DOI: http://dx.doi.org/10.33948/sjps-ksu-1-10-1

أثر اتجاه الفقرة وعدد فئات الاستجابة على الأداء التفاضلي باستخدام نموذج الاستجابة المتدرجة

أ. عبد العزيز بن حمد العميري[1] أ.د. إسماعيل بن سلامة البرصان[2]

(قدم للنشر 1447/01/10هـ - وقبل 1447/03/23هـ)

المستخلص: هدفت الدراسة إلى الكشف عن أثر اتجاه الفقرة وعدد فئات الاستجابة على الأداء التفاضلي في ضوء متغير الجنس، باستخدام نموذج الاستجابة المتدرجة. وقد اعتمدت الدراسة المنهج الوصفي الارتباطي، وتمثلت أداتها في ست نسخ من مقياس تقدير الذات لروزنبرغ، صُمّمت لتختلف في اتجاه الفقرات (إيجابية/سلبية) وعدد فئات الاستجابة (ثلاثي، رباعي، خماسي)، وطُبقت على عينة مكوّنة من (7200) معلم ومعلمة تم اختيارهم بالطريقة العشوائية البسيطة. أظهرت النتائج أن الفقرات السلبية كانت أكثر عرضة لظهور الأداء التفاضلي (DIF) بين الذكور والإناث، كما كشفت عن تزايدها مع ازدياد عدد فئات الاستجابة؛ إذ ظهرت فقرتان فقط في التدرج الثلاثي، وارتفع العدد إلى ثلاث فقرات في التدرج الرباعي، بينما أظهرت جميع الفقرات السلبية الخمس أداءً تفاضليًا عند استخدام التدرج الخماسي. في المقابل، اتسمت الفقرات الإيجابية بالاستقرار؛ حيث لم يظهر فيها أداء تفاضلي عبر جميع فئات الاستجابة، باستثناء فقرة واحدة أظهرت أداءً تفاضليًا غير منتظم عند استخدام التدرج الخماسي في الصورة الإيجابية من المقياس. وفي ضوء النتائج، قدمت الدراسة عددًا من التوصيات، أبرزها: استخدام الفقرات الموجبة وتقليل الفقرات السلبية عند بناء المقاييس، ولا سيما في التدرجات الرباعية والخماسية، لضمان دقة القياس وعدالته.

الكلمات المفتاحية: اتجاه الفقرة، فئات الاستجابة، الأداء التفاضلي، نموذج الاستجابة المتدرجة.

The Effect of Items Direction and The Number of Response Categories on Differential Item Functioning Using Graded Response Model

Abdul Aziz H. Alomiri[1] *Ismail S. Albursan*[2]

(Submitted 06-07-2025 and Accepted on 15-09-2025)

Abstract: This study examined the effects of item wording direction (positive vs. negative) and response category formats on differential item functioning (DIF) across gender, using the Graded Response Model (GRM). Six versions of the Rosenberg Self-Esteem Scale, systematically varied by item wording and response categories (3, 4, and 5-point), were administered to a random sample of 7,200 teachers. Results indicated that negatively worded items were consistently more susceptible to gender-based DIF, particularly as the number of response categories increased. Specifically, DIF was observed in two items on the three-point scale, three items on the four-point scale, and all five negative items on the five-point scale. In contrast, positively worded items showed minimal evidence of DIF across response formats, with only one instance of non-uniform DIF observed on the five-point scale. These findings emphasize the role of item wording and response format in contributing to DIF and support the use of three response categories when combining positive and negative items to reduce DIF occurrence.

Keywords: Item wording direction, Response categories, Differential item functioning, Graded Response Model.

---

(1) ، (2) قسم علم النفس – كلية التربية- جامعة الملك سعود

(1), (2) Department of Psychology – College of Education – King Saud University

E-mail: alomairi1983@gmail.com *E-mail:* ibursan@ksu.edu.sa

مقدمة

تعد عدالة المقاييس في مجال القياس النفسي مؤشرًا على جودتها، حيث تضمن أن تكون الأدوات المستخدمة لتقييم الأفراد غير متحيزة وتعكس بدقة خصائصهم النفسية أو قدراتهم بغض النظر عن خلفياتهم الثقافية، أو الاجتماعية، أو العرقية، أو الاقتصادية. وتعتبر العدالة من الخصائص المهمة التي ينبغي أن تتحقق عند بناء المقاييس والاختبارات، ويشير هذا المفهوم إلى قدرة الاختبار أو المقياس على توفير فرص متكافئة لكافة الأفراد لإظهار ما يمتلكون من قدرات أو سمات.

ويُعتبر الأداء التفاضلي للفقرة Differential Item Functioning (DIF) أحد الطرق التي يمكن خلالها التحقق من عدالة الاختبار أو المقياس، ويهدف إلى ضمان أن أداء الأفراد المتساويين في القدرة لا يعتمد على المجموعة العرقية أو الثقافية التي ينتمون إليها (Ohiri et al.,2024). وعلى الرغم من أن الأداء التفاضلي والتحيز كلاهما يهتمان بعدالة الاختبار إلا أن مفهوم الأداء التفاضلي يختلف عن التحيز، فالفقرة المتحيزة تظهر أداء تفاضليا لكن الأداء التفاضلي لا يكفي للحكم على أن الفقرة متحيزة، فمفهوم التحيز أعم وأشمل، ويحمل تحيز الفقرة معنًى اجتماعيًا، وعدم العدالة أو المساواة (Camilli, 2006). وفي هذا السياق يرى دورانس وهولاند (Dorans &Holland,1993) أن الأداء التفاضلي يهتم بالخصائص السيكومترية للفقرة فهو ذو طابع إحصائي من حيث كيفية عمل الفقرة في المجموعات المختلفة، بينما يركز التحيز على الجانب الاجتماعي المترتب على تطبيقها وتفسير نتائجها.

ونظرًا لاختلاف آلية بناء المقاييس من حيث صياغة فقراتها وعدد فئات الاستجابة، قد يؤدي ذلك إلى تفاوت في استجابة الأفراد عنها مما قد يسهم في ظهور الأداء التفاضلي بين الفئات المختلفة من المفحوصين؛ إذ كشفت العديد من الدراسات (Harzing et al., 2012; Marsh,2013) أن التعامل مع فئات الاستجابة المختلفة في تدرج ليكرت يتأثر بعوامل متنوعة منها الجنس. وأكدت هذا الجانب دراسة جيتشل وآخرون (Gitchel et al.,2011) التي أظهرت أن تعامل الذكور والإناث يختلفان مع الفقرات السلبية حيث تميل الإناث للحصول على درجة عالية فيها مقارنة بالذكور مما يؤثر على صدق المقياس. ونتيجة لذلك قد يتأثر الأداء التفاضلي بصياغة الفقرة نتيجة للتباين في كيفية تعامل المستجيبين معها حيث أظهرت دراسة كانتيلو وآخرون (Cantillo et al., 2023) أن اتجاه الفقرة في المقاييس النفسية يؤثر على سلوك المستجيبين، حيث يميل بعضهم إلى عدم الموافقة على الفقرات السلبية. كما أن الأداء التفاضلي يتأثر بعدد فئات الاستجابة، إذ كشفت دراسة الحارثي (2022) أن الأداء التفاضلي يختلف بين التدرجات الرباعية والخماسية التي تشمل بديل محايد، وهذا يشير إلى أن اختلاف عدد فئات الاستجابة قد يؤثر على العدالة بين جميع المفحوصين.

وتوجد عدة طرق للكشف عن الأداء التفاضلي للفقرة منها ما يعتمد على النظرية الكلاسيكية (Classical Test Theory-CTT)، ومنها ما يستند على نظرية الاستجابة للفقرة (Item Response Theory-IRT) التي أصبحت شائعة الاستخدام في تطوير وبناء الاختبارات والمقاييس حيث يتم فيها تقدير مَعْلَمات الفقرات بمعزل عن خصائص الأفراد (Person- Free) ويتم تقدير قدرات

الأفراد بمعزل عن خصائص الفقرات (Item-Free) (دي أيالا، 2017). ولنظرية الاستجابة للفقرة نماذج متعددة تختلف تبعًا لطبيعة السمة المقاسة (أحادية أو متعددة الأبعاد) ونوع الاستجابة، فبعضها مخصص للبيانات الثنائية (0،1)، وأخرى للفقرات متعددة الاستجابة (ثلاث فئات فأكثر) ومن بينها نموذج الاستجابة المتدرجة (Graded Response Model-GRM) الذي تم اعتماده في هذه الدراسة. ويعد هذا النموذج من النماذج الشائعة في تطوير المقاييس النفسية، وقد طوّرته سامجيما Samejima عام 1969 وهو امتداد للنموذج ثنائي المَعْلَم (2PL) ويقدّر معلمتي التمييز (a_i) والصعوبة (b_j) لكل فقرة (Baker & Kim, 2017). ويتميز النموذج بوجود معلمة تمييز واحدة لكل فقرة، وعدد من العتبات (τ) يقل بواحد عن عدد فئات الاستجابة. فعلى سبيل المثال، إذا كانت فئات الاستجابة في تدرج ليكرت أربعة يكون عدد العتبات ثلاثًا ($\tau1$، $\tau2$، $\tau3$)، وتكون العتبات ثابتة داخل الفقرة ومرتبة تصاعديا (Wells, 2021).

وعند توظيف نماذج نظرية الاستجابة للفقرة (IRT) في تحليل الأداء التفاضلي، تتعدد الأساليب المستخدمة تبعًا لخصوصية كل نموذج ونوع المعلمات التي يتم تقديرها. فبعض هذه الأساليب يتلاءم مع فقرات ذات استجابات ثنائية مثل طريقة مؤشر المساحة الإشارية لراجو (Raju's Area DIF Measures)، في حين تُعدّ أساليب أخرى أكثر ملاءمة للفقرات ذات الاستجابات المتعددة التي تتراوح بين ثلاث فئات فأكثر مثل طريقة مربع كاي لورد (Lord's Chi-Square) وطريقة اختبار نسبة الأرجحية Likelihood-Ratio Test (LRT).

ونظرا لاعتماد الدراسة الحالية على نموذج الاستجابة المتدرجة (GRM) يعد اختبار نسبة الأرجحية ملائما لتقييم الأداء التفاضلي في ضوء هذا النموذج حيث تقدم الطريقة نهجا مرنا لاختبار الاختلاف في معاملات الصعوبة والتمييز (Teresi et al., 2021). وقد أشار ثيسن وزملاؤه (Thissen et al.,1993) إلى أن اختبار نسبة الأرجحية من الأساليب الإحصائية القوية للكشف عن الأداء التفاضلي، خصوصًا عند توفر حجم عينة كبير كما هو الحال في الدراسة الحالية. وتقوم فكرته على المقارنة بين نموذجين: المقيد (Constrained model): ويتضمن فرضية أن مَعْلَمة الفقرة (التمييز أو الصعوبة) متساوية بين المجموعتين، أما غير المقيد (Unconstrained model)يتضمن فرضية أن مَعْلَمة الفقرة حرة وتختلف بين المجموعتين، وتُجرى المقارنة بين النموذجين باستخدام اختبار نسبة الأرجحية (G2) ويتبع توزيع مربع كاي chi-square، ويعطى بالمعادلة الآتية (Wells,2021):

$$G^2 = -2log\left(\frac{L_A}{L_C}\right)$$

حيث L_A الأرجحية للنموذج المقيد، L_C : الأرجحية للنموذج غير المقيد

ويمكن كتابة المعادلة على الصورة الأتية:

$$G^2 = -2(logL_c - logL_A)$$

وعلى الرغم من تعدد الدراسات التي تناولت بشكل منفصل تأثير كلٍ من اتجاه الفقرة أو عدد فئات الاستجابة على الخصائص السيكومترية للمقياس، إلا أن الأدبيات تفتقر إلى دراسات بحثت أثر التفاعل بين هذين المتغيرين ضمن إطار واحد. إذ لم يُعالج هذا التفاعل بشكل منهجي لدراسة تأثيره على الخصائص السيكومترية، وبوجه خاص على الأداء التفاضلي للفقرات (DIF). ومن بين الدراسات التي تناولت تأثير اتجاه الفقرة على الخصائص السيكومترية دراسة كانتيلو وآخرين (Cantillo et al., 2023) التي أظهرت أن استخدام الفقرات الإيجابية فقط

أدى إلى تحسين مؤشرات الصدق والثبات للمقياس، بينما تسببت الفقرات السلبية في ميل المستجيبين إلى عدم الموافقة عليها، مما أثر سلبًا على الصدق والثبات. وفي السياق ذاته، توصلت دراسة بولوت وبولوت (Bulut & Bulut, 2022)، إلى أن الفقرات السلبية تربك المستجيب وتؤثر سلبًا على البنية العاملية للمقياس.

وأظهرت دراسة سيتياواتي وزملاؤه (Setiawati et al., 2022) أن استخدام الفقرات الإيجابية والسلبية معًا يؤثر على مؤشرات المطابقة في التحليل العاملي التوكيدي، بينما تحسنت هذه المؤشرات بشكل ملحوظ عند فصل الفقرات إلى نموذج ثنائي البُعد، حيث تشبعت الفقرات السلبية في عامل مستقل عن الإيجابية. كما دعمت نتائج دراسة غارسيا فيرنانديز وآخرين (Garcia-Fernandez et al., 2022) هذا التوجه، إذ تبين أن المقياس الذي يحتوي على فقرات إيجابية فقط أو سلبية فقط، أظهر خصائص سيكومترية أفضل مقارنة بالمقياس الذي تضمن كلا النوعين. ومن منظور مغاير نوعًا ما، بينت دراسة شريم (2022) أن استخدام نسبة مرتفعة من الفقرات السلبية (50%) بصياغة صريحة في مقياس الاتجاه نحو الرياضيات أدى إلى انخفاض في دالة المعلومات ومعامل التمييز، مقارنة بنموذج يحتوي على 25% فقط من الفقرات السلبية.

وفي الإطار ذاته، أوضحت دراسة النمرات (2021) أن المقياس الذي تكون فقراته إيجابية اتّسم بمستوى أعلى من الصدق والثبات، وحقق معاملات تمييز أعلى مقارنةً بالذي يتضمن فقرات سلبية. وتدعم نتائج دراسة فيجول كوليت وآخرين (Vigil Colet et al., 2020) هذه الاستنتاجات، حيث أظهرت أن المقياس ذي الفقرات الإيجابية أظهر خصائص سيكومترية أفضل من النسخة المختلطة (إيجابية وسلبية).

من جهة أخرى، أظهرت دراسات عدة أن عدد فئات الاستجابة يؤثر على الخصائص السيكومترية للمقاييس. فقد بيّنت دراسة سانتوسو وآخرين (Santoso et al., 2023) أن معاملات الصعوبة تزداد بزيادة عدد الفئات، دون تأثير على معاملات التمييز.

وكشفت دراسة آيبك وتورمان (Aybek & Toraman, 2022) أن التدرج الخماسي يوفر دالة معلومات أعلى من التدرجين الثلاثي والسباعي دون أن يُحدث فرقًا في معاملات التمييز. أما دراسة زارع والحربي (2022) فقد أظهرت أن التدرج الخماسي يقدم تقديرات أدق لمعالم الصعوبة ودالة المعلومات مقارنة بالتدرج الرباعي. وتوصلت دراسة بورجاس (Borjas, 2020) إلى أن التدرجين الخماسي والسباعي يوفران درجات متقاربة من الثبات. وفيما يتعلق بالأداء التفاضلي وجدت دراسة الحارثي (2022) أن البديل المحايد في التدرج الخماسي لم يؤثر على دالة المعلومات، لكنه أحدث تأثيرا في الأداء التفاضلي.

يظهر من استعراض الدراسات السابقة أنها تناولت أثر اتجاه الفقرة أو عدد فئات الاستجابة بشكل منفصل، وغالبها استند إلى النظرية الكلاسيكية للقياس (CTT). فبعض الدراسات أشارت إلى أن الفقرات السلبية تؤثر سلبًا في الثبات والبنية العاملية للمقاييس، في حين أوضحت دراسات أخرى أن زيادة عدد الفئات يحسّن دقة التقدير دون اتفاق واضح حول العدد الأمثل. ولم تبحث هذه الدراسات التفاعل بين اتجاه الفقرة وعدد الفئات وتأثيرهما على الأداء التفاضلي (DIF) في ضوء النظرية الحديثة للاستجابة للفقرة (IRT). ومن هنا برزت الفجوة البحثية التي تسعى الدراسة إلى معالجتها، من خلال توظيف نموذج الاستجابة المتدرجة (GRM) للكشف عن

أثر كل من اتجاه الفقرة وعدد فئات الاستجابة والتفاعل بينهما على الأداء التفاضلي وفق متغير الجنس.

مشكلة الدراسة وأسئلتها

تُعدّ جودة الخصائص السيكومترية للمقاييس النفسية من العوامل الأساسية التي تؤثر في دقة وموثوقية النتائج البحثية والتشخيصية، إلا أنها قد تتأثر بجوانب تتعلق ببناء المقياس مثل اتجاه الفقرة (إيجابي/سلبي) وعدد فئات الاستجابة. فقد أظهرت العديد من الدراسات أن اتجاه الفقرة سواء أكانت إيجابية أو سلبية أو من حيث استخدامها معاً يعد عاملاً مؤثراً على خصائص المقياس السيكومترية (Setiawati et al.,2022; Ilhan& Guler, 2017). كما تتأثر هذه الخصائص باختلاف عدد فئات الاستجابة، مما أدى إلى وجود تباين بين الباحثين حول العدد المناسب لاستخدامه عند بناء المقاييس النفسية (Alan & Kabasakal,2020; Aybek & Toraman, 2022).

ويؤثر اتجاه الفقرة وفئات الاستجابة في المقاييس النفسية على سلوك المستجيبين، حيث يميل بعضهم إلى عدم الموافقة على الفقرات السلبية (Cantillo et al.,2023). كما قد يختلف تعامل الذكور والإناث مع فقرات المقياس السلبية، حيث أظهرت دراسة إبراهيم وبكّار (Ibrahim & Bkkar,2015) أن الأنثى تجيب بعناية أكثر من الذكر عن الفقرات السلبية. أما من حيث عدد فئات الاستجابة فقد كشفت دراسة ويجترز وآخرون (Weijters et al.,2010) أن الإناث يملن إلى اختيار درجة الموافقة المتطرفة. وأظهرت دراسات أخرى أن الذكور أكثر ميلاً لها (Rammstedt et al., 2017؛ Klar et al., 2022).

وفي ضوء ما تقدم يتضح أن اتجاه الفقرة وعدد فئات الاستجابة يؤثران على الخصائص السيكومترية للمقياس بشكل عام، وهذا التأثير قد يختلف بين الذكور والإناث عند التعامل مع الفقرة، لذلك تبرز الحاجة إلى دراسة تأثير اتجاه الفقرة وفئات الاستجابة معا على الأداء التفاضلي وفقا لمتغير الجنس. كما أن العديد من الدراسات الحديثة أوصت بتقصّي تأثير الفقرات السلبية (Castilla et al., 2024; Setiawati et al., 2022; Cantillo et al., 2023) وفئات الاستجابة المختلفة (Aybek & Toraman, 2022; Donnellan & Rakhshani, 2023) على الخصائص السيكومترية باستخدام نماذج الاستجابة للفقرة، كما أوصت دراسات عربية (الحارثي، 2022؛ الخيري، 2022) بدراسة أثر الفقرات السلبية وفئات الاستجابة وفقاً لمتغير الجنس. واستجابة لهذه التوصيات، إضافة إلى قلة الدراسات التي تتناول أثر اتجاه الفقرة وعدد فئات الاستجابة على الخصائص السيكومترية بشكل عام وعلى الأداء التفاضلي بشكل خاص كونه أحد مؤشرات الصدق، تحددت مشكلة الدراسة الحالية في سعيها للإجابة عن الأسئلة الآتية:

1- ما الفقرات التي تبدي أداءً تفاضلياً وفقا لمتغير الجنس في الصورة الإيجابية والسلبية لمقياس تقدير الذات تبعاً لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي) باستخدام نموذج الاستجابة المتدرجة (GRM)؟

2- ما الفقرات التي تبدي أداءً تفاضلياً وفقا لمتغير الجنس في الصورة الإيجابية لمقياس تقدير الذات تبعاً لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي) باستخدام نموذج الاستجابة المتدرجة (GRM)؟

أهداف الدراسة

1- معرفة الفقرات التي تظهر أداءً تفاضلياً وفقاً لمتغير الجنس في الصورة الإيجابية والسلبية لمقياس تقدير الذات تبعا لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي) باستخدام نموذج الاستجابة المتدرجة (GRM).

2- تقصي الفقرات التي تظهر أداءً تفاضلياً وفقاً لمتغير الجنس في الصورة الإيجابية لمقياس تقدير الذات تبعا لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي) باستخدام نموذج الاستجابة المتدرجة (GRM).

أهمية الدراسة

تكمن الأهمية النظرية للدراسة في توظيف نظرية الاستجابة للفقرة لتقصي تأثير اتجاه الفقرة على الخصائص السيكومترية باستخدام نموذج الاستجابة المتدرجة (GRM)، مما يسهم في توسيع المعرفة حول موضوع لايزال محدود التناول. أما من الناحية التطبيقية، فقد تقدم الدراسة مؤشرات حول أثر كل من اتجاه الفقرة وعدد فئات الاستجابة على الأداء التفاضلي، مما يساعد الباحثين في اتخاذ قرارات مدروسة عند صياغة الفقرات وتحديد فئات الاستجابة المناسبة عند بناء المقاييس. كما قد تفتح الدراسة المجال أمام أبحاث مستقبلية تستكشف عوامل أخرى مؤثرة في الأداء التفاضلي للفقرة.

حدود الدراسة

الحدود الزمانية: طُبّقت الدراسة في العام الدراسي 2024/ 2025م.

الحدود المكانية: المعهد التخصصي للتدريب المهني للمعلمين في سلطنة عمان.

الحدود البشرية: اقتصرت على عينة من المعلمين المتدربين في المعهد التخصصي للتدريب المهني للمعلمين.

الحدود الموضوعية: اقتصرت على تقصي أثر اتجاه الفقرة وعدد فئات الاستجابة على الأداء التفاضلي في ضوء متغير الجنس (ذكر/ أنثى) باستخدام نموذج الاستجابة المتدرجة (GRM).

مصطلحات الدراسة

اتجاه الفقرة (Item direction): طريقة صياغة الفقرة، حيث تكون إما إيجابية تدل الموافقة عليها على ارتفاع القدرة المقاسة، أو سلبية تشير الموافقة عليها إلى انخفاضها (DeVellis & Thorpe,2022).

التعريف الإجرائي لاتجاه الفقرة: نمط صياغة فقرات مقياس تقدير الذات المستخدم في الدراسة، ويتضمن صورًا إيجابية وسلبية، حيث تدل الموافقة على الفقرة الإيجابية على ارتفاع تقدير الذات، بينما تشر الموافقة على الفقرات السلبية إلى انخفاضه.

فئات الاستجابة (Responses Categories): يقصد بها التدرجات التي تتاح للمستجيب ليعبّر عن رأيه حول فقرات المقياس، مثل تدرجات ليكرت المختلفة (Joshi et al., 2015).

ويُقصد بها إجرائيا: التدرجات الثلاثية، والرباعية، والخماسية لمقياس ليكرت التي استُخدمت للإجابة عن فقرات مقياس تقدير الذات في هذه الدراسة.

الأداء التفاضلي للفقرة Differential item functioning (DIF): يقصد به أن احتمالية استجابة الأفراد المتساوين في القدرة عن الفقرة تعتمد على المجموعة التي ينتمون إليها (Ohiri et al.,2024).

ويعرَّف إجرائيا: بأنه الفروق في استجابات الذكور والإناث متساويي القدرة في تقدير الذات، ويُحسب وفق نموذج الاستجابة المتدرجة (GRM) واختبار نسبة الأرجحية (G^2).

منهجية الدراسة وإجراءاتها

منهجية الدراسة

اتبعت الدراسة الحالية المنهج الوصفي الارتباطي لكونه الأكثر ملاءَمةً لتحقيق أهدافها. وتم توزيع الصور المختلفة للمقياس اعتمادًا على تصميم المجموعات العشوائية Random Groups Design من خلال تعيين الممتحنين بشكل عشوائي في التوزيع حيث يُعطى مثلا الممتحن الأول النموذج X، والنموذج Y للممتحن الثاني، ويعطى الممتحن الثالث النموذج X، وهكذا يستمر التوزيع

وتنتج هذه الطريقة مجموعات متكافئة (Kolen & Brennan, 2013).

مجتمع الدراسة

يشمل جميع المتدربين المستهدفين في البرامج طويلة المدى والقصيرة في المعهد التخصصي للتدريب المهني للمعلمين في سلطنة عمان في العام الدراسي 2024-2025 وعددهم (16988) معلّمًا ومعلّمةً منهم (13210) معلمة، و(3538) معلم (المعهد التخصصي للتدريب المهني للمعلمين، 2024).

عينة الدراسة

تمثلت عينة الدراسة في (7200) معلم ومعلمة أي بواقع (47%) من حجم مجتمع الدراسة تم اختيارهم "بطريقة العينة العشوائية البسيطة، وقد تم مراعاة تساوي أعداد أفراد العينة وفقا للنسخ الستّ لمقياس تقدير الذات بواقع (1200) معلم ومعلمة لكل نسخة. وقد حُدِّد هذا العدد لضمان ملاءمة التحليلات الإحصائية، حيث يقترح دي أيالا (2017) أن تكون العينة عند استخدام نموذج الاستجابة المتدرجة لا تقل عن 500 مفحوص. ويوضح جدول (1) توزيع أفراد العينة وفقا لمتغيرات الدراسة.

جدول 1

توزيع أفراد العينة وفقا لمتغيرات الدراسة

فئات الاستجابة	اتجاه الفقرة	الذكور	الإناث	المجموع
ثلاثي	إيجابية وسلبية	500	700	1200
	إيجابية	500	700	1200
رباعي	إيجابية وسلبية	500	700	1200
	إيجابية	500	700	1200
خماسي	إيجابية وسلبية	500	700	1200
	إيجابية	500	700	1200
المجموع	6 صُور	3000	4200	7200

أداة الدراسة

اعتمدت الدراسة على الصورة العربية لمقياس روزنبرغ لتقدير الذات (Rosenberg, 1965) بترجمة وتقنين جرادات (2006)، نظراً لملاءمته لأهداف الدراسة واحتوائه على فقرات إيجابية وسلبية موزعة بالتساوي (خمس فقرات لكل نوع) تقيس تقدير الذات العام كعامل واحد. وقد تم استخدام صورتين للمقياس: الأصلية التي تشمل الفقرات الإيجابية والسلبية، ونسخة معدّلة أعدّها الباحث، تحتوي على الفقرات الإيجابية فقط، بعد تحويل الفقرات السلبية إلى نظيراتها الإيجابية. طُبقت كل صورة باستخدام ثلاثة مستويات من تدرج ليكرت: الثلاثي (موافق، محايد، لا أوافق)، والرباعي (أوافق بشدة، أوافق، لا أوافق، لا أوافق بشدة)، والخماسي (أوافق بشدة، أوافق، محايد، لا أوافق، لا أوافق بشدة).

صدق أداة الدراسة

الصدق الظاهري(Face validity): تم التحقق منه من خلال عرض النسخة المعدّلة من المقياس التي تتضمن فقرات إيجابية، على مجموعة من المختصين في القياس وعلم النفس التربوي، بلغ عددهم تسعة محكّمين. طُلب منهم تقييم مدى مكافئة الفقرات الإيجابية في النسخة المعدّلة لنظيراتها السلبية في النسخة الأصلية، بالإضافة إلى مدى وضوحها وسلامتها اللغوية. وقد اعتُمدت نسبة (80%) معيارًا لقبول الفقرة، وأُجريت تعديلات طفيفة بناءً على ملاحظاتهم.

الصدق البنائي (Construct Validity): يُعد مقياس روزنبرغ لتقدير الذات أحادي البُعد، وقد تم التحقق من صدقه البنائي في الدراسة الحالية عبر التحليل العاملي التوكيدي لجميع النسخ الست، باستخدام ستة محكات

لجودة المطابقة كما أشار إليها (Gana & Broc,2019).

ويوضح الجدول (2) قيم مطابقة النموذج أحادي البُعد.

جدول 2

مؤشرات حسن المطابقة المستخرجة من التحليل العاملي التوكيدي للعامل الواحد

اتجاه الفقرة		الصورة الإيجابية			الصورة الإيجابية والسلبية		
عدد فئات الاستجابة		ثلاثة	أربعة	خمسة	ثلاثة	أربعة	خمسة
المؤشر	معيار التحقق						
RMSEA	RMSEA≤0.08	0.05	0.03	0.05	0.07	0.05	0.02
(GFI)	GFI≥0.90	0.97	0.98	0.97	0.94	0.97	0.99
(AGFI)	AGFI≥0.90	0.98	0.97	0.95	0.91	0.96	0.98
(RMR)	RMR≤0.08	0.01	0.03	0.02	0.01	0.02	0.02
(CFI)	CFI≥0.90	0.95	0.98	0.97	0.91	0.96	0.99
TLI	TLI ≥0.90	0.94	0.97	0.97	0.89	0.95	0.99

يتضح من الجدول (2) أن جميع مؤشرات المطابقة حققت المعايير المحددة، مع استثناء طفيف في مؤشر تاكر لويس (TLI) في نسخة الفقرات الإيجابية والسلبية للتدرج الرباعي، حيث بلغ (0.89)، وهو قريب من الحد المقبول (0.90). وبالنظر إلى *تحقق* بقية المؤشرات، يمكن التغاضي عن هذا الانخفاض البسيط، مما يشير إلى تحقق أحادية البعد في جميع الصور الستة لمقياس تقدير الذات.

ثبات الصور الستّ لمقياس تقدير الذات

يتمتع مقياس تقدير الذات لروزنبرغ بثبات مناسب حيث بلغ ثباته (0.92) في نسخته الأصلية، وبلغ ثباته (0.73) في نسخته العربية التي قام بتقنينها جرادات (2006)، وقد تم إعادة التحقق من ثبات الصُور المختلفة باستخدام معامل ثبات ألفا لكرونباخ (Cronbach's Alpha Coefficient)، ويُوضح الجدول (3) النتائج.

جدول 3

معاملات ثبات ألفا لكرو نباخ للصور الستة لمقياس تقدير الذات

اتجاه الفقرة	فئات الاستجابة	عدد المفردات	معامل ألفا لكرونباخ
فقرات إيجابية	ثلاثي	10	0.84
	رباعي	10	0.78
	خماسي	10	0.88
فقرات إيجابية وسلبية	ثلاثي	10	0.81
	رباعي	10	0.84
	خماسي	10	0.84

يتضح من الجدول (3) أن قيم معامل ألفا لكرونباخ لجميع الصور الستّ للمقياس تراوحت بين (0.78– 0.88)، وجميع هذه القيم تدل على أن المقياس في جميع صوره يتمتع بثبات مناسب لتحقيق غرض الدراسة الحالية.

التحقق من ملائمة البيانات لاستخدام نماذج نظرية الاستجابة للفقرة

أولاً: التحقق من أحادية البُعدUnidimensionality

يقصد بأحادية البُعد وجود قدرة واحدة (θ) تفسّر أداء الأفراد على المقياس. ووفقًا لنتائج التحليل العاملي التوكيدي في جدول (2)، تبيّن التحليل العاملي أن مقياس روزنبرغ بجميع صوره الست المستخدمة في الدراسة يحقق مؤشرات مطابقة مناسبة لنموذج العامل الواحد.

ثانيا: الاستقلال الموضعي Local Independence

يتحقق افتراض الاستقلال الموضعي عند تحقق أحادية البُعد (Hambleton & Swaminathan, 1985)، وهو ما تأكد في بيانات هذه الدراسة. كما تم التحقق منه باستخدام اختبار Q3 ليَن (Yen)، الذي يقيس الارتباط بين بواقي أزواج الفقرات، وذلك عبر برنامج R باستخدام حزمة mirt. ويعرض الجدول (4) معاملات الارتباط بين البواقي لجميع نسخ المقياس الستّ.

جدول 4

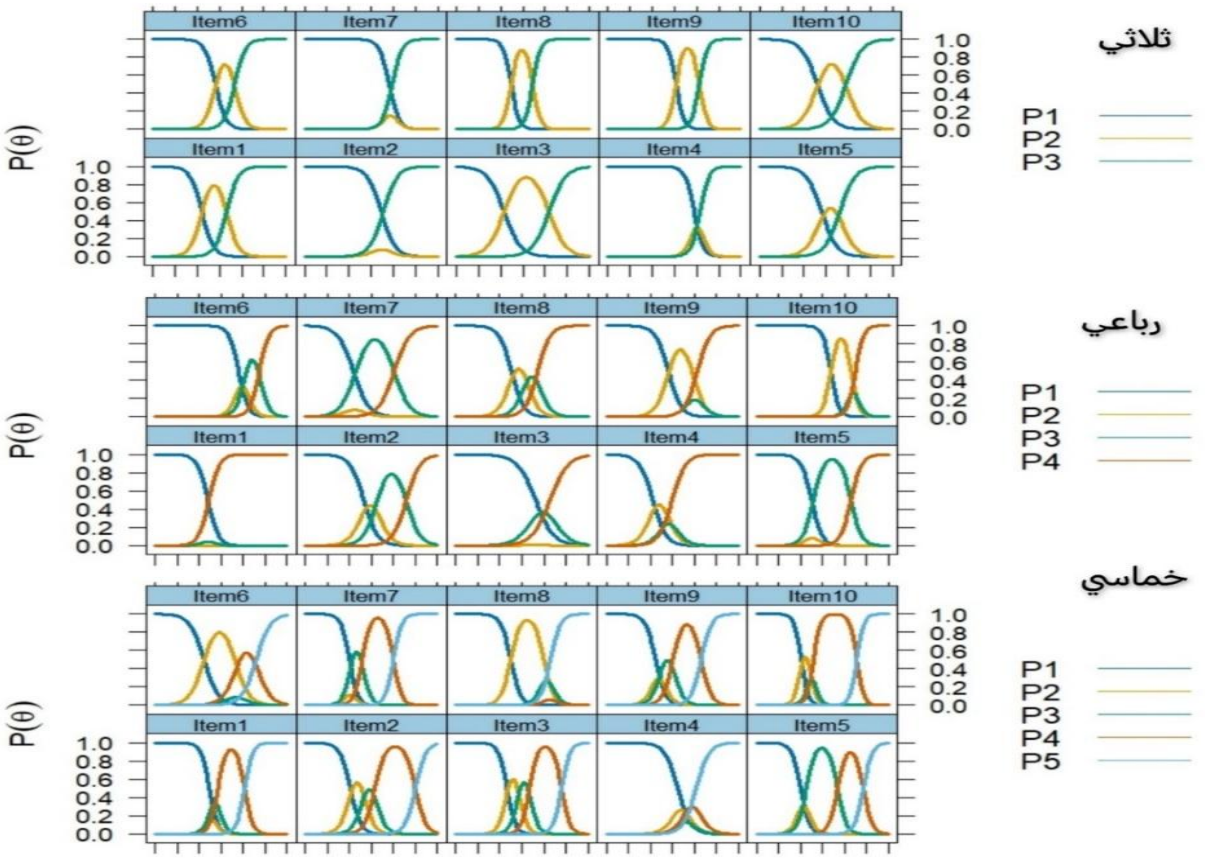
نتائج اختبار Q3 للتحقق من الاستقلال الموضعي الصور الستة للمقياس

اتجاه الفقرة	فئات الاستجابة	أدنى قيمة	أعلى قيمة
فقرات إيجابية	ثلاثي	-0.02	0.18
	رباعي	-0.03	0.20
	خماسي	-0.19	0.01
فقرات إيجابية وسلبية	ثلاثي	-0.18	0.01
	رباعي	-0.19	0.08
	خماسي	-0.17	0.01

أظهرت نتائج جدول (4) أن ارتباطات البواقي في النسخ الست تراوحت بين -0.19 و0.20، وأغلبها قريب من الصفر. ووفقًا لاختبار يَن (Yen)، فهي ضمن النطاق المقبول (±0.20)، مما يدل على تحقق الاستقلال الموضعي (دي أيالا، 2017).

ثالثاً: افتراض منحنى خصائص الفقرة Item Characteristic Curves

تفترض نماذج نظرية الاستجابة للفقرة ومنها نموذج الاستجابة المتدرجة (GRM)، أن لكل فقرة منحنى خصائص لوجستي تصاعدي، يزداد فيه احتمال اختيار الفئات الأعلى مع ارتفاع مستوى القدرة الكامنة (θ)، حيث يمثل كل منحنى فئة استجابة تغطي على نطاق محدد من القدرة، ويتم الانتقال بين الفئات بشكل تدريجي ومنطقي (Baker & Kim, 2017). وقد تم التحقق من هذا الافتراض عبر منحنيات خصائص الفقرات لجميع النسخ الست، وأظهرت النتائج مطابقة الفقرات للنموذج مع انتقال منتظم بين الفئات وخلوّها من الأنماط غير المنطقية. ونظرًا لتشابه أنماط المنحنيات عبر النسخ المختلفة، اقتصر الشكل (1) على عرض منحنيات الاستجابة البديلة (ORFs) لفقرات الصورة الإيجابية والسلبية، كمثال توضيحي.



شكل 1: منحنيات الاستجابة البديلة (ORFs) لفقرات الصورة الإيجابية والسلبية وفقًا لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي)

رابعا: التحرر من السرعة Speediness

يفترض نموذج الاستجابة المتدرجة (GRM) أن عدم استجابة الفرد للفقرة يعود إلى نقص في قدرته، وليس بسبب محدودية الوقت، ويتحقق هذا الافتراض بوجود أُحادية البُعد (Hambleton & Swaminathan, 1985). وفي الدراسة الحالية، تحقق هذا الشرط نظرًا لطبيعة المقياس وعدم ارتباطه بسرعة الأداء، حيث أُتيح للمشاركين وقتٌ كافٍ للإجابة.

التحقق من مطابقة الأفراد لنموذج الاستجابة المتدرجة

تم استخدام برنامج (R) عبر حزمة (mirt) لتدريج بيانات العينة الكلية (ن=1200) لكل نسخة من النسخ الست للمقياس، باستخدام نموذج الاستجابة المتدرجة (GRM)، وذلك بالاعتماد على مؤشري Infit وOutfit وفق متوسط مربعات البواقي (MNSQ). ويعرض الجدول (5) نسبة الأفراد غير المطابقين للنموذج في كل نسخة حسب اتجاه الفقرات وعدد فئات الاستجابة.

جدول 5

نسبة الأفراد غير المطابقين لنموذج الاستجابة المتدرجة (GRM)

الصورة	اتجاه الفقرات	فئات الاستجابة	عدد الافراد	النسبة%
1	موجبة	ثلاثي	29	2.42
2	موجبة	رباعي	76	6.33
3	موجبة	خماسي	30	2.50
4	موجبة وسلبية	ثلاثي	32	2.67
5	موجبة وسلبية	رباعي	29	2.42
6	موجبة وسلبية	خماسي	41	3.42

يوضح الجدول أن نسبة الأفراد غير المطابقين لنموذج الاستجابة المتدرجة (GRM) جاءت بين (2.42% -6.33%) وهي نسب تُعدّ قليلة ومقبولة إحصائيًا. كما جرّب الباحث حذف الأفراد غير المطابقين ولاحظ أن نتائج الأداء التفاضلي لم تختلف جوهريًا عن النتائج المستخرجة من العينة الكلية، مما يعزز استقرار النتائج. وبناءً على ذلك، قرر الباحث الإبقاء على جميع أفراد العينة دون حذف، وذلك للحفاظ على توازن حجمها بين صور المقياس.

مطابقة الفقرات لنموذج الاستجابة المتدرجة

اعتمدت الدراسة على مؤشر المطابقة ($S\text{-}X^2$) الذي طوّره أورلاندو وثيسن (Orlando & Thissen, 2000) والذي ينبغي أن يكون غير دال إحصائيًا، مما يدل على مطابقة جيدة. ويعرض الجدول (6) مؤشرات مطابقة الصورة الإيجابية من المقياس لنموذج الاستجابة المتدرجة، بحسب عدد فئات الاستجابة.

جدول 6

إحصاءات المطابقة للصورة الإيجابية لمقياس تقدير الذات لنموذج الاستجابة المتدرجة حسب عدد فئات الاستجابة

الفقرة	عدد فئات الاستجابة					
	ثلاثة		أربعة		خمسة	
	$S\text{-}X^2$	القيمة الاحتمالية (p)	$S\text{-}X^2$	القيمة الاحتمالية (p)	$S\text{-}X^2$	القيمة الاحتمالية (p)
1	32.1	0.962	36.08	0.789	43.38	0.891
2	39.26	0.060	54.49	0.183	45.65	0.915
3	36.84	0.430	34.79	0.938	52.28	0.311
4	38.88	0.475	70.78	0.140	55.47	0.708
5	51.33	0.129	47.64	0.113	59.17	0.110
6	43.78	0.122	110.36	**0.000**	53.3	0.685
7	25.11	0.763	42.1	0.467	43.23	0.629
8	51.97	0.139	51.29	0.309	46.5	0.534
9	38.42	0.586	47.93	0.355	59.46	0.637
10	24.61	0.98	162.68	**0.000**	44.95	0.432

لنموذج الاستجابة المتدرجة (GRM) لفقرات مقياس تقدير الذات.

كما يوضح الجدول (7) مؤشرات المطابقة للصورة الإيجابية والسلبية حسب فئات الاستجابة

جدول 7

مؤشرات المطابقة للصورة الإيجابية والسلبية حسب فئات الاستجابة لنموذج الاستجابة المتدرجة حسب عدد فئات الاستجابة

الفقرة	عدد فئات الاستجابة					
	ثلاثة		أربعة		خمسة	
	$S\text{-}X^2$	القيمة الاحتمالية (p)	$S\text{-}X^2$	القيمة الاحتمالية (p)	$S\text{-}X^2$	القيمة الاحتمالية (p)
1	25.64	0.059	32.04	0.953	81.97	0.480
2	15.96	0.143	45.22	0.021	64.51	0.424
3	13.83	0.679	38.62	0.352	54.46	0.796
4	16.86	0.816	38.43	0.496	62.62	0.898
5	21.86	0.408	49.28	0.176	138.82	0.000
6	21.74	0.415	43.57	0.126	81.9	0.199
7	16.71	0.213	25.58	0.741	70.15	0.751

8	16.28	0.843	49.94	0.187	119.55	0.007
9	17.14	0.445	39.19	0.551	81.98	0.603
10	18.63	0.814	23.34	0.991	61.61	0.779

يتضح من الجدولين (6) و(7) أن معظم فقرات مقياس تقدير الذات مطابقة لنموذج الاستجابة المتدرجة (GRM)، مع استثناءات محدودة. ففي الجدول (6)، جاءت جميع الفقرات في الصورة الإيجابية مطابقة للنموذج، باستثناء الفقرتين (6) و(10)، وذلك عند استخدام تدرج رباعي لفئات الاستجابة. أما في الصورة التي تتضمن فقرات إيجابية وسلبية، فكانت الفقرة (2) غير مطابقة عند استخدام أربع فئات، والفقرتان (5) و(8) غير مطابقتين عند استخدام خمس فئات. ورغم ظهور بعض الفقرات غير مطابقة لنموذج الاستجابة المتدرجة، فقد أبقي عليها استنادًا إلى ما أوضحه (Sinharay &Haberman, 2014) من أن الفقرات التي تُظهر عدم مطابقة إحصائية قد تكون ضرورية للحفاظ على توازن المحتوى أو لتحقيق أغراض عملية مثل تقصي الأداء التفاضلي (DIF) أو المعادلة (equating). كما أكدت دراسة (Fahrmann et al., 2022) أن عدم المطابقة لا يبرر الحذف ما لم تكن له آثار عملية جوهرية على النتائج، وأن حذف البنود قد يؤدي إلى إضعاف صدق المحتوى.

نتائج الدراسة ومناقشتها

نتائج السؤال الأول: ما الفقرات التي تبدي أداءً تفاضلياً وفقا لمتغير الجنس في الصورة الإيجابية والسلبية لمقياس تقدير الذات تبعا لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي) باستخدام نموذج الاستجابة المتدرجة؟

للإجابة عن السؤال تم تحليل البيانات باستخدام نموذج الاستجابة المتدرجة (GRM)، وتم التحقق من الأداء التفاضلي (DIF) باستخدام اختبار نسبة الأرجحية، وذلك بهدف اختبار فرضيات تكافؤ معلمات التمييز (a)، الصعوبة (b)، وكذلك اختبار تكافؤ المعلمتين معًا بين الذكور والإناث. وقد أُجريت التحليلات باستخدام حزمة mirt ضمن البرمجة الإحصائية R. ويعرض الجدول (8) هذه النتائج.

جدول 8

نتائج اختبار الأداء التفاضلي للفقرات (DIF) تبعًا لمتغير الجنس في الصورة الإيجابية والسلبية للمقياس حسب عدد فئات الاستجابة

عدد فئات الاستجابة		ثلاثة			أربعة			خمسة		
الفقرة	الفرضية H_0	G^2	p	اتجاه DIF	G^2	p	اتجاه DIF	G^2	p	اتجاه DIF
1	تساوي التمييز (a)	0.23	0.870	لا يوجد	0.03	0.921	لا يوجد	0.20	0.800	لا يوجد
	تساوي الصعوبة (b)	4.20	0.409		5.27	0.307		2.73	0.604	
	تساوي الكل	4.20	0.788		6.24	0.365		3.84	0.598	
2	تساوي التمييز (a)	0.01	0.924	ذكور	6.16	0.039	غير منتظم	4.12	0.088	ذكور
	تساوي الصعوبة (b)	8.38	0.076		88.89	0.01>		50.41	0.01>	
	تساوي الكل	22.83	0.01>		90.15	0.01>		52.59	0.01>	
3	تساوي التمييز (a)	0.66	0.870	لا يوجد	2.01	0.351	لا يوجد	0.05	0.832	لا يوجد
	تساوي الصعوبة (b)	0.89	0.801		2.20	0.761		3.16	0.590	
	تساوي الكل	0.95	0.903		2.45	0.917		3.67	0.598	
4	تساوي التمييز (a)	0.01	0.924	لا يوجد	0.01	0.921	لا يوجد	0.34	0.800	لا يوجد
	تساوي الصعوبة (b)	2.17	0.676		0.17	0.982		4.41	0.560	
	تساوي الكل	2.99	0.788		0.17	0.996		4.54	0.598	

عدد فئات الاستجابة		ثلاثة			أربعة			خمسة		
الفقرة	الفرضيةH_0	G^2	p	اتجاهDIF	G^2	p	اتجاهDIF	G^2	p	اتجاهDIF
5	تساوي التمييز(a)	0.92	0.870	لا يوجد	9.08	0.012	غير منتظم	38.41	0.01>	غير منتظم
	تساوي الصعوبة(b)	1.78	0.685		30.24	0.01>		355.30	0.01>	
	تساوي الكل	1.80	0.903		33.18	0.01>		368.37	0.01>	
6	تساوي التمييز(a)	0.17	0.870	لا يوجد	0.23	0.808	لا يوجد	4.06	0.088	غير منتظم
	تساوي الصعوبة(b)	0.25	0.979		1.49	0.856		426.80	0.01>	
	تساوي الكل	0.27	0.965		1.51	0.917		457.93	0.01>	
7	تساوي التمييز(a)	2.96	0.428	لا يوجد	0.32	0.808	لا يوجد	0.43	0.800	لا يوجد
	تساوي الصعوبة(b)	2.45	0.676		0.46	0.982		3.97	0.560	
	تساوي الكل	3.10	0.788		1.84	0.917		3.99	0.598	
8	تساوي التمييز(a)	8.23	0.041	غير منتظم	0.68	0.410	لا يوجد	16.55	0.01>	غير منتظم
	تساوي الصعوبة(b)	12.57	0.019		6.00	0.279		387.75	0.01>	
	تساوي الكل	14.14	0.014		6.07	0.365		430.13	0.01>	
9	تساوي التمييز(a)	0.28	0.870	لا يوجد	14.86	0.01>	غير منتظم	29.83	0.01>	غير منتظم
	تساوي الصعوبة(b)	0.00	1.000		81.19	0.01>		49.43	0.01>	
	تساوي الكل	1.27	0.903		92.06	0.01>		52.72	0.01>	
10	تساوي التمييز(a)	0.15	0.870	لا يوجد	0.31	0.808	لا يوجد	0.13	0.800	لا يوجد
	تساوي الصعوبة(b)	1.05	0.801		4.76	0.317		3.70	0.560	
	تساوي الكل	1.05	0.903		5.74	0.365		3.76	0.598	

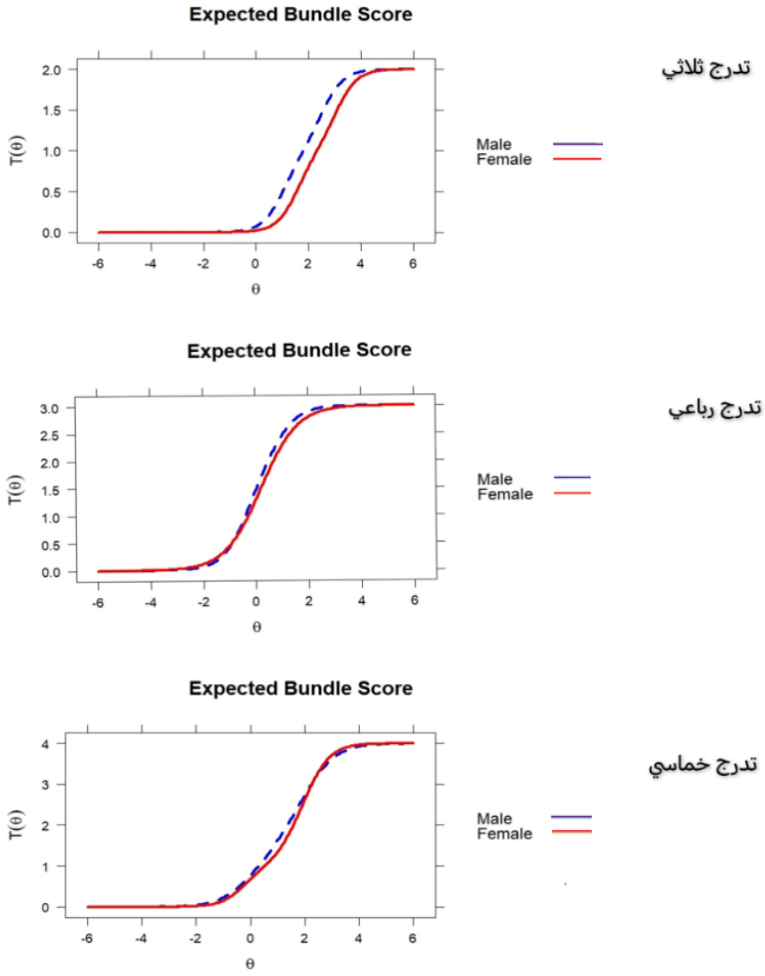
يتضح من الجدول (8) وجود اختلافات ذات دلالة إحصائية بين الجنسين في الفقرات ذات الصياغة السلبية وهي (2، 5، 6، 8، 9) حيث ظهرت فروق في بعض معاملات الصعوبة (b)، وكذلك في معاملات التمييز (a) في بعض الحالات، مما يشير إلى استجابة غير متكافئة بين الذكور والإناث لهذه الفقرات باختلاف عدد فئات الاستجابة. كما يلاحظ من الجدول أن زيادة عدد فئات الاستجابة يؤدي إلى زيادة الفقرات السلبية التي تُظهر أداءً تفاضليًا. فعند استخدام ثلاث فئات استجابة، ظهرت فقرتان فقط (2، 8) بأداء تفاضلي، بينما ارتفع العدد إلى ثلاث فقرات (2، 5، 9) عند استخدام أربع فئات استجابة. أما عند استخدام خمس فئات استجابة، فقد أظهرت جميع الفقرات السلبية (2، 5، 6، 8، 9) أداءً تفاضليًا. ويُعزى ذلك إلى التعقيد المعرفي للفقرات السلبية، الذي يتفاقم مع تعدد فئات الاستجابة، مما يؤدي إلى تباين في أنماط المعالجة بين الأفراد ويدعم هذا التوجه ما ورد في الأدبيات السابقة (Garcia-Fernandez et al.,2022;Santoso et al., 2023) حول تأثير الصياغة السلبية على خصائص الفقرة، مثل احتمالية ظهور الغموض أو سوء الفهم لدى المستجيب. وقد أكدت دراسات حديثة (Fukudome & Takeda, 2024؛ Rokeman, 2024) أن الفقرات السلبية تفتقر إلى الاتساق وتُظهر خصائص سكومترية غير مستقرة، مما يزيد من احتمالية ظهور أداءً تفاضلي (DIF).

كما يمكن تفسير ازدياد عدد الفقرات السلبية التي تُظهر أداءً تفاضليًا (DIF) مع ارتفاع عدد فئات الاستجابة باتساع مدى الاختلاف بين آراء المستجيبين وتوزعها عبر جميع الفئات، وهو ما ينعكس في زيادة التباين في معاملات الصعوبة والتمييز. وقد أكد (Taherdoost, 2019) أن تدرجات ليكرت ذات الفئات الأكثر تُبرز نطاقًا

أوسع من استجابات الأفراد مقارنة بالتدرجات الأقصر، مما يسمح برصد فروق أدق وأكثر تنوعًا. وبالمثل، أشار (Allahyari et al., 2016) إلى أن اعتماد أربع أو خمس فئات بدلاً من ثلاث يعزز قدرة الاختبار على الكشف عن البنود التي تُظهر أداءً تفاضليًا. وعليه، ففي سياق الدراسة الحالية قد يكون تعدد العتبات التقديرية في نموذج الاستجابة المتدرجة (GRM) أسهم في رفع حساسية كشف الفروق في أنماط الاستجابة بين الجنسين.

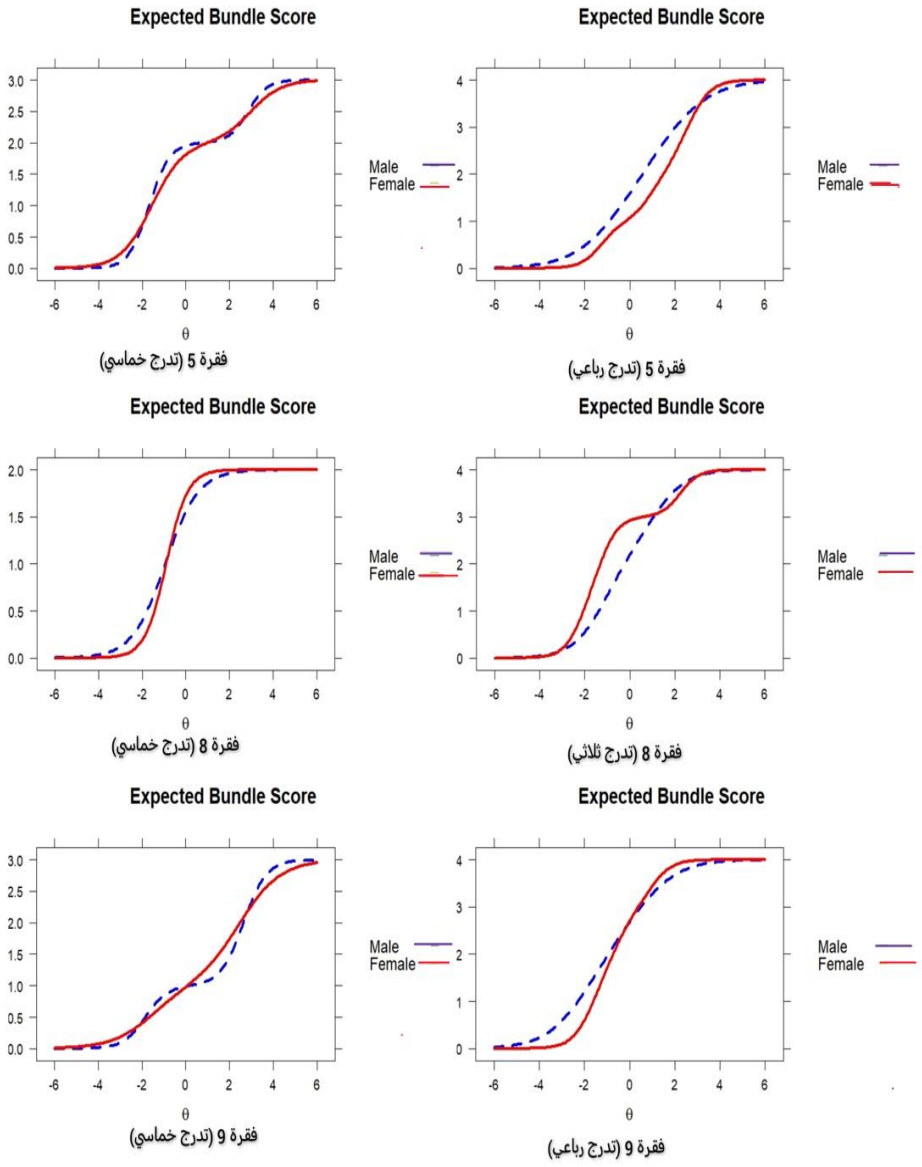
كما أن الفقرات السلبية تُظهر فروقًا أكبر في معامل التمييز (Alvarez et al.,2018)، مما يُمكن من الكشف عن أداء تفاضلي أكثر وضوحًا، خصوصًا بين الذكور والإناث حيث أشار باسمان (Basman, 2023) أن اختلاف معامل التمييز بين المجموعات يؤدي إلى احتمالية كبيرة لظهور الأداء التفاضلي.

ويلاحظ من الجدول (8) أن الفقرات السلبية أظهرت جميعها أداءً تفاضليا غير منتظماً باستثناء الفقرة (2) في النسختين ذات فئات الاستجابة الثلاثية والخماسية، حيث ظهر الأداء التفاضلي بشكل منتظم لصالح الذكور، كما هو موضح في الشكل (2).



شكل 2: اتجاه الأداء التفاضلي للفقرة (2) حسب فئات الاستجابة (ثلاثة، أربعة، خمسة)

يوضح الشكل (2) أن الفقرة (2)، والتي نصها: "أعتقد أنني لست جيدًا"، أظهرت أداءً تفاضليًا متسقًا لصالح الذكور في كُلٍّ من التدرجين الثلاثي والخماسي. أما في التدرج الرباعي، فقد كان الأداء التفاضلي غير منتظم، إلا أنه اقترب بشكل كبير من النمط المنتظم لصالح الذكور. ويعكس هذا النمط من النتائج اتساقًا واضحًا يشير إلى وجود تحيّز في الفقرة لصالح الذكور، بغضّ النظر عن عدد فئات الاستجابة المستخدمة. ويمكن تفسير ذلك بسبب وجود البديل المحايد، فقد تميل بعض الإناث إلى اختياره كاستجابة تحفظية في بعض الفقرات لتجنّب تبنّي موقف سلبي صريح تجاه الذات، بينما يميل الذكور إلى اختيار استجابات مباشرة. أما الفقرات السلبية الأخرى (5، 8، 9) فقد ترتبط بمضامين مختلفة لا تستدعي بالضرورة اللجوء إلى الحياد، ولهذا ظهر فيها الأداء التفاضلي بصورة غير منتظمة. ويدعم هذا التفسير ما توصلت إليه بعض الدراسات السابقة من أن وجود البديل المحايد يدفع بعض الأفراد إلى استخدامه بدلاً من تبنّي إجابة صريحة (de Rezende & de Medeiros, 2022)، وقد يُستعمل كخيار هروب في الأسئلة ذات الطبيعة الحساسة اجتماعيًا (Kankaraš & Capecchi, 2024)، وهو ما يفسر ظهور الأداء التفاضلي في هذه الفقرة تحديدًا. ويوضح الشكل (3) الفقرات (5، 8، 9) التي أظهرت أداءً تفاضليًا غير منتظم.



شكل 3: الأداء التفاضلي غير المنتظم للفقرات (5، 8، 9) حسب اختلاف فئات الاستجابة

أظهرَ الشكل (3) نمطًا متسقًا من الأداء التفاضلي غير المنتظم في بعض الفقرات عند اختلاف عدد فئات الاستجابة، كما في الفقرتين (5، 9) اللتين أظهرتا تفوقًا للذكور عند المستويات المنخفضة للقدرة، في حين تقاربت التوقعات عند المستويات العليا، لكنها مالت قليلاً لصالح الإناث. أما في التدريج الخماسي، فتغيرت الفروق بشكل ملحوظ، مع تقاطع منحنيات الأداء بين الجنسين في أكثر من نقطة، مما يعكس تغيرًا في اتجاه العلاقة بين الجنس ومستوى الأداء. وتُظهر الفقرة (8) نمطًا أكثر تعقيدًا؛ إذ بدت متكافئة في التدريج الثلاثي عند القدرات المتدنية، ثم مالت لصالح الإناث في المستويات المتوسطة، قبل أن تتحول لصالح الذكور في المستويات العليا. أما في التدريج الخماسي، فقد تفوق فيها الذكور عند المستويات المنخفضة، بينما مالت لصالح الإناث في المستويات

المرتفعة. وتشير هذه الأنماط إلى احتمالية وجود تباين في معلمات التمييز والعتبات بين الذكور والإناث، حيث تُظهر منحنيات الأداء المتوقع استجابات مختلفة عند نفس مستوى السمة الكامنة (θ)، ما قد يعكس اختلافًا في مواقع العتبات أو قدرة الفقرة على التمييز بين المجموعتين. وتدعم ذلك دراسة إلهان وجولر (Ilhan & Guler, 2017) التي أوضحت صعوبة تمييز فئات ليكرت في الفقرات السلبية، خاصة عندما تتطلب صياغتها معالجة معرفية أكثر تعقيدًا أو أنماط تفكير تختلف بين الجنسين إذ تستدعي عملية فهم النفي والتفكير المعكوس للاستجابة عنها، وتكون أكثر عرضة لسوء الفهم أو التأويل الانفعالي. وعلى الرغم من اختلاف متغيرات الدراسة الحالية، فإن نتائجها تتفق جزئيًا مع ما توصلت إليه دراسة سانتوسو وآخرين (Santoso et al., 2023)، من حيث تأثير عدد فئات الاستجابة على خصائص الفقرات باستخدام نموذج الاستجابة (GRM).

نتائج السؤال الثاني: ما الفقرات التي تبدي أداءً تفاضليا وفقا لمتغير الجنس في الصورة الإيجابية لمقياس تقدير الذات تبعا لعدد فئات الاستجابة (ثلاثي، رباعي، خماسي) باستخدام نموذج الاستجابة المتدرجة؟

للإجابة عن السؤال تم تدريج البيانات ذلك باستخدام نموذج الاستجابة المتدرجة (GRM) عبر حزمة (mirt) في برنامج R، وتم استخراج مؤشرات الأداء التفاضلي من خلال اختبار نسبة الأرجحية (Likelihood Ratio Test) وفقًا لثلاث فرضيات: الفرضية الخاصة بتساوي معاملات التمييز(a)، وتساوي معاملات الصعوبة(b)، وتساوي كلا المعاملين. والجدول (9) يوضح النتائج.

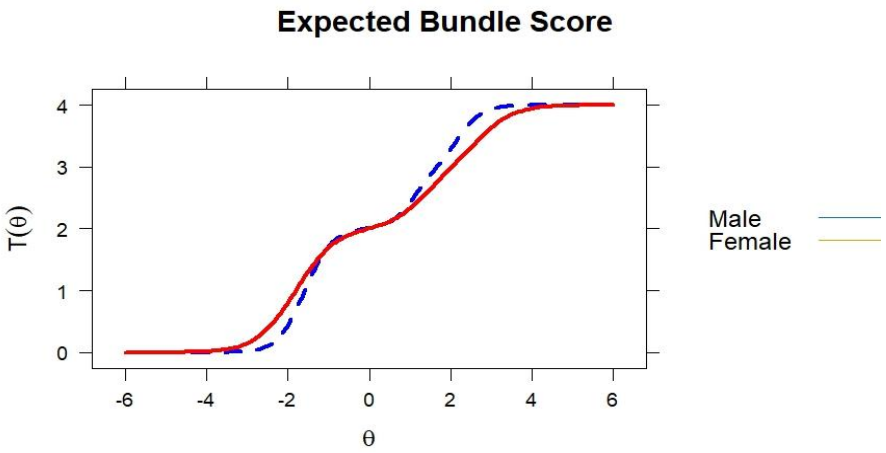
جدول 9

نتائج اختبار الأداء التفاضلي للفقرات (DIF) تبعًا لمتغير الجنس في الصورة الإيجابية لمقياس تقدير الذات حسب عدد فئات الاستجابة

عدد فئات الاستجابة		ثلاثة		أربعة		خمسة	
الفقرة	الفرضية	G^2	p	G^2	p	G^2	p
1	تساوي التمييز(a)	1.01	0.524	0.83	0.777	0.03	0.920
	تساوي الصعوبة(b)	0.45	0.921	0.76	0.860	2.32	0.961
	تساوي الكل	4.05	0.548	2.24	0.883	2.37	0.982
2	تساوي التمييز(a)	0.19	0.735	2.38	0.613	0.26	0.876
	تساوي الصعوبة(b)	0.33	0.921	2.20	0.844	0.62	0.961
	تساوي الكل	0.37	0.982	6.13	0.883	0.72	0.982
3	تساوي التمييز(a)	0.03	0.872	0.11	0.822	1.66	0.496
	تساوي الصعوبة(b)	0.16	0.921	2.50	0.844	1.22	0.961
	تساوي الكل	0.17	0.982	4.13	0.883	2.22	0.982
4	تساوي التمييز(a)	2.39	0.306	1.34	0.777	0.99	0.615
	تساوي الصعوبة(b)	2.28	0.921	2.82	0.844	0.84	0.961
	تساوي الكل	2.58	0.767	3.08	0.883	1.56	0.982
5	تساوي التمييز(a)	3.79	0.172	0.74	0.777	9.76	0.018
	تساوي الصعوبة(b)	1.37	0.921	0.96	0.860	6.91	0.762
	تساوي الكل	3.95	0.548	1.13	0.890	10.51	0.619
6	تساوي التمييز(a)	3.87	0.172	2.69	0.613	3.45	0.316
	تساوي الصعوبة(b)	1.59	0.921	1.63	0.844	6.70	0.762
	تساوي الكل	3.89	0.548	3.46	0.883	7.25	0.982
7	تساوي التمييز(a)	0.62	0.615	0.03	0.875	0.01	0.920
	تساوي الصعوبة(b)	0.48	0.921	1.53	0.844	3.03	0.961

عدد فئات الاستجابة		ثلاثة		أربعة		خمسة	
الفقرة	الفرضية	G^2	p	G^2	p	G^2	p
	تساوي الكل	0.63	0.982	1.68	0.883	3.34	0.982
8	تساوي التمييز (a)	4.26	0.172	0.30	0.794	0.81	0.615
	تساوي الصعوبة(b)	8.60	0.136	1.72	0.844	0.71	0.961
	تساوي الكل	8.99	0.295	1.73	0.883	1.12	0.982
9	تساوي التمييز (a)	0.42	0.65	0.25	0.794	1.65	0.496
	تساوي الصعوبة(b)	0.21	0.921	1.66	0.844	4.20	0.961
	تساوي الكل	1.08	0.982	2.74	0.883	4.96	0.982
10	تساوي التمييز (a)	1.68	0.39	0.23	0.794	0.03	0.920
	تساوي الصعوبة(b)	4.42	0.55	1.56	0.844	1.84	0.961
	تساوي الكل	4.44	0.548	2.52	0.883	2.44	0.982

أظهرت نتائج الجدول (9) عدم وجود فروق دالة إحصائيًا بين المجموعة المرجعية (الإناث) والمجموعة المستهدفة (الذكور) في جميع الفقرات، وذلك في جميع نماذج فئات الاستجابة الثلاثية والرباعية والخماسية، باستثناء الفقرة (5) التي نصها: «أشعر أن لدي الكثير مما أفخر به»، عند استخدام التدرج الخماسي، إذ أظهرت الفرضية الخاصة بتكافؤ معاملات التمييز فرقًا دالًا إحصائيًا، مما يشير إلى وجود أداء تفاضلي غير منتظم في هذه الفقرة، كما يوضحه الشكل (4).



شكل4: الأداء التفاضلي غير المنتظم للفقرة (5) في الصورة الإيجابية للمقياس ذات فئات التدرج الخماسي

يتضح من الشكل (4) أن الفقرة (5) أظهرت أداءً تفاضليًا غير منتظم في صيغة التدريج الخماسي فقط، إذ سجلت الإناث درجات أقل عند المستويات المنخفضة من القدرة (θ < 1.5)، ثم انعكس النمط لصالحهن عند المستويات العليا. ويُعزى هذا النمط إلى أن زيادة عدد فئات الاستجابة تتيح الكشف عن فروق دقيقة في أنماط الاستجابة بين الجنسين (Allahyari et al.,2016). كما يمكن أن يعكس هذا التباين في الأداء اختلافًا في أساليب الاستجابة، إذ يميل الذكور إلى اختيار الفئات المتطرفة مقارنة بالإناث (Zhang et al., 2022)، مما يرفع درجاتهم المتوقعة عند القدرة المنخفضة، بينما يظهر تفوق الإناث عند المستويات العليا نتيجة تحفظهن في استخدام الفئات القصوى. وقد يكون لمضمون الفقرة دور في ذلك، إذ يختلف الذكور والإناث في التعبير عن الفخر، حيث يرتبط

الفخر لدى الذكور غالبًا بالنجاح الخارجي، في حين تميل الإناث لربطه بالجوانب الشخصية أو الاجتماعية، كما قد تتجنب بعضهن التعبير القوي عن الفخر لأسباب اجتماعية (Chaplin, 2015). وفي هذا السياق، أكدت دراسة بروان وزملاؤه (Brown et al.,2020) أن كشف الأداء التفاضلي يصبح أكثر تعقيدًا في الفقرات التي تقيس سمات نفسية ذاتية، إذ قد تعكس الفروق في الاستجابة اختلافات حقيقية في البنية النفسية بين المجموعات، وليس تحيزًا في أداة القياس.

وبناءً على ما سبق، يتضح أن مقياس تقدير الذات في صورته الإيجابية يتمتع بدرجة عالية من الإنصاف بين الذكور والإناث، بغض النظر عن عدد فئات الاستجابة المعتمدة، إذ لم تُظهر الفقرات أداءً تفاضليًا سوى في حالة واحدة. وتتفق نتائج هذه الدراسة جزئيًا مع ما توصلت إليه دراسات سابقة، مثل دراسة آيبك وتورمان (Aybek & Toraman, 2022) ودراسة سانتوسو وآخرين (Santoso et al., 2023)، التي أشارت إلى أن عدد فئات الاستجابة لا يؤثر في خصائص الفقرات عند تحليلها باستخدام نموذج الاستجابة المتدرجة (GRM). وقد أظهرت نتائج الدراسة الحالية بدورها استقرار معاملات الصعوبة والتمييز عبر اختلاف عدد الفئات، مما يعزز ما ذهبت إليه تلك الدراسات، ويدل على أن خصائص الفقرات الإيجابية ظلت ثابتة إلى حد كبير رغم تغيير فئات الاستجابة.

التوصيات

1. يُوصى باعتماد ثلاث فئات للاستجابة عند تصميم أدوات تحتوي على فقرات إيجابية وسلبية، نظرًا لقلة عدد الفقرات التي أظهرت أداءً تفاضليًا مقارنة بالتدرجين الرباعي والخماسي، مما يعزز عدالة القياس.
2. يُفضّل استخدام الفقرات الإيجابية في بناء أدوات القياس النفسي والتربوي، لما تتمتع به من خصائص سيكومترية مستقرة وخلوّها من مؤشرات الأداء التفاضلي بين الجنسين، بغض النظر عن عدد فئات الاستجابة.
3. ينبغي تحديد عدد فئات الاستجابة بما يتلاءم مع خصائص العينة وطبيعة السمة المقاسة، نظرًا لتأثيره في حساسية نماذج IRT وقدرتها على الكشف عن الفروق الدقيقة في الاستجابات.

المقترحات للبحوث المستقبلية

1. دراسة أثر نسبة الفقرات السلبية في أدوات القياس على الأداء التفاضلي، من خلال مقارنة أدوات أخرى تتضمن نسبًا مختلفة (25%، 50%، 75%)، مع الإشارة إلى أن الدراسة الحالية استخدمت مقياسًا متوازنًا من حيث اتجاه الفقرات.
2. بحث تأثير التفاعل بين اتجاه الفقرة، وعدد فئات الاستجابة، وحجم العينة على الأداء التفاضلي في ضوء نماذج أخرى من نماذج نظرية الاستجابة للفقرة (IRT).
3. إجراء دراسات مماثلة على عينات مختلفة، مع تجريب تدرجات استجابة متنوعة (مثل 3، 4، 5، 6، 7، 9، 11) واستخدام طُرق أخرى لاختبار الأداء التفاضلي للتحقق من ثبات الأثر في سياقات متعددة.

أولاً: المراجع العربية

جرادات، عبد الكريم. (2006). العلاقة بين تقدير الذات والاتجاهات اللاعقلانية لدى الطلبة الجامعيين. *المجلة الأردنية في العلوم التربوية*، *2*(3)، 143 - 153.

الحارثي، حصة. (2022). أثر البديل "محايد" في تدريج ليكرت على الأداء التفاضلي ودالة المعلومات لمقياس الاتجاه نحو الحاسب الآلي. *مجلة العلوم التربوية*، *30*(3)، 245 – 268.

الخيري، محمد. (2022). أثر استجابة "محايد" على الخصائص السيكومترية للمقاييس النفسية وعلاقة ذلك ببعض المتغيرات الديموغرافية. *المجلة العربية للنشر العلمي*،22، 111-98.

دي أيالا، ر. (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة*. (عبد الله الكيلاني، إسماعيل البرصان، مترجم). دار جامعة الملك سعود.

زارع، نسرين، والحربي، أمل. (2022). أثر اختلاف فئات الاستجابة وحجم العينة على الخصائص السيكومترية لقائمة التفكير البنائي باستخدام نموذج سلم التقدير. *المجلة السعودية للعلوم النفسية*، *21*(1)، 177 -195.

شريم، رامي. (2022). *أثر طريقة صياغة الفقرات السالبة في مقياس اتجاه على الخصائص السيكومترية للفقرات والمقياس ودقة تقديرات القدرة في ضوء اختلاف المرحلة العمرية* (رسالة دكتوراه غير منشورة). جامعة اليرموك.

المعهد التخصصي للتدريب المهني للمعلمين (2024). *النشرة الإحصائية*. دائرة شؤون المتدربين.

النمرات، راميا. (2021). *أثر طريقة صياغة الفقرات السلبية في مقياس اتجاهات على خصائصه السيكومترية*. (رسالة ماجستير غير منشورة). جامعة اليرموك.

ثانيا: المراجع الأجنبية

Alan, U., & Kabasakal, K. (2020). Effect of number of response options on the psychometric properties of Likert-type scales used with children. *Studies in Educational Evaluation*, 66, 100895.

Al-Harithi, H..(2022) The Effect of the "Neutral" Alternative in Likert Scale on The Differential Functioning and Information Function of the Scale of Attitude for Computer. (in Arabic). *Journal of Educational Sciences*, *30* (3), 245–268.

Al-Khayri, M. (2022). The effect of the "neutral" response on the psychometric properties of psychological scales and its relationship with some demographic variables (in Arabic). *Arab Journal for Scientific Publishing*, 22, 98–211.

Allahyari, E., Jafari, P., & Bagheri, Z. (2016). A simulation study to assess the effect of the number of response categories on the power of ordinal logistic regression for differential item functioning analysis in rating scales. *Computational and Mathematical Methods in Medicine*, 5080826.

Al-Nimrat, R. (2021). *The Effect of Wording Negative Items Method of the Attitudes Scale on its Psychometric Properties* (in Arabic). (Unpublished master's thesis). Yarmouk University.

Alvarez, J., Pedrosa, I., Lozano, L., Garcia Cueto, E., Cuesta Izquierdo, M., & Muniz Fernandez, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, *30*(2), 149-158.

Aybek, E., & Toraman, C. (2022). How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory. *International Journal of Assessment Tools in Education*, *9*(2), 534-547.

Baker, F., & Kim, S. (2017). *The basics of item response theory using R*. Springer.

Basman, M. (2023). A comparison of the efficacies of differential item functioning detection methods. *International Journal of Assessment Tools in Education*, *10*(1), 145–159.

Borjas, M. (2020). *An Evaluation of the Number of Response Options for Scales in Psychology*. Senior Honors Theses. University of Houston.

Brown, G., Alqassab, M., Bulut, O., & Xiao, J. (2020). Detecting Differential Item Functioning of Polytomous Items in Small Samples: Comparison of MIMIC with a Pure An-chor and MIMIC-Interaction Methods. *Psychological Test and Assessment Modeling*,*62*(4), 415-525.

Bulut, H., & Bulut, O. (2022). Item wording effects self-report measures and reading achievement: Does removing careless respondents help? *Studies in Educational Evaluation*, 72, 101-126.

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.

Cantillo, J., Martin, J., & Román, C. (2023). Understanding consumers' perceptions of aquaculture and its products in Gran Canaria Island: Does the influence of positive or negative wording matter? *Aquaculture*, 562,738754.

Castilla, D., Jaen, I., Suso-Ribera, C., Garcia-Soriano, G., Zaragoza, I., Breton-Lopez, J., & Garcia-Palacios, A. (2024). Psychometric properties of the Spanish full and short forms of the system usability scale: Detecting the effect of negatively worded items. *International Journal of Human - Computer Interaction*, *40*(15), 4145-4151.

Chaplin, T. (2015). Gender and emotion expression: A developmental contextual perspective. *Emotion Review*, *7*(1), 14-21.

De Rezende, N. A., & de Medeiros, D. D. (2022). How rating scales influence responses' reliability, extreme points, middle point and

Measurement and Evaluation in Education and Psychology, *8*(3), 321-343.

Jaradat, A. (2006). The relationship between self-esteem and irrational attitudes in university students (in Arabic). *Jordan Journal of Educational Sciences, 2* (3), 143–153.

Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert scale: Explored and explained. British *Journal of Applied Science & Technology*, 7(4), 396-403.

Kankaras, M., Capecchi, S. (2024). Neither agree nor disagree use and misuse of the neutral response category in Likert-type scales. *METRON* 83, 111–140.

Klar, A., Costello, S., Sadusky, A., & Kraska, J. (2022). Personality, culture and extreme response style: A multilevel modelling analysis. *Journal of Research in Personality*, 101, 104301.

Kolen, M., & Brennan, R. (2013). *Test equating: Methods and practices*. Springer Science & Business Media.

Kusmaryono, I., Wijayanti, D., & Maharani, H. (2022). Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. *International Journal of Educational Methodology*, *8*(4), 625-637.

Marsh, K. (2013). *The effects of item and respondent characteristics on midpoint response option endorsement: A mixed-methods study*. ProQuest LLC.

Ohiri, S., Momoh, O., & Ikeanumba, C. (2024). Differential Item Functioning Detection Methods: *International Journal of Research Publication and Reviews*, *5*(2), 1555-1564.

Orlando, M., & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, *24*(1), 50-64.

Rammstedt, B., Danner, D., & Bosnjak, M. (2017). Acquiescence response styles: A multilevel model explaining individual-level and country-level differences. *Personality and individual differences*, 107, 190-194.

Rokeman, N. (2024). Likert measurement scale in education and social sciences: explored and explained. *EDUCATUM Journal of Social Sciences*, *10*(1), 77-88.

Rosenberg, M. (1965). *Society and adolescent self-image*. Princeton, NJ: Princeton University Press.

Santoso, P., Setiawati, F., Ismail, R., & Suhariyono, S. (2023). Comparing IRT properties among different category respondent's preferences. *Journal of Business Research*, 138, 266-274.

DeVellis, R., & Thorpe, C. (2022). *Scale development: Theory and applications*. Sage publications.

Donnellan, M., & Rakhshani, A. (2023). How does the number of response options impact the psychometric properties of the Rosenberg Self-Esteem Scale? *Assessment*, *30*(6), 1737-1749.

Dorans, N., & Holland, P. (1993). *Differential item functioning. In R. L. Linn (Ed.), Educational measurement* (pp. 283–314). American Council on Education and Macmillan.

Fahrmann, K., Köhler, C., Hartig, J., & Heine, J.-H. (2022). Practical significance of item misfit and its manifestations in constructs assessed in large-scale studies. *Large-scale Assessments in Education*, *10*(7). 1-21.

Fukudome, K., & Takeda, T. (2024). The influence of cognitive reflection on consistency of responses between reversed and direct items. *Personality and Individual Differences*, 230, 112811.

Gana, K., & Broc, G. (2019). *Structural equation modeling with lavaan*. John Wiley & Sons.

Garcia-Fernandez, J., Postigo, A., Cuesta, M., Gonzalez-Nuevo, C., Menendez-Aller, A., & Garcia-Cueto, E. (2022). To be direct or not: Reversing Likert response format items. *The Spanish Journal of Psychology*, 25, e24.

Gitchel, W., Roessler, R., & Turner, R. (2011). Gender effect according to item directionality on the perceived stress scale for adults with multiple sclerosis. *Rehabilitation Counseling Bulletin*, *55*(1), 20-28.

Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer Science & Business Media.

Harzing, A., Brown, M., Koster, K., & Zhao, S. (2012). Response style differences in cross-national research: dispositional and situational determinants. *Management International Review*, 52, 341-363.

Ibrahim. A., & Bkkar, B. (2015). The Impact of the Polarity of Questionnaire Items on Factor Structure and Measurement of the Underlying Trait: The Case of the Psychological Mindedness Scale. *Journal of Educational Psychological Sciences*, *16*(3), 638-656.

Ilhan, M., & Guler, N. (2017). The Number of Response Categories and the Reverse Scored Item Problem in Likert-Type Scales: A Study with the Rasch Model. *Journal of*

(Eds.). *Differential Item Functioning* (pp. 67–113). Lawrence Erlbaum Associates.

Teresi, J., Wang, C., Kleinman, M., Jones, R., & Weiss, D. (2021). Differential item functioning analyses of the Patient-Reported Outcomes Measurement Information System (PROMIS) measures: Methods, challenges, advances, and future directions. *psychometrika*, *86*(3), 674-711.

Vigil Colet, A., Navarro González, D., & Morales Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*,*32*(1), 108-114.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Wells, C. (2021). *Assessing measurement invariance for applied research*. Cambridge University Press.

Zarea, N.& Alharbi, A. (2022). The Effect of the different Response Categories and Sample Size on psychometric properties of Constructive Thinking Inventory using Rating Scale model (in Arabic). *Saudi Journal of Psychological Sciences*, *21*(1), 177-195.

Zhang, Y., Yang, Z., & Wang, Y. (2022). The impact of extreme response style on the mean comparison of two independent samples. *SAGE Open*, *12*(3). 1-19.

numbers: a case from attitudinal measurement on physics education research. *Discover Psychology*, *3*(1), 39-51.

Setiawati, F., Nurhayati, S., & Amelia, R. (2022). Study on the Threats of Reverse-Worded Items to the Psychometric Properties of the Marital Quality Scale. *The Open Psychology Journal*, *15*(1). e2208150.

Shuraym, R. (2022). *The Effect of Negative Items Negation Methods in Attitude Scale on Psychometric Properties of Items and Scale and the Accuracy of Ability Estimates According to Respondents' Age Group* (in Arabic). (Unpublished doctoral dissertation). Yarmouk University.

Sinharay, S., & Haberman, S. (2014). How often is the misfit of item response theory models practically significant? *Educational Measurement: Issues and Practice*, *33*(1), 23–35.

Specialized Institute for Professional Training of Teachers. (2024). *Statistical bulletin*. Department of Trainees' Affairs.

Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. *International Journal of Academic Research in Management*,*8*(1) 1-10.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In Holland, P. W., & Wainer, H.