# أثر وجود اداء تفاضلي في الفقرات المرساوية على دقة المعادلة العمودية لاختبار اوتيس لينون للقدرة العقلية

**راتب صايل الخضر الرحيل**

**كلية التربية، جامعة**

***الملخص_*** *هدفت هذه الدراسة إلى الكشف عن أثر وجود أداء تفاضلي للجنس في الفقرات المرساوية (anchor items) على دقة المعادلة العمودية لاختبار اوتيس لينون للقدرة العقلية بمستوييه المتوسط والمتقدم الصورة "K" والمعدلتين للبيئة الأردنية. تكونت عينة الدراسة من (1001) طالب وطالبة مقسمين إلى (500) في مستوى القدرة المتوسط و(501) في مستوى القدرة المتقدم. تم استخدام تصميم الفقرات المشتركة (المرساوية) والمجموعات غير المتكافئة لجمع البيانات حيث طبق اختبار المستوى المتوسط على أفراد مستوى القدرة المتوسطة، وطبق اختبار المستوى المتقدم على أفراد القدرة المتقدمة، وتضمن كل من الاختبارين فقرات مشتركة بينهما بلغ عددها (25) فقرة مشتركة، وبعد تصحيح الاستجابات وتجهيز ملف بيانات خاص بكل مستوى تم استخدام طريقة مانتل هانزل للكشف عن الفقرات ذات الأداء التفاضلي تبعاً لمتغير الجنس في مستويي الاختبار بشكل منفصل، وبعد ذلك تم إجراء المعادلة العمودية باستخدام برمجية BILOG-MG وفقاً للنموذج أحادي المعلمة في نماذج نظرية استجابة الفقرة بعد التحقق من افتراضات هذا النموذج ومطابقة البيانات له، وأجريت المعادلة العمودية بين مستويي الاختبار المتوسط والمتقدم مرتين: المرة الأولى في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي للجنس، والمرة الثانية في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس، وذلك اعتمادًا على درجات القدرة للمفحوصين. ولتحقيق المقارنة بين عمليتي المعادلة تم استخدام معامل الصدق التقاطعي (Cross-Validation) معياراً لمدى استقرار النتائج لدرجات القدرة المعادلة ومؤشراً على فعّالية المعادلة.*

*أشارت النتائج إلى ان إجراء المعادلة في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي زاد من دقة المعادلة العمودية بالاستناد إلى قيم معاملات الصدق التقاطعي.*

***الكلمات المفتاحية:*** *نظرية استجابة الفقرة، نموذج أحادي المعلمة، الفقرات المرساوية، الأداء التفاضلي للفقرات، طريقة مانتل هانزل، المعادلة العمودية، الصدق التقاطعي.*

## 1. المقدمة

إن من أهم القضايا التي يواجهها أي مجتمع في عصرنا هي استخدام الاختبارات لغايات مقارنة الأفراد وتصنيفهم، وهنا لا بد من التركيز على أمرين رئيسيين في الاختبار هما: محتوى الاختبار وصعوبة الاختبار. فإن كان الاختبار المطبق مثلاً في عدة جامعات يتضمن محتوى موحد وصعوبة فقراته موحدة عندها تصبح تلك اختبارات متوازية. ويرى أنجوف [1] أن معظم درجات الاختبار، إن لم تكن جميعها، تتضمن جزأين: الجزء الأول قدرة المفحوص والجزء الثاني صعوبة نموذج الاختبار المطبق على ذلك المفحوص. ويمكن إزالة الجزء الثاني من الدرجات من خلال معادلة نماذج الاختبار ووضعها على مقياس معياري موحد، فعملية المعادلة تمكن من مقارنة الدرجات المأخوذة من نماذج مختلفة دون التخوف من اختلاف درجات الصعوبة للنماذج.

معادلة الاختبارات Test Equating:

إن بطاريات الاختبارات عادة تستخدم نماذج مختلفة لنفس

الاختبار، حيث أن سرية الاختبار وامتداد الاختبار ليغطي مدى واسع من القدرات يتطلب تعدد النماذج. وهناك حاجة لمقارنة درجات المفحوصين على نماذج مختلفة للاختبار من خلال المعادلة. لقد قدمت معايير الاختبارات النفسية والتربوية [2] تعريفاً لمعادلة الاختبارات على أنها طريقة تستخدم لتحويل درجات أحد نماذج الاختبار إلى درجات النموذج الآخر بحيث تصبح هذه الدرجات متكافئة أو متوازية. وأن الهدف من المعادلة هو السماح بإجراء مقارنة للنتائج التي تحصل عليها من نماذج مختلفة للاختبار.

وقد أوضح أنجوف [3] أن بواسطة المعادلة يمكن قياس النمو، ورسم التوجهات، ومطابقة البيانات المشتقة من نماذج مختلفة للاختبار رغم اختلاف خصائص الفقرات، وبالتالي مقارنة أداء الأفراد على نماذج مختلفة من الاختبار. إن المعادلة المناسبة تمثل ضماناً لعدالة وثبات وصدق الاختبار. فمثلاً إن طُبق نموذجين للاختبار على فردين وكان هناك اختلاف في صعوبة الفقرات بين النموذجين فإن المعادلة تضمن إزالة فارق التحصيل العائد لاختلاف صعوبة النموذجين وتجعل المقارنة بينهما عادلة. ويكون الاختبار ثابتاً عندما تسمح المعادلة بالمقارنة بين الدرجات عبر النماذج، وعندها يكون صادقاً لأن النماذج المختلفة تعطي تعميمات وتفسيرات متشابهة حول قدرات المفحوصين.

أنواع المعادلة لدرجات الاختبارات:

1. المعادلة الأفقية Horizontal Equating: وهي معادلة نماذج الاختبارات المصممة لقياس نفس السمة على نفس المستوى من القدرة ولنفس المجتمع [4].
2. المعادلة العمودية Vertical Equating: لقد جذب هذا النوع من المعادلة اهتمام العديد من الباحثين. وقد عرفه سلند ولن [4] على أنه عملية تحويل علامات على نماذج للاختبار صممت لمجتمعات عند مستويات قدرة مختلفة. وقد ينظر للمعادلة العمودية على أنها الحصول على مقياس موحد (Common scale) يمتد عبر عدد من الصفوف المدرسية أو الأعمار المختلفة للطلبة.

تصاميم المعادلة Equating Design: ويقصد بتصميم المعادلة الطريقة التي يتبعها الباحث في جمع البيانات. وقد وضحت هذه التصاميم من قبل كل من أنجوف، ولورد، كوك وأيجنور وهاميلتون وسوامنثان [3,5,6,7]

ويمكن تلخيص هذه التصاميم في أربعة تصاميم شائعة الاستخدام هي:

1. تصميم المجموعات العشوائية Random Groups Desig: ويتألف من مجموعتين عشوائيتين ونموذجين مختلفين للاختبار. ويأخذ كل مفحوص نموذج واحد فقط من نماذج الاختبار.
2. تصميم المجموعة الواحدة Single Group Design: حيث يوجد نموذجين للاختبار ويأخذ كل مفحوص كلا النموذجين، وقد يدخل عامل الإرهاق في هذا النمط.
3. تصميم المجموعة الواحدة المتوازنة عكسياً Single Group: with Counter Balancing Design وفي هذا التصميم يتم تقسيم مجموعة المفحوصين إلى مجموعتين متساويتين تقسيماً عشوائياً وتطبق صورتي الاختبار عليهما بطريقة التوازن العكسي؛ أي أن إحدى المجموعتين تأخذ الصورة الأولى للاختبار والمجموعة الثانية تأخذ الصورة الثانية. ثم تأخذ كل مجموعة الصورة التي لم تأخذها بالمرة الأولى مباشرة بعد أخذها للصورة السابقة، بشرط أن يكون الوقت المحدد للصورتين متساوياً للمجموعتين. والفكرة وراء هذا التصميم هي ضمان أن عوامل التعب وآثار التعلم السابق والممارسة والخبرة لها نفس التأثير في صور الاختبار قيد المعادلة.
4. تصميم المجموعات ذات الاختبار المشترك – Common items Groups Design أو ما يسمى تصميم المرساة

(Anchor-Design) في هذا النمط يطبق نموذج واحد من نماذج الاختبار على كل مجموعة ويتضمن كل نموذج عدد من الفقرات المشتركة (Common items)؛ وهي فقرات من كلا النموذجين المراد معادلتهما، ويمكن تطبيق الفقرات المشتركة منفصلة أو مع النماذج وقد تكون الفقرات المشتركة داخلية أو خارجية. فالفقرات المشتركة الداخلية يتم تضمينها في صورتي الاختبار المراد معادلتهما أما الخارجية فتطبق بشكل منفصل عن صورتي الاختبار وقد يلجأ بعض الباحثين بمعادلة نماذج الاختبار عبر استخدام مجموعة من الأفراد المشتركين بدلاً من الفقرات المشتركة. إن اختيار أي من هذه التصاميم السابقة يعتمد على عوامل متعلقة بالتطبيق وليس فقط لرغبات إحصائية [3].

تقويم دقة المعادلة وفعاليتهاTest Equating Evaluation:

أشار كولن [8] إلى أنه لا يوجد محك متفق عليه وحاسم للحكم على دقة المعادلة وفعاليتها، كما ذكرت هاريس وكروس [9] إن هناك عدد من المعايير للحكم على معادلة الاختبار وأن هذه المعايير قد لا تناسب جميع طرق وأساليب المعادلة. ونذكر ومن هذه المعايير:

محك الصدق التقاطعي (Cross-Validation): وهو متوسط مربعات الانحرافات للعلامات المتعادلة عندما يطبق على عينات مختلفة لمعرفة مدى استقرار النتائج. ويمكن حسابه من المعادلة التالية:

$$C = \frac{\sum (Xi - Yi)^2}{NK} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

حيث:

C: معامل الصدق التقاطعي.

Xi: العلامات المعادلة في العينة الأولى.

Yi: العلامات المعادلة في العينة الثانية.

N: عدد العلامات الملاحظة في توزيع الصدق التقاطعي.

K: عدد الفقرات في الاختبار المتعادل.

ثالثاً: الأداء التفاضلي للفقرات Differential Item (DIF) functioning:

الأداء التفاضلي للفقرة هو دلالة مشتقة إحصائياً للتعبير عن الفروق في احتمالات الاستجابة الصحيحة على الفقرة بين مجموعتين من المفحوصين ممن هم في المستوى نفسه من القدرة [10]. وقد عمل المختصون في بنا الاختبارات على استخدام مفهوم الأداء التفاضلي للفقرة في الأدب التربوي للتعبير عن الطرق والمعاني الإحصائية المستخدمة للكشف عن تحيز الفقرات من منطلق أن مصطلح التحيز يحمل جوانب سلبية ومشحونة سياسياً واجتماعياً [11].

طريقة مانتل – هانزل Mantel – Haenzel Method:

تعتبر هذه الطريقة من أكثر الطرق التقليدية انتشاراً واستخداماً في الكشف عن الأداء التفاضلي للفقرات [12]. حيث تعمل هذه الطريقة على استقصاء الأداء التفاضلي بين مجموعتين إحداهما تسمى المجموعة المرجعية Reference Group والأخرى تسمى المجموعة المستهدفة Focal Group وهي المجموعة التي يعتقد أنها تتأثر بالأداء التفاضلي للفقرات. ومن خصائص هذه الطريقة سهولة تنفيذ خطواتها، وقوتها عندما تتبع الفقرة للنموذج أحادي المعلم (نموذج راش).

وتعتمد هذه الدراسة على مستويين من مستويات اختبار اوتيس لينون للقدرة العقلية هما المستوى المتوسط والمستوى المتقدم للصورة (K)

## 2. مشكلة الدراسة

### أ. مشكلة الدراسة واسئلتها

ان اختبارات الاستعداد واختبارات القدرات العقلية غالبا ما تبدي بعض فقراتها اداء تفاضليا للجنس [13] ويلجأ البعض لإجراء معادلة لنماذج من الاختبارات لوضعها على مقياس موحد يساعد في مقارنة درجات المفحوصين بشكل مباشر، وإذا علمنا ان

عملية المعادلة قد تعتمد بشكل اساسي على وجود فقرات مرساوية تربط بين نماذج الاختبار المراد معادلتها فانه من المهم التعرف على خصائص تلك الفرات التي ستقوم عليها عملية المعادلة، ومن هذه الخصائص التي قد يهملها البعض الاداء التفاضلي للفقرات المرساوية. وتهدف هذه الدراسة إلى الكشف عن الفقرات ذات الأداء التفاضلي لمتغير الجنس لاختبار أوتيس لينون للقدرة العقلية المستوى المتوسط والمستوى المتقدم, كما تهدف هذه الدراسة إلى تقصي مدى تحقق فعالية المعادلة العمودية لمستويي الاختبار في ظل وجود الفقرات المرساوية ذات الأداء التفاضلي لمتغير الجنس وفي ظل حذف تلك الفقرات وذلك وفقاً لنظرية استجابة الفقرة.

1. ما مدى تحقق معايير الفاعلية للمعادلة العمودية لمستويي الاختبار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي للجنس؟
2. ما مدى تحقق معايير الفاعلية للمعادلة العمودية لمستويي الاختبار عند حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس؟

**ب. أهمية الدراسة**

تبرز أهمية هذه الدراسة من فائدة وضع مقياس موحد لمستويات مختلفة للقدرة والذي بدوره يوفر الوقتٍ والجهد في عملية القياس، مما يمكننا ذلك من مقارنة الأفراد ذوي القدرات المختلفة بشكل مباشر، وملاحظة تطور هذه القدرات واستخدام درجات الاختبارات بشكل متبادل من خلال المعادلة العمودية – vertical equating. كما تبرز أهمية هذه الدراسة من هدفها في السعي لتقصي أثر الفقرات المرساوية ذات الأداء التفاضلي على دقة المعادلة العمودية للاختبارات باستخدام النموذج أحادي المعلمة من نماذج نظرية استجابة الفقرة، وذلك لتوضيح مدى تأثر فعالية المعادلة العمودية بوجود هذه الفقرات وبالتالي تقدم هذه الدراسة مساهمة في مجال بناء وتطبيق الاختبارات وتفسيرها وكيفية التعامل مع وجود أداء تفاضلي عند مقارنة قدرات مختلفة خلال المعادلة. كما أن هذه الدراسة تقدم مساهمة في توسيع الفهم حول المعادلة العمودية لدرجات الاختبار باستخدام نظرية استجابة الفقرة وذلك لقلة الدراسات العربية حول هذا الموضوع.

**ج. حدود البحث**

1. استخدام نموذج احادي المعلمة.
2. استخدام طريقة مانتل هانزل في الكشف عن أداء الفقرات التفاضلي للجنس.

**د. مصطلحات البحث**

**معادلة درجات الاختبار:**

تحويل نظام الدرجات الخاص بإحدى نموذجي الاختبار إلى نظام درجات خاص بالنموذج الآخر.

**المعادلة العمودية (Vertical – equating):**

تحويل الدرجات على نماذج للاختبار صممت لمجتمعات عند مستويات قدرة مختلفة إلى تدريج مشترك.

**الأداء التفاضلي للفقرة (Differential Item functioning):**

دلالة مشتقة إحصائياً للتعبير عن الفروق في الاستجابة للفقرة بين مجموعتين من المفحوصين ممن هم في المستوى نفسه من القدرة.

**اختبار القدرة العقلية:**

اختبار يقيس القدرة العقلية، والاختبار المستخدم في هذه الدراسة هو اختبار أوتيس – لينون للقدرة العقلية المستويين المتوسط صورة "K"، والعالي (المتقدم) صورة "K" المعدلتين للبيئة الأردنية وتشكل جزءاً من مجموعة سلسلة اختبارات أوتيس – لينون التي تغطي المراحل العمرية من رياض الأطفال حتى الصف الثاني عشر.

**المعايرة بالتلازم concurrent calibration:**

طريقة للحصول على تقديرات لمعالم نماذج استجابة الفقرة واقعة على مقياس مشترك، حيث يتم التعامل مع جميع البيانات الاختبارية في تشغيل واحد، وفي حالة تصميم الفقرات المشتركة

يتم ضم الاستجابات على الفقرات المشتركة مع الاستجابات على كل مستوى من مستويات الاختبار خلال المعايرة للحصول على معالم للفقرات والأفراد واقعة على مقياس مشترك.

**معامل الصدق التقاطعي:**

متوسط مربعات الانحرافات للعلامات المتعادلة عندما يطبق الاختبار (النماذج) على عينات مختلفة لمعرفة مدى استقرار النتائج.

**الفقرات المرساوية (anchor items):**

فقرات مشتركة بين نموذجي الاختبار وتستخدم في معادلة الاختبار.

### 3. الاطار النظري والدراسات السابقة

قام هوجوان [14] بدراسة هدفت للمقارنة بين عدة طرق مستخدمة في المعادلة العمودية وفق نظرية استجابة الفقرة وقد كانت هذه الطرق: المعايرة المنفصلة والتي تحتاج لإجراءات ربط بين النماذج، وطريقة المزاوجة بالتلازم Pair - Wise concurrent، وطريقة شبه التلازم Semi – concurrent وطريقة المعايرة بالتلازم Concurrent calibration والتي لا تحتاج لإجراءات ربط بعدية لتحويل الدرجات بين نماذج الاختبار، وأظهرت النتائج أن طريقة المعادلة باستخدام المعايرة بالمزاوجة وطريقة المعادلة بشبه التلازم أعطت نتائج جيدة تلتها المعادلة بالتلازم ثم المعادلة بالمعايرة المنفصلة وذلك اعتماداً على مؤشرات الخطأ المعياري للمعادلة، والتحيز المطلق Absolate biase.

أجرى دراسة الحواري [15] هدفت إلى تعرف الخصائص السيكومترية لصور مختارة من اختبارات الرخصة الدولية لقيادة الحاسوب في الأردن، ومعادلة درجاتها، حيث استخدمت كل من المعادلة الخطية الأفقية والمعادلة المئينية الأفقية كطريقتين للمعادلة وفق النظرية الكلاسيكية في القياس، وباستخدام نموذج راش الأحادي المعلمة من اختبار الرخصة الدولية لقيادة الحاسوب المكون من سبعة اختبارات فرعية تتعلق بالمهارات الأساسية لاستخدام الحاسوب وهي من الاختبارات الثنائية التدريج وتكون مجتمع الدراسة من جميع المعلمين الملتحقين ببرنامج الرخصة الدولية لقيادة الحاسوب في مدينة إربد خلال الفترة الزمنية 2004–2006. أما عينة الدراسة فتكونت من 400 مفحوص لكل نموذج من نماذج الاختبار الفرعي تم اختبارهم بالطريقة العشوائية واعتمد تصميم المجموعات العشوائية المتكافئة في جمع البيانات. وتم تقدير معالم الفقرات ومعلم القدرة للأفراد، وحساب الفاعلية النسبية لكل اختبار فرعي ولجميع النماذج، وتم اعتماد معياري الخطأ المعياري ومعامل الصدق التقاطعي في الحكم على فاعلية المعادلة أظهرت النتائج أن المعادلة وفق نموذج راش الأحادي المعلمة هي الأفضل والأكثر فاعلية في معادلة درجات المفحوصين من خلال وضعهم على ميزان مشترك واحد، ثم يليها المعادلة الخطية وأخيراً المعادلة المئينية.

وقام سنهاري وهولاند [16] بدراسة لتقصي الاعتقاد القائل بأن اختبارات الربط التي تستخدم في دمج مفردات اختبارين في تدريج واحد مشترك لإجراء معادلة الاختبارات المختلفة بأنها اختبارات مصغرة من حيث تمثيل المحتوى والخصائص الإحصائية للاختبارات التي تجري عليها المعادلة، كما هدفت الدراسة لتقصي نتائج العديد من الدراسات التي اهتمت بمعادلة الاختبارات والمستخدمة لخصائص إحصائية مختلفة لاختبارات الربط ونسب مختلفة من تمثيل المحتوى الكلي لمجال الاختبار. وقد استخدمت الدراسة في إجراءاتها على بيانات حقيقية تم الحصول عليها من بعض الدراسات السابقة وبيانات مولدة لتحقق أهداف الدراسة، وتم إعداد اختبارات ربط مختلفة من حيث اتساع مدى صعوبة المفردات واختلاف هذا المدى عن مدى صعوبة الاختبار الكلي المستهدف في إجراء معادلة له، وكذلك اختبارات ربط تختلف من حيث نسب تمثيلها لمجال محتوى الاختبار الكلي، تم استخدام انحرافات المعادلة والخطأ المعياري للتحقق من كفاءة عمليات

وإجراءات المعادلة، وأظهرت النتائج أن اختبار الربط يجب أن يتصف بنفس الخصائص التي يتصف بها الاختبار الكلي.

وأجرى اوتيروجيك وفالن دراسة [17] هدفت إلى الكشف عن أسباب ومصادر التحيز لفقرات اختبار التحصيل النهائي للمرحلة الأساسية الخاص بالطلبة المهاجرين إلى ألمانيا، وهو اختبار في الرياضيات واللغة ومعالجة المعلومات. تم اخذ ثلاثة صور للاختبار بواقع (180) فقرة لكل صورة، واستخدمت طريقة مانتل هانزل، وطريقة الإجراءات التحكيمية للكشف عن المصادر المحتملة للتحيز، وذلك بتحليل محتوى الفقرات المتحيزة لصالح الطلبة المهاجرين او ضدهم ومن ثم تقييم الفقرات من قبل محكمين. وقد أظهرت طريقة مانتل هانزل تحيز (48) فقرة لصالح الطلبة المهاجرين، وكشفت الدراسة عن عدة مصادر للتحيز، أهمها: المكونات اللغوية للفقرات، المهارات اللغوية، الاختلاف في المهارات الرياضية، مؤشرات الإجابة للفقرات. وأشار المحكمون إلى أنه من الصعب أحيانا تأكيد السبب الذي يجعل فقرة ما متحيزة دون أخرى.

كما قام ميندس وإيركان [18] بدراسة للتعرف على مصدر الأداء التفاضلي وفقاً للجنس في اختبار الرياضيات الأساسية British Columbia Principles of mathematical Exam (SIBTEST) المطبق على طلبة الصف العاشر، وكانت فرضيتهم أن محتوى الفقرة، والإدراك، والسياق الذي تقدم فيه الفقرة قد تكون مصدراً للأداء التفاضلي وفقاً للجنس. وبينت النتائج أن الفقرات المتضمنة حزمة من المشكلات واللوغاريتم، والأسس كانت مفضلة لدى الذكور، بنما الفقرات المتضمنة علاقات رياضية، والفقرات المتضمنة أوصافاً اسمية كانت مفضلة لدى الإناث، كما أظهرت النتائج تفوق الذكور في الاستجابة على الفقرات التي تتطلب عمليات إدراكية عليا ورسوما وصور.

أجرى دريانا [19] دراسة هدفت إلى الكشف عن الفقرات ذات الأداء التفاضلي المرتبط بمتغير الجنس ومتغير المنطقة في اختبار التفوق الرياضي للصف التاسع والمكون من (40) فقرة من نوع الاختيار من متعدد والذي يطبق في ولاية أوهايو ولهذا الغرض تم استخدام البيانات الناتجة عن تطبيق الاختبار عام 2005 على عينة من الطلبة بلغت (19198) من كلا الجنسين في المدارس العامة في أوهايو، تم استخدام خطوات مانتل هانزل للكشف عن الأداء التفاضلي المنتظم للفقرات، بينما استخدم اختبار بريسلو– داي للتجانس (Breslow – Day test of homogeneity) للكشف عن الأداء التفاضلي غير المنتظم. أشارت النتائج إلى وجود أربع فقرات ذات أداء تفاضلي مرتبط بالجنس، بينما لم تظهر أي من الفقرات أداء تفاضلي مرتبط بالمنقطة.

من خلال ما ورد في الدراسات السابقة يتبين لنا عدم وجود اتفاق على طريقة مثالية لإجراء المعادلة لدرجات الاختبارات، وهذا ما أكده سكاجز وليستز [20] بعد مراجعتهما نتائج عدد من الدراسات التي تناولت موضوع المعادلة الأفقية وتوصلا لعدم وجود طريقة وحيدة ومثلى في المعادلة، إلا أن النظرية الحديثة تفوقت على الطرق الكلاسيكية عندما تحققت الافتراضات الخاصة بها.

أما بالنسبة لهذه الدراسة فقد اعتمدت على النموذج أحادي المعلمة (نموذج راش) في إجراء المعادلة العمودية، كما اعتمدت على طريقة مانتل – هانزل في الكشف عن الأداء التفاضلي للفقرات وفق متغير الجنس في اختبار القدرة العقلية وصولاً لإجراء معادلة عمودية بين نموذج الاختبار المستوى المتوسط "k" ونموذج المستوى المتقدم "k"، مرة ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي ومرة أخرى بعد إزالة الفقرات المرساوية ذات الأداء التفاضلي لمتغير الجنس، ومقارنة مدى تحقق دقة المعادلة العمودية بين حالتي المعادلة اعتمادا على معيار الصدق التقاطعي. وذلك بعد التحقق من مطابقة البيانات مع نموذج راش الأحادي المعلم، وفحص تحقق افتراضات نظرية استجابة الفقرة وفق الخطوات التي سيتم ذكرها في الطريقة والإجراءات.

## 4. الطريقة والإجراءات

### أ. مجتمع الدراسة وعينتها:

يتكون مجتمع الدراسة من جميع طلاب وطالبات الصفوف: السابع والثامن والتاسع والعاشر والحادي عشر والثاني عشر والبالغ عددهم (9298) طالباً وطالبة، منهم (4520) طالباً و(4778) طالبة، موزعين على (45) مدرسة حكومية؛ منها (23) مدرسة للذكور و(22) مدرسة للإناث كما ورد في سجلات مديرية تربية بني كنانة للعام 2011/2010.

**جدول 1**

**يبين توزيع مجتمع الدراسة حسب متغيرات الجنس والمستوى التعليمي ومستوى القدرة (*)**

| | مستوى القدرة المتوسط | | | مستوى القدرة المتقدم | | | |
|---|---|---|---|---|---|---|---|
| | السابع | الثامن | التاسع | العاشر | الحادي عشر | الثاني عشر | المجموع |
| ذكور | 917 | 965 | 1013 | 981 | 332 | 312 | 4520 |
| إناث | 905 | 915 | 984 | 964 | 515 | 495 | 4778 |
| المجموع | 1822 | 1870 | 2007 | 1945 | 847 | 807 | 9298 |

***المصدر: وزارة التربية والتعليم/ مديرية التربية والتعليم للواء بني كنانة/ قسم التخطيط للعام/ 2011/2010.**

تم اختيار عينة الدراسة بالطريقة العشوائية العنقودية من مجتمع الدراسة. والجدول رقم (2) يبين توزيع عينة الدراسة حسب متغيرات الجنس والمستوى التعليمي ومستوى القدرة.

**جدول 2**

**توزيع عينة الدراسة حسب متغيرات الجنس والمستوى التعليمي ومستوى القدرة**

| المستوى | الصف | الجنس | | المجموع |
|---|---|---|---|---|
| | | ذكر | أنثى | |
| | | العدد | العدد | |
| متوسط | سابع | 76 | 84 | 160 |
| | ثامن | 95 | 99 | 194 |
| | تاسع | 59 | 87 | 146 |
| | المجموع | 230 | 270 | 500 |
| متقدم | عاشر | 99 | 78 | 177 |
| | حادي عشر | 64 | 101 | 165 |
| | ثاني عشر | 67 | 92 | 159 |
| | المجموع | 230 | 271 | 501 |

### ب. أداة الدراسة:

لقد تم اختيار اختبار أوتيس–لينون للقدرة العقلية بمستوييه المتوسط والمتقدم الصورة "K" المعدلتين للبيئة الأردنية لاستخدامه في هذه الدراسة، وذلك لما يتمتع به هذا الاختبار من دلالات صدق وثبات بدرجة مقبولة.

### صدق أداة الدراسة:

تتوفر لاختبار أوتيس– لينون دلالات صدق وثبات مقبولة تم وصفها في خلفية هذه الدراسة من خلال وصف الاختبار. وقام الباحث بالتحقق من صدق الاختبار من خلال حساب معاملات الارتباط بين درجات المفحوصين على كل نموذج من نماذج الاختبار ومعدلاتهم في المدرسة لكافة المباحث الدراسية حيث بلغ معامل الارتباط هذا (0.79) للمستوى المتوسط للاختبار

و(0.75) للمستوى المتقدم مما يعتبر مؤشراً على الصدق المحكي لنموذجي الاختبار.

**ثبات أداة الدراسة:**

تم التحقق من ثبات مستويي الاختبار من خلال معادلة كودر– ريتشارسون (K-R20) حيث بلغ لنموذج المستوى المتوسط (0.87) وبلغ للمستوى المتقدم (0.79). وهذا مؤشر على الاتساق الداخلي لنموذجي الاختبار. كما قام الباحث بالتحقق من الثبات الأمبريقي Emperica Reliability لكل مستوى من مستويي الاختبار. وبلغ (0.87) للنموذج المتوسط و(0.79) للمتقدم مما يشير إلى توفر دلالات ثبات مقبولة لنموذجي الاختبار المراد معادلتهما. والجدول رقم (3) يبين مؤشرات الثبات لمستويي الاختبار.

**جدول 3**

**معاملات ثبات الاتساق الداخلي والثبات الأمبريقي لأداة الدراسة**

| مستوى الاختبار | KR–20 | الثبات الأمبريقي |
|---|---|---|
| المتوسط | 0.87 | 0.87 |
| المتقدم | 0.79 | 0.79 |

الفقرات المشتركة Common Item: إن اعتماد تصميم الفقرات المشتركة في هذه الدراسة يتطلب تحديداً لهذه الفقرات والتي تعتبر جزءاً من كلا النموذجين المراد معادلتهما. لقد ذكر كولن وبرينان [21] أنه لا يمكن وضع حدود لعدد الفقرات المشتركة خلال عمليات المعادلة. إلا أن انغوف [22] قد ذكر أن العدد المناسب للفقرات المشتركة هو (20%) من العدد الكلي لفقرات الاختبار مهما كان طول الاختبار. واعتبر رايت [22] أن (10%) من الفقرات كافية. وفي هذه الدراسة تم اعتماد الفقرات المشتركة الواردة في نموذجي الاختبار للمستوى المتوسط والمستوى المتقدم وعددها (23) فقرة مشتركة، وتشكل ما نسبته (28%) من كل مستوى من مستويي الاختبار، ويبين الجدول رقم (4) أرقام الفقرات المشتركة بحسب ترتيبها في كل مستوى من مستويات الاختبار.

**جدول 4**

**الفقرات المشتركة وأرقامها كما وردت في مستويي الاختبار**

| مستوى الاختبار | رقم الفقرة | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **متوسط** | 11 | 17 | 20 | 26 | 34 | 37 | 39 | 41 | 48 | 50 | 51 | 52 | 53 | 56 | 57 | 58 | 65 | 67 | 68 | 69 | 72 | 76 | 80 |
| **متقدم** | 30 | 1 | 2 | 22 | 20 | 31 | 12 | 15 | 40 | 33 | 39 | 62 | 37 | 6 | 34 | 32 | 55 | 50 | 54 | 53 | 59 | 70 | 73 |

وقد تم ايجاد الإحصاءات الوصفية لمعلمة الصعوبة والخطأ المعياري للفقرات المشتركة في مستويي الاختبار. والجدول رقم (5) يبين هذه الإحصاءات.

**جدول 5**

**الإحصاءات الوصفية لمعلمة الصعوبة والخطأ المعياري للفقرات المشتركة في مستويي الاختبار**

| الإحصائي | نموذج الاختبار | | | |
|---|---|---|---|---|
| | المتوسط | | المتقدم | |
| | معلمة الصعوبة | الخطأ المعياري لمعلمة الصعوبة | معلمة الصعوبة | الخطأ المعياري لمعلمة الصعوبة |
| **العدد** | 23 | 23 | 23 | 23 |
| **المتوسط** | −0.690 | 0.058 | −0.805 | 0.053 |

**تابع جدول (5)**

| | | | | |
|---|---|---|---|---|
| **الوسيط** | −0.303 | 0.058 | −0.379 | 0.051 |
| **المنوال** | −3.422 | 0.052 | −3.749 | 0.050 |
| **الانحراف المعياري** | 1.172 | 0.009 | 1.267 | 0.009 |
| **التباين** | 1.372 | 0.000 | 1.606 | 0.000 |
| **الإلتواء** | −1.275 | 0.650 | −1.282 | 1.741 |
| **التفلطح** | 0.767 | −0.527 | 0.733 | 2.231 |
| **القيمة الصغرى** | −3.42 | 0.046 | −3.75 | 0.045 |
| **القيمة العظمى** | 1.24 | 0.076 | 1.64 | 0.074 |

ويتبين من الجدول (5) ان المتوسط الحسابي لمعلمة الصعوبة للفقرات المشتركة قد بلغ (−0.69) و(−0.805) في المستوى المتوسط للاختبار والمستوى المتقدم على التوالي، مما يشير لسهولة هذه الفقرات لدى الافراد ذات مستوى القدرة المتقدم مقارنة مع الافراد من القدرة المتوسطة.

تصميم الدراسة:

تم استخدام تصميم الفقرات المشتركة والمجموعات غير المتكافئة Common Items Nonequivalent Group Design في هذه الدراسة، وكما وصفه أنجوف [3] فإن هذا التصميم يتضمن اختبارين ومجموعتين، وتعطى كل مجموعة اختبار أو نموذج واحد يتضمن عدداً من الفقرات المشتركة. ووصفه رايت وستون [23] بأنه اقتصادي ولا حاجة لأن يأخذ المفحوص أكثر من اختبار واحد، بل يتم ربط الاختبارات عبر شبكة من الفقرات المشتركة ويوضح الجدول رقم (9) توزيع أفراد العينة على أداة الدراسة.

**جدول 6**

**توزيع أفراد عينة الدراسة على اداة الدراسة**

| **الاختبار**<br>**أفراد العينة** | **نموذج المستوى المتوسط + الفقرات المشتركة** | **نموذج المستوى المتقدم + الفقرات المشتركة** | **المجموع** |
|---|---|---|---|
| **ذكور** | 230 | 230 | 460 |
| **إناث** | 270 | 271 | 541 |
| **المجموع** | 500 | 501 | 1001 |

إجراءات الكشف عن الأداء التفاضلي للفقرات:

تم الكشف عن الأداء التفاضلي لفقرات أداة الدراسة (نموذج المستوى المتوسط صورة "K" ونموذج المستوى المتقدم "K" لاختبار أوتيس–لينون للقدرة العقلية والمعدلتين للبيئة الأردنية) باستخدام إجراءات طريقة مانتل–هانزل بعد ضبط مستوى القدرة. واعتبرت الإناث هي المجموعة المستهدفة والذكور المجموعة المرجعية.

وقد تم إيجاد نسب الارجحية المشتركة للفقرات باستخدام برمجية spss وللحكم على اتجاه الأداء التفاضلي لصالح أي مجموعة حولت نسب الأرجحية المشتركة إلى لوغاريتم طبيعي (LNMH) حول الصفر، حيث تشير القيمة السالبة للوغاريتم الطبيعي للأرجحية المشتركة المقابلة للفقرة إلى أن الفقرة تظهر أداء تفاضلياً لصالح المجموعة المستهدفة (الإناث) أي انهم يجدون الفقرة أسهل. والقيمة الموجبة تشير إلى أن الفقرة تظهر أداء تفاضلياً لصالح المجموعة المرجعية (الذكور) [24].

إجراءات المعادلة العمودية:

1. المعادلة العمودية لمستويي الاختبار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي:

لقد تم استخدام برنامج BILOG-MG في إجراء المعادلة العمودية لمستويي اختبار أوتيس– لينون للقدرة العقلية، حيث استخدمت طريقة المعادلة العمودية بأسلوب المعايرة بالتلازم Concurrent Calibration، بحيث يتم وضع مستويي الاختبار معاً على مقياس مشترك موحد باستخدام الفقرات المرساوية (المشتركة) ووفقاً للنموذج أحادي المعلمة. فقد تم استخدام طريقة بيز (Bayes) في تقدير معالم الفقرات، كما استخدمت طريقة بيز في تقدير قدرات الأفراد وهي تقديرات متوقعة بعدية Expected A posterior (EAP) كون هذه الطريقة تصلح لجميع أنماط الاستجابة ولها أقل متوسط خطأ بالنسبة لباقي الطرق الأخرى المتوفرة في برنامج BILOG-MG [25].

1. وتعمل طريقة المعادلة بالتلازم في المعايرة على وضع مستويي الاختبار بشكل تلقائي على مقياس موحد دون الحاجة لإجراءات ربط مستقلة كما هو الحال في طريقة المعادلة بالمعايرة المنفصلة Separate Calibration حيث تحتاج لإيجاد ثوابت المعادلة من خلال الفقرات المشتركة. وقد أكد هانسون وبيغون [26] فعالية هذه الطريقة حيث أعطت أقل خطأ معياري، كما أن استخدام المعايرة بالتلازم بواسطة BILOG-MG يتم فيها تقدير معالم الفقرات المشتركة على عينة أكبر.

2. المعادلة العمودية لمستويي الاختبار في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي:

تم إتباع نفس الخطوات السابقة لإجراء المعادلة العمودية بين مستويي الاختبار بعد حذف الفقرات المرساوية ذات الأداء التفاضلي من مستويي الاختبار.

إجراءات التحقق من فعّالية المعادلة العمودية:

تم اعتماد معامل الصدق التقاطعي كمؤشر على دقة المعادلة، حيث تم اختيار عينتين مستقلتين بشكل عشوائي من العينة الكلية بواقع 200 فرد لكل عينة من كل من مستويي الاختبار، ثم طبقت إجراءات المعادلة العمودية السابقة لمعادلة العينتين في ظل الاحتفاظ بالفقرات المرساوية ذات الاداء التفاضلي، ثم حُسب معامل الصدق التقاطعي للمعادلة باستخدام المعادلة 17. وطبقت نفس الإجراءات أيضاً على عينتين مستقلتين اختيرتا بشكل عشوائي بواقع 200 فرد لكل عينة من مستويي الاختبار. وبعد المعادلة على العينتين تم حساب معامل الصدق التقاطعي بعد حذف جميع الفقرات المرساوية ذات الأداء التفاضلي. وبذلك تصبح لدينا القدرة على مقارنة مدى تحقق فعالية المعادلة في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي وفي ظل حذف الفقرات المرساوية ذات الأداء التفاضلي.

المعالجات الإحصائية:

أولاً: المعالجات الإحصائية المتعلقة بفحص افتراضات نظرية استجابة الفقرة:

1. فحص أحادية البعد Unidimensionality: تم التحقق من افتراض أحادية البعد لمستويي الاختبار بإيجاد مؤشر تاناكا index Tanaka لحسن المطابقة، وذلك باستخدام برمجية NOHARM.فإذا كانت قيمة هذا المؤشر أكبر من 0.8 فإن أحادية البعد تعتبر متحققة [27].

2. التحقق من افتراض الاستقلال الموضعي Local independence: تم التحقق من الاستقلال الموضعي لمستويي الاختبار بحساب الإحصائي الموضعي (Q3) الذي يمثل معامل الارتباط للبواقي لأزواج الفقرات بعد ضبط السمة المقدرة، وبالاعتماد على برنامج LDID يتحقق الاستقلال الموضعي للاختبار عندما تكون القيمة المطلقة للإحصائي Q3 أقل من (0.05) او وقوعه ضمن فترة الثقة الخاصة به كما تم فحص مصفوفة معاملات الارتباط للبواقي بين أزواج الفقرات والناتجة عن برمجية NOHARM للتحقق من الاستقلال الموضعي للفقرات.

ثانياً: المعالجات الإحصائية المتعلقة بمطابقة البيانات للنموذج:

1. مطابقة الافراد للنموذج: تم مطابقة الأفراد لنموذج راش من خلال برنامج BILoG–MG واعتماداً على مؤشر كاي تربيع.

2. مطابقة الفقرات للنموذج: تم مطابقة فقرات مستويي الاختبار لنموذج راش من خلال برنامج BILoG-MG واعتماداً مؤشر كاي تربيع للفقرات.

ثالثاً: المعالجات الإحصائية المتعلقة بالأداء التفاضلي:

تم استخراج قيم إحصائي مانتل هانزل باستخدام برمجية SPSS للكشف عن الأداء التفاضلي للفقرات وفقاً لمتغير الجنس، حيث تظهر الفقرة أداء تفاضلياً إذا كان مستوى الدلالة الإحصائية المقابلة لها أقل من أو تساوي (0.05). كما تم إيجاد نسبة الأرجحية المشتركة والتي حولت إلى اللوغاريتم الطبيعي لتسهيل تفسيرها حيث تبدي الفقرة أداء تفاضلياً لصالح المجموعة المرجعية إذا كانت قيمت اللوغاريتم الطبيعي لنسبة الأرجحية المشتركة المقابلة لها موجبة بينما إذا كانت سالبة فإن الفقرة تبدي أداء تفاضلياً للمجموعة المستهدفة.

رابعاً: المعالجات الإحصائية المتعلقة بالمعادلة العمودية:

حساب محك الصدق التقاطعي Cross – Validation:

وهو متوسط مربعات الانحرافات للعلامات المتعادلة عندما تطبق على عينات مختلفة لمعرفة مدى استقرار النتائج ويحسب من المعادلة (17). وكلما كانت قيمة معامل الصدق التقاطعي أقل من واحد صحيح كانت المعادلة أكثر دقة.

## 5. النتائج

أولاً: النتائج المتعلقة بالتحقق من افتراضات نظرية استجابة الفقرة:

1. النتائج المتعلقة بالتحقق من افتراض أحادية البعد (Unidimensionality):

مع أن اختبار أوتيس – لينون يقيس قدرة عامة وهي القدرة العقلية إلا أن الباحث قام بالتحقق من أحادية البعد بالاعتماد على مؤشر تاناكا Tanaka لحسن المطابقة وذلك باستخدام برمجية NOHARM، وتعتبر أحادية البعد متحققة إذا كانت قيمة مؤشر تاناكا أكبر من 0.8 [27]. وقد وجدت قيمة مؤشر تاناكا لنموذج الاختبار المتوسط 0.969 ولنموذج الاختبار المتقدم 0.806 وعليه فإن نموذجي الاختبار يقيس كل منهما سمة واحدة رئيسة وتتحقق فيها أحادية البعد.

2. النتائج المتعلقة بالتحقق من افتراض الاستقلال الموضعي (Local independence).

أشار هامبلتون وسوامينثان [7] إلى أن الاستقلال الموضعي مكافئ لأحادية البعد ويعتبر متحقق بتحققه، إلا أن آخرون اعتبروا أن الاستقلال الموضعي لا يكافئ أحادية البعد [28]. ولذلك فقد تم التحقق من افتراض الاستقلال الموضعي اعتماداً على الإحصائي الوصفي(Q3) الذي يمثل معامل الارتباط للبواقي لأزواج الفقرات، وأشار لي [29] إلى أن Q3 قد أثبت فاعليته في الكشف عن الارتباط الموضعي للفقرات. وكما يشير الجدول رقم (10) ادناه فإن قيمة الوسط الحسابي للإحصائي Q3 للنموذج المتوسط (0.0130) وللنموذج المتقدم (0.0129) وكلاهما يقع ضمن فترة الثقة الخاصة به وهذا مؤشر على تحقق الاستقلال الموضعي لكلا النموذجين من الاختبار.

كما تم التحقق من الاستقلال الموضعي لنموذجي الاختبار من خلال فحص مصفوفة معاملات الارتباط للبواقي بين أزواج الفقرات والناتجة عن برمجية NOAARM، حيث وجدت هذه القيم صغيرة وقريبة من الصفر مما يشير لتحقق افتراض الاستقلال الموضعي.

3. النتائج المتعلقة بمطابقة الأفراد للنموذج:

لقد، حيث استخدام برنامج BILOG-MG لتحليل البيانات لكل نموذج بشكل مستقل وبالاعتماد على عينة الدراسة أظهرت النتائج عدم مطابقة فرداً واحداً من أفراد عينة نموذج الاختبار المتوسط والبالغ عددهم 500 فرداً وكان رقمه (24) وذلك اعتماداً على مؤشر كاي تربيع عند مستوى الدلالة $\alpha = 0.01$)، حيث كان مقدار الاحتمالية لهذا الفرد أقل من (0.01) وبذلك استقر

عدد الافراد المستجيبين على المستوى المتوسط عند (499) فردا مطابقا للنموذج احادي المعلمة.

وأظهرت نتائج المطابقة لمستوى الاختبار المتقدم على عينة الدراسة البالغة 501 فرداً مطابقة جميع الأفراد للنموذج.

4. النتائج المتعلقة بمطابقة الفقرات للنموذج:

بعد حذف استجابة الفرد غير المطابقة للنموذج أعيد التحليل باستخدام برنامج BILOG- MG، لمطابقة الفقرات مع النموذج المستخدم وأظهرت النتائج عدم مطابقة الفقرات ذات الأرقام 12، 29، 60، 79 في النموذج المتوسط وذلك اعتماداً على مؤشر كاي تربيع عند مستوى الدلالة الإحصائية ($\alpha = 0.01$)، حيث كانت الاحتمالية لهذه الفقرات أقل من 0.01، أما بالنسبة لمستوى الاختبار المتقدم فقد أظهرت النتائج عدم مطابقة الفقرات رقم 35، 41، 46، 48، وذلك اعتماداً على مؤشر كاي تربيع عند مستوى الدلالة ($\alpha= 0.01$)، وبذلك فقد استقرت عدد الفقرات لاختبار المستوى المتوسط عند (76) فقرة مطابقة للنموذج احادي المعلمة وكذلك للمستوى المتقدم واستقرت الفقرات المشتركة عند (23) فقرة مطابقة للنموذج.

ثانيا: النتائج المتعلقة بالإجابة عن السؤال الاول: ما مدى تحقق معايير الفاعلية للمعادلة العمودية لمستويي الاختبار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي للجنس؟

من أجل التحقق من فاعلية المعادلة العمودية لمستويي الاختبار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي للجنس تم إجراء معادلة عمودية لمستويي الاختبار بوجود تلك الفقرات والتي تم الكشف عنها بطريقة مانتل هانزل. وقد أتبعت خطوات المعادلة العمودية بالمعايرة بالتلازم باستخدام برنامج BILOG – MG والجدول (7) يبين تقديرات القدرة والخطأ المعياري في قياسها قبل وبعد المعادلة لعينة من درجات القدرة لمستويي الاختبار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي.

**جدول 7**

**تقديرات القدرة والخطأ المعياري لمستويي الاختبار قبل وبعد المعادلة في ظل الاحتفاظ بالفقرات ذات الأداء التفاضلي**

| النموذج المتوسط | | المقياس المشترك | | النموذج المتقدم | |
|---|---|---|---|---|---|
| القدرة | الخطأ المعياري | القدرة | الخطأ المعياري | القدرة | الخطأ المعياري |
| −1.340 | 0.3 | −1.9 | 0.34 | −2.170 | 0.36 |
| −1.342 | 0.3 | −1.8 | 0.29 | −2.050 | 0.35 |
| −1.170 | 0.29 | −1.7 | 0.39 | −1.820 | 0.33 |
| −1.171 | 0.29 | −1.6 | 0.47 | −1.710 | 0.32 |
| −1.010 | 0.28 | −1.5 | 0.57 | −1.610 | 0.32 |
| −1.011 | 0.28 | −1.3 | 0.64 | −2.170 | 0.36 |
| −0.930 | 0.28 | −1.1 | 0.64 | −1.520 | 0.31 |
| −0.780 | 0.27 | −0.8 | 0.53 | −1.330 | 0.3 |
| −0.781 | 0.27 | −0.7 | 0.45 | −1.240 | 0.3 |
| −0.700 | 0.27 | −0.6 | 0.31 | −1.160 | 0.29 |
| −0.350 | 0.26 | −0.5 | 0.17 | −0.910 | 0.28 |
| −0.220 | 0.26 | −0.4 | 0.39 | −0.830 | 0.28 |

| النموذج المتوسط | | المقياس المشترك | | النموذج المتقدم | |
|---|---|---|---|---|---|
| القدرة | الخطأ المعياري | القدرة | الخطأ المعياري | القدرة | الخطأ المعياري |
| −0.150 | 0.26 | −0.3 | 0.49 | −0.460 | 0.27 |
| −0.153 | 0.26 | −0.2 | 0.53 | −0.600 | 0.27 |
| −0.090 | 0.26 | −0.1 | 0.59 | −0.530 | 0.27 |
| −0.020 | 0.26 | 0 | 0.63 | −0.530 | 0.27 |
| −0.022 | 0.26 | 0.1 | 0.65 | −0.460 | 0.27 |
| 0.050 | 0.26 | 0.2 | 0.63 | −0.380 | 0.27 |
| 0.053 | 0.26 | 0.3 | 0.61 | −0.310 | 0.27 |
| 0.110 | 0.26 | 0.4 | 0.55 | −0.460 | 0.27 |
| 0.111 | 0.26 | 0.5 | 0.47 | −0.310 | 0.27 |
| 0.180 | 0.26 | 0.6 | 0.38 | −0.310 | 0.27 |
| 0.240 | 0.26 | 0.7 | 0.17 | −0.100 | 0.27 |
| 0.710 | 0.26 | 0.8 | 0.34 | 0.250 | 0.26 |
| 0.780 | 0.26 | 0.9 | 0.45 | 0.320 | 0.26 |
| 0.782 | 0.26 | 1 | 0.54 | 0.390 | 0.27 |
| 0.784 | 0.26 | 1.1 | 0.6 | 0.390 | 0.27 |
| 0.850 | 0.27 | 1.2 | 0.64 | 0.460 | 0.27 |
| 0.856 | 0.27 | 1.3 | 0.64 | 0.460 | 0.27 |
| 0.920 | 0.27 | 1.4 | 0.64 | 0.530 | 0.27 |
| 0.922 | 0.27 | 1.5 | 0.62 | 0.460 | 0.27 |
| 0.924 | 0.27 | 1.6 | 0.6 | 0.600 | 0.27 |
| 0.990 | 0.27 | 1.7 | 0.52 | 0.530 | 0.27 |
| 1.070 | 0.27 | 1.8 | 0.46 | 0.670 | 0.27 |
| 1.220 | 0.28 | 1.9 | 0.35 | 0.750 | 0.27 |
| 1.300 | 0.28 | 2 | 0.2 | 0.890 | 0.27 |
| 1.470 | 0.29 | 2.1 | 0.35 | 1.120 | 0.28 |

وللتحقق من فاعلية المعادلة تم إيجاد معامل الصدق التقاطعي، حيث تم اختيار عينيتين مستقلتين عشوائياً من مستويي الاختبار وأجري لهما معادلة عمودية في ظل وجود جميع الفقرات ذات الأداء التفاضلي ثم تم حساب معامل الصدق التقاطعي بناءاً على درجات القدرة المتعادلة من العينتين وقد كانت قيمة معامل الصدق التقاطعي في ظل الاحتفاظ بالفقرات ذات الأداء التفاضلي 0.00052.

ثالثا: النتائج المتعلقة بالإجابة عن السؤال الثاني: ما مدى تحقق معايير الفاعلية للمعادلة العمودية لمستويي الاختبار في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس؟

للإجابة على هذا السؤال تم إجراء معادلة عمودية لمستويي الاختبار بعد حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس من مستويي الاختبار والجدول رقم (8) يبين تقديرات القدرة والخطأ المعياري في قياسها قبل وبعد المعادلة لعينة من درجات القدرة في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس.

**جدول 8**

**تقديرات القدرة والخطأ المعياري قبل وبعد المعادلة لعينة من درجات القدرة في ظل حذف الفقرات ذات الأداء التفاضلي**

| النموذج المتوسط | | المقياس المشترك | | النموذج المتقدم | |
|---|---|---|---|---|---|
| القدرة | الخطأ المعياري | القدرة | الخطأ المعياري | القدرة | الخطأ المعياري |
| −1.573 | 0.353 | −2.2 | 0.538 | −2.703 | 0.652 |
| −1.448 | 0.354 | −2 | 0.529 | −2.29 | 0.611 |
| −0.967 | 0.336 | −1.3 | 0.494 | −1.616 | 0.372 |
| −0.856 | 0.33 | −1.1 | 0.486 | −1.487 | 0.349 |
| −0.749 | 0.322 | −0.9 | 0.479 | −1.369 | 0.339 |
| −0.463 | 0.3 | −0.5 | 0.463 | −1.147 | 0.327 |
| −0.373 | 0.302 | −0.4 | 0.46 | −1.042 | 0.322 |
| −0.28 | 0.305 | −0.2 | 0.458 | −0.939 | 0.322 |
| −0.187 | 0.305 | −0.1 | 0.456 | −0.835 | 0.324 |
| −0.095 | 0.302 | 0 | 0.456 | −0.73 | 0.325 |
| −0.004 | 0.301 | 0.1 | 0.455 | −0.625 | 0.323 |
| 0.087 | 0.303 | 0.2 | 0.456 | −0.521 | 0.322 |
| 0.18 | 0.309 | 0.4 | 0.456 | −0.417 | 0.323 |
| 0.277 | 0.313 | 0.5 | 0.457 | −0.313 | 0.324 |
| 0.473 | 0.313 | 0.8 | 0.459 | −0.104 | 0.321 |
| 0.571 | 0.313 | 0.9 | 0.46 | −0.002 | 0.318 |
| 0.67 | 0.315 | 1.1 | 0.462 | 0.201 | 0.319 |
| 0.868 | 0.313 | 1.4 | 0.466 | 0.401 | 0.312 |
| 0.965 | 0.31 | 1.5 | 0.469 | 0.497 | 0.31 |
| 1.061 | 0.311 | 1.7 | 0.473 | 0.593 | 0.311 |
| 1.159 | 0.315 | 1.8 | 0.478 | 0.691 | 0.313 |
| 1.364 | 0.323 | 2.1 | 0.491 | 0.888 | 0.312 |
| 1.469 | 0.326 | 2.3 | 0.5 | 1.081 | 0.312 |
| 1.577 | 0.332 | 2.5 | 0.51 | 1.18 | 0.317 |

وللتحقق من فاعلية المعادلة تم إيجاد معامل الصدق التقاطعي، حيث تم اختيار عينتين مستقلتين عشوائيا من كل من مستويي الاختبار وطبقت إجراءات المعادلة العمودية في ظل حذف جميع الفقرات ذات الأداء التفاضلي ثم تم حساب معامل الصدق التقاطعي بناءاً على درجات القدرة المتعادلة من العينتين وقد وجد يساوي 0.00033.

**جدول 9**

**قيم معاملات الصدق التقاطعي للمعادلة في ظل حالتي الإبقاء، والحذف للفقرات المرساوية ذات الأداء التفاضلي للجنس**

| الطريقة | معامل الصدق التقاطعي للمعادلة العمودية |
|---|---|
| في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي | 0.00060 |
| في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي | 0.00033 |

## 6. مناقشة النتائج

1. مناقشة النتائج المتعلقة بالإجابة عن السؤال الاول:

ما مدى تحقق معايير الفاعلية للمعادلة العمودية لمستويي الاختبار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي للجنس؟

للإجابة عن السؤال قام الباحث بإجراء معادلة عمودية بين مستويي الاختبار المتوسط والمتقدم باستخدام برنامج BILOG-MG في ظل الاحتفاظ بجميع الفقرات المرساوية ذات الأداء التفاضلي للجنس، وللتحقق من فعالية المعادلة تم إيجاد معامل الصدق التقاطعي من خلال عينيتين عشوائيتين طبقت عليها إجراءات المعادلة في ظل الاحتفاظ بجميع الفقرات المرساوية ذات الأداء التفاضلي للجنس، وبلغت قيمته (0.00060) والتي تشير إلى توفر الفاعلية للمعادلة العمودية لمستويي الاختبار وفقاً للنموذج أحادي المعلمة بدرجة مقبولة في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي، وذلك اعتمادا على مؤشر معامل الصدق التقاطعي، فكلما قلت قيمة هذا المؤشر وكانت أقل من واحد زادت دقة المعادلة.

2. مناقشة النتائج المتعلقة بالإجابة عن السؤال الثاني:

ما مدى تحقق معايير الفاعلية للمعادلة العمودية لمستويي الاختبار في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس؟

للإجابة عن هذا السؤال تم إجراء معادلة عمودية بين مستويي الاختبار المتوسط والمتقدم باستخدام برنامج BILOG-MG في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس قبل إجراء المعادلة، وللحكم على مدى تحقق فاعلية المعادلة تم إيجاد معامل الصدق التقاطعي من خلال عينيتين عشوائيتين طبقت عليهما إجراءات المعادلة العمودية في ظل حذف الفقرات ذات الأداء التفاضلي، وبلغت قيمته (0.00033)، ويشير إلى توفر الفاعلية للمعادلة العمودية لمستويي الاختبار المتوسط والمتقدم في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس ووفقاً للنموذج أحادي المعلمة. وبمقارنة قيم معامل الصدق التقاطعي الناتجة عن إجراء المعادلة العمودية لمستويي الاختيار في ظل الاحتفاظ بالفقرات المرساوية ذات الأداء التفاضلي للجنس مع معامل الصدق التقاطعي الناتج عن المعادلة في ظل حذف الفقرات المرساوية ذات الأداء التفاضلي للجنس، نلاحظ أن قيمة هذا المؤشر قد قلت عند حذف الفقرات ذات الأداء التفاضلي للجنس مما يدل على زيادة دقة المعادلة العمودية لمستويي الاختيار عند استبعاد الفقرات المرساوية ذات الأداء التفاضلي من عملية المعادلة، ويتفق ذلك مع ما ذكره هيدالغو و لوبير [30] من أن الفقرات ذات الأداء التفاضلي قد تلوث المعادلة العمودية وتؤثر فيها، ومن المهم تحديد هذه الفقرات وتقليل أثرها في المعادلة. كما تشير قيم الصدق التقاطعي إلى فعالية نموذج راش في إجراء المعادلة العمودية من خلال استخدام طريقة المعادلة بالتلازم. ويمكن القول بأن وجود فقرات مرساوية ذات أداء تفاضلي يؤثر في دقة المعادلة العمودية لمستويات الاختبار، حيث تتحسن درجة دقة المعادلة بحذف الفقرات ذات الأداء التفاضلي ولذلك ينبغي اخذ الأداء التفاضلي للفقرات المرساوية بعين الاعتبار عند إجراء المعادلة لدرجات الاختبارات وذلك لتقليل اثر الأداء التفاضلي على فاعلية المعادلة.

## 7. التوصيات:

يوصي الباحث بما يلي :

1. زيادة الوعي حول خصائص الفقرات المرساوية المستخدمة في عملية معادلة الاختبارات ومنها الأداء التفاضلي لما لذلك من اثر في دقة المعادلة العمودية لمستويات الاختبارات.
2. إجراء دراسة مماثلة تأخذ بعين الاعتبار اختلاف طريقة المعادلة واختلاف النموذج المستخدم، فقد تتأثر دقة المعادلة بالنموذج المستخدم أو بالطريقة المستخدمة في المعادلة.
3. إجراء دراسة مماثلة تأخذ بعين الاعتبار متغيرات أخرى قد تبدي الفقرات أداء تفاضلياً وفقاً لها غير متغير الجنس كالمنطقة الجغرافية والطبقات الاجتماعية والاقتصادية.

## المراجع

**أ. المراجع العربية:**

[15] الحواري، أروى عيسى. (2007). الخصائص السيكومترية لصور مختارة من اختبارات الرخصة الدولية لقيادة الحاسوب في الأردن، ومعادلة درجاتها، رسالة ماجستير، جامعة اليرموك، 2007، الأردن.

**ب. المراجع الأجنبية:**

[1] Angoff, W.H. (1986). Some contributions of the College Board SAT to Psychometric theory and Practice. Educational Measurement: Issues and Practice, 5 (3), 7-11.

[2] American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1985). Standards for educational and Psychological testing. Washington, D.C.: Author.

[3] Angoff, W.H. (1971). Scales, norms and equivalent scores. in R.L. Thorndike (Ed.), Educational Measurement (pp508-600). Washington, D.C: American Council on education.

[4] Slind, J.A., & Linn, R.L. (1977). Vertically equating tests: Facts or phantom? Journal of Educational measurement, 14, 23 -32.

[5] Lord, F.M. (1975). Asurvey of equating methods based on item characteristic theory. Research Bulletin 75-13. princeton, NJ: Educational Testing service.

[6] Cook, L.L. & Eignor, D.R. (1983). IRT Equating Methods, Educational measurement: Issues and practice 11(4), 37-45.

[7] Hambleton, R. & Swaminathan, H. (1985). Item response theory: Principles and application. Boston, MA: Kluwer, NIJ Publishing.

[8] Kolen, M.J (1981). Comparison of Traditional and Item Response Theory Methods for Equating Tests. Journal of Educational Measurement 18: 1-11.

[9] Harris, A.M & Crous, J.D.(1993).Astudy of criteria used in equating. Applied Meatusement in education,6,195-240.

[10] Gierl, M., Jodoin, M., & Teresi, T. (2001). Performance of Mantel-Haenszel, simultaneous item bias test ,and logistic regression when the proportion of DIF items is large. paper presented at the annual meeting of American Educational Research Association. Orleans, Louisiana, April, 24-27, 2000.

[11] Ellis, B. & Raju, S. (2003). Test and item bias: What aren't and How to detect them measuring up. Assessment issues for teachers, counselors and administrator, (ERC Document Reproduction service No. ED 480042).

[12] NandaKumar, R. (1993). Simultaneous DIF amplification and Cancellation: Shealy – Stout's test for DIF. Journal of Educational Measurement, 1973-90.

[13] Harris, A.M. & Carlton, S.T. (1995). Pattern of gender differences on mathematics items on the scholastic Aptitude. Applied Measurement in Education, 6 (2), 137 – 150.

[14] Huijuan, D. (2007). Acomparison study of IRT Calibration Methods for Mixed-Format Tests In Vertical Scaling ,Doctoral Thesis, Iwa university. Proquest information and Learning company: UML: 3383534.

[16] Sinharay, S., & Holland, P. (2007). The correlation Between the Scores of Atest, Research Report, Ets, March.

[17] Uiterwijk, H., & Vallen, T. (2005). Linguistic sources of item bias for second generation immigrants in dutch test. Language Testing, 22(2), 211-234.

[18] MendesBarnet, S., & Ercikan, K. (2006). Examining sources of gender DIF mathematics using a confirmatory multidimensional model approach [Electronic version]. Applied Measurement in Education, 19(4), 289-304.

[19] Driana, E. (2007). Gender Differential Item Functioning on aninth – grade Mathematics Proficiency test in application OHIO. Doctoral Thesis, Ohio University, Proquest information and Learning Company, UM: 3280034.

[20] Skaggs, G. & Lissitz, R. (1986). IRT Test Equating Relevant Issues and a Review of Recent Research. Review of Educational Research, 56 (4), 495-529.

[21] Kolen, M. & Brennan, R. (1995). test equating: methods and practices. N. Y.: spring

[22] Wright, B.D. (1997). Solving Measurement Proplems with the Rsch Model. Educational measurement,14,97-116.

[23] Wright, D. & Stone, M., (1979). Best Test Design, A Handbook for Resh measurement, Chicago: MESA Press.

[24] Holland, P.W., & Thayer, D.T.(1988). Differential item performance and the mantel – Haenszel Procedure. In wainer, H., & Braun, H.I. (eds.), test validity (pp.129-145) Hillsade, NJ: Lawrence Erlbaum.

[25] Zimowski, M., Muraki, E. Mislevy, R., Robert, J.; & Bock, D. (1996). BILOG-MG3: Multiple group IRT analysis and test maintenance for binary items (computer program). chicago: scientific software.17.no.1

[26] Hanson, B.A., & Beguin, A.A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. Applied Psychological Measurement. 26, 3-24.

[27] Clarke, D. Mackinnon, F. Mckenzie, F. Herrman, H.(2000). Dimentions of psychopathology in the medical I11 ALatent Trait Analysis. psychosomatics, 41:5

[28] Meara,k., Robin, F. & Sireci, S. (2000). Using multidimensional scaling to Asses the Dimensionality of Dichotomous Item Data. Multivariate Behavioral Research, 35 (2), 229-259.

[29] Lee, Y. (2004). Examining Passage – Related Local Item dependence (LID) and Measurement construct Using Q3 Statistics in an EFL reading comprehension test. Language testing, 21 (1), 74 – 100.

[30] Hidalgo – Montesinos, M.D, & Lopez – pina, J.A. (2002). Two Stage Equating in Differential Item Functioning Detection Under the Graded Response and the Lord Statistic. Educational and Psychological Measurement, 62 (1), 32.

# THE EFFECT OF ANCHOR ITEMS DIFFERENTIAL FUNCTIONING ON THE ACCURACY OF OTIS LENNON MENTAL ABILITY TEST VERTICAL EQUATING

**Rateb Sayel ALkider ALRuhail**

**Faculty of Education,**

**Abstract-**This study aimed to investigate the effect of anchor items differential functioning (DIF) du to gender on the accuracy of Otis Lennon Mental Ability Test Vertical Equating. For this Purpose a Jordanian version of Otis – Lennon Mental Ability Test (OLMAT) intermediate level and Advance levels form "K" was used. The sample consisted of (1001) students (500) of intermediate level and (501) of advance level of ability. Mantel – Haenzel procedure used for determined the DIF items vertical equating of two test levels conducted using common items – non equivalent groups design by using BILOG- MG Program according to item response one parameter model. Vertical equating of two test level conducted in two cases; first: when the DIF anchor items included in two test levels and second when DIF anchor items excluded from two test levels. Ability scores used for equating. For the purpose of comparing the two equating cases cross – validation coefficient was used as a criterion for studying the extent of stability of the results for equated scores and as the accuracy index of vertical equating. The results indicated that the DIF items affect the accuracy of equating, so that, when DIF anchor items excluded from two test forms to be equated, the accuracy of equating increased depending on cross – validation coefficient.

**Keywords:** Item response theory, one parameter model, differential Items function, Vertical equating, Cross-validation.