أثر حجم العينة في الأداء التفاضلي للمفردة وفقا لنظرية الاستجابة للفقرة

الباحث: زياد العبدالله / كلية الإلهيات – جامعة غازي عنتاب – تركيا

استلام البحث: ٨/ ١ /٢٠٢٠ قبول النشر: ٢٠٢١/٦/١٨ تاريخ النشر : ٢ /١ /٢٠٢٢

https://doi.org/10.52839/0111-000-072-004

الملخص:

بحثت الدراسة تأثير اختلاف حجم العينة على دقة الكشف عن الأداء التفاضلي للمفردة ، حيث تم استخدام ثلاثة حجوم مختلفة من العينات (١٠٠٠،٥٠٠،٣٠٠) فضلاًعن أختبار مكون من عشرين مفردة متعددة التدريج وذات خمس فئات، وتم استخدام أنموذج الاستجابة المتدرجة كأحد نماذج نظرية الاستجابة للمفردة متعددة التدريج لتقدير معالم المفردات والأفراد، كما تم اعتماد طريقة مانتل–هانز كإحدى طرق الكشف عن الأداء التفاضلي للمفردات عبر كل حالة من حجوم العينة المختلفة، وقد أظهرت نتائج الدراسة العلاقة العكسية بين حجم العينة وعدد المفردات التي أظهرت أداءا تفاضلياً.

الكلمات المفتاحية: الأداء التفاضلي للمفردة، نماذج الاستجابة للمفردة متعددة التدريج، نموذج الاستجابة المتدرجة، حجم العينة.

The Effect of Sample Size on the Item Differential Functioning in the Context of Item Response Theory

Ziad Abdullah

Theology College – Gaziantep University – Turkey

E-mail

Ziadsy@gmail.com

Ziadsy@gantep.edu.tr

Abstract

The current study examined the effect of different sample sizes to detect the Item differential functioning (DIF). The study has used three different sizes of the samples (300, 500, 1000), as well as to test a component of twenty polytomous items, where each item has five categories. They were used Graded Response Model as a single polytomous item response theory model to estimate items and individuals' parameters. The study has used the Mantel-Haenszel (MH) way to detect (DIF) through each case for the different samples. The results of the study showed the inverse relationship between the sample size and the number of items, which showed a differential performer.

Keywords: Item Differential Functioning (IDF), sample size, polytomous item response theory models, Graded Response Model (GRM).

مقدمة:

إن أي عملية تقويم لابد وأن ترتكز على الوحدة الأساسية المكونة للاختبار وهي المفردة، لذلك دأب خبراء القياس والتقويم على تطوير أساليب إحصائية عدة تهتم بتحديد معالم المفردة وصولا الى التحديد الدقيق لمواصفات الاختبار والذي يعد أداة القياس الرئيسة لمعرفة مدى تحقق نواتج التعلم التي ستبنى بناء عليها القرارات والتقييمات التوصيفية للأفراد، وأفرزت هذه الجهود النظرية التقليدية في القياس Classical Test Theory (CTT) والتي تعاملت مع الاختبار ككتلة واحدة مترابطة، حيث اهتمت بدراسة متغيرات الاختبار والعوامل المؤثرة عليه سواء المتعلقة بحجم العينة أو عدد بنوده أو نوعية المفردات، ونظراً للاهتمام المتزايد بعملية قياس نواتج التعلم فقد أظهرت هذه النظرية العديد من نقاط الضعف كان أبرزها نسبية عملية القياس وارتباط نتائجها بحجم ونوع العينة الخاضعة للتقييم، وهذا مادفع الكثير من المهتمين في هذا المجال للتفكير بطريقة تجعل الاختبار مستقلاً عن العينة التي يُطَبَّق عليها، فنتج عن ذلك النظرية الحديثة في القياس Item Response Theory (IRT) والتي جعلت المفردة الواحدة هي محط الاهتمام من خلال تحديد معالمها من صعوبة وتمييز بين الافراد وتخمين للإجابة الصحيحة، وبذلك تصبح عملية قياس نواتج التعلم من قدرات ومهارات عند الأفراد أكثر دقة وموضوعية على اعتبار أن مفردات الاختبارات بُنيت على أساس المعايير التي وضعها خبراء المجال والتي من خلالها سيتم الحكم على الأفراد، ومن خلال الاهتمام بالاختبارات على مستوى المفردة الواحدة ظهرت أيضاً بعض العقبات، من أهمها مفهوم تحيز المفردة Item Bias وأثره الكبير في قياس نواتج التعلم للأفراد، ومع ظهور المقاييس النفسية ومقاييس الذكاءات المتعددة سواء ثنائية أو متدرجة الاستجابة التي يمكن تطبيقها على عينات وبيئات مختلفة ثقافيا وعرقيا برز مفهوم الأداء التفاضلي للمفردة Item Differential Functioning (DIF) الذي يهتم بالكشف عن المفردات التي تتأثر باختلاف عينة التطبيق ومن ثم العمل على تحسين هذه المفردات أو استبعادها بهدف تحقيق أفضل درجة ممكنة من العدالة بين الافراد عند قياس نواتج التعلم التي من المفترض أن يكونوا قد اكتسبوها أثناء مرحلة التعليم والتأهيل.

هدف الدراسة:

تهدف الدراسة الحالية إلى دراسة أثر حجم العينة في الكشف عن الأداء التفاضلي للمفردات التي تؤدي دورا متباينا بين الافراد المتساوين بالقدرة والمختلفين بسمة ما كالعرق أو الجنس أو الثقافة. ويمكن تلخيص هدف الدراسة بالإجابة عن التساؤل الآتي: ما تأثير اختلاف حجم العينة في الكشف عن الأداء التفاضلي للمفردة؟ ويشتق من هذا التساؤل الفرضية التالية: لا يوجد فرق ذو دلالة إحصائية في الأداء التفاضلي للمفردات باختلاف حجم العينة.

أهمية الدراسة:

تبرز أهمية الدراسة الحالية في استخدامها للنظرية الحديثة في القياس بهدف الكشف عن المفردات التي تتسبب في حدوث تحيزات في تقييم نواتج التعلم وذلك عند اختلاف حجم العينة المستخدمة، وبالتالي ستسلط الضوء على الاهتمام بحجم العينة وأثرها عند بناء الاختبارات أو المقاييس بناء على المعايير المحددة من قبل خبراء المجال، وهذا ماسيدفع المعنيين في بناء الاختبارات والمقاييس إلى أخذ الحيطة في قياس نواتج التعلم وخاصة عند استخدام عينات صغيرة الحجم والتي من الممكن أن تلعب دوراً سلبيا في عملية التقييم. ومن جهة أخرى تبرز أهمية هذه الدراسة من خلال استخدامها لنموذج الاستجابة المتدرجة Graded Response Model كأحد النماذج متعددة التدريج Polytomous في النظرية الحديثة، وذلك على اعتبار أن معظم الدراسات التي أجريت في هذا المجال تناولت النماذج ثنائية التدريج Dichotomous.

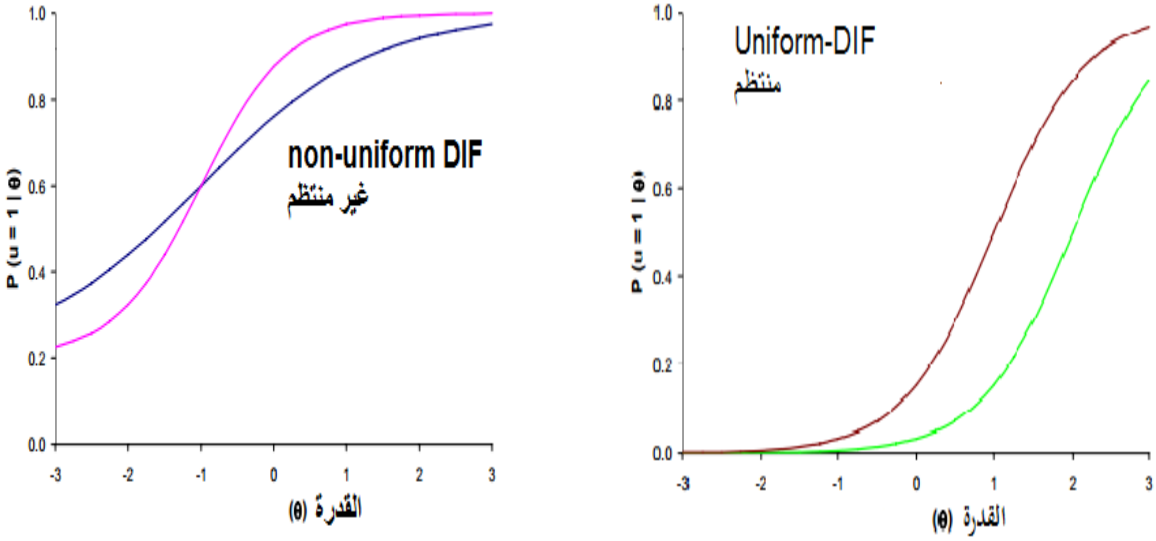
مفهوم الأداء التفاضلي للمفردة Item Differential Functioning (DIF):

يعرف الأداء التفاضلي للمفردة بشكل عام بأنه الاختلاف في أداء مجموعتين من الأفراد –تملكان المستوى نفسه من القدرة– في الاستجابة عن المفردة ذاتها علما بأن هاتين المجموعتين من الأفراد مختلفتان في بعض المتغيرات كالعرق أو الثقافة أو اللغة أو الجنس. و تسمى المجموعة الأولى بالمجموعة المرجعية Refrence group، والمجموعة الثانية المجموعة التجريبية Focal group، وقد تعددت الأساليب والتقنيات الإحصائية المعلمية واللامعلمية[١] التي عملت على كشف الأداء التفاضلي للمفردة، حيث اشتملت على نوعين من الإجرائيات: إجرائيات اهتمت بالمفردات ثنائية التدريج (٠ ، ١) Dichotomous وإجرائيات اهتمت بالمفردات متعددة التدريج Polytomous (٠ ، ١ ، ٢ ، ٣ ، ....)

(Hidalgo & Gómez–Benito, 2010; Millsap & Everson, 1993; Potenza & Dorans, 1995). وتجدر الإشارة إلى أنه عادة مايتم التمييز بين نوعين من الـ DIF الأول هو المنتظم Uniform DIF والذي يشترط أن يكون معلم التمييز للمفردة نفسه بين مجموعتي المقارنة، بمعنى أنه لايوجد تقاطع بين إجابات مجموعتي المقارنة المرجعية والتجريبية عبر كل مستويات القدرة الموجودة على متصل القدرة ، أما الثاني فهو غير المنتظم Non–Uniform DIF وفيه يكون معلم التمييز للمفردة مختلف بين مجموعتي المقارنة المرجعية والتجريبية (Finch, 2005) ، أي أنه يحصل تقاطع بين مجموعتي المقارنة عند بعض مستويات القدرة وهذا يعني أن تجانس احتمالات استجابة الأفراد بين المجموعتين ليس محققا دائما (Mellenbergh, 1982) والشكل (١) يوضح الاختلاف بين الطريقتين المذكورتين. وستعمل الدراسة الحالية على الكشف عن الأداء التفاضلي

[١] طرق DIF المعلمية تعتمد على التقديرات الدقيقة لمعالم الأفراد والمفردات عبر استخدام إحدى نماذج نظرية الاستجابة للمفردة وذلك أثناء الكشف عن الأداء التفاضلي للمفردة، أما طرق DIF اللامعلمية فتعتمد على درجة الاختبار الملاحظة بعد استجابة الفرد عن مفردات الاختبار وبالتالي لاتستخدم القيم المقدرة لمعالم الأفراد والمفردات كما هو الحال في الطرق المعلمية (Raju & Ellis, 2002)

لمفردات متعددة التدريج من خلال طريقة مانتل–هانزل مع تثبيت معلم التمييز لكل مفردة بين مجموعتي المقارنة.



الشكل (١) يوضح الفرق بين الأداء التفاضلي المنتظم وغير المنتظم لمفردة بين مجموعتين

طرق الكشف عن الأداء التفاضلي للمفردة DIF Detection Methods:

توجد خطوات لابد من اتباعها عند محاولة الكشف عن الأداء التفاضلي للمفردة وذلك بغض النظر عن الطريقة المستخدمة (وهذا ضمن الطرائق التي تستند الى نظرية الاستجابة للمفردة)، في البداية لابد من اختيار إحدى نماذج الاستجابة للمفردة من أجل تقدير معالم الأفراد والمفردات التي سيتم استخدامها أثناء خطوات الكشف عن الأداء التفاضلي للمفردات، حيث يتم اختيار نموذج الاستجابة بناءً على اعتبارات نظرية وعملية، فأما النظرية فتتعلق بطبيعة مفردات الاختبار (معرفية أو لامعرفية) وكذلك بنوع المفردات (ثنائية التدريج أو متعددة التدريج) وأيضاً بنوع المعالم اللازمة لطريقة الكشف المستخدمة DIF سواء معلم الصعوبة أو التمييز أو التخمين، أما الاعتبارات العملية فتتمثل في متغيرات حجم العينة وطول الاختبار وغيرها من المتغيرات الأخرى. وقد تنوعت كثيراً الإجراءات والطرق الإحصائية التي تناولت الكشف عن الأداء التفاضلي للمفردة (Hidalgo & Gómez–Benito, 2010;Millsap & Everson, 1993; Potenza & Dorans, 1995)، وكذلك طرق اختصت بالمفردات متعددة التدريج (Penfield & Lam, 2000; Penfield & Camilli, 2007)، ومن بين هذه الطرق طريقة مانتل–هانزل Mantel–Haenszel (MH) والتي تعتبر من أكثر طرق الكشف عن الأداء التفاضلي للمفردة انتشاراً نظراً

لسهولة استخدامها وملاءمتها للعينات الصغيرة، كما أنها استخدمت في اختبارات Educational Testing Service (ETS) المشهورة وأثبتت جدوى في هذا المجال من خلال الكثير من الأبحاث التي أجريت عليها (Guilera, Gómez-Benito & Hidalgo, 2009; Ackerman & Evans, 1992; Allen & Donoghue, 1996; Clauser, Mazor, & Hambleton, 1991; Clauser, 1993; Mazor, Clauser, & Hambleton, 1992; Mazor et al., 1994; Uttaro & Millsap, 1994)، إلا أن معظم هذه الأبحاث تناولت طريقة مانتل-هانزل من خلال مقارنتها مع طرق كشف أخرى لمعرفة مدى جدواها وفاعليتها، وأما الدراسة الحالية فستركز على الطريقة ذاتها ومدى تأثرها بمتغير حجم العينة الذي يعتبر من المتغيرات المهمة جدا في تحديد معالم الأفراد والمفردات ضمن نظرية الاستجابة للمفردة، هذا فضلاً عن تأثر الأداء التفاضلي للمفردة المتوقع بهذا المتغير.

طريقة مانتل-هانز (MH) للكشف عن الأداء التفاضلي للمفردة:

تعدّ طريقة مانتل هانزل (Mantel & Haenszel, 1959) من أكثر طرائق الكشف عن الاداء التفاضلي للمفردة انتشاراً، حيث أول ما ظهرت ضمن الأبحاث الطبية للكشف عن مدى استقلالية المتغيرات، و يعتمد عمل هذه الإجرائية على المقارنة بين معالم المفردة ضمن المجموعة المرجعية Reference والمجموعة التجريبية Focal ، وقد اهتمت هذه الأجرائية في بادئ الأمر بالمقارنات الثنائية فقط (2x2) أي أن الجدول مكون من صفين وعمودين وبالتالي المفردة من النمط الثنائي (٠، ١) ثم طوّرت فيما بعد لتشمل المفردات متعددة التدريج لتصبح مقارنات الجداول من النمط (2 x C x K) حيث (2) هي عدد الصفوف الممثلة للمجموعين التجريبية والمرجعية و (C) عدد الأعمدة الممثلة لعدد فئات الاستجابة، و (K) عدد المقابلات بين الأعمدة والصفوف لكل مفردة من مفردات الاختبار، ففي حال كانت المفردة ثنائية التدريج تكون عدد المقابلات (2x2)، والجدول (١) التالي يوضح كيفية توزيع جدول المقارنات أثناء حساب الاداء التفاضلي للمفردة متعددة التدريج .

الجدول (١) يوضح توزع معالم المفردة (i) متعددة التدريج بين المجموعتين المرجعية والتجريبية لحساب الـ DIF

المجموعة	R_1	R_2	R_3	.....	R_c	المجموع
المرجعية	N_{r1j}	N_{r2j}	N_{r3j}	.....	N_{rcj}	$N_{r.j}$
التجريبية	N_{f1j}	N_{f2j}	N_{f3j}	.....	N_{fcj}	$N_{f.j}$
المجموع	$N_{.1j}$	$N_{.2j}$	$N_{.3j}$	.....	$N_{.cj}$	$N_{..j}$

مع الإشارة إلى أن (R_c) تمثل الإجابة الصحيحة عن الفئة رقم (c) من المفردة (i).

N_{rcj} تمثل الإجابة الصحيحة عن الفئة (c) للمفردة (i) عند مستوى القدرة (j) ضمن المجموعة المرجعية (r).

N_{fcj} تمثل الإجابة الصحيحة عن الفئة (c) للمفردة (i) عند مستوى القدرة (j) ضمن المجموعة التجريبية (f).

$N_{.cj}$ تمثل مجموع الإجابات الصحيحة عن الفئة (c) للمفردة (i) ضمن المجموعتين المرجعية (r) والتجريبية (f) عند مستوى القدرة (j).

$N_{r.j}$ تمثل مجموع الإجابات الصحيحة عن كل فئات المفردة (i) ضمن المجموعة المرجعية (r) عند مستوى القدرة (j).

$N_{f.j}$ تمثل مجموع الإجابات الصحيحة عن كل فئات المفردة (i) ضمن المجموعة التجريبية (f) عند مستوى القدرة (j).

وباعتبار أن الدراسة الحالية تهتم بالمفردات متعددة التدريج لذلك سيتم الاكتفاء بعرض الدالة الرياضية التي تمثل طريقة حساب معامل مانتل–هانزل والتي تعطى بالعلاقة الآتية:

$$\text{MH} = \frac{\left[\sum_{j=1}^{k} F_j - \sum_{j=1}^{k} E(F_j)\right]^2}{\sum_{j=1}^{k} Var(F_j)} \dots\dots\dots\dots\dots (1)$$

حيث: F_j تمثل الدرجة الكلية للمجموعة التجريبية (F) عند مستوى القدرة (k) وتحسب بالعلاقة: $= \sum_{c=1}^{C} R_c N_{Fcj}$

$E(F_j)$:هو توقع الاجابة الصحيحة للمجموعة التجريبية (F) عند مستوى القدرة (k) ويحسب: $= \sum_{c=1}^{C} P_{jc} \times C$

أما (P_{jc}) فهي الدالة الاحتمالية التي يتم من خلالها حساب احتمال الاجابة الصحيحة عن الفئة (c) للمفردة المفروضة وهي في هذه الحالة تعتبر الدالة الممثلة لنموذج الاستجابة المتدرجة (GR) الذي تم اعتماده في الدراسة الحالية. وأما (C) فتمثل العدد الكلي لفئات المفردة (i).

$Var(F_j)$ هو تشتت الدرجة الكلية للمجموعة التجريبية عند مستوى القدرة (k) وأيضاً يتم حسابه من خلال الدالة الاحتمالية لنموذج الاستجابة المتدرجة.

وبعد الانتهاء من حساب احصاءة (MH) ومن أجل اختبار الفرضية الصفرية بين المجموعتين التجريبية والمرجعية لابد من الاعتماد على توزيع كاي تربيع بدرجة حرية (c−1) حيث (c) هي عدد فئات المفردة، ولمعرفة حجم تأثير الاختلاف الناتج بين المجموعتين يتم حساب القيمة (β) اعتمادا على نسب الأخطاء الشائعة common odds ratio وذلك من خلال العلاقة:

$$\beta_{MH} = \frac{\frac{\sum_{j=1}^{k} A_j D_j}{N_{..j}}}{\frac{\sum_{j=1}^{k} B_j C_j}{N_{..j}}} \quad \dots\dots\dots\dots\dots\dots (2)$$

حيث: (A_j) هي عدد الاجابات الصحيحة عن المفردة (i) عند المستوى (j) للمجموعة المرجعية.

حيث: (B_j) هي عدد الاجابات الخاطئة عن المفردة (i) عند المستوى (j) للمجموعة المرجعية.

حيث: (C_j) هي عدد الاجابات الصحيحة عن المفردة (i) عند المستوى (j) للمجموعة التجريبية.

حيث: (D_j) هي عدد الاجابات الخاطئة عن المفردة (i) عند المستوى (j) للمجموعة التجريبية.

$N_{.j}$ هي عدد الإجابات الكلية عن المفردة(i) عند المستوى (j) للمجموعتين المرجعة والتجريبية معاً.

ثم بعد ذلك ومن أجل سهولة الحكم على المفردة –فيما إذا كانت تبدي أداء تفاضلياً بين المجموعتين أم لا– يتم إجراء التحويل الرياضي الآتي (Holland and Thayer, 1988):

$$\Delta_{\beta_{MH}} = -2.35 \,.\, \ln(\beta_{MH}) \quad \ldots\ldots\ldots\ldots (3)$$

والقمية الموجبة لـ $\Delta_{\beta_{MH}}$ تشير بأن المفردة أكثر صعوبة في المجموعة المرجعية، أما القيمة السالبة لها فتشير إلى أن المفردة أكثر صعوبة في المجموعة التجريبية. ويوجد معيار أكثر تفصيلاً للحكم على المفردة من خلال العلاقة الأخيرة (Zwick & Ercikan, 1989)، حيث تم تقسيم حجم التأثير إلى ثلاثة أنماط:

١. مفردات من النمط (A): وفيه يكون حجم التأثير للأداء التفاضلي للمفردة بين المجموعتين مهملاً (أي أن القيمة ليست دالة إحصائياً)، ويحدث عند تحقق الشرط التالي: $\Delta_{\beta_{MH}} < |1|$

٢. مفردات من النمط (B): وفيه يكون حجم التأثير للأداء التفاضلي للمفردة بين المجموعتين متوسطاً (أي أن القيمة دالة إحصائياً ولكن المفردة لاتصنف على أنها تعاني أداءً تفاضلياً)، ويحدث عند تحقق الشرط الآتي:

$$|1| \leq \Delta_{\beta_{MH}} \leq |1.5|$$

٣. مفردات من النمط (C): وفيه يكون حجم التأثير للأداء التفاضلي للمفردة بين المجموعتين كبيراً (أي أن القيمة دالة إحصائياً)، ويحدث عند تحقق الشرط التالي: $\Delta_{\beta_{MH}} > |1.5|$

وتجدر الإشارة إلى أنه في الدراسة الحالية سيتم الاعتماد على محكٍّ مكافئ للمعيار السابق تم استخدامه في برنامج EasyDIF ، حيث وجد كوماغاي (Kumagai, 2012) أن حساب المساحة (K) المحصورة بين منحنيي المفردة الناتجين عن استجابة المجموعة المرجعية والمجموعة التجريبية والموضحة بالشكل (٢) يكافئ الشرط السابق، ويتم ذلك وفق الخطوات التالية:

١. يتم حساب القيمة (K) من خلال العلاقة التالية:

$$k=\int_{-\infty}^{+\infty}|P_1(\theta) - P_2(\theta)| \,.\, g_T(\theta).d(\theta) \approx \sum_{q=1}^{Q}|P_1(\theta) - P_2(\theta)|.\, g_T.(\theta_q) \quad \ldots\ldots\ldots(4)$$

حيث:

$P_1(\theta)$ احتمالية الإجابة الصحيحة عن المفردة ضمن المجموعة المرجعية.

$P_2(\theta)$ احتمالية الإجابة الصحيحة عن المفردة ضمن المجموعة التجريبية.

g_T يتم حسابها من خلال العلاقة التالية:

$$g_T = \frac{N_1}{N}.g_1 + \frac{N_2}{N}.g_2$$

حيث: (N) العدد الكلي للأفراد في المجموعتين، و (N_1) و (N_2) عدد أفراد المجموعة المرجعية (g_1) والتجريبية (g_2)على التوالي.

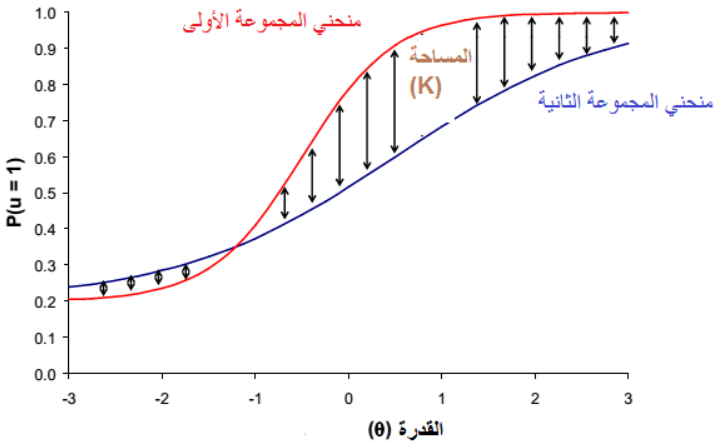
١. بعد حساب القيمة (K) الممثلة للمساحة المحصورة بين المنحنيين يتم مقارنتها بالقيمة التالية:

$$W= (C-1) \times 0.1 \quad \text{.....(5)}$$

حيث (C) هي عدد فئات المفردة، فمثلاً إذا كانت المفردة مؤلفة من خمس فئات كما في الدراسة الحالية فإن:

$$W= (5-1) \times 0.1=0.4$$

٢. إذا كان: K > W فعندها نقول بأن المفردة تبدي أداءً تفاضلياً بين المجموعتين المرجعية والتجريبية، ويشير برنامج EasyDIF إلى ذلك بوضع الإشار (###) بجانب المفردة.



الشكل (٢) يوضح المساحة (K) المطلوب حسابها بين منحنيي المجموعتين المرجعية والتجريبية

مع الإشارة دائما إلى أن أنموذج الاستجابة المتدرجة هو الأنموذج المعتمد في الدراسة الحالية، وهو أحد النماذج الملائمة للمفردات متعددة الفئات والذي تم استخدامه في برنامج EasyDIF لتقدير معالم الافردا والمفردات.

أنموذج الاستجابة المتدرجة Graded Response Model (GRM) :

استخدمت الدراسة الحالية نموذج الاستجابة المتدرجة كأحد نماذج الاستجابة للمفردة متعددة التدريج، وقد قامت بتطوير هذا النموذج ساميجيما (Samejima, 1972) حيث من الممكن استخدامه مع المفردات المكونة من فئتين أو أكثر كما هو الحال في معظم مقاييس الاتجاهات (ليكرت) Likert-type scale والدالة الرياضية الممثلة لهذا الأنموذج هي:

$$p(x \geq k \,/\, \theta) = \frac{e^{D a_i (\theta - b_{ik})}}{1+e^{D a_i (\theta - b_{ik})}} \quad \text{.....(6)}$$

حيث: k هي رقم فئة المفردة (i) .

D ثابت مقياس التدريج وقيمته هي ١.٧

a_i معلم تمييز المفردة (i).

b_iمعلم صعوبة المفردة (i).

ويمكن عدّ هذا الأنموذج تعميماً لنموذج راش ثنائي المعلم، إلا أنه يختلف عنه في اعتماده على طريقة ثورستون Thurstone's approach في تقدير المعالم والتي تقوم على تقسيم فئات الاستجابة للمفردة إلى سلسلة من المفردات الثنائية، أي أنه يستخدم كل المعلومات المتوافرة حول فئات المفردة الواحدة ليصل إلى تقدير المعلم الخاص بكل فئة (Anderson, 2003)، فمثلاً لو افترضنا بأن لمفردة ما أربع فئات استجابة (٤,٣,٢,١) فعندها يعالج هذه المفردة بأن يقسّم هذه الفئات إلى ثلاث مفردات ثنائية لتظهر بالشكل التالي:

$$\underbrace{1 \text{ VS. } 2,3,4}_{b_{i2}} \text{ و } \underbrace{1,2 \text{ VS. } 3,4}_{b_{i3}} \text{ و } \underbrace{(1,2,3 \text{ VS. } 4)}_{b_{i4}}$$

ثم يبدأ بحساب احتمالية الإجابة الصحيحة لكل من هذه المجموعات الثلاثة من خلال تطبيق النموذج الرياضي للاستجابة المتدرجة والموضح بالعلاقة (٦) الاخيرة.

الدراسات السابقة:

وجد كل من (González–Romá et al., 2006) بأن حجم العينة للمجموعة التجريبية إذا كان أقل من (١٠٠) فإن دقة الكشف عن الأداء التفاضلي للمفردة تتناقص، وتتحسن هذه النتيجة عندما ترتفع حجم العينة التجريبية إلى (٢٠٠) ، وقد أرجعت هذه الدراسة سبب ضعف الكشف عن الـ DIF باستخدام العينات الصغيرة إلى افتقار هذه العينات إلى الدقة في تقدير معلم المفردة عندما تكون حجومها صغيرة وخاصة في حال وجود تفاوت مع حجم العينة المرجعية، وفي دراسة أخرى اقترح كل من (Kim, Cohen & Kim, 1994) بأن حجم العينة اللازمة للكشف الدقيق عن الأداء التفاضلي للمفردة يجب ألا يقل عن (٨٠٠) وذلك عند استخدام نموذج راش ثلاثي المعلم (3PLM)، وكذلك وجد كل من (Roussos & Stout, 1996) – في دراسة محاكاة – بأن تناقص حجم العينة التجريبية يقابله تناقص في دقة الكشف عن الاداء التفاضلي للمفردة وذلك عند استخدام طريقة مانتل–هانزل وطريقة SIBTEST

الخطوات العملية للدراسة

أولاً: توليد البيانات

تعدّ الدراسة الحالية دراسة محاكاة Simulation study حيث اختار الباحث توليد البيانات حاسوبياً عبر برنامج WinGen (Han, & Hambleton, 2007) نظراً للحاجة الملحة إلى ضبط متغير حجم العينة (المتغير المستقل للدراسة) لمعرفة مدى تأثيره على طريقة مانتل–هانزل (المتغير التابع) للكشف عن الأداء التفاضلي للمفردة، وتم الاعتماد في ذلك على عدد من الدراسات التي تناولت هذا النوع من البيانات (Ching, 2002; Daniel & Joshua, 2009; Shudong, 1999; Fitzpatrick & Wendy, 2001; Kinsey, 2003)، كما أن بعض الدراسات (العبدالله , 2012; Williams, 2003) قارنت بين نوعي البيانات المولدة والحقيقية ووجدت بأن لا فروق دالة إحصائياً بينهما، وكذلك دراسة (Harwell et al., 1996) والتي بينت بأن استخدام البيانات المولدة (MonteCarlo studies) هو أمر مهم وخاصة في معالجة طرائق الكشف

عن الأداء التفاضلي للمفردات وخاصة في حال كانت حجوم العينة للمجموعة التجريبية صغيرة جداً. وقد تم توليد البيانات حاسوبياً وفقاً للشروط المحددة الآتية:

١. الافتراض بأن البيانات تقيس سمة واحدة عند الافراد، أي أن البيانات من النوع أحادي البعد وذلك لأن تعدد الابعاد يجعل هذا المتغير مؤثراً في الكشف عن الأداء التفاضلي للمفردات وهذا مالا تبحثه الدراسة الحالية، على اعتبار أن متغير حجم العينة هو المتغير المستقل الوحيد.
٢. تم توليد البيانات الخاصة بالأفراد وفقاً للتوزيع الطبيعي بمتوسط صفر وانحراف معياري بمقدار واحد.
٣. تم توليد البيانات الخاصة بالمفردات مطابقة لأنموذج الاستجابة المتدرجة (GRM) كما هو موضح بالجدول (٢) بحيث تتكون المفردة من خمس فئات، ويكون متوسط معلم الصعوبة لكل فئة مساوياً للصفر وبانحراف معياري مقداره الواحد وفقاً للتوزيع الطبيعي N(0 , 1). مع الاشارة الى أن معلم التمييز يكون للمفردة ككل وليس لكل فئة على حده.
٤. من خصائص المنحني المميز للمفردة الخاص بنموذج الاستجابة المتدرجة أن معلم التخمين ثابت لكل مفردات الاختبار؛ لذلك تم التركيز على معلمي الصعوبة والتمييز فقط.
٥. عدد المفردات (طول الاختبار) المعتمد (٢٠) مفردة وجميعها مكونة من خمس فئات.
٦. تم توليد ثلاثة حجوم مختلفة من العينة بناءً على معالم المفردات المولَّدة، حيث تراوحت حجوم العينة بين (٣٠٠) و (٥٠٠) و (١٠٠٠) على التوالي، وذلك مناصفة بين المجموعتين المرجعية والتجريبية بحيث تكونت كل مجموعة من (١٥٠) و (٢٥٠) و (٥٠٠) على التوالي.

الجدول (٢) يوضح معالم المفردات وفئاتها

المفردة	النموذج	عدد الفئات	معلم الصعوبة (خطوة "١")	معلم الصعوبة (خطوة "٢")	معلم الصعوبة (خطوة "٣")	معلم الصعوبة (خطوة "٤")	معلم التمييز
1	GRM	5	−0.361	−0.992	0.160	0.862	1.322
2	GRM	5	0.417	−1.134	−1.080	0.349	1.282
3	GRM	5	−0.839	0.235	0.376	0.881	1.393
4	GRM	5	−0.199	−1.505	−0.729	−0.451	0.234
5	GRM	5	2.229	−0.831	−0.255	−0.142	0.807
6	GRM	5	0.701	−1.916	−0.805	−0.127	0.639
7	GRM	5	−0.270	−0.196	0.516	0.996	1.679
8	GRM	5	−0.174	−0.875	−0.396	0.745	1.545

9	GRM	5	−0.692	−1.061	−0.562	0.464	0.780
10	GRM	5	−0.282	0.274	1.254	2.040	2.655
11	GRM	5	−0.356	−0.065	0.854	0.884	1.043
12	GRM	5	−1.478	−0.766	−0.092	0.188	1.453
13	GRM	5	0.000	−1.882	0.303	0.392	0.987
14	GRM	5	0.178	−1.261	−0.569	−0.412	0.373
15	GRM	5	1.411	−1.012	−0.249	−0.146	1.212
16	GRM	5	−0.549	−0.402	−0.142	0.237	0.404
17	GRM	5	−0.764	−1.029	−1.013	−0.971	0.312
18	GRM	5	1.881	0.351	0.395	0.878	1.401
19	GRM	5	−0.117	−1.091	−0.487	−0.007	1.309
20	GRM	5	−1.078	−0.739	−0.260	−0.192	0.220

ثانياً: التحقق من افتراضات نظرية الاستجابة للمفردة

بعد الانتهاء من خطوات توليد البيانات اللازمة لاختبار الأداء التفاضلي للمفردات بين المجموعتين المرجعية والتجريبية، كان لابد من التحقق من افتراضات نظرية الاستجابة للمفردة والمتمثلة بأحادية البعد، والاستقلال الموضعي، وتوازي المنحنيات المميزة للمفردة، وتعدّ هذه الخطوة شرطاً أساسياً لابد منه بغض النظر عن طبيعة متغيرات الدراسة أو النموذج المستخدم فيها، إلا أن الدراسة الحالية ستعتبر أن هذه الشروط محققة مسبقاً لأن البيانات تم توليدها وفق لشروط تراعي افتراضات نظرية الاستجابة للمفردة من حيث المبدأ؛ فبرنامج توليد البيانات يتيح للمستخدم اختيار نوع البيانات المولدة فيما اذا كانت أحادية أو متعددة التدريج، كما يتيح إمكانية اختيار نوع التوزيع للأفراد والمفردات، بالإضافة إلى تحقيقه لخاصية الاستقلال الموضعي للمفردة بغض النظر عن عينة الاستجابة. وبذلك يمكننا الانتقال الى الخطوة التالية وهي اختبار المتغير المستقل للدراسة والمتمثل في حجم العينة وتأثيره على الاداء التفاضلي للمفردة (DIF) .

ثالثاً: تحليل الأداء التفاضلي للمفردات بين المجموعتين المرجعية والتجريبية بطريقة مانتل−هانزل:

يتم استيراد الملف النصي للبيانات المولَّدة بالخطوة السابقة الى برنامج التحليل EasyDIF، ومن ثم تحديد بعض الخيارات البرمجية كطريقة قراءة البرنامج لملف البيانات من خلال تحديد رقم العمود الأول للقراءة، وكيفية تعامله مع الإجابات المهملة والتي يُملأ مكانها بحرف بدلاً من الرقم فمثلاً في الدراسة الحالية تم اختيار الحرف (p) بحيث أن البرنامج سيهمل أي إجابة عن المفردات تحمل هذا الرمز، ثم يتم تحديد عدد مجموعات المقارنة وهي في دراستنا مجموعتان (المجموعة المرجعية والمجموعة التجريبية)، ثم تحديد الخيار المتعلق بعدد فئات

المفردة، وأخيراً استخدام خيار قراءة ملف البيانات للتأكد من سلامته قبل البدء بعملية التحليل، وبعد الانتهاء من تجهيز ملف البيانات وفق كل الخطوات السابقة نأتي إلى الخطوة الأهم وهي تحديد المفردات التي سيتم معالجتها لمعرفة الأداء التفاضلي لها عبر المجموعتين، وفي هذه الخطوة يتيح لنا البرنامج ثلاث خيارات: الأول (All USE) وباختياره نكون قد وضعنا كل المفردات تحت الكشف عن الأداء التفاضلي لها عبر المجموعتين، والخيار الثاني (NonDIF) والذي يلغي الكشف عن الاداء التفاضلي لمفردة معينة يتم اختيارها قبل البدء بالتحليل، والخيار (Study) ويعمل عكس الخيار الأخير حيث يضع المفردة المختارة تحت الكشف والتحليل عن أدائها التفاضلي، وأخيراً الخيار (All not USE) وباختياره نكون قد ألغينا عملية الكشف عن الأداء التفاضلي لجميع مفردات الاختبار. وفي الدراسة الحالية تم وضع جميع المفردات تحت الكشف عن الأداء التفاضلي للمفردة وذلك باختيار الخيار الأول (All USE) وذلك في جميع مراحل التحليل لحجوم العينات المختلفة. وبعد الانتهاء من تحديد كل شروط التحليل الموضحة سابقاً يُعطى البرنامج أمر البدء بالكشف عن الأداء التفاضلي للمفردة لنحصل بذلك على مخرجات التحليل والتي تتلخص بمعالم فئات الصعوبة لكل مفردة وكذلك معلم التمييز بالإضافة إلى المحكّ الأهم في دراستنا وهو المحك (K) المعني بالحكم على المفردة فيما إذا كانت تعاني أداءً تفاضلياً أم لا بين المجموعتين المرجعية والتجريبية، ومن مخرجات البرنامج ايضاً معلم القدرة اللازمة للإجابة بشكل صحيح عن كل مفردة من مفردات الاختبار وكل ذلك بالنسبة لمجموعتي المقارنة عبر كل مراحل التحليل. وسنعرض فيما يأتي لنتائج التحليل وفقاً لحجوم العينات الثلاثة لمعرفة مدى تأثر الأداء التفاضلي لكل مفردة باختلاف حجم العينة:

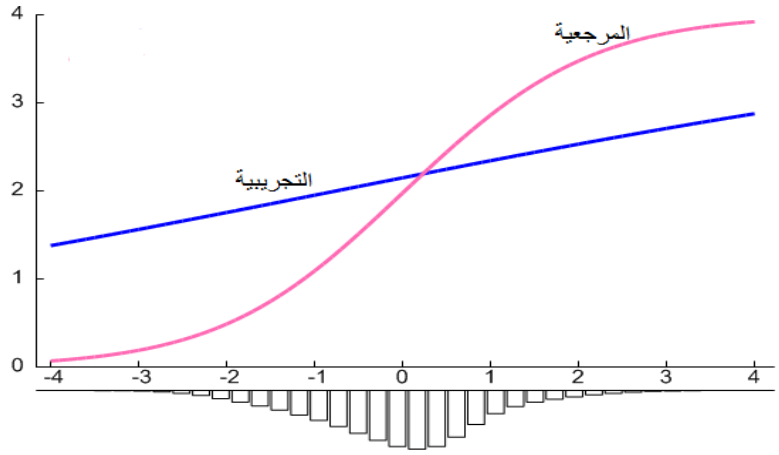
١. بالنسبة لحجم العينة (N=300):

في حال كان حجم العينة (٣٠٠) مناصفة بين المجموعتين المرجعية والتجريبية، وبعد حساب محك تقييم الأداء التفاضلي للمفردة (K) من خلال تطبيق العلاقة رقم (٤) ومقارنته بالقيمة (0.4=0,1x(1−5)) التي تم حسابها من العلاقة رقم (٥) كما سبق ووضحنا، نتج لدينا ثلاث مفردات أحدثت فروقاً بين المجموعتين اللتين تملكان متوسط القدرة ذاته وهي المفردات الثانية والثالثة والرابعة المشار إليهم بالجدول (٣) من خلال العمود الأخير، مما يدل على أن هذه المفردات أحدثت أداءً تفاضليا بين المجموعتين المرجعية والتجريبية. وبملاحظة الرسم البياني لهذه المفردات والموضحة بالشكل (٣) يتبين مدى الاختلاف الحاصل في تقييم المفردة ذاتها بين مجموعتي المقارنة وذلك من خلال الفروقات الواضحة بين المنحنيين والتي تدل على اختلاف الاستجابات بين المجموعتين على الرغم من تساوي متوسط القدرة بينهما.

الجدول(٣) يوضح المفردات التي تعاني أداءً تفاضلياً بين المجموعتين المرجعية والتجريبية

التعريف برموز الجدول: (وتنسحب على جميع الجداول التالية)

I : رقم المفردة، G: نوع المجموعة حيث (F) المجموعة التجريبية و (R) المجموعة المرجعية، DIF الأداء التفاضلي للمفردة، b1 و b2 و b3 و b4 تمثل معالم صعوبة الخطوات الأولى والثانية والثالثة والرابعة على التوالي للمفردة، (a) هي معلم التمييز للمفردة، (K) يمثل محك تقييم الأداء التفاضلي للمفردة وقد سبق توضيح طريقة حسابه.



الأداء التفاضلي للمفردات الثانية والثالثة والرابعة عبر المجموعتين التجريبية والمرجعية

الشكل (٣) يوضح الاداء التفاضلي للمفردات الثانية والثالثة والرابعة عبر مجموعتي المقارنة التجريبية والمرجعية عندما (N=300)

١) بالنسبة لحجم العينة (N=500):

في حال كان حجم العينة (٥٠٠) فردا مناصفة بين المجموعتين المرجعية والتجريبية، وبعد حساب محك تقييم الأداء التفاضلي للمفردة (K)، ومقارنته بالقيمة (0.4) كما سبق ووضحنا، نتج لدينا مفردة واحدة أحدثت فروقاً بين المجموعتين اللتين تملكان متوسط القدرة ذاته وهي المفردة الخامسة المشار اليها بالجدول(٤) من خلال العمود الأخير، مما يدل على أن هذه المفردة أحدثت أداءً تفاضليا بين المجموعتين المرجعية والتجريبية. وبملاحظة الرسم البياني لهذه المفردة والموضح بالشكل (٤) يتبين مدى الاختلاف الحاصل في تقييم المفردة ذاتها بين مجموعتي المقارنة، حيث أن الاختلاف واضح بين منحنيي المفردة عبر المجموعتين التجريبية والمرجعية.

الجدول(٣) يوضح المفردات التي تعاني أداءً تفاضلياً بين المجموعتين المرجعية والتجريبية

N = 300												
DIF		a		b4		b3		b2		b1		الفئة
الحالة	*K*	*F*	*R*	*F*	*R*	*F*	*R*	*F*	*R*	*F*	*R*	G / I
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1**
###	0.506	0.614	0.119	0.937	1.742	0.299	-0.245	-0.540	-2.120	-0.572	-2.398	**2**
###	0.555	0.614	0.119	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**3**
###	0.555	0.614	0.119	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**4**
	0.127	1.285	1.445	0.744	0.532	-0.252	-0.277	-0.374	-0.391	-1.257	-0.843	**5**
	0.202	0.264	0.455	0.490	0.416	-0.900	-0.253	-1.718	-0.810	-3.175	-2.195	**6**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**7**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**8**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**9**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**10**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**11**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**12**
	0.217	0.264	0.455	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**13**
	0.204	0.264	0.115	0.000	-0.15	0.000	-0.282	0.000	-0.554	0.000	-0.828	**14**
	0.167	0.805	0.712	1.108	1.073	-0.296	-0.574	-0.434	-0.666	-1.455	-1.906	**15**
	0.071	0.805	0.712	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**16**
	0.071	0.805	0.712	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**17**
	0.219	1.048	0.878	1.414	1.701	0.804	0.949	0.150	0.468	0.0543	0.442	**18**
	0.105	1.048	0.878	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**19**
	0.105	1.048	0.878	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**20**

الشكل (٣) يوضح الاداء التفاضلي للمفردات الثانية والثالثة والرابعة عبر مجموعتي المقارنة التجريبية والمرجعية عندما (N=300)

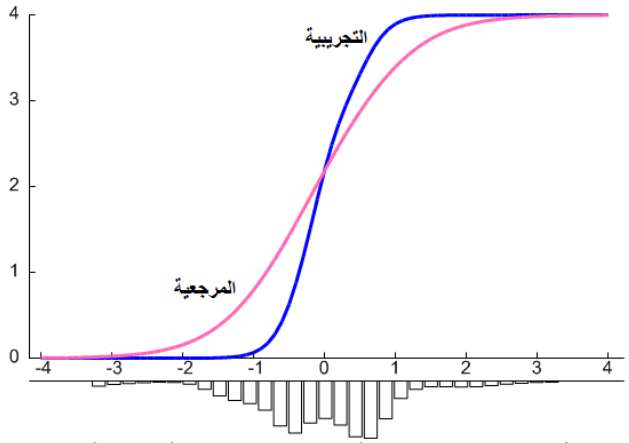
١. بالنسبة لحجم العينة (N=500):

في حال كان حجم العينة (٥٠٠) فردا مناصفة بين المجموعتين المرجعية والتجريبية، وبعد حساب محك تقييم الأداء التفاضلي للمفردة (K)، ومقارنته بالقيمة (0.4) كما سبق ووضحنا، نتج لدينا مفردة واحدة أحدثت فروقاً بين المجموعتين اللتين تملكان متوسط القدرة ذاته وهي المفردة الخامسة المشار اليها بالجدول(٤) من خلال العمود الأخير، مما يدل على أن هذه المفردة أحدثت أداءً تفاضليا بين المجموعتين المرجعية والتجريبية. وبملاحظة الرسم البياني لهذه المفردة والموضح بالشكل (٤) يتبين مدى الاختلاف الحاصل في تقييم المفردة ذاتها بين مجموعتي المقارنة، حيث أن الاختلاف واضح بين منحنيي المفردة عبر المجموعتين التجريبية والمرجعية.

الجدول(٤) يوضح المفردات التي تظهر أداءً تفاضلياً بين المجموعتين المرجعية والتجريبية

N = 500												
DIF		A		b4		b3		b2		b1		الفئة
الحالة	***K***	***F***	***R***	***F***	***R***	***F***	***R***	***F***	***R***	***F***	***R***	G/I
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1**
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**2**
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**3**
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**4**
###	0.463	1.154	3.064	0.780	0.571	-0.111	-0.068	-0.212	-0.117	-0.880	-0.430	**5**
	0.103	0.327	0.370	1.197	0.749	0.237	-0.275	-0.967	-0.889	-2.503	-1.998	**6**
	0.000	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**7**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**8**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**9**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**10**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**11**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**12**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**13**
	0.052	0.327	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**14**
	0.162	0.705	0.580	1.457	1.824	0.143	-0.184	0.019	-0.242	-0.823	-1.103	**15**
	0.112	0.705	0.580	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**16**
	0.112	0.705	0.580	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**17**
	0.097	1.233	1.193	1.325	1.476	0.340	1.015	0.379	0.505	0.337	0.423	**18**
	0.020	1.233	1.193	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**19**
	0.020	1.233	1.193	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**20**

مع الإشارة إلى أن رموز الجدول تم توصيفها آنفاً.



الأداء التفاضلي للمفردة الخامسة عبر المجموعتين التجريبية والمرجعية

الشكل (٤) يوضح الاداء التفاضلي للمفردة الخامسة عبر مجموعتين المقارنة التجريبة والمرجعية عندما (N=500)

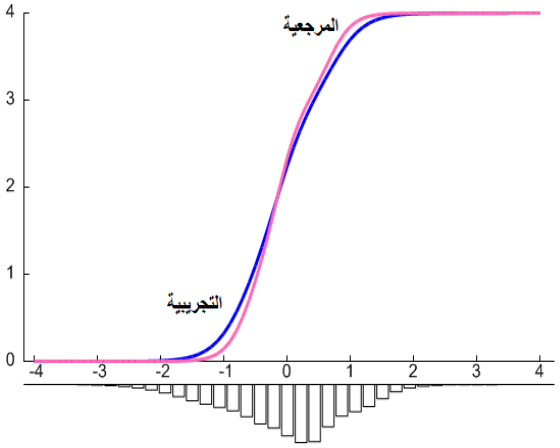
٢. بالنسبة لحجم العينة (N=1000):

في حال كان حجم العينة (١٠٠٠) فرد مناصفة بين المجموعتين المرجعية والتجريبية، وبعد حساب محك تقييم الاداء التفاضلي للمفردة (K) من خلال تطبيق العلاقة رقم (٤) ومقارنته بالقيمة (0.4) التي تم حسابها من العلاقة رقم (٥) يتضح بعدم وجود أية مفردة أحدثت أداءً تفاضلياً بين المجموعتين التجريبية والمرجعية، وهذا ما يبينه العمود الأخير من الجدول (٥)، وهذا يدل على أن جميع المفردات متجانسة بإجاباتها عبر مجموعتي المقارنة، وبأخذ مثال توضيحي للرسم البياني لإحدى هذه المفردات يتبين لنا ذلك، فمثلاً الشكل (٥) يوضح بأنه لاتوجد فروقات دالة بين استجابة المجموعتين عن المفردات الثانية والثالثة والرابعة والتي كانت تعاني من أداء تفاضلي عندما كان حجم العينة مساوياً لـ (٣٠٠)، وكذلك الأمر بالنسبة لأداء المفردة الخامسة والتي كانت تعاني من أداء تفاضلي عندما كان حجم العينة (٥٠٠).

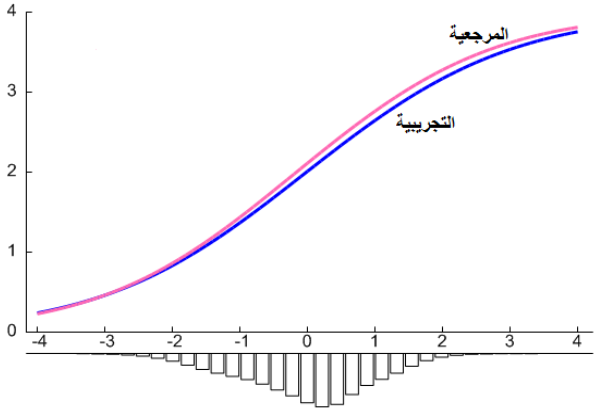
الجدول(٥) يوضح المفردات التي تعاني أداءً تفاضلياً بين المجموعتين المرجعية والتجريبية

N = 1000												
DIF		a		b4		b3		b2		b1		الفئة
الحالة	*K*	*F*	*R*	*F*	*R*	*F*	*R*	*F*	*R*	*F*	*R*	G / I
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1**
	0.095	0.461	0.429	1.138	1.279	0.370	0.363	-1.061	-0.808	-1.074	-0.861	**2**
	0.034	0.461	0.429	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**3**
	0.034	0.461	0.429	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**4**
	0.109	3.160	2.299	0.676	0.754	-0.096	-0.092	-0.170	-01.66	-0.640	-0.729	**5**
	0.194	0.578	0.843	0.503	0.356	-0.230	-0.226	-0.929	-0.697	-2.289	-1.660	**6**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**7**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**8**

	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**9**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**10**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**11**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**12**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**13**
	0.217	0.578	0.843	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**14**
	0.047	1.628	1.458	1.075	1.057	-0.102	-0.139	-0.202	-0.252	-0.884	-0.953	**15**
	0.062	1.628	1.458	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**16**
	0.062	1.628	1.458	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**17**
	0.017	2.230	2.068	1.313	1.270	0.828	0.859	0.383	0.392	0.368	0.382	**18**
	0.037	2.230	2.068	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**19**
	0.037	2.230	2.068	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**20**



الأداء التفاضلي للمفردة الخامسة عبر المجموعتين التجريبية والمرجعية



الأداء التفاضلي للمفردات الثانية والثالثة والرابعة عبر المجموعتين المرجعية والتجريبية

الشكل (٥) يوضح المفردات التي كانت تعاني من أداء تفاضلي عندما كانت حجوم العينة (٣٠٠) و (٥٠٠)

تفسير النتائج:

من خلال ماسبق يظهر التأثير الواضح والمهم لحجم العينة أثناء تحليل الأداء التفاضلي للمفردة، ففي الحالة الأولى عندما كان حجم العينة (٣٠٠) نتج عن التحليل ثلاث مفردات ذات أثر تفاضلي بين مجموعتي المقارنة، وعند زيادة حجم العينة الى (٥٠٠) اختلفت نتائج التحليل من خلال انخفاض عدد المفردات التي تعاني أداءً تفاضلياً الى مفردة واحدة ومختلفة عن المفردات الثلاثة السابقة، وعند متابعة زيادة حجم العينة المطبقة على المفردات ذاتها في المرة الثالثة الى (١٠٠٠) لاحظنا كيف اختفى الأداء التفاضلي للمفردات بحيث أظهرت جميع المفردات تجانساً بالاستجابات بين مجموعتي المقارنة في حال كانت معالم القدرة متساوية، وهذا يدلل الى العلاقة العكسية بين زيادة حجم العينة وعدد المفردات التي تظهر أداء تفاضلياً أثناء التحليل، وبهذه النتيجة يتأكد لنا مدى أهمية حجم عينة التطبيق عند محاولة اختبار أداء المفردات ضمن الاختبارات والمقاييس التي تستخدم لتقييم أداء الأفراد أو لقياس نواتج التعليم ضمن المؤسسات التعليمية. ولو حاولنا التدقيق في أسباب

حدوث هذه الفوارق بين النتائج لوجدنا بأن العوامل الأساسية التي تسهم في رسم المخطط البياني لمجموعتي المقارنة هي عدد أفراد العينة، وبناء عليه كلما زاد هذا العدد وصلنا الى شكل بياني أكثر إيضاحاُ ولمَّا كانت التقنية المستخدمة للكشف عن الاداء التفاضلي للمفردة تستند إلى حساب هذه المساحة فهذا يعني أن العنصر المؤثر في حسابها (والذي هو عدد الأفراد) هو صاحب الأثر المباشر في التحكم بمقدار الفروقات بين مجموعتي المقارنة، مما يدلل دون شك إلى ضرورة الانتباه دائما إلى عدد أفراد المجموعات التي هي محل المقارنة، فكما بينت الدراسة الحالية بأن حجم العينة إذا نقص عن (٥٠٠) فرد فإن طريقة مانتل – هانزل أبدت فروقا معتبرة في الأداء التفاضلي للمفردات، ونستطيع أن نوصي بناء على هذه النتائج بأن لا يقل حجم العينة عن هذا الحجم، والأفضل دائما أن يتم استخدام أكثر من طريقة واحدة للتأكد من المفردات التي تبدي أداء تفاضليا، وبذلك يمكن أن نتوصل إلى قرارات أكثر دقة فيما يتعلق بتعديل أو استبعاد المفردات التي تبدي أداء تفاضليا، وتظهر ضرورة هذه الخطوة في تقنين المقاييس النفسية واختبارات الذكاءات المتعددة؛ حيث من الممكن لبعض المفردات أن تتأثر بالثقافة الأصلية التي طُبّق المقياس في رحابها، وهذا ما يفسر اختلاف النتائج المترتبة عن تطبيق كثير من الاختبارات وخاصة الأجنبية، حيث يلجأ الباحث إلى ترجمة هذه المقاييس من اللغة الأم التي صُمِّم فيها المقياس إلى اللغة المراد تطبيقه عليها، ثم يبدأ بالمقارنة بين نتائجه ونتائج الدراسات الأخرى معتبراً أن أداة القياس ثابتة تماماً، وهذا من شأنه أن يؤدي إلى الوقوع في خطأ تفسير النتائج والتعليق عليها.

التوصيات:

تعددت طرائق القياس بين الاختبارات وبطاقات الملاحظة والمقاييس وغيرها من الطرائق المعروفة عند أصحاب الاختصاص، إلا أن المشكلة تكمن في مدى تحقيق أداة القياس للعدالة بين الأفراد، حيث وجد كل من (Tan & Gierl, 2005) بأن الكشف عن الأداء التفاضلي للمفردة قبل تطبيقها لتقييم أداء الفرد هي إحدى أهم خطوات تحقيق العدالة أثناء قياس قدرات الأفراد، فإذا كانت المفردة تؤدي دوراً متبايناً فإن ذلك سيشكل خللاً كبيراً في تحقيق العدالة والمساواة سواء في قياس ناتج التعليم أو في التقييم لأداء فرد معين بهدف اختياره للعمل في مؤسسة أو شركة ما، لأن التقييم يكون بناء على استجابة الفرد عن مفردات تم تحديد مواصفاتها بدقة من حيث الصعوبة والتمييز والتخمين والاداء التفاضلي، وبعد عملية التحديد هذه يصبح للمفردة معلم واضح يقابله القدرة المطلوبة للإجابة بشكل صحيح عن هذه المفردة، وبعد ذلك يتم تخزينها فيما يسمى ببنك الأسئلة لكي يتم الاستفادة منها في عملية قياس القدرات، فإذا كانت المفردة تعاني خللاً ما في إحدى معالمها فإن ذلك سينعكس سلباً على عملية القياس برمتها لنصل في النهاية إلى قرارٍ خاطئ قد يؤثر على سير عمل المؤسسة ومدى نجاحها، لذلك توصي الدراسة الحالية بالحرص الدقيق على معرفة العوامل المؤثرة في تحديد مواصفات المفردة وأدائها التفاضلي بين المجموعات المختلفة سواء بالجنس أو بالعرق أو بالثقافة، وذلك قبل تخزين هذه المفردة في بنك الأسئلة والاستفادة منها في قياس القدرات، وتبرز أهمية هذه العملية ضمن اختبارات المواءمة المحوسبة Computerized Adaptive Testing (CAT) والتي تعتمد على قياس القدرات بناء على اختبار آني قبلي

يتم من خلاله تحديد قدرة الفرد ومن ثم اعطائه صورة اختبارية مناسبة لقدرته، فإذا كانت المفردات المستخدمة في التقييم المبدئي لقدرة الفرد تعاني أداءً تفاضلياً فإن هذا سيترتب عليه إعطائه صورة اختبارية ليست متناسبة مع قدرته الأمر الذي سيؤدي دون شك لاتخاذ قرارٍ خاطئ يمس موضوعية وعدالة عملية القياس برمَّتها.

المقترحات

من خلال النتائج التي توصلت إليها الدراسة يقترح الباحث إجراء الدراسات التالية كأبحاث مستقبلية:

١. إعادة الدراسة الحالية من خلال استخدام بيانات حقيقية ومقارنتها مع البيانات المولدة حاسوبيا للوقوف على دقة النتائج.
٢. مناقشة مسألة تأثير حجم العينة على الأداء التفاضلي للمفردات من خلال استخدام نماذج أخرى من نماذج الاستجابة للمفردة أنموذج التقدير الجزئي وأنموذج التقدير الجزئي المعمم وغيره من النماذج المستخدمة في هذا المجال.
٣. إجراء دراسة مقارنة للأداء التفاضلي للمفردات في حال استخدام نماذج ثنائية التدريج ونماذج متعددة التدريج لأن تأثير حجم العينة يختلف باختلاف هذه النماذج عموما؛ فالنماذج متعددة التدريج عادة ما تتطلب حجوم عينة أكثر مما هو عليه في النماذج ثنائية التدريج.
٤. دراسة أثر طول الاختبار على الأداء التفاضلي للمفردة، فقد يكون عاملا مؤثرا في الحكم على أداء المفردات.

Recommendations

The current study recommends careful attention to the factors affecting determining the specifications of the item and its differential performance between different groups, whether by gender, race or culture, before storing this item in the item bank and benefiting from it in measuring capabilities, and the importance of this process is highlighted in computerized adaptive testing, which depends on the measurement of capabilities based on a pre-test to determine the individual's ability and then examine him with a test appropriate to his ability. If the used item in the pre-assessment of an individual's ability suffers from a differential performance, this will give him a test that is disproportionate to his ability, and this will undoubtedly lead to a wrong decision that affects the objectivity and fairness of the measurement process.

Suggestions for future research

Through the results of the study, the researcher suggests conducting the following studies as future research:

1. Conducting a study using real data and comparing it with simulation data to determine the accuracy of the results.
2. Discuss the issue of the effect of sample size on the differential item functioning through using other models of item response models such as the partial credit model, generalized partial credit model and other models in this field.
3. Carrying out a study to compare the differential item functioning when the item response models differed between polytomous and dichotomous, since polytomous models typically require a larger sample size than they do in dichotomous models.
4. Studying the effect of the length of the test on the differential item functioning, it may play an influencing role in judging the item performance.

المصادر

١. زياد أحمد العبدالله (٢٠١٢). أثر بعض طرق التقدير على دقة تقدير المعالم ضمن نماذج الاستجابة للمفردة متعددة التدريج. رسالة دكتوراه: معهد الدراسات التربوية. جامعة القاهرة.

1. Abdullah, Z. (2012) Effect of some Estimation Methods on Accuracy of Estimating Parameters in Polytomous Item Response Models. Unpublished Doctoral Dissertation, Institute of Educational Studies, Cairo University.
2. Ackerman, T. A. & Evans, J. A. (1992). An investigation of the relationship between reliability, power, and the Type I error rate of the Mantel-Haenszel and simultaneous item bias detection procedures. Annual Meeting of the National Council on Measurement in Education, San Francisco.
3. Allen, N. L., & Donoghue, J. R. (1996). Applying the Mantel-Haenszel procedure to complex samples of items. Journal of Educational Measurement, 33 (2), 231-251.
4. Ching-Fung B. Si. (2002). Ability Estimation Under Different Item Parameterization and Scoring Models. Unpublished Doctoral Dissertation, University of North Texas.
5. Clauser, B. E., Mazor, K., & Hambleton, R. K. (1991). An examination of item characteristics on Mantel-Haenszel detection rates. Annual Meeting of the National Council on Measurement in Education, Chicago.
6. Daniel, J., Joshua, G. (2009). A Comparison of IRT Parameter Recovery in Mixed Format Examinations Using PARSCALE and ICL. Poster presented at the Annual meeting of Northeastern Educational Research Association.
7. Finch, W. H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel_Haenszel, SIBTEST and IRT likelihood ratio. Applied Psychological Measurement, 29, 278-295.
8. Fitzpatrick, A. R., Wendy, M. Y. (2001). The Effects of Test Length and Sample Size on the Reliability and Equating of Tests Composed of Constructed Response Items. Applied Measurement In Education, 14(1), 31-57.

9. González-Romá, V., Hernández, A., & Gómez-Benito, J. (2006). Power and Type I error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. Multivariate Behavioral Research, 41(1), 29-53.
10. Guilera, G.; Gómez-Benito, J. & Hidalgo, M.D. (2009). Scientific production on the Mantel-Hanszel procedure as a way of detecting DIF. Psicothema, 21 (3), 492-498.
11. Han, K. T., & Hambleton, R. K. (2007). User's Manual: WinGen (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts, School of Education.
12. Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. Applied Psychological Measurement, 20(2), 101-125.
13. Hidalgo, M. D., & Gómez-Benito, J. (2010). Education measurement: Differential item functioning. In P. Peterson, E. Baker, & B. McGaw (Eds.), International Encyclopedia of Education (3rd edition). USA: Elsevier - Science & Technology.
14. Hidalgo, M. D., & Gómez-Benito, J. (2010). Education measurement: Differential item functioning. In P. Peterson, E. Baker, & B. McGaw (Eds.), International Encyclopedia of Education (3rd edition). USA: Elsevier - Science & Technology.
15. Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum
16. Kim, S. H., Cohen, A. S., & Kim, H. O. (1994). An investigation of Lord's procedure for thedetection of differential item functioning. Applied Psychological Measurement, 18(3),217-228.
17. Kinsey, Tari L. (2003). A Comparison of IRT and Rasch Procedures in a Mixed Item Format Test. Unpublished Doctoral Dissertation, University of North Texas.

18. Kumagai, R. (2012) A new method for estimating differential item functioning (DIF) for multiple groups and polytomous items: Development of index K and the computer program "EasyDIF". Japanese Journal of Psychology, 83(1), 35-43. (in Japanese)

19. Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from

retrospective studies of disease. Journal of the -ational Cancer Institute, 22, 719-748.

20. Mazor, K., Clauser, B. E., & Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. Educational and Psychological Measurement, 54 , 284-291.

21. Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. Journal of Educational Statistics 7, 105-118.

22. Millsap, R. E. & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. Applied Psychological Measurement 17, 297-334.

23. Millsap, R. E. & Everson, H. T.(1993). Methodology review: Statistical approaches for assessing measurement bias. Applied Psychological Measurement 17, 297-334

24. Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao, & S. Sinharay (Eds.), Psychometrics (pp. 125-168; 5). Amsterdam: Elsevier

25. Penfield, R. D., & Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment: review and recommendations. Educational Measurement: Issues and Practice, 19 (3), 5-15.

26. Potenza, M. T. & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. Applied Psychological Measurement,19, 23-37.

27. Potenza, M. T. & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. Applied Psychological Measurement, 19, 23-37.

28. Raju, N. S. & Ellis, B. B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), Measuring and Analyzing Behavior in Organizations: Advances in Measurement and Data Analysis (pp. 156-188). San Francisco, CA: Jossey-Bass.

29. Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample sizeand studied item parameters on SIBTEST and Mantel-Haenszel type I errorperformance. Journal of Educational Measurement, 33(2), 215-230.

30. Shudong, W. (1999). The Accuracy of Ability Estimation Methods for Computerized Adaptive Testing Using The Generalized Partial Credit Model. Unpublished Doctoral Dissertation, University of Pittsburgh.

31. Tan, X., & Gierl, M. J. (2005). Using local DIF analyses to assess group differences on multilingual examinations. Poster presented at the annual meeting of the National Council on Measurement in Education.

32. Uttaro, T., & Millsap, R. E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. Applied Psychological Measurement, 18, 15-25.

33. Williams, N. J. (2003). Item and Person Parameter Estimation using Hierarchical Generalized Linear Models and Polytomous Item Response Theory Models. Unpublished Doctoral Dissertation, University of Texas at Austin.

34. Zwick, R. and Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. Journal of Educational Measurement, 26, 55-66.