

العلاقة بين الأداء التفاضلي للموهبات والأداء التفاضلي للفقرات في اختبار رياضيات من نوع الاختيار من متعدد

يوسف السوالمه* وجهاد العجلوني**

تاريخ قبوله 2018/11/15

تاريخ تسلم البحث 2018/6/24

The Relationship Between the Differential Distractors Functioning and the Differential Item Functioning in a Multiple-Choice Mathematics Test

Yousef Al Sawalmeh, Yarmouk University, Jordan.
Jehad Al Ajlouni, Ministry of Education, Jordan.

Abstract: This study aimed at examining the relationship between the differential distractors functioning (DDF) and the differential item functioning (DIF) in a multiple-choice mathematics test. The data for the study consisted of the responses of (1540) fifth grade students to (26) multiple choice test items, each with four alternatives measuring basic mathematical skills. Because there was a statistically significant difference in performance on the test due to both, the sex of the student and the location of the school, DDFS software has been used to examine the DDF and DIF in terms of the two variables, through applying the Mantel- Hanzel odds ratio MH-LOR method. The results of the differential item and distractor functioning analysis showed that there are (8) DIF items and (15) DDF distractors in the test in terms of gender or location. With regard to the relationship between DIF and DDF, the results showed a positive relationship between them. It is expected that the item will more likely have DIF with increased probability of containing DDF distractors.

(Keywords: Differential Item Functioning, Differential Distractor Functioning, Mathematics Test, Fifth Grade, Relationship Between DDF and DIF)

ولتقويم فائدة الاختبارات وضمان تحقيقها لأغراضها نحتاج إلى إجراءات معيارية في بناء الاختبارات وإدارتها وتقويمها. فالاختبارات الجيدة يتم التخطيط لها وتطوير فقراتها وتدقيقها وتحليلها بأساليب مختلفة تسمح بانتقاء أفضل الفقرات لتضمينها في النسخة النهائية من الاختبار. وينطوي فن كتابة الفقرات على التخطيط الدقيق والتحليل لضمان الحصول على أسئلة فاعلة لا غموض فيها (Allen & Yen, 1979). ويعتمد صدق الاختبار على جودة خصائص الفقرات المتضمنة فيه، فالفقرات ذات الخصائص السيكومترية الضعيفة في الاختبار تؤثر سلباً على خصائصه السيكومترية؛ وعليه فإن تحديد مثل هذه الفقرات الإشكالية من خلال تحليل الفقرات له أهمية كبيرة في مراجعة الاختبار وتطويره (Wiberg, 2007).

ملخص: هدفت الدراسة الحالية إلى فحص العلاقة بين الأداء التفاضلي للموهبات DDF، والأداء التفاضلي للفقرات DIF في اختبار رياضيات من نوع الاختيار من متعدد. تكونت بيانات الدراسة من استجابات (1540) طالباً وطالبة في الصف الخامس على (26) فقرة اختيار من متعدد، لكل منها أربعة بدائل تقيس مهارات أساسية في الرياضيات. ونظراً لوجود فرق ذي دلالة إحصائية في الأداء على الاختبار يعزى لكل من جنس الطالب وموقع المدرسة؛ فقد تم استخدام برمجية DDFS للكشف عن الأداء التفاضلي للفقرات والموهبات وفق طريقة نسبة الأرجحية لامنتل هانزل (MH-LOR) بحسب المتغيرين. وبينت نتائج التحليل وجود (8) فقرات و(15) موهباً ذات أداء تفاضلي، سواء فيما يتعلق بجنس الطالب أو موقع المدرسة. وفيما يخص العلاقة بين الأداء التفاضلي للفقرات DIF والأداء التفاضلي للموهبات DDF، فقد أظهرت النتائج وجود علاقة موجبة بينهما. إذ يتوقع أن يزداد احتمال أن تكون الفقرة ذات أداء تفاضلي بازدياد احتمال أن تتضمن موهبات ذات أداء تفاضلي.

(الكلمات المفتاحية: الأداء التفاضلي للفقرة، الأداء التفاضلي للموهبة، اختبار الرياضيات، الصف الخامس، العلاقة بين DDF و DIF)

مقدمة: الاختبار هو أداة للحصول على عينة من سلوك المتعلم. وتستخدم الاختبارات التحصيلية والنفسية على نطاق واسع في المدارس والجامعات والعيادات والدوائر الرسمية وغير الرسمية لأغراض مختلفة تتمثل في الاختيار والتصنيف والتقويم والإرشاد والبحث العلمي. وتهدف الاختبارات التحصيلية إلى تقييم المعارف والمهارات المكتسبة من تعلم سابق والتنبؤ بفرص النجاح في تعلم مستقبلي. ويكون الاختبار صادقاً بالدرجة التي تكون الاستنتاجات المبنية عليه مناسبة وذات دلالة وفائدة.

وينظر لمفهوم صدق الاختبار تقليدياً باعتباره ينقسم إلى ثلاثة أنواع منفصلة، هي: صدق المحتوى والصدق المرتبط بمحك وصدق البناء. ويرى ميسيك (Messick, 1994) أن هذه النظرة قاصرة وغير مكتملة، لأنها تفشل في أن تأخذ بالاعتبار الأدلة على المعاني القيمية للدرجات والعواقب الاجتماعية المترتبة على استخدامها. ويعنى المفهوم الشامل للصدق في تصور ميسيك بدمج متكامل لمصادر الأدلة التي تفسر درجات الاختبار وتبين دلالتها سواء أكان المصدر محتوى الاختبار (صدق المحتوى) أم المحك (الصدق المرتبط بمحك) مع ما يترتب على استخدام الدرجات من تبعات اجتماعية (صدق المفهوم). ولا بد من توافر مواصفات في الاختبار التحصيلي كي يؤدي إلى نتائج صادقة تساعد المسؤول على اتخاذ قرارات ثابتة، وموضوعية، وعادلة.

* جامعة اليرموك، الأردن.

** وزارة التربية والتعليم، الأردن.

© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن.

الاختبار المتحرر من التحيز فقرات متحيزة لكنها تكون متوازنة في اتجاه التحيز، أي نصفها تقريباً يتحيز للمجموعة الأولى بينما يتحيز النصف الثاني للمجموعة الثانية (Hong & Roznowski, 2001).

ويعد فحص الأداء التفاضلي للفقرة (Differential Item Functioning: DIF) جزءاً مهماً في تحليل الفقرات. وقد أكد أحدث إصدار من معايير الاختبارات التربوية والنفسية (AERA, APA & NCME, 2014) على أن الأداء التفاضلي للفقرة هو دليل صدق قائم على البنية الداخلية، ويشير مصطلح الأداء التفاضلي للفقرة إلى اختلافات في الأداء على الفقرة عبر مجموعات من الأفراد بعد مزاجية تلك المجموعات فيما يتعلق بالسمة التي يفترض أن تقيسها الفقرة والاختبار الذي تظهر فيه (Dorans & Holland, 1993). ويمثل وجود أداء تفاضلي للفقرة حضور أبعاد أخرى مزعجة تتطفل على السمة المصممة لقياسها، وهو شكل من أشكال التغير غير المرغوب في القياس، ويشير إلى اختلاف ما تقيسه الفقرة من مجموعة إلى أخرى. وهذا قد يشكل مصدرًا للتحيز في علامات الاختبار (Ackarman, 1992).

ومن المهم التأكيد على أن أداء الفقرة التفاضلي هو فرق غير متوقع بين مجموعتين بعد المطابقة بينهما على القدرة الكامنة التي تستهدف الفقرة قياسها، وهو مرادف للتحيز الإحصائي المتمثل في إعطاء تقدير أعلى أو أدنى من القيمة الفعلية لمعلم أو أكثر في النموذج الإحصائي (Camilli, 2006). وهو خاصية إحصائية للفقرة تختلف عن تحيز الفقرة. فالفقرة المتحيزة تظهر أداءً تفاضلياً لكن الأداء التفاضلي لا يكفي لأن تكون الفقرة متحيزة، فمفهوم تحيز الفقرة أعم ويكمن في التفسير (Camilli & Shepard, 1994).

ويستخدم مفهوم الأداء التفاضلي للفقرة للتعبير عن الطرق والمعاني الإحصائية المستخدمة للكشف عن تحيز فقرات الاختبار (Ellis & Raju, 2003). ويشير الأداء التفاضلي للفقرة DIF إلى الفرق في الاحتمال المشروط لإجابتها إجابة صحيحة بين مجموعتين (Penfield, 2008). ويعني عدم وجود DIF أن احتمالية الاستجابة الصحيحة متساوية للأفراد الذين ينتمون إلى مجموعتين مختلفتين ولكنهم متمثلون في القدرة (θ) (Penfield, 2007) (& Camilli, 2007). ويرى لين (Linn, 1993) أن الفقرات السهلة أكثر احتمالاً لأن تكون ذات أداء تفاضلي من الفقرات الصعبة.

وذكر هامبلتون وسواميناثان وروجرز (Hambleton, Swaminathan & Rogers, 1991) أن الفقرة المتحيزة تقيس سمة غير السمة الأصلية في مجموعتين مختلفتين من الأفراد لكنهما متكافئتان في العلامة الكلية على الاختبار. أي أن الأداء التفاضلي شرط ضروري لكنه غير كافٍ للتحيز، فالفقرة التي تظهر أداءً تفاضلياً تكون غير متحيزة إذا كان الفرق عليها يعزى إلى اختلاف المجموعتين في القدرة المطلوبة للإجابة عن الفقرة، ويعد الأداء التفاضلي تحيزاً عندما يعزى فرق الأداء إلى عوامل ليس لها علاقة بالسمة المقیسة (مبارك، 2010).

ويوفر كل أسلوب من أساليب تحليل الفقرات طريقة لاستخدام بيانات التجريب الأولى لتقويم جودة فقرات الاختبار وفائدتها. فمقياس صعوبة الفقرة مفيد في تقويم ما إذا كانت الفقرة مناسبة لمستوى قدرات المفحوصين. ويشير مؤشر تمييز الفقرة إلى درجة ارتباط الاستجابات عليها مع مجموع الاستجابات على الفقرات الأخرى.

وهو مؤشر على فاعليتها في قياس ما تقيسه الفقرات الأخرى في الاختبار. كما يمثل الفرق بين درجة صعوبة الفقرة لدى ذوي التحصيل العالي (المجموعة العليا) ودرجة صعوبتها لدى ذوي التحصيل المنخفض (المجموعة الدنيا). وهو مؤشر على فاعليتها في التمييز بين فئات التحصيل المختلفة. ويجب أن تكون قيمة مؤشر تمييز الفقرة موجبة. وفي فقرات الاختبار من متعدد تعتمد فاعلية الفقرة على فاعلية المموهات أو المشتتات (Distractors) فيها. والمموه هو البديل الخاطيء للإجابة، ويكون المموه فعالاً عندما يميز باتجاه معاكس لتمييز الفقرة. أي هو المموه الذي يكون تمييزه سالباً. ورياضياً، يلاحظ أن مؤشر تمييز الفقرة بمفهومه التقليدي (نسبة الإجابات الصحيحة في المجموعة العليا - نسبة الإجابات الصحيحة في المجموعة الدنيا) هو المعكوس الجمعي لمجموع مؤشرات تمييز مموهاتها. فمثلاً، إذا كان البديل a هو البديل الصحيح في فقرة ما وكانت البدائل b, c, d هي المموهات في تلك الفقرة، فإن:

$$D_a = -(D_b + D_c + D_d)$$

حيث يشير الرمز D_i إلى تمييز البديل i.

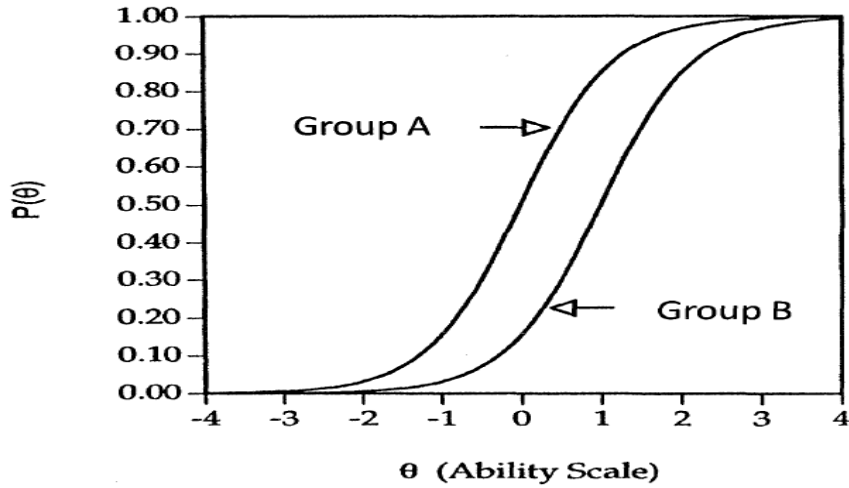
وبالأرقام، إذا توزعت إجابات أفراد المجموعتين العليا والدنيا على البدائل a, b, c, d بالنسب: 0.03, 0.07, 0.20, 0.70، وبالنسب: 0.20, 0.40, 0.30, 0.10 على الترتيب في المجموعة العليا، وبالنسب: 0.20, 0.40, 0.30, 0.10 على الترتيب في المجموعة الدنيا، فإن معاملات التمييز للبدائل الأربعة على الترتيب هي: 0.50, -0.20, -0.23, -0.07. ويلاحظ من هذه القيم أن معامل تمييز الفقرة 0.50 يساوي مجموع معاملات التمييز للمموهات الثلاثة بعد عكس إشارتها. وهذا يدل على أن تحسين المموهات وزيادة فاعليتها يزيد من فاعلية الفقرة.

ويرى وايبيرج (Wiberg, 2007) أن ضمان عدالة الاختبار من التحديات الرئيسية في استخدامه. وتتحقق العدالة عندما يكون هناك دالة تناظر بين الدرجات والقدرة الكامنة. وتعد عدالة الاختبار من وجهة نظر كين (Kane) (as cited in Cains, Bridgiall, & Chatterji, 2014) جزءاً من صدق الاختبار، وتدعم المساواة في التعامل مع المتقدمين للاختبار والاتساق في معاني الدرجات للجميع. والكشف عن الفقرات ذات الأداء التفاضلي لا يكفي من أجل ادعاء عدم العدالة في الاختبار. فعدالة الاختبار تعتمد على كيفية استخدامه، وتحيز الاختبار هو فرق منتظم في الدرجة الكلية للاختبار ضد مجموعة معينة (Camilli, 2006). وقد يتضمن

والمستهدفة B متوازيين، أي أن قيمة الفرق بين المجموعتين في احتمال الإجابة الصحيحة للفقرة له نفس الاتجاه والمقدار عبر مستويات القدرة المختلفة، كما في الشكل (1). ويلاحظ من الشكل أن الفرق في احتمال الإجابة الصحيحة ثابت ولصالح المجموعة A على الدوام وعند جميع مستويات القدرة θ .

ويتم عادة تقييم مقدار الأداء التفاضلي ونوعه، وهناك نوعان للأداء التفاضلي للفقرة تبعاً لطبيعة التفاعل بين مستوى القدرة وعضوية المجموعة، هما (Camilli & Shpard, 1994):

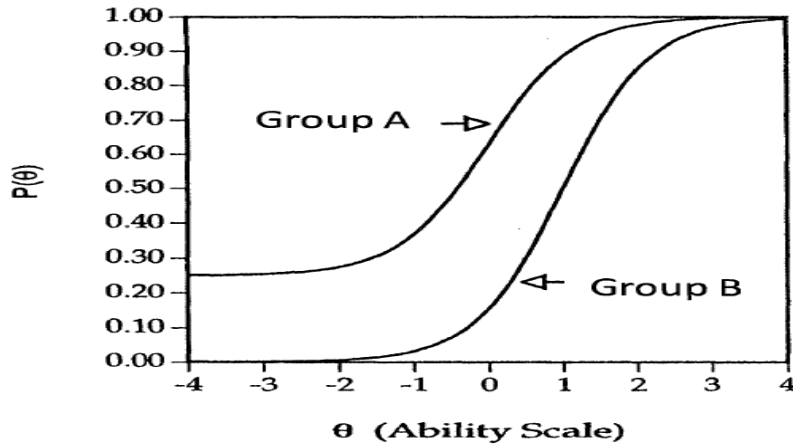
(1) الأداء التفاضلي المنتظم (Uniform): وهو الذي يظهر عندما لا يكون هناك تفاعل بين مستوى القدرة θ وعضوية المجموعة، أي عندما يكون منحنيًا خاصة الفقرة (Item characteristic curves) في المجموعتين المرجعية A



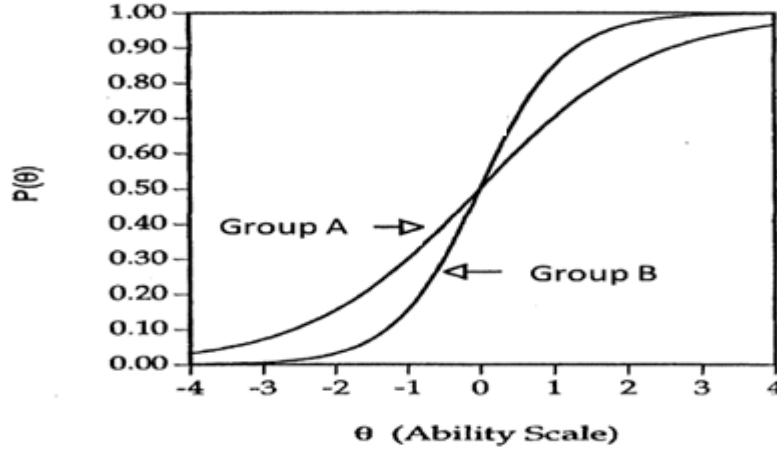
الشكل 1: الأداء التفاضلي المنتظم للفقرة

الصحيحة للفقرة يختلف في الاتجاه أوالمقدار عبر مستويات القدرة المختلفة. فقد يكون الاختلاف في المقدار كما في الشكل (2)، أو في الاتجاه والمقدار كما في الشكل (3) الذي يعبر عن نوع آخر من الأداء التفاضلي غير المنتظم يعرف بالأداء التفاضلي التقاطعي.

(2) الأداء التفاضلي غير المنتظم (Nonuniform): وهو الذي يظهر عندما يكون هناك تفاعل بين مستوى القدرة θ وعضوية المجموعة؛ أي عندما يكون منحنيًا خاصة الفقرة في المجموعتين المرجعية A والمستهدفة B غيرمتوازيين. أي أن قيمة الفرق بين المجموعتين في احتمال الإجابة



الشكل 2: الأداء التفاضلي غير المنتظم للفقرة



الشكل 3: الأداء التفاضلي غير المنتظم (التقاطعي) للفقرة

وتقوم طريقة MH-LOR في تقييمها للأداء التفاضلي للفقرة على مقارنة أرجحية (Odds) الاستجابة الصحيحة لها في المجموعة المرجعية بأرجحية الاستجابة الصحيحة لها في المجموعة المستهدفة عند كل مستوى من مستويات القدرة، ثم يتم حساب الوسط الموزون لهذه المقارنات عبر مختلف مستويات القدرة للحصول على قيمة للإحصائي MH-LOR لكل فقرة. أما في تقييمها للأداء التفاضلي للمموه، فإنها تقوم على مقارنة أرجحية الاستجابة الصحيحة للفقرة بوجوب الاختيار بين الجواب الصحيح والمموه المعني في المجموعة المعيارية بأرجحية الاستجابة الصحيحة للفقرة بوجوب الاختيار بين الجواب الصحيح والمموه المعني في المجموعة المستهدفة عند كل مستوى من مستويات القدرة. ثم يتم حساب الوسط الموزون لهذه المقارنات عبر مستويات القدرة للحصول على قيمة للإحصائي MH-LOR لكل مموه.

وقد تناول العديد من الدراسات الأداء التفاضلي لل فقرات، حيث اهتم بعضها بإجراء مقارنات بين طرق الكشف عن الأداء التفاضلي لل فقرات (أبوحماد، 2008؛ أيوب، 2008؛ البستنجي، 2004؛ جرادات، 2003؛ فريحات، 2014؛ الناجي، 2011؛ Chiu، 2008؛ Yildirm، 2008)، واهتم بعضها الآخر بالكشف عن الأداء التفاضلي لل فقرات في اختبارات ذات أهمية خاصة كاختبارات القبول أو بعض الاختبارات الدولية (حمادنة، 2007؛ العطوي، 2004؛ مبارك، 2010؛ Uiterwijk & Pae، 2004؛ Karakaya، 2012؛ Vallen، 2005)، واستقصت دراسة عبيدات (2014) الأداء التفاضلي لل فقرات تبعاً لدرجة مخاطرة المفحوصين.

أما الأداء التفاضلي للمموهات فقد كان موضوعاً لعدد قليل من الدراسات؛ إذ اهتمت بعض الدراسات بمقارنة طرق الكشف عن الأداء التفاضلي للمموهات. فقد هدفت دراسة بانكس (Banks، 2004) إلى معرفة ما إذا كانت فقرات القراءة وفنون اللغة تعمل بشكل مختلف بين المفحوصين البيض والسود في الصفين الخامس والسابع باستخدام طريقتين (نسبة الأرجحية الخطية واختبار تحيز الفقرة المتوافق SIBTEST) للكشف عن DDF. وقد توافقت

ويصنف حجم الأداء التفاضلي في فئات: صغير، ومتوسط، وكبير، باستخدام محكات مختلفة لحجم الأثر تعتمد على الأسلوب المستخدم في الكشف عنه. وفي العادة لا يتم اتخاذ أي إجراء على الفقرة في حالة الأداء التفاضلي الصغير، وينصح في حال الأداء التفاضلي الكبير بحذف الفقرة أو مراجعتها. وتكمن النقطة الأساسية في عملية المراجعة في تحديد الأسباب الممكنة للأداء التفاضلي خاصة ما يتعلق بالفقرة نفسها. وبين بنفيلد (Penfield، 2010a) أن الأداء التفاضلي للفقرة قد يفسر جزئياً من خلال دراسة استجابات المفحوصين لمموهات الفقرة. ويؤكد بارتون (Barton) (as cited in Middleton & Laitusis، 2007) أن فقرة الاختيار من متعدد على درجة كبيرة من الأهمية، فهي توفر مصدرًا إضافيًا للمعلومات حول أداء الطلبة.

ويسعى مصممو الاختبارات إلى أن تكون مموهات الفقرة جذابة للمستجيبين بغض النظر عن خصائصهم الديموغرافية. ويشير الاختلاف في جاذبية المموه عند مستوى القدرة نفسه باختلاف المجموعة التي ينتمي لها المستجيب إلى مفهوم الأداء التفاضلي للمموه Differential Distractor Functioning (DDF) (Penfield، 2008). وتحليل الأداء التفاضلي للمموهات DDF مكمل لتحليل الأداء التفاضلي لل فقرات DIF ويتم بعده.

وهناك طرق متعددة لتحديد الأداء التفاضلي لل فقرات (DIF) وللمموهات (DDF) مثل الانحدار اللوجستي (LR)، نظرية الاستجابة للفقرة (IRT)، ونسبة الأرجحية لمانتل-هانزل (MH-LOR). وتتميز طريقة نسبة الأرجحية لمانتل-هانزل (MH-LOR) على الطرق الأخرى بأنها طريقة لأملمعية، ويمكن تنفيذها بسهولة لفحص DIF وDDF. وتعد من الطرق الأقوى للكشف عن الأداء التفاضلي (Haladyna، 2004)، وتوفر معلومات كافية وذات معنى عن الأداء التفاضلي للفقرة والمموهات، فهي تعطي مقدار حجم الأثر واتجاهه (Effect size) بحسب معايير محددة، ولا تحتاج إلى عينات كبيرة (Penfield، 2008).

تؤدي إلى أداء تفاضلي منتظم أو غير منتظم أوتقاطعي للفقرة. وتسلط النتائج من الناحية النظرية الضوء على الأسباب للأشكال المختلفة من DIF. فأثر DIF المنتظم ينتج عن أثر DDF ثابت في المقدار والاتجاه (الإشارة) عبر كل مموهات الفقرة، وقد يكون مصدر هذا الأداء التفاضلي المنتظم للفقرة في متن الفقرة أو في الإجابة الصحيحة. فعلى سبيل المثال قد يكون بعض المحتوى في متن السؤال غير مألوف لإحدى المجموعتين، أو قد يكون هناك بعض الغموض في الإجابة الصحيحة، والتي قد تؤثر على تلك المجموعة بشكل منتظم. وهذا يعني أنه يجب أن يكون هناك مراجعة لمحتوى السؤال ومحتوى الإجابة الصحيحة لأي فقرة لديها DIF منتظم ودال إحصائياً. أما أثر DIF التقاطعي فينتج عن أثر DDF المختلف في الإشارة عبر مموهات الفقرة، وأثر DIF غير المنتظم وغير التقاطعي فينتج عن أثر DDF المختلف في المقدار عبر مموهات الفقرة. وتكون المموهات في هذه الحالة هي المصادر المحتملة للأداء التفاضلي للفقرة. وهذا الأمر يقودنا إلى ضرورة مراجعة محتوى المموهات وليس محتوى متن الفقرة. وتدل النتائج على أهمية إجراء تحليل DDF إلى جانب إجراء تحليل DIF فهو مهم في عملية المراجعة لل فقرات.

وفي دراسة أخرى بين بنفيلد (Penfield, 2010b) كيف يتحدد نوع ومقدار DIF بدلالة DDF من خلال توليد بيانات تخص 1000 فقرة اختيار من متعدد افتراضية في ظل 15 ظرفاً مختلفاً في المقدار والإشارة لأنماط DDF. وقد أظهرت الدراسة النتائج الآتية: (1) وجود DDF مختلف عن الصفر يولد على الدوام DIF مختلفاً عن الصفر، (2) قيمة DDF تحدد السقف الأعلى لقيمة DIF، (3) هناك احتمال قليل للحصول على قيمة DIF عالية اعتماداً على DDF لمموه واحد. وتتعارض النتيجة الأولى مع الفكرة البديهية بأن DDF الموجب والسالب التي تلغي بعضها ستؤدي إلى عدم وجود DIF.

و درس غرينبرغ (Greenberg, 2010) العلاقة والتطابق بين DIF و DDF اعتماداً على استجابات 271 طفلاً و 236 طفلة من ذوي الدخل المنخفض على اختبار علوم في مرحلة ما قبل المدرسة. وأظهرت النتائج أن حجم وشكل DIF يعتمد على تأثير DDF. وتم الكشف عن أربع فقرات يوجد بها DIF و DDF وخمس فقرات بها DDF فقط. وتم استعراض محتوى الفقرات التي يوجد بها تحيز للوقوف على الأسباب التي أدت إلى ذلك التحيز.

وفي الأردن، قام مركز تنمية الموارد البشرية بتنفيذ التقييم الوطني لاقتصاد المعرفة (National Assessment for Knowledge Economy Study) (المركز الوطني لتنمية الموارد البشرية، 2011) بهدف التعرف على مستويات أداء طلبة الصفوف الخامس والتاسع والحادي عشر في مهارات اقتصاد المعرفة ضمن محاور الرياضيات والعلوم والقراءة. وتم اختيار العينة باستخدام أسلوب العينة العشوائية العنقودية بحسب عدة متغيرات. وقد أظهرت نتائج الدراسة الوطنية فيما يخص طلبة الصف الخامس في

النتائج في الطريقتين إلى حد كبير. واهتمت دراسة كيون (Koon, 2010) بمقارنة ثلاث طرق (نسبة الأرجحية، الانحدار اللوجستي المتعدد، والطريقة المعيارية) في الكشف عن الأداء التفاضلي للمموهات بحسب المشاركة في برنامج غذائي (مجموعة مشاركة، مجموعة غير مشاركة). وأظهرت النتائج اختلافات بسيطة بين الطرق الثلاث فيما كشفته من مموهات ذات أداء تفاضلي. وتناولت دراسة كيون وكاماتا (Koon & Kamata, 2013) الطرق الثلاث نفسها في فحص الأداء التفاضلي في اختبار رياضيات يطبق على مستوى ولاية فلوريدا بهدف الكشف عما إذا كان مقدار ونمط أثر DDF ثابتاً عبر الطرق الثلاث، وهل يدعم نتائج DIF. وبالرغم من العثور على بعض الاختلافات في النتائج، فإن جميع الطرق الثلاثة تقدم خياراً قابلاً للتطبيق في تحسين فقرات الاختبار المدرجة في برامج التقييم على مستوى الولاية من خلال الاستفادة من المعلومات التي تتوافر من تحليل DDF، وبشكل خاص الفقرات التي تتضمن بعض المموهات الإشكالية (Problematic distractors).

واهتمت دراسة مدلتون ولايتوسس (Middleton & Laitusis, 2007) بفحص DDF بين ذوي صعوبات التعلم الذين حصلوا على مساعدة (قراءة بصوت عال) والذين لم يحصلوا على مساعدة باستخدام النموذج اللوغارتمي الخطي. وأظهرت الدراسة وجود 10 فقرات ذات أداء تفاضلي لصالح الذين حصلوا على مساعدة منها 7 فقرات تتضمن أداء تفاضلياً للمموهات. وهدفت دراسة عبيدي وليون وكاو (Abedi, Leon & Kao, 2007) إلى الكشف على الفقرات التي تظهر أداء تفاضلياً للمموهات بين الطلبة العاديين وذوي الإعاقة باستخدام الانحدار اللوجستي، وتبين أن معظم المموهات ذات الأداء التفاضلي تقع في النصف الثاني من الاختبار، وقد كان احتمال اختيار ذوي الإعاقة للمموه أعلى من الطلبة العاديين، ولهذا النتيجة انعكاسات على طول الاختبار يجب أن تؤخذ بالاعتبار في حالة وجود طلبة من ذوي الحاجات الخاصة. كما قام كاتو وموين وثيرلو (Kato, Moen & Thurlow, 2009) بدراسة الأداء التفاضلي للفقرات والمموهات في اختبار للقراءة للصفين الثالث والخامس بين فئات إعاقة مختلفة مقارنة بالعادين باستخدام الانحدار اللوجستي المتعدد. وتبين منها وجود أنماط مختلفة من DIF و DDF.

ومن الناحية النظرية، توجد علاقة بين الأداء التفاضلي للمموهات (DDF) والأداء التفاضلي للفقرات (DIF)، فقد أكد بنفيلد (Penfield, 2008) أهمية فحص الأداء التفاضلي للمموهات في توفير معلومات تتعلق بموقع تغير القياس وأسبابه في اختبار اختيار من متعدد. وقد اقترح تقديراً لنسبة الأرجحية لأثر DDF. وبالاعتماد على بيانات مولدة تحقق من محافظة ذلك التقدير على خصائص إحصائية مستقرة في ظل ظروف مختلفة. وقد درس بنفيلد (Penfield, 2010a) كيفية نمذجة أثر DIF كدالة بدلالة أثر DDF لكل المموهات. وقد مكن ذلك من الكشف عن الظروف التي

متحررة من القيود الاقتصادية والعمرية على المستجيبين، بينما كانت البيانات في دراسة جرينبيرغ (Greenberg, 2010) تخص ذوي الدخل المتدني في عمر ما قبل المدرسة. وكذلك فإنها أوسع نطاقاً من دراسة تساويس وسيدرديس والصاوي (Tsaousis, Sideridis, & Al-Saawi, 2018) في السعودية، حيث تدرس العلاقة بين أنماط DDF وأثر DIF على مستوى جميع الفقرات في الاختبار، وعدم اقتصرها فقط على الفقرات ذات الأداء التفاضلي كما في الدراسة السعودية. وعليه، فقد سعت الدراسة الحالية لفحص العلاقة بين الأداء التفاضلي للموهبات والأداء التفاضلي للفقرات اعتماداً على بيانات حقيقية وفرتها الدراسة التقييمية الوطنية في الأردن حول أداء طلبة الصف الخامس الأساسي في اختبار الرياضيات.

مشكلة الدراسة وسؤالها

يعتمد الاختبار الجيد على فقرات جيدة، وتعتمد الفقرة الجيدة من نوع الاختبار من متعدد على موهبات جيدة. ومن معايير الفقرة الجيدة أن تكون ذات فاعلية وعادلة، وتقاس الفاعلية للفقرة وموهباتها بقدراتها التمييزية. ويُعد الأداء التفاضلي للفقرة والموهبات من المؤشرات والأساليب المستخدمة في بحث عدالة الفقرة التي تعد جزءاً مهماً من صدقها. وفي الوقت الذي توجد فيه علاقة على صورة تركيب جمعي خطي بين القدرة التمييزية للفقرة والقدرات التمييزية لموهباتها، فإن العلاقة بين الأداء التفاضلي للفقرة والأداء التفاضلي للموهبات ليست بنفس الدرجة من الوضوح من ناحية عملية. فدراسات بنفيلد ذات طبيعة نظرية أو تقوم على بيانات مولدة، وقد أدت إلى بعض النتائج التي تتعارض مع بعض الأفكار البديهية. كما أنه، لم يسبق تحديد العلاقة على نطاق واسع- حسب علم الباحثين- في أي دراسة عربية سابقة؛ لذلك سعت الدراسة الحالية إلى كشف العلاقة بين الأداء التفاضلي للفقرة والأداء التفاضلي للموهبات. وتتميز الدراسة الحالية بأنها ستفيد في تحديد أنماط الأداء التفاضلي للموهبات التي تنتج أداءاً تفاضلياً للفقرة، وهذا يفتح الباب لدراسة الفقرة ككل ومراجعتها بعمق من خلال الدخول إلى جزئيات الفقرة، حتى تكون أكثر عدالة بين المجموعات المختلفة من المفحوصين، بعد ضبط القدرة المستهدفة في القياس لدى تلك المجموعات. وبالتحديد تحاول الدراسة الإجابة عن السؤال الآتي:

ما العلاقة بين الأداء التفاضلي للموهبات والأداء التفاضلي للفقرة في اختبار رياضيات للصف الخامس من نوع الاختبار من متعدد؟

مصطلحات الدراسة

1. الأداء التفاضلي للفقرة (DIF): هو مؤشر إحصائي للتعبير عن الفروق في الاستجابة الصحيحة على الفقرة بين مجموعتين من المفحوصين ممن لهم المستوى نفسه من القدرة (Camilli & Shepard, 1994).

الرياضيات أن هناك فرقاً ذا دلالة إحصائية في الوسط الحسابي بين الذكور والإناث لصالح الإناث، وبين طلبة مدارس المدن والقرى لصالح طلبة مدارس المدن. ونظراً لعدم وجود أدلة يقينية كافية لتفسير مثل هذه الفروق التي قد تعزى لوجود اختلافات في القدرات الرياضية بين مجموعات المقارنة، أو لاختلافات في أداء الفقرات والاختبار في المجموعات المختلفة. وفي مسعى لفهم الفروق في الأداء على اختبار الرياضيات التي تعزى لجنس الطالب وموقع المدرسة، قام العجلوني والسوالمية (مقبول للنشر) بفحص الأداء التفاضلي لفقرات ذلك الاختبار بحسب جنس الطالب وموقع المدرسة. وقد تبين وجود 8 فقرات ذات أداء تفاضلي لجنس الطالب، و8 فقرات ذات أداء تفاضلي لموقع المدرسة. ولإعطاء صورة أوضح عن الاختبار، قام العجلوني والسوالمية (2016) بفحص الأداء التفاضلي لموهبات فقرات الاختبار وفقاً لجنس الطالب وموقع المدرسة. وقد تبين نتيجة الفحص وجود 15 موهماً ذات أداء تفاضلي بالنسبة لمتغير الجنس، ووجود 15 موهماً ذات أداء تفاضلي لمتغير موقع المدرسة.

وفي السعودية، درس تساويس وسيدرديس والصاوي (Tsaousis, Sideridis & Al-Saawi, 2018) الأداء التفاضلي للمموه كوسيلة لتحسين الخصائص السيكومترية لاختبار الكيمياء ضمن اختبار القبول التحصيلي المعياري (Standard Achievement Admission Test)، وقد بينت الدراسة أن هناك 5 فقرات ذات أداء تفاضلي. ولفهم أسباب الأداء التفاضلي لهذه الفقرات، تم فحص الأداء التفاضلي لموهباتها. ويفيد تحليل الأداء التفاضلي للموهبات في تسليط الضوء على المصادر المحتملة للتباين غير المرتبط بالبناء (Construct-Irrelevant Variance). وبناء على نتائج التحليل أجريت مجموعة من التعديلات على الفقرات ذات الأداء التفاضلي من أجل تحسين الخصائص السيكومترية للاختبار.

وتأتي الدراسة الحالية استكمالاً لدراسة العجلوني والسوالمية (2016) واستجابة لتوصيتها الثانية المتعلقة بتحديد العلاقة بين الأداء التفاضلي للموهبات والأداء التفاضلي للفقرات بحسب جنس الطالب وموقع المدرسة. وتتشابه هذه الدراسة مع دراسات بنفيلد (Penfield, 2008, 2010a, 2010b) ودراسة جرينبيرغ (Greenberg, 2010) ودراسة تساويس وسيدرديس والصاوي (Tsaousis, Sideridis, & Al-Saawi, 2018) من حيث الهدف المتمثل ببحث العلاقة بين أثر DDF وأثر DIF. ولكنها تختلف عن دراسات بنفيلد (Penfield, 2008, 2010a, 2010b) باعتبارها على بيانات فعلية تتمثل باستجابات عينة من طلبة الصف الخامس على فقرات الاختبار من متعدد في اختبار الرياضيات في الدراسة التقييمية الوطنية لاقتصاد المعرفة في الأردن. بينما اتسمت دراسات بنفيلد (Penfield, 2008, 2010a, 2010b) بكونها ذات طبيعة نظرية (نمذجة رياضية)، وبالاعتماد على بيانات مولدة. كما تختلف عن دراسة جرينبيرغ (Greenberg, 2010) بأن بياناتها

مرحلتين، في المرحلة الأولى اختيرت المدرسة وفي المرحلة الثانية اختيرت الشعبة، وروعي في الاختيار السلطة التربوية المشرفة وموقع المدرسة وجنسها. وقد تم استبعاد 18 طالباً وطالبة نظراً لعدم اكتمال البيانات الخاصة بهم. وبذلك اقتصرت البيانات التي جرى التعامل معها على استجابات (1540) طالباً وطالبة. ويبين الجدول (1) توزيع العينة حسب متغيري الجنس وموقع المدرسة.

جدول 1: توزيع العينة حسب الجنس وموقع المدرسة

المتغير	المستويات	عدد الطلبة	النسبة المئوية
الجنس	ذكور	703	45.6
	إناث	837	54.4
موقع المدرسة	ريف	420	27.3
	مدينة	1120	72.7
المجموع		1540	100

أداة الدراسة

تكونت أداة الدراسة من اختبار الرياضيات للصف الخامس في الدراسة الوطنية لتقييم مهارات اقتصاد المعرفة (المركز الوطني لتنمية الموارد البشرية، 2011). يتضمن الاختبار (26) فقرة من نوع الاختيار من متعدد، و(7) فقرات مقالية. وتقيس فقرات الاختبار مهارات أساسية في الرياضيات تتمثل بالكفايات التعليمية التالية: مهارة التواصل، وإدارة المعلومات، واستخدام الأرقام والرموز، والتفكير وحل المشكلات. ولأغراض الدراسة الحالية، تم الاعتماد على فقرات الاختيار من متعدد. وقد تم التحقق في الدراسة الوطنية من دلالات صدق الاختبار وثباته. وقد توفر للاختبار أدلة صدق البناء القائم على المحتوى النظري لمهارات المعرفة المعتمدة في اختبار PISA الدولي، وتم بناء جدول مواصفات، وكتابة الفقرات بناءً عليه وتحكيمها من مجموعة الخبراء القائمين على الدراسة الوطنية. كما تم التوصل إلى أدلة صدق بناء بالاستناد إلى محك خارجي، من خلال مقارنة ثلاث مجموعات من النسب المئوية: النسب المئوية لمستويات أداء الطلبة على الاختبار وفق النظرية الكلاسيكية في القياس، والنسب المئوية لمستويات الأداء على الاختبار وفق نظرية استجابة الفقرة، والنسب المئوية لمستويات أداء الأردنيين على اختبار PISA. وقد كانت النسب الثلاث متقاربة لدرجة كبيرة في مستويات الأداء المختلفة. ويعد هذا التقارب دليلاً على صدق الاختبار. وقد تم التحقق من دلالات الثبات له باستخدام معامل ثبات الاتساق الداخلي (كرونباخ ألفا)، حيث بلغ 0.78، وتعد هذه القيمة مناسبة لتوفير بيانات مناسبة لغرض الدراسة الحالية.

المعالجة الإحصائية

تم الكشف عن الأداء التفاضلي للفقرات والمموهات باستخدام برمجية (DDFS) بحسب متغيري جنس الطالب وموقع المدرسة، بناءً على المؤشرات والمعايير التي حددتها مؤسسة خدمات

2. المموهات (Distractors): هي بدائل الإجابات غير الصحيحة للفقرات التي تكون من نوع الاختيار من متعدد.
3. الأداء التفاضلي للمموهات (DDF): هو مؤشر إحصائي للتعبير عن الفروق في الاستجابة على المموه بين مجموعتين من المفحوصين ممن لهم المستوى نفسه من القدرة في كل مجموعة (Penfield, 2008).

أهمية الدراسة

تبرز أهمية الدراسة الحالية من الناحية النظرية في تسليط المزيد من الضوء على العلاقة بين الأداء التفاضلي لفقرة الاختبار والأداء التفاضلي لمموهاتها. ومن الناحية التطبيقية تفيد الدراسة في تحديد أنماط الأداء التفاضلي للمموهات المؤثرة في الأداء التفاضلي للفقرات. ويفيد ذلك في تحديد بعض أسباب الأداء التفاضلي للفقرات، وبمعالجة تلك الأسباب خلال عملية المراجعة للفقرات لتحسن عدالة الاختبار وصدقه. وتفيد المعلومات التي يوفرها تحليل DIF وتحليل DDF معا في جعل عمليتي كتابة الفقرات ومراجعتها أكثر كفاءة وفعالية، وينعكس ذلك بصورة إيجابية على عدالة الاختبار. كما تفيد في إعطاء صورة أوضح عن أداء فقرات الاختبار ومموهاتها في مجموعات المقارنة المختلفة، وتوفير الأدلة التي ترشد باني الاختبار في عملية صياغة فقرات الاختيار من متعدد وانتقاء بدائلها لتحقيق قدر معقول من العدالة لدى المجموعات المختلفة في مجتمع الاختبار. فمراجعة المموهات ذات الأداء التفاضلي وتعديلها يحسن من الخصائص السيكومترية للاختبار، وهذه قضية على درجة كبيرة من الأهمية لكونها تتعلق بصدق الاختبار وعدالته التي تدعم المساواة في التعامل والاتساق في المعنى والتفسير.

محددات الدراسة

تقتصر الدراسة الحالية على فحص الأداء التفاضلي لمتغيري الجنس وموقع المدرسة لطلبة الصف الخامس الأساسي في الأردن في اختبار الرياضيات في التقييم الوطني لاقتصاد المعرفة (Nafke, 2011). كما تقتصر على استخدام طريقة نسبة الأرجحية لمانتل هانزل في الكشف عن الأداء التفاضلي للمموهات (DDF) والفقرات (DIF).

الطريقة

بيانات الدراسة

توفرت بيانات الدراسة الحالية من الدراسة الوطنية لتقييم مهارات الاقتصاد المعرفي التي نفذها المركز الوطني لتنمية الموارد البشرية في عام 2011. وتتكون البيانات المستخدمة من استجابات 1558 طالباً وطالبة في الصف الخامس على 26 فقرة اختيار من متعدد تقيس مهارات اقتصاد المعرفة الأساسية في مبحث الرياضيات. وقد تم اختيارهم بطريقة عشوائية عنقودية على

3) مؤشر BD: يدل على نوع الأداء التفاضلي
القيمة فوق 3.84 ($BD > 3.84$) تعني أن الأداء التفاضلي غير
منتظم.
القيمة دون 3.84 ($BD < 3.84$) تعني أن الأداء التفاضلي
منتظم.
وسيتم استخدام اختبار مربع كاي للاستقلال لفحص العلاقة
بين الأداء التفاضلي للموهبات والأداء التفاضلي للفقرات.

النتائج

أولاً: الأداء التفاضلي DIF و DDF بحسب الجنس

أظهرت نتائج تحليل الأداء التفاضلي بحسب جنس الطالب
لموهبات (78 موهماً) وفقرات الاختبار (26 فقرة) أن هناك ثمانين
فقرات ذات أداء تفاضلي (6 فقرات ذات أداء تفاضلي منتظم وفقرتان
ذات أداء تفاضلي غير منتظم) من ضمنها الفقرات 1، 8، و 26 التي
تعد فقرات إشكالية؛ نظراً لكون الأداء التفاضلي لها متوسط الحجم،
بينما كان صغيراً لبقية الفقرات. والأداء التفاضلي في الفقرتين 1
و 26 من النوع المنتظم، بينما كان من النوع غير المنتظم في الفقرة
8. كما أظهرت النتائج أن هناك خمسة عشر موهماً ذات أداء
تفاضلي. وقد توزعت الموهبات ذات الأداء التفاضلي على عشر
فقرات. وبالمحصلة فإن الاختبار يتضمن أربع عشرة فقرة نقية من
أي أداء تفاضلي. أي أنها ليست ذات أداء تفاضلي، ولا تتضمن
موهبات ذات أداء تفاضلي. ويبين الجدول (2) توزيع الفقرات
حسب كونها ذات أداء تفاضلي أو لا (DIF, NO DIF)، وحسب
كونها تتضمن موهبات ذات أداء تفاضلي أو لا (DDF, NO
DDF).

الاختبارات التربوية Penfield & Educational Testing Services
(Camilli, 2007):

1) مؤشر LOR (log-odds ratio) ويدل على اتجاه وحجم الأداء
التفاضلي بحيث:

• الاتجاه (Direction)

القيمة الموجبة ($LOR > 0$) تعني أن الأداء التفاضلي لصالح
المجموعة المرجعية.

القيمة السالبة ($LOR < 0$) تعني أن الأداء التفاضلي لصالح
المجموعة المستهدفة.

• الحجم (Magnitude)

القيمة المطلقة دون 0.43 تعني أن أثر الأداء التفاضلي
صغير.

القيمة المطلقة بين 0.43 و 0.64 تعني أن أثر الأداء
التفاضلي متوسط.

القيمة المطلقة فوق 0.64 تعني أن أثر الأداء التفاضلي كبير.

2) مؤشر (LOR) Z: يدل على وجود أداء تفاضلي للفقرة عند

مستوى الدلالة الإحصائية $\alpha = 0.05$ عندما ($Z > 2$) أو ($Z < -2$)

و للمموه عندما ($Z > 2.4$) أو ($Z < -2.4$). ويلاحظ هنا

اختلاف درجة القطع للمموه عن درجة القطع للفقرة. ويعود

ذلك إلى وجود 3 موهبات وإجابة صحيحة واحدة.

والاستخدام المتكرر للاختبار مع الموهبات يرفع من قيمة α .

وللمحافظة على قيمة α (مثلاً 0.05) تكون قيمتها الاسمية

الأولية $\frac{0.05}{3}$ تقريباً وهي تقابل القيمة الحرجة 2.4.

جدول 2: توزيع الفقرات حسب متغيري الأداء التفاضلي للموهبات والأداء التفاضلي للفقرات بحسب الجنس

المجموع	هل تتضمن الفقرة موهبات ذات أداء تفاضلي؟		نعم DIF	لا DIF
	لا DDF	نعم DDF		
8	2	6	نعم DIF	هل الفقرة ذات أداء تفاضلي؟
18	14	4	لا DIF	
26	16	10		المجموع

3، الذي يبين التوزيع التكراري لعدد الموهبات ذات الأداء
التفاضلي بحسب الجنس.

ويلاحظ من الجدول 3 انه يمكن أن تكون الفقرة ذات أداء
تفاضلي دون أن تتضمن موهبات ذات أداء تفاضلي لكن باحتمال
ضعيف، ويمكن أن تتضمن موهماً واحداً ذا أداء تفاضلي ولا تكون
ذات أداء تفاضلي.

ولفحص دلالة العلاقة بين كون الفقرة تتضمن موهبات ذات
أداء تفاضلي DDF وكونها ذات أداء تفاضلي DIF، تم استخدام
اختبار مربع كاي للاستقلال، وتبين أن قيمته تساوي 6.518، وهي
ذات دلالة إحصائية عند مستوى الدلالة $\alpha = 0.05$. وهذا يشير
إلى وجود علاقة موجبة بين المتغيرين؛ إذ يتوقع أن يزداد احتمال
أن تكون الفقرة ذات أداء تفاضلي بازدياد احتمال أن تتضمن
موهبات ذات أداء تفاضلي. ويتضح ذلك بصورة جلية في الجدول

جدول 3: التوزيع التكراري لعدد المموهات ذات الأداء التفاضلي في الفقرة بحسب الجنس

عدد المموهات ذات الأداء التفاضلي في الفقرة	عدد الفقرات	عدد الفقرات ذات الأداء التفاضلي	احتمال أن تكون الفقرة ذات أداء تفاضلي
0	16	2	0.125
1	6	2	0.33
2	3	3	1.0
3	1	1	1.0

ويبين الجدول 4 أمثلة على استجابات البنين والبنات على الفقرات في أنماط مختلفة من الأداء التفاضلي للفقرة أو المموهات بحسب الجنس.

جدول 4: أمثلة على استجابات الطلاب في أنماط مختلفة من الأداء التفاضلي للفقرة أو المموهات بحسب الجنس*

أنماط DIF و DDF	رقم الفقرة والمموهات	المجموعة ذات الأفضلية	مقدار الأداء التفاضلي	نوع الأداء التفاضلي	عدد البنات	عدد البنين
No DDF, No DIF (14 فقرة)	10	بنين	صغير	منتظم	180	168
	A	بنين	صغير		53	36
	B	بنين	صغير		420	319
	C	بنين	صغير		98	80
	بدون إجابة				86	100
No DDF, DIF (فقرتان)	5	بنات	صغير	منتظم	567	414
	A	بنات	متوسط		23	27
	C	بنات	صغير		95	90
	D	بنات	صغير		123	121
	بدون إجابة				29	51
DDF, No DIF (4 فقرات)	6	بنات	صغير	منتظم	377	283
	A	بنات	صغير		258	207
	B	بنات	كبير		73	110
	C	بنين	صغير		81	57
	بدون إجابة				48	46
DDF, Small DIF (3 فقرات)	11	بنين	صغير	منتظم	261	246
	A	بنين	صغير		143	118
	B	بنات	صغير		90	107
	C	بنين	متوسط		305	169
	بدون إجابة				38	63
DDF, Moderate DIF (3 فقرات إشكالية)	1	بنات	متوسط	منتظم	525	363
	A	بنات	صغير		135	129
	B	بنات	متوسط		75	97
	C	بنات	متوسط		61	74
	بدون إجابة				41	40
	8	بنات	متوسط	غير منتظم	651	450
	A	بنات	متوسط		45	57
	C	بنات	كبير		78	107
	D	بنات	صغير		39	36
	بدون إجابة				24	53
بدون إجابة	26	بنين	متوسط	غير منتظم	126	150
	A	بنين	كبير		136	89
	B	بنين	صغير		392	311
	D	بنين	كبير		110	68
بدون إجابة				73	85	

* الخط الغامق يشير أن الأداء التفاضلي للفقرة أو المموه دال إحصائياً عند مستوى الدلالة $\alpha=0.05$.

يلاحظ من جدول 4 وجود خمسة أنماط مختلفة من الأداء التفاضلي للفقرات والمموهات بحسب الجنس الطالب في اختبار الرياضيات كما في العمود الأول. ويلاحظ أنه من المحتمل أن تكون الفقرة ذات أداء تفاضلي للجنس حتى لو لم تكن المموهات فيها ذات أداء تفاضلي، ومن المحتمل كذلك أن لا تكون الفقرة ذات أداء تفاضلي بالرغم من كون أحد مموهاتها ذات أداء تفاضلي، وأنه حتى

غير المنتظم في الفقرتين 7 و9. وأظهرت نتائج تحليل الأداء التفاضلي DDF للمموهات (78 مموه) أن هناك خمسة عشر مموها ذات أداء تفاضلي. وقد توزعت المموهات ذات الأداء التفاضلي على إحدى عشرة فقرة. وبالمحصلة فإن الاختبار يتضمن خمس عشرة فقرة نقيه من أي أداء تفاضلي. أي أنها ليست ذات أداء تفاضلي ولا تتضمن مموهات ذات أداء تفاضلي. ويبين الجدول (5) توزيع الفقرات حسب كونها ذات أداء تفاضلي أو لا (DIF, NO DIF) وحسب كونها تتضمن مموهات ذات أداء تفاضلي أو لا (DDF, NO DDF).

تكون الفقرات إشكالية لا بد من وجود مموهين على الأقل فيها ذوي أداء تفاضلي من الحجم المتوسط المقدار على الأقل.

ثانياً: الأداء التفاضلي بحسب موقع المدرسة

أظهرت نتائج تحليل الأداء التفاضلي DIF بحسب موقع المدرسة لفقرات الاختبار (26 فقرة) أن هناك ثماني فقرات ذات أداء تفاضلي (6 فقرات ذات أداء تفاضلي منتظم وفقرتان ذاتي أداء تفاضلي غير منتظم) من ضمنها الفقرات 6، 8، و16 التي تعد فقرات إشكالية نظراً لكون الأداء التفاضلي لها متوسطاً أو كبير الحجم، بينما كان صغيراً لبقية الفقرات. والأداء التفاضلي في الفقرات 5، 6، 8، 10، 16، 22، 26 من النوع المنتظم بينما كان من النوع

جدول 5: توزيع الفقرات حسب متغيري الأداء التفاضلي للمموهات وللفقرات بحسب موقع المدرسة

المجموع	هل تتضمن الفقرة مموهات ذات أداء تفاضلي؟		هل الفقرة ذات أداء تفاضلي؟
	لا DDF	نعم DDF	
8	1	7	نعم DIF
18	14	4	لا DIF
26	15	11	المجموع

يبين التوزيع التكراري لعدد المموهات ذات الأداء التفاضلي بحسب موقع المدرسة.

ويلاحظ من الجدول 6 أنه يمكن أن تكون الفقرة ذات أداء تفاضلي دون أن تتضمن مموهات ذات أداء تفاضلي لكن باحتمال ضعيف جداً، ويمكن أن تتضمن مموهاً واحداً ذا أداء تفاضلي ولا تكون ذات أداء تفاضلي.

ولفحص دلالة العلاقة بين كون الفقرة تتضمن مموهات ذات أداء تفاضلي DDF وكونها ذات أداء تفاضلي DIF، تم استخدام اختبار مربع كاي للاستقلال، وتبين أن قيمته تساوي 9.656، وهي دالة إحصائياً عند مستوى الدلالة $\alpha = 0.01$. وهذا يشير إلى وجود علاقة موجبة بين المتغيرين؛ إذ يتوقع أن يزداد احتمال أن تكون الفقرة ذات أداء تفاضلي بازدياد احتمال أن تتضمن مموهات ذات أداء تفاضلي. ويتضح ذلك بصورة جلية في الجدول 6، الذي

جدول 6: التوزيع التكراري لعدد المموهات ذات الأداء التفاضلي بحسب موقع المدرسة

عدد المموهات ذات الأداء التفاضلي في الفقرة	عدد الفقرات	عدد الفقرات ذات الأداء التفاضلي	احتمال أن تكون الفقرة ذات أداء تفاضلي
0	15	1	0.066
1	8	4	0.50
2	2	2	1.0
3	1	1	1.0

ويبين الجدول 7 أمثلة على استجابات الطلاب في مدارس المدن والريف على الفقرات في أنماط مختلفة من الأداء التفاضلي للفقرة أو المموهات بحسب موقع المدرسة.

جدول 7: أمثلة على استجابات الطلاب في أنماط مختلفة من الأداء التفاضلي للفقرة أو المموهات بحسب الموقع*

استجابات المدن	استجابات الريف	نوع الأداء التفاضلي	مقدار الأداء التفاضلي	المجموعة ذات	الفقرة والمموهات	أنماط DIF و DDF
157	73	منتظم	صغير	ريف	4	No DDF, No DIF (فقرة 14)
317	116		صغير	ريف	B	
251	98		صغير	ريف	C	
152	46		متوسط	ريف	D	
243	87				بدون إجابة	
365	142	منتظم	صغير	ريف	11	No DDF, DIF (فقرتان) (فقرة واحدة)
200	61		صغير	ريف	A	
132	65		صغير	مدن	B	
357	117		صغير	ريف	C	
66	35				بدون إجابة	
342	126	منتظم	صغير	ريف	13	DDF, No DIF (4 فقرات)
229	99		صغير	مدن	A	
256	105		صغير	ريف	B	
186	48		متوسط	ريف	D	
107	42				بدون إجابة	
751	230	منتظم	صغير	مدن	5	DDF, Small DIF (4 فقرات)
37	13		صغير	مدن	A	
134	51		صغير	مدن	C	
154	90		متوسط	مدن	D	
44	96				بدون إجابة	
534	126	منتظم	كبير	مدن	6	DDF, Moderate & Large DIF (3 فقرات إشكالية)
326	139		متوسط	مدن	A	
99	84		كبير	مدن	B	
97	41		متوسط	مدن	C	
64	30				بدون إجابة	
847	254	منتظم	متوسط	مدن	8	
57	45		كبير	مدن	A	
126	59		متوسط	مدن	C	
44	31		كبير	مدن	D	
46	31				بدون إجابة	
132	59	منتظم	متوسط	ريف	16	
292	77		كبير	ريف	A	
353	160		صغير	ريف	B	
203	64		متوسط	ريف	C	
140	60				بدون إجابة	

* الخط الغامق يشير أن الأداء التفاضلي للفقرة أو المموه دال إحصائياً عند مستوى الدلالة $\alpha = 0.05$.

يلاحظ من جدول 7 وجود خمسة أنماط مختلفة من الأداء التفاضلي للفقرات والمموهات بحسب موقع المدرسة في اختبار الرياضيات كما في العمود الأول. ويلاحظ أنه من المحتمل أن تكون الفقرة ذات أداء تفاضلي لموقع المدرسة حتى لو لم تكن المموهات فيها ذات أداء تفاضلي، ومن المحتمل كذلك أن لا تكون الفقرة ذات أداء تفاضلي بالرغم من كون أحد مموهاتها ذا أداء تفاضلي، وأنه حتى تكون الفقرات إشكالية لا بد من وجود مموهين على الأقل فيها ذوي أداء تفاضلي من الحجم المتوسط المقدار على الأقل.

حتى تكون الفقرات إشكالية لا بد من وجود مموهين على الأقل فيها ذوي أداء تفاضلي من الحجم المتوسط المقدار على الأقل.

مناقشة النتائج

تشير نتائج تحليل الأداء التفاضلي للفقرات إلى أن عدد الفقرات الإشكالية (ذات أداء تفاضلي متوسط أو كبير الحجم) في الاختبار لا تزيد على 3 فقرات من بين 26 فقرة، سواء فيما يتعلق بالجنس أو موقع المدرسة. فكل من الفقرة 1 التي تتعلق باستخراج قيمة ما يدفعه المواطن من رسوم تلفزيون في فاتورة الكهرباء، والفقرة 8 التي تتعلق باستخراج قيمة عددية من رسم بياني بطريقة الأعمدة، هي ذات أداء تفاضلي لصالح الإناث. مما يعني أن لدى الإناث مهارة أعلى من الذكور في استخراج قيم عددية مباشرة من الجداول والرسومات البيانية المختلفة. أما الفقرة 26 التي تتعلق بالتحويلات المركبة فيما بين المقاييس (مسطرة، سم، متر) فهي ذات أداء تفاضلي لصالح الذكور. وتتعلق كل من الفقرتين 6، 8 بالرسوم البيانية، وهي ذات أداء تفاضلي لصالح طلبة المدن. أما الفقرة 16 التي تتعلق بأبعاد ومساحة مزرعة فهي ذات أداء تفاضلي لصالح طلبة الريف.

وفيما يخص العلاقة بين الأداء التفاضلي للفقرة DIF والأداء التفاضلي لمموماتها DDF فقد تبين وجود علاقة موجبة بينهما. ويصعب تحديد صورة رياضية للعلاقة بين DIF و DDF على غرار الصورة الرياضية للعلاقة بين تمييز الفقرة وتمييز الممومات. وقد يعزى ذلك إلى اختلاف مستوى القدرة في حالة التمييز - حيث توجد مجموعة عليا ومجموعة دنيا - بينما يتم ضبط مستوى القدرة في تحليل الأداء التفاضلي بين المجموعة المعيارية والمجموعة المستهدفة.

كما قد يعزى إلى اختلاف الأساس الرياضي الذي يقوم عليه تمييز الفقرة والممومة عن الأساس الرياضي الذي يقوم عليه الأداء التفاضلي للفقرة وللممومة. فتمييز الفقرة يقوم على الفرق بين نسبة الإجابات الصحيحة في المجموعة العليا ونسبة الإجابات الصحيحة في المجموعة الدنيا، بينما يقوم الأداء التفاضلي على نسبة أرجحية الإجابة الصحيحة في المجموعة المعيارية إلى ما يناظرها في المجموعة المستهدفة. وتتفق نتائج هذه الدراسة مع نتائج دراسة جرينبيرغ (Greenberg, 2010) ودراسات بنفيلد (Penfield, 2010a, 2010b) التي أكدت جميعها أن تأثيرات DDF تحدد حجم ونوع DIF. كما تتفق مع نتائج دراسة كاتو وشركاه (Kato et al., 2009)، حيث تبين وجود أنماط مختلفة من DIF و DDF. والأنماط المحتملة هي:

- النمط (No DDF, NO DIF) وهذا يمثل الفقرات النقية من الأداء التفاضلي للفقرة وللممومات، ويسعى مطورو الاختبارات

لأن تكون نسبة هذه الفقرات في الاختبار عالية. وفي هذه الفقرات يكون الفرق في أرجحية الإجابة الصحيحة بين المجموعة المعيارية والمجموعة المستهدفة صغيراً وغير جوهري. وكذلك الحال في جميع الممومات للفقرة.

- النمط (No DDF, DIF) وهذا يمثل الفقرات ذات الأداء التفاضلي التي تخلو مموماتها من الأداء التفاضلي، وهي قليلة الاحتمال. ويحدث هذا النمط نتيجة اختلاف واضح بين المجموعة المرجعية والمجموعة المستهدفة في أرجحية ترك الفقرة بدون إجابة، أو في المساهمة الكلية للممومات مجتمعة في أرجحية الإجابة الصحيحة.

- النمط (DDF, NO DIF) وهذا يمثل الفقرات النقية من الأداء التفاضلي للفقرة، بينما تتضمن بعض الممومات ذات الأداء التفاضلي. ويحدث هذا النمط عندما تكون الممومات ذات أداء تفاضلي ضعيف، أو متضادة في الاتجاه.

- النمط (DDF, DIF) وهذا يمثل الفقرات ذات الأداء التفاضلي التي تتضمن ممومات ذات أداء تفاضلي. وتشكل هذه الفقرات مصدر تهديد لمصداقية الاختبار، وهي فقرات إشكالية وبالأخص عندما يكون الأداء التفاضلي متوسطاً أو كبيراً. ويجدر دراسة هذه الفقرات دراسة معمقة وتنقيحها.

وعليه فإن هناك أهمية لاستخدام تحليل الأداء التفاضلي للممومات (DDF) جنباً إلى جنب مع تحليل DIF، فهو يعطي صورة أوضح عن أنماط اختيار المفحوصين لبدائل الإجابة. ويفيد في فهم الأداء التفاضلي للفقرات (DIF). فالأثر الكبير الناتج عن الأداء التفاضلي لأحد الممومات مقارنة مع الأثر الناتج عن بقية الممومات، يعني أن ذلك الممومة كان جذاباً لإحدى المجموعتين دون الأخرى، مما يدل على أن سبب وجود الأداء التفاضلي للفقرة يتمثل في خصائص هذا الممومة، أو في التفاعل بين خصائص هذا الممومة وبين محتوى الفقرة، ويتطلب الوضع في هذه الحالة إجراء تحليل لمحتوى خصائص هذا الممومة ولتفاعل هذه الخصائص مع محتوى الفقرة نفسه، وإجراء تعديل للممومة وليس للفقرة أو للإجابة الصحيحة. وتساعد المعلومات المتوافرة عن الأداء التفاضلي للممومات في مراجعة الفقرات ذات الأداء التفاضلي، وينعكس ذلك بشكل إيجابي على الاختبار، يتمثل في ضمان درجة أعلى في عدالته.

التوصيات

في ظل النتائج التي توصلت إليها الدراسة الحالية، يمكن الخروج بالتوصيات التالية:

1. إعطاء مزيد من الاهتمام باختبار الرياضيات لتقييم مهارات اقتصاد المعرفة ومراجعتها وتطويره بما يحسن من عدالة الاختبار، من خلال تنقيح الفقرات الإشكالية ذات الأداء التفاضلي المتوسط أو الكبير للجنس وموقع المدرسة.
2. نظراً لمحدودية نتائج هذه الدراسة، فإنه يوصى بإجراء دراسات أخرى حول العلاقة بين الأداء التفاضلي للفقرات والأداء التفاضلي للمموهات باستخدام طرق أخرى للكشف على الأداء التفاضلي واختبارات أخرى.
3. توصي الدراسة بأن يكون تحليل الأداء التفاضلي للفقرات DIF وتحليل الأداء التفاضلي للمموهات DDF من ضمن إجراءات بناء وتطوير الاختبارات الوطنية.

المراجع

- العجلوني، جهاد والسؤالمة، يوسف. (2016). فحص الأداء التفاضلي لمموهات فقرات اختبار الرياضيات في التقييم الوطني الأردني لاقتصاد المعرفة وفقاً لجنس الطالب وموقع المدرسة. *مجلة جامعة النجاح للابحاث (العلوم الانسانية)*، 30(11)، 2265-2286.
- العجلوني، جهاد والسؤالمة، يوسف. (مقبول للنشر). فحص الأداء التفاضلي لفقرات اختبار الرياضيات للصف الخامس في الدراسة الأردنية التقييمية الشاملة لمهارات اقتصاد المعرفة وفقاً لجنس الطالب وموقع المدرسة. *مجلة اتحاد الجامعات العربية للتربية وعلم النفس*.
- العيوي، إيمان. (2004). *الكشف عن الأداء التفاضلي لمتغير الجنس لفقرات اختبار تحصيلي في العلوم العامة للصف الثامن الأساسي في مديرية تربية وتعليم عمان الرابعة*. رسالة ماجستير غير منشورة، جامعة عمان العربية للدراسات العليا، عمان، الأردن.
- فريحات، نجود. (2014). *الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في العلوم والرياضيات للصف العاشر الأساسي تبعاً لطريقة تقديم الاختبار*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.
- مبارك، وائل. (2010). *الأداء التفاضلي لفقرات اختبار العلوم في الدراسة الدولية بيزا 2006*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.
- المركز الوطني لتنمية الموارد البشرية. (2011). *الدراسة الوطنية التقييمية الشاملة لمهارات اقتصاد المعرفة*. عمان، الأردن.
- الناجي، عادل. (2011). *فاعلية طريقة التحليل العائلي المقيد مقارنة بعدة طرق للكشف عن الأداء التفاضلي للفقرات*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.
- Abedi, J.; Leon, S.; & Kao, J. (2007). *Examining differential distractor functioning in reading assessments for students with disabilities*. Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment. Retrieved from: <https://nceo.umn.edu/docs/OnlinePubs/PARA/examiningddf/examiningDDReport>.
- Ackerman, T. (1992). A didactic explanation of item bias, item impact and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29, 67-91.
- Allen, M. & Yen, W. (1979). *Introduction to measurement theory*. Monterey, California: Brooks/Cole Publishing Company.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- أبو حماد، كوثر. (2008). *تقصي فاعلية منحنى خصائص الفقرة في الكشف عن الأداء التفاضلي لمتغير الجنس في عينة مختارة من اختبار في الرياضيات*. رسالة ماجستير غير منشورة، جامعة اليرموك، إربد، الأردن.
- أيوب، عبد الكريم. (2008). *أثر خصائص توزيع القدرة وطول الاختبار وحجم العينة في الكشف عن الدالة التفاضلية للفقرات: دراسة محاكاة*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.
- البيستنجي، محمود. (2004). *مقارنة أربع طرائق للكشف عن الأداء التفاضلي لمتغير الجنس في فقرات اختبار قدرات عقلية خاصة للفئة العمرية من 15-16 سنة في الأردن*. أطروحة دكتوراه غير منشورة، جامعة عمان العربية للدراسات العليا، عمان، الأردن.
- جرادات، عبد الله. (2003). *مقارنة بين طريقتي مانتل هانزل والصعوبة المحولة في الكشف عن تحيز الفقرات*. رسالة ماجستير غير منشورة، جامعة مؤتة، الكرك.
- حمادنة، إياد. (2007). *الأداء التفاضلي لفقرات اختبارات "تحديد الكفاءة اللغوية" في اللغة الإنجليزية في الجامعات الأردنية: دراسة مقارنة*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.
- عبيدات، عبدالله. (2014). *استقصاء الأداء التفاضلي لفقرات اختبار اختيار من متعدد تبعاً لدرجة مخاطرة المفحوصين*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، إربد، الأردن.

- Kato, K.; Moen, E., & Thurlow, L. (2009). Differentials of a state reading assessment: Item functioning, distractor functioning, and omission frequency for disability categories. *Educational Measurement: Issues and Practice*, 28(2), 28-40.
- Koon, S.(2010). *A comparison of methods for detecting differential distractor functioning*. (Order No. 3415232, The Florida State University). ProQuest Dissertations and Theses, 93. Retrieved from [http://search.proquest.com/docview/734610226? Accountid=48928. \(734610226\)](http://search.proquest.com/docview/734610226?Accountid=48928. (734610226)).
- Koon, S. & Kamata, A.(2013). An applied examination of methods for detecting differential distractor functioning. *Quantitative Research In Education*. Retrieved from [http://purl.flvc.org/fsu/fd/FSU_mi gr_fcr-pubs-0005](http://purl.flvc.org/fsu/fd/FSU_mi_gr_fcr-pubs-0005).
- Linn, R. (1993). The use of differential item functioning statistics: A discussion of current practice and future applications. In P. W. Holland & H. Wainer (Eds.). *Differential item functioning* (pp. 349-364). Hillsdale, NJ: Lawrence Erlbaum.
- Messick, S.(1994). Validity of psychological assessment:Validation of inferences from persons' responses as scientific inquiry into score meaning. ETS Research Report Series. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2333-8504.1994.tb01618.x>.
- Middleton, K. & Laitusis, C. (2007). Examining test items for differential distractor functioning among students with learning disabilities. *Educational Testing Service (ETS)*. RR-07-43.
- Pae, T. (2004). DIF for examinees with different academic backgrounds. *Language Testing*, 21(1), 53-72.
- Penfield, D. (2008). An odds ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement*, 45(3), 247-269.
- Penfield, R. (2010a). Modeling DIF effects using distractor level invariance effects: Implications for understanding the causes of DIF. *Applied Psychological Measurement*, 34(3), 151-165.
- Penfield, R. (2010b). How are the form and magnitude of DIF effects in multiple choice items determined by distractor level invariance effects. *Educational and Psychological Measurement*, xx(x),1-14. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.832.2315&rep=rep1&type=pdf>
- Penfield, R. & Camilli, G. (2007). Differential item functioning and item bias. In S. Sinharay, & C. Rao (Eds.),*Handbook of statistics, Vol.26:Psychometrics* (pp.125-167).NewYork: Elsevier.
- Banks, K.(2004). *Exploring racial differences in items that differ in cultural characteristics through differential bundle and distractor Functioning*. (OrderNo. 3138573, The University of Wisconsin–Milwaukee). ProQuest Dissertations and Theses. Retrieved from:[http://search.proquest.com/docview/305111981?accountid=48928. \(305111981\)](http://search.proquest.com/docview/305111981?accountid=48928. (305111981)).
- Cains, J.;Bridglall, B.;& Chatterji,M.(2014).Understanding validity and fairness issues in high-stakes individual testing situations. *Quality Assurance in Education* ,22(1),5-18.
- Camilli, G. (2006). Test fairness. In R. L. Brennan(Ed.) *Educational Measurement*(4th ed., Vol. 4, pp. 221-256). Westport: American Council on Education& Praeger Publishers.
- Camilli, G., & Shepard, A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: SAGE Publications.
- Chiu, C. (2008). *The Effect of English Proficiency on Mathematics Performance: A comparison of item response theory-Based area and Mantel-Haenzel methods*. Thesis from proquest information and learning company, (UMI NO.1450675).
- Dorans, J. & Holland, W. (1993). DIF detection description, Mantel-Haenzel and standardization. In W. Holland & H. Wainer(Eds.).*Differential item functioning*(pp.35-66).Hillsdale,NJ:Lawrence Erlbaum.
- Ellis, B. & Raju, S. (2003).*Test and item bias: What they are, what aren't, and how to detect them*. Measuring up. Assessment issue for teachers, counselors and administrator. (ERIC Document Reproduction Service No.ED480042).
- Greenberg, C. (2010). Fighting bias with statistics: Detecting gender differences in responses to items on a preschool science assessment. (Order No. 3424789,University of Miami). ProQuest Dissertations and Theses,104. Retrieved from [http://search.proquest.com/docview/756923372?accountid=48928. \(756923372\)](http://search.proquest.com/docview/756923372?accountid=48928. (756923372)).
- Haladyna, M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates
- Hambleton, K.; Swaminathan, H., & Rogers, J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications.
- Hong, S. & Roznowski, M. (2001). An investigation of the influence of internal test bias on regression slope. *Applied Measurement in Education*, 14, 351-368.
- Karakaya, I. (2012). An Investigation of item bias in science and technology subtests and mathematic subtests in Level Determination Exam. *Educational Science: Theory &Practice*, 12(1), 222-229.

- Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods*. EM No 60. Retrieved August 20, 2017 from [www.edusci.umuse/digitalAssets/159/59534-em-no -60](http://www.edusci.umuse/digitalAssets/159/59534-em-no-60).
- Yildirim, S. (2008). Comparison of restricted-factor analysis with likelihood-ratio and Mantel-Haenzel Methods in DIF analysis. *Journal of Education*, 34, 297-307.
- Tsaousis, I.; Sideridis, G., & Al-Saawi, F.(2018). Differential distractor functioning as a method for explaining DIF: The case of a national admissions test in Saudi Arabia. *International Journal of Testing*, 18(1), 1-26. Retrieved from [https:// eric.ed.gov/?id=EJ1168025](https://eric.ed.gov/?id=EJ1168025)
- Uiterwijk, H.& Vallen, T.(2005). Linguistic sources of item bias for second generation immigrants in Dutch tests. *Language Testing*, 22(2), 211-234.