أثر الفقرات ذات الأداء التفاضلي المنتظم في تقديرات معالم فقرات الاختبار والأفراد في ضوء بعض المتغيرات باستخدام بيانات مولدة وفقاً لنموذج استجابة الفقرة ثلاثي المعلمة

*علي صالح رجا نوافله**

ملخص

هدفت هذه الدراسة إلى بيان أثر الفقرات ذات الأداء التفاضلي المنتظم في تقدير معالم فقرات الاختبار وقدرات الافراد وفق نموذج الاستجابة للفقرة ثلاثي المعلمة باستخدام بيانات مولدة، حيث تضمن تصميم الدراسة ظروفاً مختلفة بناءً على متغيري الدراسة، وهما (نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة)، فقد كانت نسبة الفقرات ذات الأداء التفاضلي الموجودة في الاختبار هي (10%,20%,40%)، وكان مستويا الأداء التفاضلي للفقرة هما: (مستوى متدنٍ بمقدار 0.5 ومستوى عالٍ بمقدار 1.5). وبناءً على ذلك تكون تصميم الدراسة من (2*3=6) مجموعات مستهدفة تختلف فيما بينها بنسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة، بالإضافة إلى المجموعة المرجعية التي لا تشتمل على أي فقرة ذات أداء تفاضلي، والتي تمثل القاعدة الأساسية للمقارنة، مما نتج (7) مجموعات للدراسة. وقد تم استخدام برنامج (Wingen) لتوليد استجابات (510) أفراد في كل مجموعة، وقد تكونت العينة الكلية من (3570) مفحوصاً، ومن ثم تم استخدام برنامج (Bilog MG) للتحقق من افتراض أحادية البعد، ومطابقة كل من الفقرات والأفراد للنموذج، وتقدير معالم الفقرات والمفحوصين، وذلك حسب كل مجموعة من المجموعات السابقة. وقد أشارت النتائج بوجود فروق ذات دلالة إحصائية بين متوسطات تقديرات معالم الصعوبة للفقرات تُعزى إلى نسبة الفقرات ذات الأداء التفاضلي، ولصالح المجموعة المستهدفة التي تشتمل على (40% من الفقرات ذات الأداء التفاضلي)، كما وتوجد فروق ذات دلالة إحصائية بين متوسطات تقديرات معالم التمييز ومعالم التخمين للفقرات تُعزى إلى مستوى الأداء التفاضلي للفقرة، ولصالح المجموعة المرجعية، وفيما يتعلق بقدرات الأفراد بيّنت النتائج عدم وجود فروق ذات دلالة إحصائية بين متوسطات تقديرات معالم قدرات الأفراد تُعزى إلى متغيري الدراسة، ووجود فروق ذات دلالة إحصائية بين متوسطات العلامات الملاحظة للأفراد تُعزى إلى متغيري الدراسة (نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة) والتفاعل بينهما.

الكلمات الدالة: نسبة الفقرات ذات الأداء التفاضلي في الاختبار، مستوى الأداء التفاضلي للفقرة، المجموعة المرجعية، المجموعة المستهدفة، معالم الفقرات، القدرة.

المقدمة

يستخدم الاختبار كأداة من أدوات جمع البيانات والمعلومات، بهدف إصدار الأحكام وصنع القرارات التربوية، وهو أداة قياس تحوي مجموعة من المثيرات التي تمثل السمة أو القدرة المرغوب قياسها، حيث تركزت جهود الباحثين في بناء الاختبارات وتطويرها على استخراج خصائص الفقرات من حيث الصعوبة والتمييز والتخمين. وبالرغم من أهمية هذه المعالم إلا أنها ليست كافية للحكم على صلاحية فقرات الاختبار للغرض الذي أُعدّت لأجله؛ وذلك لأن الاستجابة لفقرات الاختبار قد تتأثر بعوامل أخرى بالإضافة إلى قدرة المفحوصين، كالجنس أوالعرق أومكان السكن أو اللغة أو المستوى الاقتصادي والاجتماعي بما يؤثر سلباً على نتيجة الاختبار، وبالتالي على القرارات المبنية عليه. وبناءً على ذلك توصف الفقرات بأنها ذات أداء تفاضلي نحو مجموعة أو فئة دون أخرى (Jensen, 1980).

وقد أدى هذا الاهتمام الى اعتبار أن خلو الاختبار أو فقراته من الأداء التفاضلي من أهم الشروط التي ينبغي توافرها في الاختبار قبل نشره، وقد وضعت الجمعية الامريكية للبحث التربوي (American Educational Research Association AERA, 1999) الأداء التفاضلي ضمن قواعد النشر للاختبارات، وأشارت إلى أنه إذا استخدمت الاختبارات في اتخاذ القرارات، فلا بد من الحصول على أدلة

* هيئة اعتماد مؤسسات التعليم العالي وضمان جودتها. تاريخ استلام البحث 2015/2/16، وتاريخ قبوله 2015/8/18.

©2017 عمادة البحث العلمي وضمان الجودة/ الجامعة الأردنية. جميع الحقوق محفوظة.

تجريبية واضحة من أجل دعم الإدعاءات بأن فقرات الاختبار لا تعكس أداءً تفاضلياً للفقرة عبر المجموعات من المفحوصين.

ولقد أشار فراري وزيمر (Frary & Zimmer, 1984) إلى أن احتواء الاختبار على فقرات ذات أداء تفاضلي له أثار سلبية على القرار المتخذ في ضوء علامة الاختبار وعلى بعض السمات الإحصائية لهذه العلامات؛ حيث ينخفض الصدق والثبات نتيجة لوجود فقرات ذات أداء تفاضلي في الاختبار. ويرى كاماتا وشو (kamata & Chu, 2003) أن الوضع الأمثل يتمثل بحذف الفقرات ذات الأداء التفاضلي من الاختبار لضمان عدالته، وهذه الطريقة أبسط واقل إثارة للجدل من استخدام المعلومات التي توفرها الفقرات ذات الأداء التفاضلي.

ولقد أُجريت العديد من الدراسات على الأداء التفاضلي للفقرة ومن هذه الدراسات الدراسة التي أجراها غامر وانجلهارد (Engelhard & Gamer ,1999) هدفت للكشف عن الفروق بين أداء الجنسين في إختبار قدرة رياضية، وهو إختبار يتضمن (60) فقرة، من نوع الإختيار من متعدد تتوزع على النحو الآتي : (14) فقرة من موضوع تحليل البيانات، و (23) فقرة في موضوع الهندسة والقياس، و (11) فقرة في موضوع الحساب، و (12) فقرة في موضوع الجبر، كما يتضمن الإختبار (8) فقرات ذات إستجابة محددة البنا، ثم طبق الإختبار على (3970) طالباً من طلبة الصف الحادي عشر، وأشارت نتائج الدراسة إلى أن الطالبات أظهرن فرقاً ذو دلالة إحصائية وثبات على الطلاب في فقرات الإختيار من متعدد في الجبر، بينما أظهر الذكور فرقا أقل ثباتا في الفقرات المتعلقة بالهندسة والإعداد والعمليات الحسابية، وكانت متوسطات الدرجات للطلاب أعلى بدرجة دالة إحصائيا من الإناث في اثنتين من ثماني من الفقرات متعددة التدريج.

واجرى زومبو (Zumbo, 2000) دراسة هدفت للكشف عن أثر الفقرات ذات الأداء التفاضلي على إحصائيات نظرية الاختبار الكلاسيكية، وهي: معاملات الارتباط، ومعامل كرونباخ ألفا، والخطأ المعياري في القياس، ومعامل الإرتباط المصحح لأثر الإنخفاض(وهو معامل الارتباط المعدل من الاخطاء العشوائية بين نتائج الافراد على الاختبار المتنبئ وعلى الاختبار المحك)، واستخدم في هذه الدراسة نتائج اختبار التوفل (TOFEL)، وذلك باستخدام النموذج اللوجستي ثلاثي المعلمة في ظل مجموعة متنوعة من ظروف الاختبار. وكان حجم الاختبار ثابتاً مقداره (38) فقرة، وعدد الفقرات التي تبدي أداءً تفاضلياً هي فقرة واحدة فقط، ونوع الاداء التفاضلي هو المنتظم، ومستوى الأداء التفاضلي للفقرة هو(1.0 ,0.5 ,0.2). وهذا يعدّ أحد الظروف التجريبية المحبذة من أجل الكشف فيما إذا كان هناك فقرة واحدة من الفقرات ذات الأداء التفاضلي قادرة على التأثير في إحصائيات الاختبار في ضوء النظرية الكلاسيكية في القياس. وركزت الدراسة على نوع واحد من الأداء التفاضلي وهو الأداء التفاضلي المنتظم، وهذا يعكس أحد المواقف التي يمكن من خلالها عزو وجود أداء الفقرة التفاضلي في الفروقات في صعوبة الفقرة فقط. وأشارت النتائج إلى أن أداء الفقرة التفاضلي لم يكن ذا أثر دال إحصائياً على إحصائيات الاختبار في ضوء النظرية الكلاسيكية في القياس..

وأجرى زومبو (Zumbo, 2003) دراسة هدفت للكشف فيما إذا كان أداء الفقرة التفاضلي على مستوى الفقرة يظهر من خلال التحليلات على مستوى الاختبار مثل استخدام التحليل العاملي متعدد المجموعات، والأحادي، وحساب معامل ألفا لكل مجموعة. وتم تقديم بيانات خاصة لدراسة توليد البيانات باستخدام معالم الفقرة التي تم الحصول عليها من اختبار اللغة الإنجليزية توفل (TOFEL) نتيجة لترجمة هذه الاختبارات من لغة إلى لغة اخرى من أجل الكشف فيما إذا كان الأداء التفاضلي للفقرة على مستوى الفقرة يظهر من خلال التحليل على مستوى المقياس. وولدت بيانات استجابات المفحوصين باستخدام برنامج (MIRTGN)، وذلك باستخدام نموذج الاستجابة للفقرة اللوجستي ثلاثي المعلمة تحت ظروف مختلفة من الاختبارات، وهي: الفروق في مستويات الأداء التفاضلي للفقرة، والتي تراوحت بين مستويات متوسطة إلى مستويات مرتفعة، واستخدام الأداء التفاضلي المنتظم فقط، كما واستخدمت أعداد مختلفة من الفقرات ذات الأداء التفاضلي، وهي (16 , 8 , 4 , 1) فقرة من مجموع الفقرات الكلي الذي بلغ (38) فقرة. وأشارت النتائج أنه بغض النظر عن مستوى الأداء التفاضلي للفقرة، ونسبة الفقرات ذات الأداء التفاضلي في الاختبار، فقد أشار التحليل العاملي التوكيدي للمجموعات المتعددة (وهو تصميم تجريبي لعدة مجموعات مع عدة اختبارات) أن التشبعات العاملية وتباينات الخطأ كانت متطابقة من الناحية الإحصائية للاختبار الأصلي والاختبار المترجم على الرغم أن الاختبار المتُرجم قد إحتوى على ما نسبته (24.1%) من الفقرات التي تظهر مستويات عالية من الأداء التفاضلي للفقرة.

وأجرى كاماتا وشو (kamata & chu, 2003) دراسة هدفت إلى دقة تقديرات معالم الصعوبة للفقرات ومعالم القدرة للأفراد ومعادلة درجات الاختبار باستخدام نموذج model Rasch بوجود فقرات ذات أداء تفاضلي. كما تم أيضاً اختبار عدد الفقرات ذات الأداء التفاضلي وكانت (10,5,2)، ومستوى

الأداء التفاضلي حسب مؤشر mantel-Haenzel (2.00, 1.2, 0.8, 06). وأشارت النتائج إلى أن فاعلية نموذج راش الهرمي أفضل من نموذج راش التقليدي فيما يتعلق بدقة تقدير معالم الصعوبة للفقرات وقدرات الأفراد بوجود الفقرات ذات الأداء التفاضلي، فقد كان متوسط مربع الخطأ لتقدير معالم الصعوبة لفقرات الاختبار باستخدام نموذج راش الهرمي أقل من النموذج الأساسي (راش) الذي لا يحتوي على فقرات ذات أداء تفاضلي، وتأثر نموذج راش (النموذج الأساسي) بشكل كبير بعدد الفقرات ذات الأداء التفاضلي، وكذلك بمستوى الأداء التفاضلي للفقرة.

وفي دراسة أجراها زهان وزومبو (Zhan & Zumbo, 2009) هدفت للكشف عن أثر الفقرات ذات الأداء التفاضلي على الخطأ من النوع الأول وحجم الأثر للعينات المستقلة لاختبار (T-Test) من خلال علامات الاختبار الكلية الملاحظة. واستخدمَ النموذج اللوجستي ثلاثي المعلمة لتقدير المعالم بوجود ثلاثة مستويات للأداء التفاضلي للفقرة (مستوى متدنٍ بمقدار 0.4 ، مستوى متوسط بمقدار 0.6، مستوى عالٍ بمقدار 0.8). كما استخدم أيضاً أربع حالات من اعدد الفقرات ذات الأداء التفاضلي في الاختبار، وكانت على الشكل التالي: (1, 4, 8, 16) من عدد الفقرات الكلية في الاختبار والتي بلغت (38) فقرة. إضافة لذلك، ولأغراض المقارنة؛ فقد تم دراسة أن عدم وجود الفقرات ذات الأداء التفاضلي يمثل بيانات القاعدة الخطية. نتيجة لهذه الظروف فقد كانت تجربة توليد البيانات الناتجة على الشكل التالي (4*3). وأشارت النتائج عند وجود الفقرات ذات الأداء في الاختبار تزداد الفروقات في حجم الأثر، فكلما ازداد عدد فقرات ذات الأداء التفاضلي في الاختبار وزاد مستوى الأداء التفاضلي للفقرة ازدادت الفروقات في حجم الأثر، حيث تكون الفروق في المتوسط الحسابي للعلامة الملاحظة متضخمة نتيجة لوجود فقرات ذات الأداء التفاضلي.

كما وأجرى إرحيل (2011) دراسة هدفت للكشف عن أثر الفقرات ذات الأداء التفاضلي للجنس على دقة المعادلة العمودية لإختبار أوتيس لينون للقدرة الفعلية بمستوييية المتوسط والمتقدم ، حيث تألفت عينه الدراسة من (1001) طالب وطالبة من طلبة الصف السابع والثامن والتاسع , مقسمين إلى (500) في مستوى القدرة المتوسط, و (501) في مستوى القدرة المتقدم، واستخدم في هذه الدراسة طريقة مانتل هانزل (MH) للكشف عن الفقرات ذات الأداء التفاضلي في كل مستوى من مستويات الاختبار بشكل منفصل , وبعد ذلك تم إجراء المعادلة العمودية باستخدام برمجية (Bilog-MG) وفقا للنموذج أحادي المعلمة، وإجريت المعادلة في ضوء الإحتفاظ بالفقرات ذات الأداء التفاضلي وفي ضل حذف الفقرات ذات الأداء التفاضلي. وحيث أشارت النتائج إلى وجود أثر للأداء التفاضلي على دقة المعادلة العمودية.

الادب النظري

نظرية الاستجابة للفقرة

تستند نماذج نظرية الاستجابة للفقرة إلى بعض الافتراضات، والتي يجب تحقيقها في البيانات لكي تؤدي إلى نماذج يمكن الوثوق بها، وهذه الافتراضات هي: أُحادية البعد (Uunidimensionality) وافتراض الإستقلال الموضعي (local Independence)، والعلاقة الوتيرية بين القدرة والاداء على الفقرة (Item Characteristic Curve)، وافتراض التحرر من السرعة في الأداء (Speediness)، وقد انبثق عنها مجموعة من النماذج، وهذه النماذج ملائمة للفقرات ثنائية التدريج، اذ يعد نموذج راش (Rasch Model) اكثر هذه النماذج شهرة.

وتهدف نظرية الاستجابة للفقرة والنماذج الرياضية المرتبطة بها إلى تقدير جميع معالم الفقرة والقدرة، وكلما كان هنالك مواءمة (Fitting) بين النموذج المستخدم ومجموعة البيانات أدى ذلك إلى الحصول على تقديرات دقيقة لهذه المعالم. ويشير هاملتون وسوامينثان (Hambleton & Swaminath , 1985)، أن احدى الطرق التي تمكننا من خلال نماذج السمات الكامنة بناء اختبارات أفضل، بحيث تعطي معلومات أكثر حول السمة الكامنة هو استخدام اقتران المعلومات (information)، ويمكن استخدامها ايضاً للتأكد من دقة تقدير معالم الفقرة ومعلمة القدرة، واقتران معلومات الاختبار هو مقدار يتناسب عكسياً مع الخطأ المعياري في التقدير، وبالتالي تعتبر تقديراً لثبات الاختبار لأنه بزيادة الخطأ المعياري في التقدير يقل الثبات والعكس صحيح.

وقد بيّن هاملتون وسوامينثان وروجرز (Rogers, 1991 Hambleton & Swaminathan) أنه للحصول على تقديرات المعالم بدقة عالية يجب المواءمة بين صعوبة الفقرة من جهة، وقدرات المفحوصين من جهة أخرى، وإن معالم الفقرات السهلة لا يتم تقديرها بدقة عالية في المجموعات ذات القدرات العالية، ومعالم الفقرات الصعبة لا يتم تقديرها بدقة عالية في المجموعات ذات القدرات المتدنية.

التحيز والأداء التفاضلي للفقرة: (Bias and Differential Item Functioning).

إن الأداء التفاضلي للفقرة دلالة مشتقة إحصائياً للتعبير عن الفروق في احتمالات الاستجابة الصحيحة للفقرة بين

مجموعتين مختلفتين في نفس مستوى القدرة (Teresi, 2000; Jodoin & Geirl & 2001)، والأداء التفاضلي للفقرة من منظور نظرية الاستجابة للفقرة يتم من خلال فحص العلاقة بين الاحتمالات المشروطة للاستجابات الصحيحة على الفقرة وبين السمة الكامنة المراد قياسها (θ)، بحيث يتم مقارنة منحنيات خصائص الفقرة في المجموعتين الفرعيتين(المجموعة المرجعية والمجموعة المستهدفة)، والتي يتم تمثيلها وفق النموذج اللوجستي المستخدم.

بناءً على العديد من نتائج الدراسات (pae, 2004 ; Hambleton & Rogers, 1989) ظهر نوعان من الأداء التفاضلي هما: الأداء التفاضلي المنتظم (Uniform DIF): ينتج الأداء التفاضلي المنتظم لفقرة عندما لا يحدث تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد. والأداء التفاضلي غير المنتظم (Nonuniform DIF): يظهر عندما يكون هنالك تفاعل بين مستوى القدرة والمجموعة التي ينتمي إليها الفرد، فنمط الفروق في احتمالية الاستجابة الصحيحة ليست متشابهة في كل مستويات القدرة.

يوجد العديد من الطرق للكشف عن الأداء التفاضلي للفقرة، فبعضها يعتمد على النظرية الكلاسيكية في القياس، مثل طريقة تحليل التباين (Analysis of variance)، وطريقة مربع كآي (Chi-Square, χ^2)، وطريقة مانتل هانزل (Mantel-Method Haenzel)، وبعضها الآخر يستند إلى النظرية الحديثة في القياس، مثل طريقة فرق المساحة لمنحنى خصائص الفقرة (Item Characteristic Curve Methods, ICC)، وطريقة نسبة الأرجحية : Ratio Tests For DIF− IRT Likelihood). حيث (IRTLRDIF. تعتمد الطرق المتبعة في الكشف عن الأداء التفاضلي للفقرة وفق نماذج منحنى خصائص الفقرة على التشابه في منحنيات خصائص الفقرة للمجموعات الفرعية المختلفة، وقد بين راجو (Rajua, 1988) أنه يمكن حساب المساحة بين المنحنيات لخصائص الفقرة، سواء ذات المعلمة الواحدة أو المعلمتين أو الثلاثة معالم. ففي النموذج ثلاثي المعالم يُعبر عن المساحة بالمعادلة التالية:

$$Area = (1-c)\left|\frac{2(a_2-a_1)}{D\,a_1a_2}\ln\left(1+\exp\left(\frac{D\,a_1a_2(b_2-b_1)}{a_2-a_1}\right)\right)-(b_2-b_1)\right| \quad \ldots\ldots(1)$$

حيث يمثل الرمز (C) معلمة التخمين، ويمثل الرمز (a_1) معلمة التميز للمجموعة الأولى، و (a_2) معلمة التميز للمجموعة الثانية، ويمثل الرمز (b_1) معلمة الصعوبة للمجموعة الأولى، و (b_2) معلمة الصعوبة للمجموعة الثانية، والرمز (D) الذي يشير إلى قيمة ثابتة وتساوي (1.7).

وقد اقترح زيكي (Zieky) المشار إليه في دريانا (Driana, 2007) باستخدامه طريقة "مانتل هانزل" مقاييس لحجم الأثر، والتي يمكن استخدامها من أجل تكميم مستوى الأداء التفاضلي للفقرة ضمن سياق نظرية الاستجابة للفقرة على الشكل التالي:

1. مستوى (A) من الفقرات ذات الأداء التفاضلي، وهي الفقرات ذات الأداء التفاضلي المهمل.
2. مستوى (B) من الفقرات ذات الأداء التفاضلي، وهي الفقرات ذات الأداء التفاضلي المتوسط، إذا كانت Δ_{aMH} دالة احصائياً عند مستوى الدلالة 0.05 و $1 \le \left|\Delta_{aMH}\right| < 1.5$.
3. مستوى (C) من الفقرات ذات الأداء التفاضلي: أداء تفاضلي عالٍ، إذاكانت قيمة Δ_{aMH} دالة إحصائياً عند مستوى الدلالة 0.05 و $\left|\Delta_{aMH}\right| \ge 1.5$.

مشكلة الدراسة وأسئلتها

هُناك الكثير من الأبحاث التي أُجريت للكشف عن الأداء التفاضلي للفقرة، فبعضها يعتمد على النظرية الكلاسيكية وبعضها يعتمد على النظرية الحديثة في القياس. ومعظم هذه الأبحاث ركزت على المقارنة بين الطرق المختلفة للكشف عن الأداء التفاضلي للفقرة، ونوع الاداء التفاضلي(منتظم−غير منتظم)، ومصدر الاداء التفاضلي، بمعنى آخر أكثر هذه الدراسات تناولت مشكلة الاداء التفاضلي كمتغير تابع ومن هذه الدراسات دراسة غامر وانجلهارد (Gamer & Engelhard ,1999) ودراسة ارحيل 2011، من جهة اخرى فان القليل من الدراسات تناولت مشكلة الاداء التفاضلي كمتغير مستقل مثل دراسة زومبو (Zumbo, 2003) ودراسة زهان وزومبو (Zhan & Zumbo, 2009)، حيث أن هنالك قبول واسع بأن وجود الأداء التفاضلي في الفقرة يؤثر على تقدير معلمة الصعوبة على مستوى الفقرة نفسها لأن الفقرة التي تظهر أداءً تفاضلياً تكون أصعب لمجموعة من مجموعة أخرى(اصعب للمجموعة للمستهدفة من المجموعة المرجعية)، لذلك جاءت هذه الدراسة لبيان اثر الفقرة أو الفقرات ذات الاداء التفاضلي على تقدير المعالم لباقي الفقرات الموجودة في الاختبار وذلك من خلال الاجابة عن الاسئلة الآتية:

اسئلة الدراسة

1. ما مدى مطابقة الأفراد ومطابقة الفقرات للنموذج اللوجستي ثلاثي المعلمة تبعاً لنسبة الفقرات ذات الأداء

التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة؟

2. ما مدى تحقق افتراضات النموذج (احادية البعد والاستقلال الموضعي) في الاختبار تبعاً لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة؟

3. هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α=0.05) بين متوسطات تقديرات معالم فقرات الاختبار (الصعوبة والتمييز والتخمين) تُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة أو التفاعل بينهما؟

4. هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α=0.05) بين متوسطات تقديرات قدرات الأفراد على الاختبار تُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة أو التفاعل بينهما؟

أهمية الدراسة

تعمل هذه الدراسة على محدد من أهم محددات بناء الاختبارات، حيث إن موضوع الأداء التفاضلي للفقرة يهدد صدق البناء والثبات وتقدير المعالم للقدرة والفقرة، وبالتالي تلوث النتائج، وبذلك يصعب اتخاذ قرارات تربوية في ضوء بيانات غير صادقة. وكل ما سبق يعطي أهمية للنتائج بحيث تساعد المؤسسات والجهات التي تقوم ببناء وتطوير الاختبارات والمهتمين بموضوع الأداء التفاضلي للفقرة في وضع آلية لتقييم نتائجهم عند اشتمال اختباراتهم على فقرات ذات أداء تفاضلي على اختلاف نسبتها ومستواها في الاختبار في الحكم على مدى حجم الأثر للفقرات ذات الأداء التفاضلي في اختباراتهم.

مصطلحات الدراسة وتعريفاتها الاجرائية

نسبة الفقرات ذات الأداء التفاضلي: هي عدد الفقرات ذات الأداء التفاضلي الموجودة في الاختبار مقسوماً على عدد الفقرات الكلية في الاختبار، حيث كان هنالك ثلاث نماذج للاختبار حسب أعداد الفقرات ذات الأداء التفاضلي في الاختبار، وهي (5,10,20) فقرة من أصل العدد الكلي لفقرات الاختبار والبالغ عدد فقراته (50) فقرة، وكانت هذه الاعداد تشكل النسب التالية للفقرات ذات الأداء التفاضلي الموجودة في الاختبار، وهي (10%,20%,40%).

مستوى الأداء التفاضلي للفقرة: هو مقدار الزيادة في قيمة معلمة الصعوبة للفقرة في المجموعات المستهدفة، حيث كان في هذه الدراسة مستويان لمقدار الزيادة في قيمة معلمة الصعوبة هما: مستوى متدنٍ بزيادة مقدارها (0.5) في قيمة معلمة الصعوبة لدى المجموعة المستهدفة، ومستوى عالٍ بزيادة مقدارها (1.5) في قيمة معلمة الصعوبة لدى المجموعة المستهدفة.

المجموعة المرجعية: هي المجموعة التي لا تشتمل على أي فقرة ذات أداء تفاضلي، وهذه المجموعة كانت بمثابة القاعدة الأساسية للمقارنة مع باقي المجموعات المستهدفة.

المجموعة المستهدفة: هي المجموعة التي تشتمل على فقرات ذات أداء تفاضلي، ونظراً لوجود ثلاث حالات للاختبار حسب نسب الفقرات ذات الأداء التفاضلي وهي (10%,20%,40%)، ومستويان للأداء التفاضلي للفقرة (مستوى متدنٍ بمقدار 0.5، ومستوى عالٍ بمقدار 1.5)؛ فقد كان هنالك (6) مجموعات مستهدفة تختلف فيما بينها بنسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة.

الطريقة والاجراءات

تتناول اجراءات الدراسة وصفاً لكيفية توليد البيانات وإجراءات تقدير المعالم للفقرات وللإفراد المستخدمة للإجابة عن أسئلة الدراسة.

الظروف التي تستند إليها إجراءات توليد البيانات.

1. من حيث تصميم البيانات المستخدم في هذه الدراسة، فقد تكون من (7) مجموعات أو حالات وهما (6) مجموعات مستهدفة تختلف فيما بينها بنسبة الفقرات ذات الاداء التفاضلي ومستوى الاداء التفاضلي للفقرة، ومجموعة مرجعية وهذه المجموعة لا يوجد بها أي فقرة ذات أداء تفاضلي، وقد كانت هذه المجموعة تمثل بيانات القاعدة الأساسية للمقارنة مع باقي المجموعات، والمجموعات المتبقية جميعها مجموعات مستهدفة تحتوي على فقرات ذات أداء تفاضلي بظروف مختلفة.

2. من حيث عدد الفقرات الكلية في الاختبار، بلغ عددها (50) فقرة في كل مجموعة من المجموعات، وكانت الفقرات التي لا تظهر أداءً تفاضلياً مكررة كما هي في جميع المجموعات، حيث كان الاختلاف بين المجموعات في الفقرات التي تظهر أداءً تفاضلياً، ففي كل مجموعة من المجموعات ظرفية خاصة تختلف بها عن المجموعات الأخرى.

3. عدد الأفراد في كل مجموعة من المجموعات. حيث بلغ عدد الأفراد في كل مجموعة (510) فردًا، وكان

التصميم المستخدم في هذه الدراسة يشتمل على (7) مجموعات. وبناءً على ذلك؛ فقد بلغ عدد الأفراد الكلي (3570) مفحوصاً.

4. نسبة الفقرات ذات الأداء التفاضلي الموجودة في الاختبار. لقد تم اختبار ثلاث حالات للاختبار حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار، وهي (10%,20%,40%).

5. مستوى الأداء التفاضلي الفقرة. أُعتمد في هذه الدراسة حالتين لمستوى الأداء التفاضلي للفقرة، وهي (مستوى متدنٍ

بمقدار0.5، ومستوى عالٍ بمقدار1.5)، وذلك حسب الإحصاءات التي قدمها زيكي (2007).

6. نوع الأداء التفاضلي. اقتصرت هذة الدراسة على الأداء التفاضلي المنتظم.

إجراءات توليد البيانات

استخدم برنامج Wingen لتوليد معالم (50) فقرة في ظل ظروف مختلفة، وهي مستوى الأداء التفاضلي (0.5،1.5)، ونسبة الفقرات ذات الأداء التفاضلي الموجودة في الاختبار، وهي (10%,20%,40%)، كما وتم توليد البيانات متمثلة باستجابات المفحوصين باستخدام النموذج اللوجيستي ثلاثي المعالم، وركزت الدراسة الحالية على الفقرات ثنائية التدريج، وكانت عملية توليد هذه الاستجابات في هذه المرحلة مكونة من خطوتين رئيسيتين تم إعتمادها بشكل متكرر من أجل الإجابة عن أسئلة الدراسة.

الخطوة الأولى

في هذه الخطوة تم توليد البيانات الخاصة بالمجموعة المرجعية، حيث تم توليد قيم القدرة للمفحوصين من خلال استخدام التوزيع المنتظم (uniform)، وكانت أعلى قدرة (maximam) هي (3+)، وأدنى قدرة (minimam) هي (-3)، وذلك باستخدم النموذج اللوجيستي ثلاثي المعلمة. وتم توليد معلمة الصعوبة باستخدام التوزيع المنتظم (uniform)، وكانت أدنى قيمة هي (1.5−)، وأعلى قيمة هي (1.5+). كما تم توليد معلمة التمييز باستخدام التوزيع المنتظم، وكانت أدنى قيمة (0.2) وأعلى قيمة (1.2). واستخدم أيضاً التوزيع المنتظم في توليد معلمة التخمين حيث كانت أدنى قيمة (0.2)، وأعلى قيمة (0.25). ويبين الملحق (1) معالم الفقرات التي تم توليدها في المجموعة المرجعية.

الخطوة الثانية

في الخطوة الثانية، تم اعتماد نفس الفقرات التي تم توليدها في الخطوة الأولى باستثناء إجراء تغييرات (الزيادة في قيمة معلمة الصعوبة) على بعض الفقرات من أجل أن تعكس أداءً تفاضلياً، وحسب كل مجموعة من مجموعات البيانات التي تم توليدها. ويبين الجدول (1) أرقام الفقرات التي تم اختيارها من المجموعة المرجعية المبينة في الملحق (1) لكي تظهر أداءً تفاضلياً في المجموعات المستهدفة.

الجدول(1):أرقام الفقرات ذات الأداء التفاضلي وفقاً لنسبتها في الاختبار (10%,20%,40%).

النموذج الأول الذي يشتمل على 10% من الفقرات ذات الأداء التفاضلي بلاضافة إلى أرقام هذه الفقرات	النموذج الثاني الذي يشتمل على 20% من الفقرات ذات الأداء التفاضلي بلاضافة إلى أرقام هذه الفقرات	النموذج الثالث الذي يشتمل على 40% من الفقرات ذات الأداء التفاضلي بلاضافة إلى أرقام هذه الفقرات
1	1	1
		3
	6	6
	8	8
		14
		15
	18	18
	21	21
		23
		26
		28
29	29	29
		32

النموذج الأول الذي يشتمل على 10% من الفقرات ذات الأداء التفاضلي بلاضافة إلى أرقام هذه الفقرات	النموذج الثاني الذي يشتمل على 20% من الفقرات ذات الأداء التفاضلي بلاضافة إلى أرقام هذه الفقرات	النموذج الثالث الذي يشتمل على 40% من الفقرات ذات الأداء التفاضلي بلاضافة إلى أرقام هذه الفقرات
		34
35	35	35
39	39	39
		40
42	42	42
	44	44
		50

كيفية اختيار الفقرات الموجودة في الجدول (1) لكي تظهر أداءً تفاضلياً.

لقد تم اختيار الفقرات لتظهر أداءً تفاضلياً من خلال ترتيب معاملات الصعوبة للفقرات في المجموعة المرجعية(التي تمثل القاعدة الاساسية للمقارنة) تصاعدياً من الأدنى إلى الأعلى والبالغ عددها(50) فقرة ، ومن ثم تم تقسيمها إلى خمسة مستويات، واختير من كل مستوى الفقرات التي تتوسط الترتيب من حيث الصعوبة لكي تظهر أداء تفاضلي، كما هو مبين في الجدول (1).

وبعد اختيار الفقرات لكي تظهر أداءً تفاضلياً تم توليد الأداء التفاضلي لهذه الفقرات من خلال الزيادة في قيمة معلمة الصعوبة (b) في المجموعة المستهدفة. ونتيجةً لذلك فإن هذه الفقرة تكون أصعب للمجموعة المستهدفة عبر جميع مستويات القدرة ولصالح المجموعة المرجعية. وحسب تصميم الدراسة المستخدم يكون هنالك ثلاث حالات لنسب الفقرات ذات الاداء التفاضلي في الاختبار هي (10%,20%,40%). ويكون هنالك أيضاً مستويات مختلفة من الأداء التفاضلي للفقرة وهي (متدنٍ بمقدار 0.5، وعالٍ بمقدار 1.5).

وبناءً على ذلك؛ فإن التصميم المستخدم في هذه الدراسة قد تكون من (6) مجموعات مستهدفة كل مجموعة لها خصائص مختلفة بناءً على نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة، كما هو مبين في الجدول (2)، بلإضافة إلى دراسة مجموعة واحدة تخلو من أي فقرات ذات أداء تفاضلي، وتمثل المجموعة المرجعية(القاعدة الأساسية للمقارنة).

الجول(2).نسبة الفقرات ذات الأداء التفاضلي، ومستوى الأداء التفاضلي للفقرة لدى المجموعات المستهدفة.

نسبة الفقرات ذات الأداء التفاضلي في الاختبار	مستوى الأداء التفاضلي للفقرة	
	مستوى متدنٍ بمقدار 0.5	مستوى عالٍ بمقدار 1.5
%10	C1	C2
%20	C3	C4
%40	C5	C6

يمثل الرمز Ci (Cisimulation conditions) مجموعات توليد البيانات في الدراسة أو تصميم الدراسة، الذي يتكون من (7) مجموعات توليد للبيانات، وهي:

- ست مجموعات مستهدفة تشتمل على فقرات ذات أداء التفاضلي تتمثل في (3 حالات لنسب الفقرات ذات الأداء التفاضلي * حالتين لمستوى الأداء التفاضلي للفقرة)
- مجموعة واحدة لا تشتمل على وجود أي فقرة ذات أداء تفاضلي، والتي تمثل القاعدة الأساسية للمقارنة (المجموعة المرجعية).

عرض النتائج ومناقشتها

كان الغرض من هذه الدراسة تبيان أثر الفقرات ذات الأداء التفاضلي المنتظم في ضوء مستوى الأداء التفاضلي للفقرة ونسبة الفقرات ذات الأداء التفاضلي الموجودة في الاختبار في تقدير معالم الفقرات والافراد باستخدام النموذج اللوجستي ثلاثي المعلمة.

1. النتائج المتعلقة بمدى مطابقة الأفراد ومطابقة الفقرات للنموذج اللوجستي ثلاثي المعلمة تبعاً لنسبة الفقرات ذات

الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة؟"؛ للاجابة عن هذا السؤال، فقد تم القيام بالخطوات الآتية للتحقق من اجراءات المطابقة وفقاً للنموذج اللوجستي ثلاثي المعلمة:

أ. مطابقة الأفراد لبيانات الاختبار في كل مجموعة وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، في ضوء النموذج اللوجستي ثلاثي المعلمة:

استخدم برنامج (Bilog-MG3) للكشف عن مطابقة الأفراد تبعاً لمتغير نسبة الفقرات ذات الأداء التفاضلي في الاختبار، وهي (10%,20%,40%)، ومستوى الأداء التفاضلي (0.5,1.5). حيث تم حذف الأفراد غير المطابقين في كل مجموعة، ثم تم التحقق من دلالة الفرق بين نسب الأفراد المطابقين للنموذج الثلاثي المعلمة في نظرية الاستجابة للفقرة ما بين مجموعات الدراسة، كما في الجدول (3).

الجدول (3): أعداد وأرقام الأفراد غير المطابقين في كل مجموعة وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، ونتائج الفرق بين نسب الأفراد المطابقين عبر جميع المجموعات.

حالة الفقرة	تكرار الأفراد غير المطابقين	أرقام الأفراد غير المطابقين	التكرار المتوقع	الباقي	كاي^2 الفرعية	درجة الحرية	الدلالة الإحصائية
المرجعية	1	379	1.57	-0.57	0.21	1	0.649
المستهدفة (النسبة 10%، المستوى 0.5)	0		1.57	-1.57	1.57	1	0.210
المستهدفة (النسبة 10%، المستوى 1.5)	2	387،299	1.57	0.43	0.12	1	0.732
المستهدفة (النسبة 20%، المستوى 0.5)	0		1.57	-1.57	1.57	1	0.210
المستهدفة (النسبة 20%، المستوى 1.5)	0		1.57	-1.57	1.57	1	0.210
المستهدفة (النسبة 40%، المستوى 0.5)	4	2, 140, 251, 490	1.57	2.43	3.75	1	0.053
المستهدفة (النسبة 40%، المستوى 1.5)	4	275, 298, 363, 430	1.57	2.43	3.75	1	0.053
الكلي	**11**		**11**	**0**	**12.55**	**6**	**0.051**

أظهرت نتائج التحليل المبين في الجدول (3) أعداد وأرقام الأفراد غير المطابقين في كل مجموعة, حيث تبين من خلال الجدول أن أعداد الأفراد غير المطابقين لا يتجاوز (4) أفراد من أصل (510) أفراد في كل مجموعة. فقد تم حذف الأفراد غير المطابقين في كل مجموعة، وذلك بسبب عدم مطابقتهم للنموذج ثلاثي المعلمة لتعذر حساب الخطأ المعياري في تقدير القدرة لهم عند مستوى قدراتي ما.

كما يُلاحظ من خلال الجدول (3)، أنه لا توجد فروق ذات دلالة إحصائية بين نسب الأفراد المطابقين للنموذج الثلاثي المعلمة بين جميع المجموعات تُعزى لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، حيث بلغت احتمالية الدلالة الإحصائية (0.051).

ب. مطابقة الفقرات في كل مجموعة وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، في ضوء النموذج ثلاثي المعلمة.

تم التحقق من مطابقة فقرات الاختبار لنموذج الاستجابة للفقرة ثلاثي المعلمة للبيانات المولدة في كل مجموعة من المجموعات تبعاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، حيث تم حذف الفقرات غير المطابقة من بيانات الاختبار في كل مجموعة، والجدول (4) يبين إحصائيات الفقرات غير المطابقة.

الجدول (4): إحصائيات مطابقة الفقرات للمجموعات التي ظهر لديها فقرات غير مطابقة وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة.

حالة الفقرة في المجموعة:	رقم الفقرة غير مطابقة	معلمة التمييز	الخطأ المعياري لمعلمة التمييز	معلمة الصعوبة	الخطأ المعياري لمعلمة الصعوبة	معلمة التخمين	الخطأ المعياري لمعلمة التخمين	χ^2	درجة الحرية	احتمالية المطابقة
المستهدفة (النسبة 10%، المستوى 0.5)	22	1.145	0.220	-0.276	0.318	0.365	0.085	**36.1**	8	0.000
المستهدفة (النسبة 20%، المستوى 0.5)	1	1.476	0.459	1.126	0.218	0.487	0.045	**26.5**	9	0.002
المستهدفة (النسبة 40%، المستوى 0.5)	37	1.890	0.297	-0.516	0.184	0.313	0.068	**36.5**	8	0.000
المستهدفة (النسبة 40%، المستوى 1.5)	16	1.128	0.285	0.440	0.298	0.382	0.074	**22.0**	9	0.009
	50	2.124	0.346	0.282	0.119	0.242	0.045	**24.3**	8	0.002

يُلاحظ من نتائج مطابقة الفقرات في جدول (4)، ارقام الفقرات غير المطابقة والمجموعة التي تنتمي اليها وتم حذف جميع هذه الفقرات حيث كانت احتمالية المطابقة لجميع هذه الفقرات أقل من (0.01)، وهذا مؤشر على عدم مطابقة هذه الفقرات. وبالنسبة لباقي المجموعات فقد كانت جميع فقراتها مطابقة للنموذج.

كما يُلاحظ من الجدول(4) أن أعداد الفقرات غير المطابقة تراوحت عبر جميع المجموعات من (2-0) فقرة من أصل (50) فقرة في كل مجموعة، ويتبين من خلال ذلك أنه لا يوجد فروق في مطابقة الفقرات بين المجموعات تُعزى إلى نسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة. وتدل هذه النتائج المتعلقة بمطابقة الفقرات والأفراد للنموذج اللوجستي ثلاثي المعلمة، على أن نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة، لم يكن لهما أثر دال إحصائياً على مطابقة الفقرات والأفراد. كما تؤكد هذه النتيجة قوة الإجراءآت لمطابقة الفقرات والأفراد التي تقوم عليها نظرية الاستجابة للفقرة، ولأن البيانات أصلاً تم توليدها بما يتطابق مع النموذج ثلاثي المعلمة.

2. النتائج المتعلقة بمدى تحقق افتراضات النموذج (احادية البعد والاستقلال الموضعي) في الاختبار تبعاً لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة؟"؛ للاجابة عن هذا السؤال فقد تم القيام بالخطوات الآتية:

أ. التحقق من إفتراض أحادية البعد (Unidimensionality) لبيانات الإختبار في كل مجموعة من المجموعات بإستخدام برنامج NOHARM (Normal Ogive Harmonic Analysis Robust Method) ، ويبين الجدول (5) نتائج التحليل العاملي لكل مجموعة.

الجدول (5): نتائج التحليل العاملي لفقرات الإختبار النهائية في كل مجموعة وفقاً لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي.

حالة الفقرة	مجموع مربعات البواقي (دون القطر الرئيسي)	الجذر التربيعي لمتوسط مجموع مربعات البواقي (دون القطر الرئيسي)		مؤشر Tanaka لحسن المطابقة	عدد الأفراد المطابقين
		قيمته	المعيار		
المرجعية	0.09285	0.00871	**0.18173**	0.97736	509
المستهدفة (النسبة 10%، المستوى 0.5)	0.08520	0.00851	**0.18155**	0.98828	510
المستهدفة (النسبة 10%، المستوى 1.5)	0.09422	0.00877	**0.18191**	0.97986	508
المستهدفة (النسبة 20%، المستوى 0.5)	0.08697	0.00860	**0.18155**	0.98054	510
المستهدفة (النسبة 20%، المستوى 1.5)	0.09066	0.00860	**0.18155**	0.98859	510
المستهدفة (النسبة 40%، المستوى 0.5)	0.09583	0.00903	**0.18227**	0.98629	506
المستهدفة (النسبة 40%، المستوى 1.5)	0.08637	0.00875	**0.18227**	0.98172	506

أظهر التحليل في الجدول (5) بيانات عينة التطبيق النهائي المطابقة للنموذج ثلاثي المعلمة والمتعلقة بإستجابات الأفراد في كل مجموعة من المجموعات المستهدفة والمجموعة المرجعية، وتم التحقق من إفتراض أحادية البعد في كل مجموعة من خلال مؤشرين هما: مؤشر تاناكا (TANAKA)، فقد كانت أقل قيمة لهذا المؤشر(0.977)، وكانت لدى المجموعة المرجعية، وأعلى قيمة لهذا المؤشر(0.9885) وكانت لدى المجموعة المستهدفة التي تتمثل في أن (20% من فقراتها ذات أداء تفاضلي ومستوى الاداء التفاضلي لهذه الفقرات 1.5)، وهما قيمتان تشيران إلى تحقق إفتراض أحادية البعد وفقاً لـ (Jaspar, 2010)، وحيث كانت جميع القيم أكبر من (0.94)، والمؤشر الآخر؛ هو مؤشر الجذر التربيعي لمتوسط مربعات البواقي (RMSR)، وقد بلغت أقل قيمة محسوبة هي (0.00850) وكانت لدى المجموعة المرجعية وهي أقل من القيمة الحرجة لها البالغة قيمتها (0.18190) التي تحسب من خلال المعادلة $(4.1/\sqrt{509})$، وكذلك بلغت أعلى قيمة محسوبة لـ (RMSR) هي (0.0090) لدى المجموعة المستهدفة التي تتمثل في أن (40%من فقراتها ذات أداء تفاضلي ومستوى الاداء التفاضلي لهذه الفقرات 0.5)، وهي أقل من القيمة الحرجة لهُ البالغة قيمتها (0.1817) التي تحسب من خلال المعادلة $(4.1/\sqrt{509})$، وعند مقارنة القيمتان الناتجتان من خلال التحليل مع أعلى قيمة لهذا المؤشر تبين أن القيمتان الناتجتان تقتربان من الصفر وهذا مؤشر على تحقق إفتراض أحادية البعد (Jasper, 2010). ويُلاحظ من خلال الجدول أن أعلى فرق بين جميع الحالات لمؤشر تاناكا كانت أقل من (0.01) بمعنى لم يكن هنالك فروق ظاهرية لمؤشر أحادية البعد بين جميع المجموعات المستهدفة والمجموعة المرجعية يُعزى لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة.

ب. التحقق من إفتراض الإستقلال الموضعي (Local Independence) لبيانات الإختبار في كل مجموعة باستخدام برنامج LDID (Local Dependence Indices for Dichotomous Items) على بيانات العينة النهائية المطابقة للنموذج الثلاثي في كل مجموعة، وذلك كما هو مبين في الجدول (6).

الجدول (6). مؤشرات الإستقلال الموضعي في كل مجموعة من المجموعات المستهدفة والمجموعة المرجعية وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة.

حالة الفقرة	حالة الاستقلال الموضعي				النسبة المتوقعة	الكلي		الباقي	كا2 الفرعية	درجة الحرية	الدلالة الإحصائية
	معتمد		مستقل								
	عدد الأزواج	النسبة المئوية	عدد الأزواج	النسبة المئوية		عدد الأزواج	النسبة المئوية				
المرجعية	76	6.2	1149	93.8	92.1	1225	100.0	1.67	0.030	1	0.862
المستهدفة (النسبة 10%، المستوى 0.5)	93	7.9	1083	92.1	92.1	1176	100.0	-0.03	0.000	1	0.998
المستهدفة (النسبة 10%، المستوى 1.5)	77	6.3	1148	93.7	92.1	1225	100.0	1.57	0.027	1	0.870
المستهدفة (النسبة 20%، المستوى 0.5)	111	9.4	1065	90.6	92.1	1176	100.0	-1.53	0.025	1	0.873
المستهدفة (النسبة 20%، المستوى 1.5)	71	5.8	1154	94.2	92.1	1225	100.0	2.07	0.047	1	0.829
المستهدفة (النسبة 40%، المستوى 0.5)	85	7.2	1091	92.8	92.1	1176	100.0	0.67	0.005	1	0.944
المستهدفة (النسبة 40%، المستوى 1.5)	139	12.3	989	87.7	92.1	1128	100.0	-4.43	0.213	1	0.645
الكلي			**7679**	**644.9**	**644.9**			**0.000**	**0.347**	**6**	**0.999**

يُلاحظ من الجدول (6)، أن النتائج الخاصة به قد كانت على النحو الآتي:

عدم وجود إنتهاك لافتراض الاستقلال الموضعي للفقرات في كل المجموعات، فقد كانت أعلى قيمة لعدد الازواج المرتبطة التي وقعت خارج المدى (139) بنسبة (12.3%)، بينما كان عدد أزواج الفقرات المستقلة التي وقعت ضمن المدى (989) بنسبة (87.7%)، ويتبين من هذا أن عدد أزواج الفقرات التي حققت الإستقلالية أعلى بـ (8) أضعاف تقريباً من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر قوي على تحقق إفتراض الإستقلال الموضعي. وكانت أدنى

قيمة لعدد الأزواج المرتبطة التي وقعت خارج المدى (71) زوج بنسبة (5.8%)، بينما كان عدد أزواج الفقرات المستقلة التي وقعت ضمن المدى (1154) بنسبة (94.2%)، ويتبين من هذا أن عدد أزواج الفقرات التي حققت الإستقلالية أعلى بـ (16) ضعف تقريباً من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر قوي على تحقق إفتراض الإستقلال الموضعي. كما يُلاحظ من خلال الجدول عدم وجود فروق ذات دلالات إحصائية لمؤشرات الإستقلال الموضعي بين جميع المجموعات تُعزى لنسبة الفقرات ذات الاداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، حيث كانت قيمة إحتمالية الدلالة الإحصائية (0.99). وتدل هذه النتائج على عدم وجود إنتهاك لافتراض الاستقلال الموضعي وأحادية البعد يُعزى لنسبة الفقرات ذات الاداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، ويعود السبب في ذلك إلى قوة هذه الافتراضات التي تستند اليها نظرية الاستجابة للفقرة.

3. النتائج المتعلقة بالسؤال الثالث الذي نصَّ على:" **هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α=0.05) بين متوسطات تقديرات معالم فقرات الاختبار (الصعوبة والتمييز والتخمين) تُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة أو التفاعل بينهما؟**"

للاجابة عن هذا السؤال فقد تم تقدير معالم فقرات الاختبار(الصعوبة والتمييز والتخمين) والأخطاء المعيارية لها في كل مجموعة من المجموعات المستهدفة والمجموعة المرجعية، علاوةً على حساب قيمة اختبار χ^2 لحسن المطابقة لكل فقرة من فقرات الاختبار. ولأغراض المقارنة فقد تم إيجاد قيم المتوسطات الحسابية والانحرافات المعيارية لكل معلمة من معالم فقرات الاختبار (التمييز والصعوبة والتخمين)، بالإضافة الى إجراء تحليل التباين الثنائي، وذلك على النحو التالي:

أ. فيما يتعلق بمعلمة الصعوبة

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعلمة الصعوبة لفقرات الاختبار في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة، وذلك كما هو مبين في الجدول (7).

الجدول(7) . الوسط الحسابي والانحراف المعياري لمعلمة الصعوبة لفقرات الاختبار في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة.

نسبة الفقرات ذات الاداء التفاضلي في الاختبار	مستوى الأداء التفاضلي للفقرة	المتوسط الحسابي لمعلمة الصعوبة	الانحراف المعياري لمعلمة الصعوبة
المرجعية	المرجعية	0.273	0.41
	الكلي	0.273	0.41
%10	0.5	0.344	0.48
	1.5	0.420	0.56
	الكلي	3820.	0.52
%20	0.5	0.312	0.50
	1.5	0.493	0.58
	الكلي	0.403	0.54
%40	0.5	0.540	0.49
	1.5	0.672	0.78
	الكلي	0.606	0.65
الكلي	المرجعية	0.382	0.41
	0.5	0.419	0.50
	1.5	0.477	0.66

يلاحظ من الجدول (7)، وجود فروق ظاهرية بين المتوسطات الحسابية لمعلمة الصعوبة لفقرات الاختبار ناتجة عن اختلاف مجموعات متغيري الدراسة، حيث كان أعلى قيمة للوسط الحسابي والانحراف المعياري لمعلمة الصعوبة لفقرات الاختبار في المجموعة المستهدفة التي تتمثل في أن (40% من فقراتها ذات أداء تفاضلي ومستوى الأداء التفاضلي لهذه الفقرات 1.5)، وبلغت القيم على الترتيب (0.672,0.78)، وكانت أقل قيمة للوسط الحسابي والانحراف المعياري لمعلمة الصعوبة عند المجموعة المرجعية، وقد بلغت القيم بنفس الترتيب (0.273,0.41). كما يُلاحظ من خلال الجدول بأنه كلما زاد نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة زادت قيمة الوسط الحسابي والانحراف المعياري لمعلمة الصعوبة، وللتحقق من جوهرية الفروق الظاهرية؛ تم إجراء تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة الصعوبة لفقرات الاختبار وفقاً لمتغيري الدراسة، وذلك كما في الجدول (8).

الجدول (8). نتائج تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة الصعوبة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
نسبة الفقرات ذات الأداء التفاضلي	3.803	2	1.902	**6.242**	0.002
مستوى الأداء التفاضلي للفقرة	0.268	1	0.268	0.880	0.349
التفاعل	1.400	2	0.700	2.298	0.102
الخطأ	102.965	338	0.305		
الكلي	108.603	344			

يتضح من الجدول (8)، وجود فروق دالة إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية لمعلمة الصعوبة تُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار. أما مستوى الأداء التفاضلي للفقرة والتفاعل بين المتغيرين فلم يكن ذا دلالة إحصائية. ولكون المتغير (نسبة الفقرات ذات الأداءالتفاضلي) متعدد المجموعات؛ فقد تم إجراء اختبار Levene للتحقق من إنتهاك تجانس التباين للمتوسطات الحسابية لمعلمة الصعوبة؛ بهدف تحديد أنسب اختبار مقارنات بعدية للكشف عن جوهرية الفروق بين المتوسطات الحسابية لمعلمة الصعوبة. وقد بلغت قيمة ف المحسوبة الخاصة باختبار Levene (3.788) وهي قيمة دالة إحصائياً عند مستوى الدلالة (α=0.05) عند درجتي حرية (6 للبسط و338 للمقام)، مما يُعني وجود إنتهاك لتجانس التباين لمعلمة الصعوبة يُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار؛ بما يفيد ضرورة إجراء اختبار Games-Howell للمقارنات البعدية المتعددة للكشف عن جوهرية الفروق بين المتوسطات الحسابية لمعلمة الصعوبة وفقاً لنسبة الفقرات ذات الأداء التفاضلي، وذلك كما في الجدول (9).

الجدول (9). نتائج اختبار Games-Howell للكشف عن مواقع الفروق بين المتوسطات الحسابية لمعلمة الصعوبة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار.

نسبة الفقرات ذات الأداء التفاضلي		المرجعية	%10	%20	%40
Games-Howell	المتوسط الحسابي	0.273	0.382	0.403	0.606
المرجعية	2730.				
%10	0.382	0.109			
%20	0.403	0.130	0.021		
%40	0.606	**0.267**	0.224	0.203	

يتضح من الجدول (9) بأنه لا يوجد إلا فرقاً واحداً دال إحصائياً بين المتوسطات الحسابية لمعلمة الصعوبة وكان بين المجموعة المستهدفة التي تشتمل على 40% من الفقرات ذات الاداء التفاضلي وبين المجموعة المرجعية.

حيث يتبين من خلال النتائج السابقة المتعلقة بمعلمة الصعوبة، بأنه كلما زادت نسبة الفقرات ذات الأداء التفاضلي زادت قيمة الوسط الحسابي والانحراف المعياري لمعلمة الصعوبة. وهذه النتيجة منطقية فمن المعلوم أن الفقرات ذات الأداء التفاضلي تكون أكثر صعوبةً للمجموعة المستهدفة مقارنةً مع المجموعة المرجعية، وبالتالي كلما زادت نسبة الفقرات ذات الأداء التفاضلي في الاختبار زادت قيمة المتوسط الحسابي لمعلمة صعوبة الفقرات في الاختبار.

ب. فيما يخص معلمة التمييز:

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعلمة التمييز لفقرات الاختبار وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، وذلك كما هو مبين في الجدول (10).

الجدول (10). الوسط الحسابي والانحراف المعياري لمعلمة التمييز لفقرات الاختبار في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة.

نسبة الفقرات ذات الأداء التفاضلي	مستوى الأداء التفاضلي للفقرة	المتوسط الحسابي لمعلمة التمييز	الانحراف المعياري لمعلمة التمييز
المرجعية	المرجعية	2.069	0.69
	الكلي	2.069	0.69
%10	0.5	2.007	0.70
	1.5	1.757	0.49
	الكلي	1.883	0.62
%20	0.5	1.798	0.64
	1.5	1.817	0.56
	الكلي	1.808	0.60
%40	0.5	1.905	0.70
	1.5	1.591	0.55
	الكلي	1.746	0.64
الكلي	المرجعية	2.069	0.69
	0.5	1.903	0.68
	1.5	1.721	0.54

يُلاحظ من الجدول (10)، وجود فروق ظاهرية بين المتوسطات الحسابية لمعلمة التمييز لفقرات الاختبار ناتجة عن اختلاف مجموعات متغيري الدراسة؛ حيث بلغت أعلى قيمة للوسط الحسابي لمعلمة التمييز للفقرات (2.069)، وكانت لدى المجموعة المرجعية، وأقل قيمة للوسط الحسابي لمعلمة التمييز بلغت (1.591)، وكانت لدى المجموعة المستهدفة التي تشتمل على (40% من الفقرة ذات أداء تفاضلي ومستوى الأداء التفاضلي1.5) بفارق مقداره (0.48)، والفرق يعود لصالح المجموعة المرجعية. حيث يُلاحظ من خلال ذلك عند إشتمال الاختبار على فقرات ذات أداء تفاضلي فإن معدل التمييز لهذه الفقرات سينخفض، ويزداد مقدار الإنخفاض بزيادة مستوى الأداء التفاضلي للفقرة، وللتحقق من جوهرية الفروق الظاهرية؛ تم إجراء تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة التمييز وفقاً لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، وذلك كما في الجدول (11).

الجدول(11): نتائج تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة التمييز لفقرات الاختبار حسب نسبة الفقرات ذات الأداءالتفاضلي ومستوى الأداء التفاضلي للفقرة.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
نسبة الفقرات ذات الأداء التفاضلي	0.881	2	0.441	1.133	0.323
مستوى الأداء التفاضلي للفقرة	2.437	1	2.437	**6.266**	0.013
التفاعل	1.551	2	0.775	1.994	0.138
الخطأ	131.466	338	0.389		
الكلي	139.174	344			

يتضح من الجدول (11)، وجود فروق دالة إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية لمعلمة التمييز تُعزى لمستوى الأداء التفاضلي للفقرة، أما متغير نسبة الفقرات ذات الأداء التفاضلي والتفاعل بين المتغيرين فلم تكن ذات دلالة إحصائية، ولكون المتغير مستوى الأداء التفاضلي للفقرة متعدد المجموعات؛ فقد تم إجراء اختبار Levene للتحقق من إنتهاك تجانس التباين للمتوسطات الحسابية لمعلمة التمييز لفقرات الاختبار وفقاً لمستوى الأداء التفاضلي للفقرة؛ بهدف تحديد أنسب اختبار مقارنات بعدية للكشف عن جوهرية الفروق بين المتوسطات الحسابية لمعلمة التمييز، حيث تبين عدم وجود دلالة إحصائية عند مستوى الدلالة (α=0.05)، مما يُعني عدم وجود إنتهاك لتجانس التباين لمعلمة التمييز يُعزى لمستوى الأداء التفاضلي للفقرة؛ بما يفيد ضرورة إجراء اختبار Scheffe للمقارنات البعدية المتعددة للكشف عن جوهرية الفروق بين المتوسطات الحسابية لمعلمة التمييز، وذلك كما في الجدول (12).

الجدول (12). نتائج اختبار Scheffe للكشف عن موا قع الفروق بين المتوسطات الحسابية لمعلمة التمييز لفقرات الاختبار وفقاً لمستوى الأداء التفاضلي للفقرة.

مستوى الأداء التفاضلي للفقرة		1.5	0.5	المرجعية
Scheffe	المتوسط الحسابي	1.721	1.903	2.069
1.5	1.721			
0.5	1.903	**0.182**		
المرجعية	2.069	**0.347**	0.165	

يتضح من الجدول (12)، وجود فرق دال إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية لمعلمة التمييز لفقرات الاختبار وفقاً لمتغير الدراسة (مستوى الأداء التفاضلي للفقرة)؛ حيث تأثر قيمة الوسط الحسابي لمعلمة التمييز للفقرات في الاختبار المشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (1.5) تأثراً جوهرياً، بحيث انخفضت قيمة الوسط الحسابي لمعالم تمييز الفقرات في الاختبار مقارنةً المجموعة المرجعية، كما ويوجد فرق ذو دلالة إحصائية بين المجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (0.5) والمجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (1.5).

يُلاحظ من خلال بيانات الجدول (12) أنه عند إشتمال الاختبار على فقرات ذات أداء تفاضلي فإن معدل التمييز لفقرات الاختبار سينخفض، ويزداد مقدار الإنخفاض بزيادة مستوى الأداء التفاضلي للفقرة. ويعود السبب في ذلك؛ بأن الفقرات ذات الأداء التفاضلي تكون أكثر صعوبةً مما يترتب على ذلك إنخفاض معدل التمييز. حيث أشار فريسبي (Frisbie,1973) بأن مدى الصعوبة يؤثر على الدلالة التمييزية لفقرات الاختبار، وأن الفقرات متوسطة الصعوبة يكون تمييزها أعلى منه في حالة الفقرات السهلة جداً أو الصعبة جداً، وهذا يتفق مع ما جاءت به هذه الدراسة؛ بأنه كلما زاد مستوى الأداء التفاضلي للفقرة أزداد معدل الصعوبة لفقرات الاختبار،

مما ترتب على ذلك انخفاض معدل التمييز.

ج. فيما يتعلق بمعلمة التخمين:

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعلمة التخمين لفقرات الاختبار في كل مجموعة من مجموعات الدراسة وفقاً لمتغيري الدراسة، وذلك كما هو مبين في الجدول (13).

الجدول (13). الوسط الحسابي والانحراف المعياري لمعلمة التخمين لفقرات الاختبار في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة.

نسبة الفقرات ذات الأداء التفاضلي في الاختبار	مستوى الأداء التفاضلي للفقرة	المتوسط الحسابي لمعلمة التخمين	الانحراف المعياري لمعلمة التخمين
المرجعية	المرجعية	0.380	0.08
	الكلي	0.380	0.08
%10	0.5	0.381	0.07
	1.5	0.342	0.06
	الكلي	0.362	0.07
%20	0.5	0.361	0.05
	1.5	0.358	0.08
	الكلي	0.360	0.06
%40	0.5	0.382	0.07
	1.5	0.332	0.07
	الكلي	0.357	0.08
الكلي	المرجعية	0.380	0.08
	0.5	0.375	0.07
	1.5	0.344	0.07

يُلاحظ من الجدول (13)، وجود فروق ظاهرية بين المتوسطات الحسابية لمعلمة التخمين لفقرات الاختبار ناتجة عن اختلاف مجموعات متغيري الدراسة، وللتحقق من جوهرية الفروق الظاهرية؛ تم إجراء تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة التخمين وفقاً لمتغيري الدراسة، وذلك كما في الجدول (14).

الجدول (14). نتائج تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة التخمين لفقرات الاختبار حسب نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف	الدلالة الإحصائية
نسبة الفقرات ذات الأداء التفاضلي	0.001	2	0.001	0.113	0.893
مستوى الأداء التفاضلي للفقرة	0.068	1	0.068	**14.415**	0.000
التفاعل	0.014	2	0.007	1.513	0.222
الخطأ	1.618	338	0.005		
الكلي	1.720	344			

يتضح من الجدول (14)، وجود فروق دالة إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية لتقديرات معلمة التخمين لفقرات الاختبار تُعزى لمستوى الأداء التفاضلي للفقرة، أما متغير نسبة الفقرات ذات الأداء التفاضلي والتفاعل

بين المتغيرين فلم يكن ذو دلالة إحصائية، ولكون متغير الدراسة (مستوى الأداء التفاضلي للفقرة) متعدد المجموعات؛ فقد تم إجراء اختبار Levene للتحقق من إنتهاك تجانس التباين للمتوسطات الحسابية لمعلمة التخمين؛ بهدف تحديد أنسب اختبار مقارنات بعدية للكشف عن جوهرية الفروق بين المتوسطات الحسابية لمعلمة التخمين، حيث تبين عدم وجود دلالة إحصائية عند مستوى الدلالة (α=0.05)، مما يُعني عدم وجود انتهاك لتجانس التباين لمعلمة التخمين يُعزى لمستوى الأداء التفاضلي للفقرة؛ بما يفيد ضرورة إجراء اختبار Scheffe للمقارنات البعدية المتعددة للكشف عن جوهرية الفروق بين المتوسطات الحسابية لمعلمة التخمين لفقرات الإحتبار، وذلك كما في الجدول (15).

الجدول (15). نتائج اختبار Scheffe للكشف عن مواقع الفروق بين المتوسطات الحسابية لمعلمة التخمين لفقرات الاختبار وفقاً لمستوى الأداء التفاضلي للفقرة.

مستوى الأداء التفاضلي		1.5	0.5	مرجعية
Scheffe	المتوسط الحسابي	0.344	0.375	0.380
1.5	0.344			
0.5	0.375	**0.030**		
مرجعية	0.380	**0.036**	0.006	

يتضح من الجدول (15)، وجود فرق دال إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية لمعلمة التخمين وفقاً لمستوى الأداء التفاضلي للفقرة؛ حيث تأثر قيمة الوسط الحسابي لمعلمة التخمين في الاختبار المشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (1.5) تأثراً جوهرياً، بحيث انخفضت قيمة الوسط الحسابي لمعالم التخمين لفقرات الاختبار مقارنة مع قيمة الوسط الحسابي لمعالم التخمين في المجموعة المرجعية، كما ويُلاحظ من خلال الجدول عدم وجود فرق ذو دلالة إحصائية للأوساط الحسابية لمعلمة التخمين بين المجموعة المرجعية والمجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (0.5)، كما ويوجد فرق ذو دلالة إحصائية للأوساط الحسابية لمعلمة التخمين بين المجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (0.5) والمجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (1.5). حيث يُلاحظ من خلال ذلك عند إشتمال الاختبار على فقرات ذات أداء تفاضلي فإن معدل التخمين لفقرات الإختبار سينخفض ، ويزداد مقدار الإنخفاض بزيادة مستوى الأداء التفاضلي للفقرة.

4. النتائج المتعلقة بالسؤال الرابع الذي نصَّ على: " **هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة (α=0.05) بين متوسطات تقديرات قدرات الأفراد على الاختبار تُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة أو التفاعل بينهما؟**"؛ للاجابة عن هذا السؤال تم حساب المتوسطات الحسابية والانحرافات المعيارية لتقديرات قدرات الأفراد على الاختبار وفقاً لمتغيري الدراسة، وذلك كما هو مبين في الجدول (16).

الجدول(16): الأوساط الحسابية والانحرافات المعيارية لتقديرات قدرات الأفراد على الاختبار في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة.

نسبة الفقرات ذات الأداء التفاضلي في الاختبار	مستوى الأداء التفاضلي للفقرة	الوسط الحسابي للقدرة	الانحراف المعياري للقدرة
المرجعية	صفر	0.02	1.07
	الكلي	0.02	1.07
%10	0.5	-0.03	1.18
	1.5	0.01	1.12
	الكلي	-0.01	1.15
%20	0.5	-0.03	1.20
	1.5	-0.02	1.14

نسبة الفقرات ذات الأداء التفاضلي في الاختبار	مستوى الأداء التفاضلي للفقرة	الوسط الحسابي للقدرة	الانحراف المعياري للقدرة
	الكلي	-0.02	1.17
%40	0.5	-0.02	1.15
	1.5	-0.06	1.23
	الكلي	-0.04	1.19
الكلي	صفر	0.02	1.07
	0.5	-0.03	1.17
	1.5	-0.02	1.16

يُلاحظ من الجدول (16)، وجود فروق ظاهرية بين الأوساط الحسابية لتقديرات قدرات الأفراد على الاختبار ناتجة عن اختلاف مجموعات متغيري الدراسة، حيث كان أعلى وسط حسابي لقدرات الأفراد لدى المجموعة المرجعية وبلغت قيمته (0.02)، وأدنى قيمة للوسط الحسابي لقدرات الأفراد يقع لدى المجموعة المستهدفة التي تشتمل على (40% من فقرات الاختبار ذات أداء تفاضلي ومستوى الأداء التفاضلي لهذه الفقرات 1.5) وبلغت قيمته (0.06-) بفارق مقداره (0.08). ويُلاحظ من خلال الجدول أنه كلما زادت نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة قل الوسط الحسابي لقدرات الأفراد، وللتحقق من جوهرية الفروق الظاهرية؛ تم إجراء تحليل التباين الثنائي لتقديرات قدرات الأفراد على الاختبار وفقاً لمتغيري الدراسة، وذلك كما في الجدول (17).

الجدول (17). نتائج تحليل التباين الثنائي لتقديرات قدرات الأفرادعلى الاختبار حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
نسبة الفقرات ذات الأداء التفاضلي	0.474	2	0.237	0.178	0.837
مستوى الأداء التفاضلي للفقرة	0.004	1	0.004	0.003	0.958
التفاعل	1.009	2	0.504	0.378	0.685
الخطأ	4737.704	3551	1.334		
الكلي	4740.211	3557			

يتضح من الجدول (17)، عدم وجود فروق دالة إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية لتقديرات قدرات الأفراد على الاختبار تُعزى لمتغيري الدراسة والتفاعل بينهما. وتتفق هذه النتيجة مع الافتراضات التي تقوم عليها نظرية الاستجابة للفقرة، وهي أن تقدير القدرة (θ) متحررة من عينة الفقرات. وكما أشار هاملتون وسوامينثان (Hambelton & Swaminathan, 1985) بأن نظرية الاستجابة للفقرة تفترض بأنه يتم الحصول على نفس التقديرات لمعالم قدرات الأفراد بغض النظر سواءً كانت الفقرات سهلة أم صعبة، وهذه النتيجة تتفق مع النتيجة التي أجراها زهان وزومبو(Zhan & Zumbo, 2009). والتي تشير إلى أن الأداء التفاضلي للفقرة لا يؤثر على تقديرات قدرات الأفراد وإنما يؤثر على العلامات الملاحظة للأفراد.

كما تم حساب المتوسطات الحسابية والانحرافات المعيارية للعلامات الملاحظة للأفراد على الاختبار في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة. وذلك كما هو مبين في الجدول (18).

الجدول (18). الأوساط الحسابية والانحرافات المعيارية للعلامات الملاحظة للأفراد على الاختبار في كل مجموعة وفقاً نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة

نسبة الفقرات ذات الأداء التفاضلي	مستوى الأداء التفاضلي للفقرة	العدد	الوسط الحسابي	الانحراف المعياري
المرجعية	صفر	509	64.354	18.91
	الكلي	509	64.354	18.91
%10	0.5	510	64.122	18.77
	1.5	508	63.917	18.05
	الكلي	1018	64.020	18.40
%20	0.5	510	64.126	18.58
	1.5	510	61.365	17.86
	الكلي	1020	62.745	18.27
%40	0.5	506	61.773	18.06
	1.5	505	57.005	16.82
	الكلي	1011	59.391	17.61
الكلي	صفر	509	64.354	18.91
	0.5	1526	63.344	18.49
	1.5	1523	60.771	17.80

يُلاحظ من الجدول (18)، وجود فروق ظاهرية بين الأوساط الحسابية لأداء الأفراد على الاختبار ناتجة عن اختلاف مجموعات متغيري الدراسة، فقد كان أعلى وسط حسابي للعلامات الملاحظة للأفراد لدى المجموعة المرجعية (64.4)، وأقل وسط حسابي كان لدى المجموعة المستهدفة التي تشتمل على (40% من فقرات الاختبار ذات أداء تفاضلي ومستوى 1.5)، وقد بلغت قيمتة (57)، ويتبين من خلال ذلك أن الوسط الحسابي للعلامات الملاحظة في المجموعات يقل بشكل نمطي بزيادة نسبة الفقرات ذات الأداء التفاضلي ومستوى الأداء التفاضلي للفقرة مقارنةً مع المجموعة المرجعية، وللتحقق من جوهرية الفروق الظاهرية؛ تم إجراء تحليل التباين الثنائي للعلامات الملاحظة للأفراد على الاختبار وفقاً لمتغيري الدراسة، وذلك كما في الجدول(19).

الجدول (19). نتائج تحليل التباين الثنائي للعلامات الملاحظة للأفراد على الاختبار وفقاً لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ولمستوى الأداء التفاضلي للفقرة.

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف	الدلالة الإحصائية
نسبة الفقرات ذات الأداء التفاضلي	11600.381	2	5800.190	**17.582**	0.000
مستوى الأداء التفاضلي للفقرة	5065.623	1	5065.623	**15.355**	0.000
التفاعل	2654.668	2	1327.334	**4.024**	0.018
الخطأ	1171439.996	3551	329.890		
الكلي	1193026.849	3557			

يتضح من الجدول (19)، وجود فروق دالة إحصائياً عند مستوى الدلالة (α=0.05) بين الأوساط الحسابية للعلامات الملاحظة للأفراد على الاختبار تُعزى لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة والتفاعل بينهما، ولكون متغيري الدراسة متعددي المجموعات؛ فقد تم إجراء اختبار Levene للتحقق من إنتهاك تجانس التباين للعلامات الملاحظة للأفراد على الاختبار؛ بهدف تحديد أنسب اختبار مقارنات بعدية للكشف عن جوهرية الفروق بين الأوساط الحسابية للعلامات الملاحظة للأفراد على الاختبار، حيث بلغت قيمة ف المحسوبة الخاصة باختبار Levene (2.735) وهي قيمة دالة إحصائياً عند مستوى الدلالة (α=0.05) عند درجتي حرية (6 للبسط و 3551 للمقام)، مما يُعني وجود إنتهاك لتجانس التباين للعلامات الملاحظة للأفراد على الاختبار تُعزى لنسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة؛ بما يفيد ضرورة إجراء اختبار Games-Howell للمقارنات البعدية المتعددة للكشف عن جوهرية الفروق بين الأوساط الحسابية للعلامات الملاحظة للأفراد على الاختبار، وذلك كما في الجدولين (20) (21).

الجدول (20). نتائج اختبار Games-Howell للكشف عن مواقع الفروق للعلامات الملاحظة للأفراد على الاختبار وفقاً لنسبة الفقرات ذات الأداء التفاضلي في الاختبار.

نسبة الفقرات ذات الأداء التفاضلي		%40	%20	%10	مرجعية
Games-Howell	المتوسط الحسابي	59.391	62.745	64.020	64.354
%40	59.391				
%20	62.745	**3.354**			
%10	64.020	**4.628**	1.275		
مرجعية	64.354	**4.962**	1.608	0.334	

الجدول (21). نتائج اختبار Games-Howell للكشف عن مواقع الفروق للعلامات الملاحظة للأفراد على الاختبار وفقاً لمستوى الأداء التفاضلي للفقرة.

مستوى الأداء التفاضلي للفقرة		1.5	0.5	المرجعية
Games-Howell	المتوسط الحسابي	60.771	63.344	64.354
1.5	60.771			
0.5	63.344	**2.574**		
المرجعية	64.354	**3.583**	1.009	

يتضح من الجدول (20)، وجود فرق دال إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية للعلامات الملاحظة للأفراد على الاختبار وفقاً لنسبة الفقرات ذات الأداء التفاضلي، حيث تأثرت العلامات الملاحظة للأفراد على الاختبار المشتمل على 40% من الفقرات ذات الأداء التفاضلي تأثراً جوهرياً، بحيث تدنى أداءهم مقارنةً بمن تقدموا للاختبار في كلٍّ من المجموعات: المجموعة المرجعية، والمجموعة التي تقدمت للاختبار المشتمل على 10%من الفقرات ذات الأداء التفاضلي، والمجموعة التي تقدمت للاختبار المشتمل على 20%من الفقرات ذات الأداء التفاضلي.

كما ويتضح من الجدول (21)، وجود فروق دالة إحصائياً عند مستوى الدلالة (α=0.05) بين المتوسطات الحسابية للعلامات الملاحظة للأفراد على الاختبار وفقاً لمستوى الأداء التفاضلي للفقرة، حيث تأثرت العلامات الملاحظة للأفراد على الاختبار المشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (1.5) تأثراً جوهرياً بحيث تدنت علاماتهم الملاحظة مقارنةً بمن تقدموا للاختبار في كلٍّ من المجموعات: المجموعة المرجعية، و المجموعة التي تقدمت للاختبار المشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (0.5).

يتبين من خلال ذلك أن العامل الأهم في التأثير على العلامات الملاحظة للأفراد هو زيادة نسبة الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة، فعندما كانت نسبة الفقرات ذات الأداء التفاضلي في الاختبار(20%,10%)، ومستوى الأداء التفاضلي للفقرة (0.5)

لم يكن هنالك أثر ذا دلالة احصائية على العلامات الملاحظة للأفراد، حيث ظهر الأثر عندما كانت نسبة الفقرات ذات الأداء التفاضلي في الاختبار (40%)، ومستوى الأداء التفاضلي للفقرة (1.5)، وهذه النتيجة تتفق مع دراسة زهان وزومبو (Zhan & Zumbo, 2009) التي توصلت الى أنه كلما زاد عدد الفقرات ذات الأداء التفاضلي في الاختبار ومستوى الأداء التفاضلي للفقرة تزداد الفروقات في المتوسط الحسابي للعلامة الملاحظة نتيجةً لوجود الفقرات ذات الأداء التفاضلي.

واخيراً لا بد لمُعد الاختبار أن يأخذ بعين الاعتبار مسألة الأداء التفاضلي للفقرة، وأن تُحذف الفقرات ذات الأداء التفاضلي الموجودة في الاختبار، ويعاد تقدير معالم الفقرات من جديد، وذلك عندما تزيد نسبة الفقرات ذات الأداء التفاضلي في الاختبار عن (20%)، وعندما يزيد مستوى الأداء التفاضلي للفقرة عن (0.5).

المصادر والمراجع

المراجع العربية

إرحيل، راتب. (2011). **أثر أداء الفقرات التفاضلي للجنس في دقه المعادلة العامودية لاختبارأوتيس لينون للقدرة العقلية.** رسالة دكتوراة غير منشورة، جامعة اليرموك، اربد– الاردن.

المراجع الأجنبية

Driana, E. (2007). *Gender Diffrential item functioning on A Ninth-Grade Mathematics proficiency Test in Applachin ohio*. unpublished Doctoral Dissertation, ohio university.

Engelhard, G. & Gamer, M. (1999). Gender difference in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Eeducation,12(1),29-43.*

Frary, R. & Zimmer, M. (1984). Eliminating of Bias in tests Scores: Effect on Reliability. *Educational and psychological Measurment, 44(1),25-31.*

Frisbie, J. (1973). Multiple –Choice Versus True-False: A comparison of Reliability and Concurrent Validities. *Journal of Educational Measurement,10(4),247-304*.

Hambleton, R. & Rogers, J. (1989). Detecting potentially bias test item: comparison of IRT aria and mantel-haenszel methods. *Applied Measurement in Education,2(4),313-334.*

Hambleton, R. & Swaminathan, H. (1985). *Item Response theory*: principles and applications. Boston, MA: Kluwek NIJ publishing.

Hambleton, R. & Swaminathan, H. & Rogers, J. (1991). *Fundamentals of item response theory*. Newbury perk California: stage publications.

Jensen, R. (1980). *Bias in mental testing*. new York: Macmillan publishing co. Inc.

Jodoin, m. & Gierl , M. & Ackerman, T. (2000). Performance of mantel-Haenszel , simultane ous item bias test , and logisitic regression when the proportion of Dif item is large. *paper presented at the annual meeting of the American educational Researeh Association Orleans*, Louisiana, April,24-27,2000.

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

pae, T. (2004). Dif for examinees with different academic backgrounds. *Language Testing,21(1),53-72.*

Raju, N. (1988). The area between two item characteristic curves. *Psychometrika*, 53,495-502.

Swaminathan, H. & Rogers, J. (1990). Detecting differential item using Logistic regression procedures. *Journal of Educational Measurement* ,127(4).

Teresi, J. (2001). Statistical methods of examination of differential item functioning with applications to cross-clutural measurement of functional, *physical and mental health. J mental Health and Aging*,7,31-4o.

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

Zumbo, B. (2003). Does item-Level Dif manifest itself in scale-Level analyses? Implications for translating languge tests. *Language Testing*,20(2),136-147.

Zumbo, B. (2000). The Effect of DIF and Impact on Classical Test Statistics: Undetected DIF and Impact, and the Reliability and Interpretability of Scores from a Language Proficiency Test. *Presented at the annual conference of the nationl counic on measurement in Education (NCME),* April, 2000 new Orleans, LA.

The Effect of Paragraphs with Differential Functioning Of Uniform on Estimating Paragraphs Parameters and Persons Using a Stimulated Data According To the Three Parameters Model

Ali Saleh Raja Nawafleh *

ABSTRACT

The study aimed to indetify the effect of Paragraphs with Differential Functioning on the Test Paragraphs Characteristics and Individual's Abilities based on three parameters model using a stimulated data study. The study used different test conditions based on the two conditions used in the current study (percentage and level of paragraphs with Differential Functioning). The percentage of Paragraphs with Differential Functioning used in the current study is (10%, 20%, 40%). The level of Paragraphs with Differential Functioning were (low 0.5, high 1.5). Accordingly, the design of the study, (2*3=6) focal groups differ among the percentage Items with Differential Functioning and the level of Differential Functioning of Paragraphs. The researcher also used a reference group that does not contain any paragraph that has Differential Functioning, accordingly, resulting in (7) study data sets. The researcher used (Wingen) statistical software to stimulate the responses of (510) examinees in each group, and the total sample consisted of (3570) examinees. Then, (Bilog MG) was used to verify the unidimensionality of the resulting and the fit of paragraphs and examinees, paragraphs and examinees parameters estimation, according to each of previous groups. Results of the study indicated significant differences in paragraph difficulty estimation due to the variables of the study (40% of Paragraphs with Differential Functioning). Significant differences were found in paragraph discrimination and guessing estimation due to the variable of the study (level of paragraphs with Differential Functioning) in favor of the reference group. There were no significant differences in examinees abilities estimation due to the variables of the study (a percentage and level of Paragraphs with Differential Functioning), and significant differences were found in the observed scores of examinees due to the variables of the study (a percentage and level of Paragraphs with Differential Functioning) and the interaction in favor of the reference group.

Keywords: percentage of Paragraphs with Differential Functioning, level of Paragraphs with Differential Functioning, reference group, focal group, Paragraphs Parameters, Ability.

*Accreditation and Quality Assurance for Higher Education Institutions. Received on 16/2/2015 and Accepted for Publication on 18/8/2015.