أثر القيم المتطرفة على الأداء التفاضلي لفقرات اختبار الرياضيات في الدراسة الدولية "تيمس" تبعا لمتغير الجنس

زايد صالح بني عطا *

ملخص

هدفت الدراسة إلى التحقق من أثر القيم المتطرفة على الأداء التفاضلي لفقرات اختبار الرياضيات في الدراسة الدولية "تيمس" للصف الثامن تبعا لمتغير الجنس. ولتحقيق الغرض من الدراسة تم تحليل استجابات عينة مختارة عشوائيا على أحد كراسات الاختبار وهو الكراس الثاني والبالغ عددهم (554) طالبا وطالبة ممن تقدموا للاختبار للعام 2011. تكونت الصورة النهائية للاختبار من (26) فقرة بعد حذف الفقرات غير مطابقة للنموذج ثلاثي المعلمة. استخدمت طريقة توكي (الرسم الصندوقي) وطريقة مؤشر الأداء للفقرة والاختبار للكشف عن القيم المتطرفة، والأداء التفاضلي لفقرات الاختبار. كشفت نتائج الدراسة عن وجود (26) قيمة متطرفة منها (17) قيمة للإناث و(9) قيم للذكور، وبينت النتائج كذلك وجود سبع فقرات أبدت أداء تفاضليا بنوعيه المنتظم وغير المنتظم بوجود القيم المتطرفة من أصل (26) فقرة بنسبة مئوية (27%) تبعا لمتغير الجنس. وبعد حذف القيم المتطرفة من مجموعة البيانات وإعادة التحليل أظهرت النتائج وجود أربع فقرات قد أبدت أداء تفاضليا بنوعيه المنتظم وغير المنتظم بنسبة مئوية (15%)، حيث تناقصت هذه النسبة بمقدار (12%) عندما استبعدت القيم المتطرفة من التحليل.

الكلمات الدالة: الأداء التفاضلي للفقرة، النموذج اللوجستي ثلاثي المعلمة، اختبار الرياضيات في الدراسة الدولية، نظرية الاستجابة للفقرة.

المقدمة

يعد الانتشار الواسع لاستخدام الاختبارات التحصيلية العالمية المقننة بمثابة دراسات يشترك فيها كثير من دول العالم؛ بهدف عقد مقارنات بين تحصيل الطلبة على مستوى الدول جميعها، وعلى مستوى الدولة الواحدة، ومن هذه الدراسات دراسة التوجهات العالمية في العلوم والرياضيات (TIMSS) التي تنفذها الرابطة الدولية لتقييم التحصيل التربوي (The International Association of Educational Achievement) (IEA). وتعد دراسة التوجهات العالمية في العلوم والرياضيات (TIMSS 2011) الأحدث في سلسلة الدراسات الدولية التي هدفت إلى مقارنة تحصيل الطلبة في الصفين الرابع الأساسي والثامن الأساسي في الرياضيات والعلوم ضمن أنظمة تربوية متباينة في خلفياتها الثقافية، والاجتماعية والاقتصادية (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009).

وقد صممت اختبارات الدراسة الدولية في العلوم والرياضيات لتقديم تقديرات صحيحة وصادقة عن التحصيل العلمي للطالب، لذلك اعتمد في تصميم هذه الاختبارات مجموعة من مجمعات الفقرات (Item blocks)؛ إذ تم تطوير (14) مجمعا في الرياضيات ومثلها للعلوم، حيث تم توزيع هذه المجمعات على (14) كراسا، يتضمن كل كراس اختبارا من جزأين، الجزء الأول يحتوي على مجمعين للرياضيات، والجزء الثاني على مجمعين للعلوم. وقد استهدف اختبار الرياضيات في الدراسة الدولية للصف الثامن في هذه الدراسة أربعة مجالات لمحتوى الرياضيات، هي: الأعداد، والجبر، والهندسة، والبيانات والاحتمالات بحيث تقيس مجالات معرفية، هي: المعرفة، والتطبيق، والتبرير الرياضي. وشارك الأردن في هذه الدراسة الدولية (TIMSS 2011) من بين (45) دولة من الدول المشاركة، بالإضافة إلى مشاركته في الدورات الثلاث السابقة (1999، 2003، 2007) (المركز الوطني لتنمية الموارد البشرية، 2013).

ولضمان تحقيق العدالة لمثل هذه الاختبارات، خصوصا لما يترتب عليها من قرارات متعلقة بالنظام التربوي للدولة المشاركة في مقارنة مستويات التحصيل لدى طلبتها سواء كان على المستوى الخاص بالدولة أم تحصيل طلبتها مقارنة مع طلبة الدول

* كلية التربية – جامعة اليرموك. تاريخ استلام البحث 2015/12/29، وتاريخ قبوله 2016/4/5.

© 2018 عمادة البحث العلمي / الجامعة الأردنية. جميع الحقوق محفوظة.

الأخرى المشاركة، وللتأكد من عدالة الاختبار يتم العمل على أن تتطابق جميع الإجراءات المتعلقة بالاختبار مع المعايير الموضوعة بدءا من اختيار العينات وانتهاءً بتصحيح الاستجابات وتحليلها (Mullis, et al., 2009). ولأن تحليل البيانات المترتبة من استجابات المفحوصين على هذه الاختبارات يتطلب أساليب إحصائية مختلفة تحدد بخصائص البيانات، فقد تتضمن هذه البيانات قيما متطرفة تنعكس على أساليب التحليل المستخدمة.

ويعد موضوع القيم المتطرفة في البيانات من الموضوعات التي اهتم بها البحث السيكومتري والإحصائي لدراسة أثر هذه القيم في نتائج التحليلات الإحصائية ودقتها، حيث تعد أهمية تشخيص القيم المتطرفة في البيانات وطرق معالجتها خطوة مهمة في عملية التحليل الإحصائي، واتخاذ القرارات المترتبة عن تلك العملية. وينظر إلى القيمة المتطرفة كما عرفها جونسون وويشرن (Johnson & Wichern, 2007) على أنها قيمة غير متسقة مع بقية البيانات، في حين عرفها بارنت ولويس (Barnet & Lewise, 1994) بأنها قيمة أو أكثر لا تتسق مع مجموعة البيانات الكلية بحيث تبدو بعيدة عن بقية البيانات سواء كانت قيمة كبيرة أم صغيرة، وقد وصفا نوعين من القيم المتطرفة، النوع الأول: القيمة الملوثة (Contaminant value)، وهي قيمة تأتي من توزيع مختلف، والنوع الثاني: القيمة المتطرفة (Extreme value)، وهي إما أن تكون قيمة كبيرة أو صغيرة، لكنها من نفس التوزيع. بينما نظر إيفانز (Evavs,1999) للقيم المتطرفة على أنها قيم غير عادية يمكن أن تؤدي إلى حدوث تغييرات سلبية في نتائج التحليل الإحصائي. مما سبق يتضح بأن القيمة المتطرفة هي قيمة تخرج عن النسق المميز لمجموعة البيانات.

وقد أشار بارنت ولويس (Barnet & Lewise, 1994) بأن تطرف البيانات يعزى إما لأخطاء حسابية، أو أخطاء قراءة، أو أخطاء تسجيل، في حين أشار غرين (Green, 1976) إلى أن القيم المتطرفة في مجموعة البيانات قد تظهر بسبب أن البيانات تعود إلى توزيعات غير متماثلة، بمعنى قد يكون فيها التواء عال إما نحو اليمن أو نحو اليسار. بينما عزا هاوكينس (Hawkins, 1980) تطرف البيانات على اعتبار بأن البيانات تأتي من توزيعين، أحدهما التوزيع الأساسي (Basic distribution)، حيث يولد بيانات جيدة، بينما الآخر هو التوزيع الملوث (Contaminating distribution) الذي يولد قيما متطرفة.

ولأن القيم المتطرفة لها انعكاسات سلبية على النتائج ودقتها، فقد اهتم البحث الإحصائي بطرق الكشف عن مثل تلك القيم، وطرق معالجتها لتجنب آثارها السلبية على تحليل البيانات، حيث صنفت هذه الطرق في نوعين: النوع الأول، وهي الطرق التحليلية (Analytical Methods)، والنوع الثاني، وهي الطرق البيانية (Graphical Methods) (Ampanthong & Suwtte, 2009; Seo, 2006; Turkan, Cetin & Toktamis, 2012). وسيتم استخدام إحدى الطرق البيانية في هذا البحث، وهي طريقة الرسم الصندوقي (Box Plot)؛ للكشف عن القيم المتطرفة في مجموعة البيانات، وتعرف بطريقة توكي (Tukey, 1977). ويعود السبب في استخدام هذه الطريقة للكشف عن القيم المتطرفة بأنها لا تعتمد على الوسط الحسابي والانحراف المعياري، وهي طريقة قوية عند التعامل مع بيانات ذات حجم كبير، ويمكن استخدامها من أجل إجراء مقارنات بين عدد من المجاميع.

وقد تركزت جهود الباحثين في بناء المقاييس والاختبارات على استخراج الخصائص السيكومترية للاختبار والفقرات المتعلقة بصدق وثبات الاختبارات، وصعوبة الفقرات وتمييزها. وعلى الرغم من أهمية ذلك، إلا أن تان وجيرل (Tan & Jeral, 2005) أشارا بأن أحد أهم الخطوات في هذه العملية هو التحقق من عدالة الاختبار من خلال تحليل الأداء التفاضلي لفقراته. لذلك عمل المختصون في مجال القياس النفسي والتربوي على استخدام مفهوم الأداء التفاضلي للفقرة منذ بداية الثمانينيات من القرن الماضي؛ وذلك للتعبير عن الطرق والمعاني الإحصائية في الكشف عن تحيز فقرات الاختبار من منطلق أفضلية استخدام مصطلحات سيكومترية فنية بدلا من استخدام مصطلحات مشحونة سياسيا واجتماعيا (Ellis & Raju , 2003).

ويعد مفهوم الأداء التفاضلي للفقرة مؤشرا إحصائيا يستعمل للتعبير عن الفروق في احتمالات الاستجابة الصحيحة عن الفقرة بين مجموعتين مختلفتين من مجتمع ما لهما نفس مستوى القدرة (Gierl, Jodoin & Ackerman, 2000; Teresi, 2001). ومن منظور نظرية الاستجابة للفقرة ينظر إلى الأداء التفاضلي للفقرة من خلال فحص العلاقة بين الاحتمالات المشروطة للاستجابات الصحيحة على الفقرة والقدرة الكامنة المراد قياسها لمجموعتين فرعيتين من مجتمع ما (Gybles, 2004). ومن الجدير بالذكر بأن هناك من يستخدم مصطلح الأداء التفاضلي للفقرة كمصطلح مرادف لتحيز الفقرة، حيث أشار وليماز (Williams, 1997) بأن الفقرة تبدي أداءً تفاضليا إذا كانت احتمالات الاستجابة الصحيحة عنها مختلفا بين مجموعتين لهما نفس القدرة، بينما تعد الفقرة متحيزة عندما تقيس غير ما هو متوقع أن تقيسه. وقد ميز دوارنز وهولاند (Durance & Hollond, 1994) بين مصطلح الأداء التفاضلي للفقرة وتحيز الفقرة، فالأداء التفاضلي للفقرة يهتم بالخصائص السيكومترية لها من حيث

الكيفية التي تعمل بها في المجموعات المختلفة، في حين ينظر إلى تحيز الفقرة على انه مصطلح يتضمن معاني سلبية ترتبط بفكرة عدم العدالة والمساواة.

وفي الوقت التي تميز فيه أدبيات القياس بين الأداء التفاضلي للفقرة والتحيز، فقد ميزت بين الأداء التفاضلي والتحيز على الاختبار، حيث يشير تحيز الاختبار على انه خطأ منتظم في القياس، وهذا ما يجعل تأثيره في صدق الاختبار أكثر من ثباته، في حين يمثل الأداء التفاضلي للاختبار على الفرق في الأداء على الاختبار بين المجموعات المختلفة (Jensen, 1980 ; Osterlind, 1983)، وميز الأدب السيكومتري كذلك بين نوعين من الأداء التفاضلي للفقرة، الأول: الأداء التفاضلي المنتظم، حيث لا ينتج عن تفاعل مستوى القدرة مع عضوية المجموعة؛ بمعنى أن احتمال الإجابة الصحيحة لمجموعة من المفحوصين يكون إما أعلى أو أقل منه للمجموعة الأخرى، بينما النوع الثاني، وهو الأداء التفاضلي غير المنتظم، ويظهر عندما يكون هناك تفاعل بين مستوى الأداء وعضوية المجموعة، حيث سيكون الاختلاف بين منحى الاستجابة للفقرة في معلمة التمييز (Hambleton & Rogers, 1989; Swaminathan & Rogers, 1990; Pae, 2004).

وقد اتجهت دراسات الأداء التفاضلي في إيجاد الطرق المختلفة للكشف عن الفقرات التي تبدي أداء تفاضليا، حيث كان بعض هذه الطرق يعتمد على نظرية القياس التقليدية، وبعضها الآخر يعتمد على نظرية الاستجابة للفقرة، ونظرا للاستخدام المتزايد لنظرية الاستجابة للفقرة فقد انبثق منها طرق متعددة للكشف عن الأداء التفاضلي للفقرة (Hambleton & Swaminathan, 1985). ومن الطرق الحديثة في الكشف عن الأداء التفاضلي طريقة راجو وفاندر ليندن وفلير المطورة التي تسمى مقياس الأداء التفاضلي للفقرة والاختبار، حيث تقدم هذه الطريقة مؤشرين للأداء التفاضلي للفقرات، هما (Raju, Van Der Linden & Fleer, 1995):

أولا: مؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي(Noncompensatory DIF:NCDIF)، وهو معدل مربع المسافة ما بين منحني خصائص الفقرة للمجموعتين المستهدفة والمرجعية، ويحسب وفق المعادلة التالية:

$$NCDIF_i = E_F[d_i(\theta)^2] = \sigma^2_{d_i} + \mu^2_{d_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots 1$$

E_F: التوقع بالنسبة لتوزيع القدرة (θ) للمجموعة المستهدفة وقيمة (θ) d_i تعطى وفق العلاقة التالية

$$d_i(\theta) = P_{if}(\theta) - P_{ir}(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots 2$$

ثانيا: مؤشر الأداء التفاضلي على مستوى الفقرة التعويضي (Compensatory DIF:CDIF)، حيث يأخذ بالاعتبار التباين المشترك للفقرة، ويراعي التحيز الموجود في الفقرات الأخرى، ومن خلاله يمكن كشف الأداء التفاضلي على مستوى الاختبار باستخدام ما يسمى مؤشر الأداء التفاضلي على مستوى الاختبار(Differential Test Functioning :DTF)، ويتم حسابه وفق العلاقة التالية:

$$CDIF_i = Cov(d_i, D) + \mu_{d_i}\mu_{D_I} \quad \ldots\ldots\ldots\ldots\ldots\ldots 3$$

وتحسب قيمة D وفق المعادلة التالية

$$D = \sum_{i=1}^{n} d_i \quad \ldots\ldots\ldots\ldots\ldots\ldots 4$$

ومن أجل اختبار الدلالة الإحصائية لمؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF) تم ابتكار طريقة تكرار توليد معلمة الفقرة (Item Parameter Replication)، وتستند هذه الطريقة على اشتقاق نقاط قطع مصممة لتتوافق مع مجموعة معينة من الفقرات والبيانات المتعلقة بها، حيث تبدأ بمعالم الفقرات المقدرة للمجموعة المستهدفة، والتباينات، والتباينات المصاحبة لتلك المعالم يتم الحصول عليها من إحدى البرامج الخاصة بنظرية الاستجابة للفقرة، وبالاعتماد على هذه المعالم المبدئية يتم تكرار توليدها عدد من المرات بافتراض وجود مجتمعين لهما نفس معالم الفقرات، بحيث يمثل أحدهما المجموعة المستهدفة، والآخر المجموعة المرجعية، وينتج عن عملية التكرار أزواج من العينات التي تتضمن معالم الفقرات لكل من المجموعة المرجعية والمستهدفة، واعتمادا على ذلك يتم حساب مؤشر الأداء التفاضلي على مستوى الفقرة، واشتقاق علامة قطع

خاصة لكل فقرة لمستويات الدلالة الإحصائية (Oshima, Raju & Nanda, 2006). وقد تم استخدام هذه الطريقة في الكشف عن الأداء التفاضلي للفقرة وذلك لميزات هذه الطريقة من خلال استخدامها مع البيانات ثنائية التدريج ومتعددة التدريج، وتوفيرها مؤشرين للأداء التفاضلي على مستوى الفقرة والاختبار(Oshima & Morris, 2008).

ولأن الاختبارات التحصيلية العالمية في الرياضيات والعلوم (TIMSS) لها قيمتها التربوية في تمكين الدول المشاركة من فهم أنظمتها التربوية، وتقييم التجديدات التربوية التي تدخلها في أنظمتها التربوية، فقد اهتمت كثير من الدراسات بالوقوف على الفروق التحصيلية بين البلدان، ومقارنتها، ومصادر التحيز للفقرات، وكذلك بالفروق التحصيلية بين الجنسين. فمن الدراسات التي اهتمت بالفروق بين الجنسين في الرياضيات دراسة راين وشيو(Rayn & Chiu, 1996)، ودراسة فريدمان (Fridman, 1996)، حيث بينت نتائج هاتين الدراستين بأن الذكور أفضل من الإناث في حل المسائل، في حين كانت الإناث أفضل في الخوارزميات.

وراجع جونسون (Johnson, 1996) ثماني دراسات اهتمت بالمقارنة بين تحصيل الذكور والإناث في الرياضيات تناولت مستويات متباينة من العمر من (6–7) سنوات، ولغاية (16–17) سنة. استنتج من تلك الدراسات بأن أداء الإناث كان أفضل من أداء الذكور في المسائل الحسابية والعددية، في حين كان أداء الذكور أفضل في الجبر وحل المسائل.

وعلى صعيد الدراسات التي اهتمت بالكشف عن الأداء التفاضلي للفقرات سواء لاختبارات TIMSS، أو اختبارات تحصيلية عالمية أو وطنية في الرياضيات تبعا لمتغير الجنس، فقد أشارت نتائج دراسة هاريس وكارلتون (Haris & Carlton, 993) التي هدفت للكشف عن الفقرات التي تظهر أداء تفاضليا تبعا لمتغير الجنس باستخدام طريقة مانتل – هانزل في فقرات مجال الرياضيات من اختبار الاستعداد المدرسي، بأن أداء الذكور كان أعلى من أداء الإناث في الفقرات التي كانت تتطلب إجراءات غير روتينية لحلها, وفقرات الحساب والهندسة.

وأجرى إنجلهارد وغامر (Engelhard & Gamer, 1999) دراسة هدفت إلى الكشف عن الفروق بين أداء الذكور والإناث على اختبار قدرة رياضية المكون من (60) فقرة من نوع الاختيار من متعدد يقيس موضوعات الجبر والحساب والهندسة والقياس. بينت نتائج الدراسة بأن الذكور قد تفوقوا على الإناث في موضوعات القياس والهندسة، وتحليل البيانات، في حين بينت النتائج تفوق الإناث على الذكور في موضوع الجبر.

وفي دراسة مماثلة لميندس واريكان (Mendes & Erickan, 2006) التي هدفت إلى البحث عن مصادر الأداء التفاضلي في اختبار الرياضيات الأساسية (British Columbia Principles of Mathematical Exam) تبعا لمتغير الجنس الذي يطبق على طلبة الصف العاشر الأساسي. كشفت نتائج الدراسة بأن الفقرات التي تتضمن مشكلات ولوغارتيمات وأسسس كانت مفضلة لدى الذكور أكثر من الإناث، بينما الفقرات التي تتضمن علاقات رياضية وأوصافا اسمية كانت مفضلة لدى الإناث أكثر من الذكور، وأظهرت النتائج بشكل عام تفوق الذكور على الإناث في الفقرات التي تتطلب حلها عمليات عقلية عليا ورسوما.

واستخدم دودين وعنابي (Dodeen & Annabi, 2008) طريقة مانتل – هانزل في الكشف عن الفقرات التي تظهر أداء تفاضليا تبعا لمتغير الجنس في اختبار الرياضيات تيمس (TIMESS) لعام (1999) لطلبة الصف الثامن الأساسي الذين تم اختيارهم في الأردن والبالغ عددهم (5299) طالبا وطالبة. كشفت نتائج التحليل وجود (37) فقرة أبدت أداء تفاضليا من أصل (124) فقرة من نوع الاختيار من متعدد، منها (17) فقرة لصالح الذكور و(20) فقرة لصالح الإناث. وأشارت النتائج كذلك بأن معظم الفقرات التي أبدت أداء تفاضليا لصالح الإناث كانت تقيس موضوعات الجبر والإحصاء, أما الفقرات التي أبدت أداء تفاضليا للذكور فكانت تنتمي لموضوع القياس.

وفي السياق ذاته هدفت دراسة كوثر(2008) إلى تقصي فاعلية منحنى خصائص الفقرة المتمثلة بمؤشر المساحة للنموذج ثلاثي المعلمة، وثنائي المعلمة، ومؤشر فرق المساحة (b) في الكشف عن الفقرات التي تبدي أداء تفاضليا لعينة مختارة من الدراسة الدولية الثالثة (TIMSS)، حيث شكلت الباحثة اختبارا مكونا من (40) فقرة من نوع الاختيار من متعدد تقيس محتويات الهندسة, والجبر والقياس، تم تطبيقه على عينة مؤلفة من (800) طالب وطالبة من طلبة الصف الثامن الأساسي. بينت نتائج الدراسة وجود أداء تفاضلي متفاوت في الدرجة في عدد الفقرات تبعا لطبيعة المحتوى الرياضي الذي تقيسه من خلال الطرق الثلاث، حيث كان معظمه من نوع الأداء التفاضلي غير المنتظم، وبينت النتائج كذلك وجود أداء تفاضلي متفاوت الدرجة تبعا لمتغير الجنس.

إن المتصفح لنتائج الدراسات السابقة يجد بأنها اهتمت بالبحث عن مصادر الأداء التفاضلي سواء المتغيرات المتعلقة بالمحتوى الرياضي، أم المصادر الديموغرافية كمتغير الجنس الذي كان محور اهتمام الكثير من الدراسات والأبحاث لدراسة الفروق

التحصيلية في الرياضيات، وكذلك ركزت الدراسات على إيجاد ومقارنة طرق الكشف عن الأداء التفاضلي للفقرات. ومع ذلك لم يجد الباحث دراسة _ في حدود اطلاعه_ على مستوى الوطن العربية بشكل عام، وعلى الصعيد المحلي بشكل خاص دراسات اهتمت بدراسة أثر القيم المتطرفة على الأداء التفاضلي لفقرات الاختبار؛ مما جعل هناك فجوة في هذه الدراسات، وخصوصا لما يترتب من آثار سلبية على وجود مثل هذه في التحليلات الإحصائية، وما طرق الكشف عن الفقرات التي تبدي أداء تفاضليا إلا نوع من أنواع طرق التحليل الإحصائي. لذا جاءت الحاجة لإجراء هذه الدراسة من خلال تناولها أثر القيم المتطرفة على الأداء التفاضلي لفقرات الاختبار العالمي الذي يمس السياسة التعليمية في الأردن، وربما تكون أيضا مثل هذه القيم مصدرا آخر من مصادر التحيز التي لم تعن بها الأبحاث والدراسات.

مشكلة الدراسة وأسئلتها

انصب اهتمام الباحثين بشكل أساسي على دراسة أثر القيم المتطرفة على التحليلات الإحصائية من خلال مقارنة أساليب كشفها، وطرق معالجتها خصوصا في الاستدلال الإحصائي المبني على أساس التوزيع الطبيعي الذي يكون حساسا لمثل هذه القيم (Obsome & Overbay, 2004). ولأن القيم المتطرفة تعد قضية محورية ضمن مجموعة البيانات لما لها من آثار على قضايا التحليل الإحصائي كما يشير الأدب النظري الذي اهتم بدراسة أثر مثل هذه القيم على نتائج ودقة التحليلات الإحصائية الاختيار، إلا أن المتصفح للأدب السيكومتري المتعلق بدراسة أثر مثل هذه القيم على خصائص الاختبارات والمقاييس ودقة تقديرات الخصائص السيكومترية لها ولفقراتها يجد ندرتها – في حدود اطلاع الباحث– خصوصا في الأدب العربي، وخاصة فيما يتعلق بقضية الأداء التفاضلي للفقرات والاختبار.

وتجدر الإشارة كذلك بأن هذه الدراسة تأتي في ضوء كثير من المؤشرات التي تدل على انخفاض مستوى التحصيل الدراسي لدى طلبة الصف الثامن الأساسي في الرياضيات، تحديدا على الاختبار العالمي للرياضيات والعلوم في الدراسة الدولية TIMSS، كما أشارت نتائج التقرير الوطني لتنمية الموارد البشرية (2013)، حيث تبين تراجع وانخفاض مستوى أداء الطلبة في الأردن في هذه الدورة (TIMSS 2011)، مقارنة مع بقية الدول المشاركة وعن الدورة السابقة (2007) بشكل ملحوظ، وتراجع الوسط الحسابي في الرياضيات (21) نقطة عن الدورة السابقة (2007)، وتفوق الإناث على الذكور في العلوم والرياضيات في جميع الدورات (1999، 2003، 2007، 2011)؛ مما شكل ظاهرة تحتاج إلى التمعن والتبصر عن الأسباب المسؤولة عن هذا التفوق، التي يمكن أن تكون من ضمنها تحيز الفقرات والقيم المتطرفة، وبالتحديد فإن هذه الدراسة سعت للوصول إلى إجابة للتساؤلات الآتية:

1. ما القيم المتطرفة في درجات طلبة الصف الثامن الأساسي في اختبار الرياضيات في الدراسة الدولية تيمس (TIMSS 2001)؟
2. ما فقرات اختبار الرياضيات في الدراسة الدولية تيمس (TIMSS 2011) للصف الثامن الأساسي التي تبدي أداء تفاضليا تبعا لمتغير الجنس بوجود القيم المتطرفة؟
3. ما فقرات اختبار الرياضيات في الدراسة الدولية تيمس (TIMSS 2011) للصف الثامن الأساسي التي تبدي أداء تفاضليا تبعا لمتغير الجنس بدون القيم المتطرفة؟

أهمية الدراسة

تكمن أهمية هذه الدراسة في معالجة موضوع غفلت عنه الدراسات من حيث تناولها القيم المتطرفة، وانعكاسها على الأداء التفاضلي للفقرات التي تعد من القضايا المهمة؛ نظرا للاستخدام الواسع للاختبارات العالمية، وما يترتب عليها من قرارات للوصول إلى نتائج أكثر دقة وكفاءة في التقويم التربوي، ووضع خطط مستقبلية لتطوير وتحسين الانظمة التربوية، ودراسة فعالية المناهج الدراسية وطرق التدريس، والبحث عن مصادر أخرى للتحيز والأداء التفاضلي للفقرات عبر المجموعات الفرعية من المجتمع نفسه. وتأتي كذلك أهمية هذه الدراسة بالإسهام العلمي المتواضع في التراكم المعرفي في هذا المجال، حيث ستسهم نتائج الدراسة في قيام بعض الباحثين بالمزيد من الدراسات المشابهة على نطاق أوسع. ومن الناحية العملية فإن هذه الدراسة ستسهم بمساعدة المؤسسات والجهات التي تقوم بتطوير الاختبارات العالمية الاهتمام بالقيم المتطرفة في مجموعة البيانات الناتجة من تطبيق تلك الاختبارات وانعكاساتها على خصائص الفقرات ومنها الأداء التفاضلي، وأخذ ذلك بالاعتبار عند تفسير الدرجات، ومقارنة أداء الطلبة سواء على المستوى المحلي أو العالمي.

مصطلحات الدراسة

- **القيمة المتطرفة:** هي مشاهدة تكون بعيدة عن بقية المشاهدات الأخرى سواء كانت قيمة كبيرة أو صغيرة (Barnet & Lewise, 1994)، وتم الكشف عنها في هذه الدراسة باستخدام طريقة توكي (الرسم الصندوقي).
- **الأداء التفاضلي للفقرة:** هو مؤشر إحصائي يعبر عن دلالة الفروق في احتمالية الإجابة الصحيحة بين مجموعتين من المفحوصين ممن هم في نفس مستوى القدرة (Teresi, 2001). وفي هذه الدراسة سيتم الكشف عن الأداء التفاضلي للفقرة باستخدام طريقة راجو المطورة.
- **النموذج ثلاثي المعلمة:** هو أحد نماذج نظرية الاستجابة للفقرة ثنائية التدريج، حيث يمكن لهذا النموذج تقدير ثلاث معلمات للفقرة، هي: معامل صعوبة الفقرة، ومعامل التمييز، ومعامل تخمين الفقرة وفق المعادلة التالية:

$$P_i(\theta) = C_i + \frac{1 - C_i}{1 + e^{Da_i(\theta - b_i)}} (i = 1,2,...., n)$$ (Hambleton & Swaminathan, 1985).

- **الأداء التفاضلي المنتظم:** وهو الذي يظهر عندما ينتج تفاعل رتبي بين عضوية المجموعة ومستوى القدرة، بمعنى أن احتمال الإجابة الصحيحة عن الفقرة لمجموعة من المفحوصين يكون إما أعلى أو أقل منه للمجموعة الأخرى Swaminathan (& Rogers, 1990).
- **الأداء التفاضلي غير المنتظم:** ويظهر عندما يكون هناك تفاعل بين مستوى القدرة وعضوية المجموعة، بمعنى أن احتمال الإجابة الصحيحة عن الفقرة ليست ثابتة عند مستويات القدرة .(Swaminathan & Rogers, 1990).
- **الدراسة الدولية في العلوم والرياضيات TIMSS:** وهي عبارة عن أداء اختبارات عالمية لتقييم التوجهات في مدى تحصيل الطلبة في مبحثي العلوم والرياضيات للصفين الرابع الأساسي والثامن الأساسي، الأمر الذي يساعد صانعي القرار والمختصين في المناهج فهم أداء أنظمتهم التربوية من خلال ما توفره هذه الدراسة من قاعدة بيانات لكل دولة من الدول المشاركة في هذه الدراسة وتكون تحت إشراف الهيئة الدولية لتقييم التحصيل التربوي (IEA) (Mullis, et al., 2009).

محددات الدراسة

اقتصرت هذه الدراسة على:

- استخدام طريقة واحدة عن الكشف عن الأداء التفاضلي للفقرة وهي طريقة راجو باستخدام مؤشر الأداء التفاضلي اللاتعويضي (NCDIF).
- عينة الدراسة من طلبة الصف الثامن الأساسي الذين تقدموا على الكراس الثاني لاختبار الرياضيات في الدراسة الدولية.

الطريقة والإجراءات

منهج الدراسة

لتحقيق الغرض من الدراسة استخدم الباحث المنهج الوصفي الذي يعد الأكثر ملاءمة لمثل هذا النوع من الدراسات والأبحاث.

مجتمع الدراسة وعينتها

تكون مجتمع الدراسة من جميع طلاب وطالبات الصف الثامن الأساسي في المدارس الحكومية والخاصة ووكالة الغوث الدولية التابعة لوزارة التربية والتعليم خلال العام الدراسي 2011/2010، والذين تقدموا لاختبار الرياضيات والعلوم في الدراسة الدولية (TIMSS 2011)، والبالغ عددهم (7694) طالبا وطالبة، وقد اعتمد في اختيار هؤلاء الطلاب إجراءات محددة وفق دليل المعاينة الذي تم تطويره لأغراض الدراسة الدولية في الرياضيات والعلوم، بحيث يتم مراعاة جنس المدرسة، وموقعها الجغرافي، والسلطة المشرفة عليها بالتنسيق مع المركز الوطني لتنمية الموارد البشرية. والجدول (1) يوضح توزيع مجتمع الدراسة في الأردن حسب الجنس

جدول 1: توزيع طلبة الصف الثامن الأساسي في الدراسة الدولية حسب الجنس

المتغير	الفئة	عدد الطلاب	النسبة المئوية
الجنس	ذكور	3604	47
	إناث	4090	53
المجموع		**7694**	**100**

ولغايات تحقيق الهدف من الدراسة قام الباحث باختيار استجابات الطلبة على إحدى كراسات الاختبار المكونة من (14) كراسا عشوائيا، وكان في هذه الدراسة الكراس الثاني، وبذلك تكون عينة الدراسة مكونة من (554) طالبا وطالبة، منهم (264) طالبا و(290) طالبة.

بيانات الدراسة

لأغراض الدراسة تم استخدام البيانات المتوفرة عن فقرات الرياضيات في الدراسة الدولية للكراس الثاني بعد الحصول على كل البيانات المتعلقة بالدراسة الدولية للعلوم والرياضيات (TIMSS 2011) من المركز الوطني لتنمية الموارد البشرية. وتغطي فقرات الكراس الثاني لاختبار الرياضيات في الدراسة الدولية للصف الثامن الأساسي بعدين، هما: الأول بعد المحتوى، ويغطي محور الأعداد, والهندسة، والجبر، والبيانات والاحتمالات. أما البعد الثاني، وهو بعد المجالات المعرفية فيتضمن مجال المعرفة، والتطبيق، والتبرير الرياضي. ويتضمن اختبار الرياضيات في الدراسة الدولية للكراس الثاني فقرات من نوع الاختيار من متعدد، والإجابة القصيرة، والصواب والخطأ على غرار بقية فقرات الكراسات الأخرى لتصبح عدد فقرات الرياضيات في هذا الكراس (32) فقرة.

ومن الأمثلة على فقرات الكراس الثاني لاختبار الرياضيات في الدراسة الدولية فقرة رقم (MO52228) من نوع الاختيار من متعدد تقيس المستوى التطبيقي ضمن محور الأعداد، ومحتوى الفقرة يتعلق بمهارة طرح الكسور ومتن الفقرة هو (أي من التالي يبين طريقة صحيحة للحصول على نتيجة $\frac{1}{3}-\frac{1}{4}$). وكمثال أخر على فقرة من فقرات الكراس الثاني لاختبار الرياضيات في الدراسة الدولية فقرة رقم (MO42086) تقيس المستوى التطبيقي في المجال المعرفي، وتغطي محور الجبر ويتعلق محتوى الفقرة بالمعادلات والاقترانات ومتن الفقرة هو (ا+ب = 25 ما قيمة 2أ + 2ب +4).

إجراءات الدراسة

لغايات الإجابة عن أسئلة الدراسة وتحقيق الهدف المنشود منها قام الباحث بالإجراءات التالية:

1- تم الحصول على استجابات الطلبة على فقرات الكراس الثاني لاختبار الرياضيات والعلوم في الدراسة الدولية، وبعد ذلك تم حذف استجابات الطلبة عن فقرات العلوم والاحتفاظ باستجاباتهم عن فقرات الرياضيات المكون من (32) فقرة، بالإضافة إلى حذف جميع المتغيرات الأخرى التي ليست لها علاقة بالدراسة الحالية باستثناء متغير جنس الطالب.

2- استبعاد ثلاث فقرات متعددة التدريج من فقرات الاختبار، وتحمل الأرقام (MO52002, MO42059, MO42207)، والإبقاء على الفقرات ثنائية التدريج ليصبح الاختبار مكونا من (29) فقرة، حيث تم تحويل استجابات الطلبة على صورة (صفر وواحد) لغايات التحليل بالاستناد إلى مفتاح التصحيح المستخدم بالاختبار.

3- أدخلت البيانات إلى ذاكرة الحاسوب، واستخدم البرنامج الإحصائي (SPSS)، وبرنامج (Bilog- Mg3)، وبرنامج Excel، وبرنامج Equate المصمم من قبل بيكر (Baker, 1995)، وبرنامج DFIT الذي تم تطويره من قبل أوشيما وراجو وكوشبار وسكوت (Oshima, Kushubar ,Scott & Raju, 2009) لإجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة الدراسة، وقد تمت عملية التحليل وفق المراحل التالية:

أولا: التحقق من افتراضات النموذج

تم التحقق من افتراضات النموذج اللوجستي ثلاثي المعلمة، وهو أحادية البعد لما له من أثر على دقة التقديرات، وذلك باستخدام التحليل العاملي (Principle Component Analysis). ويظهر الجدول (2) قيم الجذر الكامن (Eigen Values)، ونسب التباين المفسر(Explained Variance) للعامل الأول والثاني، وناتج قسمة الجذر الأول على العامل الثاني.

جدول 2: قيم الجذر الكامن ونسب التباين المفسر للعامل الأول والثاني وناتج قسمة قيمة الجذر الأول على العامل الثاني

العامل	الجذر الكامن	التباين المفسر	ناتج القسمة
الأول	18.283	5.302	3.695
الثاني	4.949	1.435	

يتضح من النتائج الواردة في الجدول (2) بأن التباين المفسر من العامل الأول كان قريبا من (20)، وأن ناتج قسمة الجذر الكامن للعامل الأول على العامل الثاني أكبر من (2)، مما يشير إلى وجود سمة سائدة للاختبار؛ وهذا يعني تحقق افتراض أحادية البعد للاختبار (Hatti, 1984; Reckase, 1997).

ثانيا: مطابقة البيانات للنموذج

بهدف التحقق من درجة مطابقة الأفراد للنموذج اللوجستي ثلاثي المعلمة، تم استخدام برنامج Bilog-Mg 3 (Zimowski, Muraki, Mislevey, Robert & Book, 2002)؛ لتحليل استجابات الطلبة على فقرات الكراس الثاني لاختبار الرياضيات في الدراسة لدولية، ويتم الحكم على درجة مطابقة استجابة الفرد أو عدم مطابقتها مع النموذج من خلال القيمة الاحتمالية لمربع كاي، أو إذا تعذر تقدير الخطأ المعياري لقدرة الفرد، وتدل عليه القيمة (999.000). أظهرت نتائج التحليل مطابقة جميع استجابات الطلاب في كراس الاختبار الثاني باستثناء استجابات طالبين من طلاب عينة الدراسة، ويحملان الأرقام المتسلسلة (34, 399)، حيث كانت قيم كاي تربيع دالة إحصائيا عند مستوى الدلالة (α =0.01) .

وفيما يتعلق بفحص مطابقة الفقرات للنموذج المستخدم، أعيد التحليل باستخدام برنامج Bilog-Mg 3 بعد حذف استجابة الطالبين غير مطابقين للنموذج، وبناءً على مؤشر مربع كآي عند مستوى الدلالة (α =0.01) الذي يعد الفقرة غير مطابقة للنموذج إذا كان مقدار الاحتمالية لهذه الفقرة أقل من (0.01)، حيث أظهرت نتائج التحليل عدم مطابقة الفقرات ذوات الأرقام (1، 7، 15) التي أرقامها الأصلية في الكراس الثاني (M052216, M05302, M042032) على التوالي. وبعد حذف الفقرات غير المطابقة للنموذج اللوجستي ثلاثي المعلمة والأفراد غير المطابقين، أصبح اختبار الرياضيات في الدراسة الدولية يتكون من (26) فقرة بصورته النهائية، وبذلك يصبح مدى العلامات من (صفر) إلى (26) على هذا الاختبار، وعدد أفراد عينة الدراسة (552) طالبا وطالبة، بحيث تشكل استجاباتهم على فقرات الاختبار بصورته النهائية بيانات الدراسة.

ثالثا: تقدير قيم معالم الفقرات للاختبار حسب متغير الجنس

تم تقدير معالم الفقرات لفقرات الاختبار العالمي في الرياضيات المكون من (26) فقرة وفق النموذج ثلاثي المعلمة لمجموعتي الدراسة حسب متغير الجنس، كذلك تم تقدير معلمة القدرة لكل طالب حسب متغير الجنس بعد مطابقة مجموعة البيانات والفقرات للنموذج، وقد اعتبرت مجموعة الذكور في هذه الدراسة المجموعة المرجعية ومجموعة الإناث المجموعة المستهدفة من أجل الكشف عن الأداء التفاضلي للفقرات. والجدول (3) يوضح قيم معالم الفقرات لكل فقرة حسب مجموعات الدراسة (متغير الجنس).

جدول 3: قيم معالم الفقرات (الصعوبة، التمييز، التخمين) لدى مجموعتي الدراسة الذكور والإناث

رقم الفقرة المتسلسل	رقم الفقرة في الكراس	المحتوى	الذكور			الاناث		
			الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين
1	M052231	الأعداد	0.691	2.799	0.037	0.438	2.139	0.080
2	M052061	الأعداد	1.339	2.161	0.068	1.677	2.461	0.065
3	M052228	الأعداد	1.080	4.105	0.119	0.692	3.454	0.103
4	M052214	الأعداد	2.014	1.780	0.248	1.861	1.734	0.308
5	M052173	الجبر	2.917	1.823	0.082	2.331	2.742	0.063
6	M052362	الهندسة	0.745	3.422	0.043	1.309	1.488	0.061
7	M052408	الهندسة	1.132	2.047	0.076	1.514	2.412	0.110
8	M052084	الهندسة	1.459	1.639	0.101	1.106	4.930	0.146

رقم الفقرة المتسلسل	رقم الفقرة في الكراس	المحتوى	الذكور			الاناث		
			الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين
9	M052206	الهندسة	1.962	2.135	0.041	1.783	3.255	0.036
10	M052429	إحصاء واحتمالات	1.471	1.596	0.096	1.376	4.371	0.191
11	M052503A	إحصاء واحتمالات	2.037	1.488	0.025	2.366	1.533	0.069
12	M052503B	إحصاء واحتمالات	2.307	1.419	0.044	2.231	1.535	0.097
13	M042031	الأعداد	1.180	1.750	0.170	1.170	3.147	0.272
14	M042186	الأعداد	0.576	2.964	0.034	0.710	5.390	0.090
15	M042236	الجبر	0.952	4.908	0.232	0.769	4.799	0.266
16	M042226	الجبر	0.739	3.910	0.032	0.516	2.654	0.049
17	M042103	الجبر	1.823	2.591	0.014	1.521	3.446	0.023
18	M042086	الجبر	1.237	1.943	0.044	1.562	2.202	0.037
19	M042228	الجبر	1.589	1.834	0.040	1.803	2.944	0.130
20	M042245	الجبر	1.319	2.835	0.088	1.112	3.053	0.113
21	M042270	الهندسة	1.331	1.824	0.108	1.370	1.593	0.121
22	M042201	الهندسة	1.353	2.372	0.037	1.041	3.946	0.062
23	M042152	الهندسة	1.321	1.063	0.118	1.417	1.531	0.315
24	M042269	إحصاء واحتمالات	1.193	2.340	0.421	1.104	1.916	0.404
25	M042179	إحصاء واحتمالات	1.041	1.206	0.084	0.851	1.808	0.193
26	M042177	إحصاء واحتمالات	1.614	1.423	0.181	1.379	2.814	0.215
الوسط الحسابي			1.401	2.289	0.099	1.346	2.819	0.139
الانحراف المعياري			0.536	0.945	0.090	0.525	1.136	0.101

يتضح من النتائج الواردة في الجدول (3) تباينا ملحوظا في مدى صعوبة الفقرات، حيث كان المدى في صعوبة الفقرات للذكور تتراوح قيمته من (0.567) إلى (2.917)، بينما كان المدى للإناث يتراوح ما بين (0.438) ولغاية (2.366). أما بالنسبة لمعاملات التمييز فقد تراوحت قيمها بالنسبة للذكور من (1.063) إلى (4.904)، بينما كانت قيم معاملات التمييز للإناث تتراوح ما بين (1.488) إلى (5.390). وبينت النتائج كذلك بأن مدى معلمة التخمين كانت متقاربة للذكور والإناث؛ وربما يعود هذا التقارب في معلمة التخمين لصعوبة الفقرات كما هو وارد في الجدول (3) مما جعل الطلبة يلجئون إلى التخمين في الإجابة عن بعض فقرات الاختبار.

4- تم استخدام المعالجات الإحصائية التالية للإجابة عن أسئلة الدراسة:

- إيجاد القيم المتطرفة في استجابات الطلبة على فقرات الاختبار باستخدام طريقة توكي الرسم الصندوقي من خلال استخدام برنامج Excel.
- وضع تقديرات معالم الفقرات المقدرة لكلا المجموعتين على مقياس مشترك من خلال إجراء عملية معادلة (Equating) باستخدام برنامج Equate لاستخراج ثوابت المعادلة في حالة وجود القيم المتطرفة، وبعد حذف القيم المتطرفة.
- استخدام برنامج DIFTT للكشف عن الأداء التفاضلي للفقرات والاختبار، ورسم منحنى خصائص الفقرة للفقرات التي أبدت أداء تفاضليا لتحديد نوع الأداء التفاضلي للفقرة.

عرض النتائج ومناقشتها

للإجابة عن السؤال الأول تم إيجاد الإحصاءات الوصفية لعلامات عينة الدراسة على فقرات الكراس الثاني لاختبار الرياضيات في الدراسة الدولية، والجدول (4) يوضح هذه الإحصاءات.

جدول 4: الإحصاءات الوصفية لعلامات عينة الدراسة على فقرات اختبار الرياضيات في الدراسة الدولية

العدد	أقل علامة	أعلى علامة	الوسط الحسابي	الانحراف المعياري	الوسيط	المئين 75	المئين 25	المدى الربعي
552	00.0	23	6.93	4.45	6	9	4	5

يتضح من النتائج الواردة في الجدول (4) بأن الوسط الحسابي لعلامات عينة الدراسة بلغ (6.93) بانحراف معياري (4.45)، في حين كانت قيمة المئين75 والمئين25 تساوي (9 , 4) على التوالي. وبالنظر إلى نتائج الجدول (4) يتضح تدني مستوى أداء الطلبة على الاختبار، وجاءت هذه النتيجة منسجمة مع توصل إليه نتائج تقرير المركز الوطني لتنمية الموارد البشرية (2013) بتدني مستوى أداء الطلبة الأردنيين على الاختبار الرياضيات في العلوم والرياضيات في الدراسة الدولية (TIMSS 2011) بشكل ملحوظ، وتراجع ترتيب الأردن إلى المرتبة (35) من بين (45) دولة مشاركة في هذه الدورة.

وبالاعتماد على نتائج الجدول (4) ولغايات تحديد القيم المتطرفة لمجموعة البيانات من خلال استخدام طريقة توكي تم إيجاد الحدود الفاصلة الحد الأعلى والحد الأدنى كمحكات يستند عليها في تحديد القيم المتطرفة، حيث تم حساب الحد الأعلى الذي هو عبارة عن (المئين75 +1.5 * المدى الربعي) فبلغت قيمته (16.5). أما بالنسبة للحد الأدنى الذي هو عبارة عن (المئين 25 – 1.5 * المدى الربعي) فكانت قيمته (3.5 -). واستنادا إلى هذه الحدود الفاصلة فقد بلغ عدد القيم المتطرفة (26) قيمة، والجدول (5) يوضح الأرقام المتسلسلة للطلاب والقيم المتطرفة.

جدول 5: الأرقام المتسلسلة للطلاب وقيم العلامات المتطرفة في بيانات الكراس الثاني لاختبار الرياضيات تيمس للصف الثامن

الرقم المتسلسل	العلامة	الجنس	الرقم المتسلسل	العلامة	الجنس	الرقم المتسلسل	العلامة	الجنس
6	17.00	طالبة	314	18.00	طالب	468	19.00	طالبة
129	17.00	طالبة	484	18.00	طالب	507	19.00	طالب
130	17.00	طالبة	491	18.00	طالبة	51	21.00	طالب
176	17.00	طالبة	526	18.00	طالبة	547	21.00	طالبة
548	17.00	طالب	14	19.00	طالبة	550	21.00	طالب
9	18.00	طالبة	22	19.00	طالبة	251	22.00	طالبة
53	18.00	طالبة	145	19.00	طالبة	457	22.00	طالب
72	18.00	طالبة	152	19.00	طالبة	538	23.00	طالب
241	18.00	طالبة	294	19.00	طالب			

يتضح من النتائج الواردة في الجدول (5) بأن عدد القيم المتطرفة بلغ (26) قيمة منها (17) قيمة خاصة بالإناث و(9) قيم خاصة بالذكور، حيث كانت جميع القيم أعلى من قيمة الحد الأعلى، في حين لم تكن هناك أي قيمة متطرفة أقل من الحد الأدنى؛ لكونه كان أقل من الصفر. والمتصفح لهذه القيم يجد بأنها جاءت غير متسقة مع مجموعة البيانات، وكانت بعيدة كثيرا عن الوسط الحسابي للعلامات (6.93)؛ مما جعل عدد العلامات المتطرفة كبيرا. ويبدو أن هؤلاء الطلبة أيضا كانت قدراتهم عالية بحيث حصلوا على أعلى العلامات؛ مما جعل علاماتهم متطرفة عن بقية العلامات، خاصة أن أداء الطلبة الأردنيين على اختبار الرياضيات في الدراسة الدولية (TIMSS 2011)) قد تراجع كثيرا عن أداء الدورة السابقة (TIMSS 2007)، حيث تراجع الأداء بمقدار 21 علامة (المركز الوطني لتنمية الموارد البشرية، 2013). وباعتقاد الباحث بأن مثل هذا التراجع قد يكون أحد الأسباب التي أدت إلى انخفاض الوسط الحسابي، وكثرة العلامات المتطرفة في بيانات الكراس الثاني لفقرات اختبار الرياضيات في الدراسة الدولية.

أما بالنسبة للسؤال الثاني فقد تم استخدام برنامج DFIT8 للكشف عن الأداء التفاضلي للفقرات بوجود القيم المتطرفة تبعا لمتغير الجنس، ومن أجل ذلك فقد تم اعتماد مجموعة الطالبات كمجموعة مرجعية ومجموعة الطلاب كمجموعة مستهدفة أثناء

تشغيل برنامج Equate؛ وذلك لأن برنامج DIFT8 يتعامل مع المجموعة المرجعية كمجموعة مستهدفة والمجموعة المستهدفة كمجموعة مرجعية بعكس برنامج Equate، وبعد استخدام البرنامج لغايات معادلة المعالم فقد تم الحصول على ثوابت المعادلة حيث كانت قيمة الثابت الضربي (A=1.1487) وقيمة الثابت الجمعي (K=0.114) على اعتبار أن جميع فقرات الاختبار (الكراس الثاني) المؤلف من (26) فقرة كانت بمثابة فقرات الجذع المشترك.

وبهدف تنفيذ المرحلة الأولى للكشف عن الأداء التفاضلي فقد تم تنفيذ برنامج DIFT باستخدام معاملي الربط السابقين، حيث يستخدم طريقة تكرار توليد معلمة الفقرة من أجل حساب نقاط القطع الخاصة بكل فقرة من فقرات الاختبار؛ وذلك للحكم على الأداء التفاضلي لكل فقرة بالاعتماد على تلك القيم يتم الحكم على أن الفقرة تبدي أداء تفاضليا أم لا من خلال مؤشر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF). وتجدر الإشارة كذلك بأنه يتم تسجيل الرمز (ns) في مخرجات البرنامج في العمود الخاص ليدل على عدم وجود دلالة إحصائية (not significant) إن لم تبد الفقرة أداءً تفاضليا دالا إحصائيا. ويوضح الجدول (6) الدلالات الإحصائية للأداء التفاضلي للفقرات على مستوى الفقرة والاختبار للمرحلة الأولى.

جدول 6: الدلالات الإحصائية للأداء التفاضلي لفقرات الاختبار المرحلة الأولى بوجود القيم المتطرفة

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
1	M052231	0.139	0.105	0.139	0.105	0.001	0.18223	0.03036	0.005
2	M052061	-0.046	0.057	0.046	0.057	-0.019	-0.07982	0.00544	ns
3	M052228	0.158	0.211	0.171	0.200	0.048	0.25405	0.06932	0.001
4	M052214	0.075	0.027	0.075	0.027	0.009	0.10631	0.00631	ns
5	M052173	-0.016	0.022	0.02	0.018	0.003	-0.01778	0.00074	ns
6	M052362	-0.097	0.161	0.123	0.143	-0.052	-0.17893	0.03551	0.01
7	M052408	-0.04	0.082	0.068	0.061	-0.026	-0.07864	0.00829	ns
8	M052084	0.079	0.129	0.082	0.127	0.033	0.13622	0.02282	0.01
9	M052206	0.004	0.050	0.018	0.047	0.010	0.0144	0.00252	ns
10	M052429	0.072	0.071	0.076	0.066	0.004	0.09797	0.01013	ns
11	M052503A	0.030	0.022	0.035	0.012	-0.007	0.03118	0.00134	ns
12	M052503B	0.057	0.013	0.057	0.013	0.003	0.07702	0.00337	ns
13	M042031	0.077	0.043	0.077	0.043	0.001	0.10261	0.00788	ns
14	M042186	0.006	0.077	0.066	0.041	0.001	0.00798	0.00603	ns
15	M042236	0.113	0.121	0.113	0.121	0.022	0.16911	0.02730	0.05
16	M042226	0.127	0.131	0.127	0.131	0.007	0.17308	0.03330	0.005
17	M042103	0.045	0.104	0.045	0.104	0.028	0.08677	0.01274	0.05
18	M042086	-0.061	0.059	0.061	0.059	-0.019	-0.09848	0.00716	ns
19	M042228	0.053	0.047	0.066	0.028	-0.016	0.05378	0.00509	ns
20	M042245	0.087	0.094	0.087	0.094	0.029	0.14239	0.01642	0.05
21	M042270	0.030	0.016	0.030	0.016	0.005	0.04436	0.00117	ns
22	M042201	0.101	0.136	0.101	0.136	0.040	0.17220	0.02862	0.005
23	M042152	0.134	0.042	0.134	0.042	-0.011	0.16392	0.01959	ns
24	M042269	0.021	0.039	0.034	0.028	0.012	0.03935	0.00194	ns

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
25	M042179	0.124	0.049	0.124	0.049	0.015	0.17719	0.01767	ns
26	M042177	0.038	0.066	0.04	0.065	0.015	0.06434	0.00580	ns

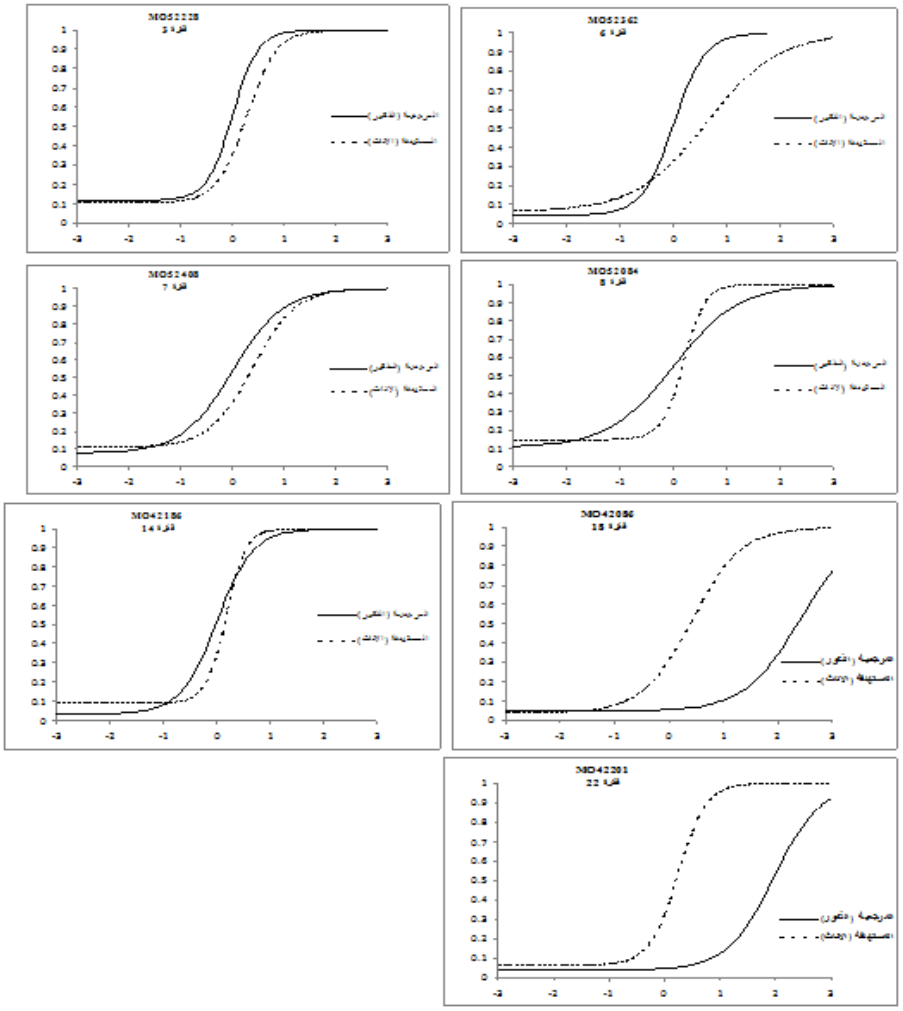
يتضح من نتائج الجدول (6) بأن الفقرات (1 ، 3 ، 6 ، 8 ، 15 ، 16 ، 17 ، 20 ، 22) قد أبدت أداء تفاضليا عند مستوى الدلالة (α = 0.05) أو أقل، ولذلك سيتم استثناء هذه الفقرات من فقرات الجذع المشترك لإيجاد معامل الثابت الضربي والجمعي من أجل تنفيذ المرحلة الثانية للكشف عن الأداء التفاضلي للفقرات بشكل نهائي. ولتحقيق ذلك فقد تم إعادة معادلة القيم المقدرة لمعالم الفقرات باستخدام برنامج Equate بعد استبعاد الفقرات التي أبدت أداء تفاضليا في المرحلة الأولى للحصول على معاملي ربط جديدين، حيث بلغت قيمة معامل الثابت الضربي (A= 1.1430)، ومعامل الثابت الجمعي (K= 0.2032). وقد تم استخدام هذين المعاملين في تنفيذ المرحلة الثانية باستخدام برنامج DIFT8 ومعالم الفقرات للمجموعتين. والجدول (7) يوضح الدلالات الإحصائية للأداء التفاضلي على مستوى الفقرة لفقرات اختبار الرياضيات المرحلة الثانية.

جدول 7: الدلالات الإحصائية للأداء التفاضلي لفقرات الاختبار المرحلة الثانية بوجود القيم المتطرفة

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
1	M052231	0.075	0.050	0.075	0.05	-0.011	0.02827	0.00821	Ns
2	M052061	-0.075	0.090	0.075	0.09	0.022	-0.01719	0.01367	ns
3	M052228	0.115	0.158	0.128	0.148	-0.071	-0.01108	0.03821	0.001
4	M052214	0.068	0.017	0.068	0.017	0.001	0.03607	0.00491	ns
5	M052173	-0.016	0.021	0.021	0.017	0.003	-0.00588	0.00071	ns
6	M052362	-0.164	0.206	0.182	0.19	0.081	-0.00376	0.06921	0.001
7	M052408	-0.079	0.118	0.104	0.097	0.043	0.00177	0.02022	0.05
8	M052084	0.055	0.115	0.074	0.103	0.014	0.04209	0.01616	0.05
9	M052206	-0.004	0.038	0.018	0.034	0.009	0.00675	0.00150	ns
10	M052429	0.048	0.076	0.075	0.050	0.033	0.05727	0.00806	ns
11	M052503A	0.019	0.034	0.034	0.020	0.003	0.01278	0.00155	ns
12	M052503B	0.05	0.008	0.050	0.008	0.002	0.02844	0.00259	ns
13	M042031	0.045	0.060	0.066	0.035	0.027	0.05044	0.00563	ns
14	M042186	-0.059	0.140	0.103	0.111	0.041	0.01089	0.02299	0.005
15	M042236	0.064	0.052	0.064	0.052	-0.021	0.01217	0.00688	ns
16	M042226	0.057	0.061	0.058	0.060	-0.014	0.01479	0.00691	ns
17	M042103	0.035	0.084	0.035	0.084	0.011	0.02867	0.00822	ns
18	M042086	-0.096	0.093	0.096	0.093	0.032	-0.01724	0.01781	0.05
19	M042228	0.033	0.071	0.070	0.034	0.015	0.03180	0.00612	ns
20	M042245	0.059	0.056	0.059	0.056	-0.012	0.01869	0.00663	ns
21	M042270	0.002	0.013	0.012	0.006	0.004	0.00437	0.00017	ns

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
22	M042201	0.072	0.103	0.072	0.103	-0.011	0.02579	0.01581	0.05
23	M042152	0.106	0.059	0.106	0.059	0.026	0.08012	0.01469	ns
24	M042269	-0.001	0.014	0.013	0.006	-0.006	-0.00677	0.00020	ns
25	M042179	0.087	0.041	0.087	0.041	-0.002	0.04343	0.00923	ns
26	M042177	0.020	0.060	0.040	0.049	0.019	0.02892	0.00399	ns

يظهر من النتائج الواردة في الجدول (7) وجود (7) فقرات أبدت أداء تفاضليا على مستوى الفقرات من أصل (26) فقرة تبعا لمتغير جنس الطالب بنسبة مئوية (27%) من الاختبار الكلي، وكانت الفقرات هي (3 ، 6، 7 ، 8 ، 14، 18، 22). وقد أبدت الفقرات (6 ، 7 ، 18 ، 22) أداء تفاضليا تبعا لمتغير الجنس عند مستوى الدلالة (α = 0.05)، في حين أبدت الفقرتان (3 ، 6) أداء تفاضليا عند مستوى الدلالة (α = 0.001)، والفقرة (14) أبدت أداء تفاضليا عند مستوى الدلالة (α = 0.005). ولتحديد نوع الأداء التفاضلي لهذه الفقرات فقد تم رسم منحنى خصائص الفقرة لكل فقرة من الفقرات التي أبدت أداء تفاضليا، والشكل (1) يوضح منحنى خصائص هذه الفقرات.



الشكل (1): منحنى خصائص الفقرات التي أبدت أداء تفاضليا تبعا لجنس الطالب

يتضح من الشكل (1) بأن الفقرات التي أرقامها (3 ، 7) قد أبدت أداء تفاضليا منتظما لصالح المجموعة المرجعية (الذكور)، حيث يتضح أن احتمالات الاستجابة الصحيحة على هاتين الفقرتين كانت أعلى للذكور من الإناث عند جميع مستويات القدرة، في حين أبدت الفقرتان (18، 22) أداء تفاضليا لصالح الإناث، بحيث أن احتمالات الاستجابة الصحيحة على هاتين الفقرتين كان أعلى للإناث من الذكور عند جميع مستويات القدرة. ويتضح كذلك من الشكل (1) بأن الفقرات (6 ، 8 ، 14) قد أبدت أداء تفاضليا غير منتظم بالنسبة لاحتمالات الاستجابة الصحيحة، حيث يتضح بأن الفقرة (6) كانت احتمالات الاستجابة الصحيحة بالنسبة للذكور أعلى منها للإناث ضمن مدى القدرة (3 , 0.5-)، في حين كانت لصالح الإناث في مستويات القدرة المتدنية. ويظهر كذلك من الشكل نفسه بأن الفقرتين (8،14) قد أظهرتا فروقا متمايزة بين الذكور والإناث في احتمالات الاستجابة الصحيحة، حيث كانت تعملان لصالح الذكور ضمن مستويات القدرة المتوسطة، ولصالح الإناث ضمن مستويات القدرة العالية والمتدنية.

وبالنظر إلى الفقرات التي أبدت أداء تفاضليا من خلال نتائج الجدول (7) والشكل (1) بأن الفقرتين (3، 7) قد أبدتا أداء تفاضليا منتظما لصالح الذكور عند جميع مستويات القدرة، وبالرجوع إلى محتوى الفقرتين نجد بأن متن الفقرة (3) التي تحمل الرقم (M052228) في الكراس الأصلي يتعلق بإجراء العمليات الحسابية على الكسور العادية ضمن محور الأعداد، بينما كان متن الفقرة (7) ورقمها في الكراس الأصلي (M052408) ينصب على إجراء القياسات المجهولة في الأشكال الهندسية الرباعية اعتمادا على معرفة العلاقات بين الزوايا في الأشكال الهندسية ضمن محور الهندسة. وقد جاءت هذه النتيجة متفقة مع الاتجاه العام للدراسات التي تناولت الفروق التحصيلية بين الإناث والذكور بأن أداء الذكور بشكل عام أفضل من أداء الإناث ضمن محور الهندسة والأعداد (Engelhard & Gamer, 1999; Haries & Carlton, 1993)، حيث إن الإناث يعتمدن في الغالب على الاستخدام الروتيني للقوانين المتعلمة داخل الغرفة الصفية، بينما الذكور لديهم أساليب ذاتية تتيح لهم تعميم المعرفة على مسائل غير مألوفة، خصوصا أن هاتين الفقرتين تقيسان مستويات معرفية عليا كالتطبيق، والتبرير الرياضي، وتتطلبان استخدام استراتيجيات خوارزمية غير روتينية.

وبينت النتائج كذلك بأن الفقرتين (18، 22) قد أبدتا أداء تفاضليا منتظما لصالح الإناث عند جميع مستويات القدرة، وبالنظر إلى محتوى الفقرتين نجد بأن متن الفقرة (18، M04286) جاء ضمن سياق إيجاد قيمة أحد المتغيرات في معادلة ما إذا أعطيت القيمة العددية للمتغيرات الأخرى ضمن محور الجبر. وجاءت هذه النتيجة متفقة مع نتائج الدراسات السابقة (Dodeen & Annabi, 2006 ; Engelhard & Gamer, 1999) التي أشارت بأن الفقرات المتعلقة بموضوع الجبر كانت أسهل لدى الإناث من الذكور، وكذلك حل مثل هذه المسألة يتطلب إجراءات روتينية مماثلة للمواقف الصفية. بينما كانت الفقرة (M02201 ,22) التي يركز متنها على إيجاد القيمة المجهولة في الشكل الهندسي المعطى ضمن محور الهندسة. وجاءت هذه النتيجة متعارضة مع دراسة (Haris & .Carlton, 1993) ويعزو الباحث السبب في ذلك إلى جدية الإناث في الإجابة عن أسئلة الفقرات، وهذا يبدو واضحا من الشكل (1)، حيث إن هناك فروقا عالية في احتمالات الاستجابة الصحيحة عن الفقرة عند جميع مستويات القدرة، وكذلك كان الوسط العام للإناث أفضل من الذكور وعدد العلامات العالية بالنسبة للإناث أكثر من الذكور كما هو وارد في نتائج الجدول (4) التي اعتبرت قيما متطرفة. كذلك ليس مسلما به دائما بأن أداء الذكور بالهندسة أفضل من الإناث؛ لأنه ربما تكون عوامل أخرى تظهر الأداء التفاضلي للفقرات.

وقد بينت النتائج كذلك بأن الفقرات (6 ، 8، 14) التي تحمل الأرقام (M052362, M042186, M042084) على التوالي في الكراس الأصلي قد أظهرت فروقا غير منتظمة في احتمالات الاستجابة الصحيحة بين الذكور والإناث، فمحتوى الفقرة (6) يتركز حول إيجاد قيمة زاوية بالاعتماد على حالات التطابق في المثلثات ضمن محور الهندسة، حيث يتضح بأن احتمالات الاستجابة الصحيحة للذكور كانت أعلى وبشكل منتظم عند جميع مستويات القدرة المتوسطة والعالية، وهي بهذا كانت متفقة مع الاتجاه العام للدراسات بأن موضوع الهندسة أسهل للذكور منه للإناث (Dodeen & Annabi, 2006)، وكأنها أبدت أداء تفاضليا منتظما، حيث كانت الفروق في احتمالات الاستجابة الصحيحة عن هذه الفقرة ضئيلة جدا عند مستويات القدرة المتدنية. وبينت النتائج كذلك بأن الفقرة (8) التي تضمنت إيجاد مساحة المربع من خلال معرفة محيطه، كانت تعمل لصالح الذكور ضمن مستوى القدرة (0.00 ,2.00-)، ولصالح الإناث في مستوى القدرة العالية على الرغم من أن محتوى الفقرة يقع ضمن محور الهندسة. وهذا يبدو واضحا في جدية الإناث، وحصولهن على علامات أعلى من الذكور التي جعلت مثل هذه العلامات تبدو متطرفة بالنسبة لمجموعة البيانات، وانعكس بالتالي على احتمالات الاستجابة الصحيحة عن الفقرة، ولكن بالنظر إلى الفقرة (14)

التي جاءت ضمن سياق معرفة النمط الذي يتوافق مع أنماط محتوى الفقرة وتقع ضمن محور الأعداد، على الرغم من أنها أبدت أداء تفاضليا غير منتظم، حيث عملت لصالح الذكور ضمن مستوى القدرة (0.00 ,1.00-)، إلا أن الفروق في احتمالات الاستجابة الصحيحة عنها ضئيلة ومتقاربة ضمن المستويات الأخرى من القدرة، وعلى الرغم من دلالتها الإحصائية، وهذا يبدو من معاملات الصعوبة للفقرة لكل من الذكور والإناث، إلا أنها كانت أكثر تمييزا لدى الإناث.

ومن أجل الكشف عن أثر القيم المتطرفة على الأداء التفاضلي لفقرات الاختبار وذلك للإجابة عن السؤال الثالث، فقد تم حذف القيم المتطرفة الواردة في الجدول (3) من مجموعة بيانات الذكور والإناث، وتم إعادة تقدير معالم الفقرات والقدرة لكل من الذكور والإناث باستخدام النموذج اللوجستي ثلاثي المعلمة من أجل وضعها على تدريج مشترك لغايات إيجاد معاملي الثابت الضربي والجمعي. وبنفس الإجراءات التي استخدمت في الإجابة عن السؤال الثاني فقد تم استخدام برنامج Equate لإيجاد المعامل الضربي والجمعي للمعادلة، حيث كانت قيمة المعامل الضربي (A=1.1487) وقيمة الثابت الجمعي (k= -0.1140) بعد اعتبار جميع فقرات الاختبار(26 فقرة) كجذع مشترك. ولتنفيذ المرحلة الأولى لبرنامج DIFT8 من أجل إيجاد الفقرات التي تبدي أداء تفاضليا باستخدام المعاملين السابقين، فقد تم الحصول على الدلالات الإحصائية للأداء التفاضلي لفقرات اختبار الرياضيات بعد حذف القيم المتطرفة للمرحلة الأولى، والجدول (8) يوضح هذه الدلالات.

جدول 8: الدلالات الإحصائية للأداء التفاضلي لفقرات الاختبار المرحلة الأولى بعد حذف القيم المتطرفة

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
1	M052231	0.108	0.063	0.108	0.063	-0.021	0.08445	0.01562	ns
2	M052061	-0.031	0.041	0.038	0.035	0.001	-0.02885	0.00264	ns
3	M052228	0.147	0.188	0.161	0.175	-0.021	0.12268	0.05682	0.001
4	M052214	0.076	0.030	0.076	0.030	0.006	0.08045	0.00669	ns
5	M052173	-0.016	0.001	0.016	0.001	0.000	-0.01594	0.00027	ns
6	M052362	-0.108	0.161	0.136	0.139	0.004	-0.10139	0.03772	0.05
7	M052408	-0.071	0.131	0.097	0.113	-0.010	-0.07912	0.02221	ns
8	M052084	0.027	0.055	0.037	0.049	0.016	0.04178	0.00379	ns
9	M052206	-0.006	0.011	0.007	0.010	-0.001	-0.00675	0.00015	ns
10	M052429	0.040	0.058	0.063	0.031	0.003	0.04233	0.00496	ns
11	M052503A	0.033	0.012	0.033	0.012	0.001	0.03345	0.00127	ns
12	M052503B	0.056	0.005	0.056	0.005	0.001	0.05495	0.00313	ns
13	M042031	0.046	0.050	0.057	0.038	0.009	0.05440	0.00464	ns
14	M042186	-0.030	0.116	0.092	0.077	0.034	0.00495	0.01444	0.05
15	M042236	0.098	0.096	0.098	0.096	-0.010	0.08510	0.01872	0.05
16	M042226	0.105	0.090	0.105	0.090	-0.026	0.07622	0.01905	0.05
17	M042103	0.018	0.020	0.018	0.020	0.004	0.02148	0.00070	ns
18	M042086	-0.038	0.059	0.051	0.048	0.017	-0.01983	0.00491	ns
19	M042228	0.049	0.074	0.078	0.041	-0.009	0.03872	0.00781	ns
20	M042245	0.066	0.067	0.066	0.067	0.010	0.07504	0.00892	ns
21	M042270	0.039	0.038	0.039	0.038	0.010	0.04746	0.00292	ns
22	M042201	0.108	0.171	0.108	0.171	0.032	0.13741	0.04100	0.001

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
23	M042152	0.140	0.062	0.140	0.062	0.019	0.15582	0.02346	ns
24	M042269	-0.010	0.077	0.070	0.034	0.025	0.01538	0.00609	ns
25	M042179	0.100	0.053	0.100	0.053	0.015	0.11179	0.01271	ns
26	M042177	0.030	0.052	0.038	0.047	0.016	0.04530	0.00366	ns

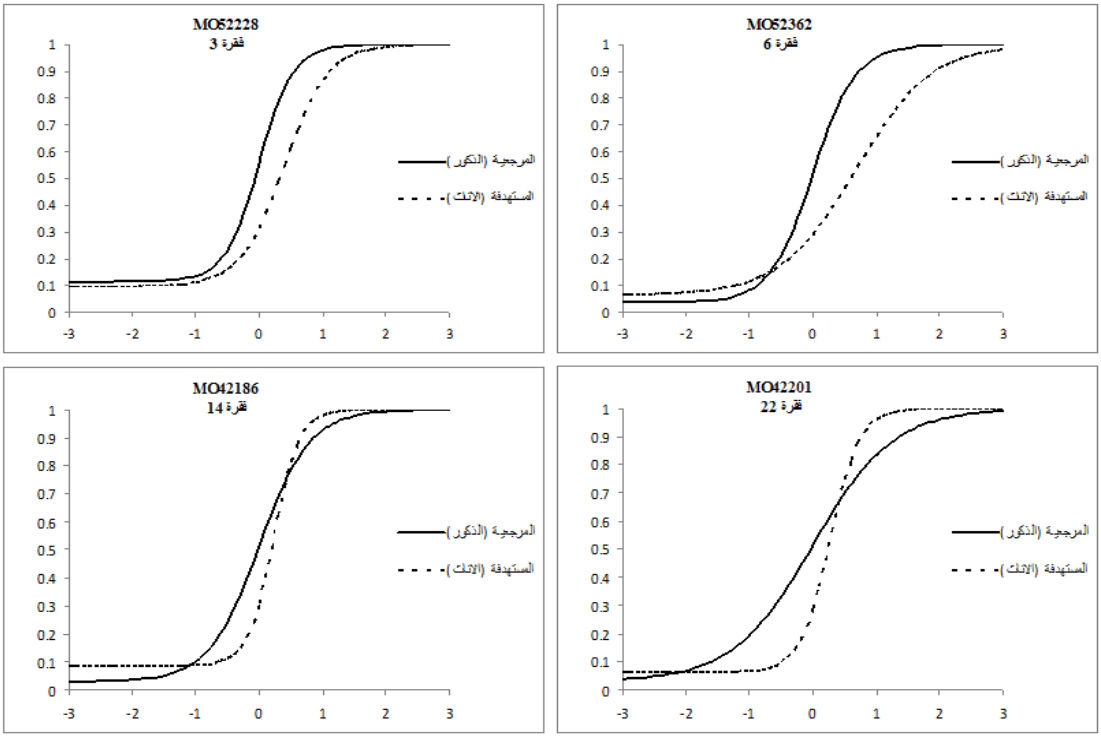
يتضح من النتائج الواردة في الجدول أن الفقرات التي تحمل الأرقام (3، 6، 14، 15، 16، 22) قد أبدت أداء تفاضليا في المرحلة الأولى عند مستوى دلالة إحصائية اقل من ($\alpha = 0.05$) بعد حذف القيم المتطرفة، حيث اشتركت الفقرات (3 ، 6، 15 ، 16، 22) في المرحلة الأولى بدون حذف القيم المتطرفة، ولكن عدد الفقرات التي أبدت أداء تفاضليا في المرحلة بعد حذف القيم المتطرفة قد قل عما هو في المرحلة الأولى بوجود القيم المتطرفة. ولغايات تنفيذ المرحلة الثانية تم استبعاد تلك الفقرات التي أبدت أداء تفاضليا من فقرات الاختبار كجذع مشترك، وتم إعادة معادلة معالم الفقرات لكلا المجموعتين للحصول على معاملي الضربي والجمعي، حيث كانت قيمة المعامل الضربي (A= 1.1862)، وقيمة المعامل الجمعي (K= -0.1981). وقد تم استخدام هذين المعاملين لإيجاد الفقرات التي تبدي أداء تفاضليا بشكل نهائي بعد تنفيذ المرحلة الثانية للحصول على الدلالات الإحصائية للأداء التفاضلي لفقرات اختبار الرياضيات بعد حذف القيم المتطرفة. والجدول (9) يوضح هذه الدلالات.

جدول 9: الدلالات الإحصائية للأداء التفاضلي لفقرات الاختبار المرحلة الثانية بعد حذف القيم المتطرفة

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
1	M052231	0.088	0.046	0.088	0.046	-0.020	0.04619	0.00991	ns
2	M052061	-0.036	0.045	0.042	0.040	0.009	-0.01828	0.00335	ns
3	M052228	0.135	0.174	0.150	0.161	-0.050	0.05251	0.04858	0.001
4	M052214	0.075	0.029	0.075	0.029	0.005	0.06138	0.00649	ns
5	M052173	-0.017	0.001	0.017	0.001	0.000	-0.01227	0.00028	ns
6	M052362	-0.129	0.175	0.154	0.153	0.038	-0.05859	0.04734	0.01
7	M052408	-0.083	0.141	0.108	0.123	0.010	-0.05251	0.02681	ns
8	M052084	0.019	0.055	0.038	0.044	0.019	0.03373	0.00338	ns
9	M052206	-0.007	0.012	0.008	0.011	0.000	-0.00499	0.00019	ns
10	M052429	0.034	0.063	0.065	0.031	0.014	0.03926	0.00516	ns
11	M052503A	0.031	0.015	0.031	0.015	0.003	0.02637	0.00116	ns
12	M052503B	0.054	0.007	0.054	0.007	0.002	0.04242	0.00295	ns
13	M042031	0.037	0.057	0.059	0.034	0.020	0.04860	0.00463	ns
14	M042186	-0.050	0.135	0.105	0.098	0.053	0.01539	0.02058	0.005
15	M042236	0.084	0.076	0.084	0.076	-0.020	0.04285	0.01281	ns
16	M042226	0.083	0.067	0.083	0.067	-0.029	0.03327	0.01137	ns
17	M042103	0.015	0.014	0.015	0.014	0.003	0.01403	0.00042	ns
18	M042086	-0.046	0.063	0.057	0.053	0.026	-0.00887	0.00606	ns

رقم الفقرة	رقم الفقرة في الكراس	d		|d|		C(d,D)	CDIF	NCDIF	الدلالة الإحصائية
		الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري				
19	M042228	0.043	0.079	0.078	0.044	0.000	0.03273	0.00807	ns
20	M042245	0.061	0.062	0.061	0.062	0.005	0.05094	0.00759	ns
21	M042270	0.031	0.036	0.032	0.036	0.010	0.03401	0.00228	ns
22	M042201	0.103	0.168	0.103	0.168	0.021	0.09878	0.0389	0.001
23	M042152	0.132	0.066	0.132	0.066	0.030	0.12943	0.02167	ns
24	M042269	-0.019	0.077	0.071	0.035	0.034	0.01917	0.00633	ns
25	M042179	0.088	0.055	0.088	0.055	0.016	0.08277	0.01076	ns
26	M042177	0.026	0.053	0.039	0.044	0.020	0.03957	0.00348	ns

يتضح من النتائج الواردة في الجدول (8) بأن الفقرات (3 ، 6 ، 14 ، 22) قد أبدت أداء تفاضليا بعد حذف القيم المتطرفة عند مستوى دلالة أقل من ($\alpha = 0.05$). ولتحديد نوع الأداء التفاضلي للفقرات فقد تم رسم منحنى خاصية الفقرة لكل فقرة من الفقرات السابقة، والشكل (2) يوضح هذه المنحنيات.



الشكل (2): منحنى خصائص الفقرات التي أبدت أداء تفاضليا تبعا لجنس الطالب بعد حذف القيم المتطرفة

يتضح من الشكل (2) بأن الفقرة (3) قد أبدت أداء تفاضليا منتظما لصالح مجموعة الذكور، في حين أبدت الفقرات (6 ، 14، 22) أداء تفاضليا غير منتظم، حيث تمايزت احتمالات الاستجابة الصحيحة بين الذكور والإناث ضمن مستويات القدرة المختلفة. وبمقارنة النتائج الواردة في الشكل (1)، والشكل (2) يتضح بأنه بعد حذف القيم المتطرفة من مجموعة البيانات للذكور والإناث فقد تناقص عدد الفقرات التي أبدت أداء تفاضليا بمقدار ثلاث فقرات من أصل (7) فقرات ليصبح عدد الفقرات التي أبدت أداء

تفاضليا بنوعيه المنتظم وغير المنتظم أربع فقرات بنسبة مئوية (15%) من الاختبار الكلي، حيث تراجعت هذه النسبة بمقدار (12%) عندما أدخلت القيم المتطرفة في التحليل، وهو ما يؤكد أثر القيم المتطرفة على الأداء التفاضلي للفقرات، حيث زاد عدد الفقرات التي تظهر أداء تفاضليا بوجود القيم المتطرفة في مجموعة البيانات، وتقلص عدد الفقرات التي تظهر أداء تفاضليا بعد حذف هذه القيم المتطرفة من البيانات.

ومما يؤكد كذلك أثر القيم المتطرفة على الأداء التفاضلي للفقرات منحنى خاصية الفقرة (22) الوارد في الشكل (1)، حيث أبدت هذه الفقرة أداء تفاضليا لصالح الإناث قبل حذف القيم المتطرفة على الرغم من أن محتوى الفقرة جاء ضمن سياق المساحات لتغطية محور الهندسة التي تعارضت مع نتيجة دراسة (Haris & Carlton, 1993) بأن المسائل المتعلقة بالهندسة تكون أسهل للذكور من الإناث، ولكن بعد حذف القيم المتطرفة أظهرت هذه الفقرة أداء تفاضليا غير منتظم، حيث عملت لصالح الذكور ضمن مستوى القدرة (0.0 ، 3 −)، ولصالح الإناث ضمن مستوى القدرة (0.0 ، 3). وربما يعزى سبب ذلك إلى عدد القيم المتطرفة التي كان عددها في الإناث (17) قيمة، بينما في الذكور (9) قيم الذي بدوره انعكس على خصائص الفقرة، وكذلك بالنظر إلى المجال المعرفي لهذه الفقرة، فهي تقيس مستوى التطبيق تتطلب استراتيجيات خوارزمية عادية في الحل لكون ارتباط حل مثل هذه المسائل بالطرق المألوفة بالحل تكون الأفضلية للإناث.

وفي السياق نفسه بالنظر إلى الفقرة (18) في الشكل (1) التي كانت تظهر أداء تفاضليا لصالح الإناث ضمن جميع مستويات القدرة وبفارق واضح ومتمايز في احتمالات الاستجابة الصحيحة عنها قبل حذف القيم المتطرفة، ولكن بعد حذف القيم المتطرفة أصبحت هذه الفقرة لا تبدي أداء تفاضليا. وهذا أيضا يؤكد بأن وجود القيم المتطرفة في البيانات ينعكس على الأداء التفاضلي للفقرات، ومثل هذه النتيجة تؤكد حاجة البحث السيكومتري إلى دراسة أثر القيم المتطرفة على تقديرات معالم الفقرات ودقتها كون هذا الهدف لم يكن من أهداف الدراسة الحالية .

أما بالنسبة للفقرتين (6 ، 14) فقد أظهرتا أداء تفاضليا غير منتظم تبعا لمتغير الجنس بعد حذف القيم المتطرفة، حيث أظهرت الفقرة (6) أداء منتظما لصالح الذكور وبشكل متمايز ضمن مستوى القدرة (0.5 − ، 3)، ولصالح الإناث ضمن مستوى القدرة المتدني وبفروق ضئيلة لاحتمالات الاستجابة الصحيحة عنها بين نفس مستوى القدرة للذكور. ويمكن عزو تلك الفروق الضئيلة لهذه الفئة من القدرة لجوء هذه الفئة للتخمين للوصول إلى الإجابة الصحيحة. وتجدر الإشارة كذلك بأن الفقرة (14) المتعلقة بمحور الأعداد وسياقها معرفة النمط المتماثل قد أظهرت أداء تفاضليا غير منتظم بفروق قليلة في احتمالات الاستجابة الصحيحة بين الذكور والإناث بعد حذف القيم المتطرفة، ولكن بشكل عام كشفت النتائج وجود نوع من الاتساق في الأداء التفاضلي للفقرتين (6 ، 14) قبل وبعد حذف القيم المتطرفة؛ وربما يعود ذلك بأن المجال المعرفي للفقرتين هو التبرير الرياضي، وبالتالي لحل مثل هاتين المسألتين يتطلب إستراتيجيات غير روتينية جعل هذا الأداء على الفقرتين للذكور والإناث متسقا سواء قبل أو بعد حذف القيم المتطرفة.

وعلى الرغم من ظهور الأداء التفاضلي إلا أنه يمكن القول بشكل عام بأنه لا توجد قاعدة عامة يمكن من خلالها تحديد حجم الأداء التفاضلي واتجاه بشكل ثابت لمجموعة ضد الأخرى، حيث كانت الفروق في احتمالات الإجابة الصحيحة متمايزة حسب مستوى القدرة كما أظهر الشكلان (1، 2)، وربما يكون هذا التمايز له مصادره الخاصة، وأن هذا الاختبار من الاختبارات العالمية التي تخضع فقراته لترجمات مختلفة، ولسياسيات تعليمية مختلفة، ومناهج مختلفة وكفاءات تدريسية من قبل المعلمين متنوعة. ونقطة أخرى جديرة بالذكر من خلال نتائج الدراسة تبين بالإضافة إلى العوامل التي تبدو ظاهريا بأنها مسؤولة عن الأداء التفاضلي للفقرات وجود القيم المتطرفة في البيانات، حيث اتضح مثل هذا الأثر بعد حذف القيم المتطرفة بتناقص الفقرات التي تبدي أداء تفاضليا، وقد يعني ذلك أن القيم المتطرفة قد تكون أحد الأسباب التي تؤدي إلى الأداء التفاضلي للفقرات، بالإضافة إلى العوامل الأخرى التي كانت محور كثير من دراسات الأداء التفاضلي، وعلى الرغم من قبولها ظاهريا كأحد الأسباب استنادا إلى نتائج هذه الدراسة؛ مما يستدعي الحاجة لإجراء دراسات إضافية تهتم بدراسة أثر القيم المتطرفة على الأداء التفاضلي لتدعيم نتائج البحث الحالي. وكذلك على المنظمة الدولية لتقييم التحصيل التربوي (IEA) مراعاة القيم المتطرفة وأخذها بالاعتبار عند تحليل نتائج اختباراتها العالمية العلوم والرياضيات للصفين الثامن والرابع .

وتجدر الإشارة كذلك بأن الباحث تناول طريقة واحدة من طرق معالجة القيم المتطرفة، وهي طريقة الحذف، وطريقة استكشافية واحدة هي طريقة توكي، وبالتالي يتطلب ذلك أيضا إجراء مزيد من الدراسات تتناول الطرق الأخرى لكشف القيم المتطرفة وطرق معالجتها وأثر ذلك على الأداء التفاضلي للفقرات..

المراجع

أبو حماد، ك . (2008). تقصي فاعلية منحنى خصائص الفقرة في الكشف عن الأداء التفاضلي لمتغير الجنس في عينة مختارة من اختبار الرياضيات. رسالة ماجستير غير منشورة، جامعة اليرموك، الأردن.

المركز الوطني لتنمية الموارد البشرية. (2013). التقرير الوطني الأردني عن الدراسة الدولية للرياضيات والعلوم 2011 (TIMSS 2011). وحدة المتابعة والتقييم: عمان.

Ampanthong, P., & Suwttee, P.(2009). A Comparative Study of Outlier Detection Procedure in Multiple Liner Regression. The Proceeding of the International Multiconference of Engineers and Computers Scientists, IMECS, Hong Kong.

Baker, F .(1995). EQUATE (version 2.1): Computer Program for Equating two Metrics in Item Response Theory [Computer Software].Madison, WI: University of Wisconsin, Laboratory of Experimental Design.

Barnett, V., & Lewis, T .(1994). Outlier in statistical data (3ed ed).New York: John Wiley and Sons.

Douden, H., & Annab, H. (2008). Sex-Related Differential Item Functioning (DIF) Analysis of TIMSS. Dirasat , Educational Science. 35,697-705.

Ellis, B ., & Raju, S. (2003).Test and Item Bias: What they are, What aren't, and How to Detect them. (ERIC Document Reproduction Service No, ED480042).

Engelhard, G ., & Gamer, M .(1999). Gender difference in Performance on Multiple-Choice and constructed Response Mathematics Items. Applied Measurement in Education,12,29-43.

Evans, V. (1999). Strategies for detecting outliers in regression analysis: An introductory primer. In B. Thompson (Ed.), Advances in social science methodology: (Vol. 5, pp. 213-233). Stamford, CT.: JAI Press.

Friedman, L. (1996). Meta-analysis and quantitative gender differences: Reconciliation. Focus on Learning Problems in Mathematics, 18, 123-128

Gierl, M., Jodoin, M., & Ackerman, T.(2000). Performance of Mantel-Haenzel, Simultaneous Item Bias Test, and Logistic Regression when the Proportion of DIF Items is Large. Paper presented at the Annual Meeting of the American Educational Research Association, New Orland's , Louisiana,USA.

Green, R. (1976).Outlier-prone and Outlier-resistant distribution. Journal of the American Statistical Association,71,502-505.

Gybles, J. (2004). The Performance of some Observed and Unobserved Conditional Invariance Techniques for the Detection of Differential Item Functioning . Quality of Quantity, 38,681-702.

Hawkins, D . (1980). Identification of outliers. London: Chapman and Hall.

Hambleton, R ., & Rogers, J.(1989). Detecting Potentially Bias Test Items Comparison of IRT area and Mantel-Haenzel Methods. Applied Measurement in Eduacation.2 ,313-334.

Hambleton, R., & Swaminathan, H. (1985). Item Response Theory Principles and Application. Boston: Kluwer, Nijh off Publishing.

Harries, A., & Carlton, S.(1993). Patterns of gender difference on mathematics items on the Scholastic Aptitude Test. Applied Measurement in Education, 6,173-151.

Hattie, J. (1985).Methodology Review: Assessing unidimensionality of tests and items. Applied Psychological Measurement.9 ,139-164.

Jensen, R. (1980). Bias mental testing. New York: Macmillan publishing CO. inc.

Johnson, S. (1996). The contribution of large-scale assessment programs to research on gender differences. Educational Research and Evaluation, 2, 25-49.

Johnson, R., & Wichern, D. (2007) Applied multivariate statistical analysis 6th ed. New Jersey: Person Prentice Hall.

Mendes, B., & Erickan, K. (2006). Examining Sources of Gender DIF Mathematics using a Confirmatory Multi-Dimensional Model (electronic version). Applied Measurement in Eduacation.19 ,289-304.

Mullis, I., Martin, O., Ruddock, G.,O'Sullivan,C., & Preuschoff, C. (2009). TIMSS 2011 Assessment

Framework. TIMSS and PIRLS International Study Center. Boston College: USA.

Osbome, J., & Overbay, A. (2004).The Power of Outliers (and why Researchers Should Always Check for Them). Practical Assessment Research and Evaluation,9,(6)Online at http://pareonline.net/getvn.asp

Oshima, T., Kushubar, S., Scott, J., & Raju N. (2009). DFIT8 for Window User's Manual: Differential functioning of items and tests. St. Paul MN: Assessment Systems Corporation

Oshima,T., & Morris, S. (2008). An NCEM Instructional Module on Raju's Differential Functioning of Items and Tests (DIF). Educational Measurement : Issues and Practice,27,43-50.

Oshima, T., Raju, N., & Nanda. (2006).A new Method for Assessing the Statistical Significance in the Differential Functioning of Items and Tests(DFIT) Framework. Journal of Educational Mesurment,43,1-17.

Osterlind, S. (1983). Test item bias. Beverly Hills: Sage publications.

Pae, N.S .(2004).DIF for Examinees with Different Academic Background. Language testing,21,53-72.

Raju, N., Van Der Linden, & Fleer, P. (1995). An IRT-Based Internal Measures of Differential Functioning of Items and Tests. Applied Psychological Measurement,19**,** 353-368.

Ryan, K., & Chiu, S. (1996). Detecting DIF on mathematics items: The case for gender and calculator sensitivity. Paper presented at the annual meeting of the American Education Research Association, New York.

Reckase, M. (1997). The past and future of multidimensional item response theory. Applied Psychological Measurement,1**,** 25-36.

Seo, S. (2006).A Review and Comparison of Methods for Detecting Outliers in University Data Sets. M.S. University of Pittsburgh.

Swaminathan, H., & Rogers, J. (1990). Detecting Differential Item Functioning using Logistic Regression Procedure. Journal of Educational Measurement ,127,

Tan, X., & Jeral, M. (2005).Using Local Dif Analysis to Assess Group Differences on Multilingual Examinations. Applied Measurement in Education,2,313-334.

Tukey, J. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley

Turkan, S., Cetin, M., & Toktamis, O. (2012).Outlier Detection by Regression Diagnostics Based on Robust Parameter Estimates. Hacettepe Journal of Mathematics and Statistics, 41,15-90.

Teresi, J. (2001).Statistical Methods of Examination of Differential Item Functioning with Application to Cross-Cultural Measurement of functional, Physical and Mental Health. J Mental Health and Aging,7,31-40.

Williams, S. (1997). The "unbiased" anchor: Bridging the gap between DIF and item bias. Applied Measurement in Education,10,253-267.

Zimowski, M., Muraki, E., Mislevey, R., Robert, J., & Bock, D .(2002).Bilog-Mg3:Multiple Group IRT Analysis and Test Maintenance for Binary Items (Computer Program). Chicago: Scientific Software.