تقصي أثر طريقة التعويض عن البيانات المفقودة ونسبها في الكشف عن الاداء التفاضلي لفقرات اختبار باستخدام طريقة راجو

د. أيمن محمد أحمد فريحات

كلية عجلون الجامعية/ جامعة البلقاء التطبيقية/ الأردن

Investigating the Effect of Missing Data imputation Method and its Percentage on the Detection of Differential Item Functioning for Items of Test Using Raju's Method

Dr. Ayman Mohamed Ahmed Freihat

College of Ajloun University\ Balqa Applied University\ Jordan

aiman.freihat@bau.edu.jo

Abstract

The aim of this study was to investigate the effect of the method of compensation for missing data and its percentage in the detection of the differential performance of the test paragraphs using the Rajo method according to the three-parameter response model using the generated data. The design of the study included different conditions based on the study variables. The differential performance in the test is 10%, 20% and 40%. The differential performance of the paragraph is: (low level of 0.5 and high level of 1.5). Accordingly, the study was designed from (6) different groups with different performance differentials and differential performance of the paragraph, as well as a reference group that does not include any differential performance clause, which is the basis of comparison, resulting in (7) sets Study.

Keywords: compensatory value, missing data, differential performance, Rajo method.

الملخص

هدفت هذه الدراسة إلى تقصي أثر طريقة التعويض عن البيانات المفقودة و نسبها في الكشف عن الاداء التفاضلي لفقرات اختبار باستخدام طريقة راجو نسبة وفق نموذج الاستجابة للفقرة ثلاثي المعلمة باستخدام بيانات مولدة، حيث تضمن تصميم الدراسة ظروفاً مختلفة بناءً على متغيرات الدراسة، 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 تم تصميم الدراسة من (٦) مجموعات 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.

الكلمات المفتاحية: القيمة التعويضية، البيانات المفقودة، الاداء التفاضلي، طريقة راجو.

المقدمة:

حظي موضوع قياس سمات الاشخاص و خصائصهم العقلية بالبحث والاهتمام منذ القدم، وتستخدم نتائج القياس في كافة المجالات لتخدم اغراضا متعددة من التخطيط و الرصد، ويبني على نتائج القياس الكثير من القرارات، مثل تطوير عملية التعلم التعليم، ولها اهمية كبيرة في تصنيف و تسكين الاشخاص اعتماداً على نتائجهم.

وحتى يكون لدرجات الاختبارات تفسيرات صحيحة يجب أن تتميز هذه الاختبارات بدقة وصدق القياس(Kane,1996)، ومن المشاكل التي يمكن ان تواجه الباحث أثناء تطبيق المقاييس والاختبارات هي ظاهرة البيانات او الاستجابات المفقودة المتمثلة بترك بعض فقرات الاختبار او المقياس دون اجابة، فلا يكاد يخلو بحث تجريبي من بيانات مفقودة.

إن مشكلة البيانات المفقودة في الأبحاث قد تُهمل في أغلب الأحيان، وإذا لم تهمل تُعطى القليل من الانتباه والأهمية، بشكل لا يوازي حجم تأثيرها السلبي على النتائج، وتؤدي البيانات المفقودة الى تحيزات كبيرة في النتائج(Peugh & Enders, 2004)، ويؤثر ترك البيانات المفقودة دون معالجة، وتجاهل تأثيرها، إما لقصور في تقدير أثر هذه البيانات المفقودة على نتائج الدراسة، أو قلة إدراك الحلول لمشكلة البيانات المفقودة وكيفية التعامل معها، بشكل كبير على تحليل البيانات وعلى نتائج الدراسة، ويقلل من قوة الاساليب الاحصائية المستخدمة (Schafer & Graham,2002)، كما يؤدي وجود البيانات المفقودة الى العديد من المشاكل كالتقدير الخاطىء للخطأ المعياري، وزيادة معدلات الخطأ من النوع الاول (Hohensinn, & Kubinger , 2011).

مشكلة الدراسة وأسئلتها:

تعتبر ظاهرة البيانات المفقودة من الظواهر الشائعة في البحوث النفسية والتربوية، فلا يكاد يخلو بحث يتعامل مع العينات من بيانات مفقودة، فهي مشكلة عامة في هذه الابحاث (Peugh & Enders,2004)، ويؤدي اهمال التعامل مع البيانات المفقودة أو تجاهلها، الى تقديرات أقل كفاءة، و يؤدي الى تحيز نتائج الاختبارات والمقاييس المستخدمة، واضعاف القوة الاحصائية لهذه النتائج، ويمكن أن يعتبر ذلك عائقاً ومحدداً لاستخدام بعض الاساليب الاحصائية التي تشترط عدم وجود بيانات مفقودة (الدرابسة ,2012)

لذلك يجب على الباحث أن يحرص على بيانات كاملة حتى يتم الخروج بنتائج صادقة، وتكون التعميمات الخاصة بها أقرب الى الحقيقة، لأن الهدف من أي تحليل أو اجراء احصائي هو التوصل الى استدلالات صحيحة وفعالة حول مجتمع الدراسة المعني، حيث من المتوقع أن تشكل البيانات المفقودة تهديداً لهذا الهدف خاصة إذا كانت مفقودة بطريقة تجعل العينة مختلفة عن خصائص المجتمع الذي أخذت منه.

ومع زيادة الاهتمام بصدق نتائج الاختبارات ونظراً لأهمية القرارات التي تبنى عليها، زاد الاهتمام بقضية التعامل مع البيانات المفقودة، وطرق معالجتها، وخاصة مع التطور في البرامج الاحصائية الذي أدى الى تنوع الاساليب الاحصائية التي يمكن من خلالها التعامل مع البيانات المفقودة، لمساعدة الباحثين في اختيار أفضل الطرق للتعامل مع مشكلة البيانات المفقودة، وبما يكفل الوصول إلى نتائج موثوقة تؤدي إلى اتخاذ قرارات صائبة، والتي مازلت تتطلب الفحص والمقارنة حول فاعليتها ومازالت الدراسات تتباين في نتائجها حول افضل الطرق وأكثرها فاعلية للتعامل مع البيانات الفقودة (بني عواد، 2010)

ويعتبر الاداء التفاضلي للفقرة الاختبارية والمتمثل في تحيزها لصالح مجموعة دون أخرى، من القضايا الهامة ذات التأثير في صدق الاختبارات، و تلوث نتائجها، ويجعل المعلومات المستقاة منها لصالح اتخاذ القرارات مشوشة، ومشوبة بالأخطاء، وغير سليمة، وبالتالي عدم تحقيق العدالة، من القضايا التي تتأثر بوجود البيانات المفقودة حيث يقود الى نتائج مضللة لتحليلات الاداء التفاضلي (Garrett ,2009)

ولما أظهرت نتائج الدراسات التي تناولت أثر طرق التعامل مع البيانات المفقودة في دقة الكشف عن الاداء التفاضلي لفقرات الاختبارات تبايناً في أثر هذه الطرق متمثلة بالقوة الاحصائية ومعدلات الخطأ من النوع الاول سواءً بالزيادة أو والنقصان، كما أظهرت أثراً لنسبة البيانات المفقودة في دقة الكشف عن الاداء التفاضلي (Finch,2011; de Gil Patricia 2015 ; Garrett ,2009).

فإن الدراسة الحالية، جاءت لاستكمال البحث في هذا المجال من خلال تفحص فاعلية بعض طرق التعويض الضمنية عن البيانات المفقودة التي لم تتناولها الدراسات السابقة في الكشف عن الاداء التفاضلي، كما ستستخدم هذه الدراسة طريقة راجو للكشف عن الاداء التفاضلي، باعتبارها من الطرق التي تعتمد على النظرية الحديثة من خلال مقارنة منحنيات خصائص الفقرة حيث لم تستخدم اي من الدراسات السابقة هذه الطريقة عند تناولها لموضوع القيم المفقودة.

حيث تهدف هذه الدراسة لتفحص فاعلية كل من طريقة خوارزمية تعظيم التوقعات (EM) و وطريقة التعويض المتعدد (MI) و وطريقة دالة الاستجابة (RF) التي تعتبر من الطرق الضمنية للتعويض عن البيانات المفقودة في ضوء نسب فقد متعددة للاستجابات (30 %، 15%، 5%)في الكشف عن الاداء التفاضلي باستخدام طريقة راجو Area (Raju's Measures).

وبناءً على ما تقدم فإن هذه الدراسة جاءت للإجابة عن الاسئلة التالية:

١- هل توجد فروق ذات دلالة أحصائية عند مستوى (0.05) بين نسب الفقرات التي تبدي أداء تفاضليا تعزى لاختلاف طريقة التعويض عن القيم المفقودة (تعظيم التوقعات، التعويض المتعدد، دالة الاستجابة) وفق نسب فقد الاستجابات (30%، 15%، 5%، 0%)

٢- هل توجد علاقة ارتباطية ذات دلالة إحصائية عند مستوى (0.05) بين قيم الدلالة الاحصائية للاداء التفاضلي لفقرات اختبار باختلاف طريقة التعويض عن القيم المفقودة (تعظيم التوقعات، التعويض المتعدد، دالة الاستجابة) وفق نسب فقد الاستجابات (30%, 15%, 5%, 0%)؟

أهمية الدراسة:

أدى الاهتمام بمعايير الجودة التي يتم تقييم البحوث العلمية على أساسها، الى زيادة الاهتمام بمعالجة البيانات المفقودة، وكيفية التعامل معها وطرق معالجتها، وخاصة مع انتشار البرامج الإحصائية المحوسبة، وتتعدد الأساليب أو الطرق التي يمكن أن يتم فيها التعامل مع البيانات المفقودة والتي ما زالت تتطلب الفحص والمقارنة للتوصل إلى نتائج علمية تمكّن الباحثين من اختيار أفضل الطرق للتعامل مع البيانات المفقودة وكيفية توظيفها في تحسين مخرجات الأساليب الإحصائية، مقارنة مع الأساليب الإحصائية التي لا يتم فيها التعويض للبيانات المفقودة، وإثراء المخرجات وإحداث تكامل بين البيانات التي تمّ جمعها وعملية التعويض من خلال البيانات الموجودة، بحيث تخدم في النهاية الغرض الأساسي للدراسة، وهي الوصول إلى استنتاجات صحيحة قابلة للتعميم حول الظاهرة (بني عواد، ٢٠١٠).

ومن هنا فإن هذه الدراسة تكتسب أهميتها النظرية من خلال محاولتها تحديد الطريقة الاكثر فاعلية للتعويض عن البيانات المفقودة من حيث أثرها في الكشف عن الاداء التفاضلي لفقرات الاختبار، ومقارنة ما توصلت إليه نتائج الدراسة الحالية المبنية على بيانات فعلية بنتائج الدراسات السابقة المبنية على بيانات مولدة.

أما أهميتها العملية فتكمن في دراستها التجريبية والتي يرى الباحث أنها تعطي صورة أوضح عن مدى كفاءة طرق التعويض عن البيانات المفقودة من حيث أثرها في الكشف عن الاداء التفاضلي، مما يسهل على الباحثين معرفة الطريقة الاكفأ للتعامل مع البيانات المفقودة عندما يتعلق الامر بأثرها في الكشف عن الاداء التفاضلي لفقرات الاختبار، ومحاولة الوصول الى تبريرات علمية تقدم للباحثين تمكنهم من استخدام أكثر الطرق كفاءة اعتماداً على اساس تجريبي، لضمان صدق نتائج الاختبارات وعدالتها وعدم تحيزها نظراً لاهمية القرارات المبنية على النتائج المنبثقة عنها.

تعريف المصطلحات:

الأداء التفاضلي للفقرة: هـو دالة مشتقة إحصائياً للتعبير عن الفروق في الاستجابة على الفقرة بين مجموعتين مختلفتين من المفحوصين، تسمى الأولى بالمرجعية (Reference Group)، والثانية بالمستهدفة (Focal Group)، ممن لهم نفس مستوى القدرة.

البيانات المفقودة (missing data): عدم الاستجابة على بعض فقرات الاختبار أوالمقياس من قبل المفحوص، وترك هذه الاجابات فارغة.

القيمة التعويضية: القيمة التي تستبدل بها القيم المفقودة بعد أن يتم معالجتها بأحدى طرق معالجة القيم المفقودة.

محددات الدراسة:

١- إقتصرت هذه الدراسة على استخدام ثلاث طرق لتعويض القيم المفقودة هي خوارزمية تعظيم التوقعات (EM)، التعويض المتعدد (MI)، دالة الاستجابة (RF)

٢- إقتصرت هذه الدراسة على استقصاء الأداء التفاضلي لفقرات الاختبار باستخدام طريقة راجو للكشف عن الأداء التفاضلي باستخدام برمجية DFIT v.8.04.

٣- إقتصرت هذه الدراسة على استخدام نموذج الاستجابة للفقرة ثلاثي المعلمة من خلال بيانات مولدة.

الاطار النظري والدراسات السابقة:

قد يلجأ البعض للتخلص من أي فرد له قيمة مفقودة لأي من الفقرات، وتحليل بيانات الأفراد الذين استجابوا لجميع الفقرات، وهذا يعني تخفيض في حجم العينة، ويجعل مسألة تمثيل العينة للمجتمع في موضع شك، ولكي تكون العينة الباقية على الأرجح ممثلة، فإنه ينبغي أن يكون عدد الأفراد الذين تم حذفهم قبل التحليل قليل، ومع هذا فإن ذلك يؤدي إلى خسارة للمعلومات بسبب حذف بعض المفحوصين، وبالتالي يزداد احتمال التحيز (Pigott , 2001).

و تعود أسباب ظهور البيانات المفقودة الى عدم قدرة الشخص على الوصول الى الفقرة بسبب محدودية وقت الاختبار، اوعدم معرفة الاستجابة الصحيحة او نسيان الاستجابة عليها، وفي المقاييس الشخصية والاستبانات قد يعود سبب عدم الاستجابة لاسباب شخصية (Bank, 2015)، واذا لم يتم التعامل مع البيانات المفقودة بشكل صحيح فانها تؤدي الى نتائج احصائية متحيزة، و تفسيرات غير دقيقة لمعالم الفقرات وقدرات الاشخاص(Little & Rubin, 1987).

آلية فقد البيانات: Missing Data Mechanism

يشير مفهوم آلية فقد البيانات الى الكيفية التي تمت من خلالها عملية فقد بعض القيم من بعض متغيرات الدراسة، حيث تحدد آليات فقد القيم العلاقات الاحتمالية بين الفقد والقيم المشاهدة في مجموعة البيانات.

ويقسّم شيما (Cheema,2014) آليات فقد البيانات إلى ثلاث مجموعات رئيسة اعتماداً على طبيعة وخصائص القيم المفقودة والتي قد تختلف بشكل كبير عن البيانات الاصلية المشاهدة:

١-**الفقد العشوائي بالكامل** (MCAR) Missing Completely at Random.

يحدث الفقد العشوائي بالكامل عندما تكون البيانات المفقودة مستقلة عن المتغير المقاس، اي عندما تكون البيانات المفقودة مشروطة بمتغيرات اخرى غير النتائج، وكمثال على ذلك عندما يكون هناك استجابات مفقودة على المتغير Y مثل (الكفاءة الذاتية) غير مرتبطة بأي من المتنبئات في مجموعة البيانات مثل (الجنس او العرق)، وايضاً عندما تكون البيانات المفقودة على المتغير Y غير مرتبطة بقيمة المتغير Y نفسه (على سبيل المثال لايكون عدد البيانات المفقودة مرتبط بمستوى الكفاءة الذاتية مرتفع اومنخفض)، اي لايوجد آلية منتظمة لفقد البيانات، أي أن المفحوص من الممكن أن يترك الفقرة دون استجابة بكل سهولة. حيث يمكن تجاهل البيانات المفقودة بالكامل، وتكون هذه الحالة مشروعة عندما يكون حجم البيانات المتبقية مازال يمثل بدقة المجتمع المستهدف.

٢- **الفقد العشوائي** (MAR) Missing At Random.

يعبر الفقد العشوائي عن الحالة التي يكون فيها احتمال فقد القيمة الخاصة بالمتغيرY غير مرتبطة بالمتغيرY نفسه، ولكن قد تكون مرتبطة بقيمة المتغيرات الاخرى في مجموعة البيانات، فعلى سبيل المثال تكون البيانات المفقودة على المتغيرY مثل (الكفاءة الذاتية) تعتمد على متغير اخر مثل (الجنس او العرق) ولكنها غير مرتبطة بقيمة المتغيرY، وبكلمات اخرى ان الفقد العشوائي MAR هي الحالة التي تكون فيها احتمالية فقد البيانات في فقرة معينة مرتبط بالبيانات المشاهدة بشكل منتظم، أي أن بعض المفحوصين من مجموعة معينة لديهم ارجحية أعلى لترك الفقرة دون استجابة مقارنة بالمفحوصين في مجموعة أخرى.

وتعتبر البيانات المفقودة في ظل الية الفقد (MCAR) و(MAR) قابلة للتجاهل لأن البيانات المشاهدة المتبقية تمثل عينة عشوائية من البيانات الكلية المفترضة (Rubin, 1976).

٣- **الفقد غير العشوائي**Missing Not At Random (MNAR)

يعبر الفقد غير العشوائي عن الحالة التي يكون فيها احتمال فقد القيمة الخاصة بالمتغير X مرتبط بالمتغير X نفسه، وبكلمات اخرى الفقد غير العشوائي MNAR يوضح الحالة التي يكون فيه احتمال ترك الاشخاص للاستجابة على الفقرة معتمدا على اداء الافراد (Peng & Zhu, 2008). ويعني أن البيانات المفقودة مرتبطة بالمتغير نفسه، أي أن لها علاقة إما بمعالم الفقرة (صعوبة، تمييز، تخمين) أو بقدرة الأفراد، أي أن المفحوصين من الممكن أن لايستجيبوا على الفقرة لعدم قدرتهم على ذلك.

طرق التعامل مع البيانات المفقودة:

يمكن تقسيم أساليب معالجة البيانات المفقودة إلى فئتين رئيسيتين: تتضمن الفئة الأولى الطرق التي تعتمد على حذف القيم المفقودة (deletion)، بينما تتضمن الفئة الثانية طرقًا تقوم على التعويض (imputation) كما في الجدول رقم (1)

جدول (1) طرق معالجة القيم المفقودة

الطريقة	أشكال الطريقة	أنواع الطريقة
الطرق القائمة على الحذف	حذف الحالة	حذف الحالة الكاملة
		حذف الحالة المتوفرة
	طريقة الخطأ	
	طريقة غير الموجود	
الطرق القائمة على احتساب قيمة تعويضية	الطرق الصريحة	التعويض من خلال الوسط الحسابي
		التعويض من خلال الانحدار
	الطرق الضمنية	التعويض بطريقة دالة الاستجابة
		خوارزمية تعظيم التوقعات
		حساب قيم تعويضية متعددة
		حساب قيمة تعويضية للوسط المصحح للفقرة
		حساب قيمة تعويضية من توزيع مشروط
		حساب قيمة تعويضية من توزيع غير مشروط

الأداء التفاضلي للفقرة: (Differential Item Functioning)

تحتاج دراسة الاداء التفاضلي الى بيانات مكتملة، ومن العوامل التي تؤثر في دقة نتائج الكشف عن الاداء التفاضلي لفقرات الاختبارت هي ظاهرة البيانات المفقودة (Bernhard et al , 1998).

ويحدث الاداء التفاضلي عندما لايحصل المفحوصين الاعضاء في مجموعة فرعية على احتمال للاجابة على الفقرة بشكل صحيح مساوي لاحتمال الاجابة على نفس الفقرة لاعضاء اخرين في مجموعة اخرى من نفس القدرة، ويشير الاداء التفاضلي الى أن البنية الداخلية لفقرات الاختبار(مثل خصائص الفقرة كالصعوبة والتمييز) ليست متساوية للفحوصين من نفس القدرة في المجموعات المختلفة (Woods, C. M. 2008)، وهذا الاختلاف غير المتوقع من الممكن أن يسبب استنتاجات خاطئة حول السمة لبعض لمفحوصين، وبالتالي يؤثر في صدق الاختبار.

يعرّف الأداء التفاضلي (Differential Item Functioning, DIF) للفقرة على أنه اختلاف احتمال استجابة المفحوصين ممن هم في نفس مستوى القدرة على الفقرة استجابة صحيحية نتيجة انتمائهم لمجموعات مختلفة(Hambleton & Rogers, 1989).

ويشير جيبلس (Gybles, 2004) الى أن الكشف عن الأداء التفاضلي للفقرة وفق نظرية استجابة الفقرة (Item Response Theory, IRT)، يتم من خلال مقارنة منحنيات خصائص الفقرة في المجموعات الفرعية لمجتمع معين لكشف الأداء التفاضلي فيها، من خلال فحص العلاقة بين الاحتمالات المشروطة للإجابة الصحيحة على الفقرة، وبين السمة الكامنة المراد قياسها.

أنواع الأداء التفاضلي للفقرة: (Types of Differential Item Functioning)

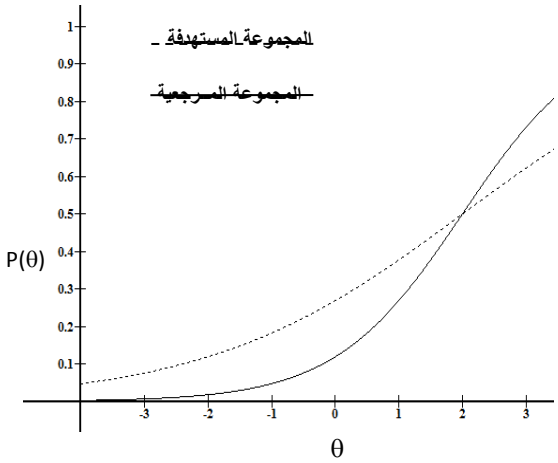
يصنف نينغ وسوزان (Ning & Suzanne, 1996) الاداء التفاضلي في فئتين:

١) الأداء التفاضلي المنتظم (Uniform DIF): وينتج عندما لا يحدث تفاعل بين مستوى الأداء وعضوية المجموعة؛ أي أن احتمال الإجابة الصحيحة للفقرة يكون أكبر دائماً لإحدى المجموعتين عند جميع مستويات القدرة.

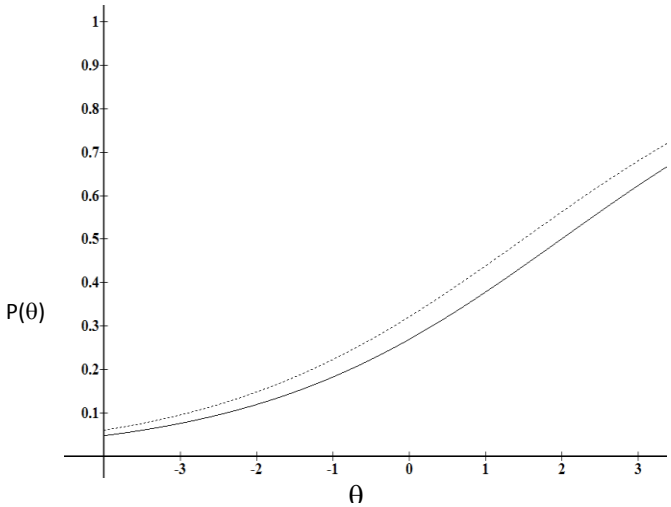
كما في الشكل رقم (١): يلاحظ عدم وجود تفاعل بين مستوى القدرة وعضوية المجموعة، حيث نجد أن احتمال الإجابة الصحيحة على الفقرة دائماً أكبر للمجموعة المستهدفة.

٢) الأداء التفاضلي غير المنتظم (Non uniform DIF): ويظهر عندما يحدث تفاعل بين مستوى القدرة وعضوية المجموعة؛ أي أن يَظْهر الأداء التفاضلي مرةً لصالح المجموعة المرجعية عند مستوى محدد من القدرة، ومرةً لصالح المجموعة المستهدفة عند مستوى آخر من القدرة.

في الشكل رقم (٢): يلاحظ أن الأداء التفاضلي لصالح المجموعة المستهدفة عند المستويات الدنيا من القدرة، ولصالح المجموعة المرجعية عند المستويات العليا. حيث يلاحظ تقاطع المنحنيين مما يدل لى وجود التفاعل بين مستوى القدرة والمجموعة، وبالتالي ظهور الأداء التفاضلي غير المنتظم(Gruijter & Kamp, 2005).



شكل (٢): الأداء التفاضلي غير المنتظم



شكل (١): الأداء التفاضلي

طرق الكشف عن الأداء التفاضلي:

تنوعت طرق الكشف عن الاداء التفاضلي نتيجة الاهتمام المتزايد بصدق نتائج الاختبارات واهمية القرارات المبنية عليها، حيث اخذت هذه الطرق اتجاهين الاول اعتمد على النظرية الكلاسيكية في القياس، والاتجاه الثاني اعتمد على النظرية الحديثة في القياس. ونظراً للاستخدام المتزايد لنظرية استجابة الفقرة (Item Response Theory,IRT)، فقد انبثقت من هذه النظرية طرقاً للكشف عن الأداء التفاضلي للفقرة مستندة بذلك على منحنى خصائص الفقرة (ICC).

وتعتمد مثل هذه الطرق في الكشف عن الاداء التفاضلي للفقرة على مدى التشابه بين منحنيات خصائص الفقرة للمجموعات الفرعية المختلفة، حيث لاتبدي الفقرة اداء تفاضلياً إذا تشابهت منحنيات خصائصها في المجموعات الفرعية المختلفة، أي أن تكون الفقرة متساوية في صعوبتها وتمييزها وتخمينها لجميع المفحوصين ممن لهم نفس مستوى القدرة في المجموعات الفرعية المختلفة، وذلك حسب النموذج المستخدم (Chung & Hui, 2004; Crocker & Algina, 1986). ويصنف هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) إجراءات الكشف عن الأداء التفاضلي للفقرة في ثلاث فئات وفق نظرية استجابة الفقرة هي:

١- الطرق القائمة على المقارنة بين المنحنيات المميزة للفقرة: (Comparison of ICC)

حيث يتم في هذه الطرق تقدير معالم الفقرات والقدرات لكل مجموعة بشكل منفصل، ثم رسم منحنى خصائص الفقرة لكل فقرة في المجموعتين، ثم حساب المساحة بين منحني خصائص الفقرة في المجموعتين فإذا كانت المساحة قليلة ولا تتعدى معياراً معيناً ممثلاً بخطأ المعاينة، عندها تكون الفقرة متحررة من الأداء التفاضلي.

٢- الطرق القائمةعلى المقارنة بين درجة المطابقة بين نموذج الاستجابة الفقرة والبيانات في المجموعات المختلفة للمفحوصين، حيث يتم في هذه الطريقة اختبار مدى مطابقة البيانات لكل فقرة في كل مجموعة فرعية لنموذج الاستجابة للفقرة المستخدم، فإذا اختلفت البيانات في كلا المجموعتين في مطابقتهما لنفس النموذج لفقرة ما، فإن هذه الفقرة تبدي عندئذٍ أداءً تفاضلياً.

٣- الطرق القائمة على المقارنة بين معالم الفقرات في المجموعات: وهذه الطرق تقوم على مقارنة معالم الفقرات والقدرات مباشرة بين المجموعات المختلفة عن طريق وضع كل معلم في مصفوفة، وحساب إحصائي خاص يتبع توزيع مربع كاي.

وتتفق هذه الطرق السابقة في اعتبار الاختلاف في المنحنيات المميزة لخصائص الفقرة للمجموعات الفرعية المتكافئة في القدرة موضع الاهتمام مؤشراً على الأداء التفاضلي لها، بالرغم من اختلاف الاجراءات المتبعة للحصول على المؤشرات الدالة على وجود التفاضلي او عدم وجوده.

حساب المساحة بين منحنيات خصائص الفقرة: (Item Characteristic Curve Method)

من البديهي ألا تتطابق منحنيات خصائص الفقرة لفقرة ما حتى وإن لم تكن تبدي اداء تفاضلياً، وذلك بسبب التقلبات العينية أو خطأ المعاينة، لذا لابد من توفر مقياس للفروق ما بين منحنيات خصائص الفقرة مثل حساب المساحة بين المنحنيين.

اقترح هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) الخطوات التالية لحساب المساحة الفاصلة بين أي منحنيين للفقرة:

١) اختيار نموذج الاستجابة للفقرة المناسب، وفيه يتم تقدير معالم الفقرات والأفراد لكل مجموعة بشكل مستقل، وتحديد طول متصل القدرة من -٣ إلى +٣ في المجموعات المختلفة.

٢) تقسيم متصل القدرة إلى فئات صغيرة عددها(K) من الفئات، حيث طول كل فئة منها(0.01).

٣) تحديد مركز كل فئة (θ_k).

٤) حساب قيم احتمالية للاجابة الصحيحة على الفقرة في المجموعتين الفرعيتين، أي $P_{i2}(\theta_k), P_{i1}(\theta_k)$ عند كل نقطة تمثل مركز الفئة، ولكل الفئات التي عددها (K).

٥) حساب الفرق في المساحة بين منحنيي خصائص الفقرة في المجموعتين الفرعيتين عن طريق إيجاد القيم المطلقة للفروق بين احتمالات الاستجابة الصحيحة على الفقرة في المجموعتين الفرعيتين عند كل مركز فئة، ثم يتم جمع الفروق ولكل مراكز الفئات من مستويات القدرة المختلفة، وفقاً للمعادلة الآتية:

$$A_{1i}\sum_{\theta=-3}^{\theta=3}\left|P_{i1}(\theta_k)-P_{i2}(\theta_k)\right|\Delta\theta =\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

حيث:

$P_{i1}(\theta_k)$: احتمال الإجابة الصحيحة على الفقرة (i)في المجموعة الأولى عند مستوى القدرة θ_K.

$P_{i2}(\theta_k)$: احتمال الإجابة الصحيحة على الفقرة (i)في المجموعة الثانية عند مستوى القدرة θ_K.

Δ_θ: طول الفئة.

٦) اتخاذ قرار بوجود الاداء التفاضلي في الفقرة (i) بناءً على قيمة A، فإن كانت قيمة A صغيرة فالتحيز بسيط، وإن كانت كبيرة فإمكانية وجود التحيز في الفقرة (i) محتملة.

ويعاب على هذه الطريقة أنها لم تأخذ بعين الاعتبار وجود أخطاء في تقدير منحنيات خصائص الفقرة للمجموعتين الفرعيتين عند مستويات القدرة المختلفة ((linn et al ,1981، وللتغلب على هذه المشكلة اقترح هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985 وزن قيم منحنيات خصائص الفقرة $P_i(\theta_k)$ باستخدام الخطأ المعياري في تقدير احتمالية الاستجابة الصحيحة على الفقرة عند قيم مراكز الفئات (θ_K) المختلفة. وعلى الرغم من ذلك فإنّ مثل هذه الطريقة تعتبر وصفية، وتفتقر لاختبارات الدلالة الإحصائية.

قدم راجو(Raju, 1988) طريقة يمكن من خلالها حساب المساحة بين منحنيات خصائص الفقرة ذات المعلمة الواحدة أو المعلمتين أو الثلاث معالم. حيث يُعبّر عن المساحة كالآتي:

$$Area = \int_{-\infty}^{\infty}|P_1(\theta)-P_2(\theta)|d\theta \ldots\ldots\ldots\ldots(12)$$

حيث:

$$P_1(\theta) = c_1 + \frac{1-c_1}{1+e^{-Da_1(\theta-b_1)}}$$

$$P_2(\theta) = c_2 + \frac{1-c_2}{1+e^{-Da_2(\theta-b_2)}}$$

فإذا كانت $a_1 \neq a_2$ ، $c = c_1 = c_2$، فإنه يمكن التعبير عن المساحة بالمعادلة:

$$\boldsymbol{Area = (1-c).\left|\frac{2(a_2-a_1)}{Da_1a_2}\ln\left(1+e^{\frac{Da_1a_2(b_2-b_1)}{a_2-a_1}}\right)-(b_2-b_1)\right| \ldots\ldots(13)}$$

حيث:

a_1: تمييز الفقرة للمجموعة الأولى، a_2: تمييز الفقرة للمجموعة الثانية.

b_1: صعوبة الفقرة للمجموعة الأولى، b_2: صعوبة الفقرة للمجموعة الثانية.

c: قيمة معلمة التخمين، D: (Scaling Factor) قيمة ثابتة وتساوي(1.7).

وإذا كانت $a_1 = a_2$ ، $c = c_1 = c_2$، ، فإنّ المساحة تعطى بالمعادلة:

$Area = (1 - c)|\, b_2 - b_1|$ ... ... (14) وعندما تكون $c_1 \neq c_2$ تصبح المساحة ما بين المنحنيين مالانهاية ولايمكن حسابها.

ويرى هامبلتون وروجرز (Hambleton & Rogers, 1989)، أنه في حال عدم تساوي معلمة التخمين للمجموعتين الفرعيتين، فإنّه عادةً ما يتم تثبيت هذه المعلمة عند (C=0.02) لحساب فرق المساحة بين منحنيات خصائص الفقرةومن ثمّ إيجاد نقطة قطِع للمساحة الإحصائية التي تعتبر معياراً للحكم على وجود الأداء التفاضلي للفقرة، أو عدم وجوده.

ويتم حساب نقطة القطع هذه على النحو الآتي:

١) تقسيم المجموعة المرجعية إلى مجموعتين عشوائيتين متكافئتين.

٢) تقدير منحنيات خصائص الفقرة لكل مجموعة بشكل منفصل.

٣) تكرار عملية اختيار المجموعة وتقسيمها لمجموعتين عشوائيتين متكافئتين لعدد من المرات، وفي كل مرة يتم تقدير معالم الفقرة لكل مجموعة منفصلة.

٤) حساب المساحة بين تلك المنحنيات في المجموعات، ويتم اعتماد أكبر قيمة للمساحة بين المنحنيين في المجموعات المختارة، كنقطة قطع تقارن بها المساحات الفعلية.

٥) إذا كانت المساحة الفعلية بين المنحنيين أكبر من القيمة العظمى لفرق المساحات بين المنحنيين في المجموعات المختارة عشوائياً (نقطة القطع)، فإن ذلك يدل على وجود الأداء التفاضلي للفقرة، وإذا كانت أقل من قيمة نقطة القطع، فإن ذلك يدل على عدم وجود أداء تفاضلي للفقرة (Hambleton & Rogers, 1995).

تتطلب طريقة راجو السابقة عند حساب المساحة بين منحنيي خصائص الفقرة للمجموعتين الفرعيتين تساوي معلمتي التخمين لهما، أو تثبيتها عند القيمة (C=0.02)، كما أنها تفترض أنّ الفرق بين منحنيي خصائص الفقرة للمجموعتين المستهدفة والمرجعية وعند جميع مستويات القدرة يسهم بشكل متساوي في قياس الأداء التفاضلي للفقرة، غير أن هذا الفرق هو موضع أهمية كبرى فيما إذا حدث في مدى من القدرة يتجمع به عدد كبير من الأفراد المفحوصين، مما يُعد قصوراً أو عيباً في تلك الطريقة، الأمر الذي دفع راجو وزميليه فاندر ليندن و فلير (Raju, van der Linden & Fleer, 1995) لتطوير ما يسمى بِـ "مقياس الأداء التفاضلي للفقرة والاختبار" (Measures of Differential Functioning of Items and Tests, DFIT).

ويشير اوشيما وموريس (Oshima & Morris, 2008) الى أن مقياس راجو للأداء التفاضلي يتمتع بالميزات الآتية:

١) يُمكن استخدامه في حالة البيانات ثنائية ومتعددة التدريج.

٢) يُمكن استخدامه لمعالجة نماذج نظرية الاستجابة للفقرة أحادية البعد ومتعددة الأبعاد.

٣) يعطي مؤشراً على الأداء التفاضلي ليس على مستوى الفقرة فقط بل وعلى مستوى الاختبار.

٤) يقدم مؤشرين للأداء التفاضلي على مستوى الفقرة هما:

- الأداء التفاضلي اللاتعويضي (Noncompensatory DIF, NCDIF)
- الأداء التفاضلي التعويضي (Compensatory DIF, CDIF)

مؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (Noncompensatory DIF, NCDIF):

يَفْترض هذا المؤشر أنّ جميع فقرات الاختبار غير تلك المراد كشف الأداء التفاضلي لها غير متحيزة، ويعرّف على أنّه معدّل مربع المسافة ما بين منحنيي خصائص الفقرة للمجموعتين المستهدفة والمرجعية، وفقاً للمعادلة الرياضية: (Raju, van der Linden & Fleer, 1995)

$$\mathrm{NCDIF}_i = \int_{-\infty}^{\infty} [P_{iF}(\theta) - P_{iR}(\theta)]^2 f_F(\theta) d\theta \text{ ... ... (15)}$$

حيث:

$f_F(\theta)$: اقتران الكثافة للقدرة (θ) في المجموعة المستهدفة.

$P_{iF}(\theta)$: احتمال الاستجابة الصحيحة على الفقرة i عند مستوى القدرة (θ) في المجموعة المستهدفة.

$P_{iR}(\theta)$: احتمال الاستجابة الصحيحة على الفقرة i عند مستوى القدرة (θ) في المجموعة المرجعية.

أو وفقاً للمعادلة:

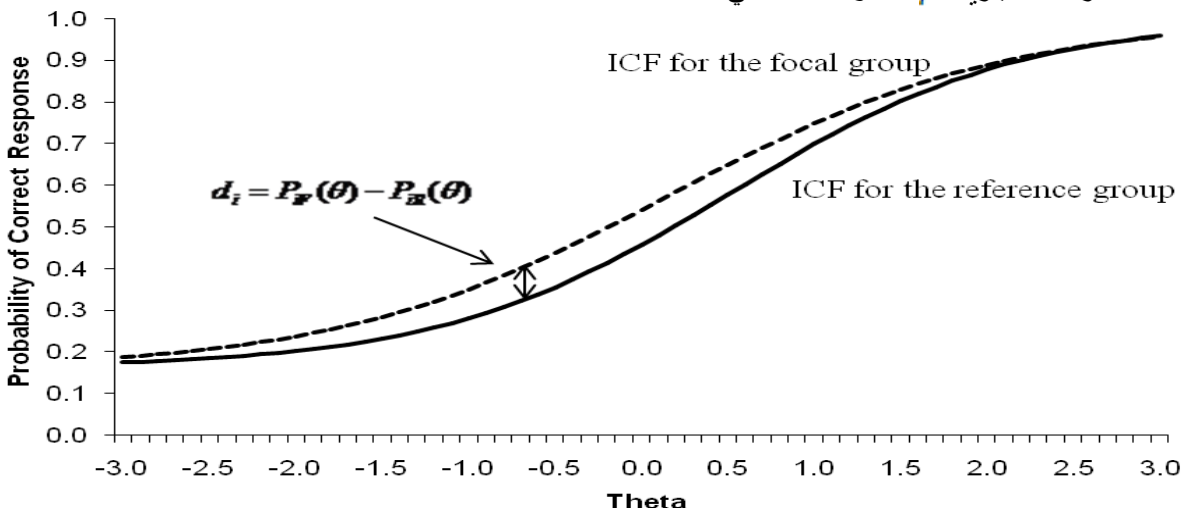
$$NCDIF_i = E_F[d_i(\theta)^2] = \sigma_{d_i}^2 + \mu_{d_i}^2 \dots\dots\dots\dots (16)$$

حيث:

E_F: التوقع بالنسبة لتوزيع (θ) للمجموعة المستهدفة.

$d_i(\theta) = P_{iF}(\theta) - P_{iR}(\theta)$ في حالة نماذج الاستجابة للفقرة ثنائية التدريج (انظر الشكل 3).

σ: الانحراف المعياري. μ: المتوسط الحسابي.



شكل (٣):المسافة ما بين منحنيي خصائص الفقرة

مؤشر الأداء التفاضلي على مستوى الفقرة التعويضي (Compensatory DIF, CDIF):

يأخذ هذا المؤشر بعين الاعتبار التباين المشترك للفقرة، كما يراعي التحيز الموجود في الفقرات الأخرى للاختبار، ويعرّف وفقاً للمعادلة الآتية:

$$CDIF_i = E_F(d_i D) = Cov(d_i, D) + \mu_{d_i}\mu_D \dots\dots\dots\dots (17)$$

حيث:

E_F: التوقع بالنسبة لتوزيع (θ) للمجموعة المستهدفة.

$d_i(\theta) = P_{iF}(\theta) - P_{iR}(\theta)$

cov: التباين المشترك. μ: المتوسط الحسابي.

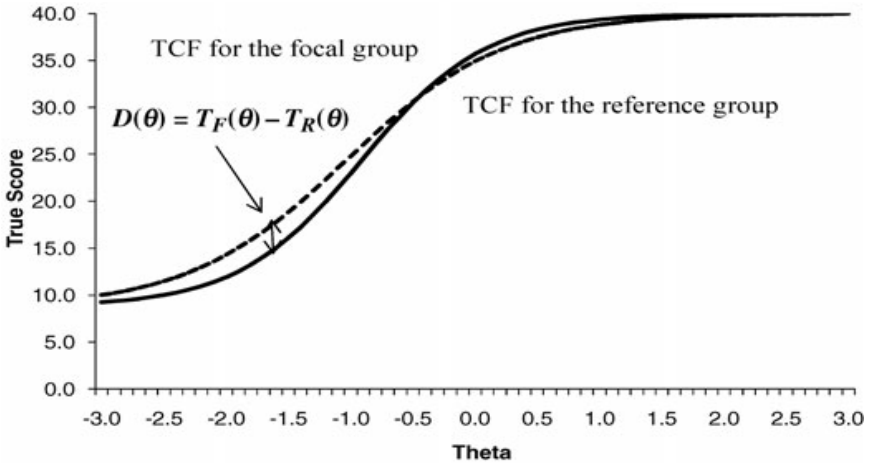
$D = T_F(\theta) - T_R(\theta) = \sum_{i=1}^{n} d_i$ (انظر الشكل 4)

$T_F(\theta)$: منحنى خصائص الاختبار للمجموعة المستهدفة عند مستوى القدرة (θ).

$T_R(\theta)$: منحنى خصائص الاختبار للمجموعة المرجعيــة عند مستوى القدرة (θ) حيث:

$$T = \sum_{i=1}^{n} P_i(\theta)$$

شكل (٤): المسافة ما بين منحنيي خصائص الاختبار



مؤشر الأداء التفاضلي على مستوى الاختبار Differential Test Functioning, DTF:

قدم راجو (Raju et all, 1995) هذا المؤشر لربط الأداء التفاضلي على مستوى الفقرة بالأداء لتفاضلي على مستوى الاختبار، حيث عّرف الأداء التفاضلي على مستوى الاختبار وفقاً للمعادلة:

$$DTF = E_F[D(\theta)^2] = \sigma_D^2 + \mu_D^2 \dots\dots\dots\dots (18)$$

حيث أنّ $D = T_F(\theta) - T_R(\theta) = \sum_{i=1}^{n} d_i$

ويمكن التعبير عن هذا المؤشر باستخدام مؤشر الأداء التفاضلي على مستوى الفقرة التعويضي (Noncompensatory DIF, CDIF) على الصورة:

$$DTF = \sum_{i=1}^{n} CDIF_i \dots\dots\dots\dots (19)$$

حيث يظهر من هذه المعادلة أنّ الأداء التفاضلي على مستوى الاختبار ما هو إلا مجموع الأداء التفاضلي على مستوى الفقرة – التعويضي. فإذا كان الأداء التفاضلي– التعويضي لفقرتين من فقرات الاختبار مثلاً متساوياً في المقدار ومتعاكساً في الإشارة، فإنّ هاتين الفقرتين تسهمان بما مجموعه صفر في الأداء التفاضلي على مستوى الاختبار.

اختبارات الدلالة الإحصائية المستخدمة في "مقياس الأداء التفاضلي للفقرة والاختبار":

Significance Tests in the DFIT Framework

طـوّر راجو (Raju et all, 1995) اختبارات للدلالة الإحصائية لمقياسه قائمة على إحصائي مربع كاي (chi–square statistic)، وذلك لاختبار الدلالة الإحصائية في كل من مؤشر الأداء التفاضلي على مستوى الاختبار (DTF)، ومؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF)، أما بالنسبة لمؤشر الأداء التفاضلي على مستوى الفقرة التعويضي (CDIF) فـإن دلالتــه لا تختبر مباشرة، فبدلاً من ذلك يتم حذف الفقرات ذات الأداء التفاضلـي العالـي (أي عند $\alpha = 0.005$, $\alpha =$ 0.001) واحدة تلو الأخرى حتى يصبح الأداء التفاضلي على مستوى الاختبار (DTF) غير دال إحصائياً، ومن ثمّ تعتبر الفقرات

المحذوفة والتي أبدت أداءً تفاضلياً تعويضياً (CDIF) دالة إحصائياً، بالرغم من استخدام إحصائي مربع كاي لاختبار دلالة مؤشر الأداء التفاضلي على مستوى الاختبار (DTF)، ومؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF)، إلا أن هذا الإحصائي يعتبر حساساً لحجوم العينات الكبيرة، وبالتالي فهو يميل في حالة العينات الكبيرة لجعل الفقرات المتحررة من الأداء التفاضلي تُبْدي أداءً تفاضلياً ذو دلالة إحصائية. الأمر الذي دفع راجو لابتكار علامة قطع من خلال دراسات قائمة على بيانات مولّدة لاختبار دلالة مؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF) لتكون:

$NCDIF > 0.006$ في حالة الفقرات ثنائية التدريج.

$NCDIF > 0.006 \times (k - 1)2$ في حالة الفقرات متعددة التدريج، حيث K: فئة الاستجابة.

إلا أنّ علامة القطع هذه قد افتقرت إلى قضية تعميمها، حيث أجريت دراسات مثل دراسة بلوت(Blot, 2002)، ودراسة تشامبلي(Chamblee, 1998) أثبتت أنها تعتمد على عدة عوامل أخرى مثل حجم العينة المستخدمة، ونوع النموذج المستخدم، الأمر الذي دفع راجو وزميليه أوشيما و ناندا (Oshima, Raju & Nanda, 2006)لابتكار اختبار دلالة إحصائية جديدة لمؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي(NCDIF) يصلح للفقرات ثنائية التدريج سمي بـ"تكرار توليد معلمة الفقرة" (Item Parameter Replication, IPR).

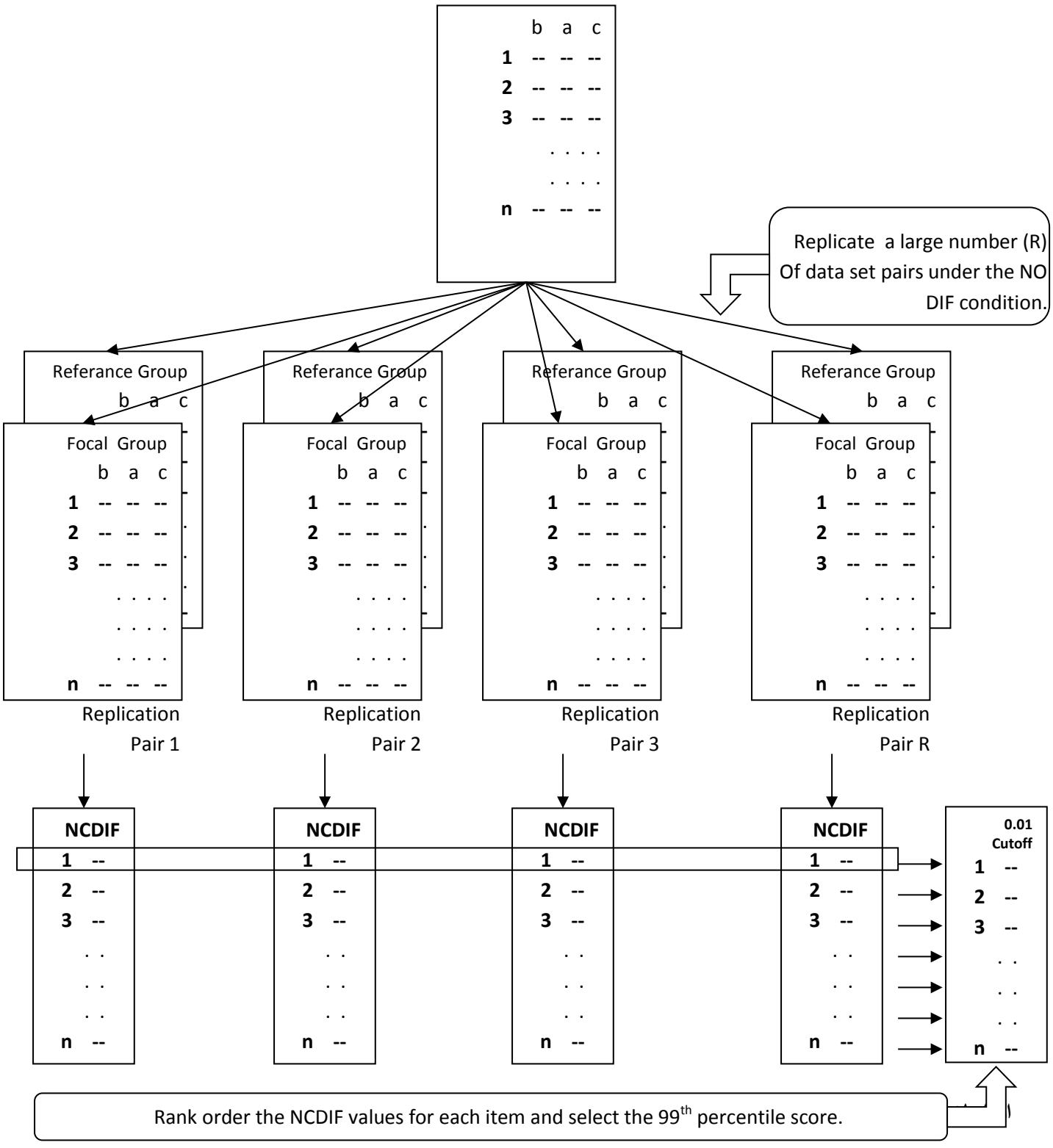
طريقة تكرار توليد معلمة الفقرة (IPR):

توفر هذه الطريقة وسيلة لاشتقاق نقاط قطع مصممة لتتوافق مع مجموعة معينة من الفقرات والبيانات المتعلقة بها، حيث تبدأ بمعالم الفقرات المقدرة للمجموعة المستهدفة، والتباينات والتباينات المشتركة لعينة تلك الفقرات والتي يتم الحصول عليها من إحدى البرامج الخاصة بنظرية الاستجابة للفقرة(IRT) مثل برنامج (Zimowski, Muraki, Mislevy, & Bock, 2002) BILOG-MG3، وبالاعتماد على هذه التقديرات المبدئية لمعالم الفقرات يتم تكرار توليدها عدداً كبيراً من المرات(على سبيل المثال ١٠٠، ١٠٠٠، أو ١٠٠٠٠ مرة) ولكل من المجموعة المستهدفة والمرجعية والموضوعتين أصلاً على نفس المتصل مع مراعاة أنّ التوقع (Expectation) للتقديرات الجديدة لمعالم الفقرات هي نفسها للتقديرات المبدئية، وبنفس التباينات والتباينات المشتركة الأولية، مما يجعل الفروق في هذه التقديرات ولأنها ناتجة من نفس توزيع المجتمع تُعْزى للخطأ العيني(Sampling Error) لا لغيره.(Oshima & Morris, 2008)

ينتج من العملية السابقة كما يلاحظ أزواجاً من العينات التي تتضمن معالم الفقرات المولدة لكل من المجموعة المستهدفة والمرجعية في كل عملية تكرار، ومن ثمّ يمكن حساب مؤشر الأداء التفاضلي اللاتعويضي(NCDIF) على مستوى الفقرة لكل زوج من تلك العينات بافتراض أنّ الفرضية الصفرية تدعي أن معالم الفقرات متماثلة لكل من المجموعة المستهدفة والمرجعية، وهنا يتم الحصول على توزيعات عينية امبريقية لمؤشر الأداء التفاضلي اللاتعويضي(NCDIF) على مستوى الفقرة، مما يتيح المجال لترتيب قيم مؤشر الأداء التفاضلي اللاتعويضي(NCDIF) على مستوى الفقرة واشتقاق علامة قطع لكل فقرة، عن طريق حساب المئين المتعلق بمستوى الدلالة(Alpha Level) المرغوب(على سبيل المثال المئين ٩٩ لمستوى الدلالة $\alpha = 0.01$).

والشكل (5) يوضح طريقة تكرار توليد معلمة الفقرة (**IPR**)

الشكل (5) طريقة تكرار توليد معلمة الفقرة (IPR)



تناولت العديد من الدراسات موضوع أثر طريقة التعويض عن القيم المفقودة في الاداء التفاضلي لفقرات الاختبارات والمقاييس ومن هذه الدراسات:

دراسة دي جيل باتريشا (de Gil Patricia , 2015) التي هدفت لدراسة أثر ست طرق للتعويض عن البيانات المفقودة (طريقة التعويض المتعدد (MI)، طريقة التعويض باستخدام متوسط استجابات الشخص (PMS)، طريقة التعويض باستخدام الارجحية العظمى(ML)، طريقة التعويض باستخدام الانحدار المنفرد (SRS)، طريقة التعويض باستخدام الوسط النسبي (RMS) طريقة الحذف (LD) على معدلات الخطأ من النوع الاول والقوة الاحصائية للكشف عن الاداء التفاضلي للفقرات متعددة الاستجابة في ظل نموذج الاستجابة المتدرجة (GRM)، باستخدام طريقة نسبة الارجحية (IRT–LR) للكشف عن الاداء التفاضلي باستخدام بيانات مولدة (simulation data) لـ (20 فقرة) وفق مقياس ليكرت في ظل نموذج الاستجابة (GRM)، وفي ظل ظروف متنوعة من نسب فقد الاستجابات (10%، 20%، 40%)، وحجم العينة للمجموعة المرجعية والمستهدفة 250:250، 500:500، 300:200، (600:400) وحجم الاداء التفاضلي (، 0.75 0.50 0.25)، وأظهرت نتائج الدراسة أن معدلات الخطأ من النوع الاول كانت قريبة من المستوى الاسمي لهذا الخطأ لجميع طرق التعويضعن البيانات المفقودة، و أن طريقة التعويض باستخدام طريقتي الارجحية العظمى (ML) و متوسط استجابات الشخص(PMS) تعطي معدلات خطأ من النوع الاول قريبة من معدلات الخطأ التي تعطيها في حالة البيانات الكاملة، وكانت طريقتي التعويض المتعدد (MI) والانحدار المنفرد(SRS) ذات فاعلية منخفضة في الكشف عن الاداء التفاضلي حيث اظهرت معدلات متضخمة للخطأ من النوع الاول، في حين كانت طريقة الحذف (LD) هي أقل الطرق فاعلية في الكشف عن الاداء التفاضلي، وأظهرت تحليلات القوة ان النتائج كانت متسقة بالكامل عبر المستويات الاسمية ومتغيرات الدراسة وطرق التعويض عن البيانات المفقودة، وان متغيرات حجم العينة ونسبة البيانات المفقودة لها تأثير في الكشف عن الاداء التفاضلي باستخدام طريقة الارجحية (IRT–LR) عبر طرق التعويض عن البيانات المفقودة المستخدمة في الدراسة.

دراسة فنش (Finch,2011 a) التي هدفت لدراسة أثر ثلاث طرق من طرق التعويض عن البيانات المفقودة (طريقة التعويض المتعدد multiple imputation، طريقة الحذف listewise deletion، طريقة اعتبار الاجابة خاطئة omitted as incorrect) على معدلات الخطأ من النوع الاول والقوة الاحصائية للكشف عن الاداء التفاضلي المنتظم، باستخدام طريقة الانحدار اللوجستي (LR) وطريقة مانتل هانزل (MH)، وطريقة (SIBTEST) باستخدام بيانات مولدة (simulation data) لعدد من الفقرات (40) فقرة ثنائية الاستجابة وفق النموذج ثلاثي المعلمة، في ظل ظروف متنوعة من حجم العينة (500، 1000، 2000)، ونسب فقد للاستجابات (0%، 5%، 15%)، واليات فقد البيانات (فقد عشوائي (MAR)، فقد عشواي بالكامل(MCAR)، فقد غير العشوائي(MNAR))، وحجم الاداء التفاضلي (، 0.6 0.3 0)، واظهرت النتائج الدراسة ان طريقة التعويض عن القيم المفقودة باعتبارها غير صحيحية تؤدي الى معدلات مرتفعة من الخطأ من النوع الاول وفي جميع ظروف الدراسة مقارنة بطرق التعويض الاخرى المستخدمة في هذه الدراسة و كانت هي الاقل قوة في الكشف عن الاداء التفاضلي مقارنة بالطرق الاخرى، كما اظهرت النتائج عدم وجود فروق دالة احصائياً في اثر طرق الكشف عن الاداء التفاضلي على معدلات الخطأ من النوع الاول والقوة.

دراسة فنش (Finch,2011b) التي هدفت لدراسة أثر ثلاث طرق من طرق التعويض عن البيانات المفقودة (طريقة التعويض المتعدد multiple imputation، طريقة الحذف listewise deletion، طريقة التعويض عن البيانات المفقودة باستجابة خاطئة Omitted as Incorrect وطريقة التعويض باستخدام الانحدار العشوائي Stochastic Regression Imputation) على معدلات الخطأ من النوع الاول والقوة الاحصائية للكشف عن الاداء التفاضلي غير المنتظم، باستخدام طريقة الانحدار اللوجستي (LR) وطريقة اختبار نسبة الارجحية (IRTLR) وطريقة (SIBTEST) باستخدام بيانات مولدة (simulation data) لـ (20 فقرة (حسب النموذج ثلاثي المعلمة، وفي ظل ظروف متنوعة من نسب فقد للاستجابات (0%، 5%، 15%، 30%)، وحجم العينة (500، 1000، 2000)، واليات فقد البيانات (الفقد العشوائي (MAR)، الفقد العشوائي بالكامل (MCAR) الفقد غير العشوائي(MNAR))، وحجم

الاداء التفاضلي التفاضلي (، 0.8 0.4 0،)، ومستويات متفاوته من القدرة للمجموعة المرجعية والمستهدفة، واظهرت النتائج الدراسة ان طريقة التعويض عن الاستجابة المفقودة باستجابة خاطئة تؤدي الى معدلات مرتفعة من الخطأ من النوع الاول عندما لاتحوي البيانات اداءً تفاضلياً وعندما تكون آلية الفقد هي (MNAR) و (MCAR)، وان طريقة التعويض عن البيانات المفقودة بطريقة الحذف ادت الى نتائج متشابهه لمعدلات الخطأ من النوع الاول والقوة عندما تكون البيانات مكتملة، واعتبرت الطريقة الاقل قوة في الكشف عن الاداء التفاضلي مقارنة بالطرق الاخرى عند الافراد ذوي القدرة المتدنية، وبمقارنة طريقة التعويض المتعدد مع طريقة الانحدار العشوائي، فإن طريقة الانحدار العشوائي تنتج معدلات خطأ اعلى خصوصا عندما تكون الية فقد البيانات MAR، حتى عندما تقدم هذه الطريقة معدلات خطأ من النوع الاول مقبولة فإنها تكون أقل قوة من طريقة التعويض المتعدد عندما تكون الية فقد البيانات (MNAR) و(MCAR)، واظهرت النتائج ايضأ أن طريقة (MI) أقل قوة بقليل من طريقةLD عندما تكون المجموعة المستهدفة أعلى بمتوسطات القدرة من المجموعة المرجعية، واظهرت نتائج الدراسة تبايناً في مستويات الخطأ من النوع الاول والقوة باختلاف طريقة التعويض عن البيانات المفقودة عند تفاعلها مع ظروف الدراسة المختلفة وباختلاف طريقة الكشف عن الاداء التفاضلي.

وأجرى ايمنجو وفالنجوك وجايلد دراسة (Emenogu & Falenchuk & Childs, 2010) هدفت لدراسة أثر ثلاث طرق من طرق التعويض عن البيانات المفقودة (طريقةanalysiswise deletion، طريقة الحذف المزدوج listewise deletion، طريقة اعتبار الاجابة خاطئة omitted as incorrect) باستخدام طريقة مانتل هانزل (MH)، حيث استخدم الباحث حالتين الحالة الاولى باستخدام بيانات حقيقية مستمدة من بيانات اختبار تقييم مؤشرات التحصيل المدرسي لبرنامج الرياضيات (SAIP) (2001) للطلبة في اعمار13 و16 عاماً في كندا، حيث تكون الاختبار من 75 فقرة من اربعة بدائل و50 فقرة من نوع الاجابة القصيرة، وقدم الاختبار بنسختين احدهما باللغة الانجليزية والاخرى باللغة الفرنسية، وتكونت عينة الدراسة من 452 طالباً اجابو على النسخة الانجليزية و 304أجابوا على النسخة الفرنسية، كما استخدم الباحث بيانات استجابات الطلبة من عمر 12 عاما على اختبار (TIMMS) في الرياضيات للعام 1995 المقدم باللغة الفرنسية والانجليزية حيث بلغ عدد الطلبة الذين تقدموا للاختبار باللغة الانجليزية 759 و عدد الطلبة الذين تقدموا للاختبار باللغة الانجليزية 318، واستخدم الباحث في الحالة الثانية بيانات مولدة وفق النموذج ثنائي المعلمة لـ 4000 شخص اعتماداً على استجابات 313 طالباً من عمر 13 سنة على آخر 25 فقرة من اختبار(SAIP) منها15 فقرة متعددة البدائل و 10 تتطلب اجابة قصيرة منهم 207 قدموا الاختبار باللغة الانجليزية و 106 قدموا الاختبار باللغة الفرنسية، واظهرت نتائج الدراسة اختلاف في عدد ونوع الفقرات التي تبدي أداءً تفاضيلياً عند تغير طريقة التعويض عن البيانات المفقودة، وان طريقة التعامل مع البيانات المفقودة باعتبارها اجابة خاطئة omitted as incorrect ليست هي الخيار الافضل للتعويض عن القيم المفقودة في تحليلات الاداء التفاضلي بطريقة مانتل هانزل، وان افضل الطرق هي طريقة analysiswise deletion.

دراسة جاريت (Garrett ,2009) التي هدفت لدراسة أثر طريقتين من طرق التعويض عن البيانات المفقودة (طريقة التعويض المتعدد multiple imputation، طريقة التعويض باستخدام متوسط استجابات الشخص within-person mean substitution) على معدلات الخطأ من النوع الاول والقوة الاحصائية للكشف عن الاداء التفاضلي للفقرات متعددة الاستجابة وعلى مقياس حجم الأُثر في ظل نموذج الاستجابة الجزئية (PCM)، باستخدام طريقتي مانتل هانزل (MH) وطريقة الانحدار اللوجستي الرتبي (IRTLR) للكشف عن الاداء التفاضلي عندما تكون الية فقد البيانات عشوائية بالكامل (MCAR) باستخدام بيانات مولدة (simulation data) لـ(20 فقرة) في ظل نموذج الاستجابة الجزئية (PCM)، وفي ظل ظروف متنوعة من نسب حجم العينة للمجموعة المرجعية والمستهدفة (500:500,700:300,900:100)، ونسب فقد للاستجابات (10%,25%,40)، وحجم الاداء التفاضلي (0.75, 0.50 0.25)، وأظهرت نتائج الدراسة ان الاداء التفاضلي يعتمد على نسبة البيانات المفقودة، وعلى حجم الاداء التفاضلي، ونسبة حجم العينة، حيث تتغير معدلات الخطأ من النوع الاول لطريقتي الكشف عن الاداء التفاضلي باختلاف طريقة التعويض عن البيانات المفقودة، وتزداد قوة الكشف عن الاداء التفاضلي يزيادة حجم الاداء التفاضلي، لكن هناك انخفاض نسبي في القوة بزيادة نسبة البيانات المفقودة، كما

أظهرت النتائج أن وجود البيانات المفقودة وحجم الاداء التفاضلي ونسبة حجم العينة يؤثر في مقياس حجم الاثر لطريقتي الكشف عن الاداء التفاضلي، حيث يزداد حجم الاثر بزيادة حجم الاداء التفاضلي، وبزيادة نسبة البيانات المفقودة يقل حجم الاثر.

دراسة روبتزش و روب (Robitzch & Rupp,2009) التي هدفت لدراسة أثر اربع طرق من طرق التعويض عن البيانات المفقودة (طريقة الحذف listewise deletion، طريقة التعويض بصفر zero imputation، طريق التعويض الثنائي two way imputation، طريق التعويض المتعدد باستخدام المعادلات المقيدة (multiple imputation by chained equations) على التحيز و والجذر التربيعي لمتوسط الخطأ، ومعدلات الرفض، باستخدام طريقة الانحدار اللوجستي (LR) وطريقة مانتل هانزل (MH)، باستخدام بيانات مولدة (simulation data)، في ظل ظروف متنوعة من حجم العينة (500,2000,5000) فرد وعدد الفقرات (20,40) فقرة، ونسب فقد للاستجابات (10%,30%)، واليات التعويض (الفقد العشوائي (MAR)، الفقد العشوائي بالكامل (MCAR) و الفقد غير العشوائي(MNAR))، وحجم الأثر للاداء التفاضلي (0.4,0.6, 0.2 ,0)، واظهرت نتائج الدراسة أن التفاعل بين آليات فقد البيانات، طرق التعويض عن البيانات المفقودة، وحجم الأثر هي الاكثر تأثيراً في تفسير التباين في الاداء التفاضلي للفقرات، كما أظهرت النتائج الى الطريقة الخاطئة للتعويض عن البيانات المفقودة من الممكن أن تزيد معدلات الخطأ من النوع الاول والثاني، كما أظهرت النتائج ان اختلاف طريقة الكشف عن الاداء التفاضلي وعدد فقرات الاختبار وحجم العينة ليس له أثر في نتائج الدراسة.

من خلال ما تمّ استعراضه من الدراسات السابقة المتعلقة بأثر طريقة التعويض عن القيم المفقودة في الكشف عن الاداء التفاضلي، تبين مايلي:

١- إن طريقة (LD) الحذف هي الاقل فاعلية في دقة الكشف عن الاداء التفاضلي، حيث تقود الى معدلات متضخمة للخطأ من النوع الاول، في حين كانت طريقة التعويض المتعدد (MI) أفضل الطرق في دقة الكشف عن الاداء التفاضلي.

٢- تباين أثر طرق التعامل مع البيانات المفقودة في دقة الكشف عن الاداء التفاضلي بتباين ظروف الدراسة من حجم العينة، وحجم الاداء التفاضلي، وعدد فقرات الاختبار.

٣- زيادة نسبة البيانات المفقودة يؤدي الى زيادة معدلات الخطأ من النوع الاول ويقلل من القوة الاحصائية للكشف عن الاداء التفاضلي، دراسة (de Gil Patricia ,2015) و (Garrett ,2009).

الطريقة والإجراءات:

للإجابة عن أسئلة الدراسة من خلال الإعتماد على بيانات مولدة وكيفية توليد البيانات للاستجابة على الفقرات. ويتضمن تصميم توليد البيانات ووصفاً لكل من مستوى الفقرات ذات الأداء التفاضلي، ونسبة الفقرات ذات الأداء التفاضلي في الاختبار. كما تناول وصفاً لكيفية توليد البيانات وإجراءات تقدير المعالم للفقرات وللإفراد. ووصفاً للمعالجات الإحصائية المستخدمة للإجابة عن أسئلة الدراسة حسب الاجراءات التالية:

أولا: الظروف التي تستند إليها إجراءات توليد البيانات.

١. من حيث تصميم البيانات المستخدم في هذه الدراسة، فقد تكون من (7) مجموعات، وهي مجموعة مرجعية واحدة، 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.

٢. 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.

٣. عدد الأفراد في كل مجموعة من المجموعات. حيث بلغ عدد الأفراد في كل مجموعة (510) أفراد، وكان التصميم المستخدم في هذه الدراسة يشتمل على (7) مجموعات. وبناءً على ذلك؛ فقد بلغ عدد الأفراد الكلي (3570) مفحوصاً.

٤. 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%).

٥. 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).

٦. 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 الجدول (1) معالم الفقرات التي تم توليدها في المجموعة المرجعية.

الجدول (١): قيم معالم الفقرات ثنائية التدريج من حيث (الصعوبة والتمييز والتخمين) لدى المجموعة المرجعية تبعاً للنموذج اللوجستي ثلاثي المعلمة.

رقم الفقرة	معالم الفقرة		
	التمييز A	الصعوبة B	التخمين C
1	0.313	-0.898	0.222
2	0.324	-0.693	0.25
3	0.265	-1.379	0.215
4	0.124	0.535	0.202
5	0.215	-1.114	0.269
6	0.321	0.299	0.221
7	0.311	0.088	0.265
8	0.302	0.811	0.278
9	0.345	0.939	0.207
10	0.359	0.693	0.262
11	0.405	0.361	0.253
12	0.321	-1.282	0.246
13	0.449	0.66	0.218
14	0.468	0.876	0.286
15	0.496	-0.231	0.226
16	0.508	-0.204	0.22
17	0.528	-1.053	0.271
18	0.531	-0.806	0.264
19	0.562	-0.208	0.229
20	0.57	-0.16	0.249
21	0.592	-0.299	0.24
22	0.603	-1.467	0.295
23	0.624	-0.76	0.276
24	0.614	1.172	0.253

25	0.642	-0.635	0.242
26	0.649	0.334	0.298
27	0.679	-1.456	0.252
28	0.712	-1.358	0.269
29	0.715	-1.373	0.242
30	0.787	-0.711	0.207
31	0.79	0.019	0.25
32	0.839	-0.463	0.27
33	0.896	-0.675	0.27
34	0.9	0.702	0.288
35	0.902	-0.332	0.238
36	0.903	-0.218	0.282
37	0.914	-1.315	0.23
38	0.95	-1.074	0.265
39	0.99	0.166	0.237
40	0.997	0.093	0.228
41	1.001	0.503	0.295
42	1.046	0.778	0.285
43	1.047	1.232	0.238
44	1.106	-1.368	0.289
45	1.123	-0.741	0.289
46	1.15	-1.13	0.289
47	1.171	-0.566	0.287
48	1.179	-1.396	0.266
49	1.195	0.381	0.205
50	1.196	-0.982	0.265

الخطوة الثانية

في الخطوة الثانية، تم توليد بيانات المجموعات المستهدفة باستخدام نفس الإجراءات في الخطوة الأولى باستثناء إجراء تغييرات على قيم معالم الفقرات في المجموعات المستهدفة من أجل أن تعكس أداءً تفاضلياً على بعض الفقرات المختارة، وحسب كل مجموعة من مجموعات البيانات التي تم توليدها. ويبين الجدول (2) أرقام الفقرات التي تم اختيارها من المجموعة المرجعية الجدول (1) لكي تظهر أداءً تفاضلياً في المجموعات المستهدفة، فكانت على النحو التالي:

الجدول(٢):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
		34

35	35	35
39	39	39
40		
42	42	42
44	44	
50		

كيفية اختيار الفقرات الموجودة في الجدول (2) لكي تظهر أداءً تفاضلياً.

لقد تم اختيار الفقرات لتظهر أداءً تفاضلياً من خلال ترتيب معاملات الصعوبة للفقرات في المجموعة المرجعية تصاعدياً من الأدنى إلى الأعلى، 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 (2).

وبعد إختيار 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. وفي هذا الموقف يكون هنالك عدة فقرات ذات أداء تفاضلي ونسب هذه الفقرات في الاختبار هي (10%، 20%، 40%). ويكون هنالك أيضاً مستويات مختلفة من الأداء التفاضلي للفقرة وهي (متدنٍ بمقدار0.5، وعالٍ بمقدار1.5). وفي هذه المواقف وحين يتوفر عدد معين من الفقرات ذات الأداء تفاضلي في كل مجموعة، فإن هذه الفقرات تمتلك نفس المستوى من الأداء التفاضلي، وهذا يعنى من حيث سهولة تفسير البيانات فلم يتم البحث في أثر وجود مستويات مختلفة للأداء التفاضلي للفقرة عند دراسة كل مجموعة منفردة تبعاً لنسبة الفقرات ذات الأداء التفاضلي.

النتائج ومناقشتها:

السؤال الاول: هل توجد فروق ذات دلالة أحصائية عند مستوى (0.05) بين نسب الفقرات التي تبدي أداء تفاضليا تعزى لاختلاف طريقة التعويض عن القيم المفقودة (تعظيم التوقعات، التعويض المتعدد، دالة الاستجابة) وفق نسب فقد الاستجابات (30%، 15%، 5%، 0%)

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 (٣).

الجدول (٣): 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 ١٠%، المستوى ٠,٥)	0		1.57	-1.57	1.57	1	0.210
المستهدفة (النسبة ١٠%، المستوى ١,٥)	2	299, 387	1.57	0.43	0.12	1	0.732
المستهدفة (النسبة ٢٠%، المستوى ٠,٥)	0		1.57	-1.57	1.57	1	0.210
المستهدفة (النسبة ٢٠%، المستوى ١,٥)	0		1.57	-1.57	1.57	1	0.210
المستهدفة (النسبة ٤٠%، المستوى ٠,٥)	4	2, 140, 251, 490	1.57	2.43	3.75	1	0.053
المستهدفة (النسبة ٤٠%، المستوى ١,٥)	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 (٣) أعداد وأرقام الأفراد غير المطابقين في كل مجموعة حسب نسبة الفقرات ذات الأداء التفاضلي، ومستوى الأداء التفاضلي للفقرة، حيث تبين من خلال الجدول أن أعداد الأفراد غير المطابقين لا يتجاوز (4) أفراد

من أصل (510) أفراد في كل مجموعة. كما يُلاحظ بأنه لم يكن هنالك تقاطعاً بأرقام الأفراد غير المطابقين عبر جميع المجموعات، 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 (٤)، أنه لا توجد فروق ذات دلالة إحصائية بين نسب الأفراد المطابقين للنموذج الثلاثي المعلمة في نظرية الاستجابة للفقرة بين جميع المجموعات تُعزى لنسبة الفقرات ذات الأداء التفاضلي ولمستوى الأداء التفاضلي للفقرة، حيث بلغت احتمالية الدلالة الإحصائية (0.051).

أ. 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 (٤) يبين إحصائيات الفقرات غير المطابقة.

الجدول (٤): 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 ١٠%، المستوى ٠,٥)	22	1.145	0.220	-0.276	0.318	0.365	0.085	**36.1**	8	0.000
المستهدفة (النسبة ٢٠%، المستوى ٠,٥)	1	1.476	0.459	1.126	0.218	0.487	0.045	**26.5**	9	0.002
المستهدفة (النسبة ٤٠%، المستوى ٠,٥)	37	1.890	0.297	-0.516	0.184	0.313	0.068	**36.5**	8	0.000
المستهدفة (النسبة ٤٠%، المستوى ١,٥)	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)، أن هنالك فقرة واحدة لم تطابق في المجموعة المستهدفة التي تتمثل في أن (١٠%) من فقراتها ذات أداء تفاضلي، ومستوى الأداء التفاضلي لهذه الفقرات (0.5)، وتم حذفها من المجموعة وكان رقمها (22). كما تم حذف الفقرة رقم (1) في المجموعة المستهدفة التي تتمثل في أن (20%من فقراتها ذات أداء تفاضلي، ومستوى الأداء التفاضلي لهذه الفقرات (0.5) بسبب عدم مطابقتها، وحذفت الفقرة رقم (37) في المجموعة المستهدفة التي تتمثل في أن (40%) من فقراتها ذات أداء تفاضلي، ومستوى الأداء التفاضلي لهذه الفقرات (0.5)، وتم حذف فقرتين رقمهما (16,50) في المجموعة المستهدفة التي تتمثل في أن (40%من فقراتها ذات أداء تفاضلي، ومستوى الأداء التفاضلي لهذه الفقرات (0.5)، 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 (٥) 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.

السؤال الثاني: هل توجد علاقة ارتباطية ذات دلالة إحصائية عند مستوى (0.05) بين قيم الدلالة الاحصائية للاداء التفاضلي لفقرات اختبار باختلاف طريقة التعويض عن القيم المفقودة (تعظيم التوقعات، التعويض المتعدد، دالة الاستجابة) وفق نسب فقد الاستجابات (30 %، 15%، 5%، 0%)؟

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)، بلإضافة الى إجراء تحليل التباين الثنائي، وذلك على النحو التالي:

معلمة 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 (٥).

الجدول(٥). 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	٣٨٢0.	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 (٥)، 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)، بفارق مقداره بنفس الترتيب (0.4، 0.37). 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 (٦).

الجدول (٦). 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 (٦)، 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 (٦ 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 (7).

الجدول (7). 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	٢٧٣0.				
%١٠	0.382	0.109			
%20	0.403	0.130	0.021		
%40	0.606	**0.267**	0.224	0.203	

يتضح من الجدول (٧) بأنه لا يوجد إلا فرقاً واحداً دال إحصائياً بين المتوسطات الحسابية لمعلمة الصعوبة وكان بين المجموعة المستهدفة التي تشتمل على ٤٠% 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.

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 (٨).

الجدول (٨). 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 (٨)، 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 (٩).

الجدول(٩): نتائج تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة التمييز لفقرات الاختبار حسب نسبة الفقرات ذات الأداء التفاضلي 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 (٩)، 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 (١٠).

الجدول (١٠). 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 (١٠)، 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 (١,٥) تأثراً جوهرياً، بحيث انخفضت قيمة الوسط الحسابي لمعالم تمييز الفقرات في الاختبار مقارنةً مع قيمة الوسط الحسابي لمعالم تمييز الفقرات في المجموعة المرجعية، كما يُلاحظ من خلال الجدول عدم وجود فرق ذو دلالة إحصائية للأوساط الحسابية بين المجموعة المرجعية والمجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (٠,٥)، كما ويوجد فرق ذو دلالة إحصائية للأوساط الحسابية بين المجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (٠,٥) والمجموعة التي تشتمل على فقرات ذات مستوى أداء تفاضلي مقداره (١,٥).

يُلاحظ من خلال بيانات الجدول (١٠) 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.

معلمة 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 (١١).

الجدول (١١). 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 (١١)، 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 (١٢).

الجدول (١٢). 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 المحسوبة	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 (١٢)، 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.

المراجع العربية

بني عواد، علي.(2010). مقارنة طرق التعامل مع البيانات المفقودة في تقدير معالم الفقرات وقدرات الافراد. رسالة دكتوراة غير منشورة. جامعة اليرموك. الاردن.

الدرابسة، رياض.(2010). أثر طريقة تقدير القدرة وطريقة التعامل مع القيم المفقودة على دقة تقدير معالم الفقرات والافراد. رسالة دكتوراة غير منشورة. جامعة اليرموك. الاردن.

عبدالعزيز، عبدالله.(2006). فاعلية طرق منحنى خصائص الفقرة في الكشف عن الاداء التفاضلي لمتغير الجنس في فقرات اختبار رياضيات متعددة التدريج وثنائية التدريج. رسالة دكتوراة غير منشورة. جامعة اليرموك. الاردن.

المراجع الاجنبية

Acock, A. (2005).Working With Missing Values. Journal of Marriage and Family, 15 1012–1028.

Allison, P. D. (2002). Missing data (Vol. 136). Sage publications.

Banks, K. (2015). An Introduction to Missing Data in the Context of Differential Item Functioning. Practical Assessment, Research & Evaluation, 20

Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. Multivariate Behavioral Research, 35, 321 – 364.

Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M.,... & Hürny, C. (1998). Missing quality of life data in cancer clinical trials: serious problems and challenges. Statistics in medicine, 17(5-7), 517-532.

Blot, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education, 15*(2), 113-141.

Chamblee, M. C. (1998). *A Monte Carlo investigation of conditions that impact type 1 error rates of differential functioning of items and tests***.** Unpublished doctoral dissertation, Georgia State University, Atlanta.

Cheema, J. R. (2014). A review of missing data handling methods in education research. Review of Educational Research, 84(4), 487-508.

Crocker, L.& Algina , J. (1986). *Introduction to Classical and Modern Test Theory*. California: Cole Publishing Company.

Chung Wang, W. & Hui Su, Y. (2004). Effects of average signed area between two item characteristic curves and purification procedures on the DIF detection via the Mantel-Haenzel method. *Applied Measurement in Education, 17*(2), 113-144.

de Gil Patricia, R. (2015). An Empirical Comparison of the Effect of Missing Data on Type I Error and Statistical Power of the Likelihood Ratio Test for Differential Item Functioning: An Item Response Theory Approach using the Graded Response Model(Unpublished doctoral dissertation).. University of South Florida.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1-38.

Enders, C. K. (2010). Applied missing data analysis. Guilford Press.

Emenogu, B. C., Falenchuk, O., & Childs, R. A. (2010). The effect of missing data treatment on Mantel-Haenszel DIF detection. Alberta Journal of Educational Research, 56(4), 459

Finch, H. (2011a). The use of multiple imputation for missing data in uniform DIF analysis: Power and type I error rates. Applied Measurement in Education, 24, 281-301.

Finch, W. H. (2011b). The impact of missing data on the detection of nonuniform differential item functioning. Educational and Psychological Measurement, 71(4), 663-683.

Garrett, P. (2009). A Monte Carlo study investigating missing data, differential item functioning, and effect size (Unpublished doctoral dissertation). Georgia State University.

Gruijter, D. & Kamp, L. (2005). *Statistical test theory for education and psychology*. Retrieved from: http://irt.com.ne.kr/data/test_theory.pdf.

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

Hambleton, R. & Rodgers, J. (1995). Item bias review. *Practical Assessment, Research & Evaluation*,4(6).RetrievedMarch9,2014fromhttp://PAREonline.net/getvn.asp?v=4&n=6.

Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. Applied Measurement in Education, 2(4), 313-334.

Hambleton, R., & Swaminathan, H. (1985). *Item Response Theory: principles and applications*. Boston: Kluwer–Nijhoff publishing.

Hohensinn, C., & Kubinger, K. D. (2011). On the impact of missing values on item fit and the model validness of the Rasch model. Psychological Test and Assessment Modeling, 53(3), 380-393.

Kane, M. (1996). The precision of measurements. Applied Measurement in ducation, 9 (4), 355-379.

Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. Applied Psychological Measurement, 5(2), 159-173.

Little, R. J., & Rubin, D. (1987). Analysis with missing data.. John Wiley & Sons.

Little, R. J., & Rubin, D. B. (2002). Statistical analysis with missing data 2ed edition. Hobokin,New York: John Wiley.

Ning, Wang & Lane, Suzanne. (1996). Detection of gender-related differential item functioning in mathematics performance assessment. *Applied Measurement in Education. 9*, 171-211.

Oshima, T. C. & Morris, S. B. (2008). An NCME Instructional Module on Raju's Differential Functioning of Items and Tests (DFIT). *Educational Measurement: Issues and Practice 27* (3): 43–50.

Oshima, T. C., Raju, N. S. & Nanda. (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of Educational Measurement*, *43*, 1-17.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. Review of educational research, 74(4), 525-556.

Peng, C. Y. J., & Zhu, J. (2008). Comparison of two approaches for handling missing ovariates in logistic regression. Educational and Psychological Measurement, 68(1), 58-77.

Pigott, T. D. (2001). A review of methods for missing data. Educational research and evaluation, 7(4), 353-383.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*(4),495-502.

Raju, N. S., van der Linden & Fleer, P. F. (1995). An IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19*(4), 353-368.

Robitzsch, A., & Rupp, A. A. (2009). Impact of missing data on the detection of differential item functioning: The case of Mantel-Haenszel and logistic regression analysis. Educational and Psychological Measurement, 69(1), 18-34.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.Schafer, J. L., & Graham, J. W. (2002). Missing data: Ourview of thestate of the art. Psychological Methods, 7,147–177.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art.

Psychological Methods, 7(2), 147-177.

Witta, E. L. (2000). Effectiveness of Four Methods of Handling Missing Data Using Samples from a National Database.

Woods, C. M. (2008). Likelihood-ratio DIF testing: Effects of nonnormality. Applied

Psychological Measurement, 32(7), 511-526.

Zimowski, M., Muraki, E. Mislevy, R. Robert, J., & Bock, D. (1996). Bilog-Mg3**:** *Multiple group IRT analysis and test maintenance for binary items* (computer program) Chicago: Scientific Software. 17. No.1.