الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف العاشر تبعا لمتغير الجنس*

Differential Item Functioning of the National Educational Quality Control Test in Mathematics for 10th Grade According to Gender

حسان العمري*[1]، وعبدالله شطناوي[2]

Hassan Alomari & Abdallah shatnawi

[1] قسم علم النفس التربوي، كلية العلوم التربوية، الجامعة الأردنية، الأردن

[2] قسم القياس والتقويم، كلية العلوم التربوية، جامعة جدارا، الأردن

*الباحث المراسل، بريد الكتروني: ha.omari@ju.edu.jo

تاريخ التسليم: (2015/9/17)، تاريخ القبول: (2016/3/13)

ملخص

هدفت هذه الدراسة إلى الكشف عن الأداء التفاضلي لفقرات الاختبار الوطني الأردني لمادة الرياضيات للصف العاشر تبعا لمتغير الجنس، وبلغت عينة الدراسة (2400) طالباً وطالبة بواقع (1200) طالب و(1200) طالبة ممن تقدموا للاختبار للعام الدراسي 2014/2013 ولتحقيق هدف الدراسة تم الحصول على البيانات المتعلقة باستجابات الطلبة على الاختبار المكوّن من (39) فقرة، وتم استخدام (مؤشر الأداء التفاضلي اللاتعويضي NCDIF) في الكشف عن الأداء التفاضلي للفقرات حيث بلغ عدد الفقرات في صورتها النهائية بعد مطابقة الفقرات للنماذج المستخدمة (25) فقرة مطابقة للنموذج الأحادي المعلمة و(34) فقرة مطابقة للنموذج الثنائي المعلمة وكشفت النتائج عن وجود أداء تفاضلي منتظم لـ (9) فقرات من أصل (16) فقرة هي عدد الفقرات المشتركة بين الطلاب والطالبات وعن وجود أداء تفاضلي غير منتظم لـ (15) فقرة من أصل (26) فقرة هي عدد الفقرات المشتركة بين الطلاب والطالبات وخلصت هذه الدراسة إلى جملة من التوصيات تدعو إلى إجراء دراسات مختلفة تتناول الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم وخصوصا في مباحث أخرى غير الرياضيات، والكشف عن الأداء التفاضلي للاختبارات الوطنية الأخرى التي تم تطبيقها الكترونيا على الطلبة.

* بحث مستل من رسالة ماجستير غير منشورة.

الكلمـات المفتاحيـة: الأداء التفاضـلي للفقـرة، نظريـة الاسـتجابة للفقـرة، الاختبـار الـوطني لضـبط نوعيـة التعليم، اختبـار الرياضـيات، النمـوذج أحـادي المعلمـة، النمـوذج الثنـائي المعلمـة، مؤشر NCDIF.

Abstract

This study aimed to reveal the gender related differential item functioning for Jordanian national test items for mathematics learning quality control for tenth grade. The sample of the study consisted of 2400 students (1200 males and 1200 females) who took the exam for the year 2013/2014. To achieve the purpose of the study we used the students responses at the test which consist of (39) items by using Measures of Differential Functioning of Items and Tests (Noncompensatory: NCDIF), where the items in its final form was (25) items for one-parameter logistic model and (34) items for two-parameter logistic model. The results revealed that there is a uniform differential functioning for (9) items out of (16) common items and revealed non- uniform differential functioning for (15) items out of (26) common items. This study concluded a number of recommendations claiming different studies in other subjects related to differential item functioning in the national test of the learning quality control. And Detection of differential item functioning of other national tests that have been applied to students electronically.

Kay Words: Differential Item Functioning, Item Response Theory, The National Educational Quality Control Test, Mathematics Test, One-Parameter Logistic Model, Two-Parameter Logistic Model. NCDIF.

المقدمة

اهتم المختصون في مجال القياس النفسي والتربوي بقضية تحيّز الاختبار ويرجع ذلك إلى أهمية هذا الموضوع وتأثيره بشكل مباشر على القرارات مثل اختيار الأفراد للوظائف المختلفة أو تصنيفهم تبعاً لقدراتهم أو تحصيلهم.

وعُرّف الاختبار غير المتحيّز بأنه الاختبار الصادق نسبياً لجميع المجموعات والأفراد ويعطي فرصاً متساوية لجميع المفحوصين لإظهار المهارات والمعارف التي اكتسبوها والتي تتصل بغرض الاختبار (Roever, 2005). ويتحقق ذلك عندما يكون أداء المفحوصين الذين

لديهم القدرة نفسها متماثلاً على فقرات الاختبار بدون تحيز للجنس أو العرق أو الثقافة. لذلك فإن التحقق من خلو الاختبار وفقراته من التحيّز يعد خطوة مهمة من خطوات بناء الاختبار (Howe, 1995) وقد عمل المختصون في بناء الاختبارات على استخدام مفهوم الأداء التفاضلي للفقرة للتعبير عن الطرق والمعاني الإحصائية المستخدمة للكشف عن تحيّز الفقرات.

وعلى الرغم من استخدام الباحثين للأساليب القياسية التقليدية في بناء أدوات القياس بشكل واسع وحتى هذه اللحظة، إلا أن هذه الأساليب واجهت انتقادات بسبب بعض نواحي القصور التي تركزت على موضوعية هذه الأساليب في قياس ما أعدت لقياسه، ومن بين هذه الانتقادات تقييد الدرجة الكلية للاختبار بفقرات هذا الاختبار، بحيث تتغير هذه الدرجة مع تغيير أو تبديل الاختبار بآخر يقيس نفس القدرة فضلا عن عدم وجود وحدة ثابتة للقياس، واعتماد إحصاءات الفقرة على عينة المفحوصين وتغيّر معنى ومضمون فقرات الاختبار بمرور الزمن (Hambelton, 1987).

ثم جاءت نظرية استجابة الفقرة (Item Response Theory, IRT) لتحل أوجه القصور التي عانت منها نظرية الاختبار (Classical Test Theory, CTT) من خلال افتراضاتها واستخدامها لبرمجيات متطورة مثل برمجية (DFIT) المستخدمة في هذه الدراسة والتي لها مميزات كثيرة منها أنه يُمكن استخدامها في حالة البيانات ثنائية ومتعددة التدريج ويُمكن استخدامها لمعالجة نماذج نظرية الاستجابة للفقرة أُحادية البعد ومتعدّدة الأبعاد ومميزات أخرى غيرها.

مشكلة الدراسة وأسئلتها

تكمن مشكلة هذه الدراسة في الكشف عن فقرات الاختبار الوطني التي تحمل أداءً تفاضلياً والتأكد من أن هذا الاختبار غير متحيّز لفئة على حساب آخرى مما يساهم في تحقيق مبدأ العدالة بين الطلاب، وذلك باستخدام برنامج (DFIT v. 8.04) حيث يختــص هذا البرنامج في كشف الأداء التفاضلي على مستوى الفقرة للنَّماذج البارامترية القائمة على نظرية الاستجابة للفقرة، ثنائية التدريج ومتعددة التدريج، حيث يتمُّ استخدام مؤشِّر الأداء التفاضلي اللاتعويضي (Noncompensatory DIF, NCDIF)، وقد جاءت هذه الدراسة للإجابة عن السؤالين الآتيين:

1. ما فقرات الاختبار الوطني لمادة الرياضيات التي تظهر أداءً تفاضلياً منتظما ذو دلالة إحصائية لمتغير الجنس في نفس مستويات القدرة؟
2. ما فقرات الاختبار الوطني لمادة الرياضيات التي تظهر أداءً تفاضليًا غير منتظم ذو دلالة إحصائية لمتغير الجنس في نفس مستويات القدرة؟

أهمية الدراسة

تكتسب هذه الدراسة أهميتها من أهمية الاختبار المستخدم فيها نفسه كونها تبحث في الأداء التفاضلي لفقرات الاختبار الوطني لمادة الرياضيات لمتغير الجنس وهو الاختبار الذي ترتكز عليه وزارة التربية والتعليم في تشخيص واقع التعليم في الأردن، حيث ستفيد هذه الدراسة في تقديم معلوماتٍ حول خصائص الفقرات الاختبارية التي يتضمنها هذا الاختبار وكشف الفقرات المتحيزّة منها، لتجنب وجود مثل هذه الفقرات في المرات القادمة التي يعقد فيها الاختبار.

هدف الدراسة

هدفت هذه الدراسة إلى الكشف عن الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم للرياضيات تبعا لمتغير الجنس باستخدام طريقة تعتمد على نظرية الاستجابة للفقرة والتي تعتبر من الطرق الحديثة وغير المستخدمة بشكل كبير في الدراسات العربية.

مصطلحات الدراسة

الاختبار الوطني لضبط نوعية التعليم في الرياضيات: اختبار سنوي تعده وزارة التربية والتعليم/مديرية الاختبارات لقياس امتلاك الطلبة في الصف المستهدف لمهارات التعلم الأساسية ومستويات أداء الطلبة بدلالة مؤشرات الأداء، ومدى امتلاكهم لمهارات اقتصاد المعرفة في المباحث التي اختبروا فيها.

الأداء التفاضلي للفقرة: هو مؤشر إحصائي يستخدم للتعبير عن الفروق في احتمال الاستجابة الصحيحة على الفقرة بين مجموعتين فرعيتين من المفحوصين لمجتمع إحصائي ما، ممن لهم نفس مستوى القدرة، تسمَّى إحداها بالمجموعة المرجعية (Reference Group)، فيما تسمَّى الأخرى بالمجموعة المستهدفة (Focal Group).

الأداء التفاضلي المنتظم (Uniform – DIF): وهو الأداء التفاضلي الذي لا يتفاعل فيه مستوى القدرة مع عضوية المجموعة عند جميع مستويات القدرة.

الأداء التفاضلي غير المنتظم (Nonuniform-DIF): وهو الأداء التفاضلي الذي يظهر عندما يحدث تفاعلاً بين مستوى الأداء وعضوية المجموعة.

مؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (Noncompensatory DIF): وهو معدَّل مربَّع المسافة ما بين منحنيي خصائص الفقرة للمجموعتين المستهدفة والمرجعية.

الاختبار الوطني: هو اختبار تعدّه وزارة التربية والتعليم الأردنية /إدارة الامتحانات والاختبارات كلّ عام لصف أو أكثر في مباحث محددة يهدف إلى ضبط نوعية التعليم من خلال قياس مهارات التعلم الأساسية المتوقع أن يمتلكها الطلبة (وزارة التربية والتعليم, 2014).

الإطار النظري

يشير الأدب النظري المتعلق بالأداء التفاضلي للفقرات الاختبارية إلى نوعين رئيسيين هما الأول: الأداء التفاضلي المنتظم (Uniform – DIF) وهو الأداء التفاضلي الذي يكون فيه احتمال الإجابة الصحيحة لمجموعة من المفحوصين أعلى أو أقل منه في المجموعة الأخرى عند جميع مستويات القدرة. بمعنى انه لا يتفاعل فيه مستوى القدرة مع عضوية المجموعة. والثاني: الأداء التفاضلي غير المنتظم (Nonuniform -DIF) وهذا النوع من الأداء التفاضلي يظهر عندما يكون هنالك تفاعلا بين مستوى الأداء وعضوية المجموعة، أي أن الأداء التفاضلي سيظهر لصالح المجموعة المستهدفة عند مستوى قدرة معين ولصالح المجموعة المرجعية عند مستوى قدرة أخر ويكون الاختلاف بين منحنيات الاستجابة للفقرة في معلمة التمييز (Gruijter, kamp, 2005).

لقد أشار أوشيما وموريس (Oshima& Morris, 2008) إلى وجود مؤشرين للأداء التفاضلي على مستوى الفقرة هما:

أولاً : مؤشّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (Noncompensatory DIF, NCDIF)

يَفْتـرض هذا المؤشّر كما أشار راجو وآخرون (Raju, *et al*., 1995) أنَّ جميع فقرات الاختبار (بافتراض أنَّ لدينا n من الفقرات) غير متحيّزة باستثناء الفقرة (i) المراد كشف الأداء التفاضلي لها، ويعرَّف مؤشر الأداء التفاضلي اللاتعويضي على أنَّه معدَّل مربَّع المسافة ما بين منحنيي خصائص الفقرة للمجموعتين المستهدفة والمرجعية، وِفقاً للمعادلة الرياضية (1) الآتية:

$$NCDIF_i = \int_{-\infty}^{\infty} \left[P_{iF}(\theta) - P_{iR}(\theta)\right]^2 f_F(\theta) d\theta \text{ ....(1)}$$

حيث:

$f_F(\theta)$: اقتران الكثافة للقدرة θ في المجموعة المستهدفة.

$P_{iF}(\theta)$: احتمال الاستجابة الصحيحة على الفقرة i عند مستوى القدرة θ في المجموعة المستهدفة.

$P_{iR}(\theta)$: احتمال الاستجابة الصحيحة على الفقرة i عند مستوى القدرة θ في المجموعة المرجعية.

أو وفقاً للمعادلة (2):

$$NCDIF_i = E_F\left[d_i(\theta)^2\right] = \delta^2_{d_i} + \mu^2_{d_i} \text{ .... (2)}$$

E_F: التوقع بالنسبة لتوزيع θ للمجموعة المستهدفة.

$d_i(\theta) = P_{iF}(\theta) - P_{iR}(\theta)$ وذلك في حالة نماذج الاستجابة للفقرة ثنائية التدريج.

σ: الانحراف المعياري، μ: الوسط الحسابي.

ثانياً: مؤشِّر الأداء التفاضلي على مستوى الفقرة التعويضي
(Compensatory DIF, CDIF)

ويأخذ هذا المؤشِّر بعين الاعتبار التباين المشترك للفقرة، كما يراعي التحيُّز الموجود في الفقرات الأخرى للاختبار.

أما في ما يتعلق باختبارات الدلالة الإحصائية المستخدمة في "مقياس الأداء التفاضلي للفقرة والاختبار" (Significance Tests in the DFIT Framework) فقد طوّر راجو (Raju *et al*, 1995) اختبارات للدلالة الإحصائية قائمة على الإحصائي كاي- تربيع (Chi-Square statistic)، وذلك لاختبار الدلالة الإحصائية في كل من مؤشِّر الأداء التفاضلي على مستوى الاختبار(DTF)، ومؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF)، أما بالنسبة لمؤشِّر الأداء التفاضلي على مستوى الفقرة التعويضي (CDIF) فإن دلالته لا تُخْتبر مباشرة (Oshima & Morris, 2008).

وعلى الرغم من استخدام إحصائي كاي- تربيع لاختبار دلالة مؤشِّر الأداء التفاضلي على مستوى الاختبار(DTF)، ومؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي(NCDIF)، إلَّا أنَّ هذا الإحصائي يُعتبر حسَّاساً لحجوم العينات الكبيرة، وبالتالي فهو يميل في حالة العينات الكبيرة لجعل الفقرات المتحررة من الأداء التفاضلي تُـبْـدي أداءً تفاضلياً ذو دلالة إحصائية، الأمر الذي دفع راجو لابتكار علامة قطِع من خلال دراسات قائمة على بيانات مولَّدة (Simulated Data) لاختبار دلالة مؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (NCDIF) لتكون كما يلي:

- $NCDIF > 0.006$: في حالة الفقرات ثنائية التدريج.
- $NCDIF > 0.006\times(k-1)^2$: في حالة الفقرات متعدّدة التدريج، حيث K: فئة الاستجابة.

إلَّا أنَّ علامة القطع هذه قد افتقرت إلى قضية تعميمها، حيث أجريت دراسات مثل دراسة بلوت (Blot, 2002)، ودراسة تشامبلي (Chamblee, 1998) أثبتت

أنَّهـا تعـتمد علـى عـدة عوامل أخـرى مثـل حجـم العينـة المسْتخـدمة، ونـوع النَّمـوذج المسْتخـدم، الأمر الذي دفع أوشيما وزميليه راجو وناندا (Oshima, Raju & Nanda, 2006)لابتكار اختبار للدلالة الإحصائية لمؤشِّر الأداء التفاضلي على مستـوى الفقـرة اللاتعـويضي (NCDIF) يَصْلُح للفقـرات ثنائيـة التدريج سُمِّـيَّ بـِ: **"تكرار توليد معلمة الفقرة" IPR** ((Item Parameter Replication).

تعتبر طـريقة تكـرار توليد معلمة الفقـرة (IPR) وفق ما أشار له أوشيما وموريس (Oshima & Morris, 2008) بأنَّها الطـريقة الأمثل التي توفِّر وسيلة لاشتقاق نقاط قطِع مصمَّمة لتتوافق مع مجموعة معينة من الفقرات والبيانات المتعلِّقة بها، حيث تبدأ بمعالم الفقرات المقدَّرة للمجموعة المستهدفة (Focal Group)، والتباينات المصاحبة لتلك المعالم المقدَّرة لعينة تلك الفقرات، والتي يتمُّ الحصول عليها من إحدى البرامج الخاصَّة بنظرية الاستجابة للفقرة (IRT) مثل برنامج (BILOG-MG3)، الموضوع من قبل كلٍّ من زيموسكي، وموراكي ومسليفي وبوك (Zimowski, Muraki, Mislevy, & Bock, 2002).

بالاعتمـاد علـى هـذه التقديرات المبدئيـة لمعالـم الفقـرات (Initial estimated Parameters) يتمُّ تكـرار توليـدها عـدداً كبيراً مـن المـرات (على سبيل المثال 100، 1000،...، أو 10000 مرة) بافتراض وجود مجتمعين لهما نفس معالم الفقرات بحيث يمثِّل أحدهما المجموعة المستهدفة في حين يمثل الآخر المجموعة المرجعية، وبافتراض أنَّ الأداء التفاضلي الحقيقي بين المجموعة المستهدفة والمرجعيـة معدوماً (True DIF = Zero)، مـع مراعاة أنَّ التوقُّع (Expectation) للتقديرات الجديدة لمعالم الفقـرات هـي نفسها للتقـديرات المبدئيـة، وبنفـس التباينـات والتباينات المصاحبـة الأوليَّـة، مما يجعـل الفـروق فـي هـذه التقـديرات ولأنَّها ناتجـة مـن نفـس تـوزيع المجتمـع تُعْـزى للخطـأ العينـي (Sampling Error) فقط لا لغيره. إن ما تنتجه هذه العملية يمثِّل أزواجاً من العينات التي تتضمن معالم الفقرات المولَّدة لكلٍّ من المجموعة المستهدفة والمجموعة المرجعية في كل عملية تكرار (Replication)، ومن ثمَّ يتمُّ حساب مؤشِّر الأداء التفاضـلي علـى مستـوى الفقـرة اللاتعـويضي (NCDIF) لكل زوج من تلك العينات مع الحصول على توزيعات عينية إمبريقية لمؤشِّر الأداء التفاضـلي علـى مستـوى الفقـرة اللاتعـويضي (NCDIF Distribution)، ولكل فقرة من فقرات الاختبار كلٌّ على حِدة، مما يتيح المجال لتـرتيب قيم مؤشِّر الأداء التفاضـلي علـى مستـوى الفقـرة اللاتعـويضي (NCDIF) لكل فقرة، واشتـقاق عـلامة قطِـع خاصَّـة بتلك الفقرة عـن طـريق حساب المئين المتعلِّق بمستـوى الدلالة (Alpha Level) المرغوب.

أمَّا ميزات مقياس الأداء التفاضلي للفقرة والاختبار فتتمثل في استخدامه في حالة البيانات ثنائية ومتعددة التدريج، وإمكانية استخدامه لمعالجة نماذج نظرية الاستجابة للفقرة أُحادية البعد ومتعددة الأبعاد كما وأنه يعطي مؤشِّراً عن الأداء التفاضلي ليس على مستوى الفقرة فقط، بل وعلى مستوى الاختبار ويُستخدم في تطبيقات متعددة مثل: حُزم الأداء التفاضلي (Differential

(Bundle Functioning) والأداء التفاضلي الشرطي (Conditional DIF) (Oshima & Morris, 2008).

الدراسات السابقة

في دراسة قام بها البرصان (2013) هدفت إلى الكشف عن الأداء التفاضلي لمتغير الجنس لفقرات الاختبار الوطني الأردني لضبط نوعية التعليم لمادة الرياضيات للصف العاشر، وذلك لكل من المستويات الأربعة التي تصنف وزارة التربية والتعليم طلبتها فيها وهي المستوى المتقدم، والماهر، والماهر جزئياً، والمبتدئ. بلغ عدد أفراد عينة الدراسة (43468) طالباً وطالبة ممن تقدموا للاختبار للعام الدراسي 2009/2008 اختيروا عشوائيا من طلبة الصف العاشر الأساسي بواقع (21330) طالباً و(22138) طالبة. ولتحقيق أهداف الدراسة تم الحصول على استجابات الطلبة للاختبار الوطني الأردني لمادة الرياضيات المكون من (30) فقرة، ثم تم تحليلها لمستويات الطلبة (اربعة مستويات) باستخدام طريقة مانتل-هانزل للكشف عن الأداء التفاضلي، وكشفت نتائج الدراسة عن الفقرات التي تظهر أداء تفاضلياً في كل مستوى، وقد كان عدد الفقرات يزيد كلما اتجه الباحث من المستوى الأعلى إلى المستوى الأدنى، حيث تراوحت نسبة الفقرات التي تظهر أداء تفاضليا بين 30% للمستوى المتقدم و67% للمستوى المبتدئ.

وأجرت أبو حماد (2008)، دراسة هدفت إلى تقصي فاعلية طرق منحنى خصائص الفقرة (مؤشر المساحة للنموذج ثلاثي، ثنائي، فوق المعلم) في الكشف عن الأداء التفاضلي لمتغير الجنس في عينة مختارة من الاختبار الدولي في الرياضيات من نوع الاختيار من متعدد، كما هدفت إلى تحري درجة التوافق بين طرق منحنى خصائص الفقرة في اعتبار الفقرة تظهر أو لا تظهر أداءً تفاضلياً في كل فقرات الاختبار والكشف عن القدرة الرياضية الذي تقيسه الفقرات التي تظهر أداءً تفاضلياً لمتغير الجنس في كل فقرات الاختبار. وتكون الاختبار من (40) فقرة من نوع الاختيار من متعدد لكل فقرة أربعة بدائل وتكونت عينة الدراسة من (800) طالباً وطالبة اختيروا من مدارس بني كنانة وأظهرت النتائج وجود أداء تفاضلي متفاوت في الدرجة في (27) فقرة للنموذج ثلاثي المعلم و(28) فقرة للنموذج ثنائي المعلم و(19) فقرة لمؤشر فوق المعلم، ووجدت أيضاً أن الطرق الثلاثة لم تتفق في تحديد الفقرات التي أظهرت أداءً تفاضلياً وقد بلغ عدد الفقرات التي أظهرت أداءً تفاضلياً بموجب الطرق الثلاثة (12) فقرة. وجدت أن هنالك أداءً تفاضلياً متفاوتاً في الدرجة في عدد من الفقرات وفق طبيعة المحتوى الرياضي الذي تقيسه بموجب الطرق الثلاثة حيث كان في معظمه من النوع غير المنتظم.

كما قام ميندس وايركان (Mendes, Erickan, 2006) بدراسة للتعرف على مصدر الأداء التفاضلي وفقاً للجنس في اختبار الرياضيات الأساسية (British Columbia Principles of Mathematical Exam) المطبق على طلبة الصف العاشر وكانت فرضيتهم أن محتوى الفقرة والإدراك والسياق الذي تقدم فيه الفقرة قد تكون مصدراً للأداء التفاضلي وفقاً للجنس، وبينت النتائج أن الفقرات المنظمة حزمة من المشكلات واللوغاريتم والأسس كانت مفضلة لدى الذكور بينما الفقرات المتضمنة علاقات رياضية والفقرات المتضمنة أوصافاً إسمية

كانت مفضلة لدى الإناث كما أظهرت النتائج تفوق الذكور في الاستجابة على الفقرات التي تتطلب عمليات إدراكية عليا ورسوماً وصور.

كما وأجرت كاثرين وشوان (Katherine & Shuan, 2001) دراسة هدفت إلى الكشف عن تجمع الفقرات التي تظهر الأداء التفاضلي حسب متغير الجنس وتأثير تغير مواقعها في نموذج اختبار مستوى في الرياضيات (Midwestern Mathematics Placement Exam) يتكون من (40) فقرة من نوع الاختيار من متعدد طبق على (2731) مفحوصا (1066 طالبا، 1665 طالبة)، وبيّنت نتائج الدراسة أن فقرات موضوع الجبر كانت أسهل على الجنسين، في حين أنَّ الفقرات في موضوعي الهندسة والمثلثات كان أسهل عند الذكور منها عند الإناث.

وأجرى غامر وانغلهارد (Gamer & Englhard, 1999) دراسة هدفت إلى الكشف عن الفقرات التي تظهر أداء تفاضلياً لمتغير الجنس في اختبار قدرة رياضية، وقد استخدم البرنامج الإحصائي (FACETS) في الكشف عن الأداء التفاضلي حيث تكون من 60 فقرة من نوع الاختيار من متعدد وقد أظهرت نتائج الدراسة أن الذكور يتفوقون على الإناث في موضوعي الهندسة والقياس وتحليل البيانات, بينما أظهرت الإناث تفوقا على الذكور في موضوع الجبر، أما بالنسبة للفقرات محددة البناء، فقد كان متوسط علامات الذكور أعلى من متوسط علامات الإناث في فقرتين فقط. في حين لم تظهر فروق دالة إحصائيا بين الجنسين على بقية الفقرات من نفس المجال.

مجتمع الدراسة وعينتها

تكون مجتمع الدراسة من جميع طلاب وطالبات الصف العاشر الأساسي للعام 2014/2013 الذين تقدموا للاختبار الوطني لضبط نوعية التعليم في مادة الرياضيات بنسخته الورقية في مدارس وزارة التربية والتعليم الحكومية والخاصة، والمدارس التابعة لوكالة الغوث الدولية، والمدارس التابعة لمديرية التعليم والثقافة العسكرية، حيث يتم تطبيق الاختبار على جميع الطلبة والطالبات من قبل وزارة التربية والتعليم. وبعد الرجوع إلى ملف البيانات الذي تم الحصول علية من قسم الاختبارات التابع لوزارة التربية والتعليم كان عدد المتقدمين للاختبار (22048) طالباً وطالبة، تم اختيار عينة عشوائية منه لهذه الدراسة مقدارها (2400) طالبا وطالبة وذلك باستخدام برمجية (*SPSS*) وبواقع (1200) طالباً و(1200) طالبة.

أداة الدراسة

تكونت أداة الدراسة من "الاختبار الوطني لضبط نوعية التعليم" للصف العاشر الأساسي للعام الدراسي 2014/2013 حيث تكون الاختبار من (39) فقرة جميعها من نوع الاختيار من متعدد من اربعة بدائل، توزعت هذه الفقرات على اربعة محاور هي:الأعداد والعمليات عليها، والجبر، والقياس، والهندسة، والإحصاء والاحتمالات.

وبالنسبة إلى صدق وثبات الاختبار فيتم التأكد من صدق الاختبار من قبل فريق وزارة التربية والتعليم في قسم الاختبارات وذلك بعرضه على عدد من المختصين في القياس والتقويم

وعدد من المشرفين والمعلمين كما تم تقدير معامل الثبات من قبل فريق وزارة التربية في إدارة الامتحانات باستخدام معامل كودر ريشاردسون (KR-20) حيث بلغ ثبات هذا الاختبار (0.817).

إجراءات الدراسة

بعد الحصول على البيانات الضرورية للدراسة تم التحقق من افتراض أُحادية البُعد وِفقاً لنظرية الاستجابة للفقرة باستخدام برنامج (NOHARM) على عينة الدراسة باستخدام مؤشر تاناكا لحسن مطابقة البيانات للنموذج ومؤشر الجذر التربيعي للوسط الحسابي لمربعات البواقي (RMSR) الذي يحسب من خلال المعادلة (3):

$$RMSR \leq \frac{4.1}{\sqrt{N}} \ldots\ldots\ldots (3)$$

حيث: N عدد أفراد كل مجموعه

ثم التحقق من افتراض الاستقلال الموضعي (Local Independence) لفقرات الاختبار البالغ عددها (39) فقرة باستخدام برنامج *LDID* (Local Dependence Indices for Dichotomous Items) على بيانات عينة التطبيق النهائي وفقًا للنموذجين اللوجستيين (الأحادي المقترن بالأداء التفاضلي المنتظم، والثنائي المقترن بالأداء التفاضلي غير المنتظم)، ومن ثم التحقق من مطابقة الأفراد والفقرات للنموذجين اللوجستي أحادي وثنائي المعلمة وِفقاً لنظرية الاستجابة للفقرة باستخدام برنامج (BILOG-MG3) حيث تم حذف الفقرات التي لم تطابق النموذج المستخدم ثم تقدير معالم الفقرات لكل من مجموعتي الطلبة وفقا للنموذجين، ومن ثم حساب معاملي الربط باستخدام برنامج (*Equate v.2.1*) ثم البدء باستخدام برنامج (*DFIT v. 8.04*)، للإجابة عن سؤالي الدراسة.

نتائج الدراسة

أولاً: نتائج التحقق من افتراضات نظرية الاستجابة للفقرة

تمّ التحقق من افتراض أُحادية البُعد (Unidimensionality) باستخدام برنامج (Normal Ogive Harmonic Analysis Robust Method: *NOHARM*)، الذي وضعه فراسر (Fraser, 1988) باستخدام مؤشر تاناكا (TANAKA) لحسن مطابقة البيانات للنموذج ومؤشر RMSR (الجذر التربيعي للوسط الحسابي لمربعات البواقي) وكانت النتائج كما في الجدول (1).

جدول (1): نتائج مؤشِّري أُحادية البُعد Tanaka وRMSR للاختبار لدى مجموعتي الطلاب والطالبات.

مؤشري أحادية البعد	القيمة المحسوبة للإحصائي (ذكور)	نقطة القطع لاحتمالية قبول قيمة الإحصائي القيمة الحرجة	القيمة المحسوبة للإحصائي (إناث)	نقطة القطع لاحتمالية قبول قيمة الإحصائي القيمة الحرجة
مؤشر RMSR	0.00653	0.11836	0.00655	0.11836
مؤشر TANAKA	0.96785	0.96	0.96988	0.96

نلاحظ من الجدول (1) أن مؤشر TANAKA زادت قيمته عن (0.96) لدى طلبة المجموعتين، حيث بلغت للطلاب 0.96785 وللطالبات 0.96988 وهي قيمة تشير إلى تحقق أُحادية البُعد بمستوى جيد من مطابقة البيانات لافتراض أحادية البعد (Jasper, 2010). كما بلغت القيمة المحسوبة لمؤشِّر (RMSR) حوالي (0.00653) تقريباً لدى الطلاب وهي أقل من القيمة الحرجة له البالغة قيمتها (0.11836)، فيما بلغت قيمته المحسوبة (0.00655) تقريباً لدى مجموعة الطالبات وهي أقل من القيمة الحرجة له البالغة قيمتها (0.11836). كما أنَّ القيمتان الناتجتان (0.00653)، (0.00655) لمجموعتي الدِّراسة الطلاب والطالبات على الترتيب منخفضتان وتكادان تقتربان من الصفر وهذا مؤشِّر كاف على تحقق افتراض أُحادية البُعد (Fraser & McDonald, 1988).

وتم أيضا التحقق من افتراض الاستقلال الموضعي (Local Independence) لفقرات الاختبار البالغ عددها (39) فقرة، باستخدام *برنامج (LDID) على بيانات عينة التطبيق النهائي وفقًا للنموذجين اللوجستيين (الأحادي والثنائي) المعلمة.*

جدول (2): التكرارات والنِّسب المئويَّة لحالتي الاستقلال الموضعي لفقرات الاختبار لدى مجموعتي الطلاب والطالبات وفقًا للنموذجين اللوجستيين (الأحادي، والثنائي) المعلمة.

الأداء التفاضلي	حالة التبعية الموضعية	ذكور		إناث	
		التكرار	النسبة المئوية	التكرار	النسبة المئوية
منتظم	غير مستقل	224	30.2	251	33.9
	مستقل	517	69.8	490	66.1
	الكلي	**741**	**100.0**	**741**	**100.0**
غير منتظم	غير مستقل	206	27.8	227	30.6
	مستقل	535	72.2	514	69.4
	الكلي	**741**	**100.0**	**741**	**100.0**

تشير نتائج الجدول أعلاه إلى أن عدد أزواج الفقرات التي وقعت خارج مدى فترة الثقة لدى الطلاب بالنسبة للنموذج الأحادي المعلمة (224) زوجاً؛ أي ما نسبته (30.2%) وللطالبات

(251) زوجاً؛ أي ما نسبته (33.9%) من عدد الأزواج الكلي، بينما كان عدد أزواج الفقرات التي وقعت ضمن مدى فترة الثقة لدى الطلاب (517) زوجاً؛ أي ما نسبته (69.8%) وللطالبات (490) زوجاً؛ أي ما نسبته (66.1%) من عدد الأزواج الكلي، وهذا يبين أنَّ عدد أزواج الفقرات التي حققت الاستقلالية الموضعية أعلى بحوالي ضعفين من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي (Kim, Cohen & Lin, 2005).

أما بالنسبة للنموذج اللوجستي ثنائي المعلمة المقترن بالأداء التفاضلي غير المنتظم فإن عدد أزواج الفقرات التي وقعت خارج مدى فترة الثقة لدى الطلاب (206) زوجاً، أي ما نسبته (27.8%) وللطالبات (227) زوجاً، أي ما نسبته (30.6%) من عدد الأزواج الكلي، بينما كان عدد أزواج الفقرات التي وقعت ضمن مدى فترة الثقة للطلاب (535) زوجاً أي ما نسبته (72.2%) وللطالبات (514) زوجاً، أي ما نسبته (69.4%) من عدد الأزواج الكلي، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي.

كما تم التحقق من مطابقة الأفراد والفقرات للنموذجين اللوجستي أحادي وثنائي المعلمة من خلال مطابقة بيانات الطلبة لافتراضات النَّموذج أحادي المعلمة باستخدام برنامج (BILOG-MG3) بهدف مطابقة الأفراد والفقرات لكلٍّ من مجموعتي الطلاب والطالبات كلٌّ على حِدة، حيث تمَّ استخدام البرنامج لكلِّ مجموعة على حِدة ثلاث مرات، ففي المرة الأولى التي هدفت إلى حذف الأفراد الذين لم تتطابق استجاباتهم مع توقعات النَّموذج اللوجستي أحادي المعلمة لم يتم حذف أي طالب أو طالبة وفي المرة الثانية التي هدفت إلى فحص مطابقة فقرات الاختبار للنَّموذج اللوجستي أحادي المعلمة، فقد تبين عدم مطابقة (16) فقرة للنَّموذج المستخدم لدى مجموعة الطلاب وهي الفقرات ذوات الأرقام (1، 2، 3، 4، 6، 9، 11، 13، 16، 18، 22، 26، 30، 31، 33، 39)، حيث لم تُطابِق هذه الفقرات شروط نظرية الاستجابة للفقرة وذلك باستخدام اختبار (2χ) لحسن المطابقة, وعدم مطابقة (21) فقرة للنَّموذج المستخدم لدى مجموعة الطالبات ذوات الأرقام (1، 2، 4، 6، 8، 9، 10، 11، 12، 13، 16، 18، 19، 22، 24، 30، 31، 32، 33، 35، 39), أما تشغيل البرنامج للمرَّة الثالثة فقد هدف إلى تقدير معالم كلّ من الأفراد والفقرات لكل مجموعة على حِدة وتبين أن عدد الفقرات المشتركة وغير المشتركة بين مجموعتي الطلاب والطالبات هو (25) فقرة، وأن عدد الفقرات المشتركة بينهما هو (16) فقرة، وأن عدد الفقرات غير المشتركة من جهة الطلاب هو (7) فقرات ومن جهة الطالبات هو (فقرة) واحدة، ويقصد بالمشتركة أن الفقرة مطابقة للنموذج المستخدم للطلاب والطالبات معا، إذ تراوحت قيم معلمة الصعوبة من (-0.49 إلى 0.66) لدى مجموعة الطلاب، وبمعدَّل مقداره 0.04، ومن (-1.21 إلى 1.40) لدى مجموعة الطالبات، وبمعدَّل مقداره 0.10. أما بالنسبة للنموذج الثنائي المعلمة ففي المرة الأولى تمَّ حذف (7) أفراد من مجموعة الطلاب، وذلك لتعذر حساب الخطأ المعياري في تقدير القدرة لديهم عند مستوى قدرة ما ليصبح العدد النهائي لمجموعة الطلاب (1193) طالباً، كما تمّ حذف (14) فرداً من مجموعة الطالبات لنفس السبب السابق ليصبح العدد النهائي لمجموعة الطالبات (1186) طالبةً، وبالنسبة للتشغيل

الثاني للبرنامج فقد تبين عدم مطابقة (11) فقرة للنَّموذج المستخدم لدى مجموعة الطلاب ذوات الأرقام (1، 2، 6، 11، 13، 16، 21، 22، 26، 28، 30)، بسبب انخفاض احتمالية الخطأ لهذه الفقرات دون مستوى الدلالة (0.005)، كما تبين عدم مطابقة (7) فقرات للنَّموذج المستخدم لدى مجموعة الطالبات ذوات الأرقام (1، 4، 6، 11، 13، 22، 35)، لنفس السبب السابق. وبالتشغيل الثالث تبيّن أن عدد الفقرات المشتركة وغير المشتركة بين مجموعتي الطلاب والطالبات هو (34) فقرة، وأن عدد الفقرات المشتركة بينهما هو (26) فقرة، وأن عدد الفقرات غير المشتركة من جهة الطلاب هو (فقرتين) ومن جهة الطالبات هو (6) فقرات، وأنَّ القيم المقدَّرة لمعلمة الفقرات تتباين من حيث التمييز عبر مجموعتي الطلاب والطالبات، إذ تراوحت قيم معلمة التمييز بين (0.19 إلى 1.03) لدى مجموعة الطلاب، وبمعدَّل مقداره 0.65، وبين (0.16 إلى 1.35) لدى مجموعة الطالبات، وبمعدَّل مقداره 0.67. وأنَّ القيم المقدَّرة لمعلمة الفقرات تتباين من حيث الصعوبة عبر مجموعتي الطلاب والطالبات، إذ تراوحت قيم معلمة الصعوبة بين (0.49 إلى 5.39) لدى مجموعة الطلاب، وبمعدَّل مقداره (1.62)، وبين (-1.29 إلى 14.44) لدى مجموعة الطالبات، وبمعدَّل مقداره (2.85).

وقد تمّ حساب معامل الثبات لكلا مجموعتي الاختبار حيث بلغت قيمة معامل الثبات الإمبريقي للاختبار الخاصَّة بمجموعة الطلاب وفقًا للنموذج أحادي المعلمة (0.7060)، وللطالبات (0.6953). في حين بلغت قيمة معامل الثبات الإمبريقي للاختبار الخاصَّة بمجموعة الطلاب وفقًا للنموذج ثنائي المعلمة (0.7168)، وللطالبات (0.7708).

ثانياً: نتائج السؤال الأول ومناقشتها

هدف السؤال الأول إلى الكشف عن فقرات الاختبار الوطني لمادة الرياضيات التي تظهر أداءً تفاضليًا منتظمًا لمتغير الجنس ذات دلالة إحصائية في نفس مستويات القدرة تمَّ اعتماد مجموعة الطالبات كمجموعة مرجعية (Reference Group)، ومجموعة الطلاب كمجموعة مستهدفة (Focal Group) أثناء تشغيل برنامج (Equate v2.1) بسبب أنَّ برنامج *DFIT v. (8.04)* يعتبر المجموعة المرجعية على أنَّها مجموعة مستهدفة، والمجموعة المستهدفة على أنَّها مجموعة مرجعية بعكس برنامج (Equate v2.1) الذي تتمُّ فيه عملية المعادلة.

تمَّ استخدام برنامج (*Equate v.2.1*) الموضوع من قبل بيكر (Baker, 1995) لحساب الثابت الضربي، الذي بلغت قيمته (A=1.000)، والثابت الجمعي الذي بلغت قيمته (K=0.0351)، للتمكن من معادلة القيم المقدَّرة لمعلمة الصعوبة لفقرات الاختبار لدى مجموعة الطلاب كمجموعة مستهدفة إلى القيم المقدَّرة لمعلمة لفقرات الاختبار لدى مجموعة الطالبات كمجموعة مرجعية وذلك باستخدام جميع فقرات الاختبار المطابقة للنموذج اللوجستي أحادي المعلمة (المشتركة وغير المشتركة بين الطلاب والطالبات) البالغ عددها (25) فقرة واستخدام الفقرات المشتركة بينهما والبالغ عددها (16) فقرة كجذع مشترك في عملية المعادلة. وقد تمَّ تنفيذ برنامج (*DFIT v. 8.04*) باستخدام معاملي الربط السَّابقين، وملف القيم المقدَّرة لمعلمة الصعوبة b، والتباينات (Variances) الخاصَّة بفقرات الاختبار لدى

مجموعة الطالبات كمجموعة مرجعية والبالغ عددهن (1200) طالبةً، بالإضافة لملف القيم المقـدَّرة لمعلمة الصعوبة b الخاصَّة بفقرات الاختبار لدى مجموعة الطلاب كمجموعة مستهدفة والبالغ عددهم (1200). بالإضافة أيضاً لملـف القـدرة المقـدَّرة لمجموعة الطالبات كمجموعة مرجعية. وفي ضوء ما تقـدَّم تمَّـت معادلـة القيم المقدَّرة لمعلمة الصعوبة b لفقرات الاختبار لدى مجموعة الطلاب كمجموعـة مستهدفة باستخدام معاملي الربط الأوليين وتـمَّ ذلك مـن خـلال برنامـج (*DFIT v. 8.04*)، وباستخـدام طريقـة "تكـرار توليد معلمـة الفقـرة،(Item Parameter) Replication" المُشار لها سابقا، حِسـاب قيـم نقـاط القطِـع (Cut off Scores) الخاصَّة بكلِّ فقرة من فقرات الاختبار المشتركة بين الطلاب والطالبات البالغ عددها (16) فقرة عند مستويات دلالة مختلفة يعطيها البرنامج، وذلك للحكم على الأداء التفاضلي المنتظم من حيث دلالته الإحصائية أم لا، وعلى مستوى كل فقرة من فقرات الاختبار، بالإضافة لحساب نقاط القطع الخاصَّة بالحكم على الأداء التفاضلي على مستوى الاختبار ككل، وعند مستويات دلالة مختلفة أيضاً، والوسط الحسابي والانحراف المعياري لقيم نقاط القطع وبناءً علـى قيم نقـاط القطـع المحسـوبـة الخاصَّـة بكـل فقـرة مـن فقـرات الاختبار، تـمَّ الحكـم علـى الفقـرة إنْ كانـت تُبـدي أداءً تفاضـليًا أم لا بالاعتمـاد علـى قيمـة مؤشِّـر الأداء التفاضلي علـى مستوى الفقـرة اللاتعويضـي (NCDIF) الذي يحسب عن طريق المعادلة رقم (4):

$$NCDIF_i = E_F\left[d_i(\theta)^2\right] = \delta^2_{d_i} + \mu^2_{d_i} \text{ ....(4)}$$

المُشار لها سابقا، ولكل فقرة من فقرات الاختبار كلٌّ على حِدة، وذلك عن طريق مقارنة قيمة مؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي (**NCDIF**) لكلِّ فقـرة مـن فقـرات الاختبار، بقيم نقاط القطـع الخاصَّـة بتلك الفقـرة ابتـداءً من قيمـة القطِـع المناظِـرة لمستـوى الدلالة (α=0.001)، حيث تتوقف العملية عند أول قيمة نقطة قطع مناظِرة لمستوى دلالة معين، تكون عندها قيمة هذا المؤشِّر أكبر من تلك القيمة، وفي هذه الحالة تكون الفقرة موضع الاهتمام ذات أداء تفاضلي دال إحصائياً عند مستوى الدلالة المناظر لقيمة نقطة القطع التي تمَّ التوقف عندها، وتُسجَّل قيمة مستوى الدلالة تلك مقابل الفقرة في مخرجات البرنامج في العمود الخاص بمستوى القيمة الاحتمالية.

وتبين أنَّ كلاً من الفقرات (5، 15، 21، 23، 25، 27، 28، 29، 36، 37) قد أبدت أداءً تفاضلياً عند مستوى الدلالة α= 0.05 أو أقل من ذلك لذا سيتمُّ استثناؤها من فقرات الجذع المشترك البالغ عددها (16) فقرة في عملية إجراء المُعادلة للمرَّة الثانية باستخدام برنامج (*Equate v.2.1*) للحصول على ثابتين ضربي وجمعي جديدين ومنقَّحين بهدف تشغيل برنامج (*DFIT v. 8.04*) لتنفيذ المرحلة الثانية لكشف الأداء التفاضلي المنتظم في الفقرات بشكل نهائي. حيث تمَّ إعـادة استخدام برنامج (*Equate v.2.1*) لحساب ثابـتين ضربي وجمعي جديدين ومنقَّحين، بهدف التمكن من معادلة القيم المقدَّرة لمعلمة الصعوبة b لفقرات الاختبار لدى

مجموعة الطلاب (المجموعة المستهدفة) إلى القيم المقدَّرة لمعلمة الصعوبة b لفقرات الاختبار لدى مجموعة الطالبات (المجموعة المرجعية) بفقرات جذع مشترك متحررة من الأداء التفاضلي، وذلك باستخدام جميع فقرات الاختبار كفقرات جذع مشترك في عملية المعادلة باستثناء الفقرات التي أبـدت أداءً تفاضلياً والبالغ عددها (6) فقرات، والتي حملت الأرقام (7، 14، 17، 20، 34، 38), حيث بلغـــت قيمـــة الثابت الضربي الجديد (A=1.000)، فـــي حين بلغت قيمة الثابت الجمعي الجديد (K=0.0798).

وقد تمَّ تنفيذ برنامج (*DFIT v. 8.04*) مرَّة ثانية باستخدام قيمتي الثابتين الضربي، والجمعي المنقَّحين كما في الخطوة السابقة وفي ضوء ما تقدَّم، تمَّ حساب قيم نقـــاط القطعِ الجديدة (Cutoff Scores) الخاصَّة بكل فقرة من فقرات الاختبار عند مستويات دلالة مختلفة، من أجل الحكم على الأداء التفاضلي المنتظم من حيث دلالته الإحصائية أم لا على مستوى كل فقرة من فقرات الاختبار، بالإضافة لحساب قيم نقاط القطع الجديدة الخاصَّة بالحكم على الأداء التفاضلي المنتظم على مستوى الاختبار ككل، وعند مستويات دلالة مختلفة أيضاً، والوسط الحسابي والانحراف المعياري لقيم نقاط القطع، وذلك كما هو مبين في الجدول (3).

جدول (3): قيم نقاط القطعِ الخاصَّة بالحكم على الأداء التفاضلي المنتظم على مستوى الفقرة لفقرات الاختبار وقيم نقاطَ القطعِ الخاصَّة بالحكم على الأداء التفاضلي المنتظم على مستوى الاختبار ككل عند مستويات دلالة مختلفة والوسط الحسابي والانحراف المعياري لها

رقم الفقرة	0.001	0.005	0.01	0.05	0.1	0.5	المتوسط الحسابي	الانحراف المعياري
3								
5	0.00588	0.00414	0.00352	0.00174	0.00116	0.00018	0.00045	0.00069
7	0.00573	0.00414	0.00351	0.00168	0.00119	0.00021	0.00046	0.00067
8								
10								
12								
14	0.00661	0.0056	0.00433	0.00235	0.00172	0.00026	0.00062	0.00091
15	0.00428	0.00363	0.00319	0.00203	0.00137	0.0002	0.0005	0.0007
17	0.00718	0.00505	0.00439	0.00237	0.0017	0.00025	0.00063	0.00091
19								
20	0.00696	0.00522	0.00421	0.00223	0.00156	0.00029	0.00062	0.00089
21	0.00618	0.00523	0.00387	0.00227	0.00162	0.00027	0.00061	0.00084
23	0.00683	0.00466	0.00404	0.00234	0.00152	0.00028	0.0006	0.00086
24								

... تابع جدول رقم (3)

رقم الفقرة	0.001	0.005	0.01	0.05	0.1	0.5	المتوسط الحسابي	الانحراف المعياري
25	0.00532	0.00336	0.00281	0.00175	0.00124	0.00019	0.00045	0.00065
26								
27	0.0051	0.00424	0.00376	0.00211	0.00146	0.00025	0.00054	0.00078
28	0.00538	0.00488	0.00415	0.00241	0.00173	0.00024	0.00061	0.00089
29	0.0082	0.00457	0.00383	0.00205	0.00162	0.00027	0.00059	0.00086
32								
34	0.00542	0.00434	0.00408	0.00238	0.00158	0.00026	0.00059	0.00085
35								
36	0.00676	0.00484	0.00375	0.00247	0.00169	0.00027	0.00061	0.00088
37	0.00655	0.00453	0.00374	0.00229	0.00153	0.00025	0.00058	0.00086
38	0.00498	0.00421	0.00402	0.00207	0.00147	0.00026	0.00055	0.00078
DTF	0.09383	0.08412	0.06157	0.03164	0.02226	0.00449	0.00882	0.01249

وفي ضوء قيم نقاط القطع المبينة في الجدول (3) الخاصَّة بكل فقرة من فقرات الاختبار، تمَّ الحكم على الفقرة إن كانت تُـبدي أداءً تفاضلياً ذا دلالة إحصائية، أم لا، بالاعتماد على قيمة مؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي NCDIF الذي تمَّت الإشارة إليه آنفاً في معرض الحديث عن المرحلة الأولى، وذلك كما هو مبين في الجدول (4).

جدول (4): الدلالات الإحصائية للأداء التفاضلي على مستوى الفقرة لفقرات الاختبار.

رقم الفقرة	(d)		(ldl)		C (d,D)	CDIF	NCDIF	القيمة الاحتمالية
	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري				
3								
5	-0.103	0.042	0.103	0.042	0.003	-0.02371	0.01234	0.001
7	-0.01	0.004	0.01	0.004	0	-0.0023	0.00011	غير دالة
8								
10								
12								
14	0.024	0.006	0.024	0.006	0.001	0.00669	0.0006	غير دالة
15	-0.07	0.024	0.07	0.024	0.001	-0.0172	0.00543	0.001
17	-0.008	0.002	0.008	0.002	0	-0.00212	0.00006	غير دالة

... تابع جدول رقم (4)

رقم الفقرة	(d) المتوسط الحسابي	(d) الانحراف المعياري	(ldl) المتوسط الحسابي	(ldl) الانحراف المعياري	C (d,D)	CDIF	NCDIF	القيمة الاحتمالية
19								
20	-0.006	0.002	0.006	0.002	0	-0.00178	0.00004	غير دالة
21	0.086	0.02	0.086	0.02	0.002	0.024	0.00788	0.001
23	0.153	0.035	0.153	0.035	0.003	0.04291	0.02476	0.001
24								
25	-0.143	0.057	0.143	0.057	0.004	-0.03345	0.0237	0.001
26								
27	-0.039	0.012	0.039	0.012	0	-0.00987	0.00163	غير دالة
28	0.093	0.023	0.093	0.023	0.001	0.02561	0.00919	0.001
29	0.055	0.017	0.055	0.017	0	0.01409	0.00328	0.05
32								
34	0.026	0.007	0.026	0.007	0	0.00713	0.00075	غير دالة
35								
36	0.143	0.033	0.143	0.033	0.003	0.04013	0.02149	0.001
37	0.077	0.019	0.077	0.019	0.001	0.02092	0.00631	0.005
38	-0.02	0.007	0.02	0.007	0	-0.00508	0.00046	غير دالة

يتضح من الجدول (4) وجود (9) فقرات من أصل (16) فقرة من فقرات الاختبار قد أبدت أداءً تفاضليًا منتظمًا منسوبًا لجنس الطلبة، وهي الفقرات ذوات الأرقام (5، 15، 21، 23، 25، 28، 29، 36، 37)؛ حيث أبدت الفقرة (29) فقط أداءً تفاضليًا منسوبًا لجنس الطلبة عند مستوى الدلالة α=0.05، فيما أبدت الفقرة (37) فقط أداءً تفاضليًا منسوبًا لجنس الطلبة عند مستوى الدلالة α=0.005، كما أبدت الفقرات (5، 15، 21، 23، 25، 28، 36) أداءً تفاضليًا منسوبًا لجنس الطلبة عند مستوى الدلالة α= 0.001، حيث كانت الفقرات (5، 15، 25) تبدي أداء تفاضليا منتظما لصالح الطلاب، حيث تقع الفقرتان (5، 15) في مجال الهندسة فيما تقع الفقرة (25) تحت محور الأعداد والعمليات كذلك أبدت الفقرات (21، 23، 28، 29، 36، 37) أداء تفاضليا منتظما لصالح الطالبات, حيث تقع الفقرة (21) في محور الجبر فيما تقع الفقرة (23) تحت محور القياس والفقرات (28، 29، 36، 37) تحت محور الإحصاء.

لقد اتفقت هذه النتائج مع دراسة بردجمان وشمت (Bridgeman and Schmitt, 1997) في أن الجبر كان أسهل بالنسبة للطالبات والهندسة أسهل بالنسبة للطلاب واتفقت أيضا

مع دراسة حمزة واريفين وياسين (Hamzah, Ariffin, yassin, 2004) في أن الجبر كان أسهل بالنسبة للطالبات واختلفت معها في أن الإحصاء كان أسهل بالنسبة للطلاب واتفقت أيضا مع دراسة هاريس وكارلتون (Harris & Carlon,1993) في أن الهندسة والحساب والفقرات التي تتطلب مهارات تفكير عليا والرسوم كانت أصعب بالنسبة للطالبات، كما واتفقت أيضا مع دراسة عنابي ودودين (Innabi & Dodeen, 2006) في أن الجبر والإحصاء كانت أسهل بالنسبة للطالبات.

ثالثاً: نتائج السؤال الثاني ومناقشتها

هدف السؤال الثاني إلى الكشف عن فقرات الاختبار الوطني لمادة الرياضيات التي تظهر أداءً تفاضليًا غير منتظمًا لمتغير الجنس ذات دلالة إحصائية في نفس مستويات القدرة، فقد تمَّ استخدام برنامج *Equate v.2.1* لحساب الثابت الضربي، الذي بلغت قيمته (A=1.0405)، والثابت الجمعي الذي بلغت قيمته(K=-0.3137)، للتمكن من معادلة القيم المقدَّرة لمعلمتي الصعوبة والتمييز لفقرات الاختبار لدى مجموعة الطلاب كمجموعة مستهدفة (المرجعية في ضوء برنامج (DFIT v. 8.04) إلى القيم المقدَّرة لمعلمتي الصعوبة والتمييز لفقرات الاختبار لدى مجموعة الطالبات كمجموعة مرجعية (المستهدفة في ضوء برنامج (DFIT v. 8.04)، وذلك باستخدام جميع فقرات الاختبار المطابقة للنموذج اللوجستي ثنائي المعلمة (المشتركة وغير المشتركة بين الطلاب والطالبات) البالغ عددها (34) فقرة واستخدام الفقرات المشتركة بينهما والبالغ عددها (26) فقرة كجذع مشترك في عملية المعادلة.

تمَّ تنفيذ برنامج (***DFIT v. 8.04***) باستخدام معاملي الربط السَّابقين، وملف القيم المقدَّرة لمعلمتي الصعوبة (b) والتمييز (a)، والتباينات والتباينات المصاحبة الخاصَّة بفقرات الاختبار لدى مجموعة الطالبات كمجموعة مرجعية والبالغ عددهن (1186) طالبةً، بالإضافة لملف القيم المقدَّرة لمعلمتي الصعوبة (b) والتمييز (a) الخاصَّة بفقرات الاختبار لدى مجموعة الطلاب كمجموعة مستهدفة والبالغ عددهم (1193) طالبًا، بالإضافة أيضاً لملف القدرة المقدَّرة لمجموعة الطالبات كمجموعة مرجعية. وفي ضوء ما تقدَّم تمَّت معادلة القيم المقدَّرة لمعلمتي الصعوبة (b) والتمييز (a) لفقرات الاختبار لدى مجموعة الطلاب كمجموعة مستهدفة باستخدام معاملي الربط الأوليين.

ثُمَّ تمَّ من خلال برنامج (***DFIT v. 8.04***)، وباستخدام طريقة "**تكرار توليد معلمة الفقرة**، المُشار إليها سابقاً، حِساب قيم نقاط القطِع (Cut off Scores) الخاصَّة بكلِّ فقرة من فقرات الاختبار المشتركة بين الطلاب والطالبات البالغ عددها (26) فقرة عند مستويات دلالة مختلفة يعطيها البرنامج، وذلك للحكم على الأداء التفاضلي غير المنتظم من حيث دلالته الإحصائية أم لا، وعلى مستوى كل فقرة من فقرات الاختبار، بالإضافة لحساب نقاط القطِع الخاصَّة بالحكم على الأداء التفاضلي على مستوى الاختبار ككل، وعند مستويات دلالة مختلفة أيضاً،والوسط الحسابي والانحراف المعياري لقيم نقاط القطِع وبناءً

علـــى قيم نقــــاط القطـــع المحســـوبـــة الخاصَّـــة بكـــل فقـــــرة مـــن فقـــرات الاختبار، تـــمَّ الحكـــم علـــى الفقـــرة إنْ كانـــــت تُبـــــــدي أداءً تفاضـــــليًا أم لا بالاعتمـــاد علـــى قيمـــة مؤشِّـــر الأداء التفاضلـي علـــى مستـوى الفقـــرة اللاتعويضـي NCDIF.

وتبيّن أنَّ كلاً من الفقرات (5، 7، 8، 9، 10، 15، 18، 19، 23، 25، 27، 29، 36، 37، 39) قد أبدت أداءً تفاضلياً عند مستوى الدلالة α=0.05 أو أقل من ذلك لذا تمَّ استثناؤها من فقرات الجذع المشترك البالغ عددها (26) فقرة في عملية إجراء المُعادلة للمرَّة الثانية باستخدام برنامج (***Equate v.2.1***) للحصول على ثابتين ضربي وجمعي جديدين ومنقَّحين بهدف تشغيل برنامج (***DFIT v. 8.04***) لتنفيذ المرحلة الثانية لكشف الأداء التفاضلي غير المنتظم في الفقرات بشكل نهائي.

وعليه تمَّ إعـادة استخدام برنـامج (*Equate v.2.1*) لحساب ثابـتين ضـربي، وجمعي جديدين ومنقَّحين، بهدف التمكن من معادلة القيم المقدَّرة لمعلمتي الصعوبة b والتمييز a لفقرات الاختبار بفقرات جذع مشترك متحررة من الأداء التفاضلي،وذلك باستخدام جميع فقرات الاختبار كفقرات جذع مشترك في عملية المعادلة باستثناء الفقرات التي أبـدت أداءً تفاضلياً والبالغ عددها (11) فقرة، والتي حملت الأرقام (3، 12، 14، 17، 20، 24، 31، 32، 33، 34، 38) حيث بلغـــت قيمـــة الثابت الضربي الجديد (A=1.0789)، فـــي حين بلغت قيمة الثابت الجمعي الجديد (K=-0.3179).

ثم تمَّ تنفيذ برنامج (*DFIT v. 8.04*) مرَّة ثانية باستخدام قيمتي الثابتين الضربي، والجمعي المنقَّحين، كما في المرحلة السابقة وتمَّ حساب قيم نقـــاط القطع الجديدة (Cutoff Scores) الخاصَّة بكل فقرة من فقرات الاختبار عند مستويات دلالة مختلفة، من أجل الحكم على الأداء التفاضلي غير المنتظم من حيث دلالته الإحصائية أم لا على مستوى كل فقرة من فقرات الاختبار، بالإضافة لحساب قيم نقاط القطع الجديدة الخاصَّة بالحكم على الأداء التفاضلي غير المنتظم على مستوى الاختبار ككل، وعند مستويات دلالة مختلفة أيضاً، والوسط الحسابي والانحراف المعياري لقيم نقاط القطع، وذلك كما هو مبين في الجدول (5).

جدول (5): قيم نقاط القطعِ الخاصَّة بالحكم على الأداء التفاضلي غير المنتظم على مستوى الفقرة لفقرات الاختبار وقيم نقاط القطعِ الخاصَّة بالحكم على الأداء التفاضلي غير المنتظم على مستوى الاختبار ككل عند مستويات دلالة مختلفة والوسط الحسابي والانحراف المعياري لها.

رقم الفقرة	0.001	0.005	0.01	0.05	0.1	0.5	المتوسط الحسابي	الانحراف المعياري
2								
3	0.00599	0.00439	0.00387	0.00219	0.00169	0.00048	0.00074	0.00081
4								

... تابع جدول رقم (5)

رقم الفقرة	0.001	0.005	0.01	0.05	0.1	0.5	المتوسط الحسابي	الانحراف المعياري
5	0.00769	0.00562	0.00434	0.00219	0.00156	0.0004	0.00068	0.00087
7	0.00522	0.00297	0.00279	0.0017	0.00134	0.00038	0.00058	0.00065
8	0.00455	0.00345	0.00272	0.00181	0.00141	0.00043	0.00061	0.00064
9	0.00411	0.00311	0.00268	0.00171	0.00119	0.00034	0.00053	0.00058
10	0.00612	0.00405	0.00387	0.00234	0.00184	0.00053	0.00079	0.00085
12	0.00571	0.00412	0.00391	0.00261	0.00198	0.00054	0.00083	0.00086
14	0.00513	0.0044	0.00392	0.00247	0.00191	0.00055	0.00081	0.00083
15	0.00366	0.00318	0.00283	0.0018	0.00139	0.00042	0.0006	0.00061
16								
17	0.00498	0.00408	0.0034	0.00222	0.00175	0.00056	0.00079	0.00076
18	0.00381	0.00324	0.00277	0.00187	0.00134	0.0004	0.00059	0.0006
19	0.00812	0.00607	0.00454	0.00276	0.00199	0.00055	0.00088	0.00101
20	0.00443	0.00365	0.00321	0.00214	0.00157	0.00048	0.0007	0.00069
21								
23	0.00391	0.0036	0.0033	0.00201	0.00156	0.00047	0.00068	0.00068
24	0.00748	0.00456	0.00431	0.00259	0.00188	0.00051	0.00081	0.00094
25	0.00885	0.0048	0.00415	0.0024	0.00161	0.00037	0.00068	0.00094
26								
27	0.00428	0.00353	0.00292	0.00186	0.00131	0.00038	0.00058	0.00062
28								
29	0.00401	0.00324	0.00283	0.00176	0.0013	0.00039	0.00058	0.0006
30								
31	0.01855	0.01363	0.00914	0.00365	0.00232	0.00049	0.00103	0.00177
32	0.00452	0.00297	0.00263	0.00162	0.00132	0.00041	0.00058	0.00057
33	0.00385	0.00283	0.00254	0.00175	0.00133	0.00038	0.00056	0.00057
34	0.00494	0.00408	0.00338	0.0022	0.00164	0.00047	0.00071	0.00074
35								
36	0.00537	0.00387	0.00344	0.0023	0.00182	0.00053	0.00078	0.00078
37	0.00441	0.00353	0.00321	0.00197	0.00147	0.00044	0.00064	0.00066

... تابع جدول رقم (5)

رقم الفقرة	0.001	0.005	0.01	0.05	0.1	0.5	المتوسط الحسابي	الانحراف المعياري
38	0.00468	0.0035	0.003	0.00194	0.00153	0.00045	0.00066	0.00067
39	0.03273	0.01575	0.01115	0.00486	0.00292	0.00044	0.00119	0.00249
DTF	0.12697	0.10141	0.08624	0.0561	0.04212	0.01296	0.01886	0.01865

وفي ضوء قيم نقاط القطع المبينة في الجدول (5) الخاصَّة بكل فقرة من فقرات الاختبار، تمَّ الحكم على الفقرة فيما إذا كانت تُبدي أداءً تفاضلياً ذا دلالة إحصائية، أم لا، بالاعتماد على قيمة مؤشِّر الأداء التفاضلي على مستوى الفقرة اللاتعويضي **NCDIF** الذي تمَّت الإشارة إليه آنفاً ، وذلك كما هو مبين في الجدول (6).

جدول (6): الدلالات الإحصائية للأداء التفاضلي غير المنتظم على مستوى الفقرة لفقرات الاختبار.

رقم الفقرة	(d)		(ldl)		C(d,D)	CDIF	NCDIF	القيمة الاحتمالية
	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري				
2								
3	0.023	0.036	0.027	0.032	0.006	0.0056	0.00177	غير دالة
4								
5	-0.107	0.018	0.107	0.018	-0.004	-0.00031	0.01178	0.001
7	-0.017	0.045	0.042	0.024	0.008	0.0086	0.00234	0.05
8	0.03	0.062	0.051	0.046	0.014	0.01302	0.00471	0.001
9	-0.005	0.059	0.053	0.027	0.014	0.01382	0.00352	0.005
10	0.078	0.043	0.082	0.035	-0.008	-0.01053	0.00793	0.001
12	0.027	0.031	0.038	0.018	-0.006	-0.00651	0.00173	غير دالة
14	0.006	0.012	0.012	0.006	-0.003	-0.00273	0.00019	غير دالة
15	-0.092	0.021	0.092	0.021	-0.001	0.00226	0.00888	0.001
16								
17	-0.028	0.026	0.031	0.023	-0.005	-0.00446	0.00148	غير دالة
18	0.041	0.064	0.057	0.051	0.015	0.01328	0.00581	0.001
19	0.044	0.046	0.058	0.025	-0.008	-0.00941	0.004	0.05
20	-0.03	0.01	0.031	0.01	0.002	0.00266	0.00103	غير دالة
21								

... تابع جدول رقم (6)

رقم الفقرة	(d) المتوسط الحسابي	(d) الانحراف المعياري	(ldl) المتوسط الحسابي	(ldl) الانحراف المعياري	C(d,D)	CDIF	NCDIF	القيمة الاحتمالية
23	0.123	0.03	0.123	0.03	0.001	-0.00319	0.01602	0.001
24	0.032	0.028	0.039	0.017	-0.005	-0.00605	0.0018	غير دالة
25	-0.159	0.056	0.159	0.056	-0.011	-0.00638	0.02845	0.001
26								
27	-0.059	0.048	0.072	0.024	0.009	0.01055	0.00571	0.001
28								
29	0.031	0.036	0.035	0.033	0.008	0.00727	0.00232	0.05
30								
31	-0.009	0.022	0.016	0.018	-0.004	-0.00366	0.00057	غير دالة
32	-0.004	0.04	0.035	0.019	0.009	0.00919	0.00161	غير دالة
33	-0.03	0.019	0.033	0.012	0.004	0.00442	0.00125	غير دالة
34	0.014	0.017	0.016	0.015	0.003	0.00303	0.00048	غير دالة
35								
36	0.122	0.024	0.122	0.024	-0.002	-0.00624	0.01548	0.001
37	0.062	0.074	0.074	0.062	0.016	0.01433	0.00937	0.001
38	-0.035	0.022	0.039	0.012	0.004	0.00471	0.00171	غير دالة
39	-0.091	0.004	0.091	0.004	-0.001	0.00219	0.00832	0.05

يتضح من الجدول (6) وجود (15) فقرة من أصل (26) فقرة من فقرات الاختبار قد أبدت أداءً تفاضليًا غير منتظمًا منسوبًا لجنس الطلبة، وهي الفقرات ذوات الأرقام (5، 7، 8، 9، 10، 15، 18، 19، 23، 25، 27، 29، 36، 37، 39)؛ حيث أبدت الفقرات (7، 19، 29، 39) فقط أداءً تفاضليًا منسوبًا لجنس الطلبة عند مستوى الدلالة α=0.05، فيما أبدت الفقرة (9) فقط أداءً تفاضليًا منسوبًا لجنس الطلبة عند مستوى الدلالة α=0.005، كما أبدت الفقرات (5، 8، 10، 15، 18، 23، 25، 27، 36، 37) أداءً تفاضليًا منسوبًا لجنس الطلبة عند مستوى الدلالة α=0.001 . فالفقرات ذوات الأرقام (5، 15، 25، 39) تبـــدي أداءً تفاضليًا غير منتظمًا ظاهريًا لصالح مجموعة الطلاب، وذلك لأن احتماليـــة الاستجابـــة الصحيحـــة علـــى الفقـــرة لمجموعـــة الطلاب تتقاطع مع احتمالية الاستجابة الصحيحة على الفقرة لمجموعـــة الطالبات خارج المدى المعتمد للقدرة (-4.00 إلى 4.00)، والفقرتين (23، 36) تبـــديان أداءً تفاضليًا غير منتظمًا ظاهرياً لصالح مجموعة الطالبات، أيضاً بسبب التقاطع خارج المدى المعتمد للقدرة (-4.00 إلى 4.00)، والفقرات ذوات الأرقام (7، 8، 9، 18، 19، 29،

37) تبــدي أداءً تفاضليًا غير منتظمًا لصالح مجموعة الطالبات على حساب مجموعة الطلاب في مستويات القدرة المرتفعة والعكس من ذلك صحيح, والفقرات ذوات الأرقام (10، 19، 27) تبــدي أداءً تفاضليًا غير منتظمًا، لصالح مجموعة الطالبات على حساب مجموعة الطلاب في مستويات القدرة المنخفضة والعكس من ذلك صحيح.

لقد اتفقت هذه النتائج مع دراسة غامر وانغلهارد (Gamer & Englhard, 1999) في أن الطلاب أفضل من الطالبات في الهندسة والقياس وأن الطالبات أفضل من الطلاب في الجبر، واتفقت ايضاً مع دراسة ميندس وابيركان (Mendes, Erickan, 2006) في أن الطلاب أفضل في الاستجابة على الفقرات التي تتطلب عمليات إدراكية عليا ورسوماً وصور.

التوصيات

بناءً على نتائج الدراسة فإنه يمكن التوصية بما يأتي

1. دراسة الأداء التفاضلي لفقرات الاختبار الوطني لمواد أخرى غير الرياضيات لمتغير الجنس.
2. إجراء دراسات تكميلية تحاول تحليل الفقرات التي أظهرت تحيزا منتظما ومحاولة الكشف عن أسباب ذلك للاستفادة من هذا التحليل في كتابة الفقرات مستقبلا.
3. إجراء دراسات تكميلية تحاول الوصول الى استنتاجات حول الفقرات لتي أظهرت تحيّزا منتظما مقابل الفقرات التي أظهرت تحيّزا غير منتظم.
4. إجراء دراسات تتناول الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم للكشف عن الأداء التفاضلي للاختبارات الوطنية الأخرى التي تم تطبيقها الكترونيا على الطلبة.

References (Arabic & English)

- Abu Hammad, Kawsar. (2008). *Investigate the effectiveness of the item characteristics curve in the detection of differential functioning of the gender in a selected sample of test in math. non published thesis*. Yarmouk University. Irbid. Jordan.
- Al-prusan, Ismail. (2013). *Differential functioning of the gender in Jordanian national test items for the quality of education for mathematics to the tenth grade*. Number 79. Zagazig magazine. Egypt.
- 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.

- Blot, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education, 15(2).* 113-141.
- Bridgeman, B., & Schmitt, A. (1997). *Fairness issues in test development and administration*. Gender and fair assessment, 185-226.
- Chamblee, M. C. (1998). *A Monte Carlo investigation of conditions that impact type 1 error rates of differential functioning of items and tests***.** Unpublished doctoral dissertation, Georgia State University, Atlanta.
- De Gruijter, D. N. M., & Van der Kamp, L. J. Th. (2005). *Statistical test theory for education and psychology*.
- Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate behavioral research, 23(2)*, 267-269.
- Fraser, C. (1988). *NOHARM. [Computer software and manual]*. Armidale, New South Wales, Australia: author.
- Gamer, M. & Engelhard, G. (1999). Gender differences in performance on multiple -choice and constructed response of mathematics items. *Applied Measurement in Education, 12(1)*, 29-43.
- Hambleton, Ronald. (1987). *Item Respose theory principles and application*, NY. Nighoffpuplishing
- Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications (Vol. 7). Springer Science & Business Media.
- Hamzah, H. Ariffin, S. & Yassin, R. (2004). Girls' Overachievement in The Mathematics National Examination: Can Type Of Items Be

The Cause? *Journal of Science And Mathematics Education In S.E. Asia 29(2).*

- Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test. *Applied Measurement in Education, 6(2)*, 137-151.
- Howe, D., & Campling, J. (1995). *Attachment theory for social work practice*. Macmillan.
- Innabi, H.Dodeen, H. (2006). Content Analysis of Gender-related Differential Item Functioning TIMSS Items in Mathematics in Jordan, *School Science and Mathematics. 106(8).*
- Jasper, F. (2010). Applied dimensionality and test structure assessment with the START-M mathematics test. *International Journal of Educational and Psychological Assessment, 6(1)*, 104-125.
- Kim, S. H., Cohen, A. S., & Lin, Y.-H. (2005). *LDID: A computer program for local dependence indices for dichotomous items*. Applied Psychological Measurement.
- Katherine, R & Shuwan, C. (2001). An Examination of Item Context Efects, DIF, and Gender DIF. *Applied Measurement In Education. 14(1)*, 73-90
- Ministry of Education. (2014). *The Result of the National Educational Quality Control Test 2013/2014*, Amman Jordan
- Oshima, T. C. & Morris, S. B. (2008). An NCME Instructional Module on Raju's Differential Functioning of Items and Tests (DFIT). *Educational Measurement: Issues and Practice 27 (3)*: 43–50.
- Roever, C., (2005). That's not fair" Fairness, bias, and differential item functioning in language testing. Retrieved, February, 6, 2015, from: http // www.hawaii .edul /2ve / brounbag.doc.

- Zimowski, M., Muraki, E. Mislevy, R. Robert, J., & Bock, D. (2000). Bilog-Mg3**:** *Multiple group IRT analysis and test maintenance for binary items* (computer program) Chicago: Scientific Software No.17.