for educational and psychological testing*. Washington, DC: AERA Publications Sales.
- Alquraan, M., & Alkuwaiti. (2017). Differential item functioning in students rating of teaching effectiveness surveys in higher education according to academic disciplines: Data from a Saudi university. *Studies Psychological and Educational of J*, 4(11), 780-770.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison-Wesley.
- Camilli, G., & Shepared, L. (1994). *Methods for identifying bias test item*. Sage Publications.
- Chung, W., & Huisu, Y. (2004). Effects of average signed area between two item characteristic curves and purification procedures on the DIF detection via the Mantel-Hanzel method. *Applied Measurement in Education*, 17(2), 113-144.
- Croudace, T., & Brown, A. (2012). Measurement invariance and differential item functioning. Short Course in Applied Psychometrics, Peterhouse College, 10-12.
- Dorans, N., & Holland, E. (1993). DIF detection and description Mantel-Haenszel standardization. *Educational Testing Service*. (QAT24225) ED287526.
- Fitzpatrick, A., & Wendy, M. (2001). The effects of test length and sample size on the reliability and equating of tests composed of constructed response items. *Applied Measurement in Education*, 14(1), 31-57.
- Gabriel, E. L. (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 [Unpublished dissertation]. University of South Florida.
- Garre, F., & Vermunt, J. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*, 33(1), 43-59.
- Gierl, M. J., Jodoin, M. G., & Ackerman, T. A. (2000). Performance of Mantel-Haenszel, Simultaneous Item Bias Test, and logistic regression when the proportion of DIF items is large [Paper presentation]. Annual Meeting of the American Educational Research Association (AERA), New Orleans, LA.
- Gruijter, D., & Kamp, L. (2005). Statistical test theory for education and psychology. Retrieved December 30, 2021, from www.leidenuniv.nl/gruijterdnmde.
- Hambleton, K., & Rogers, J. (1986). Technical advances in credentialing examinations. *Evaluation & the Health Professions*, 9(2), 205-229.
- Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and application*. Boston, MA: Kluwek NIJ Publishing.

- Hambleton, R. (1990). Item response theory: Introduction and bibliography. *Psicothema*, 2(1), 97-107.
- Han, T., & Hambleton, K. (2007). User's manual: WinGen (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts, School of Education.
- Hanson, A. (1998). Uniform DIF and DIF defined by differences in item response functions. *Journal of Educational and Behavioral Statistics*, 23(2), 112-129.
- Haughbrook, R. (2020). Exploring racial bias in standardized assessments and teacher-reports of student achievement with differential item and test functioning analyses [Doctoral dissertation]. The Florida State University.
- Karami, H., & Salmani Nodoushan, A. (2011). Differential item functioning (DIF): Current problems and future directions. *Online Submission*, 5(3), 133-142.
- Lord, F., & Novick, M. (1980). *Application of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pae, I. (2004). Gender effect on reading comprehension with Korean EFL learners. *System*, 32(2), 265-281.
- Park, C. (2010). Examining the relationship between differential item functioning and differential test functioning. *Language Testing*, 23(4), 475-496.
- Raju, N., & Ellis, B. (2002). Differential item and test functioning. *Jossey-Bass*, 156–188.
- Raquel, M. (2019). The Rasch measurement approach to differential item functioning (DIF) analysis in language assessment research. In *Quantitative data analysis for language assessment*, 1, 103-131.
- Rebecca, Z., Dorothy, T., & John, M. (1997). Descriptive and inferential procedures for assessing differential item functioning in polychomous items. *Applied Measurement in Education*, 10(4), 321-344.
- Reynolds, R., & Lowe, A. (2009). The problem of bias in psychological testing. *School Psychology Handbook*, 332-374.
- Ryan, E., & Chiu, S. (2001). An examination of item context effects, DIF, and gender DIF. *Applied Measurement in Education*, 14(1), 73-90.
- Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Sijtsma, K., & Hamker, B. T. (2000). A taxonomy of IRT models for ordering of persons and items using simple sum scores. *Journal of Educational and Behavioral Statistics*, 25(2), 391-415.
- Sijtsma, K., & Molenaar, I. (2002). *Introduction to nonparametric item response theory*. Sage Publications.
- Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement*, 22(3), 3-31.

-
- Sijtsma, K., & Verweij, A. C. (1992). Mokken scale analysis: Theoretical considerations and an application to transitivity tasks. *Applied Measurement in Education*, 5(4), 355-373.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item functioning and differential test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology*, 2(6), 117-135.
- Swaminathan, H., & Rogers, J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 12(4), 361-370.
- Uiterwijk, H., & Vallen, T. (2005). Linguistic sources of item bias for second-generation immigrants in Dutch tests. *Language Testing*, 22(2), 211-234.
- Van De Vijver, F., & Tanzer, K. (2004). Bias and equivalence in cross-cultural assessment. *European Review of Applied Psychology*, 54(2), 119-135.
- Van Der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In *Handbook of modern item response theory* (pp. 1-28). New York, NY: Springer New York.
- Warne, T., Yoon, M., & Price, J. (2014). Exploring the various interpretations of "test bias". *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 570-578.
- Williams, S. (1997). The "unbiased" anchor: Bridging the gap between DIF and item bias. *Applied Measurement in Education*, 10(3), 253-267.