المجلة الأردنية في العلوم التربوية، مجلد 17، عدد 2، 2021، 251-269

الأداء التفاضلي لفقرات وموهات اختبار للتفكير الناقد لطلبة المرحلة الجامعية تبعًا لمتغير الجنس

كوثر العبد القادر و يوسف السوالمه*

Doi: //10.47015/17.2.6

تاريخ تسلم البحث: 2020/2/23 تاريخ قبوله: 2020/4/20

ملخص: هدفت الدراسة الحالية إلى فحص الأداء التفاضلي لفقرات ومموهات اختبار للتفكير الناقد على المستوى الجامعي وفقًا لمتغير جنس الطالب. تكوّن الاختبار المستهدف بالدراسة من (49) فقره من نوع الاختيار من متعدد، لكل منها أربعة بدائل، وتغطي فقراته مهارات التفكير الناقد المتمثلة في التحليل، والتفسير، والتقويم، والشرح، والاستدلال، والتبرير الإحصائي والاحتمالي، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها، والتركيز. ولتحقيق هدف الدراسة، تم تحليل استجابات عينة مكونة من (930) طالبًا وطالبة تم اختيارهم وفق أسلوب العينة العشوائية العنقودية من جامعتي اليرموك والبلقاء التطبيقية، وتم استخدام طريقة نسبة الأرجحية لمانتل-هانزل للكشف عن الأداء التفاضلي للفقرات والمموهات. وقد أشارت النتائج إلى وجود (8) فقرات ذات أداء تفاضلي منها (6) فقرات لصالح الذكور، وكان حجم الأداء التفاضلي صغيرًا في (4) فقرات ومتوسطًا في (3) فقرات وكبيرًا في فقرة واحدة. كما أشارت النتائج أن هناك (11) مموهًا ذات أداء تفاضلي منها (7) مموهات لصالح الذكور، وكان حجم الأداء التفاضلي كبيرًا في مموهين ومتوسطًا في (9) مموهات.

(**الكلمات المفتاحية**: الأداء التفاضلي للفقرات، الأداء التفاضلي للمموهات، اختبار التفكير الناقد، التحيز، طريقة مانتل-هانزل)

مقدمة: يعدَ التفكير سمة من السمات التي ميّز الله بها الإنسان على سائر المخلوقات، وحثه على التفكير في كثيرٍ من الآيات. وقد وردت كلمة تفكير أو مرادفاتها (يتفكرون – يبصرون – يعقلون – يتذكرون.. إلخ) مرات عديدة في القرآن الكريم. ويعدَ التفكير نشاطًا عقليًا هادفًا يشجع على الاهتمام بحل المشكلة وتفسيرها واتخاذ القرارات المناسبة بشأنها (McWhorter & Collins, 1992). ويبدأ الفرد في التفكير عندما يوضع في موقف محيِّر وغامض؛ أي عندما لا يعرف ما الذي يجب عليه عمله بالتحديد (Diwi,1960).

وتهتم الاتجاهات الحديثة في التدريس بتنمية مهارات التفكير الناقد وإكسابها للطلبة لمواجهة التحديات التي يتعرضون إليها، والتي أخذت تفرض وجودها في نواحي الحياة وميادين المعرفة كافة بأبعادها المختلفة، لاسيما وأن العالم يشهد تغيرات وتحولات تمس الواقع الحالي بشكل مباشر؛ إذ لا بد من توسيع آفاق الطلبة المعرفية لتمكينهم من النجاح والتعامل والتكيف مع مثل هذه المتغيرات والتحولات المتسارعة (Zoabi, 2006).

ولمفهوم التفكير الناقد أبعاد عديدة، تعكس تعقد العقل البشري وتشعب عملياته؛ إذ تتم عملية التفكير الناقد من خلال سلسلة من النشاطات العقلية التي يقوم بها الدماغ عندما يتعرض لمثير يتم استقباله من إحدى الحواس الخمس المعروفة. وتتضمن مهارة التفكير الناقد البحث عن معنى، ويتطلب التوصل إليه تأملاً في مكونات الموقف أو الخبرة التي يمر بها الفرد (Paul ,1998).

Differential Item and Distractor Functioning in a Critical Thinking Test among University Students According to Gender

Kawthar Al-Abd Al-Qader and Yousef Al-Sawalmeh,
Yarmouk University, Jordan.

Abstract: This study aimed to examine gender-related differential item and distractor functioning in a critical-thinking test among university students. The test consisted of (49) four alternative multiple-choice items, covering critical thinking-skills representing: Analysis, interpretation, evaluation, explanation, reasoning, statistical and probabilistic reasoning, identification of assumptions and concentration. The responses of a cluster random sample of (930) male and female students from Yarmouk University and Al-Balqa Applied University were analyzed. Mantel-Haenszel odds ratio method was used to examine differential item and distractor functioning. The results indicated that there were (8) items with differential functioning, in favor of males in (6) items. The size of the differential functioning was small in 4 items, moderate in (3) items and large in one item. Also, there were (11) distractors with differential functioning in favor of males in (7) distractors. The size of differential functioning was large in (2) distractors and moderate in (9) distractors.

(**Keywords**: Differential Item Functioning, Differential Distractor Functioning, Critical Thinking Test, Bias, Mantel-Haenszel)

وتتأثر عملية التفكير الناقد بالسياق الاجتماعي والنظامي الذي ينتمي إليه الفرد؛ فهي تقوم على مبدأ الوعي والإدراك لديه، وذلك من خلال تفاعله مع البيئة المحيطة به (Brookfield, 1997). ويؤكد ستيرنبرغ (Sternberg, 2004) دور المعرفة في تحسين مهارة التفكير الناقد. ويرى بيترس (Petress, 2004) أن التفكير الناقد يؤدي إلى نواتج مرغوبة وهادفة، ولذلك يمكن أن نطلق عليه التفكير الموجه نحو غاية أو هدف ما.

ويعدَ التطور الفكري للطلاب من الموضوعات المشتركة بين أنظمة التعليم العالي المختلفة، وغالبًا ما يُعبَّر عن ذلك بالتفكير الناقد. والتفكير الناقد موجود عبر عدد من المواقف النظرية المختلفة المتعلقة بأغراض التعليم العالي (Erikson & Erikson, 2019). ويمكن النظر إلى التفكير الناقد على أنه يرتبط بالوظيفة التحريرية التي يجب أن يتحلى بها التعليم العالي وأنه الأداة المركزية في تمكين الطلاب (Bowell, 2017).

* جامعة اليرموك، الأردن.
© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن.

ويرى ماكفارلين (Macfarlane, 2017) أن التفكير الناقد لا يتعلق بالاستنتاجات التي يتوصل إليها الطلاب، بل يتعلق بكيفية وصولهم إليها. كما يؤكد على أن تعزيز التفكير الناقد لدى الطلاب هو تعزيز للحرية الأكاديمية لديهم. وبموازاة اعتماد سياسات التعليم العالي في جميع أنحاء العالم لنموذج مخرجات التعلم (Havnes & Proitz, 2016)، فإنه يتوقع من الطلاب الجامعيين تطوير التفكير الناقد باعتباره نتيجة مرجوة من التعليم العالي؛ أي من نتاجات التعلم المقصودة. وقد أكدت توصيات المؤتمر الدولي لتقويم التعليم في الرياض بعنوان " مهارات المستقبل: تنميتها وتقييمها" على أهمية دمج مهارات المستقبل -ومن ضمنها مهارات التفكير الناقد- في المناهج الدراسية للتعليم العام والجامعي من أجل تعليم يهيئ الجيل الجديد لوظائف المستقبل (International Conference on Education Evaluation, 2018).

ولأن التفكير الناقد يشمل عددًا من المهارات القابلة للتحسن والتدريب، فلا بدّ من بناء مقياس يحدد مقدار تقدم وتحسن هذه المهارات. وتعدّ الاختبارات الركيزة الأساسية لعملية التقويم، حيث تستخدم نتائجها في اتخاذ القرارات المتنوعة، ويؤدي الحصول على معلومات دقيقة من الاختبارات إلى اتخاذ قرارات صائبة وذات أهمية (Oudeh, 2010). وللحصول على نتائج دقيقة من الاختبارات، يجب أن تعَدّ بشكل جيد، وأن تتمتع بخصائص الصدق والثبات والموضوعية؛ كي تساعد التربويين في اتخاذ قرارات صائبة (Sawalmeh & Kawasma, 2000).

ويتطلب قياس مهارات التفكير الناقد توفر مقاييس تتمتع بخصائص سيكومترية مقبولة، لذا فقد بنيت مقاييس عديدة للتفكير الناقد. ومن الأمثلة على اختبارات التفكير الناقد واسعة الانتشار: اختبار واطسون-جليسر (Watson-Glaser Test) الذي يتضمن خمسة اختبارات فرعية هي: المسلمات، والاستنتاج، والاستقراء، والتفسير، وتقويم الحجج؛ واختبار الاستدلال المنطقي؛ واختبار الحكم: الاستدلال المنطقي وإدراك الافتراضات (Judgment: Deductive Logic & Assumption Recognition Test)؛ واختبار كري للتفكير الناقد (Curry Test of Critical Thinking؛ واختبار مهارات الاستقصاء (Test of Inquiry Skills)؛ واختبار تقييم الملاحظات (Test on Appraising Observations)؛ واختبار كاليفورنيا لمهارات التفكير الناقد (California Critical Thinking Skill Test)؛ واختبار مينسوتا للتفكير الناقد (Minnesota Test of Critical Thinking).

وفي الأردن، تم بناء اختبار للتفكير الناقد وفق نظرية الاستجابة للفقرة، تتوفر له مؤشرات إحصائية مستقرة متحررة من الأفراد (Zoabi, 2006). وقد تم تطبيقه على طلاب المرحلة الأساسية، والمرحلة الثانوية، والمرحلة الجامعية. تكون الاختبار بصورته الأولية من (228) فقرة من نوع الاختيار من متعدد، لكل منها أربعة بدائل، وتغطي فقراته مهارات التفكير الناقد المتمثلة في التحليل، والتفسير، والتقويم، والشرح، والاستدلال، والتبرير الإحصائي والاحتمالي، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها، والتركيز. وتكونت الصورة النهائية للاختبار بعد عمليات التحكيم والتجريب من (89) فقرة. وجرى تطبيق هذه الفقرات على (444) طالبًا وطالبة في المرحلة الأساسية العليا، و(610) طلاب وطالبات في المرحلة الثانوية، و(120) طالبًا وطالبة في المرحلة الجامعية. وقد أظهرت النتائج أن المتوسطات الحسابية للمجموعات المختلفة للاختبار الكلي كانت متسقة مع الأدب التربوي فيما يتعلق بالمستوى التعليمي؛ إذ إن المتوسطات الحسابية لأداء الطلبة على الاختبار كانت (38.82) في المرحلة الأساسية العليا، و(46.74) في المرحلة الثانوية، و(55.02) في المرحلة الجامعية. وتشير هذه القيم إلى العلاقة المتزايدة وتيريًا للمتوسط الحسابي للأداء على الاختبار بازدياد المستوى التعليمي. وقد تبين من الدراسة وجود فرق دال إحصائيًا بين متوسط الذكور (56.69) ومتوسط الإناث (47.88) على المستوى الجامعي. وهذا الفرق على المستوى الجامعي لم يتسق مع الفروق التي أظهرتها النتائج في المرحلتين الأساسية العليا والثانوية، كما أنه يتعارض مع نتائج العديد من الدراسات التي أكدت عدم وجود فرق دال إحصائيًا في التفكير الناقد بين الجنسين على المستوى الجامعي في الأردن وفلسطين وغيرهما من دول العالم.

واستخدم الكركي ومحادين (Al-Karaki & Mahadeen, 2019) اختبار كاليفورنيا للتفكير الناقد في دراستهما التي أجريت على عينة عشوائية طبقية تكونت من (834) طالبًا وطالبة من جامعة مؤتة. وأشارت نتائج الدراسة إلى عدم وجود فرق دال إحصائيًا بين متوسط الذكور (14.0) ومتوسط الإناث (13.4). وفي فلسطين، طبق سرحان (Serhan, 2000) اختبار كاليفورنيا للتفكير الناقد على عينة من طلبة الجامعات الفلسطينية تكونت من (199) طالبًا وطالبة، وأشارت نتائج دراسته إلى عدم وجود فروق دالة إحصائيًا في المتوسطات الحسابية للتفكير الناقد ومهاراته الفرعية بين الذكور والإناث. وقد أجريت دراسة في إيران (Salahshoor & Raflee, 2016) بهدف معرفة أثر الجنس في التفكير الناقد. وتكونت عينة الدراسة من (186) طالبًا وطالبة من خمسة معاهد لتعليم اللغة الإنجليزية كلغة أجنبية. وقد أظهرت نتائج الدراسة عدم وجود فرق دال إحصائيًا بين المتوسط الحسابي لأداء الذكور (40.94) والمتوسط الحسابي لأداء الإناث (43.20) على اختبار واطسون-جليسر. وفي ماليزيا، استخدم مقياس سلوك التفكير الناقد (Ghadia et al., 2013) مع (951) طالبًا وطالبة من جامعة بوتر، وأشارت النتائج إلى عدم وجود فرق دال إحصائيًا بين متوسط الذكور (2.34) ومتوسط الإناث (2.26). وفي الولايات المتحدة الأمريكية، طبق ليش (Leach, 2011) اختبار كاليفورنيا لمهارات التفكير الناقد على عينة من (1455) طالبًا وطالبة في إحدى الجامعات في الجنوب الشرقي، وقد بينت النتائج عدم وجود فرق دال إحصائيًا بين متوسط الذكور ومتوسط الإناث.

وقد بدأت دراسة التحيّز في الفقرات والاختبارات في نهاية الستينات من القرن الماضي، وتسارعت بعد ذلك لاعتبارات اجتماعية ونفسية وتربوية. ويتزايد حاليًا الاهتمام بقضايا الإنصاف والعدالة في التقييم ضمن تصور أوسع لأدلة الصدق التي تلزم لتبرير استخدام درجات الاختبار لغرض معين (Sireci, 2016). وتتضمن النماذج المقترحة لتقييم جودة الاختبارات الأداء التفاضلي للفقرات كتصميم بحثي ممكن لجمع الأدلة حول صدق البناء (Evers et al., 2013). ويعني ذلك أن الاختلاف في أي معلم من معالم الفقرة بين مجموعات المفحوصين قد يهدد صدق درجات الاختبار إذا تطلب الاختلاف تفسيرات مختلفة في المجموعات المختلفة (Gomez-Benito et al., 2018). وقد تعمل الفقرات ذات الأداء التفاضلي باحتمال مرتفع على تهديد صدق درجات الاختبار إذا لم يتم التدقيق في تلك الفقرات ومراجعتها (Reeves & Marbach, 2016). كما يفيد تحليل (Differential Item Functioning ;DIF) في تحديد فجوات الأداء التي لا يتم كشفها بين المجموعات المختلفة من خلال مقارنة الدرجات الكلية لتلك المجموعات (Zeiky, 2003). ويحدث DIF عندما يكون هناك فرق بين المجموعات في السمات الثانوية التي تقيسها الفقرة، سواء بصورة منفردة أو بالإضافةِ إلى السمة الرئيسة المستهدفة (Martinkova et al., 2017). وعليه، يجب إجراء تحليل DIF بصورة روتينية عند تطوير الاختبارات، لأن تفسير نتائج DIF ضروري لتحسين الاختبار ولتحسين صدق تفسير درجاته.

وبالرغم من ظهور مصطلح الأداء التفاضلي للفقرة في الأدب التربوي بعد مصطلح تحيز الفقرة، فإنه ينظر لتحليلات DIF باعتبارها خطوة أولية في تقييم تحيز الفقرة. ويُعد وجود DIF مؤشرًا على أن الفقرة قد تكون متحيزة، لكنه لا يضمن التحيز. وهو يشير إلى وجود سمة أخرى بالإضافة للسمة الرئيسة موضع الاهتمام. وهذا يعني أنه يمكن أن تكون الفقرة ذات أداء تفاضلي، لكنها غير متحيزة. وعليه، فإنه لتحديد الفقرات المتحيزة، تتم دراسة الأداء التفاضلي أولاً، وبعد تحديد الفقرات ذات الأداء التفاضلي يتم فحص محتوى الفقرة ذات الأداء التفاضلي وشكلها ومجموعة الاستجابات لها وسياقها بحثًا عن سبب الأداء التفاضلي. فإذا تبين أن السبب ليست له صلة بالسمة المقاسة، كانت الفقرة متحيزة. وتتحقق العدالة عندما ترتبط السمة الثانوية بالسمة الرئيسة.

وقد ظهر مصطلح الأداء التفاضلي في مطلع الثمانينات من القرن الماضي كمصطلح مرادف لمصطلح تحيز الفقرة في بادئ الأمر (Camilli & Shepard, 1994). إلا أنه كان هناك اهتمام كبير في التمييز بين المصطلحين؛ لكونهما غير مترادفين على أرض الواقع. وقد جاء مصطلح الأداء التفاضلي للتأكيد على الطبيعة الإحصائية للأداء التفاضلي؛ فهو يشير إلى خاصية إحصائية للفقرة، توضح درجة إمكانية قياس الفقرة لسمات مختلفة في المجموعات المختلفة. وباختصار، تظهر الفقرة أداءً تفاضليًا عندما تختلف المجموعات المختلفة من الأفراد في احتمال إجابة الفقرة بعد ضبط الأداء الكلي على الاختبار، فيما يتضمن تحيز الفقرة جمع النتائج الإحصائية مع التفسيرات المادية المتعلقة بالسمة مدار التمثيل وأسباب الأداء التفاضلي التي لا تتعلق بتلك السمة؛ ذلك أن التحيز يحمل معنى عدم العدالة وعدم المساواة؛ بمعنى أنه ذو صبغة اجتماعية.

والتمييز بين الأداء التفاضلي والتحيز يزداد صعوبة لصعوبة أساليب DIF الإحصائية وتعقدها (Gomez-Benito et al., 2018). والهدف النهائي لدراسات الأداء التفاضلي للفقرة هو تحديد الجوانب غير المتعلقة بالسمة المستهدفة سواءً في محتوى الفقرة أو شكلها أو معايير التصحيح التي قد تؤثر تفاضليًا في درجات مجموعة واحدة أو أكثر من بين المتقدمين للاختبار. وعندما يتم الكشف عن الأداء التفاضلي، يحاول مطورو الاختبار أو مستخدموه تحديد التفسيرات المعقولة للاختلافات، وبعدها يمكن تبديل الفقرات ذات الأداء التفاضلي أو مراجعتها وتعديلها للحصول على تفسيرات سليمة لدرجات جميع المفحوصين بغض النظر عن المجموعات التي ينتمون لها.

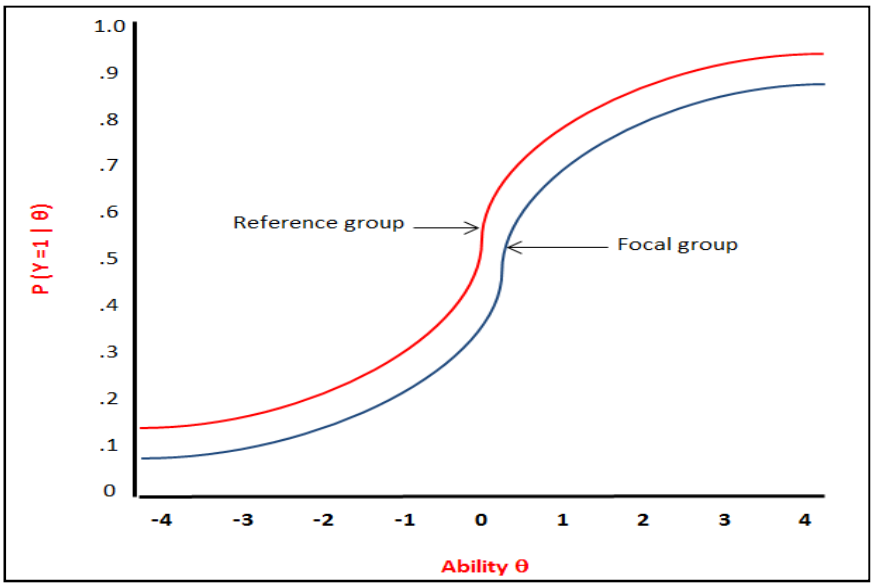
وتجدر الإشارة إلى أن الاختلاف بين المجموعات في الأداء على الفقرة الذي يمثل فرقًا في السمة المقاسة يشار إليه عادة بمصطلح التأثير (Impact)، وهو يمثل فرقًا مرتبطًا بالسمة (Camilli & Shepard, 1994). وتجنب الخلط بين التأثير والأداء التفاضلي كان ولا يزال محل اهتمام في البحث عن الفقرة المتحيزة. ويسعى تحليل DIF إلى التأشير على فقرات من المحتمل أن تكون متحيزة. وهكذا، فإن تمييز DIF عن التأثير، وتحديد ما إذا كانت الفقرات ذات الأداء التفاضلي تقيس السمة المستهدفة، هي قضايا أساسية في الصدق وفي السعي لتحقيق عدالة الاختبار (Gomez-Benito et al., 2018).

ويتم عادة تقييم نوع الأداء التفاضلي من خلال التفاعل بين مستوى قدرة الفرد والمجموعة التي ينتمي لها. وهناك شكلان للأداء التفاضلي للفقرات هما (Camilli & Shpard, 1994):

1. الأداء التفاضلي المنتظم: ينتج الأداء التفاضلي المنتظم عندما يكون احتمال الاستجابة الصحيحة عن الفقرة لإحدى المجموعتين الفرعيتين (المرجعية مثلاً) دائمًا أكبر عند جميع مستويات القدرة θ. ويتضح ذلك ضمن سياق نظرية الاستجابة للفقرة عندما يكون لمنحنى خاصية الفقرة الميل نفسه في المجموعتين (المرجعية والمستهدفة)، لكنهما تختلفان في معلمة الصعوبة كما يظهر في الشكل (1).

الشكل 1

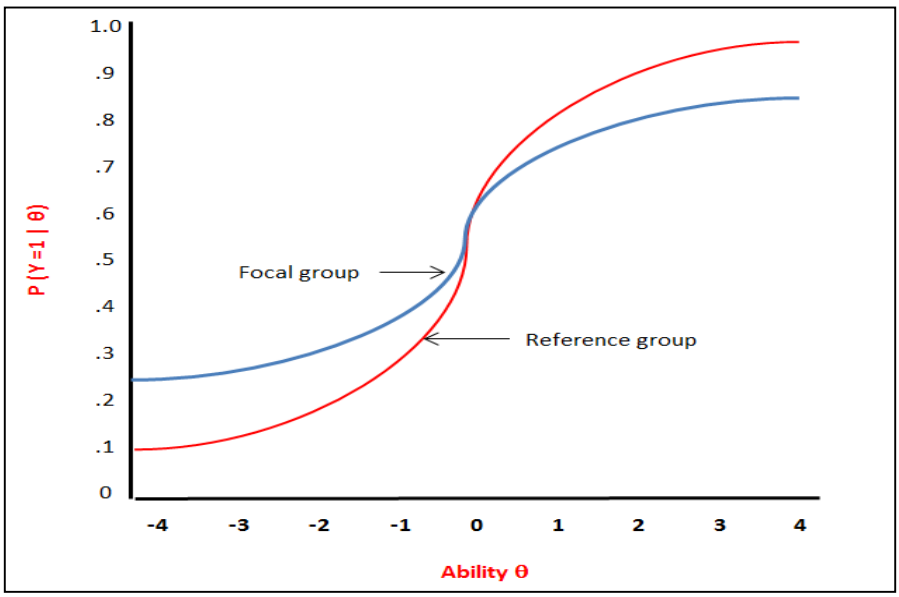
الأداء التفاضلي المنتظم للفقرة.



2. الأداء التفاضلي غير المنتظم: وينتج عند حدوث تفاعل بين المجموعة ومستوى القدرة θ؛ أي يختلف الفرق في احتمال الإجابة الصحيحة للفقرة بين المجموعتين باختلاف مستوى القدرة. ويكون الفرق لصالح إحدى المجموعتين (المستهدفة مثلاً) على النصف الأيسر لمتصل القدرة θ ولصالح المجموعة الأخرى (المرجعية مثلاً) على النصف الأيمن لمتصل القدرة θ، كما يظهر في الشكل (2).

الشكل 2

الأداء التفاضلي غير المنتظم للفقرة.



ويفيد الكشف عن DIF في تنقية الاختبار من التحيز، ويحقق ذلك مبدأ العدل والمساواة بين المجموعات المختلفة، وينعكس إيجابًا على صدق الاختبار. وتوجد عدة طرق إحصائية للكشف عن الأداء التفاضلي للفقرة (DIF) (Penfield, 2008)، والطرق الأكثر استخدامًا هي: الانحدار اللوجستي (Logistic Regression: LR)، ونظرية الاستجابة للفقرة (Item Response Theory: IRT)، وطريقة مانتل-هانزل (Mantel-Haenszel: MH)، ونسبة الأرجحية (Odds Ratio: OR)،

ونسبة الأرجحية لمانتل-هانزل (Mantel-Haenszel Log-Odds Ratio: MH-LOR).

وتحتاج طريقتا الانحدار اللوجستي (LR) ونظرية الاستجابة للفقرة (IRT) إلى عينات كبيرة، ووقت أطول، وتكلفة أكثر. وتعد طريقة نسبة الأرجحية لمانتل-هانزل (MH-LOR) أقوى الطرق للكشف عن DIF (Haladyna, 2004)، وتمتاز بأنها لا تحتاج إلى عينات كبيرة، وهي أكثر كفاءة من الطرق الأخرى؛ إذ توفر معلومات كافية وذات معنى عن الأداء التفاضلي للفقرة؛ لذا فقد اختيرت للاستخدام في الدراسة الحالية.

وتستخدم طريقة نسبة الأرجحية لمانتل-هانزل (MH-LOR) لتقييم DIF في الفقرات ثنائية التصحيح، وتعتمد على جدول التوافق (Contingency Table)، حيث يتم تصنيف الأفراد لكل فقرة وعند كل علامة على الاختبار في مصفوفة ثنائية (2×2). ويبين الجدول (1) توزيع الإجابات الصحيحة والإجابات الخطأ للمجموعتين المرجعية والمستهدفة عند مستوى القدرة s.

الجدول 1

توزيع الإجابات الصحيحة والإجابات الخطأ عند مستوى القدرة (s).

المجموعة	الإجابة الصحيحة (1)	الإجابة الخطأ (0)	المجموع
المرجعية r	n_{r1s}	n_{r0s}	n_{rs}
المستهدفة f	n_{f1s}	n_{f0s}	
المجموع	n_{1s}	n_{0s}	n_s

وتمثل كل خلية في الجدول عدد المفحوصين في كل مجموعة. فالعدد n_s هو عدد المفحوصين الذين مستوى القدرة لهم s، والعدد n_{rs} هو عدد أفراد المجموعة المرجعية في المستوى s، والعدد n_{fs} هو عدد أفراد المجموعة المستهدفة في المستوى s، والعدد n_{1s} هو عدد الذين أجابوا الفقرة إجابة صحيحة في المستوى s، والعدد n_{0s} هو عدد الذين أجابوا الفقرة إجابة خاطئة في المستوى s، والعدد n_{r1s} هو عدد أفراد المجموعة المرجعية الذين أجابوا الفقرة إجابة صحيحة في المستوى s، والعدد n_{r0s} هو عدد أفراد المجموعة المرجعية الذين أجابوا الفقرة إجابة خاطئة في المستوى s، والعدد n_{f1s} هو عدد أفراد المجموعة المستهدفة الذين أجابوا الفقرة إجابة صحيحة في المستوى s، والعدد n_{f0s} هو عدد أفراد المجموعة المستهدفة الذين أجابوا الفقرة إجابة خاطئة في المستوى s. ويفيد هذا الجدول في حساب الأرجحية (Odds) للإجابة الصحيحة على الفقرة لكل من المجموعة المرجعية والمجموعة المستهدفة عند كل مستوى من مستويات القدرة (s) من خلال المعادلتين الآتيتين:

$$O_{rs} = \frac{n_{r1}/n_r}{n_{r0}/n_r} = \frac{n_{r1}}{n_{r0}} \dots\dots\dots\dots\dots\dots (1)$$

$$O_{fs} = \frac{n_{f1}/n_f}{n_{f0}/n_f} = \frac{n_{f1}}{n_{f0}} \dots\dots\dots\dots\dots\dots (2)$$

وبناء على ذلك، تحسب نسبة أرجحية الإجابة الصحيحة لأفراد المجموعة المرجعية إلى أفراد المجموعة المستهدفة عند المستوى (s) باستخدام المعادلة التالية:

$$OR_s = \frac{n_{r1}/n_{r0}}{n_{f1}/n_{f0}} = \frac{n_{r1}\, n_{f0}}{n_{r0}\, n_{f1}} \dots\dots\dots\dots\dots\dots (3)$$

وتحسب قيمة مقدر نسبة الأرجحية (α_{MH}) من خلال المعادلة (4) (Penfield, 2008):

$$\alpha_{MH} = \frac{\sum_s \left[n_{r1s} n_{f0s}/n_s\right]}{\sum_s \left[n_{r0s} n_{f1s}/n_s\right]} \dots\dots\dots\dots\dots\dots (4)$$

فإذا كانت قيمة المقدر تساوي (1)، $\alpha_{MH} = 1$، فهذا يدل على عدم وجود أداء تفاضلي للفقرة، أما إذا كانت قيمة المقدر أقل من واحد صحيح، $\alpha_{MH} < 1$، فهذا يدل على وجود أداء تفاضلي لصالح المجموعة المستهدفة، وأما إذا كانت قيمة المقدر أكبر من الواحد، $\alpha_{MH} > 1$، فهذا يدل على وجود أداء تفاضلي لصالح المجموعة المرجعية. ولاختبار الفرضية الصفرية (H_0: $\alpha_{MH} = 1$) يستخدم اختبار Z:

$$Z = \frac{\sum_s [n_{r1s} - E(n_{r1s})] - 0.5}{\sqrt{Var(n_{r1s})}} \dots\dots\dots\dots\dots\dots (5)$$

حيث

$$E(n_{r1s}) = \frac{(n_{rs})(n_{1s})}{n_s} \dots\dots\dots\dots\dots\dots (6)$$

$$Var(n_{r1s}) = \frac{(n_{rs})(n_{1s})(n_{fs})(n_{0s})}{(n_s^2)(n_s^2 - 1)} \dots\dots\dots\dots (7)$$

ويمكن تحويل قيمة مقدر نسبة الأرجحية α_{MH} إلى اللوغارتم الطبيعي من أجل توضيح تفسير DIF وجعله أكثر فائدة، بحيث يصبح لديها قيمة على التدريج من $(-\infty$ الى $+\infty)$، علمًا بأن الصفر يدل على عدم وجود أداء تفاضلي للفقرة (Camilli & Shepard, 1994).

$$\lambda_{MH} = Log\,(\hat{\alpha}_{MH}) \dots\dots\dots\dots\dots\dots (8)$$

وتدلل λ_{MH} على اتجاه وحجم تأثير الأداء التفاضلي؛ فالإشارة الموجبة تدل على أن DIF لصالح المجموعة المرجعية، والإشارة السالبة تدل على أن الفرق لصالح المجموعة المستهدفة. أما حجم الأثر فيتم تحديده وفقًا للمعايير الآتية:

1) $|\lambda_{MH}| = 0$، فإنه لا يوجد DIF.

2) $|\lambda_{MH}| < 0.43$، يكون تأثير DIF صغيرًا.

3) $0.64 > |\lambda_{MH}| \geq 0.43$، يكون تأثير DIF متوسطًا.

4) $|\lambda_{MH}| \geq 0.64$، يكون تأثير DIF كبيرًا (Penfield & Camilli, 2007).

يطلق الأداء التفاضلي للمموهات (Differential Distractor Functioning; DDF) على البدائل غير الصحيحة في فقرات الاختيار من متعدد، التي تعرف بمصطلح مموهات أو مشتتات (Distractors). وتعتمد جودة فقرة الاختيار من متعدد على جودة مموهاتها؛ فالمفهوم التقليدي للقوة التمييزية للفقرة هو المعكوس الجمعي لمجموع قيم التمييز لمموهاتها. وهذا يدلل على تأثير خصائص مموهات الفقرة بشكل كبير في صدقها وثباتها؛ فقد تكون الفقرة ضعيفة بسبب عدم فعالية مموهاتها. فالمموه الجيد هو المموه الجذاب، وبالذات للأفراد الذين لا يعرفون الاختيار الصحيح، ويعني ذلك أنه يميز باتجاه معاكس لتمييز الفقرة، بحيث يكون عدد المفحوصين الذين يختارونه من الفئة العليا أقل من عددهم من الفئة الدنيا (Jalabi, 2005). وهذا يعني أنه يمكن السيطرة على خصائص فقرات الاختبار السيكومترية والتحكم فيها من خلال زيادة الاهتمام ببدائل الإجابة لتلك الفقرات (Oudeh, 2010).

ويختلف الأداء التفاضلي للمموهات DDF عن الأداء التفاضلي للفقرات DIF في اختبارات الاختيار من متعدد. فالفقرة قد لا تكون ذات أداء تفاضلي بينما تتضمن مموهات ذات أداء تفاضلي، وقد تكون الفقرة ذات أداء تفاضلي وفي الوقت نفسه لا تتضمن مموهات ذات أداء تفاضلي. ومن المحتمل أن تسبب بعض المموهات ذات الأداء التفاضلي أداء تفاضليًا للفقرة (Anamezie & Nnadi, 2018). ويستخدم DDF كأسلوب لتحسين جودة المقياس من خلال فهم الاستجابات المتحيزة عبر المجموعات، ويمكن له أن يلقي الضوء على المصادر المحتملة للتباين غير المرتبط بالسمة المستهدفة (Tsaousis et al., 2018)، ويوفر مصدرًا إضافيًا مهمًا للمعلومات حول أداء الطلبة تزداد أهميته مع زيادة الاهتمام بتعديل فقرات الاختبار اعتمادًا على استراتيجية حذف بعض المموهات (Middleton & Laitusis, 2007)؛ بمعنى أنه يساعد بصورة كبيرة في شرح الأداء التفاضلي للفقرات. وذكر بنفليد (Penfield, 2008) أن هناك طرقًا متعددة لتحديد الأداء التفاضلي للمموهات (DDF)، منها الانحدار اللوجستي (LR)، ونظرية الاستجابة للفقرة (IRT)، ونسبة الأرجحية لمانتل-هانزل (MH-LOR). وتم استخدام نسبة الأرجحية لمانتل-هانزل في الدراسة الحالية. ولحساب DDF، تمت مقارنة نسبة اختيار الإجابة الصحيحة بوجود الاختيار بين البديل الصحيح وكل واحد من المموهات، وتم حساب قيمة MH-LOR لكل مموه (J) ويرمز لها بالرمز (λ_{MHJ})، ويمكن من خلالها أيضًا تحديد حجم تأثير DDF، واتجاهه. فالإشارة الموجبة تدل على أن المموه أقل جذبًا للمجموعة المرجعية، وتدل الإشارة السالبة على أن المموه أقل جذبًا للمجموعة المستهدفة. ويتم تحديد حجم التأثير باستخدام المعايير نفسها التي سبقت الإشارة إليها. ويتم اشتقاق (λ_{MHJ}) عند كل مستوى s لكل مموه (J) من خلال المعادلة التالية:

$$\lambda_{MHJ} = Log(\alpha_{MHJ}) \dots\dots\dots\dots\dots\dots\dots (9)$$

وبالاعتماد على جدول التوافق (2X2) المبين في الجدول (2) يمكن حساب قيمة α_{MHJ} من خلال المعادلة التالية:

$$\alpha_{MHJ} = \frac{\sum_s \frac{n_{r1s} n_{fjs}}{n_s}}{\sum_s \frac{n_{rjs} n_{f1s}}{n_s}} \dots\dots\dots\dots\dots\dots (10)$$

الجدول 2

توزيع الإجابة الصحيحة / المموه (J) عند المستوى (s).

المجموعة	الإجابة الصحيحة (1)	اختيار البديل j	العدد الكلي
المرجعية r	n_{r1s}	n_{rjs}	n_{rs}
المستهدفة f	n_{f1s}	n_{fjs}	n_{fs}
المجموع	n_{1s}	n_{js}	n_s

وتمثل كل خلية في الجدول (2) عدد المفحوصين في كل مجموعة كما كانت في الجدول (1) باستثناء الأعداد n_{fjs}، n_{rjs} حيث هي عدد أفراد المجموعتين (المستهدفة والمرجعية) الذين اختاروا البديل j.

ويعمل الأداء التفاضلي للمموهات (DDF) كإطار لفهم الأداء التفاضلي للفقرات (DIF) بصورةٍ أفضل، حيث يساعد الأداء التفاضلي للمموهات (DDF) في توفير معلومات عن بدائل الاستجابة المسؤولة عن أثر الأداء التفاضلي للفقرة (DIF)،

ويساعد الأداء التفاضلي للمموهات (DDF) أيضًا في تحديد سبب الأداء التفاضلي للفقرات. فمثلاً يدلل ثبات الأثر الناتج عن الأداء التفاضلي للمموهات عبر جميع المموهات على وجود فروق منتظمة بين المجموعتين في احتمالية اختيار أي مموه من المموهات، مما يدل على أن سبب وجود الأداء التفاضلي للفقرة هو وجود تحيز في الإجابة الصحيحة لتلك الفقرة. وهذا يرشد إلى ضرورة الكشف عن مصدر هذا التحيز من خلال مراجعة محتوى الإجابة الصحيحة. وعندما يكون أثر الأداء التفاضلي الناتج لأحد المموهات كبيرًا بالمقارنة مع الأثر الناتج لبقية المموهات، فيدل هذا على أن ذلك المموه أكثر جاذبية لإحدى المجموعات دون الأخرى، وهذا يعني أن سبب وجود الأداء التفاضلي للفقرة يتمثل في إحدى خصائص هذا المموه، أو يكون بسبب وجود تفاعل بين خصائص المموه وبين محتوى الفقرة، وهذا يتطلب إجراء تحليل لخصائص المموه وتفاعل هذه الخصائص مع محتوى الفقرة نفسها (Penfield, 2008). لذلك قام المتخصصون في مجال بناء الاختبارات النفسية باستخدام مفهومي الأداء التفاضلي للفقرة، والأداء التفاضلي للمموهات، للتعبير عن الطرق والمعاني الإحصائية المستخدمة للكشف عن تحيز فقرات الاختبار (Ellis & Raju, 2003).

وقد اهتم عدد من الباحثين بالكشف عن الأداء التفاضلي للاختبار وفقراته (Autawi, 2004; Innabi & Dodeen, 2006; Karakaya, 2012; Mubarak, 2010; Ubaydat, 2014). وهناك دراسات اهتمت بالكشف عن الأداء التفاضلي لمموهات الاختبار، سواء من حيث دراستها بحدّ ذاتها، أو استخدامها للكشف عن الفقرات ذات الأداء التفاضلي للفقرات، أو تحديد أثر الأداء التفاضلي للمموهات على الأداء التفاضلي للفقرات (Ajlouni, 2016; Banks, 2004; Koon, 2010; Mahmoud, 2010). أما دراسة جرينبرغ (Greenberg, 2010) فقد اهتمت بالكشف عن العلاقة بين DD و DIF، والكشف عنهما في فقرات الاختبار من نوع الاختيار من متعدد، وأظهرت النتائج أن حجم وشكل DIF يعتمد على تأثير DDF. وهناك دراسات تناولت الأداء التفاضلي للفقرات في اختبارات التفكير الناقد (Jaradat, 2003; Rashwan, 2015; Sahin et al., 2015).

يتبين من الدراسات السابقة أن هناك دراستين عربيتين فقط تناولتا الأداء التفاضلي لمموهات الفقرة، سواء من حيث دراستها بحد ذاتها، أو استخدامها للكشف عن الفقرات ذات الأداء التفاضلي للفقرات، أو تحديد أثر الأداء التفاضلي للمموهات على الأداء التفاضلي للفقرات. أما بالنسبة للدراسات الأجنبية، فكانت الدراسات التي تطرقت إلى الأداء التفاضلي للمموهات قليلة مقارنة بالدراسات التي تناولت الأداء التفاضلي للفقرات. ولم يتم العثور على أي دراسة عربية تطرقت للأداء التفاضلي للفقرات والمموهات في اختبار للتفكير الناقد. لذلك، تناولت الدراسة الحالية الأداء التفاضلي للجنس على المستوى الجامعي في فقرات ومموهات اختبار التفكير الناقد المعد في البيئة الأردنية (Zoabi, 2006). وتم تناول هذا الاختبار لأنه يتميز بأنه اختبار شامل لجميع مهارات التفكير الناقد، وأنه يناسب البيئة الأردنية بشكل خاص والبيئة العربية بشكل عام، ولأنه بني وفق النظرية الحديثة في القياس التي تراعي مبادئ القياس الموضوعي. كما أنه يتمتع بخصائص سيكومترية مقبولة. ويؤخذ عليه أنه طويل نسبيًا مقارنة باختبار كاليفورنيا ويحتاج تطبيقه إلى وقت طويل مما يحول دون استخدامه. كما قد تكون عدالته على المحك ومحل تشكيك في المستوى الجامعي نظرًا لوجود فرق دال إحصائيًا بين متوسط الذكور ومتوسط الإناث. وهذا الفرق بحاجة إلى تفسير ومعرفة ما إذا كان فرقًا حقيقيًا أم يعود إلى وجود أداء تفاضلي في فقرات ومموهات الاختبار. وعليه، فقد استهدف هذا الاختبار في الدراسة الحالية لتنقيته من الفقرات ذات الأداء التفاضلي للجنس على المستوى الجامعي إن وجدت، وبذلك نتيقن من عدالته وفي الوقت نفسه يتم تقصيره مما يزيد من فرص استخدامه في قياس التفكير الناقد لدى الطلبة في الجامعات الأردنية والعربية.

مشكلة الدراسة وأسئلتها

بالتناغم مع اهتمام النظام التعليمي الأردني بالتفكير الناقد، فقد تم بناء اختبار للتفكير الناقد يناسب المجتمع الأردني في مراحل التعليم المختلفة بدءًا من المرحلة الأساسية العليا وفق نظرية الاستجابة للفقرة (Zoabi, 2006). وقد أظهرت نتائج هذا الاختبار وجود فرق دالٍ إحصائيًا على المستوى الجامعي بين متوسط الذكور ومتوسط الإناث لصالح الذكور. وكان من الصعب إعطاء تفسيرات معقولة لهذا الفرق، خاصة في ظل عدم وجود فرق في الأداء على الاختبار نفسه في المراحل السابقة للمرحلة الجامعية، وفي ظل نتائج العديد من الدراسات المحلية والإقليمية والدولية التي أكدت عدم وجود فرق دال إحصائيًا في المتوسط الحسابي للتفكير الناقد لدى الطلبة الجامعيين بين الذكور والإناث. وهذا يتطلب البحث عن أدلة صدق إضافية لهذا الاختبار تساعد في تفسير هذا الفرق ومعرفة ما إذا كان يعزى لعوامل لها علاقة بمهارات التفكير الناقد أم لعوامل ليست لها علاقة بذلك. والسبيل لذلك هو فحص الأداء التفاضلي لفقرات ومموهات الاختبار انسجامًا مع الدعوات لأن تكون دراسة الأداء التفاضلي للفقرات والمموهات من الإجراءات الروتينية في سياق عملية تطوير الاختبار، وللتحقق من خاصية استقرار القياس؛ إذ يفيد ذلك في تنقية الاختبار من بعض مهددات الصدق. فإن تبين عدم وجود فقرات ومموهات ذات أداء تفاضلي للجنس في الاختبار، فإن ذلك يعد دليلاً على أن الفرق بين الذكور والإناث هو فرق حقيقي. وإن تبين وجود فقرات ومموهات ذات أداء تفاضلي للجنس في الاختبار، فإن ذلك مؤشر على أنها تقيس سمات ثانوية زيادة على السمة الرئيسة المتمثلة في التفكير الناقد، ويعد ذلك مهددًا لصدق الاختبار ويستدعي المعالجة. لذا تولدت الحاجة إلى فحص الأداء التفاضلي للفقرات

والمموهات في هذا الاختبار للتفكير الناقد على المستوى الجامعي. وبالتحديد حاولت الدراسة الحالية الإجابة عن الأسئلة الاتية:

1- هل تظهر **فقرات** اختبار التفكير الناقد أداءً تفاضليًا لمتغير الجنس على المستوى الجامعي باستخدام طريقة نسبة الأرجحية لمانتل-هانزل؟

2- هل تظهر **مموهات** فقرات الاختيار من متعدد في اختبار التفكير الناقد أداءً تفاضليًا لمتغير الجنس على المستوى الجامعي باستخدام طريقة نسبة الأرجحية لمانتل-هانزل؟

3- هل توجد علاقة بين الأداء التفاضلي للمموهات والأداء التفاضلي للفقرات في اختبار التفكير الناقد؟

هدف الدراسة

تهدف الدراسة الحالية إلى فحص الأداء التفاضلي لفقرات ومموهات اختبار مهارات التفكير الناقد على المستوى الجامعي وفقًا لمتغير الجنس، واستقصاء العلاقة بين الأداء التفاضلي للمموهات والأداء التفاضلي للفقرات، والاستفادة من ذلك في تحسين اختبار التفكير الناقد على المستوى الجامعي وتنقيته من مهددات الصدق، إن وجدت.

أهمية الدراسة

تبرز أهمية الدراسة الحالية من أهمية التفكير الناقد وأهمية توفير اختبار لقياسه تتوفر له درجة عالية من الصدق تتصل بأدلة الصدق من مصادره المختلفة كالمحتوى، وعملية الاستجابة، والبناء الداخلي، والعلاقة مع متغيرات أخرى، وتبعات استخدام الدرجات. ويتحقق ذلك لاختبار التفكير الناقد المستهدف في الدراسة من خلال الكشف عن الأداء التفاضلي للفقرات (Gomez-Benito et al., 2018) وفهمها بعمق من خلال الأداء التفاضلي للمموهات. وستساهم عملية معالجة الفقرات ذات الأداء التفاضلي في تحقيق درجة أعلى من العدالة بين طلبة الجامعة. ومن ناحية تطبيقية، فإن الدراسة الحالية ستوفر للباحثين والجامعات اختبارًا لمهارات التفكير الناقد يتمتع بدرجة عالية من الصدق واستقرار القياس. وستسلط الضوء على العلاقة بين DIF وDDF في ظل بيانات فعلية للمساعدة في فهم DIF وتفسيره.

مصطلحات الدراسة

- **التفكير الناقد**: أسلوب في التفكير المنظم، والمركز، والمبرر، بغرض حل المسألة وصياغة الاستدلالات وحساب الاحتمالات وعمل القرارات، ويقاس بالاختبار المستخدم في الدراسة الحالية والمتمثل في مهارات التحليل والتفسير والتقويم والشرح والاستدلال والتعرف إلى الافتراضات والتركيز.
- **الأداء التفاضلي للفقرة (DIF):** هو مؤشر إحصائي للتعبير عن الفروق في الاستجابة على الفقرة بين مجموعتين من المفحوصين ممن لهم المستوى نفسه من القدرة.
- **الأداء التفاضلي للمموهات (DDF):** هو مؤشر إحصائي للتعبير عن الفروق في الاستجابة على المموه بين مجموعتين من المفحوصين ممن لهم المستوى نفسه من القدرة.

محددات الدراسة

اقتصرت الدراسة الحالية على الكشف عن الأداء التفاضلي للفقرات والأداء التفاضلي للمموهات في اختبار مهارات التفكير الناقد وفقًا لمتغير الجنس على المستوى الجامعي في عينة من طلبة جامعتي اليرموك والبلقاء التطبيقية. كما اقتصرت الدراسة الحالية على استخدام طريقة نسبة الأرجحية لمانتل- هانزل MH-LOR في الكشف عن الأداء التفاضلي للفقرات (DIF) والمموهات (DDF).

الطريقة

مجتمع الدراسة

تكون مجتمع الدراسة من طلبة جامعتي اليرموك والبلقاء التطبيقية (كلية إربد الجامعية) للعام الدراسي (2018/2017)، حيث بلغ عدد الطلبة في جامعة اليرموك (26465) طالبًا وطالبة، وعدد الطلبة في كلية إربد الجامعية (6200) طالبٍ وطالبة.

عينة الدراسة

تم اختيار (500) طالب وطالبة وفق أسلوب العينة العشوائية العنقودية من كليات الآداب والعلوم والتربية والطب في جامعة اليرموك، وتم اختيار (500) طالب وطالبة وفق أسلوب العينة العشوائية العنقودية من أقسام العلوم السياسية والعلوم التربوية والعلوم الإدارية واللغة العربية والمهن الطبية المساعدة في كلية إربد الجامعية، وكانت الشعبة هي وحدة الاختيار. وقد تم استبعاد الإجابات النمطية وغير المكتملة، وتكونت العينة في صورتها النهائية من (930) طالبًا وطالبة (526 طالبة، و404 طلاب).

أداة الدراسة

تكونت أداة الدراسة من اختبار التفكير الناقد الذي بنته الزعبي (Zoabi, 2006) للبيئة الأردنية، وقد تحققت من الاستقلال المحلي له وافتراض أحادية البعد والثبات والصدق من خلال عدد من المؤشرات المتمثلة في صدق المحتوى وصدق البناء والصدق المرتبط بالمحك. وقد بلغ معامل الثبات للاختبار الكلي (0.91) حسب معادلة KR20، وتمتع الاختبار بالقدرة على التمييز بين المستويات التعليمية وبين المستويات التحصيلية، وكان ارتباطه مع اختبار كاليفورنيا لمهارات التفكير الناقد (0.72). ولأغراض الدراسة الحالية، تم استخدام (49) فقرة من نوع الاختيار من متعدد، لكل منها أربع بدائل. وهي مجموعة الفقرات التي تطابقت مع نموذج راش أحادي المعلمة، وهي أيضًا من الفقرات التي طابقت النموذج ثنائي المعلمة ومن الفقرات التي طابقت النموذج ثلاثي المعلمة. وتغطي هذه الفقرات مهارات التفكير الناقد المتمثلة في التحليل، والتفسير، والتقويم، والشرح، والاستدلال، والتبرير الإحصائي والاحتمالي، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها، والتركيز، كما في الجدول (3).

الجدول 3

توزيع الفقرات على مهارات التفكير الناقد .

المهارة	العدد	الفقرات
التحليل	4	1,2,3,4
التفسير	6	5,6,7,8,9,10
التقويم	7	11,12,13,14,15,16,17
الشرح	12	18,19,20,21,22,23,24,25,26,27,28,29
الاستدلال	12	30,31,32,33,34,35,36,37,38,39,40,41
التبرير الاحتمالي والإحصائي	4	42,43,44,45
التعرف إلى الافتراضات	3	46,47,48
التركيز	1	49

يلاحظ من الجدول (3) أن مهارتي الشرح والاستدلال استحوذتا على العدد الأكبر من الفقرات بواقع (12) فقرة لكل منهما، بينما استحوذت مهارة التركيز على فقرة واحدة فقط.

وقد تم التحقق من صدق المحتوى لهذا الاختبار من خلال عرض فقراته على مجموعة من المحكمين من ذوي الخبرة والاختصاص. وقد كان هناك اتفاق بينهم على الفقرات وتم إبقاؤها بعد إعادة النظر في صياغة بعضها في ضوء ما قدموه من ملاحظات. كما تم تقدير ثبات الاتساق الداخلي من خلال طريقة (كرونباخ ألفا)، حيث بلغت قيمة ألفا (0.75)، وهي كافية لأغراض الدراسة الحالية.

إجراءات الدراسة

تم اختيار عدد من الشعب عشوائيًا من الجامعتين، واستكمل العدد ليصل إلى (500) في كل منهما من خلال التطبيق على عدد من الطلبة في شعبة للعلوم العسكرية تتضمن طلبة من مختلف التخصصات. وقد اقتصرت الدراسة على طلبة من هاتين الجامعتين، ويرجع ذلك إلى سهولة الوصول إليهما، وبسبب تعاون أحد أعضاء هيئة التدريس في جامعة البلقاء التطبيقية الذي سهل عملية التواصل مع أعضاء هيئة التدريس المختارة شعبهم وحثهم على التعاون مع الباحثين. وبعد تطبيق الاختبار على عينة المفحوصين والحصول على استجاباتهم، تم إدخال البيانات إلى الحاسوب بعد أن استبعدت الاستجابات النمطية والاستجابات غير التامة. وقد تم إعطاء العلامة (واحد) للإجابة الصحيحة، والعلامة (صفر) للإجابة الخطأ. وبهذا حُولت جميع البيانات بعد عملية التصحيح الآلي إلى (صفر وواحد). تم إدخال البيانات على شكل ملف (Excel) ثم تحويله إلى ملف (Text)، ثم تم إجراء التحليل للبيانات باستخدام البرمجية Differential Distractor Functioning Software: DDFS التي طورها بنفيلد (Penfield, 2010).

المعالجة الإحصائية

للإجابة عن السؤال الأول، تم استخدام طريقة نسبة الأرجحية لمانتل-هانزل المتوفرة في برمجية (DDFS)، وتم تفسير النتائج بناءً على المؤشرات التالية:

1- **مؤشر Z(LOR)**: الذي يحدد وجود أداء تفاضلي للفقرة أم لا، حيث سيكون هناك أداء تفاضلي للفقرة عند مستوى ($\propto = .050$) إذا كانت قيمة ($Z>1.96$) أو ($Z<-1.96$).

2- **مؤشر LOR (**log-odds ratio**)**: الذي يفسر **اتجاه وحجم** الأداء التفاضلي للفقرة**.** حيث يكون DIF لصالح المجموعة المرجعية ممثلة بمجموعة الذكور عندما تكون (LOR 0 >)، ويكون DIF لصالح المجموعة المستهدفة ممثلة بمجموعة الإناث عندما تكون ($LOR< 0$). ويكونDIF كبيرًا إذا كانت ($|LOR|>0.64$)، ومتوسطًا إذا كانت ($0.64 >|LOR|\geq 0.43$)، وصغيرًا إذا كانت ($|LOR| < 0.43$).

3- **مؤشر SELOR (**The standard error of the log-odds ratio**)**: الذي يحدد قيمة الخطأ المعياري لـ LOR.

4- **مؤشر BD (** The Breslow-Day chi-square test of trend in odds ratio**): الذي** يحدد شكل الأداء التفاضلي هل هو منتظم أم غير منتظم، حيث يكون شكل الأداء التفاضلي منتظمًا إذا كان مؤشر ($BD<3.84$)، وغير منتظم إذا كان مؤشر ($BD>3.84$).

وللإجابة عن السؤال الثاني، تم استخدام طريقة نسبة الأرجحية لمانتل-هانزل المتوفرة في برمجية (DDFS)، وتم تفسير النتائج بناءً على المؤشرات التالية:

1- **مؤشر LOR (log-odds ratio):** الذي يفسر **اتجاه وحجم** الأداء التفاضلي للمموه**.** حيث يكون المموه أقل جذبًا للمجموعة المرجعية (الذكور) إذا كانت ($LOR > 0$). ويكون المموه أقل جذبا للمجموعة المستهدفة (الإناث) إذا كانت ($LOR< 0$). ويكونDDF كبيرًا إذا كانت ($0.43\leq|LOR|< 0.64$) ومتوسطًا إذا كانت ($0.43\leq |LOR|\leq 0.64$)، وصغيرًا إذا كانت ($|LOR| < 0.43$).

2- **مؤشر Z(LOR)**: الذي يحدد وجود أداء تفاضلي للمموه أم لا، حيث سيكون هناك أداء تفاضلي للمموه عند مستوى ($\propto = 0.05$) إذا كانت قيمة ($Z>2.4$) أو ($Z<-2.4$).

وللإجابة عن السؤال الثالث، تم استخدام اختبار مربع كاي للاستقلال.

نتائج الدراسة

أولاً: النتائج المتعلقة بالإجابة عن السؤال الأول "*هل تظهر فقرات الاختبار من متعدد في اختبار التفكير الناقد أداءً تفاضلياً لمتغير الجنس على المستوى الجامعي باستخدام طريقة نسبة الأرجحية لمانتل- هانزل***؟"**

يبين الجدول (4) المتوسطات الحسابية والانحرافات المعيارية للأداء في اختبار التفكير الناقد حسب الجنس ودلالة الفرق بين المتوسطين.

الجدول 4

المتوسطات الحسابية والانحرافات المعيارية للأداء في اختبار التفكير الناقد حسب متغير الجنس ودلالة الفرق بين المتوسطين.

الجنس	العدد	الوسط الحسابي	الانحراف المعياري	قيمة t	مستوى الدلالة
ذكور	404	15.90	6.79	2.114	0.035
إناث	526	15.01	5.77		

يلاحظ من الجدول (4) أن المتوسط الحسابي للأداء في اختبار التفكير الناقد منخفض لكل من الذكور والإناث. وأن الفرق بين المتوسطين دال إحصائيًا عند مستوى الدلالة الإحصائية (α=0.05) لصالح الذكور. وهذا يؤكد ما توصلت له الزعبي (2006) من وجود فرق بين الجنسين على المستوى الجامعي. وتفسير هذا الفرق يتطلب فحص الأداء التفاضلي لفقرات الاختبار.

ومن أجل الكشف عن الأداء التفاضلي لفقرات الاختبار تبعًا لمتغير الجنس باستخدام طريقة نسبة الأرجحية لمانتل- هانزل، تمت مطابقة الأفراد في المجموعتين (المرجعية، والمستهدفة) بناءً على العلامة الكلية على الاختبار فقط، وتم استخدام البرمجية الإحصائية (DDFS). ويبين الجدول (5) نتائج الأداء التفاضلي لفقرات اختبار مهارات التفكير الناقد لمتغير الجنس باستخدام طريقة نسبة الأرجحية لمانتل-هانزل.

الجدول 5

نتائج الأداء التفاضلي لفقرات اختبار مهارات التفكير الناقد لمتغير الجنس باستخدام طريقة نسبة الأرجحية لمانتل _ هانزل.

الفقرة	LOR	الفئة المستفيدة	المقدار	SELOR	Z(LOR)	BD	الأداء التفاضلي
1	0.1056	ذكور	صغير	0.1646	0.6416	0.633	منتظم
2	**0.3455**	**ذكور**	**صغير**	**0.1442**	**2.396***	**0.026**	**منتظم**
3	-0.0838	إناث	صغير	0.1458	-0.5748	0.825	منتظم
4	0.1966	ذكور	صغير	0.1473	1.3347	8.848	غير منتظم
5	-0.2921	إناث	صغير	0.1524	-1.9167	16.772	غير منتظم
6	0.17	ذكور	صغير	0.1821	0.9336	3.439	منتظم
7	**0.3866**	**ذكور**	**صغير**	**0.1578**	**2.4499***	**0.071**	**منتظم**
8	0.1916	ذكور	صغير	0.1482	1.2928	0.923	منتظم
9	**0.481**	**ذكور**	**متوسط**	**0.1499**	**3.2088***	**0.195**	**منتظم**
10	0.162	ذكور	صغير	0.1519	1.0665	0.011	منتظم
11	-0.1286	إناث	صغير	0.1589	-0.8093	1.87	منتظم
12	-0.2691	إناث	صغير	0.1573	-1.7107	0.298	منتظم
13	-0.1736	إناث	صغير	0.16	-1.085	1.029	منتظم
14	-0.2274	إناث	صغير	0.1609	-1.4133	5.268	غير منتظم
15	-0.0719	إناث	صغير	0.1528	-0.4705	0.624	منتظم
16	-0.0928	إناث	صغير	0.1638	-0.5665	0.192	منتظم
17	0.0843	ذكور	صغير	0.1643	0.5131	0.381	منتظم
18	-0.1493	إناث	صغير	0.1725	-0.8655	0.855	منتظم
19	0.063	ذكور	صغير	0.1496	0.4211	0.31	منتظم
20	-0.0881	إناث	صغير	0.1488	-0.5921	1.901	منتظم
21	0.0665	ذكور	صغير	0.1536	0.4329	0.069	منتظم
22	-0.2143	إناث	صغير	0.17	-1.2606	0.134	منتظم
23	0.0588	ذكور	صغير	0.1487	0.3954	5.303	غير منتظم
24	0.1522	ذكور	صغير	0.1495	1.0181	0.283	منتظم
25	0.0587	ذكور	صغير	0.1488	0.3945	1.565	منتظم
26	-0.2698	إناث	صغير	0.1661	-1.6243	0.999	منتظم
27	**0.5452**	**ذكور**	**متوسط**	**0.1938**	**-2.8132***	**0.938**	**منتظم**
28	-0.0716	إناث	صغير	0.1827	-0.3919	1.575	منتظم
29	-0.2547	إناث	صغير	0.1525	-1.6702	1.72	منتظم
30	-0.0662	إناث	صغير	0.1529	-0.433	1.532	منتظم
31	**-0.3654**	**إناث**	**صغير**	**0.1676**	**-2.1802***	**3.273**	**منتظم**
32	0.1167	ذكور	صغير	0.1492	0.7822	6.852	غير منتظم
33	0.1757	ذكور	صغير	0.1498	1.1729	0.762	منتظم
34	**-0.3856**	**إناث**	**صغير**	**0.1627**	**-2.37***	**4.598**	**غير منتظم**

الفقرة	LOR	الفئة المستفيدة	المقدار	SELOR	Z(LOR)	BD	الأداء التفاضلي
35	0.0389	ذكور	صغير	0.1556	0.25	0.936	منتظم
36	-0.0598	إناث	صغير	0.1618	-0.3696	0.304	منتظم
37	-0.2046	إناث	صغير	0.164	-1.2476	0.276	منتظم
38	-0.0739	إناث	صغير	0.1591	-0.4645	0.122	منتظم
39	**0.6563**	**ذكور**	**كبير**	**0.1718**	**2.4232***	**1.488**	**منتظم**
40	-0.1324	إناث	صغير	0.1671	-0.7923	0.887	منتظم
41	0.2533	ذكور	صغير	0.1531	1.6545	0.726	منتظم
42	-0.0455	إناث	صغير	0.1475	-0.3085	1.342	منتظم
43	-0.0816	إناث	صغير	0.1535	-0.5316	0.526	منتظم
44	0.0538	ذكور	صغير	0.1495	0.3599	2.199	منتظم
45	0.0383	ذكور	صغير	0.1714	0.2235	1.666	منتظم
46	0.0808	ذكور	صغير	0.176	0.4591	1.193	منتظم
47	**0.557**	**ذكور**	**متوسط**	**0.1674**	**3.3274***	**5.097**	**غير منتظم**
48	0.0464	ذكور	صغير	0.1482	0.3131	9.414	غير منتظم
49	-0.1294	إناث	صغير	0.1499	-0.8632	1.102	منتظم

• الفقرة بالخط الغامق ذات أداء تفاضلي.

يظهر من الجدول (5) أن هناك (8) فقرات ذات أداء تفاضلي في اختبار مهارات التفكير الناقد تبعًا لمتغير الجنس، وهي الفقرات (2,7,9,27,31,34,39,47). وكان الأداء التفاضلي منتظمًا في جميع الفقرات ما عدا الفقرتين (34و47)، لصالح المجموعة المرجعية (الذكور) في الفقرات (2,7,9,27, 39,47)، ولصالح المجموعة المستهدفة (الإناث) في الفقرتين (31و34). وبالنسبة لحجم التأثير، فكان صغيرًا للفقرات (2,7,31,34)، ومتوسطًا للفقرات (9,27,47)، وكبيرًا للفقرة (39). ويبين الجدول (6) مهارات التفكير الناقد التي تنتمي لها الفقرات ذات الأداء التفاضلي. ويلاحظ من الجدول (6) أن الفقرات في ثلاث مهارات (التقويم، والتبرير، والتركيز) لم تبد أداءً تفاضليًا، وأن مهارة الاستدلال استحوذت على أكبر عدد من الفقرات ذات الأداء التفاضلي بواقع ثلاث فقرات من بين (12) فقرة، تلتها مهارة التفسير التي استحوذت على فقرتين لهما أداء تفاضلي من أصل ست فقرات.

الجدول 6

مهارات التفكير الناقد التي تنتمي لها الفقرات ذات الأداء التفاضلي.

مهارات التفكير الناقد	الفقرات	الفقرات ذات الأداء التفاضلي
التحليل	1,2,3,4	2
التفسير	5,6,7,8,9,10	**9**,7
التقويم	11,12,13,14,15,16,17	-
الشرح	18,19,20,21,22,23,24,25,26,27,28,29	**27**
الاستدلال	30,31,32,33,34,35,36,37,38,39,40,41	31,34, **39**
التبرير الاحتمالي	42,43,44,45	-
تعرف الافتراضات	46,47,48	**47**
التركيز	49	-

ويبين الجدول (7) المتوسطات الحسابية والانحرافات المعيارية للأداء في اختبار التفكير الناقد حسب الجنس ودلالة الفرق بين المتوسطين بعد حذف الفقرات ذات الأداء التفاضلي. والفقرات المحذوفة هي الفقرات (9,27,39,47) لكون الأداء التفاضلي لها من الحجم المتوسط أو الكبير.

الجدول 7

المتوسطات الحسابية والانحرافات المعيارية للأداء في اختبار التفكير الناقد حسب الجنس ودلالة الفرق بين المتوسطين بعد حذف الفقرات ذات الأداء التفاضلي.

الجنس	العدد	الوسط الحسابي	الانحراف المعياري	قيمة t	مستوى الدلالة
ذكور	404	14.83	6.41	1.579	0.115
إناث	526	14.20	5.55		

يلاحظ من الجدول (7) أن المتوسط الحسابي للأداء في اختبار التفكير الناقد منخفض لكل من الذكور والإناث، والفرق بين الوسطين غير دالٍ إحصائيًا عند مستوى الدلالة الإحصائية (α=0.05) بعد حذف الفقرات ذات الأداء التفاضلي. وهذا يدلل فعليًا على أن فرق الدلالة الإحصائية الذي كان بين الذكور والإناث ليس فرقًا حقيقيًا، وأن هذا الفرق يعود إلى وجود خلل في بعض فقرات الاختبار، وبالتالي فإنه عند إزالة هذه الفقرات، زال هذا الفرق. وبالتالي أصبح الاختبار أكثر عدالة.

ثانيًا: النتائج المتعلقة بالإجابة عن السؤال الثاني "*هل تظهر **مموهات** فقرات الاختيار من متعدد في اختبار التفكير الناقد أداءً تفاضليًا لمتغير الجنس باستخدام طريقة نسبة الأرجحية لمانتل-هانزل؟***"**

يبين الجدول (8) الفقرات التي تضمنت مموهات ذات أداء تفاضلي لمتغير الجنس في اختبار مهارات التفكير الناقد باستخدام طريقة نسبة الأرجحية لمانتل-هانزل.

الجدول 8

الفقرات التي تضمنت مموهات ذات أداء تفاضلي لمتغير الجنس في اختبار مهارات التفكير الناقد باستخدام طريقة نسبة الأرجحية لمانتل _هانزل.

الفقرة والمموهات	LOR	الفئة المستفيدة	المقدار	SELOR	Z(LOR)	ذكور	إناث
2							
A	**0.5919**	**ذكور**	**متوسط**	**0.1928**	**3.07***	**73**	**130**
C	0.249	ذكور	صغير	0.1913	1.3016	81	114
D	0.2241	ذكور	صغير	0.2041	1.098	64	87
دون إجابة						2	10
3							
A	**-0.6602**	**إناث**	**كبير**	**0.2731**	**-2.4174***	**53**	**32**
C	-0.0025	إناث	صغير	0.1667	-0.015	129	180
D	0.1738	ذكور	صغير	0.2283	0.7613	42	69
دون إجابة						5	15
5							
B	**-0.5496**	**إناث**	**متوسط**	**0.2119**	**-2.5937***	**85**	**56**
C	-0.1527	إناث	صغير	0.1817	-0.8404	94	133
D	**-0.614**	**إناث**	**متوسط**	**0.2413**	**-2.5446***	**53**	**47**
دون إجابة						2	10
7							
A	0.1597	ذكور	صغير	0.2295	0.6959	62	75
B	**0.5962**	**ذكور**	**متوسط**	**0.1872**	**3.1848***	**107**	**192**
C	0.2628	ذكور	صغير	0.2019	1.3016	99	122
دون إجابة						5	7

الفقرة والمموهات	LOR	الفئة المستفيدة	المقدار	SELOR	Z(LOR)	ذكور	إناث
9							
A	0.3415	ذكور	صغير	0.1918	1.7805	90	125
B	**0.6179**	**ذكور**	**متوسط**	**0.1964**	**3.1461***	**79**	**137**
D	0.4642	ذكور	متوسط	0.1955	2.3744	82	124
دون إجابة						2	5
21							
A	-0.4071	إناث	صغير	0.2209	-1.9787	73	66
B	-0.0629	إناث	صغير	0.2026	-0.3105	87	112
D	**0.5385**	**ذكور**	**متوسط**	**0.2036**	**2.6449***	**61**	**125**
دون إجابة						5	11
27							
A	0.4238	ذكور	صغير	0.2511	-1.6878	67	96
B	**0.5436**	**ذكور**	**متوسط**	**0.2101**	**-2.5873***	**183**	**201**
C	0.4143	ذكور	صغير	0.245	-2.2624	99	120
دون إجابة						2	7
31							
A	-0.2472	إناث	صغير	0.2273	-1.0875	71	90
B	-0.1092	إناث	صغير	0.2101	-0.5198	104	146
C	**-0.4694**	**إناث**	**متوسط**	**0.1929**	**-2.4334***	**134**	**142**
دون إجابة						7	13
39							
B	0.426	ذكور	صغير	0.2045	2.0831	109	170
C	0.0183	ذكور	صغير	0.2106	0.0869	108	117
D	**0.7409**	**ذكور**	**كبير**	**0.219**	**3.3831***	**66**	**125**
دون إجابة						13	19
47							
B	0.4483	ذكور	وسط	0.1957	2.2908	118	168
C	**0.5144**	**ذكور**	**متوسط**	**0.206**	**2.4971***	**98**	**150**
D	0.4143	ذكور	صغير	0.2383	2.1582	57	90
دون إجابة						13	25

• المموه بالخط الغامق ذو أداء تفاضلي.

يتبين من الجدول (8) وجود (10) فقرات تتضمن DDF وأن عدد المموهات ذات الأداء التفاضلي فيها يساوي (11) مموهًا. فالفقرة (2) أظهر فيها المموه (A) أداءً تفاضليًا متوسطًا لصالح الذكور، وفي الفقرة (3) أظهر المموه (A) أداءً تفاضليًا كبيرًا لصالح الإناث، وفي الفقرة (5) أظهر المموهان (B, D) أداءً تفاضليًا متوسطا لصالح الإناث، وفي الفقرة (7) أظهر المموه (B) أداءً تفاضليًا كبيرًا لصالح الذكور، وفي الفقرة (9) أظهر المموه (B) أداءً تفاضليًا كبيرًا لصالح الذكور، وفي الفقرة (21) أظهر المموه (D) أداءً تفاضليًا كبيرًا لصالح الذكور، وفي الفقرة (27) أظهر المموه (B) أداءً تفاضليًا متوسطاً لصالح الذكور، وفي الفقرة (31) أظهر المموه (C) أداءً تفاضليًا متوسطًا لصالح الإناث، وفي الفقرة (39) أظهر المموه (D) أداءً تفاضليًا كبيرًا لصالح الذكور. وفي الفقرة (47) أظهر المموه (D) أداءً تفاضليًا متوسطًا لصالح الذكور.

ثالثًا: النتائج المتعلقة بالإجابة عن السؤال الثالث "*ما العلاقة بين الأداء التفاضلي للمموهات والأداء التفاضلي للفقرات؟***"**

أظهرت نتائج تحليل الأداء التفاضلي لفقرات ومموهات اختبار مهارات التفكير الناقد تبعًا لمتغير الجنس باستخدام طريقة نسبة الأرجحية لمانتل-هانزل، أن هناك (8) فقرات ذات أداء تفاضلي من

أصل (49) فقرة، وأن هناك (11) مموهًا لها أداء تفاضلي من أصل (147). ويبين الجدول (9) توزيع الفقرات حسب متغيري الأداء التفاضلي للفقرات والأداء التفاضلي للمموهات حسب متغير الجنس.

الجدول 9

توزيع الفقرات حسب متغيري الأداء التفاضلي للفقرات والأداء التفاضلي للمموهات حسب متغير الجنس.

		هل تتضمن الفقرة مموهات ذات أداء تفاضلي؟		
		نعم DDF	لا DDF	المجموع
هل الفقرة ذات أداء تفاضلي؟	نعم DIF	7	1	8
	لا DIF	3	38	41
المجموع		10	39	49

يلاحظ من الجدول (9) أن هناك (7) فقرات ذات أداء تفاضلي وتتضمن مموهات ذات أداء تفاضلي، وأن هناك فقرة واحدة ذات أداء تفاضلي ولا تتضمن مموهات ذات أداء تفاضلي، وأن هناك (3) فقرات ليست ذات أداء تفاضلي بينما تتضمن مموهات ذات أداء تفاضلي، وأن هناك (38) فقرة نقية من الأداء التفاضلي للفقرات وللمموهات. ولفحص العلاقة بين الأداء التفاضلي للفقرات والأداء التفاضلي للمموهات، تم استخدام اختبار مربع كاي للاستقلال، حيث كانت قيمته تساوي (26.5)، وهي ذات دلاله إحصائية عند مستوى الدلالة (0.01). وهذا يعني وجود علاقة موجبة بين الأداء التفاضلي للفقرات والأداء التفاضلي للمموهات.

مناقشة النتائج

أظهرت النتائج وجود ثماني فقرات ذات أداء تفاضلي لمتغير الجنس، وكان الأداء التفاضلي منتظمًا في ست فقرات، ولصالح الذكور في ست فقرات، وصغيرًا في أربع فقرات، ومتوسطًا في ثلاث فقرات، وكبيرًا في فقرة واحدة. وقد يعزى الأداء التفاضلي للفقرات بين الذكور والإناث إلى عوامل ثقافية-اجتماعية شكلت اعتقادًا لدى الذكور بأنهم أفضل في المواقف التي تتطلب مهارات تفكير عليا، مما يشكل دافعية واتجاهات إيجابية نحو تلك المواقف لدى الذكور أكثر منها لدى الإناث، ومن المعلوم أن إتقان المهارات يرتبط بتقدير الذات والثقة بالنفس (Aladaileh, 1996). كما قد يعزى الأداء التفاضلي للجنس إلى نوع الفقرة، وهناك دراسات تؤكد أن أداء الذكور أفضل من أداء الإناث في فقرات الاختيار من متعدد، وأن الإناث يتفوقن على الذكور في الفقرات المفتوحة (Tayler & Lee, 2012). والفقرات المستخدمة في الاختبار الحالي هي فقرات اختيار من متعدد، وربما هذا يفسر أن عدد الفقرات المتحيزة للذكور أعلى من عدد الفقرات المتحيزة للإناث. وقد يتأثر الأداء التفاضلي للجنس بمحتوى الفقرة، حيث يتفوق الذكور على الإناث في الفقرات التي تتناول الأشكال والجداول والحجج (Garner & Engelhard, 1999).

ويؤشر الأداء التفاضلي للفقرة إلى احتمالية قياس الفقرة لسمة ثانوية، وقد تختلف درجة هذه الاحتمالية باختلاف السمة الرئيسة التي تقيسها الفقرة أو تلك التي يقيسها الاختبار الكلي. وبالفعل، فقد كانت الفقرات الخاصة بمهارات التقويم والتبرير الاحتمالي والتركيز نقية من الأداء التفاضلي للفقرات. وبمراجعة محتوى الفقرات لهذه المهارات، تبين أن معطياتها ومتطلباتها كانت قليلة نسبيًا وأكثر وضوحًا، وهذا يقلل من احتمالية أن تقيس هذه الفقرات سمات ثانوية بالإضافة للمهارة الرئيسة. أما الفقرات الخاصة بمهارات التحليل والتفسير والشرح والاستدلال وتعرف الافتراضات، فقد كانت من بينها فقرات ذات أداء تفاضلي للجنس، وبالتدقيق في محتوى تلك الفقرات يتبين تعدد معطياتها ومتطلباتها وبشكل أكبر في مهارة الاستدلال التي استحوذت على أكبر عدد من الفقرات ذات الأداء التفاضلي، وهذا يزيد من احتمالية أن تقيس الفقرة مهارة ثانوية جنبًا إلى جنب مع المهارة الأساسية. ومهارة الاستدلال تعد من أصعب مهارات التفكير الناقد، خاصة عندما تتعدد المقدمات التي منها يستدل على الاستنتاج.

وبالمحصلة، تشير النتائج إلى أن اختبار التفكير الناقد مدار الاهتمام هو اختبار جيد حيث يتضمن فقط أربع فقرات ذات أداء تفاضلي متوسط أو كبير المقدار هي الفقرات (9,27,39,47) من بين (49) فقرة. ويتوقع أن تؤثر هذه الفقرات في عدالة المقارنة بين الذكور والإناث، خاصة وأنها جميعًا تعمل لصالح الذكور. وبالفعل كان هناك فرق دال إحصائيًا في المتوسط الحسابي للتفكير الناقد بين الذكور والإناث بوجود هذه الفقرات، وأصبح الفرق غير دال إحصائيًا بعد حذف هذه الفقرات. وتجدر الإشارة إلى أن الاختبار احتفظ بنفس القيمة للثبات (α=0.75) بعد حذف هذه الفقرات، أي أن الاختبار أصبح أكثر كفاءة وعدالة.

وعدم وجود فرق بين الذكور والإناث في التفكير الناقد أكدته العديد من الدراسات السابقة (Leach, 2011; Salahshoor & Raflee, 2016; Ghadia et al., 2013). وقد وفر الأداء التفاضلي لمموهات هذه الفقرات وتحليل محتواها معلومات قيمة أشارت إلى وجود خلل في هذه الفقرات أضر بأداء الإناث عليها. فالفقرة التاسعة كانت تتعلق بإعطاء تفسير للتفاوت الكبير في عدد حوادث السير بين السائقين والسائقات، وقد تضمنت هذه الفقرة

المموه" الإناث أكثر خوفًا من الذكور مما يدفعهن للسير بسرعة أقل من سرعة الذكور"، الذي اجتذب الإناث بصورة كبيرة جدًا. وهذا يوضح أن سمة الخوف لدى الإناث تدخل كسمة ثانوية إلى جانب مهارة التفسير الرئيسة. والفقرة (27) تتعلق بالروابط المنطقية بين الانتساب للضمان الاجتماعي والحلم بمستقبل آمن، وقد تضمنت المموه "كل من يحلم بمستقبل آمن منتسب للضمان الاجتماعي "، الذي اجتذب الإناث بصورة كبيرة جدًا. وهذا يوضح أن الفقرة تقيس الميل للمبالغة كسمة ثانوية لدى الإناث إلى جانب مهارة الشرح الرئيسة. وقد تناولت الفقرة (39) مهارة الاستدلال من خلال لجان متعددة تعمل في مهن ذكورية كالحدادة والنجارة والتصوير التي يكون لدى الإناث اهتمام متدنٍ نسبيًا فيها. وتهتم الفقرة (47) بالتعرف إلى الافتراض الذي يستند إليه الاعتقاد بان "السلوك الثقافي هو عملية نقل اجتماعي، حيث لوحظ ذلك عند قردة الشمبانزي قبل سبع ملايين سنة"، وقد تضمنت المموه "الشمبانزي هو الحيوان الوحيد الذي يسلك سلوكًا ثقافيًا، "الذي اجتذب الإناث بصورة كبيرة جدا. وهذا يوضح أن الفقرة تقيس الميل للاهتمام بجزئية كسمة ثانوية لدى الإناث عوضًا عن مهارة التعرف إلى الافتراض.

وقد اتفقت نتائج الدراسة الحالية مع نتائج دراسات العجلوني (Ajlouni,2016) وجرينبيرغ (Greenberg, 2010) فيما يتعلق بالعلاقة بين DIF وDDF. فقد أكدت وجود أربعة أنماط من الفقرات: النمط الأول يمثل الفقرات النقية تمامًا من الأداء التفاضلي وهي الأولى بالاحتفاظ بها في الاختبار ومثل هذه الفقرات تحقق درجة عالية من العدالة. ويمثل النمط الثاني فقرات ذات أداء تفاضلي وتتضمن مموهات ذات أداء تفاضلي ومثل هذه الفقرات تؤثر سلبًا في عدالة الاختبار. ويمثل النمط الثالث فقرات لا تبدي أداءً تفاضليًا لكنها تتضمن مموهًا أو أكثر ذا أداء تفاضلي وهذا يعكس ضعف الأداء التفاضلي للمموه أو اختلاف اتجاه الأداء التفاضلي لمموهين مختلفين ومثل هذه المموهات تحتاج إلى إعادة نظر. ويمثل النمط الرابع فقرات ذات أداء تفاضلي دون أن تكون فيها مموهات ذات أداء تفاضلي وعلى الأغلب في مثل هذه الفقرات تكون الاختلافات بين أفراد المجموعتين المرجعية والمستهدفة قليلة في اختيارهم للمموهات، لكنها في اتجاه واحد.

التوصيات

توفر نتائج الدراسة دليلاً على جودة اختبار التفكير الناقد؛ إذ لا تتعدى نسبة الفقرات الإشكالية (ذات أداء تفاضلي كبير أو متوسط) (8%). وتحديد هذه الفقرات ومعالجتها (بالحذف أو التعديل) يزيد من كفاءة الاختبار وعدالته. واعتمادًا على ذلك، يمكن الخروج بالتوصيات الآتية:

- ضرورة التحقق من الأداء التفاضلي للفقرات والمموهات عند بناء أي مقياس.
- إجراء دراسات مقارنة بين طرق الكشف عن الأداء التفاضلي للمموهات.
- استبعاد الفقرات الإشكالية الأربع من اختبار التفكير الناقد واستخدام (45 فقرة) تحقيقًا للعدالة بين الذكور والإناث، ولزيادة كفاءة الاختبار من حيث تحسين ثبات الدرجات وتقليل وقت الاختبار.

References

Ajlouni, J. (2016). *Examining the differential distractors functioning of math test in the Jordanian national assessment of knowledge Economy for sex and school location*. Ph.D. Dissertation, Yarmouk University, Irbid, Jordan.

Aladaileh, S. (1996). *The influence of sex and educational level on confidence in learning mathematics and its relationship to academic achievement in mathematics for sixth and tenth grades in the governmental schools of Al-Karak governerate*. Master Thesis. Mutah University, Al-Karak, Jordan.

Al Karki, W., & Mahadeen, S. (2019). Critical thinking level among Mutah University students and its relationship with cognitive motivation**.** *Dirasat: Educational Sciences, 46*(1), 321-342.

Anamezie, R., & Nnadi, F. (2018). Evaluation of differential distractor functioning of physics acheivement Battery for quality assurance using multinomial log-linear model. *International Journal of Modern Management Sciences, 7*(1), 28-39.

Autawi, E. (2004). *Disclosing the differential performance of the gender variant of the eighth-grade essential general science exam items in the fourth Amman education dirrectorateal*. Master's Thesis. Amman Arab University, Amman, Jordan.

Banks, K. (2004). *Exploring racial differences in items that differ in cultural characteristics through differential bundle and distractor functioning.* (Order No. 3138573, The University of Wisconsin – Milwaukee). ProQuest Dissertations and Theses, 91-91. Retrieved from: http://search. proquest.com/ docview/305111981?accountid=48928.(30511 1981).

Bowell, T. (2017). Response to the editorial education in a post-truth world. *Educational Philosophy and Theory*, 49(6), 582–585.

Brookfield, D. (1997). Assessing critical thinking. *New Directions for Adult & Continuing Education, 75*, 17-29.

Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items.* Thousand Oaks**,** CA: SAGE Publications.

Diwi, John. (1960). *A Research for certainty* (A translation of Ahmed Al Ahwani). Eissa Al-Babi Al-Halabi Press.

Ellis, B., & Raju, S. (2003). *Test and item bias: what they are, what they aren'n and how to detect them: Measuring-up, assessment issue for teachers, counselors and administrators.* (ERIC Document Reproduction Service No. ED480042).

Erikson, G., & Erikson, M. (2019). Learning outcomes and critical thinking: Ggood intentions in conflict. *Studies in Higher Education, 44*(12), 2293-2303.

Evers, A., Muñiz, J., Hagemeister, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema, 25*(3), 283-291.

Garner, M., & Engelhard Jr, G. (1999). Gender differences in performance on multiple- choice and constructed response mathematics items. *Applied Measurement in Education, 12*(1), 29-51.

Ghadia, I., Abu Bakar, K., Alwia, N., & Taliba, O. (2013). Gender analysis of critical thinking disposition instrument among University Putra Malaysia undergraduate students. *Recent Technological Advances in Education,*27-33.Retreived on March, 31,2020 from: http://www.wseas.us/e-library/conferences/ 2013/Malaysia/EDUETE/EDUETE-03.pdf

Gomez-Benit, J., Sireci, S., Padilla. J., Hidalgo, M., & Benitez, I. (2018). Differential item functioning: Beyond validity evidence based on internal structure. *Psicothema, 30*(1), 104-109.

Greenberg, A. (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.

Haladyna, T. (2004). *Developing and validating multiple-choice test items.* Mahwah, NJ: Lawrence Erlbaum Associates.

Havnes, A., & Prøitz, S. (2016). Why using learning outcomes in higher education? Exploring the grounds for academic resistance and reclaiming the value of unexpected learning. *Educational Assessment, Evaluation and Accountability, 28*(3), 205–223.

Innabi, H., & Dodeen, H. (2006). Content analysis of gender-related differential items functioning TIMIS item in mathematics in Jordan. *School Science and Mathematics, 106*(8),147-189.

International Conference on Education Evaluation. (2018). Future skills: development and assessment. December, 4-6, 2018, Riyadh.

Jalabi, Susan. (2005). *The basics of building psychological and educational tests and standards.* Aladdin Publishing and Distribution House.

Jaradat, Abdullah. (2003). *A comparison between the Mantl-Hanzel's method and the difficulty method of disclosing paragraphs.* Unpublished Master's Thesis. Mu'tah University, Al-Karak.

Karakaya, I. (2012). An investigation of item bias in science & technology subtests and mathematics subtests in level determination exam (LDE). *Educational Science: Theory & practice, 12*(1), 222-229.

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?

Leach, B. (2011). *Critical thinking skills as related to university students' gender and academic discipline.* Dissertation, East Tennessee State University.

Macfarlane, B. (2017). *Freedom to learn: The threat to student academic freedom and why it needs to be reclaimed*. London: Routledge.

Mahmoud, F. (2010). Differential performance of grade-6 science reference test elements and MUTs designed according to the paragraph response theory. *Journal of the Open University of Jerusalem for Research and Studies*, (44), 135-123.

Martinkova, P., Drabinova, A., Liaw, Y., Sanders, E., McFarland, J., & Price, R. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessment. *CBE-Life Science Education, 16*(2), PMC free article.

McWhorter, K.T., & Collins, H. (1992). *Study & thinking skills in college* (2nd edn.). Authors.

Middleton, K., & Laitusis, C. (2007). *Examining test items for differential distractor functioning among students with learning disabilities*. Educational Testing Service. Research Report. Retrieved on February, 9, 2020 from: www.ets.org/Media/Research/pdf/ RR-07-43 pdf.

Mubarak, W. (2010). *Differential performance of the science test paragraphs in the international study of PISA 2006*. Ph.D. Dissertation, Yarmouk University, Irbid, Jordan.

Oudeh, A. (2010). *Measurement and evaluation in the pedagogical process.* 4th Edition, Dar Al-Amal.

Paul, R. (1998). Critical thinking: Placing it at the heart of ethics instruction. *Journal of Development Education, 22*(2), 36-38.

Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In: S. Sinharay, & C. R. Rao (Eds.), *Handbook of* statistics, Vol. 26*: Psychometrics* (pp. 125-167). Elsevier.

Penfield, R. 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. (2010). DDFS: Differential distractor functioning software. A*pplied Psychological Measurement, 34*(8), 646-647.

Petress, K. (2004). Critical thinking: An extended definition. *Education, 124*(3), 461-466.

Rashwan, R. (2015). Differential Performance of the "Watson and Jasper Critical Thinking Test" items using parametric and nonparametric models in the individual response theory and its relationship to diffrerntial performance of testing using multi-group analysis. *Education Magazine, 165*(2), 455-525.

Reeves, T., & Marbach, G. (2016). Contemporary test validity in theory and practice: A primer for discipline- based education researchers. *CBE-Life Science Education, 15*(1). PMC free article.

Sahin, H., French, B., Hand, B., Gunel, M. (2015). Detection of differential item functioning in the Cornell Critical Thinking test between Turkish and United States students. *European Journal of Psychological Assessment, 31*(4), 238-246.

Salahshoor, N., & Raflee, M. (2016). The relationship between critical thinking and gender: A case of Iranian EFL learners. *Journal of Applied Linguistics and Language Research, 3(*2), 117-123.

Sawalmeh, Y. & Kawasma, A. (2000). The effect of variety in the number of correct and incorrect alternatives on the psychometric properties of multiple true false items. *Journal of Damascus University*. *16*(1), 616-88.

Serhan, I. (2000). *Level of critical thinking skills and its relationship with problem solving among Palestinian university students in light of some variables*. Master Thesis. Islamic University, Ghaza, Palestine.

Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice, 23*(2), 226-235. https://doi.org/10.1080/0969594X.2015.1072084

Sternberg, R. (2004). Four alternative futures for education in the United States: It's our choice. *School Psychology Review, 33***(**1). 67-77.

Taylor, C. S., & Lee, Y. (2012). Gender DIF in reading and mathematics tests with mixed item formats. *Applied Measurement in Education, 25*(3), 246-280.

Tsaousis, L., Sideridis, G., & Al-Saawi, F.(2018). Differential distractor functioning as a method for explaining DIF: The case of a national admission test in Saudi Arabia. *International Journal of Testing, 18*(1),1-26.

Ubaydat, A. (2014). *Investigating differential performance of multiple choice test items based on the risk score of the tests*. Unpublished Ph.D. Dissertation, Yarmouk University, Irbid, Jordan.

Zeiky, M. (2003). *A DIF primer*. Princeton, NJ: Educational Testing Service. Retreived on February, 11, 2020 from: www.ets.org/s/proxis/pdf/dif-primer.

Zoabi, A. (2006). *Building critical thinking skillst Test using paragraph response theory*. Unpublished Ph.D. Dissertation. Yarmouk University, Irbid, Jordan