مدى دقة تقدير صعوبة الفقرات ثنائية التدريج وفق النموذج الهرمي باختلاف نسبة الفقرات ذات الأداء التفاضلي للجنس ومستوى ذلك الأداء

محمد عبد الجبار أبو صهيون *

ملخص

هدفت الدّراسة الحاليّة إلى مدى الدقة في تقدير معلمة صعوبة الفقرات استنادًا الى نموذج راش الهرمي بوجود الأداء التفاضلي بنسب مختلفة وبدرجات متفاوتة باستخدام بياناتٍ مولّدةٍ. ولتحقيق غرض الدّراسة تم استخدام المنهج الوصفيّ في هذه الدّراسة عن طريق توليد فقراتٍ ذات استجابةٍ ثنائيّةٍ (0، 1) لعيّناتٍ تحاكي عيّنات المجتمع الأصليّ بطريقة المونتي كارلو (Monte Carlo) باستخدام برنامج (WinGen v.3) حيث بلغ عدد الأفراد (12000) فردًا بواقع (1200) فرد لكلّ ظرف من ظروف البحث العشرة الناتجة عن تفاعل متغيّريّ الدّراسة [نسبة الفقرات ذات الأداء التّفاضليّ (10%، 25%، 50%)، مستوى الأداء التّفاضليّ (0.5، 1.0، 1.5)] حيث تغطي مستويات الأداء التفاضلي المنخفض والمتوسط والمرتفع على التوالي، علمًا بأنّ الاختبار مكونٌ من (60) فقرةً لكلّ ظرفٍ بحثيٍّ.

أظهرت نتائج الدّراسة أنّ أدقّ التقديرات لمعلمة الصّعوبة المقدّرة عندما تخلو فقرات الاختبار من الأداء التفاضليّ (اي عندما تكون نسبة الفقرات التي تبدي اداءً تفاضليًّا 0%) وكذلك عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (10%) لجميع مستويات الأداء التّفاضليّ المعتمدة في الدراسة وهي: (0.5، 1.0، 1.5)، وعندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) عند مستوى الأداء التّفاضليّ (1.5)، في حين كانت أقل التقديرات دقةً لمعلمة الصّعوبة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) لكلٍّ مستوى من مستويات الأداء التّفاضليّ (0.5، 1.0).

الكلمات الدالة: نسبة الفقرات ذات الأداء التّفاضليّ، مستوى الأداء التّفاضليّ، نموذج راش الهرميّ، معلمة صعوبة الفقرة، الخطأ المعياريّ لمعلمة صعوبة الفقرة.

المقدمة

تعدّ الاختبارات من أهمّ أدوات القياس منذ زمن بعيد، فقد اهتمّ أصحاب القرار وعلماء النفس والتربويّون بتطوير هذه الاختبارات، محاولين الحصول على أدقّ التّقديرات حول ما تحمله الفقرة من معالم. واهتمّ الكثير من العلماء بأدوات القياس بشكلٍ عامٍّ وبالاختبارات بشكلٍ خاصٍّ، فقد ركّز معظمهم على ثبات الاختبارات وصدقها وسلّط الكثير منهم الضوء على الأداء التفاضلي لما له من اثر على نتائج الدراسات والبحوث، ويعود السّبب في ذلك إلى اختلاف الجنس، أو مكان السّكن، أو العرق، أو الحالة الإقتصاديّة للمفحوصين، بمعنى أنّ بعض الفقرات لا تكون عادلةً بين الأفراد ممن لهم نفس المستوى من القدرة، ووجود مثل هذه الفقرات في الاختبار قد تؤدّي إلى نتائج مضلّلةً بالنّسبة لمعالم الفقرات.

وقد أوضح تان وجيرال (Tan & Jeral, 2005) أنّ أحد الخطوات الرّئيسة للتّحقّق من دقّة تقدير قدرات الأفراد في اختبارٍ معيّنٍ هو تقييم درجة عدالة ذلك الاختبار، عن طريق تحليل الأداء التّفاضليّ لفقرات الاختبار بافتراض تساوي قدرات المفحوصين بغضّ النّظر عن متغيّرات الجنس أو العرق. ويذكر كامب وغروجتر (Kamp & Gruijter, 2005) أنّه اذا لوحظ أنّ الفقرة تبدي أداءً تفاضليًّا فمن المهمّ تحديد أسباب حدوثه، وتحديد فيما إذا كان الأداء التّفاضليّ يتعلّق بالقدرة، فإذا كان سبب الأداء التّفاضليّ بين المجموعات يعزى إلى عوامل ليس لها علاقة بالقدرة فعندها تعتبر الفقرة متحيّزةً، كما أن خلوّ فقرات الاختبار من الأداء التّفاضليّ يعدّر مؤشّرًا إيجابيًّا على نتائج بحث ذات موثوقية عالية وميّزةً مرغوبةً لدى الباحث.

التحيّز والأداء التّفاضليّ: يرى إمبريستون ورايز (Embreston & Reise, 2000) أنّ الفقرة المتحيّزة هي الفقرة التي تعمل في مجموعةٍ بصورةٍ مختلفةٍ ضدّ مجموعةٍ أخرى مأخوذتين من مجتمعٍ واحدٍ. ويرى كاميلي وشيبارد (Camilli & Shepard, 1994) أنّ الفقرة تعتبر متحيّزةً عندما تكون أكثر صعوبةً لمجموعةٍ دون أخرى ممن لهم نفس مستوى القدرة.

* وزارة التربية والتعليم، الأردن. تاريخ استلام البحث 2015/11/22، وتاريخ قبوله 2016/08/03.

كما يرى هولاند ودورنز (Holland & Dorance, 1993) أن التحيّز يحمل المعنى الإجتماعيّ وعدم المساواة بين الأفراد، بينما يدرس الأداء التّفاضليّ خصائص الفقرة السيكومتريّة من حيث كيفيّة عملها من مجموعةٍ لأخرى. حيث يرتبط مفهوم التحيّز بعدم العدالة من النّاحية السياسيّة والاجتماعيّة أكثر من ارتباطه بالجانب السيكومتريّ؛ لذلك استخدم العلماء مفهوم الأداء التّفاضليّ منذ بداية الثمانينات للتّعبير عن تحيّز الفقرات، لأنّ استخدام المصطلحات الفنّيّة السّيكومتريّة أفضل من المصطلحات المشحونة من النّاحية السياسيّة والإجتماعيّة (Ellis & Raju, 2003). فالتحيّز خطأٌ منتظمٌ يبيّن أنّ أداء مجموعةٍ أفضل من أداء مجموعةٍ أخرى، أمّا الأداء التّفاضليّ فهو شرطٌ ضروريٌّ لكنّه غير كافٍ للتّحيّز، فالفقرة قد تظهر أداءً تفاضليًّا، لكنّها لاتكون متحيّزةً إذا كان الفرق يعزى للفروق الفعليّة في القدرة المطلوبة لإجابة الفقرة (عبيدات، 2014). ويعرّف الأداء التّفاضليّ (Differential Item Functioning: DIF) بأنّه دالّةٌ مشتقّةٌ إحصائيًّا للتّعبير عن الفرقِ في احتماليّةِ الإستجابةِ للفقرةِ استجابةً صحيحةً بين مجموعتينِ فرعيّتينِ لمجتمعٍ إحصائيٍّ ما، حيث يكون لهاتين المجموعتين نفس مستوى القدرة (Gierl & Jodion & Ackerman, 2000;Teresi, 2001). وتعتبر طرق الكشف عن الأداء التّفاضليّ التابعة لنظرية الاستجابة للفقرة من الطرق الفاعلة نظريًّا، لكن المأخذ على هذه الطرق في أنّها تدخل الفقرات ذات الأداء التّفاضليّ (التي أخفق الباحث في الكشف عنها) في نتائج البحث (Miller, 1992). ولتجاوز هذا المأخذ كان لا بد من وجود طريقة تفترض وجود فقرات ذات اداء تفاضلي وتتعامل على هذا الاساس، فكانت النماذج الهرمية التي لها عدة مستويات، ابسطها نموذج راش الهرمي.

نموذج راش الهرميّ (Hierarchical Rasch Model): يوفر النموذج الخطي الهرميّ (Hierarchical Linear Model, HLM) نموذجًا إحصائيًا يتضمن نماذج متعددة المستويات في مجموعة البحوث، المستوى الأوّل يمثل مستوى الفرد والمستوى الثّاني يمثل مستوى المجموعة، مع الأخذ بعين الاعتبار اختلاف معاملات الإنحدار في كل مجموعة والمجموعات المتعددة التي يكون فيها عدد الأفراد مختلفًا من مجموعة لأخرى، وكذلك العوامل المختلفة (المتباينة) التي لديها مواصفات متعددة يمكن تمثيلها في النموذج الهرميّ بسهولة (Gokiert & Ricker, 2004).

وعند تطبيق معادلات الإنحدار على بيانات هرميّة فإنه لا بدّ من استقلالية المشاهدات كافتراضٍ أساسيٍّ للتّحليل، لكنّ هذا الافتراض يُنتَهَك بوجود البيانات الهرميّة، ولضمان افتراضات استقلاليّة المشاهدات صمّمت النّماذج الخطّيّة الهرميّة (HLMs) (Randerbush & Bryk, 1986).

ويذكر كيم (Kim, 2003) أن البيانات المجمّعة يمكن أن تتسبّب بانتهاك الإستقلاليّة في عيّنة الدّراسة؛ لذلك فإن استخدام الأساليب الإحصائية التّقليديّة يقود إلى سوء تقدير الأخطاء المعياريّة ونتائج فحص الفرضيّات الصفريّة، حيث إن صيغة النّماذج الهرميّة أكثر مرونة فهي تمكّن الباحث من دراسة الأثر الرئيس للمتغيرات المستقلة على النّتائج، وتتمتع بأنّها الأدقّ تقديرًا للمعالم، وللخطأ المعياريّ، والأدقّ تقديرًا في تحديد فترات الثّقة، وتقديم الدّلالة الإحصائيّة الصّحيحة.

وعند افتراض أن مجموعات الذكور والإناث لهما نفس توزيع القدرة فيتمّ التّعامل مع المجموعتين على أساس أنهما مجموعةٌ واحدةٌ، ولا داعي لاستخدام مستوى للمجموعة في النموذج، وبما أن الأداء التّفاضليّ للجنس يظهر عند مستوى الفرد فإنّه من المناسب استخدام نموذج راش الهرميّ بمستويين (Chu, 2002,Kamata, 1998).

كما أنّ دراسة الأداء التّفاضليّ للفقرة (DIF) بطريقة راش تسمّى بأثر الأداء التّفاضليّ العشوائيّ لأنّها تعالج الأداء التّفاضليّ للفقرة كمعلمةٍ متباينةٍ عشوائيًا بين المجموعات، ومع ذلك فإن النّماذج التقليدية للكشف عن الأداء التّفاضليّ مثل مانتل هانزل والأنحدار اللوجستيّ ليست مناسبةً لمستوى الأداء التّفاضليّ في نموذج راش الهرميّ كمعلمةٍ متباينةٍ عشوائيًّا؛ لأنّها تقتصر على تحديد العلاقات داخل المجموعة (Luppescu, 2002).

ويشير كاماتا (Kamata, 2001) إلى أنّ النموذج الخطّيّ الهرميّ المعمّم (HGLM) المستخدم في صياغة نموذج الاستجابة للفقرة متعدّدة التّدريج لنموذجٍ بمستويين يدعى نموذج راش الهرميّ، ويعبّر عنه باقتران الإرتباط اللوغرتميّ، وينبثق عن نموذج راش الهرميّ مستويين: الأوّل يسمى مستوى الفقرة، والمستوى الثّاني يسمّى بمستوى الفرد ويعبّر عن المستوى الأوّل كما يلي:

$$log\left(\frac{P_{ij}}{1-P_{ij}}\right)=\eta_{ij}$$

$$=\beta_{0j}+\beta_{1j}X_{1ij}+\beta_{2j}X_{2ij}+\cdots+\beta_{(k-1)j}X_{(k-1)ij}$$

$$= \beta_{0j} + \sum_{q=1}^{k-1} \beta_{qj} X_{qij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

علمًا بأنّ X_{qij} المتغيّر الوهميّ رقم **q** للشخص j والفقرة i وتكون قيمته -1 عندما تكون $q=i$، وصفر عندما تكون $q\neq I$، والمعامل $\beta 0j$ هو المقطع الصادي، وβqij هي المعاملات المرتبطة مع X_{qij}، حيث $q=(1,2,3,\ldots\ldots, k-1)$ علمًا بأن عدد المتغيّرات الوهمية يقل عن عدد الفقرات k بمقدار 1 وذلك لأنّه لا داعي لتعريف الفقرة الأخيرة في عمليّة التّرميز من خلال المتغيّرات الوهميّة، ولا يفترض هذا المستوى أن القيم (من β_{ij} إلى $\beta_{(k-1)j}$) متساوية أو ثابتة بين الأفراد، بل يعدّرها أثرًا مشتركًا لجميع الفقرات، ويمكن لهذا المستوى أن يصف تأثيرات الفقرة بأنّها ثابتة بين الأفراد.

ويمكن اختصار المعادلة (2) كالتالي:

$$log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \eta_{ij} = \beta_{0j} - \beta_{qj} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

وذلك للفقرة i المرتبطة بالمتغيّر الوهمي q، ويأخذ القيمة -1 في حال $q=i$، ويأخذ القيمة 0 في حال $q \neq i$، وكذلك فإن β_{qi} تمثّل أثر المتغيّر الوهميّ q، وأثر الفقرة i عندما $q=i$، ويمكن كتابة المعادلة رقم 1 بصيغة المصفوفة كالتالي:

$$\eta_{ij} = [d_j \ \vdots \ X_j]\begin{bmatrix}\beta_{0j} \\ \vdots \\ \beta_j^*\end{bmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

حيث: و

$$W_j = [d_j \quad \vdots \quad X_j] \quad \beta_j = [\beta_{0j} \quad \beta_j^*] \quad = W_j\, \beta_j$$

ويشير هنا الرمز dj إلى مصفوفة عمودية من الرتبة $k\times 1$ التي تكون كافة عناصرها مساوية 1، ثابت مصفوفه عمودية من الرتبة $(k-1)\times 1$، تتضمن المعاملات من β_{oj} إلى β_j^* ، و β_{ij} X_j $\beta_{(k-1)j}$ هي مصفوفة من الرتبة $k \times (k-1)$ حيث تكون العناصر القطرية للصفوف من الأوّل وحتى(k-1) هي−1 وتكون بقية العناصر اصفارًا. وبالنتيجة تكون جميع عناصر الصف k في المصفوفة Xj اصفارًا، لأنّه تم افتراض أن $\beta_{kj} = 0$ وفي هذه الحالة تعد قيمة β_{oj} الاثر الكلي المشترك لكل الفقرات، وتعد قيمة β_{qj} الاثر الخاص للمتغير الوهمي q.

درس شو (Chu, 2002) الأداء التّفاضليّ مستخدمًا نموذج راش الهرميّ، حيث إنه في مستوى الأداء التّفاضليّ للشّخص افترض أن يكون الجنس بدلًا من الفرد في دراسته حيث تظهر كما يأتي:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(Gender)$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}(Gender) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$
$$\beta_{3j} = \gamma_{30}$$
$$.$$
$$.$$
$$\beta_{(k-1)_j} = \gamma_{(30)}$$

حيث γ_{00} إلى $\gamma_{(k-1)_0}$ المقطع ل β_{0j} إلى $\beta_{(k-1)j}$ وتمثّل γ_{11} و γ_{21}معاملات الجنس لكل من الفقرة 1 والفقرة 2 وتمثّل μ_{0j} أثر عامل عشوائي، من 3 إلى k-1 فقرات بدون أداء تفاضلي لكنها تحتوي معالم متقاطعة (γ_{i0}) لأنّ أثر الفقرة مع الأشخاص

ثابت كذلك بين المجموعات الفرعية، أضيف متغير الأداء التّفاضليّ ومتغير الجنس إلى الفقرات 1 و2 لضبط أثرها.

ويعدّر الجنس *(gender)* متغيرًا وهميًا يعطي القيمة (1) للإناث والقيمة (0) للذكور. ووفقًا لنماذج المستوى الأوّل والثّاني المشتركة (أي عند دمج المستويين معًا) ستكون المعادلة على النحو التالي:

$$\eta_{1j} = \gamma_{00} + \gamma_{01}(gender)_j + \mu_{0j} - [\gamma_{q0} + \gamma_{q1}(gender)_j]$$

$$= \mu_{0j} + \gamma_{00} - \gamma_{q0} - (\gamma_{q1} - \gamma_{01})(gender)_j \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots 5$$

$$= \mu_{0j} - [\gamma_{q0} - \gamma_{00} + (\gamma_{q1} - \gamma_{01})(gender)_j]$$

حيث إن:

i=q

q=(1,2,……..,k-1)

i=(1,2,………..,k)

يوضّح النموذج المشترك بأنّ: $\gamma_{q0} - \gamma_{00} + (\gamma_{q1} - \gamma_{01})(gender)_j$

تمثّل صعوبة الفقرة المرتبطة بالمتغيّر الوهميّ q لمجموعة الجنس j بينما: $-\gamma_{00} - \gamma_{01}(gender)_j$

تمثّل صعوبة الفقرة المرجعيّة.

حيث إن: $(\gamma_{q1} - \gamma_{01})$

: تمثّل الجنس على الفقرة المرتبطة مع المتغيّر الوهميّ q.

1,2,……….,(k-1) =q

γ_{01} = أثر الجنس على المجموعة المرجعيّة. $\gamma_{q0} - \gamma_{00} + (\gamma_{q1} - \gamma_{01})$

بمعنىً آخر فإنّ المعادلة: تمثّل صعوبة الفقرات لدى الأنّاث

وكذلك فإن المعادلة: $\gamma_{q0} - \gamma_{00}$ تمثّل صعوبة الفقرات لدى الذكور.

وعند دمج نموذج المستوى الأوّل ونموذج المستوى الثّاني معًا يتّضح أن معالم الفقرة ثابتة بين الأفراد ومتباينة بين الفقرات لأنّه لم يضف عناصر عشوائية إلى (β_{1j} وحتى $\beta_{(k-1)j}$) بينما تكون السّمات الكامنة متباينة بين الأفراد، وثابتة بين الفقرات، كما وأنه عند دمج نماذج المستويين فإن النموذج المتنبئ الخطّي في المعادلة (3) يصبح على النحو التالي:

$$\eta_{ij} = \gamma_{00} + \mu_{0j} - \gamma_{q0}$$

للشخص j والفقرة i المرتبطة بالمتغيير الوهمي *q*

هنا فإن احتمالية إجابة الفرد *j* على الفقرة *i* إجابة صحيحة هي:

$$P_{ij} = \frac{1}{1 + \exp\{-[\mu_{0j} - (\gamma_{q0} - \gamma_{00})]\}} \dots\dots\dots\dots\dots\dots\dots\dots 6$$

وهذا يمثّل التكافؤ الجبري لنموذج راش في المعادلة رقم 1

حيث إن:

$$i = q$$

$$\theta_j = \mu_{0j}$$

$$b_1 = \gamma_{q0} - \gamma_{00}$$

مع ملاحظة أن b_1 و $\gamma_{q0} - \gamma_{00}$ معالم ثابتة للفقرة، وأنّ θ_j معلمة ثابتة عند راش، و μ_{0j} متغيّر عشوائي في النموذج

الخطي الهرميّ المعمّم.

هذه مؤشراتٌ بأن الفقرة يمكن أن تتحيّز ضدّ واحدة من المجموعات (الذكور أو الأناث)، حيث إن الفروق لا تدلّ دائما على وجود الأداء التّفاضليّ في الفقرة؛ لأنّ اختلاف الأداء بين المجموعات قد يحدث بسبب تباينات حقيقية تعود إلى نوع الجنس، لكن اختلاف الجنس يعدّر مؤشرًا على وجود أداء تفاضلي لكنه لا يعدّر دليلًا قاطعًا على وجود الأداء التّفاضليّ.

مشكلة الدّراسة وأسئلتها

يعدّ الأداء التّفاضليّ للفقرة من أهم محددات الاختبارات، حيث إن وجود الأداء التّفاضليّ يهدد صدق البناء والثبات وتقدير المعالم للقدرة والفقرة وبالتالي تلوّث النّتائج، وبذلك يصعب اتخاذ قرارات تربوية في ضوء بيانات غير صادقّة. وان ما يستحق الإشارة اليه أن وجود الأداء التّفاضليّ على الفقرات أو عدم وجوده عبارةٌ عن مسألة نسبيّة، وأنّ حجم التأثير السلبي لوجوده يعتمد على نسبة الفقرات ذات الأداء التّفاضليّ ومستوى ذلك الأداء.

وقد جرت العادة أن يتمّ الكشف عن فقرات الأداء التّفاضليّ، ومن ثم حذف هذه الفقرات التي تبدي أداءً تفاضليًا. لكنّ عمليّة الحذف تؤثر سلبًا على صدق الاختبار خاصةً إذا كان اختبارًا قصيرًا؛ لذلك وللمحافظة على بناء الاختبار وتفادي أيّة مشاكل تؤثّر في صدق البناء وعدم دقّة التّقدير لمعلمة القدرة وزيادة التّكاليف الماديّة لتطوير الاختبارات الخالية من الأداء التّفاضليّ، فقد تمّ البحث عن طريقةٍ لتجاوز هذه السّلبيّات عن طريق تضمين معلمة الأداء التّفاضليّ في النّموذج المستخدم من خلال استخدام نموذج راش الهرميّ (Kamata, 1998).

حيث استخدم الباحث هذا النموذج لتقدير معلمة الصّعوبة بوجود الأداء التّفاضليّ وفق هذا النموذج في حالات مختلفة من مستويات الأداء التّفاضليّ ونسبة فقراته إلى مجمل الفقرات الكلي. وتسعى هذه الدّراسة لبحث أثر كل من مستوى الأداء التّفاضليّ في الفقرات ونسبة الفقرات ذات الأداء التفاضلي في دقّة تقدير معلمة الصّعوبة وفق نموذج راش الهرميّ، حيث يعدّر الأداء التّفاضليّ أحد معالم هذا النموذج، وتتلخص مشكلة الدّراسة في الاجابة عن السؤال التالي:

"**هل توجد فروق ذات دلالة إحصائية عند مستوى الدّلالة (α=0.05) في دقّة تقدير معلمة صعوبة الفقرة تعزى لمتغيّريّ الدّراسة (نسبة الفقرات ذات الأداء التّفاضليّ ومستوى الأداء التّفاضليّ) والتفاعل فيما بينهما؟**"

أهمّيّة الدّراسة

تأتي أهمية الدّراسة بسبب الاستخدام الشائع للاختبارات بشتّى أنواعها، ولأنّ وجود الأداء التّفاضليّ في الفقرة يعدّ قضيّةً مهمّةً؛ كان لا بدّ من دراسة أثره في تقدير معلمة صعوبة الفقرات، ومن المهمّ كذلك تحديد نسبة الفقرات التي تبدي أداءً تفاضليًا في الاختبار، لأنّ وجود الفقرات التي تبدي أداءً تفاضليًا قد تعطي تقديراتٍ غير دقيقةٍ حول الخصائص السيكومتريّة للفقرات.

كما ان اهمية الدراسة تبرز بسبب قلّة الدراسات التي تناولت أثر نسبة الفقرات ذات الأداء التّفاضليّ في الاختبار، ومستوى الأداء التّفاضليّ في دقّة التقدير لمعالم الفقرة اعتمادًا على نموذج راش الهرميّ، حيث إن أغلب الدراسات السابقة لم تعتمد النماذج الهرمية لنظريّة الاستجابة للفقرة، اذ أنّ النّماذج الهرميّة تبقي على الأداء التّفاضليّ موجودًا وتعمل على دراسة أثره في دقّة تقدير المعالم.

محدّدات الدّراسة

اقتصرت الدّراسة على طول اختبار ثابت يتكون من (60) فقرة، واقتصرت الدّراسة على حجم عينة ثابتة تتكون من (1200) فرد، كما انها فحصت الدّراسة دقّة معلمة الصّعوبة فقط من معالم الفقرة، واقتصرت الدّراسة على استخدام نموذج راش الهرميّ، كما انه تم استخدام ثلاث قيم فقط لمستويات الأداء التّفاضليّ (0.5، 1.0، 1.5)، واستخدام ثلاث نسب لفقرات الأداء التّفاضليّ (10%، 25%، 50%).

مصطلحات الدّراسة

مستوى الأداء التّفاضليّ: تعرّف بأنّها مقدار الزّيادة التي تطرأ على معلمة صعوبة الفقرة المصمّمة لِأَنْ تبدي أداءً تفاضليًا، وفقًا لظروف البحث الذي تنتمي إليه الفقرة.

دقّة تقدير معلمة صعوبة الفقرة: تقاس باستخدام الفرق المطلق لمعلمتي الصّعوبة المقدّرة والحقيقيّة، وتقاس كذلك بالخطأ المعياريّ لمعلمتي الصّعوبة المقدّرة والحقيقيّة.

الأداء التفاضلي: هو مؤشر احصائي للتعبير عن الفروق في الاستجابة على الفقرة بين مجموعتين من المفحوصين ممن لهم نفس المستوى من القدرة (Camilly & Shepard,1994)

اما بالنسبة للدراسات السابقة فلقد حظي موضوع الأداء التّفاضليّ باهتمام الكثير من الباحثين، فقد قام انجل وجيلدون وجوز (Angel, Gildon & Jose, 2000) بدراسةٍ هدفت للكشف عن أثر نسبة الفقرات ذات الأداء التّفاضليّ في الاختبار، وأثر طول الاختبار على قوّة وفاعليّة أسلوب مانتل هانزل في الكشف عن الفقرات ذات الأداء التّفاضليّ، حيث تمّ توليد بيانات الإستجابة للفقرات وفق النّموذج اللوجستيّ ثلاثيّ المعلمة تحت ظروفٍ مختلفةٍ، وقد تمّ اعتمادُ النّسب التالية للفقرات التي تبدي أداءً تفاضليًّا (10%, 20%, 30%) من عدد الفقرات الكلي للاختبار، واعتمدت أطوال اختبارٍ متفاوتةٍ (20, 40, 60)، وأظهرت النّتائج أنّ طريقة مانتل هانزل(MH) تكون فعّالةً وذاتُ قوّةٍ إحصائيّةٍ تكفي للكشف عن الأداء التّفاضليّ للفقرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ تساوي (10%)، بينما تنخفض فاعليّتها للنّسب (20%, 30%)، وتشير النّتائج أنّ طول الاختبار كان له أثرًا ضعيفًا على القوّة الإحصائيّة لمؤشّر مانتل هانزل.

وقام شو (Chu, 2002) بإجراء دراسةٍ هدفت إلى البحث في تقدير معالم نموذج راش الهرميّ ومعادلة الاختبارات، وتمّ استخدامُ بياناتٍ مولّدةٍ لاختبارين طولُ كلٍّ منهما (20) فقرةٍ، وقد تمّ تقييم نموذج راش الهرميّ بإضافة معلمة الأداء التّفاضليّ إلى معلمة الصّعوبة في النّموذج، ثمّ تمّ إجراء معادلة الاختبار اعتمادًا على كلٍّ من التحيّز والخطأ المعياريّ والجذر التّربيعيّ لمتوسّط مربعات الخطأ، وأشارت النّتائج أنّ نموذج راشٍ الهرميّ كان الأكثرُ دقّةً في تقدير صعوبة الفقرة وقدرة الفرد، مقارنة بالنماذج الاخرى التي اعتمدت لمعادلة الاختبار.

وأجرى لوبسكو (Luppescu, 2002) دراسةً هدفت إلى المقارنة بين النّماذج الخطيّة الهرميّة وبين الطّرق التقليديّة في القدرة على كشف الأداء التّفاضليّ للفقرة، استخدم فيها بياناتٍ مولّدة لـ (180) مجموعةٍ من البيانات تتوزّع على (5) مجموعاتٍ لكلٍّ مجموعةٍ (36) حالة (3*4*3)، 3 احجامٍ للعيّنات (100, 250, 500) *4 درجاتٍ للأداء التّفاضليّ (0.25, 0.50, 0.75, 1) *3 نسبٍ للفقرات التي تبدي أداءً تفاضليًا في المجموعات المستهدفة (10%, 25%, 50%). أشارت النّتائج إلى أنّ النّماذج الخطيّة الهرميّة تتميّز بقدرتها على ضبط قدرة الفرد بوجود أداءٍ تفاضليٍّ للفقرة، وتساوي قيم الجذر التربيعيّ لمتوسّط مربّعات الأخطاء لمعظم الحالات، إلّا عندما تكون أحجام العيّنات صغيرةً، وكانت قيم الجذر التربيعيّ لمتوسّط مربّعات الأخطاء أكبرُ في النّماذج الخطّيّة الهرميّة بالنّسبة للمجموعة المستهدفة، أمّا عندما يكون حجم العيّنة كبيرًا فإنّها تقلُّ درجة الأداء التّفاضليّ، وكذلك أشارت النّتائج أن النّماذج الخطيّةِ الهرميّةِ تعطي تقديراتٍ أعلى من الطّرق التقليديّةِ.

وقام شو وكاماتا (Chu&Kamata, 2005) بإجراء دراسةٍ بعنوان معادلة درجات الاختبار بوجود الأداء التّفاضليّ، كان الهدف منها قياس دقّة وثبات تقدير القدرة وصعوبة الفقرات وفق نموذج راش الهرميّ، حيث كانت بيانات الدّراسة مولّدةٌ لمجموعتين في كلّ مجموعةٍ (500) فردٍ، وقد ركّزت إجراءات الدّراسة على إجراء معادلة الدّرجات لكلا المجموعتين، وتمّ استخدامُ نموذجًا للمجموعة الأوّلى حيث استُبدلت الفقرات ذات الأداء التّفاضليّ بفقراتٍ لا تظهرُ أيَّ أداءٍ تفاضليٍّ، أمّا للمجموعة الثّانية فقد تمّ تثبيتُ الأداء التّفاضليّ للفقراتِ وإجراءُ تقديرِ المعالمِ، حيث أظهرتِ النّتائجُ أنّ نموذجَ راشٍ الهرميّ كان أفضلُ من النّموذجين الآخَرَينِ في تقدير القدرة ومعالم الفقرات، وكذلك كان الجذرُ التّربيعيُّ لمتوسَّطِ مربَّعات الأخطاءِ لنموذج راش الهرميّ أقلُّ من النّموذجين الآخَرَين.ومن خلال اطلاع الباحث على الدراسات السابقة ذات العلاقة بموضوع البحث، فقد وجد أنّ بعض هذه الدراسات تهدف الى المقارنة بين النماذج الهرميّة ونماذج أخرى في القدرة على الكشف عن الأداء التفاضلي كدراسة لوبسكو (2002). وغيرها ما بحث في أثر وجود الأداء التفاضلي في دقة تقدير المعالم وأثره في دقّة معادلة الاختبارات مثل دراسة شو (2002)، ومنها دراسات ركزت على اثر وجود الفقرات ذات الأداء التفاضلي في قوّة اساليب الكشف عن الأداء التفاضلي مثل دراسة انجل وجيلدون وجوز (2000)، او اثر الأداء التفاضلي في احصائيات النظرية الكلاسيكية كدراسة زومبو (2000)، كما أنّ من هذه الدراسات ما بحثت في قياس دقّة وثبات تقدير المعالم وفق النماذج الهرمية مثل دراسة شو وكاماتا (2005). حيث استفاد الباحث من الدراسات السابقة في تكوين تصوّر شامل حول الاطر النظرية وتحديد منهجية الدراسة والاجراءات اللازمة في سبيل تحقيق اهدافها، واعتبرت هذه الدراسات مصدرًا مهمًّا في توجيه الباحث نحو العديد من الدراسات والمراجع المناسبة لخدمة الهدف من هذه الدراسة. أمّا الدراسة الحالية فقد جاءت امتدادًا لهذه الدراسات السابقة ولتسدّ النقص فيها، وقد تميّزت عن غيرها من حيث

غرضها، إذ أنها بحثت في مدى الفاعليّة والقوّة التي يتمتّع بها نموذج راش الهرمي في تقدير معلمة صعوبة الفقرات في ضوء نسب متفاوتة للفقرات التي تبدي أداءً تفاضليًا ومستوى ذلك الأداء التي تبديه هذه الفقرات.

الطريقة والاجراءات

تتناول طريقة الدراسة واجراءاتها وصفًا للبيانات التي تمّ توليدها باستخدام برنامج WinGen v3، وذلك بهدف الإجابة عن سؤال الدّراسة المتعلق بمعلمة الصّعوبة المقدّرة (b) باستخدام طريقة النّماذج الخطية الهرميّة (HLM: Hierarchical Linear Models) المضمنة في برنامج (HLM v7.01) وفقًا لمتغيّريّ الدّراسة المُعتمدين في هذه الدّراسة (نسبة الفقرات ذات الأداء التّفاضليّ ومستوى الأداء التّفاضليّ).

توليد البيانات (Simulating The data): استخدمت في هذه الدّراسة طريقة المحاكاة لتوليد البيانات اللازمة لتنفيذ الدّراسة والمسماة بطريقة مونت كارلو Monte Carlo Methods (MCM)، وذلك باستخدام برنامج (WinGen V3) لتوليد البيانات وفقًا لمستويات متغيّريّ الدّراسة التي تشكّل بدورها تسعة مواقف بحثيّة، بالإضافة إلى المجموعة المرجعية (Reference Group) من أجل المقارنة بين الوسط الحسابيّ للفرق المطلق لمعلمة الصّعوبة، بما يتناسب مع سؤال الدّراسة، وذلك بهدف إظهار مدى كفاءة تقدير معلمة الصّعوبة وفقًا لمتغيّريّ الدّراسة باستخدام طريقة (HLM) المُضمّنة في برنامج (HLM v7.01).

وقد تمت مراعاة توليد القدرة الحقيقيّة بالاعتماد على التّوزيع الطبيعي (Normal Destribution) بوسط حسابيّ مقداره (0) وانحراف معياريّ مقداره (1) حيث تبقى قيم القدرة الحقيقيّة المُولّدة لأي استجابة في أي ظرف من الظروف العشرة المشمولة بالدّراسة كما هي في بقية الظروف الأخرى نتاج تفاعل متغيّريّ الدّراسة، كما تمت مراعاة توليد معلمة الصّعوبة الحقيقيّة بالاعتماد على التّوزيع الطبيعي بوسط حسابيّ مقداره (0) وانحراف معياريّ مقداره (1) حيث تبقى قيم معلمة الصّعوبة الحقيقيّة المُولّدة لأي استجابة في أي ظرف من الظروف العشرة المشمولة بالدّراسة كما هي في بقية الظروف الأخرى نتاج تفاعل متغيّريّ الدّراسة، باستثناء الفقرات التي صممت لأنّ تبدي أداءً تفاضليًا، وذلك كما هو مبيّن في الجدول (1).

الجدول (1)

الفقرات ذات الأداء التّفاضليّ وفقًا لمتغيّريّ الدّراسة

مستوى الأداء التّفاضليّ								
0.5			1.0			1.5		
نسبة الفقرات			نسبة الفقرات			نسبة الفقرات		
10%	25%	50%	10%	25%	50%	10%	25%	50%
1	1	1	1	1	1	1	1	1
	2	2		2	2		2	2
		3			3			3
		4			4			4
		5			5			5
		6			6			6
		7			7			7
11	11	11	11	11	11	11	11	11
	12	12		12	12		12	12
		13			13			13
		14			14			14
21	21	21	21	21	21	21	21	21
	22	22		22	22		22	22
		23			23			23
		24			24			24

مستوى الأداء التّفاضليّ								
0.5			1.0			1.5		
نسبة الفقرات			نسبة الفقرات			نسبة الفقرات		
31	31	31	31	31	31	31	31	31
	32	32		32	32		32	32
	33	33		33	33		33	33
		34			34			34
41	41	41	41	41	41	41	41	41
	42	42		42	42		42	42
		43			43			43
		44			44			44
51	51	51	51	51	51	51	51	51
	52	52		52	52		52	52
		53			53			53
		54			54			54
		57			57			57
	58	58		58	58		58	58
	59	59		59	59		59	59

والجدول (2) يبين الأوساط الحسابيّة والإنحرافات المعياريّة لمعلمة الصّعوبة الحقيقيّة وفقًا لمتغيّريّ الدّراسة.

الجدول (2)

الأوساط الحسابيّة والإنحرافات المعياريّة لمعلمة الصّعوبة الحقيقيّة وفقًا لمتغيّريّ الدّراسة

نسبة الفقرات ذات الأداء التّفاضليّ	مستوى الأداء التّفاضليّ	المتوسط الحسابيّ	الإنحراف المعياريّ
0	0	-0.1256	0.8874
10%	0.5	-0.0756	0.8799
	1	-0.0256	0.8982
	1.5	0.0244	0.9407
25%	0.5	-0.0006	0.9144
	1	0.1244	0.9899
	1.5	0.2494	1.1042
50%	0.5	0.1244	0.8863
	1	0.3744	0.9542
	1.5	0.6244	1.0783

الكشف عن أحادية البعد باستخدام برنامج NOHARM: تمّ التحقق من افتراض أحادية البعد باستخدام برنامج NOHARM (Normal Ogive Harmonic Analysis Robust Method) لمعالجة البيانات المُولَّدة المتعلقة بالاستجابات البالغة (1200) استجابة على اختبار مؤلف من (60) فقرة وفقًا لمتغيّريّ الدّراسة، الذي يكشف عن أحادية البعد باستخدام مؤشرين متزامنين معًا؛ هما: مؤشر (TANAKA) ومؤشر (RMSR)، والجدول (3) يبين القيم الخاصّة بكلٍّ منهما وفقًا لمتغيّريّ الدّراسة.

الجدول (3)

قيم مؤشري أحادية البعد (TANAKA) و(RMSR) المتزامنين معًا وفقًا لمتغيّريّ الدّراسة

نسبة الفقرات ذات الأداء التّفاضليّ	مستوى الأداء التّفاضليّ	مؤشرات أحادية البعد	قيمة المؤشر
0	0	مؤشر RMSR	0.005
		مؤشر Tanaka لحسن المطابقة	0.977
10%	0.5	مؤشر RMSR	0.005
		مؤشر Tanaka لحسن المطابقة	0.975
	1	مؤشر RMSR	0.005
		مؤشر Tanaka لحسن المطابقة	0.977
	1.5	مؤشر RMSR	0.005
		مؤشر Tanaka لحسن المطابقة	0.977
25%	0.5	مؤشر RMSR	0.004
		مؤشر Tanaka لحسن المطابقة	0.970
	1	مؤشر RMSR	0.004
		مؤشر Tanaka لحسن المطابقة	0.973
	1.5	مؤشر RMSR	0.005
		مؤشر Tanaka لحسن المطابقة	0.969
50%	0.5	مؤشر RMSR	0.004
		مؤشر Tanaka لحسن المطابقة	0.970
	1	مؤشر RMSR	0.004
		مؤشر Tanaka لحسن المطابقة	0.968
	1.5	مؤشر RMSR	0.004
		مؤشر Tanaka لحسن المطابقة	0.971

يلاحظ من الجدول (3) أن جميع قيم مؤشر TANAKA قد تخطت قيمة الـ 0.95؛ بشكل يتزامن مع انخفاض جميع قيم مؤشر RMSR دون القيمة الحرجة له البالغة قيمتها (0.1183) التي تحسب عن طريق تطبيق المعادلة $(4.1/\sqrt{1200})$ في مختلف الظروف المشمولة بالدّراسة وفقًا لمتغيّريّ الدّراسة؛ وتعد هذه النتيجة محققة لافتراض أحادية البعد وفقًا لطريقة برنامج NOHARM في الكشف عن أحادية البعد؛ حيث يشترط أن تزيد قيمة مؤشر TANAKA عن 0.80 بشكل يتزامن مع انخفاض قيمة مؤشر RMSR باتجاه الصفر دون زيادةٍ عن القيمة الحرجة له في كل ظرف من الظروف المشمولة بالدّراسة وفقًا لمتغيّريّ الدّراسة (Jasper, 2010).

متغيّرات الدّراسة: اشتملت الدّراسة على المتغيّرات الآتية:

أ. المتغيّرات المستقلة، وهي:

1. نسبة الفقرات ذات الأداء التّفاضليّ؛ ولها ثلاث نسب (10%، 25%، 50%).
2. مستوى الأداء التّفاضليّ؛ ولها ثلاث قيم (0.5، 1.0، 1.5).

ب. المتغيّرات التابعة، وهي معلمة صعوبة فقرات الاختبار المقدّرة، والفرق المطلق لمعلمتي الصّعوبة المقدّرة والحقيقيّة، والخطأ المعياريّ لمعلمة صعوبة فقرات الاختبار.

المعالجات الاحصائية: في البداية تم توليد البيانات باستخدام برنامج (WinGen-3) وفقًا لمتغيّريّ الدّراسة، ثمّ تم استدعاء هذه البيانات لإجراء التحليلات باستخدام برمجيات (HLM v7.1 and SPSS v22)، ثم إجراء المعالجات الإحصائية للإجابة عن سؤال الدّراسة وفق الخطوات الآتية:

1- حساب الأوساط الحسابيّة والإنحرافات المعياريّة لمعلمتي صعوبة فقرات الاختبار المقدّرة والحقيقيّة، والأوساط الحسابيّة والإنحرافات المعياريّة للفرق المطلق لمعلمتي الصّعوبة المقدّرة والحقيقيّة وفقًا لمتغيّريّ الدّراسة.

2- إجراء تحليل التّباين الثّنائيّ للأوساط الحسابيّة لمعلمة صعوبة فقرات الاختبار المقدّرة وللفرق المطلق في معلمتي الصّعوبة المقدّرة والحقيقيّة الهرميّة وفقًا لمتغيّريّ الدّراسة.

3- حساب الأوساط الحسابيّة والإنحرافات المعياريّة للخطأ المعياريّ لمعلمة صعوبة فقرات الاختبار المقدّرة وفقًا لمتغيّريّ الدّراسة.

4- إجراء تحليل التّباين الثّنائيّ للخطأ المعياريّ لمعلمة صعوبة فقرات الاختبار المقدّرة وفقًا لمتغيّريّ الدّراسة.

عرض النتائج: لعرض نتائج الدراسة فقد تم حساب الوسط الحسابيّ والإنحراف المعياريّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة، وذلك كما يظهرها الجدول (4).

الجدول (4)

الوسط الحسابيّ والإنحراف المعياريّ لمعلمة الصّعوبة المقدّرة للفقرة تبعًا لمتغيّريّ الدّراسة

نسبة الفقرات ذات الأداء التّفاضليّ	مستوى الأداء التّفاضليّ	المتوسط الحسابيّ	الإنحراف المعياريّ
0%	0	-0.1090	0.7528
10%	0.5	-0.0652	0.7339
	1	-0.0236	0.7512
	1.5	0.0104	0.7874
	الكلي	-0.0262	0.7542
25%	0.5	1.1513	1.6752
	1	1.2379	1.8022
	1.5	0.2002	0.9209
	الكلي	0.8631	1.5803
50%	0.5	0.5423	1.1531
	1	0.7981	1.3418
	1.5	0.9936	1.4284
	الكلي	0.7780	1.3185
الكلي	0	-0.1090	0.7528
	0.5	0.5428	1.3374
	1	0.6708	1.4577
	1.5	0.4014	1.1570

يلاحظ من الجدول (4) وجود فروق ظاهرية في الوسط الحسابيّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة، وللتحقق من جوهرية الفروق الظاهرية، فقد تم إجراء تحليل التّباين الثّنائيّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة، كما يظهرها الجدول (5).

الجدول (5)

نتائج تحليل التّباين الثّنائيّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة

مصدر التّباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدّلالة الإحصائية
نسبة الفقرات ذات الأداء التّفاضليّ	86.688	2	43.344	30.179	0.000
مستوى الأداء التّفاضليّ	6.537	2	3.269	2.276	0.104
نسبة الفقرات ذات الأداء التّفاضليّ×مستوى الأداء التّفاضليّ	39.562	4	9.890	6.886	0.000
الخطأ	847.365	590	1.436		
الكلي	1002.780	599			

يتبين من الجدول وجود فروق ذات دلالة إحصائية عند مستوى الدّلالة ($\alpha=0.05$) في الوسط الحسابيّ لمعلمة صعوبة الفقرة المقدّرة تُعزى لنسبة الفقرات ذات الأداء التفاضلي، وتُعزى كذلك الى التفاعل بين متغيّريّ الدّراسة؛ ولكون المتغيّريّن متعددي المستويات، فقد تم إجراء اختبار (Levene) للكشف عن انتهاك تجانس التّباين في الوسط الحسابيّ لجميع قيم معلمة الصّعوبة المقدّرة تبعًا لمتغيّريّ الدّراسة، حيث دلّت نتائج التحليل على انتهاك تجانس التّباين؛ لذلك تمّ استخدام أحد الاختبارات البعديّة التي تراعي التّباين غير المتجانس، وقد تم استخدام اختبار (Games-Howell) للمقارنات البعديّة المتعددة؛ بهدف تحديد لصالح إيٍّ من نسب الفقرات ذات الأداء التّفاضليّ قد كانت الفروق الجوهرية في الوسط الحسابيّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغير

الدّراسة، وذلك كما هو مبيّن في الجدول (6).

الجدول (6)

نتائج اختبار Games-Howell للمقارنات البعديّة المتعددة للوسط الحسابيّ

لمعلمة صعوبة الفقرة المقدّرة تبعًا لمستويات المتغيّر (نسبة الفقرات ذات الأداء التّفاضليّ)

نسبة الفقرات ذات الأداء التّفاضليّ		0%	10%	50%	25%
Games-Howell	المتوسط الحسابيّ	-0.109	-0.026	0.778	0.863
0%	-0.109				
10%	-0.026	0.083			
50%	0.778	**0.887**	**0.804**		
25%	0.863	**0.972**	**0.889**	0.085	

يتّضح من الجدول (6) وجود تقارب غير دال إحصائيًا عند مستوى الدّلالة (α=0.05) للوسط الحسابيّ لمعلمة صعوبة الفقرة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (10%) و(0%)، وعندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) و(50%). كما يتضح من الجدول (6) أن أكبر الفروق ذات الدّلالة الإحصائية عند مستوى الدّلالة (α=0.05) للوسط الحسابيّ لمعلمة صعوبة الفقرة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) و(0%).

وفي ضوء النّتائج التي تظهر في الجدول (5) تبيّن وجود فروق ذات دلالة إحصائية عند مستوى الدّلالة (α=0.05) في الوسط الحسابيّ لمعلمة صعوبة الفقرة المقدّرة تُعزى لتفاعل متغيّريّ الدّراسة. كما تم حساب الوسط الحسابيّ والإنحراف المعياريّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّريّ الدّراسة، كما يظهرها الجدول (7).

الجدول (7)

الوسط الحسابيّ والإنحراف المعياريّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّريّ الدّراسة

نسبة الفقرات ذات الأداء التّفاضليّ	**مستوى الأداء التّفاضليّ**	**المتوسط الحسابيّ**	**الإنحراف المعياريّ**
0%	0	0.1165	0.0906
10%	0.5	0.1244	0.0941
	1	0.1268	0.0918
	1.5	0.1290	0.0997
	الكلي	0.1267	0.0947
25%	0.5	1.1924	1.3510
	1	1.1765	1.2970
	1.5	0.1497	0.1312
	الكلي	0.8395	1.1837
50%	0.5	0.5185	0.7995
	1	0.5473	0.8181
	1.5	0.5051	0.7608
	الكلي	0.5236	0.7889
الكلي	0	0.1165	0.0906
	0.5	0.6118	1.0054
	1	0.6169	0.9823
	1.5	0.2613	0.4793

يلاحظ من الجدول (7) وجود فروق ظاهرية في الوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا

لمتغيّريّ الدّراسة؛ وللتحقق من جوهرية الفروق الظاهرية، تم إجراء تحليل التّباين الثّنائيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّريّ الدّراسة، وذلك كما يظهرها الجدول (8).

الجدول (8)

نتائج تحليل التّباين الثّنائيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّريّ الدّراسة

مصدر التّباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدّلالة الإحصائية
نسبة الفقرات ذات الأداء التّفاضليّ	45.925	2	22.962	**42.154***	0.001
مستوى الأداء التّفاضليّ	14.960	2	7.480	**13.731***	0.001
نسبة الفقرات ذات الأداء التّفاضليّ×مستوى الأداء التّفاضليّ	27.933	4	6.983	**12.820***	0.001
الخطأ	321.392	590	0.545		
الكلي	418.013	599			

يتبين من الجدول (8) وجود فروق ذات دلالة إحصائية عند مستوى الدّلالة (α=0.05) في الوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تعزى لمتغيّريّ الدّراسة والتفاعل بينهما؛ ولكون المتغيّريّن متعدّدي المستويات، فقد تم إجراء اختبار Levene للكشف عن انتهاك افتراض تجانس التّباين في الوسط الحسابيّ لجميع قيم الفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّريّ الدّراسة، حيث بلغت قيمة F المحسوبة لاختبار Levene ما مقداره 55.017 عند درجتي حرية (9 للبسط، و590 للمقام) بدلالة إحصائية عند مستوى الدّلالة (α=0.05)؛ مما يدلّ على انتهاك افتراض تجانس التّباين، وبناءً عليه فقد تمّ استخدام أحد الاختبارات البعديّة التي تراعي التّباين غير المتجانس وهو اختبار (Games-Howell) للمقارنات البعديّة المتعدّدة؛ بهدف تحديد لصالح إيٍّ من نسب الفقرات ذات الأداء التّفاضليّ أو مستوى الأداء التّفاضليّ قد كانت الفروق الجوهرية في الوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّريّ الدّراسة، وذلك كما هو مبيّن في الجدولين (9 و10).

الجدول (9)

نتائج اختبار Games-Howell للمقارنات البعديّة المتعددة للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمستويات المتغيّر (نسبة الفقرات ذات الأداء التّفاضليّ)

نسبة الفقرات ذات الأداء التّفاضليّ		0	10%	50%	25%
Games-Howell	المتوسط الحسابيّ	0.117	0.127	0.524	0.840
0	0.117				
10%	0.127	0.010			
50%	0.524	**0.407**	**0.397**		
25%	0.840	**0.723**	**0.713**	**0.316**	

يتضح من الجدول (9) وجود تقارب غير دال إحصائيًا عند مستوى الدّلالة (α=0.05) للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (0%) و(10%). وكما يتضح من الجدول أن أكبر الفروق ذات الدّلالة الإحصائية عند مستوى الدّلالة α=0.05 للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (0%) و(25%).

كما تم إجراء اختبار (Games-Howell) للمقارنات البعديّة المتعددة للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمستويات المتغيّر (مستوى الأداء التّفاضليّ) كما يظهرها الجدول (10).

الجدول (10)

نتائج اختبار Games-Howell للمقارنات البعديّة المتعددة للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمستويات المتغيّر (مستوى الأداء التّفاضليّ)

مستوى الأداء التّفاضليّ		0.0	1.5	0.5	1.0
Games-Howell	المتوسط الحسابيّ	0.117	0.261	0.612	0.617
0.0	0.117				
1.5	0.261	**0.145**			
0.5	0.612	**0.495**	**0.351**		
1.0	0.617	**0.500**	**0.356**	0.005	

يتضح من الجدول (10) وجود تقارب غير دال إحصائيًا عند مستوى الدّلالة (α=0.05) للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة عندما يكون مستوى الأداء التّفاضليّ (0.5) و(1.0). كما يتضح من الجدول أن أكبر الفروق ذات الدّلالة الإحصائية عند مستوى الدّلالة (α=0.05) للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة عندما يكون مستوى الأداء التّفاضليّ (0) و(1.0). وفي ضوء النّتائج التي تظهر في الجدول (10) تبين وجود فروق ذات دلالة إحصائية عند مستوى الدّلالة (α=0.05) في الوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تُعزى لتفاعل متغيّريّ الدّراسة.

كما تمّ حساب الوسط الحسابيّ والإنحراف المعياريّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة، وذلك كما يظهرها الجدول (11).

الجدول (11)

الوسط الحسابيّ والإنحراف المعياريّ للخطأ المعياريّ لمعلمة الصّعوبة المقدّرة للفقرة تبعًا لمتغيّريّ الدّراسة

نسبة الفقرات ذات الأداء التّفاضليّ	**مستوى الأداء التّفاضليّ**	**العدد**	**المتوسط الحسابيّ**	**الإنحراف المعياريّ**
0%	0.0	60	0.102	0.00
10%	0.5	60	0.102	0.00
	1	60	0.101	0.00
	1.5	60	0.101	0.00
	الكلي	180	0.101	0.00
25%	0.5	60	0.128	0.06
	1	60	0.134	0.07
	1.5	60	0.105	0.01
	الكلي	180	0.122	0.06
50%	0.5	60	0.103	0.02
	1	60	0.112	0.03
	1.5	60	0.115	0.04
	الكلي	180	0.110	0.03
الكلي	0.0	60	0.102	0.00
	0.5	180	0.111	0.04
	1	180	0.116	0.05
	1.5	180	0.107	0.02

يلاحظ من الجدول (11) وجود فروق ظاهرية في الوسط الحسابيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ

الدّراسة؛ وللتّحقّق من جوهريّة الفروق الظّاهريّة، فقد تمّ إجراء تحليل التّباين الثّنائيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة، وذلك كما يظهرها الجدول (12).

الجدول (12)

نتائج تحليل التّباين الثّنائيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّريّ الدّراسة

مصدر التّباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدّلالة الإحصائية
نسبة الفقرات ذات الأداء التّفاضليّ	0.041	2	0.020	**16.544**	0.000
مستوى الأداء التّفاضليّ	0.007	2	0.003	2.750	0.065
نسبة الفقرات ذاتالأداء التّفاضليّ×مستوى الأداء التّفاضليّ	0.025	4	0.006	**5.039**	0.001
الخطأ	0.729	590	0.001		
الكلي	0.805	599			

يتبيّن من الجدول (12) وجود فروقٍ ذات دلالةٍ إحصائيّةٍ عند مستوى الدّلالة (α=0.05) في الوسط الحسابيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تُعزى لمتغيّر الدّراسة (نسبة الفقرات ذات الأداء التّفاضليّ)؛ ولكون المتغيّر متعدّد المستويات، فقد تمّ إجراء اختبار Levene للكشف عن وجود انتهاكٍ لافتراض تجانس التّباين في الوسط الحسابيّ لجميع قيم الخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تبعًا لمتغيّر الدّراسة، حيث أفادت قيمة F المحسوبة لاختبار (Levene) بوجود انتهاك لافتراض تجانس التّباين، ممّا أوجب استخدام أحد الاختبارات البعديّة التي تراعي التّباين غير المتجانس، وقد تم استخدام اختبار (Games-Howell) للمقارنات البعديّة المتعدّدة؛ بهدف تحديد لصالح إيٍّ من نسب الفقرات ذات الأداء التّفاضليّ قد كانت الفروق الجوهرية في الوسط الحسابيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة تبعًا لمتغير الدّراسة، وذلك كما هو مبيّن في الجدول (13).

الجدول (13)

نتائج اختبار Games-Howell للمقارنات البعديّة المتعدّدة للخطأ المعياريّ

لمعلمة صعوبة الفقرة تبعًا لمستويات المتغيّر (نسبة الفقرات ذات الأداء التّفاضليّ)

نسبة الفقرات ذات الأداء التّفاضليّ		10%	0%	50%	25%
Games-Howell	المتوسط الحسابيّ	0.101	0.102	0.110	0.122
10%	0.101				
0%	0.102	0.001			
50%	0.110	**0.009**	**0.008**		
25%9	0.122	**0.021**	**0.020**	0.012	

يتّضح من الجدول (13) وجود تقارب غير دال إحصائيًا عند مستوى الدّلالة α=0.05 للوسط الحسابيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) و(50%) وعندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (0%) و(10%)، كما يتضح من الجدول (13) أن أكبر الفروق ذات الدّلالة الإحصائية عند مستوى الدّلالة α=0.05 للوسط الحسابيّ للخطأ المعياريّ لمعلمة الصّعوبة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) و(10%). وفي ضوء النّتائج التي تظهر في الجدول (13) تبين وجود فروق ذات دلالة إحصائية عند مستوى الدّلالة (α=0.05) في الوسط الحسابيّ للخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة تُعزى لتفاعل متغيّريّ الدّراسة.

مناقشة النّتائج

تتناول مناقشة النتائج التي خلصت إليها هذه الدّراسة التي هدفت إلى الكشف عن أثر وجود الأداء التّفاضليّ في فقرات الاختبار على دقّة تقدير معلمة صعوبة الفقرة، ودراسة أثر نسبة الفقرات التي تبدي أداءً تفاضليًا إلى عدد الفقرات الكلّيّ للإختبار،

إضافةً إلى درجة الأداء التّفاضليّ الذي تبديه تلك الفقرات، اعتمادًا على نموذج راش الهرميّ باستخدام برنامج (HLM.7)، وكذلك التّوصيات المنبثقة عن هذه الدّراسة ونتائجها، وفيما يلي مناقشة النّتائج: حيث أشارت النّتائج المتعلّقة بمعلمة الصّعوبة للفقرات بأنّ أفضل تقديرات معلمة الصّعوبة دقّة عندما كانت درجة الأداء التّفاضليّ (1.0)، كما أنّ أفضل تقديرات معلمة الصّعوبة دقّة عندما كانت نسبة الفقرات ذات الأداء التّفاضليّ (25%). وتؤكّد هذه النتيجة أنّ الأداء التّفاضليّ المتوسّط أو القريب من الوسط من شأنه أن يسهم في مستوى دقّة تقدير معلمة صعوبة الفقرة سواءً كان ذلك من حيث نسبة الفقرات ذات الأداء التّفاضليّ أو مستوى ذلك الأداء لدى المجموعة التي طبّق عليها الاختبار. وقد تعزى هذه النتيجة إلى أنّ الفقرات ذات الأداء التّفاضليّ المتوسّط وكذلك النّسبة المتوسّطة لهذه الفقرات ينتج عنها تحيّزًا بدرجةٍ معيّنة، ممّا ينعكس عنه ارتفاع مستوى صعوبة الفقرة، وبالتّالي تكون هذه الفقرات مناسبةً لفئةٍ محدّدةٍ من الأفراد ممّا يؤثّر على الأفراد الذين يستجيبون على الفقرة ممن لا تتحيّز الفقرة لصالحهم. أمّا بالنّسبة للتّفاعل بين متغيّريّ الدّراسة (مستوى الأداء التّفاضليّ ونسبة الفقرات ذات الأداء التّفاضليّ) فيلاحَظُ بأنّ أعلى مستوىً لمعلمة الصّعوبة قد كان عند نسبة الأداء التّفاضليّ (25%) لمستويات الأداء التّفاضليّ (0.5، 1.0) وعندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (50%) يكون أعلى مستوى لمعلمة الصّعوبة عند مستوى الأداء التّفاضليّ (1.5)، ويعزو الباحث هذه النّتيجة إلى أنّ أثر درجة الأداء التّفاضليّ المرتفع (1.5) على معلمة الصّعوبة يكون طرديًّا، ممّا قد يزيد من مستوى التّحيّز الذي قد ينتج عن ذلك الأداء التّفاضليّ، وأنّ هذا الأداء من الممكن أن ينتج عنه تحيّزًا مفترضًا لا بدّ لمطوّر الاختبار الكشف عنه ومعالجته وضبطه. كما اشارت النّتائج الخاصّة بالفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة بان اعلى مستوى كان عند نسبة الأداء التّفاضليّ (25%)، في حين كان في ادنى مستوياته لدى نسبة الفقرات ذات الأداء التّفاضليّ (0%، 10%) التي لم يكن الاختلاف بينهما كبيرا وذات دلالة احصائية. اما الفرق بين نسبة الفقرات ذات التّفاضليّ (25%، 50%) فقد كان دالا لصالح نسبة الفقرات ذات الأداء التّفاضليّ (25%). ويعلّل الباحث ذلك بأنّ الفرق المطلق لمعلمة الصّعوبة يكون منخفضًا عند احتسابه لمعلمتي الصّعوبة المقدّرة والحقيقيّة عند النّسبة المنخفضة للفقرات ذات الأداء التّفاضليّ، باعتبار ان مستوى التّحيّز الذي يمكن أن تسهم فيه النّسب المنخفضة للفقرات ذات الأداء التّفاضليّ يكون قليلًا وهذا من شانه ان يسهم في تقليل درجة الفرق المطلق، أمّا بالنسبة للنسب المرتفعة للفقرات ذات الأداء التّفاضليّ، فإنّ ذلك من شأنه أن يسهم في زيادة مستوى التّحيّز؛ وبذلك يسهم في زيادة مستوى الفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة. كما اشارت النّتائج إلى أنّ هناك اختلافٌ دالٌّ بالفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة تبعًا لمتغيّر درجة الأداء التّفاضليّ، حيث تبيّن بأنّ درجة الأداء التّفاضليّ (0.5، 1.0) كان لديها اعلى متوسط للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة، وأنّ أكبر الفروق الدالّة للوسط الحسابيّ للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة عندما تكون درجة الأداء التّفاضليّ (0.0) و(1.0). ويعلّل الباحث هذه النّتيجة بأنّ ارتفاع مستوى الأداء التّفاضليّ من شأنه أن يلعب دورًا مهمًّا في زيادة مستوى التّحيز للفقرة والخاصّة بالفرق المطلق، نتيجةً للتّباينات التي من الممكن أن يضيفها التحيز الناتج عن الأداء التّفاضليّ على التّقديرات الخاصّة بمعلمة صعوبة الفقرة، الأمر الذي يسهم في اتّساع فجوة الفرق المطلق بين تقديرات الفقرة المقدّرة والحقيقيّة. وقد أظهرت النّتائج المتعلّقة بالتّفاعل بين متغيّريّ الدّراسة (نسبة الفقرات ذات الأداء التّفاضليّ ومستوى الأداء التّفاضليّ) للفرق المطلق بين معلمتي الصّعوبة المقدّرة والحقيقيّة، أنَّ أفضل التقديرات لمعلمة الصّعوبة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (10%) عند جميع درجات الأداء التّفاضليّ المضمّنة في الدّراسة، وكذلك عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (0%) ودرجة الأداء التّفاضليّ (0)، وكذلك عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) ومستوى الأداء التّفاضليّ (1.5)، وأظهرت النّتائج كذلك أنّ أقلّ التقديرات دقّة لمعلمة الصّعوبة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) لكل من درجة الأداء التّفاضليّ (0.5، 1.0). وقد تعزى هذه النّتيجة إلى العلاقة المتبادلة بين النسبة المتوسّطة للفقرات ذات الأداء التّفاضليّ والدّرجة المتوسّطة للأداء التّفاضليّ، وبين هذين المتغيّريّن (نسبة الفقرات ذات الأداء التّفاضليّ المتوسط ومستوى الأداء التّفاضليّ) والفرق المطلق لمعلمتي الصّعوبة المقدّرة الحقيقيّة والذي يزداد مع زيادة التحيّز الذي يسهم في زيادته ارتفاع نسبة الفقرات ذات الأداء التّفاضليّ ودرجة الأداء التّفاضليّ.

كما أنّ النّتائج الخاصّة بالخطأ المعياريّ لمعلمة الصّعوبة تظهر عدم وجود فرق دالٍّ إحصائيًا لمتوسّط الخطأ المعياريّ لمعلمة صعوبة الفقرة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) و(50%)، وعندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (0%) و(10%)، وأن أكبر الفروق ذات الدّلالة الإحصائيّة كانت لمتوسّط الخطأ المعياريّ لمعلمة الصّعوبة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) و(10%). وتشير هذه النتيجة إلى أنّ نسبة الفقرات ذات الأداء التّفاضليّ (25%) تسهم في أعلى مستوىً من الأخطاء المعياريّة المقدّرة؛ ويمكن تعليل ذلك بأنّ المستوى المتوسّط من نسبة الفقرات ذات

الأداء التّفاضليّ تسهم في تحيّزها في ارتفاع مستوى الخطأ المعياريّ وقد يعود ذلك إلى انحصار التّوزيع المعياريّ الخاص بهذه الفئة، الأمر الذي قد يتأثّر بتشتّت بعض الفقرات مما يؤدّي إلى ارتفاع مستوى الخطأ المعياريّ والذي يؤدّي إلى زيادة مستوى التّحيّز الذي قد ينتج عن الأداء التّفاضليّ. ويمكن استنتاج أنّ انخفاض الأداء التّفاضليّ يزيد من قوة الاختبار.

كذلك فقد اشارت نتائج التفاعل بين متغيّريّ نسبة الفقرات ذات الأداء التّفاضليّ ومستوى الأداء التّفاضليّ للخطأ المعياريّ لمعلمة الصّعوبة بأنّ أقلّ قيم للخطأ المعياريّ لمعلمة الصّعوبة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (10%) لكل من مستويات الأداء التّفاضليّ (0.5، 1.0، 1.5)، وكذلك عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (0%) ومستوى الأداء التّفاضليّ (0.0)، وكذلك عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (50%) ومستوى الأداء التّفاضليّ (0.5)، وكما يتّضح من الشّكل (5) أنَّ أكبر قيم الخطأ المعياريّ لمعلمة الصّعوبة المقدّرة عندما تكون نسبة الفقرات ذات الأداء التّفاضليّ (25%) لكل من مسوى الأداء التّفاضليّ (0.5، 1.5). ويعلّل الباحث هذه النتيجة من خلال عاملين: الأوّل يتمثّل بضيق التّوزيع لمعلمة الصّعوبة عند نسبة الفقرات ذات مستوى الأداء التّفاضليّ المتوسّط، والعامل الثّاني يتمثّل بالتّشتّت الذي من الممكن أن يضيفه الأداء التّفاضليّ عند مستوى الأداء (0.5، 1.5)، ممّا يُسهم في ارتفاع مستوى الخطأ المعياريّ.

استناداً إلى النتائج التي توصل اليها الباحث فإنه يوصي بإجراء دراسةٍ مماثلة باستخدام نماذج هرميّة أخرى كالنّماذج الثّنائيّة أو النّماذج الثّلاثيّة؛ لدراسة معالم أخرى للفقرة كالتّمييز والتّخمين، وإجراء دراسةٍ مماثلة حيث يكون فيها أطوال إختباراتٍ متفاوتةٍ، وكذلك يوصي بإجراء بحث لدراسة أثر نسبة الفقرات ذات الأداء التّفاضليّ ومستوى الأداء التّفاضليّ بالإعتماد على عيّنتين مختلفتين من حيث عدد الأفراد، لإضافة مقارنةٍ جديدةٍ تتمثّل بأثر حجم العينة على دقّة التّقدير. واعتماد مطوّري الاختبارات على النّماذج الهرميّة في الكشف عن الأداء التّفاضليّ وضبطه، لدقّة هذه النّماذج في التّعامل مع الفقرات ذات الأداء التّفاضليّ. واخيرًا معالجة الأداء التّفاضليّ وضبطه من خلال النّماذج الهرميّة لدقّتها، بدلًا من حذف الفقرات التي تبدي أداءً تفاضليًّا؛ لأنّ ذلك يقلّل من قيم ثبات الاختبار وصدقه.

المراجع

المراجع العربية:

عبيدات، عبد الله. (2014)، استقصاء الأداء التّفاضليّ لفقرات اختبار اختيـار مـن متعدد تبعًا لدرجة مخاطرة المفحوصين من مستويات تحصيل مختلفة، أُطروحة دكتوراة غير منشورة، جامعة اليرموك.

المراجع الأجنبية:

Angel, M., Gideon, D. and Jose, M. (2000). Effects of Amount of DIF, Test Length, and Purification type on Robustness and Power of Mantel-Hanszel Procedures. Institute for Science Education, 5, (3), 44-53.

Camilli, G. and Shepard, L. A. (1994). Methods for Identifying Biased Test Items. Thousand Oasks, CA: SAGE Publications.

Chu, K. (2002). Equivalent Group Test Equating With the Presence of Differential Item Functioning. Unpublished doctoral dissertation. Florida State University.

Chu, K. and Kamata, A. (2005). Test equating in the Presence of DIF Items. Journal of Applied Measurement. Special Issues: The Multilevel Measurement Model, 6(3), 342.

Ellis, B. and Raju, S. (2003). Test and Item Bias: What they are, what aren't, and How to detect them. Measuring up. Assessment issues for teachers, counselors and administrator. (ERIC Document Reproduction Service No.ED480042).

Embreston, S. and Reise, S. (2000). Item response theory for psychologists. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Gierl, M., Jodoin, M. and Ackerman, T. (2000). Performance of Mantel-Hanszel, Simultaneous Item Bias Test, and Logistic Regression When the Proportion of DIF Items is Large. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, USA. April, 24-27, 2000.

Gokiert, R. J. and Ricker, K. L. (2004). Gender differential item functioning on the WISC-II: Analysis of the Canadian standardization sample (Centre for research in applied measurement and evaluation). Edmonton, Canada: University of Alberta.

Holland, Paul W. and Dorans, Niel J (1993). DIF Detection and Description Mantel-Hanszel and Standardization. Educational Testing Service Princeton, N.J. (QAT 24225) ED 387526.

Jasper, F. (2010). Applied Dimensionality and Test Structure Assessment with the START-M Mathematics test. The International

Journal of Educational and Psychological Assessment. 6(1), 109-126.

Kamata, A. (1998). Some Generlaizations of the Rasch Model: an Application of the Hierarchical Generalized Linear Model. Unpublished doctoral dissertation. Michigan State University.

Kamata, A. (2001). Item Analysis by the Hierarchical Generalized Linear Model. Journal of Educational Measurement, 38, 79-93.

Kamata, A. and Vaughn, B. (2004). An Introduction to Item Functioning Analysis. A Contemporary Journal, 2(2), 49-62.

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

Kim, W. (2003). Development of a differential item functioning (DIF) procedure using the hierarchical generalized linear model: A comparison study with logistic regression procedure. (Unpublished doctoral dissertation). The Pennsylvania State University, Philadelphia

Luppescu, S. (2002). DIF detection in HLM. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Miller, M. (1992). Effect of Sample Size, Number of Biased Items, And Magnitude of Biase on Two-Stage Item Biase Estimation Methods. Applied Psychological measurement. l, 16, (4): (381-388).

Tan, X. and Jeral, M. (2005). Using Local DIF analyses to assess group differences on multilingual examinations. Applied Measurement in Education, 2, 313-334.

Teresi, J. (2001). Statistical methods of examination of differential item functioning with applications to cross-cultural measurement of functional, physical and mental health. Mental Health and Aging, 7: 31-40.

Binary Item Difficulty Estimation Accuracy According to Hierarchical Model in Light of Variation of Percentage of Differential Performance Items and the Level of that Performance

Mohammad A. Abu Sahion *

ABSTRACT

This study aimed to identify the effectiveness of hierarchical Rasch model in item difficulty estimation with different cases of percentage and DIF level by using simulation data. The study used the descriptive methodology by generating binary responses (0,1) for samples that simulate the original society samples by using Monte Carlo method using WinGen v.3 programm.

Number of individuals in this study was (2400) persons (1200 in the reference group and 1200 in the focal group) in each interaction cell, covering (10) situations that arouse from the interaction between the study variables (percentage of items with DIF (10%, 25% and 50%), and DIF magnitude (0.5, 1.0 and 1.5). The test length was (60) items in each situation.

The results showed that hierarchical Rasch model gave the best accuracy estimation of difficulty parameter when the percentage of (DIF) is (0%), when the percentage of items with a DIF was (10%) with DIF magnitudes of (0.5, 1.0 and 1.5) and when the percentage of items with a DIF was (25%) with DIF magnitude (1.5), while the lowest accuracy of difficulty parameter estimation was recorded when the percentage of items with DIF was (25%) with DIF magnitudes of (0.5 and 1.0).

Keywords: Percentage of items with differential functioning, Level of items with differential functioning, Hierarchical Rasch model, Item difficulty parameter, Standard error of difficulty parameter.

* The Ministry of Education, Jordan. Received on 22/11/2015 and Accepted for Publication on 03/08/2016.