أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة.

نضال كمال الشريفين *

ملخص

هدفت الدراسة إلى الكشف عن أثر نوع الأداء التفاضلي للفقرات (منتظم، غير منتظم) على الخصائص السيكومترية للفقرات، والاختبار وفق النموذج اللوجستي ثلاثي المعلمة، والنموذج اللوجستي الثلاثي اللامعلمي لنظرية الاستجابة للفقرة. ولتحقيق هدف الدراسة تم استخدام برنامج (WINGEN) لتوليد نموذجي اختبار من نوع الاختيار من متعدد، تكون كل منهما من (50) فقرة، تضمن الأول (10) فقرات ذات أداء تفاضلي منتظم، والثاني (10) فقرات ذات أداء تفاضلي غير منتظم، وتم توليد استجابات (1000) فرد، وحللت البيانات وفق النموذج الثلاثي المعلمة باستخدام البرمجية (Bilog-Mg)، ووفق النموذج الثلاثي اللامعلمي باستخدام برنامج (TestGraf). أشارت النتائج إلى وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لكل من معالم التمييز، ومعالم التخمين لفقرات الاختبار يُعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج المعلمي مقارناً بالنموذج اللامعلمي، وأشارت النتائج أيضاً إلى وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لمعالم الصعوبة يُعزى للنموذج المستخدم؛ لصالح النموذج اللامعلمي مقارناً بالنموذج المعلمي، ووجود فروق دالة إحصائيًا بين الأوساط الحسابية لمعالم كلٍ من: الصعوبة والتخمين لفقرات الاختبار تُعزى للتفاعل بين النموذج المستخدم، ونوع الأداء التفاضلي، كما أشارت النتائج إلى وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى للنموذج، لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، حيث كانت كمية المعلومات التي يقدمها الاختبار أكبر عند جميع مستويات القدرة وفق النموذج اللامعلمي، سواء كان الأداء التفاضلي منتظماً أم غير منتظم، باستثناء أن النموذج المعلمي الثلاثي قدم معلومات أكبر من النموذج اللامعلمي عند مستويات القدرة المرتفعة، وبينت النتائج أن قيم معاملات الثبات المقدرة وفق النموذجين كانت عالية، وأن قيم معاملات الثبات المقدرة وفق النموذج المعلمي كانت الأعلى، سواء كان الأداء التفاضلي للفقرات منتظماً أم غير منتظم.

الكلمات الدالة: نظرية الاستجابة للفقرة، الأداء التفاضلي، الخصائص السيكومترية للفقرات، الخصائص السيكومترية للاختبار، النماذج المعلمية، النماذج اللامعلمية، النموذج اللوجستي ثلاثي المعلمة.

المقدمة

تتركز جهود الباحثين، ومطوري الاختبارات على بناء اختبارات تتمتع بخصائص سيكومترية مقبولة للفقرات تتمثل في الصعوبة، والتمييز، والتخمين، وخصائص جيدة للاختبار من حيث الصدق، والثبات. وبالرغم من أهميتها إلا أنها ليست كافية للحكم على صلاحية الاختبار لتحقيق الغرض الذي أعدت من أجله، ونظرا لتأثر خصائص الفقرات، والاختبار بخصائص المجموعة التي يتم التطبيق عليها؛ مما يؤدي إلى اختلاف أدائها التفاضلي باختلاف المجموعة التي يتم التطبيق عليها، فقد زاد الاهتمام بتطوير اختبارات تتمتع بخصائص سيكومترية ملائمة، لا تتأثر بعينة المفحوصين، حيث يتم تطبيقها على عينات مختلفة دون اختلاف في أدائها التفاضلي، أو تطبيق عينات مختلفة من الفقرات على المفحوصين أنفسهم دون أن تتأثر خصائص الفقرات بخصائص العينة، بالإضافة إلى خلو فقرات الاختبار من خصائص معينة تحابي مجموعة دون أخرى من المجموعات التي يتم تطبيق الاختبار عليها، وعدم تأثرها بعوامل ذات علاقة بعينة الدراسة كالجنس، والعرق، ومكان الإقامة، واللغة، والمستوى الاقتصادي والاجتماعي التي تؤثر بدورها على القرارات المتخذة بناء عليها.

وقد هدفت البحوث المتعلقة بالأداء التفاضلي إلى إزالة التحيز من الاختبارات التربوية والنفسية. فأداء الفقرة على الاختبار يجب ألا يتغير عبر المجموعات المختلفة إذا كان أفرادها لهم مستوى القدرة نفسه على السمة التي يقيسها الاختبار، ويظهر الأداء التفاضلي

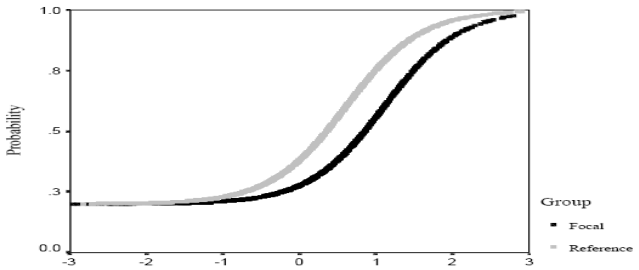
* كلية التربية، جامعة اليرموك. تاريخ استلام البحث 2016/2/4، وتاريخ قبوله 2016/6/24.

© 2018 عمادة البحث العلمي / الجامعة الأردنية. جميع الحقوق محفوظة.

للفقرة (Differential Item Functioning, DIF) عندما يختلف أداء المفحوصين المتساوين في مستوى القدرة، والذين ينتمون إلى مجموعات مختلفة، ونظرًا لاحتمالية أن يهدِّد الأداء التفاضلي صدق الاختبار، فالكشف عن الأداء التفاضلي للفقرات وإزالتها يعدُّ ضروريًا للقياس الصادق والمتحرر من التحيز، لذا فقد عد الأداء التفاضلي طريقة إحصائية للكشف عن التحيز المحتمل للفقرة، ومن هنا ظهرت العديد من الطرق الإحصائية للكشف عن الأداء التفاضلي، منها ما يعتمد على نظرية القياس الكلاسيكية: كطريقة الصعوبة المحولة، وطريقة مانتل هانزل، وطريقة الانحدار اللوجستي، ومنها ما يعتمد على نظرية الاستجابة للفقرة (Item Response Theory, IRT). كما يتم تصنيف طرق الكشف عن الأداء التفاضلي وفقًا لنظرية الاستجابة للفقرة إلى طرق معلمية (Parametric)، ولا معلمية (Nonparametric)، ويشير مصطلح معلمي إلى أن المعالم المميزة للفقرات والمفحوصين تقدَّر من خلال تحليل البيانات، التي تتطلب تحديد نموذج الاستجابة للفقرة، وتقدير المعالم ذات العلاقة لمقارنة الفقرات عبر مجموعتين تسمى الأولى: مرجعية (Referenced Group)، والثانية: مستهدفة (Focal Group)، ومن هذه الطرق طريقة نسبة الأرجحية، وطريقة راجو، وطريقة فرق المساحة بين منحنيات الخصائص للفقرة وغيرها (Park, 2010).

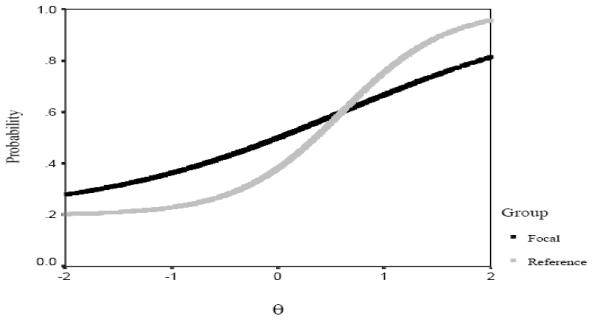
والأداء التفاضلي للفقرة مفهوم إحصائي يستدل عليه من دالة مشتقة إحصائيا للتعبير عن الفروق في احتمالات الاستجابة الصحيحة للفقرة بين مجموعتين ممن هم في مستوى القدرة نفسه (Gruijter & Kamp, 2005)، حيث إن وجود الأداء التفاضلي للفقرة مؤشر بأن هناك عوامل مرتبطة تنتمي إلى مجموعة معينة تؤثر في احتمالية الإجابة الصحيحة للفقرة، وبالتالي فإنها تهدد التقييم العادل بين الأفراد. والأداء التفاضلي يتم الكشف عنه من خلال فحص العلاقة بين الاحتمالات المشروطة للاستجابة الصحيحة على الفقرة، وبين السمة الكامنة المراد قياسها، حيث يتم مقارنة منحنيات خصائص الفقرة بين المجموعتين اللتين يتم رسمهما وفق النموذج المستخدم (معلمي أو لا معلمي) (Gybles, 2004).

وتعد طريقة فرق المساحة بين منحنيات الخصائص للفقرة من الطرق المستخدمة للكشف عن الأداء التفاضلي، فإذا تشابهت منحنيات خصائص الفقرة في المجموعات الفرعية المختلفة، فإن الفقرة لا تظهر أداءً تفاضلياً؛ أي تكون الفقرات متساوية في صعوبتها وتمييزها وتخمينها لجميع المفحوصين من مستوى القدرة نفسه في المجموعات الفرعية المختلفة، وذلك وفقًا للنموذج المستخدم (Chung & Huisu, 2004). وهناك نوعان من الأداء التفاضلي، أولهما: الأداء التفاضلي المنتظم (Uniform Differential Item Functioning, UDIF) الذي ينتج عندما لا يحدث 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، والشكل (1) يبين ذلك.



الشكل1. التمثيل البياني لمنحنى الخصائص للفقرة الذي يظهر أداء تفاضليا منتظماً بين المجموعتين

وثانيهما: الأداء التفاضلي غير المنتظم (Non-Uniform Differential Item Functioning, NUDIF)، الذي يظهر عندما يكون هناك تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد؛ أي أن نمط الفروق في احتمالية الاستجابة الصحيحة ليست متشابهة في كل مستويات القدرة، فهذه الفروق قد تكون لصالح إحدى المجموعات عند مستوى قدرة معين، ولصالح مجموعة أخرى عند مستوى قدرة آخر، حيث معالم التمييز مختلفة عبر المجموعات ومعالم الصعوبة والتخمين متساوية، والشكل (2) يبين ذلك.



الشكل2. التمثيل البياني لمنحنى الخصائص للفقرة الذي يظهر اداء تفاضليا غير منتظما بين المجموعتين.

ولكن يبقى السؤال المحير "هل هناك تأثير لنوع الأداء التفاضلي على خصائص الفقرات وبالتالي خصائص الاختبار"؟ وهذا بالتحديد ما سعت هذه الدراسة للإجابة عنه، وسعيا من المهتمين في مجال القياس النفسي والتربوي للوصول إلى اختبارات ومقاييس تتمتع بدرجة عالية من الدقة والموضوعية فقد نشأت نظريات القياس النفسي التي حاولت الوصول إلى أدوات قياس ملائمة لقياس السمات التي أعدت من أجلها، وبما يكفل الوصول إلى اتخاذ القرارات الصائبة، ومنها النظرية الكلاسيكية في القياس (Classical Test Theory, CTT)، ونظرية الاستجابة للفقرة (Item Response Theory, IRT). وما يجدر الإشارة إليه أن جميع المؤشرات الإحصائية للفقرة، وللاختبار التي تستند في حسابها على النظرية التقليدية، مثل معاملات التمييز، والصعوبة، والثبات تعتمد على خصائص عينة الأفراد التي يطبق عليها الاختبار، وتقديرات القدرة لعينة الأفراد تعتمد على عينة الفقرات، لذا جاءت نظرية الاستجابة للفقرة (Item Response Theory, IRT)؛ للتغلب على جوانب القصور في النظرية الكلاسيكية، حيث قدمت طريقة في انتقاء الفقرات تتميز بتقديم معالم ثابتة للفقرة، هي: معالم الصعوبة، والتمييز، والتخمين، إضافة إلى توحيد المقياس لكل من الصعوبة، وقدرة المفحوص لكي يتمكن مطورو الاختبارات من انتقاء الفقرات الأكثر ملاءمة لتصنيف الأفراد وفقًا لمستويات الأداء التي تتميز في حالة مطابقة البيانات للنموذج بعدم تغير معالم الفقرات بتغير عينة المفحوصين، وعدم تغير تقديرات القدرات بتغير عينة الفقرات (Hambleton & Swaminathan, 1985). وتربط نظرية الاستجابة للفقرة بين قدرات المفحوصين، واستجاباتهم على فقرات الاختبار بنماذج إحصائية (Crocker & Algina, 1986)، وتنقسم النماذج إلى نوعين رئيسين، هما: النماذج المعلمية (Parametric Models)، والنماذج اللامعلمية (Non- Parametric Models).

فقد شاع استخدام النماذج المعلمية من قبل الباحثين على الرغم من موانعها في تحليل البيانات في المستوى الرتبي، إلا أن مصداقية النتائج قد تكون موضع تساؤل عندما لا يتحقق فرض وقوع البيانات على مستوى فئوي، الأمر الذي تبرره النماذج اللامعلمية التي لا تضع قيودا حول شكل دالة الاستجابة للفقرة، مما يثير التساؤل حول مدى مطابقة النوعين للبيانات التحصيلية التي يعتبرها بعضهم رتبية، وبعضهم الآخر شبه فئوية، ومدى دقة النتائج التي تفرزها مثل هذه الاختبارات (Liang, 2010)؛ إذ يمكن تحليل بيانات النماذج المعلمية وفق المعالجات الإحصائية لتقدير معالم الفقرة، والقدرة باستخدام النماذج اللامعلمية، ولا يجوز العكس، حيث النماذج اللامعلمية تستطيع تحليل بيانات على مستوى القياس الرتبي بعكس النماذج المعلمية، التي تشترط وقوع البيانات على مستوى القياس الفئوي على الأقل؛ أي أنه إذا كان من الممكن قياس السمة على مستوى أعلى، فإنه يمكن قياسها على مستوى أدنى، والعكس لا يجوز، بسبب خاصية التصنيف الهرمي في مستويات القياس، ولكن قد يكون هناك اختلاف بين النموذجين في دقة التقديرات التي يعطيها كلاهما (Sijtsma & Hemker, 2000).

ومنذ الخمسينيات زاد الاهتمام بالإحصاء اللامعلمي لأهميته البالغة في حساب الدلالة الإحصائية، وخاصة عندما لا تصلح المقاييس المعلمية لحساب تلك الدلالة؛ لعدم توفر الشروط اللازمة لاستخدامها، وقد شاع استخدام هذا النوع من الإحصاء في العينات الصغيرة، والصغيرة جداً التي قد يلجأ إليها الباحث النفسي لاختيار أدوات قياسه بطريقة مبدئية، وسريعة، وفى التوزيعات الحرة غير المقيدة بالتوزيع الاعتدالي. هذا ولا يقتصر استخدام الإحصاء اللامعلمي على هاتين الناحيتين، بل يمتد أيضا للعينات الكبيرة، وتقترب أغلب مقاييسه في توزيعاتها من التوزيع الاعتدالي وفقًا لزيادة حجم العينة، ولذلك ينفرد بالتحليل الإحصائي لمستويات القياس الوصفي والرتبي، ويمتد أيضاً للمستويات الأخرى للقياس الدقيق مثل النسبي بينما يقتصر مجال استخدام الإحصاء المعلمي على المستويات العليا للقياس التي تتمثل في مقياس الفئات المتساوية (Sijtsma & Molenaar, 2002).

ويشير سيجتسما ومولينار (Sijtsma & Molenaar, 2002) إلى أن النماذج اللامعلمية تقوم على عددٍ من الافتراضات

التي تعد أقل تشدداً من تلك التي تقوم عليها النماذج المعلمية، وهي: أحادية البعد (Unidimensionality)، والاستقلال الموضعي (Local Independence)، والاطرادية (Monotonicity)، حيث إنه بازدياد قيمة القدرة (θ) لدى المفحوص تزداد احتمالية الاستجابة على الفقرة إجابة صحيحة، أو تبقى ثابتة ضمن مستويات القدرة المختلفة بالإضافة إلى افتراض الاطرادية المضاعفة (Double Monotonicity)، وهو يشير إلى امتلاك دوال استجابة غير متقاطعة لفقرات الاختبار التي تشكل التدريج. ويعتبر هذا الافتراض الأصعب مقارنةً بباقي الافتراضات، ويندرج في نظرية الاستجابة للفقرة اللامعلمية نموذجان لامعلميان، الأول: نموذج التجانس الاطرادي ((Monotone Homogeneous Model, (MH)، ويشير سيجتسما وفيروج (Sijtsma&Verweij, 1992) إلى أن نموذج الاستجابة للفقرة اللامعلمي معروف بنموذج موكن، ويستخدم لتحليل التدريج للاستجابات الثنائية، وتحليل التدريج لموكن عبارة عن تحليل التدريج لجتمان؛ أي يعطي تدريجا للأفراد والفقرات على مقياس رتبي أحادي البعد، واقترح موكن في نموذج التجانس الاطرادي عام (1971) تحقق ثلاثة افتراضات أساسية، وهي: افتراض أحادية البعد للمقياس، والاستقلال الموضعي لفقرات المقياس، واطرادية منحنيات خصائص الفقرة (Item Characteristic Curve: ICC)؛ أي أن تكون دالة استجابة الفقرة غير متناقصة كدالة للقدرة (IRF) بزيادة مستوى قدرة المفحوص (Ɵ)، فالافتراض الأول والثاني ينطبقان مع نماذج نظرية الاستجابة للفقرة المعلمية IRT، بينما الافتراض الثالث يفترض أن منحنيات خصائص الفقرة (ICCs) اطرادية غير متناقصة (Nozawa, 2008)، ونظراً لعدم توفر قيود إضافية كباقي النماذج كما في النموذجين ثنائي وثلاثي المعلمة يعتبر نموذج التجانس الاطرادي من النماذج الأكثر تحرراً، فهو بذلك من النماذج الفعالة؛ لأنه يسمح بترتيب الأفراد على السمة باستخدام الدرجة الكلية (Douglas 1997). والنموذج الثاني: نموذج الاطراد المضاعف (Double Monotonicity Model, (DM)، وهو نموذج موكن الثاني اللامعلمي الأكثر تعقيداً (المقيد)، فبالإضافة إلى الافتراضات الثلاثة السابقة للنموذج (MH)، يوجد افتراض رابع يتطلب تحقيقه، حيث إن منحنيات خصائص الفقرة غير متقاطعة معاً تحت هذا الافتراض (ICCs) التي تكون دوال غير متزايدة، لكن يسمح لها بالتماس في المناطق المتطرفة، كما يسمح بترتيب الأفراد على متصل السمة وفق الدرجة الكلية (Sijtsma & verweij, 1992).

وللتحقق من التجانس الاطرادي في فقرات الاختبار أو المقياس يتم إيجاد مجموعة من المعاملات تسمى بمعاملات التدريج (scalability coefficients) من الضروري توفرها وهي: (H, Hi, Hij)، حيث يشير موكن ولويس (Mokken & Lewis, 1982) إلى أن الهدف الأساسي من نماذج التدريج عامة هو تزويد الفاحص بطريقة لقياس جودة مجموعة من الفقرات في مطابقة تدريج معين، وقد عرفا المقياس على أنه مجموعة من الفقرات التي ترتبط مع بعضها إيجابيا، حيث يكون معامل التدريج (Hi) أكبر من أو يساوي قيمة ثابتة موجبة (c)، حيث (c) ثابت تعريف التدريج. ففي حال تحقق التعريف السابق على مجموعة الفقرات يطلق عليها أنها فقرات تتبع تدريج موكن، ويمكن استخراج معامل تدريج الفقرة (Hi) لكل فقرة من فقرات المقياس المكون من (K) من الفقرات، وهو معامل يناظر معلم التمييز في النماذج المعلمية ثنائية، وثلاثية المعالم، ثم يتم استخراج معامل التدريج الكلي للاختبار (H)؛ حيث إنه كلما زادت قيمة (H) دل ذلك على موثوقية أكبر في المقياس، وقدرته في ترتيب المفحوصين على التدريج باستخدام الدرجة الملاحظة كما في نظرية الاختبار الكلاسيكية؛ إذ يقتضي المحك الثاني أن تكون جميع معاملات التدريج موجبة كالتالي:$1 \geq Hij \geq 0$ لجميع قيم i,j، و $1 \geq Hi \geq 0$ لجميع قيم i مع باقي فقرات المقياس، و $1 \geq H \geq 0$. وقد ذكر كل من سيجتسما ومولينار (Sijtsma & Molenaar, 2002) أن موكن قام بتحديد معايير خاصة ببعض معاملات التدريج لضمان كفاءة التدريج، واقترح حدا أدنى لمعامل تدريج على مستوى التدريج (Hi= 0.3)، كما اقترح بعض المعايير للحكم على جودة التدريج الكلي، وهي كالتالي: مقياس غير قابل للتدريج ($H < 0.3$)، مقياس ضعيف ($0.3 \leq H \leq 0.4$)، مقياس متوسط ($0.4 \leq H \leq 0.5$)، مقياس قوي ($H \geq 0.5$).

أما نماذج نظرية الاستجابة للفقرة المعلمية فهي مجموعة من النماذج التي تعرف بنماذج السمات الكامنة (Latent Trait Models). وتعد النماذج السكونية أحادية البعد الأكثر شيوعاً في تصميم الاختبارات والمقاييس التربوية والنفسية وبنائها، وهي الأكثر ملاءمة للفقرات ثنائية التدريج، ويعد عدد المعالم التي توصف الفقرة بها الفارق الأساسي بين نماذج السمات الكامنة أحادية البعد (Gruijter & Kamp, 2005؛ علام، 2005). ويعد النموذج اللوجستي ثلاثي المعالم ((3PLM) Parameter Three- Logistic Model) المستخدم لأغراض تحليل بيانات هذه الدراسة أقل النماذج أحادية البعد تشدداً؛ إذ يسمح بأن تختلف فقرات الاختبار في صعوبتها وتمييزها، وهو ما يلاحظ عادة في بناء الاختبارات؛ إذ من الصعب إيجاد مجموعة من الفقرات التي تميز بدرجة واحدة بين مستويات السمة، أو القدرة التي يقيسها اختبار معين، وهذا الافتراض الذي اعتمده النموذج

الأحادي المعالم (One Parameter Logistic Model or Rasch Model). كذلك يفترض تأثر الإجابات بعامل التخمين، وهو المعلم الثالث الذي تميز به عن النموذج الثنائي المعالم (Two –Parameter Logistic Model)، الذي أضافه لورد، وأطلق عليه معلم الخط التقاربي الأدنى (Lower Asymptote Line) ، أو معلم التخمين (Guessing Parameter)، وهذا المعلم يحدد احتمال أن يجيب فرد يفترض أن مستوى قدرته متدن جداً، ومع هذا يجيب إجابة صحيحة عن فقرات الاختبار عن طريق التخمين، والحقيقة أن إضافة هذا المعلم الثالث يكون بمثابة تعويض عدم مطابقة المنحنيات المميزة للفقرات عند النهاية السفلى لمتصل القدرة؛ نتيجة تأثر إجابات هؤلاء الأفراد بعامل التمييز(Lord, 1980; Hambleton, 1989). ويوصف النموذج اللوجستي الثلاثي المعالم بالمعادلة الرياضية الآتية:

$$p_i(\theta) = c_i + \frac{1-c_i}{1+e^{Dai(\theta-bi)}}(i = 1,2,....n\ )$$

حيث $p_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ) عن الفقرة (i) إجابة صحيحة، D: ثابت التدريج ويساوي تقريباً 1.7، b_i: معلم الصعوبة للفقرة(i)، وهي النقطة التي تقع على متصل القدرة عندما يكون ميل منحنى خصائص الفقرة أكبر ما يمكن، a_i: معلم تمييز الفقرة، وهو ميل منحنى خصائص الفقرة عند انعطاف المنحنى، c_i: خط التقارب السفلي لمنحنى خصائص الفقرة الذي يمثل احتمالية إجابة المفحوص بأقل قدرة إجابة صحيحة على الفقرة (معلم التخمين للفقرة). ويعتمد احتمال إجابة الفرد عن الفقرة إجابة صحيحة في نماذج نظرية الاستجابة للفقرة، على قدرة الفرد (θ)، والمعالم المتعلقة بالفقرة متمثلة بمعلم الصعوبة (b)، ومعلم التمييز (a)، ومعلم التخمين (c)، ولكن هذه القيمة الاحتمالية، وهذه المعالم تكون غير معلومة، بينما إجابات الأفراد عن الفقرة تكون معلومة، لذلك فإن أساليب تقدير المعالم تهدف لتحديد قيمة (θ) لكل فرد، وكذلك قيم معالم الفقرة من إجابات الأفراد عليها.

ومن المؤشرات المستخدمة للكشف عن الخصائص السيكومترية للفقرات وللاختبار دالة معلومات الفقرة والاختبار، ويمكن توضيح ذلك للنماذج المعلمية وللنماذج اللامعلمية في نظرية الاستجابة للفقرة على النحو الآتي: فقد أشار هامبلتون وسوامينثان (Hambleton; Swaminathan& Rogers,1991) إلى أن دالة المعلومات تمثل إحدى الطرق التي تمكننا من بناء اختبارات أفضل– حيث تعطي معلومات أكثر حول السمات الكامنة المقدرة، ويمكن استخدامها أيضا للتأكد من دقة تقدير معالم الفقرة، ومعالم القدرة، وتأتي أهمية دالة المعلومات للفقرة من كون مساهمة كل فقرة بدالة معلومات الاختبار تتحدد بشكل مستقل عن باقي الفقرات في الاختبار. وفي حالة نماذج نظرية الاستجابة للفقرة المعلمية يقدم بيكر (Baker, 2001) دالة معلومات الفقرة لكل نموذج من نماذج النظرية، وتختلف باختلاف النموذج المستخدم (Loken & Rulison,2010). ومن المؤشرات الأخرى التي يستدل منها على معامل الثبات للاختبار في النظرية الحديثة في القياس هي دالة المعلومات للاختبار (Information Function)، ومنحنى دالة المعلومات للاختبار ينتج من تكوين منحنيات الفقرات فوق بعضها بعضا وفق العلاقة: $I(\theta) = \sum Ig(\theta)$

حيث ($I(\theta)$): هي كمية المعلومات للاختبار عند مستوى القدرة، $\sum Ig(\theta)$ مجموع دوال المعلومات لفقرات ذلك الاختبار عند مستوى القدرة نفسه. ولذلك فإن زيادة عدد الفقرات يعطي خطأ معيارياً صغيراً S.E (θ)، ونقصان قيمة الخطأ المعياري في تقدير القدرة عند مستوى القدرة يؤدي إلى زيادة كمية المعلومات للاختبار وفق العلاقة الآتية:

$$I(\theta) = \frac{1}{\sqrt{S.E(\theta)}}$$

أما دالة المعلومات الخاصة بالفقرة والاختبار وفق نماذج نظرية الاستجابة للفقرة اللامعلمية فيرى لوكن ورليسون (Loken & Rulison,2010) أن دالة المعلومات للاختبار، والفقرة تلعب دورا مهمًا في تقدير دقة القياس؛ حيث يتم تقدير دالة معلومات الفقرة والاختبار في نماذج نظرية استجابة الفقرة اللامعلمية، كما هو الحال في النماذج المعلمية، حيث تعطى دالة معلومات الفقرة للفقرات الثنائية في نماذج نظرية الاستجابة للفقرة كما يلي:

$$I(\theta) = \frac{\left[P^1(\theta)\right]^2}{[P(\theta)][Q(\theta)]}$$

حيث (□)I: كمية المعلومات التي نحصل عليها عند مستوى القدرة (□).

$\left[P^1(\theta)\right]^2$: مربع المشتقة الأولى لدالة الاستجابة (لمنحنى خصائص الاختبار) عند مستوى القدرة (□).

$[P(\theta)]$: نسبة الناجحين في الاختبار.

$[Q(\theta)]$: 1-p(□).

وفي النماذج اللامعلمية يتم تحديدها من خلال استخدام برنامج (Test Graf) برسم متوسط دالة معلومات الفقرة (mean item information function). أما فيما يخص تقدير معالم الفقرات وقدرات الأفراد وفق نماذج نظرية الاستجابة للفقرة اللامعلمية، فإنه يمكن تقدير كمية المعلومات للفقرة بالاعتماد على طرق الانحدار اللامعلمي (Non-Parametric Regression Methods)، وهي: Kernel Smoothing (KS)، وهي تقنية تشير إلى تقدير الانحدار اللامعلمي بالاعتماد على نظرية الاستجابة للفقرة، حيث تتيح تنظيم المفحوصين وفق تقدير القدرة. وتعد طريقة كيرنل من أفضل طرق تقدير كميات المعلومات من خلال المنحنيات اللامعلمية مقارنة بغيرها من الطرق؛ وذلك لبساطة تقديرها من خلال توفر البرامج الحاسوبية، بالإضافة إلى إمكانية استخدامها لبيانات ثنائية التدريج وذات مستويات قياس أسمية. كما أنه ليس هناك حاجة للقلق بشأن عملية التقدير؛ لأن هناك تقاربا مشتركا ما بين تقديرات القدرة الرتبية، وطريقة كيرنل في تقدير منحنيات خصائص الفقرة (ICCs) (Nozawa, 2008).

الدراسات السابقة

يتناول هذا الجزء عرضًا للدراسات ذات الصلة بموضوع أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة – ففي حدود اطلاع الباحث– لا توجد دراسات بحثت مباشرة بالموضوع نفسه، فهناك دراسات بحثت أثر الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات وفق النماذج المعلمية، ودراسات قارنت بين النماذج المعلمية والنماذج اللامعلمية، لذلك قام الباحث بتناول الدراسات التي بحثت أثر الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات وفق نماذج نظرية الاستجابة للفقرة، وقد تم عرض هذه الدراسات وفق التسلسل الزمني والمنطقي.

فقد أجرى زمبو (Zumbo, 2000) دراسة هدفت إلى الكشف عن أثر الفقرات ذات الأداء التفاضلي على الإحصائيات الخاصة بالاختبار، وهي: معامل كرونباخ ألفا، والخطأ المعياري للقياس، ومعامل الارتباط المصحح من أثر الانخفاض وغيرها، وفق النظرية الكلاسيكية في القياس. تكون الاختبار من (38) فقرة من نوع الاختيار من متعدد تضمنت فقرات أبدت أداء تفاضلياً منتظماً. بينت نتائج الدراسة أنه لم يكن هناك تأثير دال إحصائياً على الإحصائيات الخاصة بالاختبار وفق النظرية الكلاسيكية في القياس جميعها، كما قام أنجل وجيلدون وجوس (Angel,Gildon & Jose, 2000) بدارسةٍ هدفت إلى الكشف عن أثر نسبة الفقرات ذات الأداء التّفاضليّ في الاختبار، وطول الاختبار على فاعلية طريقة مانتل هانزل في الكشف عن الفقرات ذات الأداء التّفاضلي، أظهرت نتائج الدراسة أنّ طريقة مانتل هانزل (MH) تكون فعّالة وذات قوة إحصائيّةٍ تكفي للكشف عن الأداء التّفاضليّ للفقرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ المنتظم تساوي (10%)، بينما تنخفض فاعليتها للنسب(20%,30%)، وتشير النّتائج أنّ طول الاختبار كان له أثر منخفض على القوة الإحصائية لمؤشر مانتل هانزل.

وفي دراسة قام بها جودوين وجيرل (Jodoin & Gierl,2001) هدفت إلى دراسة تقييم الخطأ من النوع الأول، وقوة الاختبار الإحصائي من خلال مقياس حجم الأثر لطريقة الانحدار اللوجستي، وذلك من خلال دراسة محاكاة، بينت نتائج الدراسة انخفاض الأخطاء من النوع الأول مع زيادة حجم العينة، كما بينت النتائج أن قوة الكشف عن الأداء التفاضلي غير المنتظم أقل من الأداء التفاضلي المنتظم في جميع الحالات، وأجرى كماتا وشو (Kamata & Chu, 2003) دراسة هدفت إلى الكشف عن دقة تقديرات معالم الصعوبة للفقرات، ومعالم القدرة للأفراد، ومعادلة درجات الاختبار باستخدام نموذج راش الهرمي ونموذج راش الكلاسيكي بوجود فقرات ذات أداء تفاضلي، أشارت نتائج الدراسة إلى فاعلية نموذج راش الهرمي مقارنة مع نموذج راش التقليدي، فيما يتعلق بدقة تقدير معالم الصعوبة للفقرات، وقدرات الأفراد بوجود الفقرات ذات الأداء التفاضلي، وتأثر نموذج راش الكلاسيكي بشكل كبير بعدد الفقرات ذات الأداء التفاضلي، وكذلك بمستوى الأداء التفاضلي، كما قام هانغ وماتيوس ودورانس (Zhang, Mattews & Dorans, 2003) بدراسة هدفت إلى تحديد أثر الفقرات ذات الأداء التفاضلي في دقة المعادلة، واستخدم نتائج اختبار الاستعداد المدرسي (Scholastic Aptitudes Test, SAT)، حيث أشارت نتائج الدراسة إلى أن حذف الفقرات ذات الأداء التفاضلي كان لصالح المجموعة المرجعية في حين تأثر تدريج المجموعات المستهدفة سلباً بحذف الفقرات ذات الأداء التفاضلي.

وحول الموضوع نفسه فقد أجرى جيرل (Gierl, 2004) دراسة هدفت إلى الكشف عن فاعلية طريقة تحيز الفقرة، تحت ظروف تجريبية مختلفة عن طريق بيانات مولدة على الخطأ من النوع الأول، أشارت نتائج الدراسة إلى أن نسبة القرارات

الصحيحة تزداد كلما ازداد حجم العينة عندما يكون الأداء التفاضلي منتظماً، وتتناقص قيمها عندما يكون الأداء التفاضلي غير منتظم في معظم الحالات، وأن نسبة القرارات الصحيحة تكون أعلى في جميع الظروف عندما يكون الأداء التفاضلي غير منتظم، وعد اختبار تحيز الفقرة طريقة غير دقيقة للكشف عن الأداء التفاضلي للفقرة في جميع الظروف التي تكون فيها نسبة الفقرات ذات الأداء التفاضلي في الاختبار 40% و 60%، ويكون فيها الأداء التفاضلي غير منتظم.

وفي دراسة لكاماتا وشو (Kamata & Chu, 2005) هدفت إلى قياس دقة وثبات تقدير القدرة وصعوبة الفقرات بعد إجراء معادلة لدرجات الاختبار بوجود وعدم وجود الفقرات ذات الأداء التفاضلي وفق نموذج راش الهرمي، بينت نتائج الدراسة أن النموذج الهرمي لراش كان الأفضل بين النماذج الأخرى المستخدمة في تقدير القدرة ومعالم الفقرات، وكانت دقة المعادلة باستخدام نموذج راش الهرمي هي الأكثر دقة باستخدام معيار الجذر التربيعي لمتوسط مربعات الأخطاء للمعادلة، وقام زهان وزمبو (Zhan & Zumbo, 2009) بدراسة هدفت إلى الكشف عن أثر الأداء التفاضلي للفقرات على الخطأ من النوع الأول، وحجم الأثر، وتم استخدام النموذج اللوجستي ثلاثي المعالم لتقدير المعالم بوجود ثلاثة مستويات للأداء التفاضلي للفقرة، أشارت نتائج الدراسة إلى أن وجود فقرة واحدة ذات الأداء التفاضلي في اختبار مكون من (38) فقرة لم يؤثر في مستوى الخطأ من النوع الأول، أما حين كان عدد الفقرات ذات الأداء التفاضلي (16) فقرة فقد حصل تضخم في مستوى الخطأ من النوع الأول، كما أن زيادة الفقرات ذات الأداء التفاضلي يؤدي إلى زيادة الفروقات في حجم الأثر.

وأجرى الحياصات (2011) دراسة هدفت إلى تحديد أثر الفقرات ذات الأداء التفاضلي في تحقيق معايير الفاعلية في معادلة الاختبار باستخدام النموذج اللوجيستي ثلاثي المعالم، حيث أشارت النتائج إلى أن حذف الفقرات ذات الأداء التفاضلي يؤدي إلى زيادة دقة المعادلة وفاعليتها وفقا لطريقة "تكر" الخطية التي تتبع النظرية الكلاسيكية. وبينت النتائج أيضاً أن حذف الفقرات ذات الأداء التفاضلي يؤدي إلى زيادة الدقة للمعادلة وخاصة في الطرف العلوي من العلامات وفقا لطريقة معادلة العلاقات الحقيقية التي تتبع النظرية الحديثة، كما قام يانجو (Yanju, 2012) بدراسة هدفت إلى معرفة كيف يؤثر تمييز الفقرة ذات الأداء التفاضلي على الخطأ من النوع الأول، وتم استخدام طريقتين للكشف عن الأداء التفاضلي، هما: طريقة مانتل–هانزل وطريقة الانحدار اللوجستي، واستخدم الباحث طريقة المحاكاة باستخدام النموذج الثلاثي والثنائي المعالم، في ظل عوامل أخرى، بينت نتائج الدراسة أن هناك تضخما قليلا جدا في الخطأ من النوع الأول عند التمييز المنخفض حتى مع حجم العينة الكبير، ولكل طريقة من طرق الكشف عن الأداء التفاضلي. كما بينت نتائج الدراسة أنه عند التمييز المرتفع للفقرة كان تضخم الخطأ من النوع الأول قليلا، كما تم ملاحظة أنه عندما تم توليد البيانات وفق النموذج ثلاثي المعالم تبين أن المعلم (التخمين) أثر على الخطأ من النوع الأول.

وأجرى الحجيلي (2012) دراسة هدفت إلى الكشف عن أثر طول الاختبار وحجم العينة في دقة تقدير معالم الصعوبة للفقرة والقدرة للأفراد، ومعادلة الاختبارات بوجود فقرات ذات أداء تفاضلي باستخدام نموذج راش الهرمي، أظهرت نتائج الدراسة دقة المعادلة باستخدام النموذج الهرمي لراش، وازدياد دقة تقدير معالم الصعوبة للفقرات، وزيادة دقة التقدير لمعالم القدرة للأفراد بازدياد حجم العينة، أما طول الاختبار فقد عمل على زيادة دقة تقدير القدرة عند زيادة طول الاختبار، ودقة تقدير المعالم أفضل عند تقليص طول الاختبار، وأجرى إرحيل (2013) دراسة هدفت إلى الكشف عن أثر الفقرات ذات الأداء التفاضلي وفقًا لمتغير الجنس على دقة المعادلة العمودية لاختبار أوتيس لينون للقدرة العقلية بمستوييه المتوسط والمتقدم. أشارت النتائج إلى وجود أثر للفقرات ذات الأداء التفاضلي على دقة المعادلة العمودية.

أما دراسة جابريل(Gabriel, 2012) التي هدفت إلى المقارنة بين فعالية ثلاث طرق للكشف عن الأداء التفاضلي باستخدام نماذج معلمية وأخرى لامعلمية، وهي: طريقة اختبار تحيز الفقرة المتزامن والمتقاطع مع الأداء (CIBTEST)، وطريقة اختبار الأرجحية العظمى في نظرية الاستجابة للفقرة (IRT-LR)، وطريقة الانحدار اللوجستي (LOGREG). حيث تم استخدام ظروف تجريبية مختلفة من حيث: طول الاختبار، حجم العينة، مقدار الأداء التفاضلي للفقرة والاختبار، وتوزيع قدرات مختلفة للمجموعات المرجعية والمستهدفة. وتم توليد البيانات وفق النموذج اللوجستي ثلاثي المعلمة، وحساب الخطأ من النوع الأول، والخطأ من النوع الثاني، وقوة الاختبار الإحصائي. كشفت نتائج الدراسة أن طريقة الانحدار اللوجستي كانت أفضل الطرق في قوة الكشف عن الأداء التفاضلي، ومعدل الخطأ من النوع الأول، كما كشفت النتائج عن عدم وجود طريقة واحدة فعالة تميز بين نوعي الأداء التفاضلي.

وقام النوافلة (2013) بدراسة هدفت إلى دراسة أثر نسب الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي في تقدير معالم الفقرات وقدرات الأفراد وفق النموذج اللوجستي ثلاثي المعالم، واستخدم الباحث أسلوب المحاكاة في توليد البيانات تحت ظروف اختبارية متباينة. أشارت نتائج الدراسة إلى وجود دلالة إحصائيا 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، وعدم وجود فروق ذات دلالة إحصائية في تقديرات القدرة تعزى إلى متغيري الدراسة، أما كمية المعلومات فإنها تنخفض كلما ازدادت نسبة الفقرات ذات الأداء التفاضلي، وأجرى اجبارة (2015) دراسة هدفت إلى بحث أثر نسبة الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول، وقوة الاختبار في الكشف عنها باستخدام طريقتي الانحدار اللوجستي ونسبة الأرجحية، حيث تم استخدام أسلوب المحاكاة في توليد البيانات، أظهرت النتائج وجود فروق ذات دلالة إحصائية لمعدل الخطأ من النوع الأول تعزى لنسب الفقرات ذات الأداء التفاضلي، ولصالح النسبة الأقل. كما بيّنت النتائج وجود فروق دالة إحصائيا لمعدل الخطأ من النوع الأول باختلاف طريقة الكشف، ولصالح طريقة نسبة الأرجحية، وعدم وجود فروق دالة إحصائيا تعزى للتفاعل بين النسبة وطريقة الكشف. أما من حيث قوة الاختبار فقد كشفت النتائج عن وجود فروق دالة إحصائيا تعزى لنسب الفقرات ذات الأداء التفاضلي، وكانت النتيجة لصالح النسبة الأقل، وعدم وجود فروق دالة إحصائيا تعزى لطريقة الكشف عن الأداء التفاضلي، وعدم وجود فروق دالة إحصائيا تعزى للتفاعل بين نسب الفقرات ذات الأداء التفاضلي واختلاف طريقة الكشف عن الأداء التفاضلي.

وأما الدراسات التي تناولت المقارنة بين تقديرات القدرة ومعالم الفقرات وفق نماذج نظرية الاستجابة للفقرة المعلمية ونماذج نظرية الاستجابة للفقرة اللامعلمية، فقد أجرى كونينغ وسيجتسما وهامرز (Koning, Sijtsma & Hamers, 2002)) قارنت بين نموذجين معلميين ونموذجين آخرين لامعلميين من نماذج نظرية الاستجابة للفقرة للتعرف إلى فاعليتها في تحليل البيانات التجريبية للاختبار، ومقارنة نتائج نماذج نظرية الاستجابة للفقرة اللامعلمية مع نظيراتها نماذج نظرية الاستجابة للفقرة المعلمية، وأفضت نتائج الدراسة إلى الأفضلية في الجمع بين النوعين من النماذج المعلمية واللامعلمية، إذ أثبتت النماذج المعلمية فاعليتها في تقديم معلومات مفيدة حول معالم الفقرات، بالإضافة فاعليتها في بعض الجوانب التطبيقية كمعايرة درجات الاختبار والاختبارات التكيفية، فالنماذج المعلمية واللامعلمية قدمت معلومات متباينة باستخدام إحصائيات مختلفة، وفضلت الدراسة الجمع بينهما لتحسين نوعية الاختبار وجودة فقراته، وفي دراسة أوليفرس وجالاردو وكرامب (Olivares, Gallardo & Kramp, 2005) التي هدفت إلى إجراء مقارنة لمطابقة نماذج نظرية الاستجابة للفقرة المعلمية مع نظيرتها نماذج نظرية الاستجابة للفقرة اللامعلمية باستخدام النماذج التالية: نموذج التقدير الجزئي لماستر (Masters Partial Credit Model)، ونموذج ستينبرغ الذي يعد توسعا لنموذج ماسترز (Steinberg Extension of Masters Partial Credit Model)، ونموذج بوك الاسمي (Bock's Nominal Model)، ونموذج سيمجيما اللوجستي (Samejima's logistic model)، ونموذج ليفين اللامعلمي (Levine's Non-Parametric Model). أظهرت النتائج تفوق نموذج سيمجيما المعلمي على باقي النماذج المعلمية للعينتين في مطابقة بغض النظر عن تقنيات التقدير التي تم اعتمادها، وتفوق النموذج اللامعلمي في المطابقة لعينة المعايرة على باقي النماذج المعلمية، إلا أن النموذج المتدرج المعلمي فاقت مطابقته مطابقة النموذج اللامعلمي لعينة الصدق التقاطعي.

وقامت ديهاوس (Dyehouse, 2009) بمقارنة مطابقة البيانات باختلاف نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية، وقد بينت النتائج عدم تحقق افتراض الاطرادية المضاعفة لأي من الفقرات وهو الافتراض الأكثر صعوبة، في حين تحقق افتراض التجانس الاطرادي الأساسي لنموذج موكن لمعظم الفقرات ومجموعات الاضطراب مما يجعل من نموذج موكن نموذجا مفيدا لأغراض الاختيار ومطابقة بياناته، وفي السياق نفسه فقد أجرى لي وولاك ودوغلاس (Lee, Wollak & Douglas, 2009) دراسة هدفت إلى تقييم مطابقة البيانات للنموذج ثنائي المعالم من خلال منحنى خصائص الفقرة المعلمية مع إجراءات تقدير دالة خصائص الفقرة وفق النماذج اللامعلمية. أظهرت نتائج الدراسة بأن طرق التقدير اللامعلمية الثلاث أنتجت دوال استجابة فقرة مشابهة لتلك المقدرة وفق النموذج ثنائي المعلمة للفقرات التي تم توليدها لتطابق النموذج المعلمي، أما الفقرات التي لم تتشابه فيها دوال استجابة الفقرات، التي لم تتطابق وافتراضات النموذج اللامعلمي المستخدم، فقد كانت إجراءات تقدير (KR) هي الأقوى في الكشف عن عدم التطابق بين النموذجين ثنائي المعلمة والنموذج اللامعلمي.

وفي دراسة أجرتها المومني (2011) تهدف للكشف عن النموذج الأفضل في مطابقة فقرات الاختبار التحصيلي في مادة الرياضيات من نوع الاختيار من متعدد والمكون من (30) فقرة، من خلال مقارنة نتائج تحليل البيانات باستخدام النموذج ثنائي المعلمة مع نتائج تحليل البيانات باستخدام نموذج موكن للتجانس الاطرادي اللامعلمي، حيث بينت النتائج التي تم استخراجها باستخدام برنامج (Bilog-MG3) عدم فعالية للنموذج ثنائي المعلمة في مطابقة فقرات الاختبار. أما فيما يخص استخراج كمية معلومات الاختبار فقد أظهرت نتائج الرسم باستخدام برنامج (TESTGRAF) أن الاختبار بصورته الرتبية الأفضل في تقديم المعلومات حول السمة المقيسة، وهدفت دراسة الدلوع (2013) إلى بيان مدى التوافق بين نموذج ثلاثي المعلمة البارامتري

ونموذج موكن اللابارامتري لنوعين من فقرات اختبار تحصيلي هما (الاختيار من متعدد، صواب/خطأ). أشارت النتائج أن نموذج موكن أكثر تشدداً في افتراضاته الأساسية لمطابقة فقرات الاختبار، وأن النموذج ثلاثي المعلمة أكثر تشدداً في افتراضاته الأساسية لمطابقة عدد الأفراد، والنموذج الثلاثي المعلمة ذو اهتمام أكثر بقدرات الأفراد. كما بينت الدراسة أن كمية المعلومات التي يقدمها النموذج ثلاثي المعلمة أكبر من كمية المعلومات التي يقدمها نموذج موكن، وفي كلا الاختبارين.

وفي دراسة أجراها القيسي (2013) هدفت إلى مقارنة دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة المعلمية واللامعلمية باختلاف حجم العينة وطول الاختبار اعتمادا على مؤشري التحيز(BIAS)، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE) باستخدام بيانات مولدة. أشارت نتائج الدراسة إلى أن تقديرات معالم الفقرة والقدرة بالطريقة المعلمية أقل منها في الطريقة اللامعلمية في جميع حالات الدراسة تقريبا. وأما فيما يتعلق بمؤشر الدقة في القياس RMSE فقد أشارت النتائج إلى أن قيم الأوساط لطريقة التقدير المعلمية كانت اقل من الأوساط لطريقة التقدير اللامعلمية في كل حالات الدراسة، وقام مساعدة (2013) بدراسة هدفت إلى المقارنة بين ثلاثة أشكال من أشكال صياغة فقرات الاختبار (الاختيار من متعدد، الصواب والخطأ، التكميل) في مطابقة الفقرات لنموذج موكن اللامعلمي ونموذج ثنائي المعالم. أظهرت النتائج وجود فرق دال إحصائيًا بين متوسطات كمية المعلومات للاختبارات (الاختيار من متعدد، صواب وخطأ، تكميل) يعزى لنموذجي (موكن، ثنائي المَعلمة)، لصالح ما يقدمه نموذج موكن من كمية المعلومات لكلٍّ من الاختبارات (الاختيار من متعدد، صواب وخطأ، التكميل) مقارنة بما يقدمه النموذج ثُنائي المَعلَمة من دالة للمعلومات لكلٍّ من الاختبارات (الاختيار من متعدد، صواب وخطأ، تكميل).

في ضوء ما تقدم عرضه من الدراسات السابقة يمكن ملاحظة أن معظم الدراسات التي تم القيام بها قد بحثت في الكشف عن الفقرات ذات الأداء التفاضلي باستخدام طرق مختلفة ومتنوعة، لكن لم تسفر نتائج تلك الدراسات في تحديد الطريقة الفعالة رغم تعدد تلك الطرق؛ ويعود السبب في ذلك أن هذه الطرق تعتمد على: حجم العينة، ونوع الأداء التفاضلي، وطول الاختبار، ومصدر الأداء التفاضلي، ومستوى تمييز الفقرة وصعوبتها وغيرها. وبعض الدراسات ركز على أثر الفقرات ذات الأداء التفاضلي على دقة المعادلة (Zhang, Mattews & Dorans, 2003)، وبعضها ركز على أثر الفقرات التي تظهر أداء تفاضلي على إحصائيات النظرية الكلاسيكية للاختبار، مثل: معامل الثبات، الخطأ المعياري في القياس، ومعامل التصحيح منن أثر الانخفاض (Zumbo, 2000)، وبعضها درس أثر الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول، وحجم الأثر باستخدام اختبار (Zhan & Zumbo, 2009)، وقد تبين من نتائج تلك الدراسات أن بعض هذه الإحصائيات قد تأثرت بوجود الفقرات ذات الأداء التفاضلي، وبعضها الآخر لم يتأثر بوجود الفقرات ذات الأداء التفاضلي. ويتبين من هذه الدراسات أنها ركزت بشكل أساسي على دراسة الأثر للفقرات ذات الأداء التفاضلي باستخدام النظرية الكلاسيكية، وهناك دراسات بحثت في أثر نسب الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي في تقدير معالم الفقرات وقدرات الأفراد وفق النموذج اللوجستي ثلاثي المعالم (النوافلة، 2013)، أو في الكشف عن أثر نسبة الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول وقوة الاختبار باستخدام طريقتي الانحدار اللوجستي و نسبة الأرجحية (إجبارة، 2016)، وغالبية هذه الدراسات لم تهتم بنوع الأداء التفاضلي باستثناء دراسة جيرل (Gierl, 2004) التي اهتمت بنوع الأداء التفاضلي، وبالتالي لم تقدم إجابة حول أثر نوع الأداء التفاضلي على الخصائص السيكومترية للفقرات والاختبار كما أنه– في حدود اطلاع الباحث– لا توجد دراسات عربية حول الموضوع نفسه ، لذا فإن هذه الدراسة عالجت مثل هذا النوع من المشكلات البحثية، فإنه يؤمل من هذه الدراسة أن تضيف معلومات جديدة استكمالاً لما تم البحث فيه حول ذلك، لذا جاءت هذه الدراسة لتسهم في تقديم إضافة معرفية جديدة حول هذا الموضوع.

كما أن معظم الدراسات السابقة التي تناولت موضوع الأداء التفاضلي ركزت على نماذج استجابة الفقرة المعلمية رغم أن الدراسات التي تناولت المقارنة بين النماذج المعلمية واللامعلمية قد أظهرت تفاوتاً في نتائجها حول المطابقة، وتقدير المعالم الخاصة بالفقرات، والأفراد، وكمية المعلومات، فلكل نموذج افتراضاته وإحصاءاته الخاصة به، كما تتطلب النماذج المعلمية افتراضات أكثر يصعب تحقيقها مقارنة بنظيرتها اللامعلمية، مثل: وقوع البيانات على مقياس فئوي، وشكل دالة الفقرة، كذلك فإن نماذج نظرية استجابة الفقرة اللامعلمية تفترض أن البيانات واقعة بمستوى القياس الرتبي. لذا فإن هذه الدراسة ستتيح المجال لتوظيف نظرية الاستجابة للفقرة المعلمية واللامعلمية في تقدير خصائص الفقرات وقدرات الأفراد والمقارنة بين النماذج المعلمية والنماذج اللامعلمية.

أما عن سبب اختيار الباحث النموذج اللوجستي ثلاثي المعالم لتحقيق غرض هذه الدراسة، فهو أن هذا النموذج يعد من أكثر النماذج واقعية، كما أشارت لذلك معظم الدراسات السابقة التي بينت أن النموذج الثلاثي المعالم هو الطريقة الفضلى في تقدير

القدرة للفرد؛ نظراً لاحتوائه على معالم التخمين الخاصة بالأفراد ذوي القدرة المتدنية، كما أن هذا النموذج يشتمل إضافة إلى المعالم التي يشتملها النموذج ذو المعلمتين على معالم التخمين للفقرة، لتقدير احتمالية أن يجيب الفرد إجابة صحيحة عليها.

مشكلة الدراسة وأسئلتها

أجريت العديد من الدراسات للكشف عن الأداء التفاضلي، حيث أن الأداء التفاضلي يهدد صدق البناء، والثبات للاختبار، ودقة تقدير معالم الفقرات، ودقة تقديرات القدرة للأفراد، وبالتالي تلوث النتائج، مما يؤدي إلى صعوبة في اتخاذ القرارات التربوية في ضوء بيانات غير صادقة، كما أن معظم الدراسات ركزت على المقارنة بين الطرق المستخدمة للكشف عن الأداء التفاضلي للفقرة، وبينت نتائج معظم هذه الدراسات أن وجود الأداء التفاضلي في الفقرة يؤثر على تقديرات الصعوبة، والتمييز، والتخمين، حيث إن الفقرة التي تظهر أداء تفاضلياً تكون أصعب لمجموعة من مجموعة أخرى، وبالتالي، فإن وجود عدد من الفقرات ذات الأداء التفاضلي له تأثير على الخصائص السيكومترية للفقرات، والخصائص السيكومترية للاختبار، ولكن الدراسات لم تجب عن السؤال المحير، وهو: "ماذا لو كان الأداء التفاضلي لهذه الفقرات منتظما، أم غير منتظم، فهل سيكون له انعكاسات وتأثير على الخصائص السيكومترية للفقرات والاختبار؟"، الذي يتوقع أن تقدم هذه الدراسة إجابة له. كما سيتم استخدام النماذج المعلمية والنماذج اللامعلمية في نظرية الاستجابة للفقرة بدلاً من النماذج المعلمية فقط التي شاع استخدامها من قبل الباحثين على الرغم من موانعها في تحليل البيانات في المستوى الرتبي، إذ إن مصداقية النتائج قد تكون موضع تساؤل عندما لا يتحقق فرض وقوع البيانات على مستوى فئوي، الأمر الذي تبرره النماذج اللامعلمية التي لا تضع قيودا حول شكل دالة الاستجابة للفقرة؛ مما يثير التساؤل حول مدى مطابقة النوعين للبيانات التي يعتبرها بعضهم رتبية، وبعضهم الآخر شبه فئوية، حيث النماذج اللامعلمية تستطيع تحليل بيانات على مستوى القياس الرتبي بعكس النماذج المعلمية، التي تشترط وقوع البيانات على مستوى القياس الفئوي على الأقل، كما تتطلب النماذج المعلمية افتراضات أكثر يصعب تحقيقها مقارنة بنظيرتها اللامعلمية مثل وقوع البيانات على مقياس فئوي، وشكل دالة الفقرة. لذا، فإن هذه الدراسة ستتيح المجال لتوظيف نماذج نظرية الاستجابة للفقرة المعلمية، واللامعلمية في تقدير الخصائص السيكومترية للفقرات، وقدرات الأفراد، وبالتالي إتاحة الفرصة للمقارنة بينهما، وتحديداً تسعى هذه الدراسة للإجابة عن الأسئلة التالية:

السؤال الأول: هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار تعزى للنموذج نظرية الاستجابة للفقرة (المعلمية، واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة) والتفاعل بينها؟".

السؤال الثاني: هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار تعزى لنموذج نظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة) والتفاعل بينها؟".

السؤال الثالث: هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تعزى لنموذج نظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة) والتفاعل بينها؟".

السؤال الرابع: هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لكمية المعلومات التي يقدمها الاختبار تعزى لنموذج نظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم) والمجموعة (مرجعية، مستهدفة) والتفاعل بينها؟".

السؤال الخامس: هل تختلف معاملات الثبات للاختبار باختلاف النموذج لنظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة) ؟".

أهمية الدراسة

نظراً للأهمية البالغة للاختبارات النفسية والتربوية، التي تستخدم في تزويد صناع القرارات ببيانات مفيدة يمكن الاستناد إليها بقدر من الثقة في اتخاذ القرارات المتعلقة في انتقاء الأفراد وتصنيفهم أو تحويلهم للبرامج التربوية أو العلاجية أو التأهيلية، وغير ذلك الكثير. وجاءت هذه الدراسة استكمالاً لجهود الباحثين في الإجابة عن الاستفسارات المطروحة في هذه الدراسة، التي ستكشف عن الأساليب المناسبة في تحليل البيانات والتعامل معها، التي تشكل أهمية هذه الدراسة، حيث ستتمثل الأهمية في محورين أساسيين، أحدهما نظري والآخر تطبيقي. فيتمثل الجانب النظري في إجراء مقارنة للخصائص السيكومترية للفقرات والاختبار وفقًا لنوع الأداء

التفاضلي للفقرات (منتظم، غير منتظم) والنموذج المستخدم في تقدير تلك المعالم (المعلمي، واللامعلمي)، وبالتالي الكشف عن مدى تأثير نوع الأداء التفاضلي على دقة تقدير معالم الفقرات وقدرات الأفراد والخصائص السيكومترية للاختبار. وأما المحور التطبيقي أو العملي فيتمثل في الكشف عن المنهجية المناسبة لتطوير الاختبارات، والتحقق من خصائصها حيث توفر نموذج التحليل المناسب الذي يمكن لمطوري الاختبارات، ومراكز القياس والتقويم تبنيه من أجل تحسين نوعية الاختبارات وجودتها ودقة التقدير لقدرات الأفراد كما سيتضح النموذج المناسب في تحليل الفقرات عندما تظهر الفقرات أداء تفاضليا سواء كان منتظماً، أم غير منتظم.

التعريفات الاصطلاحية والإجرائية

نوع الأداء التفاضلي: نوع الأداء الذي تظهر فيه الفقرة دلالة مشتقة إحصائيا للتعبير عن الفروق في احتمالات الإجابة الصحيحة للفقرة بين مجموعتين ممن هم في مستوى القدرة نفسه، ويكون **منتظماً** إذا كان احتمال الإجابة الصحيحة للفقرة أكبر لمجموعة دون مجموعة أخرى عند جميع مستويات القدرة، **وغير منتظم** إذا كان احتمال الإجابة الصحيحة للفقرة متفاوتاً حيث يكون لصالح مجموعة عند مستوى قدرة معين، ولصالح المجموعة الأخرى عند مستوى قدرة آخر.

الخصائص السيكومترية للفقرات: تتمثل في تقدير معالم الصعوبة، ومعالم التمييز، ومعالم التخمين (الخط التقاربي الأدنى)، التي تم تقديرها بالاعتماد على نماذج نظرية الاستجابة للفقرة المستخدمة (المعلمية، واللامعلمية).

صعوبة الفقرة: خاصية تصف الفقرة وتأخذ قيماً نظرية تتراوح من (− ∞ إلى + ∞)، وعملياً من (− 3 إلى + 3)، وهي النقطة الواقعة على متصل القدرة التي تقابل احتمال إجابة صحيحة قدره (1+c)/2 في حالة استخدام النموذج اللوجستي ثلاثي المعلمة (حالة وجود تخمين) المستخدم في هذه الدراسة.

تمييز الفقرة: خاصية تصف الفقرة، وتأخذ قيماً نظرية تتراوح من (− ∞ إلى + ∞)، وعملياً من (− 3 إلى + 3)، ويستدل عليها من خلال شدة انحدار أو ميل المماس لمنحنى خصائص الفقرة(ICC) عند نقاط انعطاف منحنى الخصائص للفقرة.

تخمين الفقرة: خاصية تصف الفقرة، وتأخذ قيماً نظرية تتراوح بين الصفر والواحد صحيح (100%)، وتمثل احتمال الإجابة عن الفقرة إجابة صحيحة عن طريق التخمين، وتمثل تقاطع الخط التقاربي الأدنى (Lower Asymptote Line)، لمنحنى الخصائص للفقرة مع المحور العمودي الذي يمثل: احتمال أن يجيب الفرد ذو القدرة (θ) عن الفقرة (i) إجابة صحيحة.

الخصائص السيكومترية للاختبار: تتمثل في تقدير كمية المعلومات التي يقدمها الاختبار من خلال دالة المعلومات للاختبار، وتقدير الثبات للاختبار، التي تم تقديرها بالاعتماد على نماذج نظرية الاستجابة للفقرة المستخدمة (المعلمية، واللامعلمية).

نماذج نظرية الاستجابة للفقرة المعلمية: هي النماذج التي يمكن تطبيقها على فقرات ثنائية التدريج ذات شكل استجابة لوجستي محدد، وتفترض أن البيانات فيها واقعة بمستوى القياس الفئوي، ولا يوجد عدد محدد لهذه النماذج التي يمكن أن تتولد في إطار نظرية الاستجابة للفقرة، وفي هذه الدراسة تم استخدام النموذج اللوجستي ثلاثي المعلمة.

نماذج نظرية الاستجابة للفقرة اللامعلمية: هي النماذج التي يمكن تطبيقها على فقرات ثنائية التدريج ذات شكل استجابة لوجستي غير محدد، وتتميز بإمكانية استخدامها للبيانات الرتبية والفئوية، وفي هذه الدراسة تم استخدام نموذج موكن اللامعلمي.

دالة المعلومات للفقرة: يُعد من المفاهيم الأساسية في نظرية الاستجابة للفقرة سواء كانت الفقرة ثنائية التدريج، أو تتطلب استجابات متعددة التدريج، وتساعد هذه الدالة في بناء أدوات قياس تجعل دقة القياس أو المعلومات أكبر ما يمكن، وتبين مدى مساهمة الفقرة في دالة المعلومات للاختبار بشكل مستقل عن الفقرات الأخرى.

دالة المعلومات للاختبار: وهي دالة رياضية تعبر عن مجموع دوال المعلومات لجميع فقرات الاختبار عند مستوى معين من القدرة.

المجموعة المستهدفة (Focal Group): المجموعة التي تشتمل على فقرات ذات أداء تفاضلي (منتظم أو غير منتظم)، ويتم مقارنتها مع المجموعة المرجعية.

المجموعة المرجعية (Referenced Group): المجموعة التي لا تشتمل على أي فقرة ذات أداء تفاضلي، وهي المجموعة التي تقارن المجموعة المستهدفة معها.

متغيرات الدراسة

المتغيرات المستقلة

- نوع الأداء التفاضلي لفقرات الاختبار، وله مستويان: المنتظم، وغير المنتظم.
- نموذج الاستجابة للفقرة المستخدم، وله ثلاثة مستويات، هي: النموذج اللوجستي ثلاثي المعالم، ونموذج اللوجستي الثلاثي اللامعلمي.

- مجموعة الدراسة ولها مستويان: المجموعة المرجعية والمجموعة المستهدفة.

المتغيرات التابعة

تقديرات معالم الصعوبة والتمييز والتخمين باستخدام النموذج اللوجستي ثلاثي المعالم، ونموذج موكن اللامعلمي، ومعاملات الثبات، وكمية المعلومات المستخدمة في تقديرها وفقًا لنوع الأداء التفاضلي باستخدام النموذج اللوجستي ثلاثي المعالم، والنموذج الثلاثي اللامعلمي.

محددات الدراسة

اقتصرت الدراسة على: اعتماد النموذج اللوجستي الثلاثي المعالم، ونموذج موكن اللامعلمي من نماذج نظرية الاستجابة للفقرة، وكذلك على شكل واحد من أشكال فقرات الاختبار، وهي الفقرات ثنائية التدريج التي تأخذ القيم (0 ، 1)، بالإضافة إلى استخدام بيانات مولدة باستخدام برنامج WINGEN، اقتصرت الدراسة على نسبة واحدة للفقرات التي تظهر أداء تفاضلياً وهي (10%)، والمفاهيم والمصطلحات المستخدمة في هذه الدراسة، محددة في التعريفات الإجرائية، وبالتالي فإن إمكانية تعميم النتائج تتحدد في ضوء هذه التعريفات.

الطريقة والإجراءات

يتناول هذا الجزء وصفاً للإجراءات المستخدمة للإجابة عن أسئلة الدراسة من خلال الاعتماد على بيانات مولدة، حيث هدفت هذه الدراسة إلى الكشف عن أثر نوع الأداء التفاضلي للفقرات على الخصائص السيكومترية للفقرات والاختبار وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة، وتم استخدام بيانات مولدة وفق معايير محددة مسبقا لمعالم الفقرات، وأحجام العينات؛ لأن البيانات التي تم استخدامها في التحليل تحتاج لأن تكون الفقرات ذات الأداء التفاضلي معرفة مسبقاً ومحددة وفق نوعها لكي يتم دراسة أثرها على الخصائص السيكومترية للفقرات والاختبار. كما يبين وصفاً للمعالجات الإحصائية المستخدمة للإجابة عن أسئلة الدراسة وفق الإجراءات التالية:

توليد البيانات باستخدام برنامج (Wingen 3)

1. تم استخدام برنامج (WINGEN 3) لتوليد اختبار مكون من (50) فقرة، ومن هذا الاختبار تم تشكيل نموذجين يختلفان فقط في (10) فقرات، وأما باقي الفقرات (40) فقرة فهي نفسها في النموذجين، حيث كانت (10) فقرات ذات أداء تفاضلي منتظم في النموذج الأول، و(10) فقرات ذات أداء تفاضلي غير منتظم في النموذج الثاني، وفي الموقع نفسه، حيث كانت الفقرات جميعها ثنائية التدريج (0،1)، حيث تعطى الدرجة (صفر) للاستجابة الخاطئة، والدرجة (1) للاستجابة الصحيحة.
2. تم توليد فقرات الاختبار وفق النموذج ثنائي المعلمة حيث كانت معالم الصعوبة وفقًا للتوزيع الطبيعي بوسط مقداره صفر، وانحراف معياري مقداره 1، ومعالم التمييز وفقًا للتوزيع المنتظم بقيمة صغرى مقدارها 0.4، وقيمة عظمى مقدارها 1.6، ومعالم التخمين متكافئة كما يفترض النموذج، ولتشكيل النموذج الأول (الذي يتضمن 10 فقرات ذات أداء تفاضلي منتظم) تم ترتيب الفقرات وفقًا لمعالم الصعوبة تصاعديًا واختيار فقرات متباعدة بشكل منتظم عن بعضها، وهي الفقرات: (1، 5، 10، 15، 20، 25، 30، 35، 40، 45)، بما يشكل ما نسبته 20% من أصل 50 فقرة، مع مراعاة إضافة قيمة ثابتة (1) لمعلم الصعوبة للفقرات التي ستبدي أداءً تفاضليًا منتظماً في المجموعة المستهدفة، والجدول في الملحق أ يبين معالم الصعوبة والتمييز الخاصة في النموذج الأول (يتضمن 10 فقرات ذات أداء تفاضلي منتظم).
3. لتشكيل النموذج الثاني للاختبار (الذي يتضمن 10 فقرات ذات أداء تفاضلي غير منتظم)، وقد تم ترتيب الفقرات وفقًا لمعلمة التمييز تصاعديًا بشكل يوازي ترتيب الفقرات وفقًا لمعلمة الصعوبة؛ لاختيار فقرات متباعدة بشكل منتظم عن بعضها، وهي: (1، 5، 10، 15، 20، 25، 30، 35، 40، 45)، بما يشكل ما نسبته 20% من أصل 50 فقرة مزعومة، مع مراعاة إضافة قيمة (0.4) لمعلمة التمييز للفقرات التي ستبدي أداءً تفاضليًا غير منتظم في المجموعة المستهدفة. والجدول في الملحق أ يبين معالم الصعوبة والتمييز الخاصة في النموذج الثاني (يتضمن 10 فقرات ذات أداء تفاضلي غير منتظم).
4. من حيث تصميم البيانات المستخدم في هذه الدراسة، فإن عدد المجموعات المستهدفة والمرجعية وفق النماذج المعلمية والنماذج اللامعلمية لنظرية الاستجابة للفقرة، أربع مجموعات: مجموعتان (مرجعية، ومستهدفة)، في حالة الأداء التفاضلي المنتظم، ومجموعتان مرجعية ومستهدفة في حالة الأداء التفاضلي غير المنتظم.
5. تم تحديد عدد الفقرات ذات الأداء التفاضلي في مجموعات الدراسة، حيث تكون نسبة الفقرات ذات الأداء التفاضلي 20%؛ أي ما يعادل (10) فقرات من أصل (50) فقرة، وفي الموقع نفسه من حيث ترتيب الفقرات.

6. تم توليد استجابات (1000) فرد؛ وذلك لأن الطرق القائمة على نظرية استجابة الفقرة تحتاج لأن تكون أعداد المفحوصين كبيرة، بتوزيع طبيعي للقدرة وسطه الحسابي (0)، وانحرافه المعياري (1) ، وللمجموعتين المرجعية والمستهدفة.

التحقق من افتراضي أحادية البعد والاستقلال الموضعي

- تم التحقق من افتراض أحادية البعد Unidimensionality لفقرات الاختبار وفقاً للنموذج ثلاثي المعالم في نظرية الاستجابة للفقرة المعلمية كافتراض أساسي من افتراضات نظرية الاستجابة للفقرة، وتمّ التحقق من افتراض أحادية البعد باستخدام برنامج NOHARM (Normal Ogive Harmonic Analysis Robust Method) لمعالجة بيانات عينة الدراسة ، والمتعلقة باستجابات (1000) فرد، وذلك كما هو موضح في الجدول (1).

الجدول1. مؤشرا التحقق من أحادية البعد وفق النموذج ثلاثي المعالم

مجموع مربعات البواقي (دون القطر الرئيسي)		0.00547
الجذر التربيعي لمتوسط مجموع مربعات البواقي (دون القطر الرئيسي)	**قيمته**	0.00406
	المعيار	**0.129653**
مؤشر Tanaka لحسن المطابقة		0.976542

يتضح من الجدول (3) أنه للكشف عن أحادية البعد تم استخدام مؤشرين؛ هما: مؤشر تناكا (TANAKA) لحسن المطابقة، حيث زادت قيمته عن (0.95)، وتقترب من الواحد الصحيح ، وهي قيمة تشير إلى تحقق افتراض أحادية البعد وفقاً لما أشار إليه جاسبر (Jaspar, 2010)، والمؤشر الآخر؛ هو مؤشر الجذر التربيعي لمتوسط مربعات البواقي (RMSR)، حيث بلغت قيمته المحسوبة (0.00406)، وهي أقل من القيمة الحرجة له البالغة قيمتها (0.129653) التي تحسب من خلال المعادلة ((4.1/sqrt(n))، حيث (n =1000)، وعند مقارنة القيمة الناتجة من خلال التحليل مع أعلى قيمة لهذا المؤشر تبين أن القيمة الناتجة تقترب من الصفر، وهذا مؤشر على تحقق افتراض أحادية البعد (Jasper, 2010).

- التحقق من افتراض الاستقلال الموضعي Local Independence وفقاً للنموذج ثلاثي المعالم في نظرية الاستجابة للفقرة المعلمية كافتراض أساسي من افتراضات نظرية الاستجابة للفقرة، ولأغراض التحقق من افتراض الاستقلال الموضعي؛ تم استخدام برنامج (LDID: Local Dependence Indices for Dichotomous Items) على بيانات عينة الدراسة للنموذج ثلاثي المعالم95، يتم من خلاله حساب: Q3، Z_{Q3}، G^2. حيث يتم اعتماد المؤشر Z_{Q3} كحد فاصل بين أزواج الفقرات التي بينها انتهاك لافتراض الاستقلال الموضعي عن غيرها، فإذا كان المتوسط الحسابي لمؤشرات Z_{Q3} الملاحظ الخاص بأزواج الفقرات واقعا بين الحد الأدنى لمؤشر Z_{Q3} الخاص بأزواج الفقرات والحد الأعلى لمؤشر Z_{Q3} الخاص بأزواج الفقرات، فهذا مؤشر على تحقق افتراض الاستقلال الموضعي (Kim, Cohen & Lin, 2005)، حيث تبين أن قيمة الوسط الحسابي لمؤشر Z_{Q3} الخاصة بأزواج الفقرات التي تقع بين الحد الأدنى وبين الحد الأعلى لمؤشر Z_{Q3} الخاصة بأزواج الفقرات البالغ عددها (1035) ويبين الجدول (2) مؤشرات الاستقلال الموضعي وفقًا لنظرية السمات الكامنة.

الجدول 2. مؤشرات الاستقلال الموضعية وفقًا لنظرية السمات الكامنة

الإحصائي	القيمة
عدد الفقرات	50
عدد أزواج الفقرات	1225
القيمة الصغرى الملاحظة لمؤشر Z_{Q3}	− 0.453783
القيمة العظمى الملاحظة لمؤشر Z_{Q3}	0.8421350
المتوسط الحسابي لمؤشرات Z_{Q3} الملاحظة الخاص بأزواج الفقرات	0.0287918

يلاحظ من الجدول (2) أن قيمة المتوسط الحسابي لمؤشر Z_{Q3} الخاصة بأزواج الفقرات قد بلغ (0.0287918)، وهي واقعة بين الحد الأدنى والحد الأعلى لمؤشر Z_{Q3} الخاصة بأزواج الفقرات، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي (Kim, Cohen & Lin, 2005). كما تم حساب عدد أزواج الفقرات التي وقعت ضمن فترة الثقة المحققة لشروط الاستقلال الموضعي، والجدول (3) يبين مؤشرات الاستقلال الموضعي وفقاً لنظرية السمات الكامنة.

الجدول 3. مؤشرات الاستقلال الموضعي وفقاً لنظرية السمات الكامنة

حالة الاستقلال الموضعي	عدد الأزواج	النسبة المئوية
التبعية الموضعية	180	%14.7
الاستقلالية الموضعية	1045	%85.3
الكلي	1225	%100

يلاحظ من الجدول (3) أن عدد أزواج الفقرات التي وقعت ضمن حالة الاستقلالية الموضعية قد بلغت (1045) حالة بنسبة مئوية بلغت (85.3%)، في حين أن عدد الأزواج التي وقعت خارج حالة الاستقلالية الموضعية (التبعية الموضوعية) (180) وبنسبة مئوية بلغت (14.7%)، وهذا يبين أن عدد أزواج الفقرات التي حققت الاستقلالية أعلى من أربعة أمثال عدد أزواج الفقرات التي حققت التبعية الموضعية. وهذا مؤشر على تحقق افتراض الاستقلال الموضعي (Kim, Cohen & Lin, 2005)، بالإضافة لذلك فقد أشار هامبلتون وسوامنيان (Hambelton & Swaminathan, 1985) إلى أن افتراض الاستقلال الموضعي يكافئ افتراض أحادية البعد، وهذا يعني انه إذا تحقق افتراض أحادية البعد في المقياس، فإن المقياس يحقق افتراض الاستقلال الموضعي.

نتائج الدراسة ومناقشتها

أوّلًا. النتائج المتعلقة بالسؤال الأول الذي ينصَّ على: "هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التمييز لفقرات اختبار تعزى لنموذج نظرية الاستجابة للفقرة (النموذج اللوجستي ثلاثي المعالم، والنموذج اللوجستي الثلاثي اللامعلمي)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة)؟".

للإجابة عن هذا السؤال، فقد تم حساب الأوساط الحسابية والانحرافات المعيارية لمعالم التمييز لفقرات الاختبار وفقاً لنماذج نظرية الاستجابة للفقرة (المعلمية، اللامعلمية)، ولنوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة)، وذلك كما في الجدول (4).

جدول 4: الأوساط الحسابية والانحرافات المعيارية لمعالم التمييز لفقرات الاختبار وفقًا لنماذج نظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

نوع الأداء التفاضلي	المجموعة	العدد	معالم التمييز في نظريتي الاستجابة للفقرة:				الكلي	
			المعلمية		اللامعلمية			
			الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
منتظم	مرجعية	50	1.12	0.45	0.63	0.19	0.87	0.32
	مستهدفة	50	1.22	0.47	0.61	0.19	0.92	0.33
	الكلي	100	1.17	0.46	0.62	0.19	0.90	0.33
غير منتظم	مرجعية	50	1.09	0.46	0.61	0.18	0.85	0.32
	مستهدفة	50	1.24	0.57	0.73	0.28	0.99	0.43
	الكلي	100	1.17	0.52	0.67	0.24	0.92	0.38
الكلي	مرجعية	100	1.10	0.45	0.62	0.19	0.86	0.32
	مستهدفة	100	1.23	0.52	0.67	0.25	0.95	0.38
الكلي		200	1.17	0.49	0.65	0.22		

يلاحظ من الجدول (4) وجود فروق ظاهرية بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار وفقًا لنماذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة. وللكشف عن الدلالة الإحصائية لتلك الفروق؛ تم إجراء تحليل التباين الثلاثي للقياسات المتكررة بتكرار القياس على أحد العوامل، وهو نموذج النظرية (المعلمي، واللامعلمي) للأوساط الحسابية لمعالم التمييز لفقرات الاختبار، والجدول (5) يبين نتائج التحليل.

جدول 5: نتائج تحليل التباين للقياسات المتكررة الأوساط الحسابية لمعالم التمييز لفقرات الاختبار وفقًا لنماذج نظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

آثار الاختبارات لـ:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
داخل الفقرات	النموذج	27.08	1	27.08	**543.61**	0.00
تحققت الكروية باستخدام Mauchly's Test of Sphericity	النموذج× نوع الأداء التفاضلي	0.07	1	0.07	1.46	0.23
	النموذج× المجموعة	0.15	1	0.15	2.92	0.09
	النموذج ×نوع الأداء التفاضلي× المجموعة	0.04	1	0.04	0.90	0.34
	الخطأ	9.76	196	0.05		
	الكلي	37.11	200	0.19		
بين الفقرات	نوع الأداء التفاضلي	0.06	1	0.06	0.24	0.63
	المجموعة	0.81	1	0.81	3.42	0.07
	نوع الأداء التفاضلي ×المجموعة	0.22	1	0.22	0.92	0.34
	الخطأ	46.50	196	0.24		
	الكلي	47.58	199	0.24		
الكلي		84.69	399	0.21		

يتضح من الجدول (5) وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لمعالم التمييز لفقرات الاختبار يُعزى للنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج اللوجستي الثلاثي المعلمي مقارناً بالنموذج اللوجستي الثلاثي اللامعلمي. وربما تعزى النتيجة إلى أن النماذج اللوجستية المعلمية، وإن كانت حالة خاصة من نظرية استجابة الفقرة اللامعلمية، إلا أنها تراعي بنية الفقرات في الاختبار، لذلك استطاع النموذج اللوجستي الثلاثي المعلمة أن يشخص معلمة التمييز (a) بوضوح مقارنة بالنموذج اللامعلمي الثلاثي الذي يعمل على مستوى الفقرة دون تكامليتها مع بنية الاختبار؛ أي أن النتائج هنا أظهرت تفوق النماذج المعلمية على نظيرتها النماذج اللامعلمية في تقدير معلمة تمييز الفقرات. وربما يعزى السبب إلى الأساس الرياضي في طرق التقدير لمعلمة التمييز، حيث إن طريقة التقدير اللامعلمية لمعلمة التمييز التي تقوم بتحويل الدرجات الخام إلى رتب، ومن ثم تقوم برسم منحنى خصائص الفقرة في ضوء استجابات الأفراد، ومن ثم تقدر معلمة التمييز (a). وهذا التحويل من مستوى القياس شبه الفئوي إلى مستوى القياس الرتبي يؤدي إلى فقد المعلومات؛ مما يؤدي إلى انخفاض تقديرات معلمة التمييز، وعلى العكس من ذلك في طرق التقدير المعلمية، حيث إنها تقدر معلمة التمييز (a) مباشرة.

ويتبين من الجدول (5) عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار تُعزى للتفاعلات الثنائية بين نموذج نظرية الاستجابة للفقرة المستخدم من جهة، وكلٍ من: نوع الأداء التفاضلي، والمجموعة من جهة أخرى، وللتفاعل الثلاثي بين النموذج، ونوع الأداء التفاضلي والمجموعة. كما يتبين عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار تعزى لنوع الأداء التفاضلي، وكذلك عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة α=0.05 بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار تعزى

للمجموعة، وكذلك عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار تعزى للتفاعل بين نوع الأداء التفاضلي والمجموعة.

ثانيًا. النتائج المتعلقة بالسؤال الثاني للدراسة الذي ينصَّ على: "هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار تعزى لنموذج نظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة)؟ للإجابة عن هذا السؤال، فقد تم حساب الأوساط الحسابية والانحرافات المعيارية لمعالم الصعوبة لفقرات الاختبار وفقًا لنموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة، والجدول (6) يبين ذلك.

جدول 6: الأوساط الحسابية والانحرافات المعيارية لمعالم الصعوبة لفقرات الاختبار وفقًا لنماذج نظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

نوع الأداء التفاضلي	المجموعة	العدد	معالم الصعوبة في نظريتي الاستجابة للفقرة:				الكلي	
			المعلمية		اللامعلمية			
			الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
منتظم	مرجعية	50	0.14	0.99	0.42	0.95	0.28	0.97
	مستهدفة	50	0.39	0.98	0.65	1.11	0.52	1.05
	الكلي	100	0.27	0.99	0.53	1.03	0.40	1.01
غير منتظم	مرجعية	50	0.14	0.96	0.71	0.60	0.42	0.78
	مستهدفة	50	0.04	1.08	0.49	0.74	0.26	0.91
	الكلي	100	0.09	1.02	0.60	0.68	0.34	0.85
الكلي	مرجعية	100	0.14	0.97	0.56	0.80	0.35	0.89
	مستهدفة	100	0.21	1.04	0.57	0.94	0.39	0.99
الكلي		200	0.18	1.01	0.56	0.87		

يلاحظ من الجدول (6) وجود فروق ظاهرية بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار وفقًا لنماذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة. وللكشف عن الدلالة الإحصائية لتلك الفروق؛ تم إجراء تحليل التباين الثلاثي للقياسات المتكررة بتكرار القياس على أحد العوامل، وهو نموذج النظرية (المعلمي، واللامعلمي) للأوساط الحسابية لمعالم التمييز لفقرات الاختبار، والجدول (7) يبين نتائج التحليل.

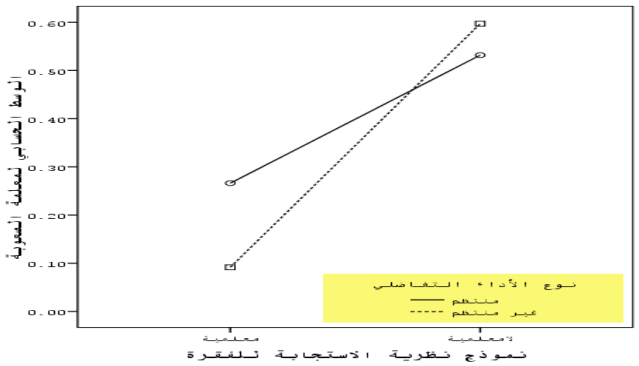
جدول 7: نتائج تحليل التباين للقياسات المتكررة للأوساط الحسابية لمعالم الصعوبة لفقرات اختبار وفقًا لنظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
داخل الفقرات	النموذج	14.87	1	14.87	**93.03**	0.00
تحققت الكروية باستخدام Mauchly's Test of Sphericity	النموذج ×نوع الأداء التفاضلي	1.44	1	1.44	**9.00**	0.00
	النموذج ×المجموعة	0.11	1	0.11	0.71	0.40
	النموذج ×نوع الأداء التفاضلي× المجموعة	0.06	1	0.06	0.40	0.53

	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
	الخطأ	31.33	196	0.16		
	الكلي	47.81	200	0.24		
بين الفقرات	نوع الأداء التفاضلي	0.30	1	0.30	0.18	0.67
	المجموعة	0.14	1	0.14	0.09	0.77
	نوع الأداء التفاضلي ×المجموعة	3.89	1	3.89	2.41	0.12
	الخطأ	316.05	196	1.61		
	الكلي	320.38	199	1.61		
الكلي		368.19	399	0.92		

يتضح من الجدول (7) وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لمعالم الصعوبة لفقرات الاختبار يُعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج اللامعلمي مقارناً بالنموذج المعلمي؛ وهذا يشير إلى أن النموذج اللامعلمي يظهر أن الفقرات أصعب مما هي عليه. وتعود النتيجة إلى أن النموذج اللامعلمي الثلاثي يختلف اختلافاً جوهرياً في تقدير معلمة الصعوبة لفقرات الاختبار عن النموذج الثلاثي المعلمة، حيث إن النموذج اللامعلمي الثلاثي يعمل بالمنطق الرتبي؛ أي يعتمد على حساب تكرارات الأخطاء المشاهدة لأزواج فقرات الاختبار وفقاً لأسلوب جتمان؛ الذي يشترط إجراء تقاطع بين كل زوج من أزواج الفقرات، حيث يتم رصد عدد التكرارات الملاحظة للإجابات الصحيحة للفقرة الأصعب، لذلك يعد النموذج الثلاثي اللامعلمي أكثر تشدداً في تقدير معلم صعوبة الفقرات مقارنة بالنماذج اللوجستية المرنة في تقدير معالم الفقرات.

كما يتضح من الجدول (7) وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة المستخدم، ونوع الأداء التفاضلي، والشكل (1) يوضح التفاعل الثنائي الرتبي بين نموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي لمعالم الصعوبة لفقرات الاختبار.



الشكل 1: التفاعل الثنائي اللارتبي بين نموذج نظرية الاستجابة للفقرة ونوع الأداء التفاضلي لمعالم الصعوبة لفقرات الاختبار.

يلاحظ من الشكل (1) وجود تفاعل لا رتبي بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار، حيث كان الوسط الحسابي لمعالم الصعوبة لفقرات الاختبار الذي يتضمن فقرات من نوع الأداء التفاضلي المنتظم أعلى عندما كان نموذج نظرية الاستجابة للفقرة المستخدم في التحليل معلمي، بينما كان الوسط الحسابي لمعالم الصعوبة لفقرات الاختبار الذي يتضمن فقرات من نوع الأداء التفاضلي غير المنتظم أعلى عندما كان نموذج نظرية الاستجابة للفقرة المستخدم في التحليل لامعلمي؛ فالنموذج اللوجستي الثلاثي اللامعلمي يعمل بالمنطق الرتبي، لذلك يعد النموذج الثلاثي اللامعلمي أكثر تشدداً في تقدير معلم صعوبة الفقرات مقارنة بالنماذج اللوجستية المعلمية المرنة في تقدير معالم الفقرات سواء كان الاختبار يتضمن فقرات ذات أداء منتظم، أم

فقرات ليست ذات أداء منتظم، في حين أن النماذج المعلمية تظهر التفوق لفقرات الاختبار التي أداؤها التفاضلي منتظم، والتفوق للنماذج اللامعلمية عندما كان الأداء التفاضلي غير منتظم. ويعزى ذلك إلى أن التفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد؛ أي نمط الفروق في احتمالية الاستجابة الصحيحة ليست متشابهة في كل مستويات القدرة، فهذه الفروق قد تكون لصالح إحدى المجموعات عند مستوى قدرة معين، ولصالح مجموعة أخرى عند مستوى قدرة آخر؛ مما يعمل على المبالغة في قيم الصعوبة وفقًا لنوع الأداء التفاضلي.

في حين يتضح من الجدول (7) عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار تُعزى للتفاعل الثنائي بين نموذج نظرية الاستجابة للفقرة المستخدم، والمجموعة، وللتفاعل الثنائي بين نموذج نظرية الاستجابة للفقرة والمجموعة، وعدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار تُعزى للأداء التفاضلي والمجموعة والتفاعل بينهما.

ثالثًا. النتائج المتعلقة بالسؤال الثالث للدراسة الذي ينصَّ على: "هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تعزى لنموذج نظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة)؟ للإجابة عن هذا السؤال، فقد تم حساب الأوساط الحسابية والانحرافات المعيارية لمعالم التخمين لفقرات الاختبار وفقًا لنموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة، والجدول (8) يبين ذلك.

جدول 8: الأوساط الحسابية والانحرافات المعيارية لمعالم التخمين لفقرات الاختبار وفقًا لنموذج نظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

نوع الأداء التفاضلي	المجموعة	العدد	معالم التخمين وفق نموذج نظرية الاستجابة للفقرة:				الكلي	
			المعلمية		اللامعلمية			
			الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
منتظم	مرجعية	50	0.05	0.02	0.13	0.13	0.09	0.07
	مستهدفة	50	0.08	0.05	0.12	0.12	0.10	0.09
	الكلي	100	0.06	0.04	0.12	0.13	0.10	0.08
غير منتظم	مرجعية	50	0.69	0.15	0.16	0.15	0.42	0.16
	مستهدفة	50	0.01	0.01	0.17	0.12	0.07	0.09
	الكلي	100	0.35	0.36	0.16	0.14	0.24	0.26
الكلي	مرجعية	100	0.37	0.34	0.14	0.14	0.25	0.24
	مستهدفة	100	0.04	0.05	0.15	0.12	0.08	0.10
الكلي		200	0.21	0.29	0.15	0.13		

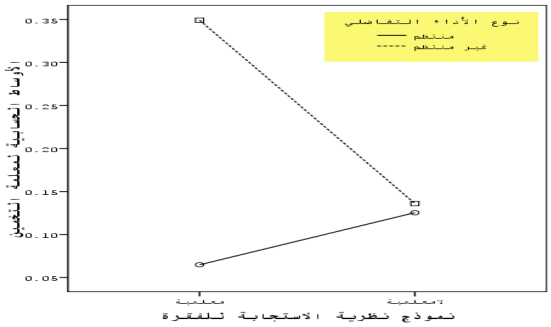
يلاحظ من الجدول (8) وجود فروق ظاهرية بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار وفقًا لنماذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة. وللكشف عن الدلالة الإحصائية للفروق بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار،؛ تم إجراء تحليل التباين الثلاثي للقياسات المتكررة بتكرار القياس على أحد العوامل، وهو نموذج النظرية (معلمي، واللامعلمي) والجدول (9) يبين نتائج التحليل.

جدول 9: نتائج تحليل التباين للقياسات المتكررة للأوساط الحسابية لمعالم التخمين لفقرات الاختبار وفقًا لنظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

آثار الاختبارات لـ:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
داخل الفقرات	النموذج	0.58	1	0.58	**32.96**	0.00
تحققت الكروية باستخدام Mauchly's Test of Sphericity	النموذج ×نوع الأداء التفاضلي	1.87	1	1.87	**106.08**	0.00
	النموذج ×المجموعة	2.35	1	2.35	**133.12**	0.00
	النموذج ×نوع الأداء التفاضلي× المجموعة	3.01	1	3.01	**170.12**	0.00
	الخطأ (معالم التخمين في النظريتين)	3.46	196	0.02		
	الكلي	11.28	200	0.06		
بين الفقرات	نوع الأداء التفاضلي	2.18	1	2.18	**211.68**	0.00
	المجموعة	2.92	1	2.92	**283.47**	0.00
	نوع الأداء التفاضلي ×المجموعة	3.35	1	3.35	**325.38**	0.00
	الخطأ	2.02	196	0.01		
	الكلي	10.46	199	0.05		
الكلي		21.74	399	0.05		

يتضح من الجدول (9) وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لمعالم التخمين لفقرات الاختبار يُعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج المعلمي مقارناً بالنموذج اللامعلمي، وهذا يعني أن تقديرات التخمين في النماذج المعلمية هي الأعلى مقارنة بالنماذج اللامعلمية، حيث يمثل التخمين احتمال الإجابة الصحيحة للأفراد من ذوي القدرة المنخفضة، حيث تزداد فرص الكشف عن التخمين العشوائي كمتغير مؤثر في الأداء في الاختبار في النموذج المعلمي عنه في النموذج اللامعلمي.

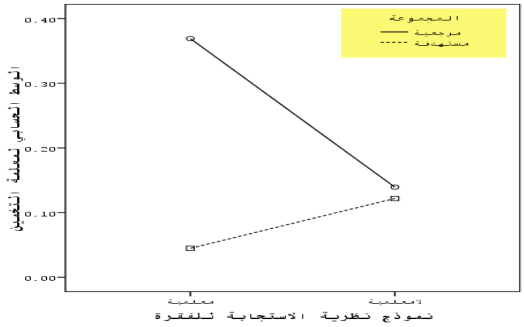
كما يتضح من الجدول (9) وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والشكل (2) يوضح التفاعل الثنائي الرتبي بين نموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي على الأوساط الحسابية لمعالم التخمين لفقرات الاختبار.



الشكل 2: التفاعل الثنائي الرتبي وفقًا لنموذج الاستجابة للفقرة ونوع الأداء التفاضلي لمعالم التخمين لفقرات الاختبار.

يلاحظ من الشكل (2) وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تُعزى لتفاعل بين نموذج نظرية الاستجابة للفقرة، والمجموعة، ونوع الأداء التفاضلي، وكان الفارق بين الوسطين

الحسابيين لتقديري معالم التخمين في نظرية الاستجابة للفقرة اللامعلمية أقل ما يمكن بين نوعي الأداء التفاضلي (منتظم، وغير منتظم). وعلى العكس من ذلك كان الفارق بين الوسطين الحسابيين لتقديري معالم التخمين في نموذج نظرية الاستجابة للفقرة المعلمية أكبر ما يمكن بين نوعي الأداء التفاضلي (منتظم، وغير منتظم)، وهذا يعني أن هناك فارقا كبيرا في التقدير للتخمين بين الفقرات ذات الأداء التفاضلي عندما كان منتظماً عنه عندما كان غير منتظم. والعكس من ذلك في حالة النموذج اللامعلمي الذي قلص الفارق لأدنى حد ممكن، وهذا يعني تفاوت في تقدير التخمين في حالة النماذج المعلمية، ولكن ذلك يتقلص بشكل ملحوظ في النماذج اللامعلمية. كما يتضح من الجدول (9) وجود فروق دالة إحصائيًا عند مستوى الدلالة α=0.05 بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة والمجموعة، والشكل (3) يوضح التفاعل الثنائي الرتبي لنموذج نظرية الاستجابة للفقرة والمجموعة على الأوساط الحسابية لمعالم التخمين لفقرات الاختبار، حيث يكون الفارق بين الوسطين الحسابيين لتقديري معالم التخمين في نظرية الاستجابة للفقرة اللامعلمية أقل ما يمكن بين نوعي الأداء التفاضلي (منتظم، وغير منتظم). وعلى العكس من ذلك يكون الفارق بين الوسطين الحسابيين لتقديري معالم التخمين في نموذج نظرية الاستجابة للفقرة المعلمية أكبر ما يمكن بين نوعي الأداء التفاضلي (منتظم، وغير منتظم)، وهذا يعني أن هناك فارقا كبيرا في التقدير للتخمين بين الفقرات ذات الأداء التفاضلي عندما كان منتظماً عنه عندما كان غير منتظم، والعكس من ذلك في حالة النموذج اللامعلمي الذي قلص الفارق لأدنى حد ممكن.



الشكل 3: التفاعل الثنائي الرتبي نظرية الاستجابة للفقرة والمجموعة لمعالم التخمين لفقرات الاختبار.

يلاحظ من الشكل (3) أن هناك تفاعلا رتبيا بين نموذج النظرية المستخدم والمجموعة، فالمجموعة المرجعية يكون الوسط الحسابي لمعلمة التخمين هو الأعلى مقارناً بالمجموعة المستهدفة سواء كان النموذج اللوجستي المستخدم معلميا أم لا معلميا. وحيث يكون نموذج نظرية الاستجابة للفقرة لامعلميا يكون الفارق بين الوسطين الحسابيين لتقديري معالم التخمين أقل ما يمكن بين المجموعتين (مرجعية، ومستهدفة)، وعلى العكس من ذلك يكون الفارق بين الوسطين الحسابيين لتقديري معالم التخمين في نموذج نظرية الاستجابة للفقرة المعلمية أكبر ما يمكن بين المجموعتين (مرجعية، ومستهدفة)، فالنماذج اللامعلمية لا تبالغ في تقدير معالم التخمين بين المجموعتين المرجعية والمستهدفة، بينما تظهر المبالغة في تقديرها عندما تكون النماذج معلمية، لذا ينصح باستخدام النماذج اللامعلمية في تقدير فقرات الاختبار تجنباً لتأثير الأداء التفاضلي في تقديرها. كما يتضح من الجدول (9) وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة، والمجموعة، ونوع الأداء التفاضلي.

رابعًا. النتائج المتعلقة بالسؤال الرابع للدراسة الذي ينصَّ على: "هل توجد فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لكمية المعلومات التي يقدمها الاختبار يعزى لنموذج نظرية الاستجابة للفقرة (المعلمية واللامعلمية)، ونوع الأداء التفاضلي (منتظم، غير منتظم)، والمجموعة (مرجعية، مستهدفة) والتفاعل بينها؟ للإجابة عن هذا السؤال، فقد تم حساب الأوساط الحسابية والانحرافات المعيارية لكمية المعلومات للاختبار وفقًا لنموذجي الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة، والجدول (10) يبين ذلك.

جدول 10: الأوساط الحسابية والانحرافات المعيارية لكمية المعلومات للاختبار وفقًا لنموذج نظرية الاستجابة للفقرة ونوع الأداء التفاضلي والمجموعة.

نوع الأداء التفاضلي	المجموعة	العدد	دالة معلومات الاختبار لنظرية الاستجابة للفقرة:				الكلي	
			المعلمية		اللامعلمية			
			الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
منتظم	مرجعية	61	3.89	2.55	16.03	13.70	9.96	8.13
	مستهدفة	61	4.12	2.94	15.72	13.10	9.92	8.02
	الكلي	122	4.01	2.74	15.87	13.35	9.94	8.05
غير منتظم	مرجعية	61	3.79	2.94	15.42	13.22	9.60	8.08
	مستهدفة	61	4.55	3.27	17.43	15.60	10.99	9.43
	الكلي	122	4.17	3.12	16.43	14.44	10.30	8.78
الكلي	مرجعية	122	3.84	2.74	15.72	13.41	9.78	8.08
	مستهدفة	122	4.34	3.10	16.58	14.37	10.46	8.74
الكلي		244	4.09	2.93	16.15	13.88		

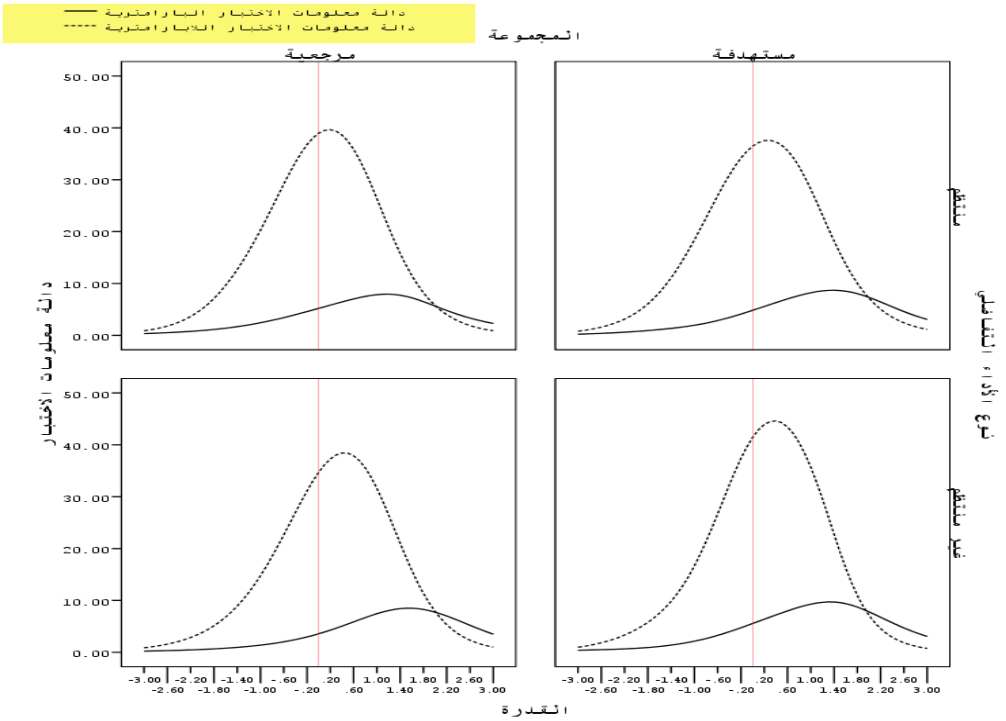
يلاحظ من الجدول (10) وجود فروق ظاهرية بين الأوساط الحسابية لكمية المعلومات للاختبار وفقًا لنموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة. وبهدف التحقق من الدلالة الإحصائية لتلك الفروق؛ تم إجراء تحليل التباين الثلاثي للقياسات المتكررة بقياسات متكررة على أحد العوامل وهو نموذج نظرية الاستجابة للفقرة (المعلمي، واللامعلمي) للأوساط الحسابية لكمية المعلومات للاختبار وفقًا لنموذج نظرية الاستجابة للفقرة، ونوع الأداء التفاضلي، والمجموعة، والجدول (10) يبين نتائج التحليل.

جدول 10: نتائج تحليل التباين الثلاثي للقياسات المتكررة لكمية المعلومات للاختبار وفقًا للنموذج ونوع الأداء التفاضلي والمجموعة.

آثار الاختبارات لـ:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
داخل الفقرات تحققت الكروية باستخدام Mauchly's Test of Sphericity	النموذج	17745.40	1	17745.40	**217.96**	0.00
	النموذج ×نوع الأداء التفاضلي	4.60	1	4.60	0.06	0.81
	النموذج ×المجموعة	3.86	1	3.86	0.05	0.83
	النموذج ×نوع الأداء التفاضلي ×المجموعة	24.40	1	24.40	0.30	0.58
	الخطأ	19539.73	240	81.42		
	الكلي	37317.99	244	152.94		
بين الفقرات	نوع الأداء التفاضلي	15.56	1	15.56	0.13	0.72
	المجموعة	56.11	1	56.11	0.46	0.50
	نوع الأداء التفاضلي× المجموعة	62.11	1	62.11	0.51	0.48
	الخطأ	29185.05	240	121.60		
	الكلي	29318.83	243	120.65		
الكلي		66636.81	487	136.83		

يتضح من الجدول (10) وجود فرق دال إحصائيًا عند مستوى الدلالة (α=0.05) بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى لنموذج نظرية الاستجابة للفقرة؛ لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي. كما يتبين من الجدول (10) عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لكمية المعلومات للاختبار تُعزى للتفاعل بين النموذج المستخدم، ونوع الأداء التفاضلي، وللتفاعل بين النموذج المستخدم والمجموعة، وللتفاعل بين نوع الأداء التفاضلي والمجموعة. وكذلك يتضح من الجدول (10) عدم وجود فروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية لكمية المعلومات للاختبار تُعزى للتفاعل بين النموذج المستخدم، ونوع الأداء التفاضلي، والمجموعة.

ولتوضيح فيما إذا كانت دالة معلومات للاختبار تختلف باختلاف النموذج المستخدم (معلمي واللامعلمي)، ونوع الأداء التفاضلي (منتظم، وغير منتظم)، والمجموعة (المستهدفة، والمرجعية) فقد تم رسم منحنيات دالة المعلومات للاختبار وفقًا للنموذج (النموذج ثلاثي المعالم والنموذج الثلاثي اللامعلمي)، والمجموعة (مرجعية ومستهدفة)، ونوع الأداء التفاضلي (منتظم، وغير منتظم)، والشكل (7) يبين ذلك.



الشكل 7. رسم بياني لدوال معلومات الاختبار وفقًا للنموذج والمجموعة ونوع الأداء التفاضلي.

يلاحظ من خلال دوال المعلومات على مستوى الاختبار عدم وجود توافق بين شكل دالة المعلومات للنموذجين المعلمي واللامعلمي سواء كان نوع الأداء التفاضلي منتظماً أم غير منتظم، وسواء كانت المجموعة مرجعية أم مستهدفة، ويلاحظ من الشكل (7) القيمة القصوى لكمية المعلومات التي يقدمها الاختبار كانت وفق النموذج اللامعلمي، سواء كان الأداء التفاضلي منتظماً أم غير منتظم، أو المجموعة مرجعية أم مستهدفة تكون أكبر ما يمكن عند مستوى القدرة المتوسطة تقريباً؛ بمعنى أن الاختبار يزودنا بمعلومات أكثر فاعلية عند الأفراد ذوي القدرة المتوسطة، بينما تكون قيم دالة المعلومات التي يقدمها الاختبار أقل ما يمكن عن مستويات القدرة العالية والمتدنية رغم اختلاف شكل التوزيع، بمعنى أن الاختبار يزودنا بمعلومات قليلة عند الأفراد ذوي القدرات العالية والقدرات المتدنية، وتشابهت مع نتائج دراسة لي، وولاك ودوغلاس (Lee, Wollak & Douglas, 2009). وتشابهت في ذلك مع نتائج الدراسة بأن طرق التقدير اللامعلمية الثلاث أنتجت كمية معلومات أكبر من كمية المعلومات وفق النموذج ثنائي المعلمة للفقرات التي تم توليدها لتطابق النموذج المعلمي، كما تشابهت مع دراسة أجرتها المومني (2011)، ودراسة القيسي (2013)، ودراسة المساعدة (2013) التي بينت أن كمية المعلومات باستخدام النماذج اللامعلمية كانت الأكبر من المعلمية، واختلفت مع دراسة الدلوع (2013) التي بينت أن كمية المعلومات التي يقدمها النموذج ثلاثي المعلمة أكبر من كمية

المعلومات التي يقدمها نموذج موكن.

كما بينت أن شكل دالة المعلومات وفق النموذج المعلمي كان ملتويا نحو اليسار؛ أي أن القيمة القصوى لكمية المعلومات التي يقدمها الاختبار وفق النموذج المعلمي سواء كان الأداء التفاضلي منتظماً أم غير منتظم، أو المجموعة مرجعية أم مستهدفة تكون أكبر ما يمكن عند مستوى القدرة العالية، وتقدم أقل كمية من المعلومات عند مستويات القدرة العالية والمتدنية، ويلاحظ أيضاً أن النموذج الثلاثي اللامعلمي يعطي أكبر قدر من المعلومات عند معظم مستويات القدرة المختلفة مقارنة بدالة معلومات الاختبار المقدرة وفق النموذج الثلاثي المعالم باستثناء القدرات العالية (التي تجاوزت 2 لوجيت). ويلاحظ أن النموذج المعلمي الثلاثي قد قدم قدرا من المعلومات أكثر من النموذج اللامعلمي عند الأفراد ذوي القدرات المرتفعة، لذلك يفضل استخدام النماذج اللامعلمية عندما تكون فقرات الاختبار ذات أداء تفاضلي بغض النظر الأداء التفاضلي منتظماً أو غير منتظم، إلا إذا كان الأفراد من ذوي القدرة العالية، فالنماذج المعلمية قدمت معلومات أكثر من اللامعلمية، وبالتحديد إذا كان الأداء التفاضلي غير منتظم.

خامساً: فيما يتعلق بالإجابة عن السؤال الخامس للدراسة الذي ينص على "ما أثر نوع الأداء التفاضلي للفقرات (منتظم، وغير منتظم)، ونموذج نظرية الاستجابة للفقرة المستخدم (النموذج الثلاثي المعالم، والنموذج الثلاثي اللامعلمي)، والمجموعة (المرجعية والمستهدفة) على ثبات الاختبار؟ فقد تم تقدير الثبات الإمبريقي (الثبات التجريبي) للبيانات باستخدام النموذج اللوجستي ثلاثي المعالم الذي يعتمد على متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد، ويظهر من خلال المرحلة الثالثة من مراحل التحليل في برنامج (Bilog-Mg3)، كما تم تقدير الثبات كذلك للنموذج اللوجستي الثلاثي اللامعلمي باستخدام برنامج TestGraf الذي يعتمد دالة المعلومات كمؤشر مباشر على مصداقية الاختبار، حيث يتم تقدير الثبات R من دالة المعلومات وفق المعادلة R= I/I+1 (Ramsay, 2000). ويبين الجدول (11) قيم معاملات الثبات وفقًا لنوع الأداء التفاضلي للفقرات (منتظم، وغير منتظم)، ونموذج نظرية الاستجابة للفقرة المستخدم (النموذج الثلاثي المعالم، والنموذج الثلاثي اللامعلمي)، والمجموعة (المرجعية والمستهدفة).

الجدول 11. معاملات الثبات وفقًا لنوع الأداء التفاضلي للفقرات ونموذج نظرية الاستجابة للفقرة المستخدم ، والمجموعة.

نوع الأداء التفاضلي	المجموعة	الثبات في نظريتي الاستجابة للفقرة:	
		المعلمية	اللامعلمية
منتظم	مرجعية	0.9097	0.8480
	مستهدفة	0.9114	0.8513
غير منتظم	مرجعية	0.9084	0.8470
	مستهدفة	0.9290	0.8482

يلاحظ من الجدول (11) أن قيم معاملات الثبات المقدرة وفق النموذجين كانت عالية، وأن قيم معاملات الثبات والمقدرة وفق نموذج نظرية الاستجابة للفقرة المعلمية كانت الأعلى، سواء كان نوع الأداء التفاضلي منتظماً أم غير منتظم، وسواء كانت المجموعة مرجعية أم مستهدفة. ولم تبين النتائج فروق ملاحظة بين قيم معاملات الثبات وفقًا لنوع الأداء التفاضلي سواء كان النموذج معلميا أم لامعلميا، وسواء المجموعة مرجعية أم مستهدفة.

الاستنتاجات والتوصيات

تبين للباحث من خلال نتائج الدراسة وجود فرق دال إحصائيًا بين الوسطين الحسابيين لمعالم التمييز لفقرات الاختبار يُعزى للنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج اللوجستي الثلاثي المعلمة مقارناً بالنموذج اللوجستي الثلاثي اللامعلمي؛ أي أن النتائج هنا أظهرت تفوق النماذج المعلمية على نظيرتها النماذج اللامعلمية في تقدير معلمة تمييز الفقرات، فقد كانت قيمتها الأقرب إلى القيمة الفعلية التي تم توليدها، وهي (1) لوجيت، وعدم وجود فروق دالة إحصائيًا بين الأوساط الحسابية لمعالم التمييز لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة المستخدم ونوع الأداء التفاضلي.

كما أظهرت النتائج وجود فرق دال إحصائيًا بين الوسطين الحسابيين لمعالم الصعوبة لفقرات الاختبار يُعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج اللامعلمي مقارناً بالنموذج المعلمي، لذا فإن النموذج المعلمي أبدى تفوقاً بشكل عام،

حيث كانت قيم متوسط معالم الصعوبة هي الأقرب لمتوسط القيمة الفعلية التي تم توليدها، وهي (0.0125) لوجيت، ووجود فروق دالة إحصائيًا بين الأوساط الحسابية لمعالم الصعوبة لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة المستخدم، ونوع الأداء التفاضلي، لذلك يعتبر النموذج الثلاثي اللامعلمي أكثر تشدداً في تقدير معلم صعوبة الفقرات مقارنة بالنماذج اللوجستية المعلمية المرنة في تقدير معالم الفقرات سواء كان الاختبار يتضمن فقرات ذات أداء منتظم، أو فقرات ليست ذات أداء منتظم، في حين أن النماذج المعلمية تظهر التفوق لفقرات الاختبار التي أداؤها التفاضلي منتظم، والتفوق للنماذج اللامعلمية عندما كان الأداء التفاضلي غير منتظم، ويظهر مرة أخرى تفاوت في التقدير لمعالم الصعوبة بين النموذجين المعلمي واللامعلمي.

كما بينت النتائج أيضاً وجود فرق دال إحصائيًا بين الوسطين الحسابيين لمعالم التخمين لفقرات الاختبار يُعزى لنموذج نظرية الاستجابة للفقرة المستخدم؛ لصالح النموذج المعلمي مقارناً بالنموذج اللامعلمي؛ وهذا يعني أن تقديرات التخمين في النماذج المعلمية هي الأعلى مقارنة بالنماذج اللامعلمية؛ وهذا يعني زيادة فرص الكشف عن التخمين العشوائي كمتغير مؤثر في الأداء في الاختبار في النموذج المعلمي عنه في النموذج اللامعلمي. وتبين وجود فروق دالة إحصائيًا بين الأوساط الحسابية لمعالم التخمين لفقرات الاختبار تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة والمجموعة ونوع الأداء التفاضلي، ووجود فروق دالة إحصائيًا تُعزى للتفاعل بين نموذج نظرية الاستجابة للفقرة والمجموعة ونوع الأداء التفاضلي، وهذا يعني تفاوت في تقدير التخمين في حالة النماذج المعلمية، ولكن ذلك يتقلص بشكل ملحوظ في النماذج اللامعلمية، وأظهرت النتائج وجود فرق دال إحصائيًا بين الوسطين الحسابيين لكمية معلومات للاختبار يعزى لنموذج نظرية الاستجابة للفقرة؛ لصالح النموذج اللامعلمي مقارنة بالنموذج المعلمي، ووجود فروق دالة إحصائيًا بين الأوساط الحسابية لكمية المعلومات للاختبار تُعزى: للتفاعل بين النموذج المستخدم ونوع الأداء التفاضلي، وللتفاعل بين النموذج المستخدم والمجموعة، وللتفاعل بين نوع الأداء التفاضلي والمجموعة، وكانت القيمة القصوى لكمية المعلومات التي يقدمها الاختبار وفق النموذج اللامعلمي، سواء كان الأداء التفاضلي منتظماً أم غير منتظم، أو المجموعة مرجعية أم مستهدفة أكبر ما يمكن عند مستوى القدرة المتوسطة تقريباً؛ بمعنى أن الاختبار يزودنا بمعلومات أكثر فاعلية عند الأفراد ذوي القدرة المتوسطة بينما كانت كمية المعلومات التي يقدمها الاختبار أقل ما يمكن عن مستويات القدرة العالية والمتدنية رغم اختلاف شكل التوزيع، بمعنى أن الاختبار يزودنا بمعلومات قليلة عند الأفراد ذوي القدرات العالية والقدرات المتدنية، بينما شكل دالة المعلومات وفق النموذج المعلمي كان ملتويا نحو اليسار؛ أي أن القيمة القصوى لكمية المعلومات التي يقدمها الاختبار وفق النموذج المعلمي سواء كان الأداء التفاضلي منتظماً أم غير منتظم، أو المجموعة مرجعية أم مستهدفة تكون أكبر ما يمكن عند مستوى القدرة العالية، ويلاحظ أن النموذج المعلمي الثلاثي قد قدم قدر من المعلومات أكثر من النموذج اللامعلمي عند الأفراد ذوي القدرات المرتفعة، لذلك يفضل استخدام النماذج اللامعلمية عندما تكون فقرات الاختبار ذات أداء تفاضلي بغض النظر الأداء التفاضلي منتظماً أو غير منتظم عند الأفراد ذوي مستويات القدرة المتوسطة، إلا إذا كان الأفراد من ذوي القدرة العالية، فالنماذج المعلمية قدمت معلومات أكثر من اللامعلمية، وبالتحديد إذا كان الأداء التفاضلي غير منتظم. وفيما يتعلق بقيم معاملات الثبات المقدرة وفق النموذجين كانت عالية، وأن قيم معاملات الثبات، والمقدرة وفق نموذج نظرية الاستجابة للفقرة المعلمية كانت الأعلى، سواء كان نوع الأداء التفاضلي منتظماً أو غير منتظم، وسواء كانت المجموعة مرجعية أو مستهدفة،، ولم تبين النتائج فروق ملاحظة بين قيم معاملات الثبات وفقًا لنوع الأداء التفاضلي سواء كان النموذج معلميا أم لامعلميا، وسواء المجموعة مرجعية أم مستهدفة.

واتفقت نتائج هذه الدراسة مع نتائج دراسة كونينغ وسيجتسما وهامرز (Koning, Sijtsma & Hamers, 2002) ، ودراسة اوليفرس وجالاردو وكرامب (Olivares, Gallardo & Kramp, 2005)، ودراسة لي، وولاك ودوغلاس (Lee, Wollak & Douglas, 2009) التي أفضت إلى الأفضلية في الجمع بين النوعين من النماذج المعلمية واللامعلمية؛ إذ أثبتت النماذج المعلمية فاعليتها في تقديم معلومات مفيدة حول تقدير بعض معالم الفقرات، بينما النماذج اللامعلمية أثبتت فاعليتها في تقدير معالم أخرى، فالنماذج المعلمية واللامعلمية قدمت معلومات متباينة باستخدام إحصائيات مختلفة، وفضلت الدراسة الجمع بينهما لتحسين نوعية الاختبار وجودة فقراته، بينما اختلفت مع دراسة ديهاوس (Dyehouse, 2009) التي أشارت إلى أن النموذج اللامعلمي مفيد بدرجة أكبر لأغراض الاختيار، ومطابقة بياناته من النموذج المعلمي.

واعتماداً على ذلك، فإن الباحث يوصي مطوري الاختبارات بإمكانية استخدام النماذج اللامعلمية في تقدير معالم الفقرات، حيث تكون دقة التقدير لمعالم الصعوبة، والتمييز، والتخمين عالية باستخدام النماذج اللامعلمية مما ينعكس إيجابا على جودة التقديرات، وربما تساهم هذه التوصية في تمكين باني الاختبارات التكيفية (المفصلة) (Adaptive Testing)، وكانت دقة التقدير أكثر وضوحاً عند استخدام كمية المعلومات كمؤشر على ذلك، حيث ظهر التفاوت الواضح في كمية المعلومات في كل مستويات

القدرة باستثناء القدرات العالية التي زادت على (2) لوجيت، كما يمكن أن تقدم هذه الدراسة دليلاً إمبريقياً بأن تقدير الثبات باستخدام النموذج الثلاثي المعلمة تبقى قيمته إلى حد بعيد مستقرة، حتى الفقرات من حيث نوع الأداء التفاضلي،. فعلى سبيل المثال، لا أهمية لأفراد يجيبون على فقرات الاختبار أو المقياس بطريقة عشوائية، إذ إن برامج التحليل وفق نظرية الاستجابة للفقرة تستبعد مثل هذه الفئة وتعدهم أفراداً غير مطابقين للنموذج، كما هو الحال في هذه الدراسة، ولذلك فالجدية والحرص والمصداقية مهمة في الإجابة على فقرات الاختبار أو المقياس عند تقدير الثبات أكثر من كون الفقرات ذات أداء تفاضلي منتظم أو غير منتظم، وربما تعد هذه النقطة تحديداً ميزة مهمة لنظرية الاستجابة للفقرة، في حين تعتبر مأخذاً على النظرية الكلاسيكية التي لا تستبعد أي فرد عند تقدير الثبات للإخبار أو المقياس. كما يوصي الباحث بإجراء دراسات أخرى مماثلة لكن باستخدام بيانات فعلية باستخدام البرامج الإحصائية، ودراسة المقارنة تحت ظروف خاصة من توزيعات مستويات القدرة (طبيعي، منتظم، ملتو التواءً سالباً، ملتو التواءً موجباً). كذلك أخذ حجم العينة وعدد الفقرات كمتغيرين مستقلين. كما يمكن إعادة الدراسة نفسها باستخدام نماذج أخرى من نماذج نظرية الاستجابة للفقرة كالنموذجين الأحادي المعلمة والثنائي المعلمة والرباعي المعلمة ونماذج لامعلمية أخرى، وكذلك استخدام مقاييس التقدير والاختبارات بأشكالها المختلفة للوصول إلى نتائج أكثر قابلية للتعميم.

المراجع

الحجيلي، أحمد (2012). أثر طول الاختبار وحجم العينة في دقة تقدير معلمة صعوبة الفقرة والقدرة ومعادلة الاختبارات بوجود الأداء التفاضلي للفقرة. أطروحة دكتوراه غير منشوره، جامعة اليرموك، إربد، الأردن.

الحياصات، خالد .(2011) .مدى تحقق معايير الفاعلية في معادلة نموذجي اختبار مع بقاء الفقرات ذات الأداء التفاضلي للجنس وحذفها . أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد–الأردن.

دلوع، أحمد عبدالعزيز (2013). مدى التوافق بين نموذج استجابة الفقرة الثلاثي المعلمة البارامتري ونموذج موكن اللابارامتري لنوعين من فقرات اختبار تحصيلي. رسالة دكتوراه غير منشورة، جامعة اليرموك، اربد.

الرحيل، راتب. (2013). اثر وجود أداء تفاضلي في الفقرات المرساوية على دقة المعادلة العمودية لاختبار اوتيس لينون للقدرة العقلية. المجلة الدولية التربوية المتخصصة، 2 (8)، 754 – 771..

علام, صلاح الدين. (2005). نماذج الاستجابة للمفردات الاختيارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي (الطبعة الأولى). القاهرة: دار الفكر العربي.

القيسي، حسين عبدالنبي. (2013). دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار، رسالة دكتوراه غير منشورة، جامعة اليرموك.

مساعدة، محمد رياض (2013). أثر شكل فقرات الاختبار في مطابقة الفقرات لنموذج موكن اللابارامتري ونموذج ثنائي المعلمة البارامتري، رسالة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.

مناصره، سوسن (2014). خصائص توزيع قدرات الأفراد ومعالم فقرات الاختبار وفق نماذج نظرية الاستجابة للفقرة المعلمية واللامعلمية: دراسة مقارنة. أطروحة دكتوراه غير منشوره، جامعة اليرموك، إربد، الأردن.

المومني، رنا ثاني(2012). مقارنة مطابقة الفقرات بين نموذج موكن اللابارامتري والنموذج ثنائي المعالم البارامتري. رسالة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.

النوافلة، علي (2013). أثر نسبة الفقرات ذات الأداء التفاضلي ومستواه في تقدير معالم فقرات الاختبار وقدرات الأفراد وفق نموذج الاستجابة ثلاثي المعلمة، أطروحة دكتوراه غير منشوره، جامعة اليرموك، إربد، الأردن.

Angel, M., Gildon, D. & Jose, M. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Hanszel procedures. Institute for Science Education, 5(3), 44-53.

Baker , Frank B. (2001). The Basic of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation.

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.

Crocker , L. & Algina , J. (1986). Introduction to classical & modern test theory. Fort Worth , TX: Harcourt Brace Jovanovich.

Douglas, J. (1997). Joint Consistency of Nonparametric Item Characteristic Curves and Ability Estimation. Psychometrika, 62, 7-28.

Dyehouse, M. (2009). A Comparison of Model-Data Fit For Parametric & Nonparametric Item Response Theory Models Using Ordinal-Level Ratings. Dissertation Abstract International. (UMI No. 3379330).

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.

Gierl, M. (2004). Performance of SIBTEST when the percentage of DIF item is large, Applied Measurement in Education,17 (3), 241- 264.

Gruijter,D. & Kamp,L.(2005). Statistical test theory for education and psychology. Retrieved December 30, 2005 from: www.leidenuniv.nl/gruijterdnmde.

Gybles,J. (2004). The performance of some observed and unobserved conditional invariance techniques for the detecting of differential item functioning. Quality & Quality, 38, 681-702.

Hambleton, R& Swaminathan, H.(1985). Item response theory: principles and applications. Boston: Kluwer.

Hambleton, R. (1989). Principles and selected application of item response theory .In Linn, Robert, L. (Ed.), Educational Measurement (3rd ed.). New York: American Council on Education, Macmillan Publishing Company.

Hambleton, R. K., Swaminatban, H. & Rogers. H. (1991). Fundamentals of item response theory. Thousand. Thousand oaks. CA: Sage Publication.

Hambleton, Ronald K; Swaminathan, Hariharan.(1985). Item response theory: principles and applications. Boston: Kluwer.

Jodoin, M & Gierl, M. (2001). Evaluation Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection .Applied Measurement in Education, 14(4),329-349.

Kamata, A. & Chu, k. (2003). Test equating in the presence of DIF items. Journal of Applied Measurement,6(3), 342-353.

Koning, E.; Sijtsma, K. & Hamers, J. (2002). Comparison of Four IRT Models When Analysis Two Tests for Inductive Reasoning, Applied Psychological Measurement, 26(3), 302-320.

Lee,S.; Wollak, J. & Douglas, J. (2009). On the Use Nonparametric Item Characteristic Curve Estimation Techniques for Checking Parametric Model Fit. Educational & Psychological Measurement. 69,(2) ,181-197.

Liang, T. (2010). Assessment of the non- parametric approach of evaluating the fit of item response model. Dissertation Abstract International and Theses, University of Massachuselts.

Loken , E. & Rulison, K. (2010). Estimation of a four – parameter item response theory model. The British Psychological Society , 63, 509-525.

Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum.

Mokken, R. & Lewis, C. (1982). A Nonparametric approach to the analysis of dichotomous item responses. Applied Psychological Measurement, 6, 417-430.

Nozawa,Y.(2008). Comparison of parametric and nonparametric IRT equating methods under the Common – item nonequivalent groups design, A thesis Submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychological and Quantitative Foundations , Educational Measurement and Statistics in the Graduate of the University of Lowa. http://books.google.jo.

Olivares, A; Gallardo, D. & Cramp, U. (2005). Parametric vs. Nonparametric IRT Modeling of Likert Type Personality Data. Multivariate Behavioral Research, 40(2), 261-279.

Park, C. (2010). Examining the relationship between deferential item functioning and differential test

functioning. Language Testing 23(4), 475 -496.

Ramsay, J.(2000) A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data, McGill University August 1, 2000

Sijtsma, K. & Hemker, B. (2000). A taxonomy of IRT Models for Ordering of Persons and Items Using Simple Sum Scores. Journal of educational and Behavioral Statistics , 25, 391-415.

Sijtsma, K. & Molenaar, I. (2002). Introduction to nonparametric item response theory. Sage Publication, International Educational and Professional Publisher. Thousand Oaks: London. New Delhi.

Sijtsma, K. Verweij, A. (1992). Mokken scale analysis: the cortical Considerations and an applications to transitive tasks. Applied Measurement in Education, 5(4),355 – 373.

Stochi, Joens & Croudence (2012). Nonparametric Extension of Item Reponses Theory Models and Its Usefulness for Assessment of Dimensionality of tests, L (2): 1-19.

Yanju, Li. (2012) .Item Discrimination and type I error rates in DIF detecting using Mantel-Haenszel and Logistic Regression. Procedure Unpublished doctoral dissertation, Ohio University.

Zhan, L. & Zumbo, B. (2009). Impact of Differential Item Functioning on Subsequent statistical conclusions Based on observed Test Score. Psychological, 30(2),343-370.

Zhang, y. & Mattews, J. & Dorans, N. (2003). Using dif dissection to assess effects of item deletion due to dif on the performance of sat I: Reasoning test subpopulations (Report). U.S, New Jersey.

Zumbo, B. (2000). The effect of DIF and impact on classical test statistics: Undetected DIF and impact of the reliability and interpretability of scores from a language proficiency test. Paper presented at the annual conference of the National Council on Measurement in Education (NCME), April,2000 new, Orleans,LA.

Impact of the Type of Differential Items Functioning on the Psychometric Properties of Items and Test According to Parametric and Non-parametric Models of IRT

*Nedal K. Al-Sharefeen**

ABSTRACT

This study aimed at investigating the impact of the type of DIF(Systematic, Unsystematic) on the psychometric properties of items and test according to 3LM parametric and 3LM non-parametric models of IRT. To achieve the study objective, (WINGEN) statistical software was used to stimulate two forms of test each form consisted of 50 items. The first form included 10 systematic DIF, the second form included 10 unsystematic DIF. The responses of (1000) persons were stimulated, the data was analyzed according to 3PLM using Bilog-Mg and according to 3LM nonparametric used TestGraf. The results showed that there was a significant different between the means of discrimination parameters and the means of guessing parameter of test items due to IRT models, in favor of parametric model compared with the nonparametric one. Also, significant difference was found between the means of difficulty parameters to the model used, in favor of the non-parametric model with the parametric one; significant differences were found between the means of the parameters of each of: Difficulty and Guessing of test items due to the interaction between the model used and type of DIF; significant differences was observed between the means of amount of test information due to the model, where the amount of test information larger at all of ability levels according to non-parametric model, whether the DIF model was systematic or unsystematic, with the exception of the 3LM presented information larger than non-parametric model at the high ability levels. In additions the results indicated that the estimated values of the reliability coefficients according to the two models were high, these values were higher for the parametric model, whether the DIF was systematic or unsystematic.

Keywords: Item Response Theory(IRT), Differential Item Functioning(DIF), Psychometric Properties of Items, Psychometric Properties of Test, Parametric and Non-parametric Models , Three Logistic model (3LM).

* Faculty of Education, Yarmouk University. Received on 4/2/2016 and Accepted for Publication on 24/6/2016.