المجلة الأردنية في العلوم التربوية، مجلد 13، عدد 4، 2017، 389 - 403

أثر أسلوب التعامل مع القيم الشاذّة في فاعلية معادلة نموذجي اختبار

محمد إبراهيم * يوسف سوالمة**

تاريخ تسلم البحث 2017/2/9 تاريخ قبوله 2017/7/13

ملخص: هدفت الدراسة إلى الكشف عن فاعلية مُعادلة نموذجي اختبار حسب نظرية الاستجابة للفقرة باختلاف أسلوب التعامل مع القيم الشاذة. ولتحقيق هدف الدراسة، تمَ استخدام الطريقة الخطية لمُعادلة نموذجي الاختبار الوطني لضبط نوعية التعليم لمبحث الرياضيات لطلبة الصف الرابع الأساسي بالاعتماد على القدرة (θ) وفق تصميم المجموعات العشوائية في المُعادلة. وتكوّن النموذج الأول من (35) فقرة من نوع الاختيار من متعدد، وتكوّن النموذج الثاني من (36) فقرة من نوع الاختيار من متعدد، وقد اختيرت عشوائيًا استجابات (2000) طالبًا وطالبة لكل نموذج. وتمَ استخدام محكي الخطأ المعياري للمُعادلة والصدق التقاطعي للحكم على فاعلية المُعادلة.

وأظهرت النتائج بأنَ استبدال القيم الشاذّة يؤدي إلى زيادة فاعلية مُعادلة الاختبارات. كما كانت هناك فروق جوهرية بين الأوساط الحسابية للأخطاء المعيارية للقدرات المُعادلة لصالح استبدال القيم الشاذّة عند جميع مستويات القدرة مقارنةً بكل من الاحتفاظ بالقيم الشاذّة وحذفها. كما أظهرت النتائج بأنَ المُعادلة تكون أكثر فاعلية كلما اتجه مستوى القدرة نحو وسط التوزيع.

(**الكلمات المفتاحية:** مُعادلة الاختبارات، القيم الشاذّة، الصدق التقاطعي، الخطأ المعياري للمُعادلة).

مقدمة: تُعَدُّ الاختبارات التربوية من أهم الأدوات المُستخدمة في العملية التعلمية التعليمية؛ إذ غالبًا ما يتم استخدام درجات المفحوصين على تلك الاختبارات من المتخصصين التربويين لتزويدهم بمعلومات تساعدهم في عملية اتخاذ قرارات تربوية متعلقة بتقويم التطور التربوي للمفحوصين، كتحويلهم إلى برامجَ خاصّةٍ سواءً أكانت إثرائية أم علاجية وبالاختيار، والنقل، والترفيع (Hu, 2004). ويواجه المتخصصون في الاختبارات العديد من المواقف التي تتطلب اختبار مجموعات مختلفة من المفحوصين باستخدام نماذج مختلفة من الاختبار كاختبارات القَبول في الجامعة واختبارات تحديد المستوى واختبارات مزاولة المهن (Kolen & Brennan, 2004). ويفترض في هذه النماذج أنها تقيس السِّمةَ نفسهَا. وهذا يتطلب من القائمين على تلك الاختبارات أن يكونوا حريصين على إعداد نماذج اختبارات متكافئة. وهذا في الواقع أمر غير قابل للتحقيق، إذ من الصعب تطوير اختبارات متكافئة في المحتوى أو الصعوبة أو الثبات، وبالتالي يصبح من غير العدل تطبيق تلك الاختبارات على مجموعات مختلفة من المفحوصين (Jimenez, 2011).

لذلك ظهرت الحاجة إلى مُعادلة النماذج المختلفة للاختبارات، والتي تهدف بشكل أساسي إلى جعل الدرجات على النماذج المختلفة متكافئة وقابلةً للمقارنة، من خلال تعديل الفروق الناتجة من معاملات الصعوبة بين تلك النماذج. وعندها تصبح المقارنة ممكنة بين الدرجات التي يحصل عليها المفحوصون على النماذج المختلفة، وتكون ذات معنى بغض النظر عن نموذج الاختبار الذي طُبِّق عليهم (Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004).

وهناك نوعان من المُعادلة: المُعادلة الأفقية، والمُعادلة العمودية. وتفترض المعادلة الأفقية أن صور الاختبار المختلفة تقيس السمة نفسَها، وتكون الفقرات للصور المختلفة متقاربةً في صعوبتها، وتُطبّق على مجموعات متقاربة في مستوى قدرتها. وبالتالي يكون الهدف منها تعديل الفروق الناتجة عن صعوبة الفقرات على صورتي الاختبار. وتفترض المعادلة العمودية أن صور الاختبار المختلفة تقيس السمة نفسَها، وتكون متباينةً في صعوبتها، وتطبق على مجموعات متباينة في مستوى قدرتها. أي أن السمة تقع على متصل واسع المدى يحتاج إلى عدة اختبارات لتغطيته. ويُفترَضُ في هذه الاختبارات أن تقيس السمة نفسَها، بحيث يغطي كل اختبار جزءاً من المتصل. ونظراً لأنه من غير المؤكد أن يكون نظام الوحدات نفسه في جميع الاختبارات، تنشأ الحاجة إلى مُعادلة هذه الاختبارات بحيث تكون العلامة المحققة على أي منها مناظرةً لموقع على متصل السمة بدلالة وحدة القياس نفسِها مهما كان ذلك الموقع (Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004).

The Effect of Outliers Treatment Method on the Effectiveness of Equating Two Test Forms

Mohamad Ibraheem, *Change Agent for Arab Development and Education Reform (CADER), Jordan.*
Yousef Sawalmeh, *Collage of Education, Yarmouk University, Jordan.*

Abstract: This study aimed at revealing the effectiveness of equating two test forms according to item response theory using different outlier treatment methods. To achieve that, the linear equating method was used to equate two equivalent test forms of the National Test to Control the Quality of Education in Mathematics for fourth grade, based on the ability (θ) according to random groups design in equating. The first form consisted of (35) multiple choice items and the second form consisted of (36) multiple choice items. Responses of (2000) students were randomly selected for each form. Standard error of equating and cross-validation were used for judging equating effectiveness.

The results showed that replacement of outliers lead to an increase in tests equating effectiveness. Also, there are significant differences between the means of standard errors for the equated abilities, in favor of replacing the outliers at all ability levels compared with either keeping or removing the outliers. It also revealed that effectiveness of equating is better whenever the ability level tends toward the center of ability distribution.

(**Keywords** Test Equating, Outliers, Cross-Validation, Standard Error of Equating).

* الكادر العربي لتطوير وتحديث التعليم، الأردن.
** كلية التربية، جامعة اليرموك، الأردن.

ويتم جمع البيانات التي تحتاجها عملية مُعادلة الاختبارات (Crocker & Algina, 1986; Kolen & Brennan, 2004)، من خلال استخدام تصميم المجموعة الواحدة، أو تصميم المجموعات المتكافئة أو العشوائية، أو تصميم المجموعات غير المتكافئة للفقرات المشتركة، أو تصميم اختبار الجذع المشترك، أو تصميم المجموعات المتكافئة للفقرات المشتركة، أو تصميم المجموعات العشوائية المتوزانة.

وتُعَدُّ الأخطاء من المصادر الأساسية التي تؤثر على دقة مُعادلة درجات الاختبارات. وقد حدد كولن وبرينان (Kolen & Brennan,2004) نوعين من أخطاء المعادلة للدرجات: الخطأ العشوائي والخطأ المنتظم. ويعزى الخطأ العشوائي إلى أسلوب المعاينة واستخدام العينات المختارة في تقدير الاوساط الحسابية والانحرافات المعيارية، والتي بدورها تستخدم في تقدير دالة المعادلة. ويساعد اختيار عينات كبيرة في ضبط هذا النوع من الخطأ وتخفيضه. ويظهر الخطأ المنتظم نتيجةً لعدم التزام الباحث بشروط المُعادلة أو انتهاك الافتراضات الإحصائية المطلوبة. وعملية ضبطه أصعب من عملية ضبط الخطأ العشوائي.

وتُعَدُّ هذه الأخطاء من أكبر التحديات التي تواجه عملية مُعادلة الاختبارات، وقد تختلف باختلاف تصميم المعادلة وشكلها؛ فلكل تصميم شروطه ولكل شكل من أشكال المعادلة افتراضاته، ومسألة الالتزام بالشروط وتحقيق الافتراضات تختلف باختلاف التصميم وشكل المعادلة. وهذا يستوجب البحث عن أفضل التصاميم والأشكال للمعادلة، بحيث يتم الحصول على أقل الأخطاء، ورفع فاعلية المُعادلة إلى الحد الأقصى.

وقد سهّلت نظرية الاستجابة للفقرة عملية مُعادلة الاختبارات. ومن أهم ميزات هذه النظرية قدرتها على وضع عدّة اختبارات وعدّة مجموعات من المفحوصين على تدريج مشترك. وقد أظهرت الدراسات (الشريفين، 2003؛ Hanson & Lee & Ban, 2010 ;Beguin, 2002; Kim& Kolen, 2006) بأن أساليب المُعادلة حسب نظرية الاستجابة للفقرة أفضل من الأساليب المعتمدة على النظرية التقليدية. وهناك العديد من أشكال المُعادلة للاختبارات حسب نظرية الاستجابة للفقرة ومن أهمها: مُعادلة العلامة الحقيقية True Score Equating، ومُعادلة العلامات الملاحظة/ الخام، ومُعادلة علامة القدرة Ability Score Equating.

وقد تم تناول أثر العديد من المتغيرات على فاعلية مُعادلة الاختبارات مثل الخصائص السيكومترية للفقرات من صعوبةٍ وتمييز(اللمع، 2012)، وطول الاختبار وحجم العينة (الحجيلي، 2012)، والأداء التفاضلي للفقرة(الرحيل،2013؛الحياصات، 2011)، وطرق تصحيح فقرات الصواب-الخطأ المتعدد (الصمادي، 2006)، وطريقة تصفية المموهات في فقرات الاختيار من متعدد (الكوفحي، 2013).

وأظهر الأدب التربوي أنّ مُعادلة الاختبارات قد تتأثر بعامل مهم آخر هو الفقرات المتطرّفة في الفقرات المشتركة (He, 2013; Jimenez, 2011; Michaelides, 2003). فقد هدفت دراسة ماكليدز (Michaelides, 2003) إلى فحص تأثير الفقرات المشتركة التي تسلك سلوكا مختلفا عن بقية الفقرات (Misbehaving common items) على عملية المُعادلة بالاعتماد على نظرية الاستجابة للفقرة (IRT-based equating). وأشارت النتائج إلى أنّ الفقرات المتطرّفة عندما تكون سهلةً جدًا أو صعبةً جدًا تؤثر على إحصاءات المُعادلة بغض النظر عن أسلوب المُعادلة.

كما قام هيو وروجرز وفوكميروفيك (Hu, Rogers & Vukmirovic, 2008) بدراسةٍ هدفت إلى مقارنة فاعلية المُعادلة في حالتي حذف الفقرات المتطرّفة أو الاحتفاظ بها. وقد أشارت نتائج الدراسة إلى عدم إختلاف فاعلية المُعادلة بينهما في حالة المجموعات المتكافئة. ولكن في المجموعات غير المتكافئة، كانت فاعلية المُعادلة أفضل عندما يتم حذف الفقرات المتطرّفة.

وهدفت دراسة جيمنيز (Jimenez, 2011) إلى فحص أثر الفقرات المشتركة ذات التقديرات المتطرّفة لمعالم الصعوبة ومعالم التمييز على مُعادلة الاختبارات. وقد أظهرت نتائج الدراسة بأن وجود الفقرات المتطرّفة في الفقرات المشتركة يعمل على زيادة أخطاء المُعادلة بغض النظر عن كون مجموعات الدراسة متكافئةً أو غير متكافئةٍ وبغض النظر عن أسلوب المُعادلة المُستخدَم.

أما دراسة هي وسوي وفانق وشن (He & Cui & Fang & Chen, 2013)، فهدفت إلى تحديد الفقرات المتطرّفة باستخدام نموذج انحدار خطي، وفحص تأثير حذفها على مُعادلة العلامة الحقيقية في نظرية الاستجابة للفقرة. وقد أظهرت نتائج الدراسة أن حذف الفقرات المتطرّفة يقلل من أخطاء المُعادلة، ويزيد من فعاليتها.

كما أجرى هي (He, 2013) دراسةً هدفت إلى تطوير أساليب جديدة للحصول على معاملات تحويل التدريج للتحقق من تأثير الفقرات المشتركة المتطرّفة، ومقارنة أداء تلك الأساليب على فاعلية مُعادلة الاختبارات. وقد أشارت نتائج الدراسة إلى أن أخطاء المُعادلة تزداد بوجود فقرات متطرّفة.

وهناك جانب آخر من التطرّف على درجة من الأهمية غير الفقرات المتطرفة هو القيم الشاذّة في درجات الأفراد Outliers التي ينبغي الكشف عنها، وتحديد أسلوب التعامل معها. ويرى ريه(Ray, 2016) أن معظم المعالم الإحصائية، مثل الأوساط الحسابية، والانحرافات المعيارية، ومعاملات الارتباط، وكل إحصائي يعتمد عليها، حساسة جداً للقيم المتطرفة. وتوجد عدّة أساليب لتحديد القيم الشاذّة من امثلتها: أسلوب الانحراف المعياري، والدرجة المعيارية، والدرجة المعيارية المعدّلة The Modified-Standard Score، وأسلوب توكي (الرسم الصندوقي) Tukey's Method(Boxplot)، والرسم الصندوقي المعدّل Adjusted

Boxplot، والانحراف المطلق للوسيط MAD_e، وقاعدة الوسيط Median Rule، واختبار جروب Grubb's test. وعند ظهور قيم متطرِفةٍ أو شاذّةٍ في مجموعة من البيانات، تظهر الحاجة إلى أساليب لاستبدالها، ومن أشهر تلك الأساليب: الوسط المبتور Trimmed Mean، والوسط التعويضي Winsorized Mean. ويعد الوسط المبتور من أكثر الأساليب فاعليةً في استبدال القيم الشاذّة لعدم تأثره بها بشكل كبير. كما يُعد مقدّرًا جيّدًا لوسط المجتمع لقدرته على تخفيض تأثير القيم الشاذّة (Mann, 2009).

وقد تمت دراسة أثر القيم الشاذة في الدرجات على الاختبارات الإحصائية المعلمية واللامعلمية (Zimmerman, 2001)، وعلى الاختبارات الإحصائية المعلمية في البيانات الخاصة بالهرمونات(Pollet & Meij, 2017)، وعلى نتائج تحليلات الارتباط (Osborne & Overbay, 2004)، وعلى تقديرات معامل ثبات كرونباخ ألفا α (Liu & Wu & Zumbo, 2010)، وتحليلات الانحدار البسيط والمتعدد (Dan & Ijeoma, 2013) . وفي حدود علم الباحِثَين فإنه لم يتم التطرق لدراسة أثر القيم الشاذة في درجات الافراد على دقة مُعادلة الاختبارات. ومن هنا جاءت فكرة الدراسة الحالية التي قامت على مقارنة فاعلية مُعادلة نموذجي اختبار باستخدام نظرية الاستجابة للفقرة في ظل اختلاف أسلوب التعامل مع القيم الشاذة في درجات الافراد(الاحتفاظ بالقيم الشاذّة في درجات الأفراد، أو حذفها، أو استبدالها) باستخدام محكي تقويم فاعلية المُعادلة وهما: الصدق التقاطعي Cross-Validation، والخطأ المعياري للمُعادلة Standard Error of Equating.

مشكلة الدراسة وأسئلتها

يُلْحَظُ تزايد الحاجة إلى العديد من الاختبارات التي تستخدم صورًا متعددةً للاختبارات نفسِها في المجالات التربوية وغير التربوية. وتقتضي هذه الصور إجراء مُعادلة للدرجات المنبثقة عنها؛ لكي تكون عملية المقارنة بين الأفراد موضوعيّةً وعادلةً، ولكي يتمكن المتخصصون من تفسير الدرجات بشكل صحيح، وبالتالي اتخاذ القرارات المناسبة. وقد أشارت الدراسات إلى أن وجود قيمٍ شاذّةٍ في مجموعةٍ من البيانات قد تؤدي إلى تضخم معدّل الأخطاء، إذ إنها تزيد الخطأ من النوع الثاني. وبالتالي تعمل على تخفيض قوة الاختبارات الإحصائية. كما تؤدي إلى تشوهات كبيرة في تقديرات الإحصاءات والمعالم سواءً كان ذلك عند استخدام اختبارات معلمية أم اختبارات لا معلمية، ومن المحتمل أن تؤدي إلى تحيُزٍ في التقديرات موضع الاهتمام (Seo, 2006; Zimmerman, 2001). وقد تؤثر التقديرات المتحيزة سلباً في فاعلية مُعادلة الاختبارات. وهذا يستدعي الكشف عن الدرجات الشاذة في درجات الاختبارات المراد معادلة الدرجات لها، وتحديد الأسلوب المناسب للتعامل معها وبما يعزز من فاعلية معادلة درجاتها. وجاءت الدراسة الحالية لفحص أثر أسلوب التعامل مع القيم الشاذّة على فاعلية مُعادلة نموذجي اختبار بالاعتماد على القدرة (θ) حسب نظرية الاستجابة للفقرة بدلالة محكي الخطأ المعياري للمعادلة والصدق التقاطعي. ولأغراض الدراسة الحالية، تم استخدام استجابات طلبة الصف الرابع الأساسي على نموذجي الاختبار الوطني لضبط نوعية التعليم في الرياضيات للعام الدراسي 2012/2011 م.

وتتحدد مشكلة الدراسة بالإجابة عن السؤال: "ما أثر أسلوب التعامل مع القيم الشاذّة في فاعلية مُعادلة نموذجي الاختبار الوطني لضبط نوعية التعليم لمبحث الرياضيات لطلبة الصف الرابع الأساسي للعام 2012/2011 م حسب نظرية الاستجابة للفقرة بدلالة محكي الخطأ المعياري للمُعادلة والصدق التقاطعي ؟"

أهمية الدراسة

تنبع أهمية الدراسة الحالية من اهمية دقة معادلة الدرجات لاختبارين أو أكثر. وتُعد دقة المعادلة للدرجات أمراً أساسياً لأي برنامج اختباري يتضمن استخدام عدة نماذج مختلفة من الاختبار سواء في مرة التطبيق الواحدة أو في مرات التطبيق المختلفة. وأنه من المهم جداً أن تتم عملية معادلة الاختبارات بحرص ودقة، ذلك أن أي خطأ في دالة تحويل الدرجات من المحتمل أن يؤثر في درجات جميع المفحوصين. وعليه فإن دقة المعادلة تتصل بكل من العدالة والصدق على حد سواء. وفي ضَوْءِ مراجعة الأدب التربوي المتعلق بدقة مُعادلة الاختبارات، اتضح بأنه تمّ دراستها في ضَوْءِ كثير من المتغيرات، والتي كانت بالمجمل مرتبطةً بخصائص الاختبار كمحتوى الفقرات، وعددها، ودرجة تمثيلها، وتوزيعات القدرة، وحجوم العينات، وطول الاختبار، والتطرف في فقرات الاختبار وفي الفقرات المشتركة. والصفة العامة لهذه المتغيرات هي ارتباطها بالاختبار نفسه. كما تبين عدم وجود دراسات حول مدى تأثر دقة معادلة الاختبارات بأسلوب التعامل مع القيم الشاذة في درجات الاختبارات المراد معادلة درجاتها. لذلك يضيف البحث الحالي معلومات جديدة تتعلق بأثر أسلوب التعامل مع القيم الشاذّة في فاعلية مُعادلة نموذجي اختبار. وتفيد نتائج الدراسة في توفير معلومات عن جدوى الكشف عن القيم الشاذة وأسلوب استبدالها في تحسين دقة المعادلة لاختبارين أو أكثر. وبذلك يتحقق الهدف الرئيس للمعادلة وهو السماح باستخدام الدرجات من الاختبارات المختلفة على نحو تبادلي، ما ينطوي عليه نتائج وفائدة كبيرة على الصعيد التطبيقي؛ إذ يوفر إمكانية التفسير والتشخيص والمقارنة العادلة واتخاذ قرارات تربوية بشأن التطور التربوي. وتساهم هذه الدراسة في دعم القاعدة النظرية للبحوث المتعلقة بتفسير نتائج مُعادلة الاختبارات وَفْقًا لنماذج نظرية الاستجابة للفقرة. كما تقدم للباحثين والتربويين المتخصصين بالاختبارات في الجامعات ووزارتي التربية والتعليم، والتعليم العالي الأردنية تصوُرًا واضحًا حول كيفية التعامل مع القيم الشاذّة عندما توجد في بيانات نماذج الاختبارات المختلفة المطبّقة على الطلبة في تلك المؤسسات ضمن برامجها المختلفة.

التعريفات الإجرائية

مُعادلة الاختبارات Tests Equating

عملية إجراء تعديل إحصائي لدرجات النموذج الأوّل، بحيث تصبح مُكافئةً لدرجات النموذج الثاني (Brennan, 2004, p.2 & Kolen). وفي الدراسة الحالية سيتم تحويل قيم القدرة θ_x على النموذج الأول إلى قيم القدرة θ_y على النموذج الثاني من خلال المُعادلة الخطية .

القيم الشاذة Outliers

هي قيم الأفراد في أي من نموذجي الاختبار المُراد مُعادلتهما، التي لم تتسق مع القيم الأخرى وتنحرف بشكل كبير وواضح عنها(Hawkins, 1980, p.1). والقيم الشاذّة في الدراسة الحالية هي قدرات الأفراد (θ) التي تقع خارج الحدود الداخلية للرسم الصندوقي.

الخطأ المعياري للمعادلة Standard Error of Equating

هو مؤشر على دقة المعادلة، وبنقصان قيمته تزداد دقة وفاعلية المعادلة. ويقاس في الدراسة الحالية بالانحراف المعياري للدرجات المعادلة خلال تكرار عملية المعادلة في 400 عينة من المفحوصين.

محددات الدراسة

عند تعميم نتائج الدراسة لا بُدّ من أخذ المحددات الآتية بالاعتبار:

1- اقتصارها على طريقةٍ واحدةٍ من طرق مُعادلة الاختبارات المنبثقة عن نظرية الاستجابة للفقرة، وهي مُعادلة علامة القدرة.

2- اقتصارها على مِحَكَّين من مِحَكَّاتِ تقويم فاعلية مُعادلة الاختبارات، وهما: الخطأ المعياري للمُعادلة والصدق التقاطعي.

3- اقتصارها على أسلوب واحدٍ من أساليبِ الكشف عن القيم الشاذّة، وهو أسلوب الرسم الصندوقي.

4- اقتصارها على أسلوب واحدٍ من أساليب استبدال القيم الشاذّة، وهو أسلوب الوسط المبتور.

5- اقتصارها على نموذجين للاختبار الوطني لضبط نوعية التعليم لمبحث الرياضيات لطلبة الصف الرابع الأساسي للعام الدراسي 2012/2011م.

6- اقتصارها على النموذج اللوجستي ثلاثي المعالم.

7- اقتصارها على عينتين عشوائيتين، الأولى تمّ اختيارها من استجابات الأفراد الذين أجابوا عن النموذج الأول، والثانية تمّ اختيارها من استجابات الأفراد الذين أجابوا عن النموذج الثاني، وحجم كل منهما (2000) طالبًا وطالبةً.

الطريقة

مجتمع الدراسة

تكوّن مجتمع الدراسة من (25%) من طلبة الصف الرابع الأساسي في مدارس المملكة الأردنية الهاشمية الحكومية والخاصة، والمدارس التابعة لمديرية التعليم والثقافة العسكرية للعام الدراسي 2012/2011م. وبلغ عدد الأفراد الكلي للمجتمع (28303) طالبًا وطالبةً، حيث أجاب (14424) طالبًا وطالبةً عن النموذج الأوّل للاختبار الوطني لضبط نوعية التعليم ، وأجاب (13879) طالبًا وطالبةً عن النموذج الثاني للاختبار الوطني لضبط نوعية التعليم (وزارة التربية والتعليم الأردنية، 2012).

عينة الدراسة

تمّ اختيار عينتين عشوائيتين من أفراد مجتمع الدراسة: تكونت العينة الأولى من الأفراد الذين أجابوا عن النموذج الأوّل بحجم (2000) طالبًا وطالبةً. وتكونت العينة الثانية من الأفراد الذين أجابوا عن النموذج الثاني بحجم (2000) طالبٍ وطالبةٍ.

أداة الدراسة

تمّ الاعتماد على بيانات الاختبار الوطني لضبط نوعية التعليم لمبحث الرياضيات لطلبة الصف الرابع الأساسي للعام الدراسي 2012/2011 م. وهو اختبار أعدّته وزارة التربية والتعليم (مديرية الاختبارات) لقياس درجة امتلاك طلبة الصف الرابع الأساسي للمهارات الرياضية الأساسية في مرحلة التعليم الأساسي(الصف الأول _ الصف الرابع)، والكشف عن مستويات أداء الطلبة بدلالة مؤشرات الأداء المعيارية للرياضيات، ودرجة امتلاكهم لمهارات اقتصاد المعرفة في مبحث الرياضيات. حيث تمّ بناء تجمع من الفقرات من فريق متخصص تمّ تشكيله لهذا الغرض، وَفْقًا لجدول مواصفات يتكون من بعدين هما:

- المحتوى: ويتضمن أربعة محاور أساسية تتمثل بالأعداد والعمليات عليها، والجبر، والهندسة والقياس، والإحصاء والاحتمالات.
- المستويات المعرفية: وتتضمن محورين يتمثلان بالقدرات العقلية الدنيا، والقدرات العقلية العليا.

وبعد تجريب الفقرات، ودراسة خصائصها الإحصائية، تمّ الإبقاء على 71 فقرةً من نوع الاختيار من متعدد، تمّ توزيعها في نموذجين: تضمن النموذج الأوّل (35) فقرةً، وتضمن النموذج الثاني (36) فقرةً. ويوضح جدول (1) توزيع فقرات نموذجي الاختبار على بعدي المحتوى والمستويات المعرفية.

الجدول 1: توزيع فقرات نموذجي الاختبار على بعدي المحتوى والمستويات المعرفية

المحتوى	النموذج الأول		النموذج الثاني	
	مهارات تفكير دنيا	مهارات تفكير عليا	مهارات تفكير دنيا	مهارات تفكير عليا
الأعداد والعمليات عليها	13	4	12	5
الجبر	4	2	3	2
الهندسة والقياس	7	2	9	1
الإحصاء والاحتمالات	3	0	3	1
المجموع	27	8	27	9

وحتى تتم معادلة الدرجات على النموذجين، لا بد من توفر بعض الشروط التي تتمثل في قياسهما للسمة نفسها، وتساوي الثبات في كليهما، ومتطلبات التماثل والمساواة واللاتغير في المجتمع. وقد تحقق للنموذجين في الدراسة الحالية قيمة عالية لثبات الاتساق الداخلي حسب معادلة كودر-ريتشاردسون 20 حيث بلغ ثبات الأول 0.82 وثبات الثاني 0.80 وبذلك يتحقق تساوي الثبات لهما لكون الفرق بين معاملي الثبات للنموذجين يقل عن (0.03). كما تؤشر عملية بناء فقرات النموذجين اعتمادا على جدول مواصفات موحد على انهما يقيسان السمة نفسها بالرغم من وجود بعض الاختلافات في توزيع الفقرات على مجالات المحتوى. ولحسن الحظ فإن الاختلافات في توزيع الفقرات على المجالات الأربعة للمحتوى لم تكن دالة إحصائيا عند مستوى (α=0.01)، حيث كانت قيمة اختبار مربع كاي للاستقلال بدرجات حرية تساوي 4 هي 0.272. ويتحقق متطلب التماثل بالتعريف في المعادلة الخطية(Victor, 2007). ويكتنف شرطي المساواة واللاتغير في المجتمع بعض الغموض وعدم الصلة وأنهما غير عمليان وصارمان (Dorans & Holand, 2000) ويجوز إجراء عملية المعادلة بدونهما.

المعالجات الإحصائية

لأغراض التحقق من افتراضات نظرية استجابة الفقرة، وتحليل استجابات الأفراد، واستخراج معالم الفقرات وتقدير القدرة، والكشف عن القيم الشاذة واستبدالها، ومُعادلة الاختبارات، وحساب أخطاء المُعادلة، تمَّ استخدام برمجيات حاسوبية متعددة:

أولًا: برمجية NOHARM (Normal Ogive Harmonic Analysis Robust Method)

استخدمت للتحقق من أحادية البعد لنموذجي الاختبار الأوّل والثاني باستخدام مؤشّر تاناكا لحسن مُطابقة البيانات لنموذج، ومؤشّر الجذر التربيعي للوسط الحسابي لمربعات البواقي RMSR. وتدلل قيمة مؤشّر تاناكا التي تزيد على (0.90) على تحقق أحادية البعد (Jasper, 2010). وتدلل قيمة مؤشّر الجذر التربيعي للوسط الحسابي لمربعات البواقي التي تقل عن أو تساوي القيمة الحرجة له ($RMSR \leq 4.1/\sqrt{N}$) على تحقق أحادية البعد (Fraser & McDonald, 1988; Jasber, 2010; Ruscio, Haslam & Ruscio, 2012)

ثانيًا: برمجية LDID (**L**ocal **D**ependence **I**ndices for **D**ichotomous Items)

استُخْدِمَتْ للتحقق من افتراض الاستقلال الموضعي لفقرات نموذجي الاختبار الأوّل والثاني، وَفْقًا لمؤشر Q3 للاستقلال الموضعي والمحول إلى قيم Z الفشرية Z_{Q_3} . فإذا كانت قيمة Z_{Q_3} ضمن فترة الثقة فإنها تدلل على الاستقلال، وإذا كانت خارج فترة الثقة فإنها تدلل على عدم الاستقلال (Kim & Cohen & Lin, 2005)

ثالثًا: برمجية الحزمة الإحصائية للعلوم الاجتماعية (SPSS V22)

استخدمت لتنظيم مخرجات برمجية (BILOG-MG3)؛ لجعلها مدخلات مناسبة لبرمجيتي (LDID, Excel).

رابعًا: برمجية (BILOG-MG3)

استخدمت للتحقق من مطابقة بيانات عينتي الدراسة لنموذجي الاختبار لافتراضات النّموذج اللوجستي الثلاثي المعلمة.

خامسًا: برمجية إكسيل (Excel)

استخدمت للكشف عن القدرات الشاذة وللإجابة عن سؤال الدراسة، وعلى النحو الآتي:

* **للكشف عن القدرات الشاذة:** تمَّ تحويل العلامات الخام إلى علامات قدرة لكلا نموذجي الاختبار الأوّل والثاني باستخدام برمجية (BILOG-MG3)، ومن ثمّ تمّ الكشف عن القدرات الشاذّة في مجموعة البيانات لنموذجي الاختبار الأوّل والثاني باستخدام أسلوب توكي (الرسم الصندوقي)، لأنّه يعّد من الأساليب الحسّاسة في الكشف عن القيم الشاذّة في مجموعة من البيانات، لاعتماده على الرباعيات (Brys & Hubert &

Rousseeuw, 2005; Dan & Ijeoma ,2013b; Olewuezi, 2011;Seo, 2006).

ولتحديد القيم الشاذّة حسب هذا الأسلوب (Seo, 2006). تمّ اتباع الخطوات الآتية :

1- حساب المدى الربيعي IQR

$$IQR = Q_3 - Q_1 \quad \text{------------(1)}$$

حيث:

Q_1: الربيع الأول، وهو القيمة التي يقل عنها أو يساويها 25% من القيم.

Q_3: الربيع الثالث، وهو القيمة التي يقل عنها أو يساويها 75% من القيم.

2- حساب الحدود الداخلية (الفترة الداخلية)، حسب العلاقة التالية:

$$[\, Q_1 - 1.5\ IQR,\ Q_3 + 1.5\ IQR \,] \quad \text{------------(2)}$$

3- حساب الحدود الخارجية (الفترة الخارجية)، حسب العلاقة التالية:

$$[\, Q_1 - 3\ IQR,\ Q_3 + 3\ IQR \,] \quad \text{-------------(3)}$$

فإذا وقعت القيمة ما بين الحدود الداخلية والحدود الخارجية، فإنها تُعَدُّ قيمةً شاذّةً Outlier value وهي موضع الاهتمام في الدراسة الحالية. أمّا إذا وقعت القيمة خارج الحدود الخارجية، فتُعَدُّ قيمةً متطرّفةً Extreme value. لذلك تمّ حساب الربيعي الأول Q_1 والربيعي الثالث Q_3 لكلا نموذجي الاختبار، ومن ثمّ تمّ حساب المدى الربيعي IQR لكلا النموذجين باستخدام المعادلة (1)، وحساب الحدين الأدنى والأعلى للرسم الصندوقي باستخدام المعادلة (2)، وبعد ذلك تمّ رصد القدرات الشاذّة في مجموعة البيانات، إذ تمّ اعتبار أي قدرة (θ) كقيمة شاذّة إذا وقعت خارج حدود الفترة حسب المعادلة (2)، في كلا نموذجي الاختبار الأول والثاني.

*** للإجابة عن سؤال الدراسة**

حتى تتمّ عملية المُعادلة الخطية لعلامتي القدرة المُراد مُعادلتهما θ_x, θ_y (Hambleton & Swaminathan, 1985)، تمّ حساب ما يأتي:

- $(\mu_{\theta x})$: الوسط الحسابي لعلامات القدرة على صورة الاختبار x
- $(\mu_{\theta y})$: الوسط الحسابي لعلامات القدرة على صورة الاختبار y
- $(\sigma_{\theta x})$: الانحراف المعياري لعلامات القدرة على صورة الاختبار x
- $(\sigma_{\theta y})$: الانحراف المعياري لعلامات القدرة على صورة الاختبار y

ونحصل على علامتي قدرة متكافئتين على صورتي الاختبار، باستخدام المعادلة التالية:

$$\theta_y - \alpha\,\theta_x + \beta \quad \text{--------(4)}$$

حيث يتم حساب كل من ثوابت المعادلة (α, β) كما يأتي:

$$\alpha = \frac{\sigma_{\theta y}}{\sigma_{\theta x}} \quad \text{--------(5)}$$

$$\beta = \mu_{\theta y} - \frac{\sigma_{\theta y}}{\sigma_{\theta x}}\mu_{\theta x} \quad \text{--------(6)}$$

وبالتالي تصبح المعادلة (4) على الصورة التالية:

$$\theta_y - \frac{\sigma_{\theta y}}{\sigma_{\theta x}}\theta_x + \left(\mu_{\theta y} - \frac{\sigma_{\theta y}}{\sigma_{\theta x}}\mu_{\theta x}\right) \quad \text{------ (7)}$$

وللحكم على فاعلية مُعادلة الاختبارات (Effectiveness Equating) في الدراسة الحالية، تمّ الاعتماد على محكي الخطأ المعياري للمُعادلة، والصدق التقاطعي. ويمثل الخطأ المعياري للمُعادلة الانحراف المعياري للعلامات المُعادلة لعينة من المفحوصين، ويتمُّ حسابه بالاعتماد على المعادلات الآتية (Kolen&Brennan, 2004).

$$\hat{s}e[\hat{e}_{\theta_y}(\theta_x)] = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\left[\hat{e}_{\theta_y,r}(\theta_x) - \bar{\hat{e}}_{\theta_y}(\theta_x)\right]^2} \quad \text{-----------(8)}$$

$$\bar{\hat{e}}_{\theta_y}(\theta_x) = \frac{1}{R}\sum_{r=1}^{R}\hat{e}_{\theta_y,r}(\theta_x) \quad \text{--------(9)}$$

حيث:

R: عدد العينات العشوائية المُختارة من الصورة Y .

$\hat{e}_{\theta_y,r}(\theta_x)$: القدرة θعلى الصورة Y المُعادلة للقدرة θ على الصورة x.

$\bar{\hat{e}}_{\theta_y}(\theta_x)$: الوسط الحسابي للقدرات المُعادلة على جميع العينات.

ويمثل الصدق التقاطعي متوسط مربعات الفروق للعلامات المُعادلة عندما يطبّق على عينتين عشوائيتين مستقلّتيْن، ويكشف مدى استقرار العلامات المُعادلة في عينتين عشوائيتين مستقلّتَيْن عن بعضهما. ويشير بترسون وكولن وهوفر(Peterson, Kolen & Hoover) إلى أنه يمكن حساب معامل الصدق التقاطعي باستخراج متوسط مربعات الفروق للعلامات المُعادلة من العينة الأولى والعينة الثانية من خلال المعادلة التالية(الكوفحي،2013):

$$C = \sum_i \frac{(\theta_{Y_i} - \theta'_{Y_i})^2}{NK} \quad \text{------------(10)}$$

حيث :

θ_{Y_i}: القدرات المُعادلة في العينة الأولى.

θ'_{Y_i}: القدرات المُعادلة في العينة الثانية.

N: عدد القدرات الملاحظة في توزيع الصدق التقاطعي.

K: عدد الفقرات في الاختبار المُعادل.

أولاً : في حال الاحتفاظ بالقيم الشاذّة

تمّ استخدام المُعادلة الخطية لمُعادلة نموذج الاختبار الأوّل إلى نموذج الاختبار الثاني بالاعتماد على القدرة. إذ تمّ حساب الوسطين الحسابيين، والانحرافين المعياريين لبيانات عينتي الدراسة لنموذجي الاختبار الأوّل والثاني. ومن ثمّ تمّ حساب الميل (α) باستخدام المعادلة (5)، وحساب المقطع الصادي (β) باستخدام المعادلة (6). وبالاعتماد على قيمهما، تمّ إجراء المُعادلة الخطية من النموذج الأوّل إلى النموذج الثاني بالاعتماد على قدرات الأفراد في عينتي الدراسة حسب المعادلة (7). وللتحقق من فاعلية المُعادلة، تمّ اختيار عينتين عشوائيتين مستقلتين من القدرات المُعادلة. تكونت كلُّ عينة منهما من 400 طالب وطالبة، وحساب مربع انحراف كل قدرة مُعادلة من العينة العشوائية المستقلة الأولى عن نظيرتها من العينة العشوائية المستقلة الثانية. ومن ثمّ حساب قيمة معامل الصدق التقاطعي للمُعادلة باستخدام المعادلة (10). وتمّ حساب قيم الأخطاء المعيارية للقدرات المُعادلة باستخدام المعادلتين (8) و(9) بعد أن تمّ اختيار 400 طالب وطالبة من عيّنةٍ عشوائيّةٍ من العينة الأولى (الأفراد الذين أجابوا عن النموذج الأوّل والتي بحجم 2000 طالبًا وطالبةً)، بحجم 100 فرد لكلٍّ منها، واختيار 400 طالب وطالبة من عيّنةٍ عشوائيّةٍ من العينة الثانية (الأفراد الذين أجابوا عن النموذج الثاني والتي بحجم 2000 طالبًا وطالبةً) ، بحجم 100 فرد لكلٍّ منها، وذلك حسب الإجراءات الآتية:

أ- بالنسبة للنموذج الأوّل، تمّ القيام بالخطوات الآتية:

1. فتح الملف الرئيس لقدرات الأفراد غير المُعادلة.
2. اختيار عينةٍ عشوائيّةٍ واحدةٍ بحجم 100 فردٍ من الملف المذكور في الخطوة (1).
3. تخزين قدرات الأفراد غير المُعادلة للعينة العشوائية في ملف مستقل ويعطى اسمًا للملف يدلل على رقم العينة العشوائيّة.
4. تكرار الخطوات من (2 وحتى 3) 399 مرّةً.
5. تجميع ملفّات البيانات الـ 400 الخاصة بالعينات العشوائية الـ 400 في ملف بيانات واحد.
6. مراعاة تثبيت رقم تسلسلي يخصُّ كلّ عيّنةٍ عشوائية يبدأ بالقيمة (1) للعينة العشوائية الأولى، وينتهي بالقيمة (400) للعينة العشوائية الأخيرة رقم (400).

ب- بالنسبة للنموذج الثاني، تم تكرار جميع الخطوات المندرجة تحت النقطة (أ).

ج- جمع ملفي بيانات القدرة للنموذجين المذكورين في النقطتين (أ) و(ب) بالاعتماد على الأرقام المتسلسلة الخاصة برقم العينة العشوائية الذي تمّ الإشارة إليه في الخطوة (6).

د- حساب الوسط الحسابي والانحراف المعياري لقدرات الأفراد غير المُعادلة للنموذجين الأول والثاني لكلِّ عيّنةٍ عشوائيةٍ بالاعتماد على ملف البيانات المنتج في النقطة (ج).

ه- حساب قيمة الميل والمقطع الصادي بالاعتماد على مخرجات النقطة (د) لكلِّ عيّنةٍ عشوائيّةٍ.

و- إجراء عملية مُعادلة قدرات الأفراد عن طريق تعويض قيم مُختارة للقدرة غير المُعادلة تبدأ من (- 4) وحتى (+4) بزيادة مقدارها (0.1) لكلِّ عيّنةٍ عشوائيّةٍ، وَفْقًا للأسلوب الآتي:

(لكلِّ عينةٍ عشوائية من العينات العشوائية الـ (400)، تمّ ضرب قيمة الميل في قيمة القدرة غير المُعادلة المُختارة من ضمن مجموعة القيم الخاصة بها التي تمّ الإشارة إليها في النقطة (و)، ثم إضافة قيمة المقطع الصادي لحاصل عملية الضرب).

ز- حساب الوسط الحسابي لجميع قيم القدرة المُعادلة عبر العينات العشوائية الـ (400) عند كلِّ مستوى من مستويات القدرة المُختارة.

ح- حساب مربع الفرق بين القدرة المُعادلة لكلِّ عيّنةٍ عشوائيّةٍ والوسط الحسابي لهنّ جميعًا.

ط- حساب مجموع مخرجات النقطة (ح).

ي- قسمة مخرجات النقطة (ط) على عدد العينات العشوائية مطروحًا منها واحد (القسمة على 399).

ك- حساب الجذر التربيعي لمخرجات الخطوة (ي).

ل- تثبيت قيم الأخطاء المعيارية للقدرات المُعادلة أمام كل مستوى من مستويات القدرة المُختارة.

ثانيًا : في حال حذف القيم الشاذّة

بعدَ أن تمّ حذف القدرات الشاذّة من عيّنتي الدراسة في النموذجين الأوّل والثاني، تمّ تكرار الإجراءات كما في حال الاحتفاظ بالقيم الشاذّة لإجراء المُعادلة الخطية من النموذج الأوّل إلى النموذج الثاني، والتحقق من فاعلية المُعادلة.

ثالثًا : في حال استبدال القيم الشاذّة

تمّ استبدال القيم الشاذّة لبيانات عينتي الدراسة للنموذجين الأوّل والثاني، حسب أسلوب الوسط المبتور، والذي يتم من خلال حذف J% من أكبر القيم و J% من أقل القيم في مجموعة البيانات بعد ترتيب البيانات تصاعديًا أو تنازليًا، وحساب الوسط الحسابي لبقية القيم (100- 2J)% في مجموعة البيانات (Mann, 2009). وأشارت الجبوري (دبدوب ويونس، 2006) إلى أنه لاستبدال القيم الشاذّة حسب أسلوب الوسط المبتور، يتمُ ترتيب القيم تصاعديًا، ومن ثمّ يتمُ حذف أكبر قيمةٍ وأصغر قيمةٍ من البيانات بعد ترتيبها، وبعد ذلك يحسب الوسطَ الحسابيَّ للقيم المتبقية، أي إيجاد وسط

مبتورٍ لـ (n -2) من القيم، يمثل الوسط المبتور القيمة التقديرية للقيمة الشاذّة. ويتم حساب الوسط المبتور حسب المعادلة التالية:

$$\bar{y}_{tk} = \frac{1}{n-2k}\sum_{i-k+1}^{n-k} y_{(i)} \quad -------- (11)$$

حيث:

$\bar{y}_{tk}$: الوسط المبتور (الوسط الحسابي بعد حذف أعلى القيم وأدنى القيم).

$y_{(i)}$: عدد القيم المتبقية بعد الحذف.

n: حجم العينة الكلي.

k: عدد القيم المحذوفة من أعلى القيم أو من أدنى القيم.

لذلك تمّ الاستبدال قدرة قدرة، من خلال ترتيب القدرات تصاعديًا، ومن ثم حذف أكبر قدرة وأصغر قدرة من البيانات بعدَ ترتيبها، وبعدَ ذلك تمّ إيجاد الوسط الحسابي للقدرات المتبقية، أي إيجاد وسط مبتورٍ لـ (n -2) من القدرات حسب المعادلة (11)؛ إذ اعتبر قيمةً تقديريّةً للقدرة الشاذّة، ومن ثم تمّ استبدال القدرة الشاذّة بقيمة الوسط المبتور الناتجة سواءً أكانت القدرة الشاذّة من الطرف السفلي للتوزيع أم من الطرف العلوي، وتمّ تكرار ما سبق مع كل قدرة شاذّة حتى تمّ استبدال جميع القدرات الشاذّة في نموذجي الاختبار.

وبعدَ أن تمّ استبدال القدرات الشاذّة، تمّ استخدام المُعادلة الخطية لمُعادلة نموذج الاختبار الأوّل إلى نموذج الاختبار الثاني بالاعتماد على القدرة، وتمّ تكرار الإجراءات كما في حالتي الاحتفاظ بالقيم الشاذّة وحذفِها؛ لإجراء المُعادلة الخطية من النموذج الأوّل إلى النموذج الثاني، والتحقق من فاعلية المُعادلة.

وقد استخدم تحليل التباين الثنائي للقياسات المتكررة، لمقارنة الأوساط الحسابية للأخطاء المعيارية للقدرات المُعادلة بحسب متغيري أسلوب التعامل مع القيم الشاذّة (احتفاظ، حذف، استبدال)، ومستوى القدرة(مرتفع $1 < \theta \leq 4$، متوسط $-1 \leq \theta \leq 1$، منخفض $-4 \leq \theta < -1$). وتمّ استخدام اختبار بونفيروني Bonferroni للمقارنات البعدية.

النتائج

تم التحقق من افتراضات النموذج ومدى ملاءمته للبيانات قبل إجراء المعالجة. وقد أظهر مؤشر تاناكا لحسن مُطابقة البيانات للنموذج أن بيانات النموذجين الأوّل والثاني أحادية البُعْدِ، حيث بلغت قيمته (0.99210) للنموذج الأوّل، و(0.99051) للنموذج الثاني وكلاهما أكبر من (0.90). وأكدَ ذلك مؤشر الجذر التربيعي للوسط الحسابي لمربعات البواقي، حيث بلغت قيمته المحسوبة (0.009) في النموذج الأوّل و(0.008) في النموذج الثاني وكلاهما أقل من القيمة الحرجة (0.11836). وبيّن مؤشر Z_{Q_3} أنّ الاستقلال الموضعي متحقق لنموذجي الاختبار الأوّل والثاني، حيث شكلّت أزواج الفقرات التي وقعت ضمن فترة الثقة ما نسبته 96.81% بالنسبة للنموذج الأول وما نسبته 96.83% بالنسبة للنموذج الثاني. وأظهرت برمجية (BILOG-MG3) أنّ هناك (9) أفراد غير مُطابقين للنموذج اللوجستي ثلاثي المعالم في نموذجي الاختبار الأوّل والثاني، وأن هناك (10) فِقْراتٍ من النموذج الأوّل للاختبار، و(14) فِقْرةً من النموذج الثاني للاختبار لا تطابق النّموذج اللوجستي الثلاثي؛ لذلك تمّ الإبقاء على 25 فقرةً في النموذج الأول، و22 فقرةً في النموذج الثاني.

وبيّن أسلوب الرسم الصندوقي أنّ هناك 28 قُدرةً شاذّةً في ضَوْءِ الاستجابة للنموذج الأول من الاختبار، وجميعها كانت تقلُّ عن الحدِّ الأدنى (-2.97) للرسم الصندوقي. كما أن هناك 14 قدرةً شاذّةً في ضَوْءِ الاستجابة للنموذج الثاني من الاختبار، توزعت بالتساوي على طرفي التوزيع، حيث كانت 7 قدراتٍ تَقِلُّ عن الحد الأدنى (-2.92) و7 قدراتٍ تزيد على الحدِّ الأعلى (2.94) للرسم الصندوقي. وفيما يلي النتائج مصنفة بحسب أسلوب التعامل مع القيم الشاذّة:

أولاً: النتائج المتعلقة بالمُعادلة الخطية للقدرات في النموذجين في حالة الاحتفاظ بالقدرات الشاذّة

يبين الجدول (2) الأوساط الحسابية والانحرافات المعيارية لقدرات الأفراد في النموذجين في حال الاحتفاظ بالقدرات الشاذّة.

الجدول 2: الأوساط الحسابية والانحرافات المعيارية لقدرات الأفراد في النموذجين في حال الاحتفاظ بالقدرات الشاذّة

النموذج	عدد الفقرات	الوسط الحسابي	الانحراف المعياري
الأول (X)	25	-0.045	1.174
الثاني (Y)	22	0.076	1.037

ولمُعادلة قدرات الأفراد في النموذجين، تمّ حساب المعاملات للمعادلة الخطية، فكانت قيمةُ الميل 0.884، وقيمةُ المقطع الصادي 0.116. وبناءً على ذلك تمت المُعادلة للنموذجين من خلال المعادلة الخطية التالية:

$$\theta_Y = 0.884\,\theta_x + 0.116$$

حيث :

θ_x: القدرة على النموذج الأول X.

θ_Y:القدرة المعادلة على النموذج الثانيY.

وتفيد هذه المعادلة في معرفة القدرة على النموذج الثاني المعادلة لأيِّ قدرة على النموذج الأول في حال الاحتفاظ بالقدرات الشاذّة.

وللتحقق من فاعلية المُعادلة الخطية، تم حساب قيمة الخطأ المعياري للمُعادلة عند مستوياتٍ مختلفةٍ من القدرة (θ) كما هو مبين في الجدول (3).

الجدول 3: الأخطاء المعيارية للقدرات المُعادلة عند مستويات مختلفة من القدرة (θ) في حال الاحتفاظ بالقدرات الشاذّة

القدرة	الخطأ المعياري للمُعادلة	القدرة	الخطأ المعياري للمُعادلة	القدرة	الخطأ المعياري للمُعادلة
-4.0	0.44713	-1.3	0.22033	1.4	0.16792
-3.9	0.43816	-1.2	0.21320	1.5	0.17309
-3.8	0.42921	-1.1	0.20626	1.6	0.17862
-3.7	0.42028	-1.0	0.19953	1.7	0.18446
-3.6	0.41138	-0.9	0.19303	1.8	0.19060
-3.5	0.40250	-0.8	0.18679	1.9	0.19700
-3.4	0.39365	-0.7	0.18083	2.0	0.20365
-3.3	0.38484	-0.6	0.17518	2.1	0.21051
-3.2	0.37605	-0.5	0.16987	2.2	0.21756
-3.1	0.36730	-0.4	0.16493	2.3	0.22480
-3.0	0.35858	-0.3	0.16041	2.4	0.23220
-2.9	0.34991	-0.2	0.15633	2.5	0.23974
-2.8	0.34128	-0.1	0.15272	2.6	0.24742
-2.7	0.33269	0.0	0.14964	2.7	0.25522
-2.6	0.32416	0.1	0.14710	2.8	0.26313
-2.5	0.31567	0.2	0.14513	2.9	0.27114
-2.4	0.30725	0.3	0.14377	3.0	0.27924
-2.3	0.29889	0.4	0.14302	3.1	0.28742
-2.2	0.29060	0.5	0.14290	3.2	0.29569
-2.1	0.28238	0.6	0.14341	3.3	0.30402
-2.0	0.27425	0.7	0.14454	3.4	0.31242
-1.9	0.26620	0.8	0.14627	3.5	0.32088
-1.8	0.25825	0.9	0.14859	3.6	0.32940
-1.7	0.25041	1.0	0.15147	3.7	0.33797
-1.6	0.24268	1.1	0.15488	3.8	0.34658
-1.5	0.23509	1.2	0.15878	3.9	0.35524
-1.4	0.22763	1.3	0.16314	4.0	0.36394

تراوحت قيم الأخطاء المعيارية بين 0.143 و 0.447 بوسط حسابي يساوي 0.253 وانحراف معياري يساوي 0.088. وهي بصورةٍ عامّةٍ مرتفعةٌ نسبيًا، مما يؤثر على فاعلية المُعادلة بدرجةٍ كبيرةٍ نسبيًا. وتمّ حساب قيمة معامل الصدق التقاطعي للمُعادلة اعتماداً على المعادلة (10) والتي تتمحور حول معرفة القيمة المُعادلة في العينة الأولى والقيمة المُعادلة في العينة الثانية؛ إذ بلغت قيمته 0.001، وتُدلّلُ هذه القيمة على استقرار القدرات المعُادلة في العينتين العشوائتين المستقلتين بصورةٍ مرتفعةٍ نسبيًا.

ثانيًا: النتائج المتعلقة بالمُعادلة الخطية للقدرات في النموذجين في حال حذف القدرات الشاذّة

يبين الجدول (4) الأوساط الحسابية والانحرافات المعيارية لقدرات الأفراد في النموذجين في حال حذف القدرات الشاذّة.

الجدول 4: الأوساط الحسابية والانحرافات المعيارية لقدرات الأفراد في النموذجين في حال حذف القدرات الشاذّة

النموذج	عدد الفقرات	الوسط الحسابي	الانحراف المعياري
الأول (X)	25	0.008	1.093
الثاني (Y)	22	0.076	0.997

ولمُعادلة قدرات الأفراد في النموذجين، تمّ حساب المعاملات للمعادلة الخطية، فكانت قيمة الميل 0.912، وقيمة المقطع الصادي 0.069. وبناءً على ذلك تمت المعادلة للنموذجين من خلال المعادلة الخطية التالية:

$$\theta_Y = 0.912\,\theta_x + 0.069$$

وتفيد هذه المعادلة في معرفة القدرة على النموذج الثاني المُعادلة لأيّ قدرةٍ على النموذج الأول في حال حذف القدرات الشاذّة.

وللتحقق من فاعلية هذه المُعادلة الخطية، تمّ حساب قيمة الخطأ المعياري للمُعادلة عند مستوياتٍ مختلفةٍ من القدرة ($\boldsymbol{\theta}$) كما هو مبين في الجدول (5).

الجدول 5: الأخطاء المعيارية للقدرات المُعادلة عند مستويات مختلفة من القدرة (θ) في حال حذف القدرات الشاذّة

القدرة	الخطأ المعياري للمُعادلة	القدرة	الخطأ المعياري للمُعادلة	القدرة	الخطأ المعياري للمُعادلة
-4.0	0.38117	-1.3	0.18808	1.4	0.14400
-3.9	0.37353	-1.2	0.18203	1.5	0.14840
-3.8	0.36590	-1.1	0.17613	1.6	0.15310
-3.7	0.35829	-1.0	0.17042	1.7	0.15808
-3.6	0.35071	-0.9	0.16491	1.8	0.16331
-3.5	0.34314	-0.8	0.15961	1.9	0.16876
-3.4	0.33561	-0.7	0.15456	2.0	0.17441
-3.3	0.32809	-0.6	0.14977	2.1	0.18026
-3.2	0.32061	-0.5	0.14527	2.2	0.18626
-3.1	0.31316	-0.4	0.14110	2.3	0.19243
-3.0	0.30573	-0.3	0.13727	2.4	0.19872
-2.9	0.29834	-0.2	0.13383	2.5	0.20515
-2.8	0.29099	-0.1	0.13079	2.6	0.21169
-2.7	0.28368	0.0	0.12819	2.7	0.21833
-2.6	0.27641	0.1	0.12606	2.8	0.22506
-2.5	0.26919	0.2	0.12442	2.9	0.23188
-2.4	0.26202	0.3	0.12328	3.0	0.23879
-2.3	0.25490	0.4	0.12267	3.1	0.24576
-2.2	0.24785	0.5	0.12259	3.2	0.25280
-2.1	0.24085	0.6	0.12305	3.3	0.25990
-2.0	0.23393	0.7	0.12402	3.4	0.26705
-1.9	0.22708	0.8	0.12552	3.5	0.27426
-1.8	0.22032	0.9	0.12751	3.6	0.28151
-1.7	0.21365	1.0	0.12997	3.7	0.28881
-1.6	0.20708	1.1	0.13288	3.8	0.29615
-1.5	0.20062	1.2	0.13621	3.9	0.30353
-1.4	0.19429	1.3	0.13992	4.0	0.31094

يُلْحَظُ من الجدول (5) أنَ قيم الأخطاء المعيارية تتراوح بين 0.123 و 0.381 بوسط حسابي يساوي 0.216 وانحراف معياري يساوي 0.075؛ إذ إنّ حذف القدرات الشاذّة، التي تتمتّعُ عادةً بقيم مرتفعة للخطأ المعياري للتقدير، جعل الأخطاء المعيارية بصورة عامة متوسطةً نسبيًا. وبصورة عامة يُلْحَظُ أن الأخطاء المعيارية للمعادلة في حال حذف القدرات الشاذّة أقل منها في حال الاحتفاظ بالقدرات الشاذّة. وينعكس هذا إيجابًا على فاعلية المعادلة. كما تمّ حساب قيمة معامل الصدق التقاطعي للمُعادلة والذي بلغت قيمته 0.00045، وتدلل هذه القيمة على استقرار القدرات المعُادلة في العينتين العشوائتين المستقلتين وبصورة مرتفعة.

ثالثًا: النتائج المتعلقة بالمُعادلة الخطية للقدرات في النموذجين في حال استبدال القدرات الشاذّة

يبين الجدول (6) الأوساط الحسابية والانحرافات المعيارية لقدرات الأفراد في النموذجين، في حال استبدال القدرات الشاذّة.

الجدول 6: الأوساط الحسابية والانحرافات المعيارية لقدرات الأفراد في النموذجين في حال استبدال القدرات الشاذّة

النموذج	عدد الفقرات	الوسط الحسابي	الانحراف المعياري
الأول (X)	25	0.008	1.085
الثاني (Y)	22	0.076	0.993

ولمعادلة قدرات الأفراد في النموذجين، تمّ حساب المعاملات للمُعادلة الخطية، فكانت قيمة الميل 0.915، وقيمة المقطع الصادي 0.069. وبناءً على ذلك تمت المعادلة للنموذجين من خلال المعادلة الخطية التالية:

$$\theta_Y = 0.915\,\theta_x + 0.069$$

وتفيد هذه المعادلة في معرفة القدرة على النموذج الثاني المُعادلة لأيِّ قدرةٍ على النموذج الأول في حال استبدال القدرات الشاذّة وَفْقًا لطريقة الوسط المبتور.

وللتحقق من فاعلية هذه المُعادلة الخطية، تمّ حساب قيمة الخطأ المعياري للمُعادلة عند مستويات مختلفة من القدرة ($\boldsymbol{\theta}$) كما هو مبين في الجدول (7).

الجدول 7: الأخطاء المعيارية للقدرات المُعادلة عند مستويات مختلفة من القدرة (θ) في حال استبدال القدرات الشاذّة

القدرة	الخطأ المعياري للمُعادلة	القدرة	الخطأ المعياري للمُعادلة	القدرة	الخطأ المعياري للمعادلة
-4.0	0.12759	-1.3	0.04962	1.4	0.04556
-3.9	0.12462	-1.2	0.04700	1.5	0.04816
-3.8	0.12164	-1.1	0.04442	1.6	0.05080
-3.7	0.11867	-1.0	0.04191	1.7	0.05349
-3.6	0.11570	-0.9	0.03947	1.8	0.05622
-3.5	0.11273	-0.8	0.03712	1.9	0.05897
-3.4	0.10977	-0.7	0.03487	2.0	0.06175
-3.3	0.10681	-0.6	0.03275	2.1	0.06456
-3.2	0.10386	-0.5	0.03078	2.2	0.06738
-3.1	0.10091	-0.4	0.02900	2.3	0.07022
-3.0	0.09796	-0.3	0.02744	2.4	0.07308
-2.9	0.09502	-0.2	0.02615	2.5	0.07595
-2.8	0.09209	-0.1	0.02515	2.6	0.07883
-2.7	0.08917	0.0	0.02449	2.7	0.08173
-2.6	0.08625	0.1	0.02419	2.8	0.08463
-2.5	0.08334	0.2	0.02428	2.9	0.08754
-2.4	0.08044	0.3	0.02474	3.0	0.09046
-2.3	0.07755	0.4	0.02555	3.1	0.09339
-2.2	0.07467	0.5	0.02669	3.2	0.09633
-2.1	0.07181	0.6	0.02811	3.3	0.09927
-2.0	0.06896	0.7	0.02977	3.4	0.10221
-1.9	0.06612	0.8	0.03164	3.5	0.10516
-1.8	0.06331	0.9	0.03367	3.6	0.10812
-1.7	0.06052	1.0	0.03585	3.7	0.11108
-1.6	0.05775	1.1	0.03815	3.8	0.11404
-1.5	0.05500	1.2	0.04054	3.9	0.11701
-1.4	0.05230	1.3	0.04301	4.0	0.11998

يُلْحَظُ من الجدول (7) أنَ قيم الأخطاء المعيارية تتراوح بين 0.024 و 0.128 بوسط حسابي يساوي 0.068 وانحراف معياري يساوي 0.031. ويُلْحَظُ أنَ الأخطاء المعيارية بصورة عامة منخفضة نسبيًا، ما ينعكس إيجابًا على فاعلية المُعادلة. وبصورة عامة يُلْحَظُ أن الأخطاء المعيارية للمُعادلة في حال استبدال القدرات الشاذّة أقل منها في حال الاحتفاظ بالقدرات الشاذّة. كما تمَّ حساب قيمة معامل الصدق التقاطعي للمعادلة؛ إذ بلغت قيمته 0.00032، وتدلل هذه القيمة على استقرار القدرات المعُادلة في العينتين العشوائتين المستقلتين بصورة مرتفعة.

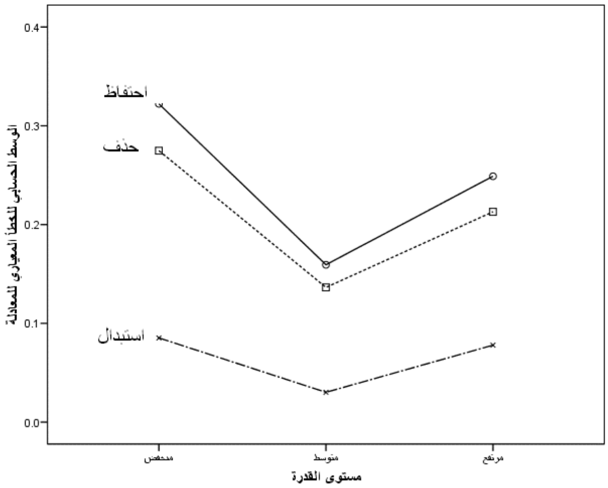
وقد تمَّ استخدام تحليل التباين الثنائي للقياسات المتكررة للأخطاء المعيارية للقدرات المُعادلة وَفْقًا لمتغيري أسلوب التعامل مع القيم الشاذّة، ومستوى القدرة، كما هو مبين في جدول (8).

الجدول 8: نتائج تحليل التباين الثنائي للقياسات المتكررة للأخطاء المعيارية للقدرات المُعادلة وَفْقًا لمتغيري أسلوب التعامل مع القيم الشاذّة ومستوى القدرة

اختبار الآثار	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية	الدلالة العملية
داخل القدرات Greenhouse-Geisser	أسلوب التعامل	1.41587	1.00039	1.41531	1697.25298	0.0000	0.96
	أسلوب التعامل×مستوى القدرة	0.08455	2.00078	0.04226	50.67517	0.0000	0.57
	الخطأ (حالة القيم الشاذّة)	0.06507	78.03056	0.00083			
	الكلي	1.56548	81.03174	0.01932			

اختبار الآثار	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية	الدلالة العملية
بين القدرات	مستوى القدرة	0.52277	2	0.26139	43.54167	0.0000	0.53
	الخطأ	0.46824	78	0.00600			
	الكلي	0.99102	80	0.01239			
	الكلي	**2.55650**	**161.03174**				

يتضح من جدول (8) وجود فروق دالة إحصائيًا عند مستوى الدلالة $\alpha = 0.01$ بين الأوساط الحسابية للأخطاء المعيارية للقدرات المُعادلة تعزى لكلٍ من أسلوب التعامل مع القيم الشاذّة، ومستوى القدرة والتفاعل بينهما. ويبين الشكل (1) أثر التفاعل بين أسلوب التعامل مع القيم الشاذّة، ومستوى القدرة على الأخطاء المعيارية للقدرات المُعادلة.



الشكل (1). أثر التفاعل بين أسلوب التعامل مع القيم الشاذّة، ومستوى القدرة على الأخطاء المعيارية للقدرات المُعادلة

يُلْحَظُ من الشكل (1) أنّ هناك تفاعلاً رتبيًا لأسلوب التعامل مع القيم الشاذّة ومستوى القدرة. فقد كان أسلوب استبدال القيم الشاذّة هو الأفضل مقارنةً مع أسلوبي حذفها والمحافظة عليها، وذلك عند جميع مستويات القدرة. ويبين الجدول (9) نتائج اختبار بونفيروني Bonferroni للمقارنات البعدية المتعددة للأخطاء المعيارية للقدرات المُعادلة وَفْقًا لأسلوب التعامل مع القيم الشاذّة.

الجدول 9: نتائج اختبار بونفيروني Bonferroni للمقارنات البعدية بين أوساط الأخطاء المعيارية للقدرات المُعادلة وَفْقًا لأسلوب التعامل مع القيم الشاذّة

أسلوب التعامل مع القيم الشاذّة		استبدال	حذف	احتفاظ
	الوسط الحسابي	0.06823	0.21601	0.25282
استبدال	0.06823			
حذف	0.21601	**0.14362**		
احتفاظ	0.25282	**0.17904**	**0.03542**	

يتضح من جدول (9) وجود فروق جوهرية عند مستوى الدلالة $\alpha = 0.05$ بين الأوساط الحسابية للأخطاء المعيارية للقدرات المُعادلة يعزى لأسلوب التعامل مع القيم الشاذّة، لصالح حال استبدالها مقارنةً بكلٍ من حالتي حذفها والاحتفاظ بها، ثم لصالح حذفها مقارنة بالاحتفاظ بها.

ويبين الجدول (10) نتائج اختبار بونفيروني Bonferroni للمقارنات البعدية للأخطاء المعيارية للقدرات المُعادلة وَفْقًا لمستوى القدرة.

الجدول 10: نتائج اختبار بونفيروني Bonferroni للمقارنات البعدية للأخطاء المعيارية للقدرات المُعادلة وَفْقًا لمستوى القدرة

مستوى القدرة		متوسط	مرتفع	منخفض
	الوسط الحسابي	0.10863	0.17992	0.22739
متوسط	0.10863			
مرتفع	0.17992	**0.07128**		
منخفض	0.22739	**0.11876**	**0.04747**	

يتضح من جدول (10) وجود فروق جوهرية عند مستوى الدلالة ($\alpha = 0.05$) بين الأوساط الحسابية للأخطاء المعيارية للقدرات المُعادلة يعزى لمستوى القدرة، لصالح مستوى القدرة المتوسط مقارنة بكلٍ من مستويي القدرة المنخفض والمرتفع، ثم لصالح مستوى القدرة المرتفع مقارنةً بمستوى القدرة المنخفض.

مناقشة النتائج

أظهرت النتائج التي تمّ التوصل إليها بأنّ هناك أفضلية لأسلوب استبدال القيم الشاذّة على أسلوبي حذف القيم الشاذّة أو الاحتفاظ بها وبشكل دال احصائيًا. وتؤكد هذه النتيجة قيمة معامل الصدق التقاطعي للقدرات المُعادلة التي كانت أقل ما يمكن في حال استبدال القيم الشاذّة، مما يدل على أنّ المُعادلة في ظل استبدال القيم الشاذّة كانت أكثر فاعليةً من المُعادلة في ظل حذفها أو الاحتفاظ بها. وتتفق هذه النتيجة مع نتائج دراسة زيميرمان (Zimmerman, 2001)، ونتائج دراسة أوسبورن وأوفرباي (Osborne & Overbay, 2004)، ونتائج دراسة برانكو وأوليفيرا وأوليفيرا وميندر (Branco, Oliveira, Oliveira & Minder, 2011) بالرغم من اختلاف موضوعاتها. وقد يعود السبب في ذلك إلى أن عملية استبدال القدرات الشاذّة تؤدي إلى تجانس القدرات أكثر مع المحافظة على حجم العينة، ما يؤدي إلى تقليل قيمة الوسط الحسابي والانحراف المعياري للأخطاء المعيارية للقدرات المُعادلة أكثر من حالتي حذف القيم الشاذّة والاحتفاظ بها. وبالرجوع إلى قيم الميل α وقيم المقطع الصادي β في غالبية العينات العشوائية التي دخلت في حساب الأخطاء المعيارية للقدرات المُعادلة، نجد بأن قيمهما كانت أقل ما يمكن في حال استبدال القيم الشاذّة، مما ساهم بتجانس القدرات بعد المُعادلة بشكل أكبر من حالتي حذف القيم الشاذّة والاحتفاظ بها، وانعكس ذلك على قيمة الوسط الحسابي للقدرات المعادلة؛ إذ ساهم في جعل قيمة بسط معادلة الخطأ المعياري للمُعادلة في حال استبدال القيم الشاذّة أقل منه في حالتي حذف القيم الشاذّة والاحتفاظ بها، مما جعل قيمة الوسط الحسابي والانحراف المعياري للأخطاء المعيارية للقدرات المُعادلة في حال استبدال القيم الشاذّة أقل منه في حالتي حذف القيم الشاذّة والاحتفاظ بها؛ لذلك تفوقت المُعادلة في حال استبدال القيم الشاذّة، وكانت أكثرَ فاعليةً من بقية الحالات.

كما توصلت النتائج إلى أنّ المُعادلة تكون أكثر فاعليةً كلما اتجه مستوى القدرة نحو الوسط، وتكون أقل فاعليةً كلما اتجه مستوى القدرة نحو الأطراف وخاصةً الطرف السفلي من التوزيع. وقد يعود السبب في ذلك الى أنّ توزيع القدرات كان ملتويًا إلى اليسار، وتمركز القدرات الشاذّة في الطرف السفلي من التوزيع.

التوصيات

في ضوء ما تم التوصل إليه من نتائج في الدراسة الحالية، يوصي الباحثان بما يلي:

1- ضرورة استبدال القيم الشاذّة في درجات الاختبارات المراد معادلتها قبل إجراء المُعادلة.

2- إجراء المزيد من الدراسات حول أثر اثر استبدال القيم الشاذّة في مجموعة من البيانات على فاعلية المُعادلة لاختبارات أخرى، أو باستخدام طرق مُعادلة أخرى، أو باسخدام محكات أخرى للحكم على فاعلية المُعادلة.

المراجع

الحجيلي، خالد. (2012). أثر *طول الاختبار وحجم العينة في دقة تقدير معلمة صعوبة الفقرة والقدرة ومعادلة الاختبارات بوجود الأداء التفاضلي للفقرة*. رسالة دكتوراة غير منشورة، جامعة اليرموك، إربد، الأردن.

الحياصات، خالد. (2011). *مدى تحقق معايير الفاعلية في معادلة نموذجي اختبار مع بقاء الفقرات ذات الأداء التفاضلي للجنس وحذفها*. رسالة دكتوراة غير منشورة، جامعة اليرموك، إربد، الأردن .

دبدوب، فرح و يونس، فرح. (2006). تأثير القيم الشاذة على نتائج تحليل الانحدار مع التطبيق على المواليد الخدج. *مجلة علوم الرافدين، 17*(1)، 62- 81.

الرحيل، راتب. (2013). أثر وجود أداء تفاضلي في الفقرات المرساوية على دقة المعادلة العمودية لاختبار أوتيس لينون للقدرة العقلية. *المجلة الدولية التربوية المتخصصة،2*(8)، 754 – 771.

الشريفين، نضال. (2003). *مدى تحقق معايير الفاعلية في معادلة اختبارين أحدهما ثنائي التدريج والآخر متعدد التدريج وفْقَ نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس*. رسالة دكتوراة غير منشورة، جامعة عمان العربية للدراسات العليا، عمان، الأردن.

الصمادي، مروان. (2006). *فاعلية طرق تصحيح اختبار الصواب-الخطأ المتعدد وتأثيرها على دقة معادلة الاختبار باستخدام نماذج النظرية الحديثة للقياس*. رسالة دكتوراة غير منشورة، جامعة اليرموك، إربد، الأردن.

الكوفحي، موسى. (2013). *أثر طريقة تصفية المموهات في فقرات الاختيار من متعدد على معادلة الاختبار*. رسالة دكتوراة غير منشورة، جامعة اليرموك، إربد، الأردن.

اللمع، أنور. (2012). *أثر مستوى الصعوبة والتمييز لفقرات الاختبار المستهدف والاختبار المشترك على دقة المعادلة باستخدام تصميم المجموعات العشوائية*. رسالة دكتوراة غير منشورة، جامعة اليرموك، إربد، الأردن.

وزارة التربية والتعليم الأردنية. (2012). *نتائج الاختبار الوطني لضبط نوعية التعليم للعام الدراسي 2011-2012*. عمان: إدارة الامتحانات والاختبارات.

Branco, F. , Oliveira, T. , Oliveira ,A. & Minder, C. (2011). The impact of outliers on the power of the randomization test for two independent groups. *Numerical Analysis and Applied Mathematics, 1389*(1), 1545-1548.

Brys,G. , Hubert, M. & Rousseeuw, P. (2005). A robustification of independent component analysis. *Journal of Chemometrics, 19,364*-375.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart & Winston Inc.

Dan, E. & Ijeoma, O. (2013a). Statistical analysis/ methods of detecting outliers in a multivariate data in a regression analysis model. *Journal of International Academic Research for Multi-disciplinary,1*(3),302-337.

Dan, E. & Ijeoma, O.(2013b). Statistical analysis/ methods of detecting outliers in a multivariate data in a regression analysis model. *International Journal of Education and Research, 1*(5),1-24.

Dorans, N. & Holand, P.(2000). Population invariance and the equalability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37,281-306.

Fraser, C. & McDonald, P.(1988). NOHARM: least squares item factor analysis. *Multivariate Behavior Research, 23,* 267-269.

Hambleton, R. & Swaminathan, H.(1985). *Item Response Theory: Principles and Applications***.** Boston: Kluwer-Nijhoff.

Hanson, B. & Béguin, 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*(1), 3-24.

Hawkins, D. (1980). *Identification of Outlier*. London: Chapman & Hall.

He, Y. (2013). *Robust Scale Transformation Methods in IRT True Score Equating Under Common-item Nonequivalent Groups Design*. Doctoral Dissertation, University of Missouri, Missouri, USA.

He, Y., Cui, Z. , Fang, Y. & Chen, H.(2013). Using a linear regression method to detect outliers in IRT common item equating. *Applied Psychological Measurement, 37*(7), 522-540.

Hu, H. (2004). *Investigation of IRT-Based Equating Methods in the Presence of Outliers*. Unpublished Doctoral Dissertation, University of Alberta, Edmonton, Canada.

Hu, H., Rogers, W. & Vukmirovic, Z. (2008). Investigation of IRT-based equating methods in the presence of outlier common items. *Applied Psychological Measurement, 32*, 311-333.

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.

Jimenez, F. (2011). *Effect of Outlier Item Parameters on IRT Characteristics Curve Linking Methods Under the Common Item Nonequivalent Groups Design.* Unpublished Doctoral Dissertation, University of Florida, Florida: USA.

Kim, S., Cohen, A. & Lin, Y. (2005). LDIP: A computer program for local dependence indices for dichotomous items. *Applied Psychological Measurement, 30*(6), 509-510.

Kim, S. & Kolen, M. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied Measurement in Education, 19*(4), 357 - 381.

Kolen, M. & Brennan, R (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd Ed.). New York: Springer- Verlag Science and Business Media.

Lee, W. & Ban, J. (2010). A comparison of IRT linking procedures**.** *Applied Measurement in Education, 23*(1), 23-48.

Ray, S. (2016). A Comprehensive Guide to Data Exploration. Data Science Central: The Online Resource for Big Data Practitioners. Retreived on October, 14, 2017 from: www.analyticsvidhya.com/blog/2016/01/guide-data-exploration.

Ruscio, J. & Haslam, N. & Ruscio, A. (2012). *Introduction to the Taxometric Method: A practical Guide*. NewYork : Taylor and Fransis Group .

Seo, M. (2006). *A Review and Comparison of Methods For Detecting Outliers in Univariate Data Sets*. Unpublished Master Thesis. University of Pittsburgh.

Victor, K. (2007). *Equating Accuracy Using Small Samples in the Random Groups Design*. Unpublished Doctoral Dissertation, Ohio University. Retrieved on May, 8, 2017 from: https://etd.ohiolink.edu.

Zimmerman, D. w. (2001). A note on the influence of outliers on parametric and nonparametric tests. *The Journal of General Psychology, 121*(4), 391-401.

Liu, Y., Wu, A. & Zumbo, B. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/scale item response**.** *Education and Psychological Measurement, 70* (1), 5-21.

Mann, S. (2009). *Introductory Statistics (*7th ed). Hoboken, NJ: Wiley and Sons.

Michaelides, M. (2003). *Effects of Common-item Selection on the Accuracy of Item Response Theory Test Equating with Nonequivalent Groups.* Unpublished Doctoral Dissertation, Stanford University, Stanford, California.

Olewuezi, N. (2011). Note on the comparison of some outlier labeling techniques. *Journal of Mathematics and Statistics, 7*(4), 353-355.

Osborne, J. & Overbay, A. (2004). The power of outliers (and why researchers should always check for them**).** *Practical Assessment, Research & Evaluation,9*(6). Retrieved October 7, 2014 from http://PAREonline.net/getvn.asp?v=9&n=6

Pollet, T. & Meij, L. (2017). To remove or not to remove: The impact of outlier handling on significance testing in testosterone data. *Adoptive Human Behavior and Psychology, 3*(1), 43-60.