مجلة الآداب
للدراسات النفسية والتربوية

OPEN ACCESS
تاريخ الاستلام: 02/ 07/ 2026م
تاريخ القبـول: 23/ 08/ 2026م

الأداء التفاضلي للفقرة في استبانة قياس رأي الطلبة حول المناهج السريرية في جامعة حضرموت باستخدام نموذج التقدير الجزئي

أ.م.د/ محمد حاتم سعيد الدعيس*
Mohammedhatem11@gmail.com

شوقي أحمد علي حزام الدعيس**
s.alduais@hu.edu.ye

الملخص

هدفت الدراسة الحالية إلى التحقق من الخصائص السيكومترية لاستبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض بجامعة حضرموت وفقاً لنموذج التقدير الجزئي، والكشف عن ظهور مؤشر الأداء التفاضلي للفقرة (DIF) تبعاً لمتغيرَي الجنس والكلية. واستخدمت الدراسة المنهج الوصفي التحليلي، وطُبِّقت الاستبانة المكوَّنة من (24) فقرة موزعة على خمسة أبعاد على عيّنة أساسية بلغت (331) طالباً وطالبة (175 ذكراً، 156 أنثى؛ 179 من كلية الطب والعلوم الصحية، 152 من كلية التمريض)، اختيروا بالطريقة الطبقية العشوائية. وتوصلت الدراسة إلى تحقق افتراضات نموذج التقدير الجزئي، ووجود مطابقة جيدة لفقرات الاستبانة، باستثناء فقرتين تم حذفهما لا تطابقا النموذج (أصبحت الاستبانة من 22 فقرة). كما أظهرت دالة معلومات الاختبار دقة قياس مرتفعة، بمعامل ثبات هامشي تقديري (0.96≈). وعلى صعيد الأداء التفاضلي، أظهرت النتائج وجود أربع فقرات ذات أداء تفاضلي تبعًا لمتغير الجنس، في حين لم تظهر أي فقرة أداءً تفاضليًا تبعًا لمتغير الكلية، وفق معيار حجم الفرق المعتمد.

الكلمات المفتاحية: الأداء التفاضلي للفقرة؛ نموذج التقدير الجزئي؛ نظرية الاستجابة للفقرة؛ دالة المعلومات؛ استبانة قياس الرأي؛ المناهج السريرية.

* أستاذ القياس والتقويم المشارك، قسم علم النفس، كلية الآداب، جامعة تعز ، اليمن
**أستاذ القياس والتقويم المشارك، قسم العلوم النفسية والتربوية، كلية التربية ، جامعة حضرموت، اليمن

للاقتباس: الدعيس، محمد حاتم سعيد؛ الدعيس، شوقي أحمد علي حزام . (2026). الأداء التفاضلي للفقرة في استبانة قياس رأي الطلبة حول المناهج السريرية في جامعة حضرموت باستخدام نموذج التقدير الجزئي، *مجلة الآداب للدراسات النفسية والتربوية*، 8(3)، 672-710.

© نُشر هذا البحث وفقًا لشروط الرخصة Attribution 4.0 International (CC BY 4.0)، التي تسمح بنسخ البحث وتوزيعه ونقله بأي شكل من الأشكال، كما تسمح بتكييف البحث أو تحويله أو الإضافة إليه لأي غرض كان، بما في ذلك الأغراض التجارية، شريطة نسبة العمل إلى صاحبه مع بيان أي تعديلات أُجريت عليه.

(EISSN): 2708-5775 ISSN: 2707-5788
DOI: https://doi.org/10.53285/artsep.v8i3.3406

OPEN ACCESS
Received: 02-07-2026
Accepted: 23-08-2026

مجلة الآداب
للدراسات النفسية والتربوية

Differential Item Functioning in a Questionnaire Measuring Students' Perceptions of Clinical Curricula at Hadhramout University: An Application of the Partial Credit Model

Associ. Prof. Mohammed Hatem Saeed Aldoais *
Mohammedhatem11@gmail.com

Shawqi Ahmed Ali Hezam Alduais**
s.alduais@hu.edu.ye

Abstract

This study aims to examine the psychometric properties of a questionnaire designed to measure students' perceptions of the clinical curricula at the College of Medicine and Health Sciences and the College of Nursing at Hadhramout University using the Partial Credit Model (PCM). It also sought to investigate Differential Item Functioning (DIF) with respect to gender and college. A descriptive-analytical research design was employed. The questionnaire initially comprised 24 items distributed across five dimensions and was administered to a main sample of 331 students, including 175 males and 156 females, of whom 179 were enrolled in the College of Medicine and Health Sciences and 152 in the College of Nursing. Participants were selected using stratified random sampling. The findings indicated that the assumptions of the Partial Credit Model were satisfied and that the questionnaire items demonstrated good model fit, with the exception of two items that were removed due to misfit, resulting in a final 22-item questionnaire. The Test Information Function indicated a high level of measurement precision, with an estimated marginal reliability coefficient of approximately 0.96. Regarding differential item functioning, four items exhibited DIF by gender, whereas no item demonstrated DIF by college, based on the adopted effect-size criterion.

Keywords: Differential Item Functioning; Partial Credit Model; Item Response Theory; Test Information Function; Opinion Survey; Clinical Curricula.

* Associate Professor of Measurement and Evaluation, Department of Psychology, Faculty of Arts, Taiz University, Yemen.

** Associate Professor of Measurement and Evaluation, Department of Psychological and Educational Sciences, Faculty of Education, Hadhramout University, Yemen.

Cite this article as: Alduais, Mohammed Hatem Saeed. Alduais, Shawqi Ahmed Ali Hezam. (2026) Differential Item Functioning in a Questionnaire Measuring Students' Perceptions of Clinical Curricula at Hadhramout University: An Application of the Partial Credit Model. *Journal of Arts for Psychological & Educational Studies 8*(3) 672-710

© This material is published under the license of Attribution 4.0 International (CC BY 4.0), which allows the user to copy and redistribute the material in any medium or format. It also allows adapting, transforming or adding to the material for any purpose, even commercially, as long as such modifications are highlighted and the material is credited to its author.

المقدمة

يشهد التعليم العالي في برامج العلوم الصحية تطوراً متسارعاً في آليات ضمان الجودة وتقويم مخرجات التعلّم، بوصفه ركيزة أساسية لإعداد كوادر صحية مؤهَّلة قادرة على تقديم رعاية آمنة وفعّالة. ويُعد التدريب السريري مكوّنًا أساسيًا في برامج إعداد طلبة العلوم الصحية، كونه الحلقة التي تربط المعرفة النظرية بالممارسة الفعلية، وتُصقل فيها الكفايات المهنية والمهارات الإكلينيكية اللازمة لمزاولة المهنة، وهو ما يجعل من تقويمه المستمر والدقيق شرطاً جوهرياً لتحسين جودة العملية التعليمية في كليات الطب والعلوم الصحية والتمريض (Strand et al., 2013).

ونظراً لما تمثله تجربة الطالب من مصدر مباشر وأصيل للمعلومات المتعلقة بجودة التدريب السريري، فقد أصبحت استبانات قياس آراء الطلبة حول البيئة السريرية من أكثر أدوات التقويم استخداماً في مؤسسات التعليم الصحي على المستوى العالمي. ومن أبرز هذه الأدوات مقياس بيئة التعلّم السريري (CLEI) الذي طوّره(2002) Chan ، ومقياس البيئة السريرية والإشراف وتقييم المعلّم (CLES+T) الذي طوّره Saarikoski و(2002) Leino-Kilpi، إضافة إلى مقياس البيئة التعليمية الطبية (DREEM) الذي طوّره Roff وآخرون (1997). وقد اعتمدت العديد من مؤسسات التعليم الصحي هذه الأدوات لتقويم جودة البيئات التعليمية السريرية، كما أكدت ذلك دراسات حديثة، من بينها دراسة التحقق من الخصائص السيكومترية للنسخة العربية من مقياس(Alanazi et al., 2025) CLES+T ، الأمر الذي يعكس اتساع استخدام هذه الأدوات في البيئات التعليمية الصحية، بما في ذلك البيئات العربية المماثلة لبيئة الدراسة الحالية.

بيد أن أغلب هذه الأدوات بُنيت وحُلِّلت إحصائياً في إطار النظرية الكلاسيكية في القياس (Classical Test Theory)، التي ترتبط فيها الخصائص السيكومترية للأداة، مثل معاملات الصعوبة والثبات والتمييز، بخصائص العينة التي طُبِّقت عليها، مما يجعل هذه الخصائص عرضة للتغير من عينة إلى أخرى. ويُعدّ ذلك من القيود المنهجية التي سعت نظرية الاستجابة للفقرة (Item Response Theory) إلى تجاوزها من خلال نماذجها الحديثة التي تتمتع بدرجة أعلى من الاستقلال عن خصائص العينة والأداة (Hambleton & Embretson & Reise, 2000؛Swaminathan, 1985 ؛ علام، 2005). ومن بين هذه النماذج نموذج التقدير الجزئي (Partial Credit Model) الذي طوّره Masters (1982) لتحليل البيانات متعددة الفئات، ولا سيما البيانات الناتجة عن مقاييس التقدير المتدرج مثل مقياس ليكرت، ويُعد من أكثر النماذج ملاءمةً لبناء وتحليل استبانات قياس الرأي والاتجاهات.

ومن المزايا المهمة التي توفرها نظرية الاستجابة للفقرة، التي تتجاوز ما تتيحه أساليب النظرية الكلاسيكية في القياس، إمكانية الكشف عن الأداء التفاضلي للفقرات (Differential Item Functioning

[DIF]) بين المجموعات المختلفة، وهو أحد المؤشرات الرئيسة المستخدمة للتحقق من عدالة الفقرات وخلوها من التحيز. ويُعد وجود الأداء التفاضلي للفقرات تهديداً لصدق التفسيرات المستندة إلى نتائج الأداة وعدالة المقارنات بين الأفراد أو المجموعات إذا لم يُكتشف ويُعالج بصورة مناسبة (Camilli & Shepard, 1994؛ Zumbo, 1999, 2007). وتزداد أهمية هذا الإجراء عندما يكون الهدف من الأداة إجراء مقارنات بين فئات مختلفة من الطلبة، كالذكور والإناث أو طلبة الكليات والتخصصات المختلفة، كما هو الحال في الدراسة الحالية. إذ إن بقاء فقرات متحيزة ضمن الأداة قد يؤدي إلى إظهار فروق ظاهرية بين المجموعات تُعزى خطأً إلى اختلاف مستويات الرضا أو الاتجاهات، في حين تكون ناجمة - في الواقع - عن خصائص الفقرة نفسها وليس عن اختلاف حقيقي بين أفراد تلك المجموعات (Holland & Wainer, 1993).

وعلى الرغم من اتساع نطاق تطبيق نموذج التقدير الجزئي في الدراسات العربية والعالمية على مقاييس الاتجاهات والقدرات العقلية (القرشي، 2016؛ الثبيتي، 2016؛ حمدان، 2025)، وتوظيف إجراءات الكشف عن الأداء التفاضلي للفقرات في مجالات تربوية ونفسية متعددة (العناتي، 2017؛ الدوسري، 2018؛ المسكري وآخرون، 2021؛ الوابل وتعلب، 2023؛ Liu et al., 2023)، فإن مراجعة الأدبيات ذات الصلة لم تُسفر عن العثور على دراسة جمعت بين تطبيق نموذج التقدير الجزئي وتحليل استبانة لقياس آراء الطلبة حول منهج سريري. فالدراسة الأقرب من حيث المجال والتطبيق هي دراسة Alanazi et al. (2025)، التي تناولت تقويم البيئة السريرية باستخدام النسخة العربية من مقياس CLES+T، إلا أنها اعتمدت على أساليب القياس الكلاسيكي والتحليل العاملي دون الاستناد إلى نماذج نظرية الاستجابة للفقرة. وفي المقابل، تُعد دراسة Hur and Seo (2023) الأقرب من حيث المنهجية، إذ طبّقت أحد نماذج نظرية الاستجابة للفقرة على عينة من طلبة الطب، غير أنها استهدفت قياس سمات شخصية ولم تتناول تقويم المناهج أو البيئات السريرية. من ثمّ، تتحدد الفجوة البحثية الحالية في محدودية الدراسات التي وظّفت نموذج التقدير الجزئي للتحقق من الخصائص السيكومترية لاستبانات تقويم المناهج السريرية والكشف عن الأداء التفاضلي لفقراتها، وهي الفجوة التي تسعى الدراسة الحالية إلى معالجتها.

مشكلة الدراسة

تتركز مشكلة الدراسة في أن استبانة قياس رأي الطلبة حول المناهج السريرية، شأنها شأن العديد من أدوات القياس التربوي والنفسي المستخدمة في مؤسسات التعليم الصحي، تُستخدم في إجراء مقارنات بين مجموعات مختلفة من الطلبة، كالذكور والإناث وطلبة الكليات المختلفة، دون التحقق من خصائصها السيكومترية في ضوء نماذج نظرية الاستجابة للفقرة، ودون التأكد من خلوّ فقراتها من الأداء التفاضلي الذي قد يؤثر في عدالة هذه المقارنات وصحة التفسيرات المستندة إلى نتائجها. وانطلاقاً من ذلك، تتمثل مشكلة الدراسة في الإجابة عن السؤال الرئيس الآتي:

ما الخصائص السيكومترية لاستبانة قياس رأي الطلبة حول المناهج السريرية لدى طلبة كليتي الطب والعلوم الصحية والتمريض بجامعة حضرموت وفقاً لنموذج التقدير الجزئي، ومدى خلوّ فقراتها من الأداء التفاضلي تبعاً لمتغيري الجنس والكلية؟

ويتفرع عن هذا السؤال الرئيس الأسئلة الآتية:

1. هل تحقق استبانة قياس رأي الطلبة حول المناهج السريرية افتراضات نموذج التقدير الجزئي؟
2. ما مدى جودة مطابقة فقرات الاستبانة لنموذج التقدير الجزئي؟
3. ما مستوى دقة القياس الذي توفره الاستبانة وفقاً لدالة المعلومات في نموذج التقدير الجزئي؟
4. هل توجد فقرات ذات أداء تفاضلي في الاستبانة تبعاً لمتغير الجنس (ذكور/إناث)؟
5. هل توجد فقرات ذات أداء تفاضلي في الاستبانة تبعاً لمتغير الكلية (الطب والعلوم الصحية/التمريض)؟

أهمية الدراسة

تتمثل أهمية الدراسة الحالية في بُعدين رئيسين:

1- **الأهمية النظرية والمنهجية**

تُسهم الدراسة في إثراء الأدبيات العربية والأجنبية المتعلقة بتطبيقات نظرية الاستجابة للفقرة في مجال تقويم التعليم الصحي، من خلال تناول موضوع لم تحظَ تطبيقاته بالقدر الكافي من الدراسة، ويتمثل في تحليل استبانات قياس آراء الطلبة حول المناهج السريرية باستخدام نموذج التقدير الجزئي. كما تقدم الدراسة تطبيقاً منهجياً للكشف عن الأداء التفاضلي للفقرات يعتمد على الجمع بين الدلالة الإحصائية وحجم الأثر، بما يحد من المشكلات المرتبطة بالاعتماد على الدلالة الإحصائية وحدها، التي قد تتأثر بحجم العينة وتؤدي إلى نتائج مضللة في بعض الحالات .(Na et al., 2015) ومن ثمّ، يمكن أن تمثل الدراسة مرجعاً منهجياً للباحثين الراغبين في توظيف نماذج نظرية الاستجابة للفقرة في تقويم الأدوات المستخدمة في البيئات التعليمية والصحية. كما تُعدّ الدراسة – في حدود ما أسفرت عنه مراجعة الأدبيات ذات الصلة – من أوائل الدراسات التي تجمع بين تطبيق نموذج التقدير الجزئي وتحليل الأداء التفاضلي للفقرات على استبانة مخصصة لتقويم المناهج السريرية في مؤسسات التعليم الصحي، الأمر الذي يعزز إسهامها في تطوير المعرفة المنهجية في هذا المجال.

2- **الأهمية التطبيقية والعملية:** توفر الدراسة أدلة سيكومترية حول جودة استبانة قياس رأي الطلبة بشأن المناهج السريرية ومدى ملاءمتها للاستخدام في بيئات التعليم الصحي، من خلال التحقق من مطابقة فقراتها لنموذج التقدير الجزئي، وتقدير دقة القياس التي توفرها، والكشف عن أي أداء تفاضلي محتمل بين الجنسين أو بين طلبة كليتي الطب والعلوم الصحية والتمريض. ومن المتوقع أن

تسهم نتائج الدراسة في دعم متخذي القرار بجامعة حضرموت والمؤسسات التعليمية الصحية المماثلة بمعلومات موضوعية تساعد على تعزيز جودة عمليات التقويم الطلابي، وتحسين تفسير نتائج الاستبانة، واتخاذ قرارات مبنية على الأدلة بشأن تطوير الأداة أو الاستمرار في استخدامها بوصفها أداة عادلة وموثوقة في تقويم تجربة التعلّم السريري.

أهداف الدراسة

تسعى الدراسة الحالية إلى تحقيق الأهداف الآتية:

1- التحقق من مدى توفر افتراضات نموذج التقدير الجزئي في استبانة قياس رأي الطلبة حول المناهج السريرية لدى طلبة كليتي الطب والعلوم الصحية والتمريض بجامعة حضرموت.
2- تحديد مدى جودة مطابقة فقرات الاستبانة لنموذج التقدير الجزئي.
3- تقدير مستوى دقة القياس التي توفرها الاستبانة في ضوء دالة المعلومات والخطأ المعياري للقياس وفق نموذج التقدير الجزئي.
4- الكشف عن وجود فقرات ذات أداء تفاضلي في الاستبانة تبعاً لمتغير الجنس (ذكور/إناث).
5- الكشف عن وجود فقرات ذات أداء تفاضلي في الاستبانة تبعاً لمتغير الكلية (الطب والعلوم الصحية/التمريض).

مصطلحات الدراسة:

الأداء التفاضلي للفقرات:(Differential Item Functioning – DIF)

يُعرَّف بأنه الحالة التي يستجيب فيها أفراد من مجموعتين مختلفتين بشكل مختلف لفقرة معينة، رغم تساويهم في مستوى السمة الكامنة التي تقيسها الأداة (Camilli & Shepard, 1994). وتُعرّفه الدراسة إجرائياً بأنه: مقدار الفرق في معلم صعوبة الفقرة بين المجموعتين المقارَنتين (الذكور/الإناث أو طلبة كليتي الطب والعلوم الصحية/التمريض)، بحيث يُعدّ هذا الفرق دالاً عملياً إذا بلغ أو تجاوز (|0.50|) لوجيت بوصفه حجم أثر وفق معيار(1999) Keeves & Alagumalai ، دون اشتراط بلوغ الدلالة الإحصائية لاختبار z، انطلاقاً من أن الاعتماد عليها وحدها قد يُفضي إلى نتائج مضللة تبعاً لحجم العينة (Na et al., 2015).

نموذج التقدير الجزئي (Partial Credit Model – PCM):

أحد نماذج نظرية راش المخصصة لتحليل البيانات متعددة الفئات، إذ يفترض أن احتمال انتقال الفرد بين فئات الاستجابة يتحدد بالفرق بين مستوى قدرته الكامنة ومعالم عتبات الفقرات. ويُتيح هذا النموذج تقدير عتبات مستقلة لكل فقرة بما يتناسب مع طبيعة مقياس التقدير المتدرج (Masters, 1982).

وتُطبّق الدراسة هذا النموذج باستخدام أحد برمجيات القياس المتخصصة في تحليل نماذج الاستجابة للفقرة.

دالة المعلومات (Item Information Function):

دالة في إطار نظرية الاستجابة للفقرة تعبّر عن مقدار الدقة التي توفرها الفقرة أو الاختبار عبر مستويات السمة الكامنة، وترتبط عكسياً بالخطأ المعياري الشرطي للقياس؛ فكلما زادت قيمة المعلومات انخفض الخطأ المعياري (Lord, 1980). وتُستخدم في هذه الدراسة لتقدير دقة القياس الكلية للاستبانة عبر نموذج التقدير الجزئي.

استبانة قياس رأي الطلبة حول المناهج السريرية:

أداة قياس مُقنّنة أعدّها مركز التطوير الأكاديمي وضمان الجودة بجامعة حضرموت، تهدف إلى قياس آراء طلبة كليتي الطب والعلوم الصحية والتمريض حول جودة المناهج السريرية وفاعليتها. وتتكون في صورتها النهائية من (24) فقرة موزعة على خمسة أبعاد رئيسة هي: محتوى المنهج السريري، وطرائق التدريس والتعلم السريري، والبيئة والموارد السريرية، والتقويم والامتحانات السريرية، والرضا العام عن تجربة التعلم السريري، ويُجاب عنها وفق مقياس ليكرت الخماسي.

المناهج السريرية:

المقررات والأنشطة التعليمية التطبيقية التي يتعرض لها طلبة كليتي الطب والعلوم الصحية والتمريض في البيئات السريرية الواقعية (مثل المستشفيات والمراكز الصحية التدريبية)، بهدف تنمية الكفايات المهنية والمهارات الإكلينيكية اللازمة لممارسة المهنة.

حدود الدراسة

تتحدد إمكانية تعميم نتائج الدراسة الحالية بالحدود الآتية:

الحدود الموضوعية :اقتصرت الدراسة على فحص افتراضات نموذج التقدير الجزئي، وجودة مطابقة فقرات استبانة قياس رأي الطلبة حول المناهج السريرية، وتقدير دالة المعلومات، والكشف عن الأداء التفاضلي للفقرات تبعاً لمتغيري الجنس والكلية، دون التطرق إلى متغيرات تصنيفية أخرى مثل المستوى الدراسي أو المعدل التراكمي.

الحدود المكانية :طُبقت الدراسة على طلبة كليتي الطب والعلوم الصحية والتمريض بجامعة حضرموت.

الحدود الزمانية :الفصل الدراسي الأول من العام الجامعي 2025–2026م.

الحدود البشرية :عينة الدراسة الأساسية المكونة من (331) طالباً وطالبة من الكليتين، الذين طُبقت عليهم الاستبانة خلال الفترة الزمنية المحددة.

الحدود المنهجية: اعتمدت الدراسة نموذج التقدير الجزئي (Partial Credit Model) في تحليل بيانات الاستبانة، واقتصر الكشف عن الأداء التفاضلي للفقرات تبعًا لمتغيري الجنس والكلية على طريقة الفرق (Difference Method)، وذلك بحساب الفرق بين تقديرات الفقرة في مجموعتي المقارنة، واعتماد القيمة المطلقة للفرق معيارًا للحكم على وجود الأداء التفاضلي، بحيث تُعد الفقرة ذات أداء تفاضلي عندما تبلغ قيمة الفرق المطلقة 0.50 لوجيت فأكثر، وفق المحك المعتمد في الدراسة.

الإطار النظري

أولاً: من النظرية الكلاسيكية إلى نظرية الاستجابة للفقرة

تقوم النظرية الكلاسيكية في القياس (Test Theory Classical) على افتراض أن الدرجة الملاحظة لمحصلة جمعية بسيطة لفقرات الأداة، وهو ما يجعل خصائصها القياسية (كالصعوبة والثبات) مرتبطة بخصائص العيّنة المُطبَّق عليها الاختبار، ومتأثرة بتغيّرها من عيّنة إلى أخرى. وقد جاءت نظرية الاستجابة للفقرة (IRT –Theory Item Response) لتُجاوز هذا القيد، عبر نماذج رياضية تربط احتمال استجابة الفرد على الفقرة بمستوى سمته الكامنة (θ) ومعالم الفقرة ذاتها، بما يجعل معالم الفقرات والأفراد مستقلة عن العيّنة نظرياً، وقابلة للمقارنة المباشرة بين مجموعات مختلفة من الأفراد (Swaminathan, 1985; Embretson & Reise, 2000 & Hambleton؛ علام، 2005). وتُعدّ هذه الخاصية الأساس النظري الذي تقوم عليه إمكانية الكشف عن الأداء التفاضلي للفقرة بين المجموعات، وهو محور الدراسة الحالية.

ثانياً: نموذج التقدير الجزئي وأساس اختياره

يُعد نموذج التقدير الجزئي (PCM –Partial Credit Model) الذي طوّره Masters (1982) أحد نماذج راش للاستجابات متعددة الفئات، ويسمح بتقدير عتبات انتقال خاصة بكل فقرة. في المقابل، يفترض نموذج سلم التقدير (RSM –Rating Scale Model) أن تكون بنية عتبات فئات الاستجابة مشتركة بين فقرات الأداة (Andrich, 1978؛ Wright & Masters, 1982؛ Masters, 1982). وبما أن الاستبانة الحالية تستخدم خمس فئات استجابة موحدة لجميع الفقرات، فإن نموذج سلم التقدير يمثل الاختيار المتسق مبدئيًا مع تصميم الأداة.

إلا أن توحيد فئات الاستجابة في صورة الأداة لا يقتضي بالضرورة أن تعمل عتبات هذه الفئات بالطريقة نفسها عبر جميع الفقرات؛ إذ يمكن المفاضلة بين نموذج سلم التقدير ونموذج التقدير الجزئي في ضوء مدى ملاءمة كل منهما للبيانات. لذلك، تم اختبار بيانات الدراسة وفق النموذجين، ومقارنة ملاءمتهما في ضوء مؤشرات المطابقة الناتجة عن التحليل. وقد أظهر نموذج التقدير الجزئي ملاءمة أفضل للبيانات؛ إذ بلغ عدد الفقرات غير الملائمة فقرتين فقط، مقابل 12 فقرة في نموذج سلم التقدير. وبناءً على ذلك،

اعتمدت الدراسة نموذج التقدير الجزئي في التحليل النهائي، مع بقاء فئات الاستجابة الخماسية الموحدة كما صُممت في الأداة.

ولضمان سلامة تطبيق النموذج وتفسير معالمه، تم التحقق من الافتراضات الأساسية للقياس، وفي مقدمتها أحادية البعد (Unidimensionality)، التي تفترض أن الاستجابات تستند إلى سمة كامنة مهيمنة، والاستقلال الموضعي (endenceLocal Indep)، الذي يعني أن استجابة الفرد على إحدى الفقرات لا تعتمد على استجابته على فقرة أخرى بعد ضبط أثر السمة الكامنة، فضلًا عن التحرر من عامل السرعة. كما تم فحص مدى اتساق البيانات مع الشكل المتوقع للعلاقة بين السمة الكامنة واحتمالات الاستجابة وفق النموذج (ambleton & Swaminathan, 1985H؛ Bond & Fox, 2015). وتكتسب هذه الافتراضات أهمية خاصة في الدراسة الحالية؛ لأنها تمثل الأساس الذي تستند إليه تقديرات النموذج المستخدمة لاحقًا في تحليل جودة المطابقة والكشف عن الأداء التفاضلي للفقرات.

ثالثاً: دقة القياس في نظرية الاستجابة للفقرة: دالة المعلومات

تُقاس دقة القياس في نظرية الاستجابة للفقرة من خلال دالة المعلومات، التي توضح مقدار المعلومات التي يوفرها الاختبار عند مستويات السمة الكامنة المختلفة، ويرتبط بها الخطأ المعياري الشرطي للقياس ارتباطًا عكسيًا؛ إذ تزداد دقة القياس بزيادة دالة المعلومات وينخفض الخطأ المعياري للقياس (Embretson & Reise, 2000)

ولتقدير مستويات السمة الكامنة لأفراد العينة، تبنّت الدراسة طريقة الإمكان الأعظم (Maximum Likelihood Estimation – MLE)، باستخدام برنامج PARSCALE 4.1، إذ تُقدَّر قيمة θ للفرد وفق نمط استجاباته على فقرات الأداة .(Baker & Kim, 2004) وبناءً على تقديرات السمة الكامنة ودالة المعلومات الناتجة عن النموذج، تم تحديد مستويات السمة التي تحقق عندها الأداة أعلى دقة في القياس وأقل خطأ معياري.

رابعاً: الأداء التفاضلي للفقرة: المفهوم والتفسير

يُعرَّف الأداء التفاضلي للفقرة (Differential Item Functioning – DIF) بأنه الحالة التي يستجيب فيها أفراد من مجموعتين مختلفتين بشكل مختلف لفقرة معينة رغم تساويهم في مستوى السمة الكامنة (Camilli & Shepard, 1994). ومع ذلك، تؤكد الأدبيات السيكومترية الحديثة أن وجود DIF لا يُعد دليلاً مباشراً على تحيز الفقرة (Item Bias)، إذ يشير Zumbo (1999, 2007) إلى أن الفقرة تُعد متحيزة فقط عندما يكون مصدر الفرق بين المجموعات غير مرتبط بالبناء المقاس(Construct-Irrelevant) ، بينما لا يُعد الفرق مؤشراً على التحيز إذا كان ناتجاً عن اختلافات مشروعة مرتبطة بالبناء. وفي السياق نفسه، يوضح

Holland & Wainer (1993)أن الأداء التفاضلي يمثل مؤشراً إحصائياً أولياً يحتاج إلى تفسير سياقي ولا يُستخدم وحده للحكم النهائي على جودة الفقرة.

وقد تطورت أساليب الكشف عن الأداء التفاضلي من الطرق التقليدية القائمة على النظرية الكلاسيكية في القياس إلى الأساليب المعتمدة على نماذج نظرية الاستجابة للفقرة، وتحديدًا ضمن نماذج راش الترتيبية للفقرات متعددة الفئات. وتُعَدّ أبرز هذه الأساليب: اختبار مربع كاي للورد (Lord's Chi-Square) الذي يقارن معلمات الفقرة المقدَّرة بين المجموعتين إحصائيًا (Lord, 1980)؛ ومقاييس المساحة لراجو (Raju's Area Measures) التي تقيس المساحة الموقَّعة أو غير الموقَّعة بين منحنيي خاصية الفقرة للمجموعتين، التي طُوِّرت لاحقًا ضمن إطار الأداء التفاضلي للفقرات والاختبار (DFIT) ليشمل الفقرات متعددة الفئات (Raju et al., 1995)؛ واختبار نسبة الإمكان (Likelihood Ratio Test) الذي يقارن مطابقة نموذج يسمح باختلاف معلمات الفقرة بين المجموعتين بنموذج يفرض تساويها (Thissen et al., 1988؛ Cohen, 1998 & Kim)؛ إضافة إلى أسلوب Poly-SIBTEST غير المعلمي (Chang et al., 1996). وقد بيّنت دراسة مقارنة أجراها(1995) Kim & Cohen تقاربًا كبيرًا في النتائج بين هذه الأساليب الثلاثة الأولى عند تطبيقها على البيانات نفسها.

وفي هذا السياق، تتبنى الدراسة الحالية أسلوب الفرق بين معلمات الصعوبة (الموقع) بين المجموعات، وهو أسلوب ينفرد بملاءمته الخاصة لنماذج عائلة راش تحديدًا (كنموذج التقدير الجزئي المستخدم في الدراسة)؛ ذلك أن معلم الصعوبة في هذه النماذج معلمٌ وحيد ومُعرَّف بمعزل عن معلم التمييز (الذي يُفترض تساويه لجميع الفقرات وفق شرط نموذج راش)، ما يجعل الفرق بين قيمتيه المعايَرتين لكل مجموعة مؤشرًا مباشرًا وواضح التفسير على الانزياح في أداء الفقرة، خلافًا لنماذج الاستجابة الأخرى (كالنموذج ثنائي أو ثلاثي المعلمة) التي يتداخل فيها أثرا الصعوبة والتمييز فيصعب معه العزل بالطريقة نفسها. ويقترح Keeves & Alagumalai (1999) قيمة ($|\Delta b| \geq 0.50$) معياراً وحيداً للحكم على وجود الأداء التفاضلي، دون الاشتراط المسبق لبلوغ الدلالة الإحصائية؛ إذ قد يُفضي الاعتماد على الدلالة الإحصائية وحدها إلى نتائج مضللة تبعاً لحجم العينة (Na et al., 2015). وتتبنى الدراسة الحالية هذا المعيار في الإجابة عن سؤالها الرابع والخامس، مع التأكيد على أن الكشف عن الأداء التفاضلي للفقرات (DIF) يُستخدم بوصفه مؤشرًا على احتمال وجود تحيز في الفقرة بين المجموعات، ولا يُعد بديلًا عن التحقق الشامل من صلاحية الفقرات وملاءمتها للغرض الذي أُعدت من أجله. كما تسترشد الدراسة بالأدبيات السابقة التي طبقت تحليلات الأداء التفاضلي للفقرات (DIF) في سياقات تربوية وصحية متنوعة (العناتي، 2017؛ الدوسري، 2018؛ المسكري وآخرون، 2021؛ الوابل وتعلب، 2023).

خامساً: تقويم بيئة التعلّم السريري واستبانات قياس رأي الطلبة

يُعد تقويم بيئة التعلّم السريري من المجالات المحورية في التعليم الصحي، نظراً لدوره المباشر في دعم اكتساب الكفايات المهنية لدى طلبة الطب والعلوم الصحية والتمريض. وقد طُورت عدة أدوات عالمية لقياس هذا المجال، من أبرزها مقياس بيئة التعلّم السريري (CLEI) الذي طوّره Chan (2002) ، ومقياس البيئة السريرية والإشراف وتقييم المعلم (CLES+T) الذي طوّره Saarikoski & Leino-Kilpi (2002)، إضافة إلى مقياس البيئة التعليمية الطبية (DREEM) الذي طوّره Roff et al. (1997). وغالباً ما بُنيت هذه الأدوات في مراحلها الأولى بالاعتماد على الأساليب الكلاسيكية في القياس أو التحليل العاملي الاستكشافي.

وتسعى الدراسة الحالية إلى إعادة فحص خصائص إحدى هذه الاستبانات باستخدام نموذج التقدير الجزئي من نظرية الاستجابة للفقرة، بما يتيح فهماً أدق لجودة فقراتها وتكافؤ أدائها القياسي بين المجموعات المختلفة، خصوصاً الذكور والإناث وطلبة كليتي الطب والعلوم الصحية والتمريض.

الدراسات السابقة

نتناول في هذا الجزء الدراسات السابقة ذات الصلة المباشرة بموضوع الدراسة مصنفة في محورين: المحور الأول دراسات الأداء التفاضلي للفقرة (DIF)، والمحور الثاني دراسات نموذج التقدير الجزئي (PCM)، وقد رتبت هذه الدراسات ضمن كل محور ترتيباً زمنياً تصاعدياً.

المحور الأول: دراسات الأداء التفاضلي للفقرة (DIF)

هدفت دراسة **Boyd et al. (2014)** إلى تطبيق نموذج راش متعدد الأوجه لسلّم التقدير للكشف عن الأداء التفاضلي حسب جنس المدرّب ومستوى الفريق المُدرَّب، على عيّنة من المدرّبين الرياضيين الأمريكيين باستخدام مقياس الكفاءة التدريبية (CES). وأظهرت النتائج أن غالبية فقرات المقياس لم تُظهر أداءً تفاضلياً حسب جنس المدرّب، باستثناء فقرتين فقط من بُعدي التحفيز وبناء الشخصية، وهو ما عُدَّ دليلاً داعماً لإمكانية تعميم المقياس واستخدامه للمقارنة بين الجنسين رغم وجود استثناءات محدودة.

دراسة **Na et al. (2015)** التي هدفت إلى بناء أداة قياس الأمن الغذائي الأسري (FAST) باستخدام نموذج التقدير الجزئي، تكونت العينة من (11992) أسرة في بنغلاديش، بهدف فحص شدة فئات الاستجابة والأداء التفاضلي للفقرة تبعاً لمستوى الإلمام بالقراءة، والعمر، وملكية الأرض، وحجم الأسرة. وأظهرت النتائج تطابق جميع الفقرات مع نموذج التقدير الجزئي، كما أظهرت ثباتاً مرتفعاً بمعامل ألفا كرونباخ بلغ (0.85). كما أظهرت جميع الفقرات تقريباً أداءً تفاضلياً دالًّا إحصائياً عبر المتغيرات المفحوصة نتيجة لحجم العيّنة الكبير جداً، إلا أن حجم تأثير هذا الأداء التفاضلي في جميع الحالات كان أقل من 1% من التباين المُفسَّر، وهو ما دفع الباحثين إلى الحكم بخلوّ المقياس من الأداء التفاضلي ذي الدلالة العملية

رغم دلالته الإحصائية. وتُعدّ هذه النتيجة سابقة منهجية مباشرة تدعم اعتماد الدراسة الحالية على معيار حجم التأثير ($|b|\leq 0.50\Delta$).

كما تناولت دراسة **العناتي (2017)** الأداء التفاضلي لفقرات اختبار الذكاء غير اللفظي (4-TONI) وفقاً لثلاثة متغيرات هي: جنس الطالب، ودخل الأسرة، والمستوى التعليمي للأم، تكونت أداة الدراسة من (60) فقرة تمثل الصورة الرابعة للاختبار، وطبقت على عيّنة بلغت (414) طالبا وطالبة في الصف السابع الأساسي في الأردن. وتوصلت الدراسة إلى وجود فقرتين ذات أداء تفاضلي دالّ حسب الجنس، إضافة إلى وجود فقرتين ذات أداء تفاضلي مرتبط بالمتغيرات الاجتماعية والاقتصادية المختلفة، وهو ما يشير إلى أن الأداء التفاضلي قد ينتج عن تقاطع عدة عوامل سياقية لا الجنس فقط.

كما هدفت دراسة **الدوسري (2018)** إلى الكشف عن الأداء التفاضلي لفقرات اختبار في الرياضيات باستخدام طريقتين إحصائيتين هما مانتل-هانزل (Haenszel-Mantel) والانحدار اللوجستي، وذلك بهدف المقارنة بين فاعلية الطريقتين في الكشف عن الفقرات المتحيّزة. وطبق اختبارًا تحصيليًا مكونا من 34 فقرة من نوع المقال على عيّنة عشوائية مكونة من 4200 من طلبة الرياضيات، بينت النتائج وجود أداء تفاضلي لبعض الفقرات على مستوى جنس الطالب، وكذلك على مستوى الصف الدراسي؛ كما أظهرت النتائج أن الأداء التفاضلي كان غير منتظم في الكثير من الفقرات، مع وجود أداء منتظم لفقرات أخرى، وذلك على مستوى الصف الدراسي أكثر منه على مستوى جنس الطالب.

كما أجرى **المسكري وآخرون (2021)** دراسة هدفت إلى الكشف عن الأداء التفاضلي لفقرات اختبار القدرة اللفظية في مقياس الخليج للقدرات العقلية المتعددة، تبعاً لمتغيري الجنس والدولة، على عيّنة من طلبة دول مجلس التعاون الخليجي. تكونت أداة الدراسة من (30) فقرة، وتكونت عينة التحليل من (4280) طالباً وطالبة، وأظهرت النتائج وجود فقرات ذات أداء تفاضلي دالّ حسب كلٍّ من الجنس والدولة، وهو ما يُعدّ ذا صلة مباشرة بالسياق الخليجي الذي تُجرى فيه الدراسة، ويدعم أهمية فحص الأداء التفاضلي قبل استخدام أي مقياس للمقارنة بين مجموعات ثقافية أو جندرية مختلفة في هذه المنطقة.

كما أجرى **Fan et al. (2022)** دراسة استخدم فيها فريق بحثي نموذج راش لفحص الخصائص القياسية ومؤشرات الأداء التفاضلي لمقياس الوصمة الذاتية المختصر، تكونت العينة من (612) فرداً من ثلاث مجموعات تشخيصية نفسية مختلفة (اضطرابات تعاطي المواد، فرط الحركة وتشتت الانتباه، الفصام). وتوصلت الدراسة إلى تحقق أحادية بالبعد، وتدرج فئات الاستجابة الرباعية بشكل منطقي عبر متّصلة الوصمة الذاتية، كما أظهرت فقرتين فقط من أصل تسع فقرات أداءً تفاضلياً دالاً عبر المجموعات التشخيصية الثلاث، وهو ما عُدّ مؤشراً على خصائص قياسية جيدة بصفة عامة للمقياس عبر هذه المجموعات.

دراسة **الو ابل وتعلب (2023)** التي هدفت إلى الكشف عن الأداء التفاضلي لفقرات اختبار الذكاء الثلاثي لستيرنبرغ تبعاً لمتغيري الجنس والتخصص الأكاديمي، بلغت العينة (201) طالباً وطالبة بجامعة القصيم (99) ذكور، و(102) إناث، و(116) طالباً وطالبة من تخصصات علمية، و(97) طالباً وطالبة من تخصصات أدبية، تم اختيارهم بطريقة عشوائية عنقودية. واستخدمت الدراسة طريقة مانتل-هانزل وحساب نسبة الأرجحية المشتركة، وتوصلت الدراسة إلى وجود (17) فقرة ذات أداء تفاضلي حسب الجنس، (6) فقرات لصالح الذكور و(11) فقرة لصالح الإناث، كما توصلت الدراسة إلى وجود (20) فقرة ذات أداء تفاضلي حسب التخصص، (7) فقرات لصالح الذكور في التخصصات العلمية، و (13) فقرة لصالح الإناث في التخصصات الأدبية.

كما طبقت دراسة **Liu et al. (2023)** نظرية الاستجابة للفقرة متعددة الأبعاد لفحص الأداء التفاضلي حسب الجنس على مقياس الرضا عن العلاقات ذي البعدين الإيجابي والسلبي، تكونت العينة من (511) طالباً جامعياً صينياً. وتوصلت الدراسة إلى وجود (5) فقرات من أصل (14) أظهرت أداءً تفاضلياً دالاً حسب الجنس، أثّر على صدق البناء العاملي للمقياس، وهو ما يُبرز أن مقاييس الرضا -على وجه التحديد- معرّضة بشكل خاص لظهور أداء تفاضلي تبعاً للجنس، ويُعزز أهمية إجراء هذا النوع من التحليل قبل استخدام أي مقياس رضا للمقارنة بين الذكور والإناث. تماماً كما هو الحال في الدراسة الحالية.

المحور الثاني: دراسات نموذج التقدير الجزئي (PCM)

قدّم **Leis et al. (2015)** تطبيقاً مباشراً لنموذج التقدير الجزئي على مقياس مشاركة الطلبة في الرياضيات، إذ استُخدم النموذج لتقييم جودة فقرات المقياس من حيث الصعوبة وترتيب فئات الاستجابة، وتوصلت الدراسة إلى أن النموذج وفّر معلومات تشخيصية أدقّ من المعالجة الكلاسيكية حول أيٍّ من فئات الاستجابة الخماسية تحتاج إلى دمج أو إعادة صياغة، فيما يُعدّ تطبيقاً مرجعياً لكيفية استخدام النموذج في تقييم جودة فقرات مقاييس الرأي والمشاركة الأكاديمية بصورة عامة.

كما هدفت دراسة **القرشي (2016)** إلى استخدام نموذج التقدير الجزئي في انتقاء فقرات مقياس اتجاهات طلبة الدراسات العليا نحو برنامج SPSS في كلية التربية في منطقة مكة المكرمة، وقد تم بناء مقياس اتجاهات نحو برنامج SPSS، تكون المقياس من (65) فقرة حسب تدريج ليكرت خماسي، تكونت العينة من (195) طالباً وطالبة من كليات التربية بجامعات الملك عبد العزيز وأم القرى والطائف، وتوصلت الدراسة إلى أن المقياس يقيس سمة واحدة، واستقلال استجابة كل فقرة عن نتائج الفقرات الأخرى، وأن المقياس متحرر من عامل السرعة، وأن هناك مطابقة بين فقرات مقياس الاتجاهات مع نموذج التقدير الجزئي. وتعد هذه الدراسة من الدراسات المنهجية المهمة لتطبيق النموذج على مقاييس الاتجاه نحو أدوات

وبرامج أكاديمية في الجامعات السعودية، بما يماثل منهجياً تطبيقه في الدراسة الحالية على استبانة رأي حول منهج سريري.

كما هدفت دراسة **الثبيتي (2016)** تدريج فقرات مقياس اتجاهات طلبة جامعة أم القرى نحو الإشراف الأكاديمي وفق نظرية الاستجابة للمفردة، وتم تطبيق مقياس اتجاهات طلبة الدراسات العليا نحو الإشراف الأكاديمي في بعض الجامعات السعودية على عينة مكونة من (200) طالب وطالبة من طلبة جامعة أم القرى في مرحلة الماجستير والدكتوراة، وتوصلت الدراسة إلى مطابقة (42) فقرة لافتراضات نموذج التقدير الجزئي كأحد نماذج النظرية الحديثة الذي يناسب الاستجابات المتعددة، كما أظهر المقياس في صورته النهائية، المكوَّنة من (42) فقرة، خصائص سيكومترية مناسبة، بما يدعم تفوق نموذج التقدير الجزئي على النظرية الكلاسيكية البحتة عند تحليل مقاييس الاتجاه الأكاديمي.

دراسة **Hur & Seo (2023)** التي هدفت إلى تطوير مقياس لقياس السمات الشخصية لدى طلبة الطب في كوريا الجنوبية. تكونت أداة الدراسة من (160) فقرة، طبقت على عينة بلغت (856) طالباً من خمس كليات طب، واستخدم *نموذج التقدير الجزئي ضمن نظرية الاستجابة للفقرة متعددة الفئات* لفحص جودة التطابق لكل فقرة. وخلصت الدراسة إلى حذف (25) فقرة باستخدام النظرية الكلاسيكية، وحذف (17) فقرة، باستخدام نموذج التقدير الجزئي، مما يوضح قدرة النموذج على الكشف عن فقرات ضعيفة لم تكشفها الطرق الكلاسيكية، وانتهت الدراسة بمقياس نهائي مكوّن من (79) فقرة تحقق منها بالتحليل العاملي التوكيدي.

كما أجرى **حمدان (2025)** دراسةً هدفت إلى مقارنة بين نموذج التقدير الجزئي ونموذج سلَّم التقدير لمقياس الكفاءة الذاتية في الإدارة الصفية على الطلبة المعلمين. تمثلت أداةُ الدراسة في مقياس الكفاءة الذاتية في الإدارة الصفية للطالب المعلم، وقد طبق المقياس على عينة بلغت (741) من طلبة كلية التربية بجامعة أسيوط. وتوصلت الدراسة إلى مطابقة الفقرات لنموذج سلم التقدير، ونموذج التقدير الجزئي؛ إذ كانت قيم المطابقة الداخلية والخارجية للفقرات تقع ضمن المدى المقبول، وتراوحت بين (0.5 و1.5)، كما أن قيم معاملات الثبات كانت مقبولة، كما توصلت الدراسة إلى وجود فرق دال إحصائيًا عند مستوى (0.01) بين النموذجين، لصالح نموذج التقدير الجزئي، مما يدل على أن نموذجَ التقدير الجزئي هو الأكثر دقة ومناسَبةً لمقياس الكفاءة الذاتية في الإدارة الصفية.

كما أجرى **Alanazi et al. (2025)** دراسة هدفت إلى التحقق من صدق وثبات النسخة العربية السعودية لمقياس بيئة التعلّم السريري والإشراف وتقييم المُعلّم السريري (SA-CLES+T)، وتكونت عينة الدراسة من (280) من طلبة وأخصائي التمريض المتدربين سريرياً (متدربون ما قبل التدريب الداخلي ومتدربون داخليون). واستُخدم التحليل العاملي الاستطلاعي بطريقة المحاور الأساسية (Principal Axis

Factoring) للتحقق من البناء العاملي للمقياس. وترجع أهمية هذه الدراسة كونها تتناول *بيئة الدراسة الحالية نفسها* (التدريب السريري) وأداة من الفئة نفسها (استبانة رأي حول بيئة التعلّم السريري)، إلا أنها اعتمدت على أساليب القياس الكلاسيكي/التحليل العاملي لا على نماذج نظرية الاستجابة للفقرة، وهو ما يُبرز الفجوة المنهجية التي تسعى الدراسة الحالية لمعالجتها باستخدام نموذج التقدير الجزئي بدلاً من ذلك.

مناقشة الدراسات السابقة

أوجه الاتفاق:

تتفق الدراسات السابقة على ثلاث نتائج رئيسة. أولاً، أن ظاهرة الأداء التفاضلي للفقرات تُعد ظاهرة شائعة في مختلف مجالات القياس، وقد ظهرت في اختبارات القدرات والذكاء (الوابل وتعلب، 2023؛ العناتي، 2017؛ المسكري وآخرون، 2021)، وفي مقاييس الرضا والكفاءة الذاتية (Liu et al., 2023؛ Boyd et al., 2014)، وكذلك في المقاييس النفسية متدرجة الاستجابة (Na et al., 2015؛ Fan et al., 2022)، بغض النظر عن طبيعة المتغيرات التصنيفية المستخدمة.

ثانياً، تتفق الدراسات التي وظّفت نموذج التقدير الجزئي(Hur & Seo, 2023) ؛ القرشي، 2016؛ الثبيتي، 2016؛ حمدان، 2025) على أن هذا النموذج يمتلك قدرة تشخيصية أعلى مقارنة بالأساليب الكلاسيكية في الكشف عن الفقرات غير الملائمة أو غير المتطابقة.

ثالثاً، تشير عدد من الدراسات (Na et al., 2015؛ Fan et al., 2022؛ Boyd et al., 2014) إلى أن ظهور الأداء التفاضلي غالباً ما يكون محدوداً في عدد قليل من الفقرات، وليس على مستوى المقياس ككل، بما يدعم إمكانية الإبقاء على الأداة مع إجراء تعديلات انتقائية عند الحاجة.

نقاط الاختلاف بين الدراسات

على الرغم من أوجه الاتفاق التي أُشير إليها سابقاً، تُظهر الدراسات تبايناً ملحوظاً في حجم الأداء التفاضلي للفقرات ومعايير الحكم عليه. فقد توصلت دراسة الوابل وتعلب (2023) إلى نسب مرتفعة نسبياً من الفقرات ذات الأداء التفاضلي (قرابة 57% وفق متغير الجنس، و67% وفق متغير التخصص)، في حين سجلت دراسة العناتي (2017) نسباً منخفضة جداً (3.3%) على أداة مختلفة وضمن عينة أصغر سناً. كما أشارت دراسة Fan et al. (2022) إلى نسبة متوسطة من الفقرات ذات الأداء التفاضلي (22%) في سياق إكلينيكي مختلف.

ويبرز التباين بشكل أوضح في دراسة(2015) Na et al. ، التي أظهرت دلالة إحصائية للأداء التفاضلي في معظم الفقرات نتيجة لكبر حجم العينة (11,992) أسرة، في حين كان حجم الأثر الفعلي محدوداً جداً(1% >) ، وهو ما يسلط الضوء على الفجوة بين الدلالة الإحصائية والأهمية العملية، وهي مسألة لم تتناولها غالبية الدراسات السابقة بالوضوح نفسه. ويمثل هذا التباين أساساً منهجياً تعتمد عليه

الدراسة الحالية من خلال تبنّي معيار حجم الأثر ($|\Delta b| \geq 0.50$) وفق ما اقترحه Keeves & Alagumalai (1999)معياراً وحيداً للحكم على وجود الأداء التفاضلي، دون الاشتراط المسبق لبلوغ الدلالة الإحصائية، انطلاقاً من أن الاعتماد على الدلالة الإحصائية وحدها في مثل هذه السياقات قد يُفضي إلى نتائج مضللة تبعاً لحجم العينة.

كما تختلف الدراسات أيضاً في المقارنات بين النماذج الإحصائية المستخدمة؛ إذ توصلت دراسة حمدان (2025) إلى تفوق نموذج التقدير الجزئي على نموذج سلّم التقدير من حيث الدقة والملاءمة، في حين اقتصرت دراسات أخرى مثل القرشي (2016) و(2015) .Na et al على فحص نموذج واحد دون إجراء مقارنة نماذج بديلة. ويشير ذلك إلى حاجة البحث - في هذا المجال - إلى مزيد من الدراسات المقارنة بين نماذج نظرية الاستجابة للفقرة، خصوصاً في سياق أدوات قياس الرأي والاتجاه.

الفجوات البحثية وموضع الدراسة الحالية منها

يتضح من العرض السابق وجود فجوة بحثية محددة تتمثل في محدودية الدراسات التي جمعت بين تطبيق نموذج التقدير الجزئي (PCM) وتحليل الأداء التفاضلي للفقرات في سياق استبانات رأي الطلبة في المجال السريري. إذ إن الدراسات التي استخدمت هذا النموذج (المحور الثاني) طُبقت - في الغالب - على اختبارات قدرات أو مقاييس اتجاهات عامة (القرشي، 2016؛ Leis et al., 2015)، أو على عينات تعليمية غير طبية (حمدان، 2025؛ الثبيتي، 2016)، باستثناء دراسة Hur & Seo (2023) التي طُبقت على طلبة الطب، إلا أنها ركزت على قياس السمات الشخصية وليس على تقييم الرأي حول المناهج السريرية.

في المقابل، فإن الدراسة التي تناولت استبانة لبيئة التدريب السريري في السياق السعودي (Alanazi et al., 2025) اعتمدت على الأساليب الكلاسيكية في التحليل الإحصائي، دون توظيف نماذج نظرية الاستجابة للفقرة أو نموذج التقدير الجزئي. وبذلك تتقاطع الدراسة الحالية مع هذين المسارين البحثيين في آنٍ واحد، من خلال تطبيق نموذج التقدير الجزئي على استبانة رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض، مع تحليل الأداء التفاضلي للفقرات تبعاً لمتغيري الجنس والكلية، وبالاستناد إلى معيار يجمع بين حجم الأثر والدلالة الإحصائية، بما يسهم في تقليل التباين الناتج عن الاعتماد على الدلالة الإحصائية وحدها كما أشارت إليه بعض الدراسات السابقة (Na et al., 2015).

وبناءً على ذلك، تسعى الدراسة الحالية إلى سد فجوة بحثية تقع عند تقاطع ثلاثة مجالات رئيسة: أدبيات الأداء التفاضلي للفقرات، وأدبيات نموذج التقدير الجزئي، وأدبيات تقييم بيئة التدريب السريري، وذلك من خلال تقديم تحليل سيكومتري أكثر دقة لاستبانة رأي الطلبة حول المناهج السريرية في السياق الجامعي محل الدراسة.

منهجية الدراسة وإجراءاتها:

استخدم الباحثان المنهج الوصفي التحليلي وذلك لتحقيق فحص الأداء التفاضلي للفقرة في استبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض باستخدام نموذج التقدير الجزئي.

مجتمع الدراسة:

تكوَّن مجتمع الدراسة من طلبة كليتي الطب والعلوم الصحية والتمريض في جامعة حضرموت البالغ عددهم (2200) طالب وطالبة، المسجلين في الفصل الدراسي الأول من العام الجامعي 2025-2026م.

عينة الدراسة:

بلغت عينة الدراسة الأساسية (331) طالباً وطالبة، تم اختيارهم بالطريقة الطبقية العشوائية، موزعين حسب الجنس والكلية، والجدول (1) يوضح ذلك:

جدول 1

توزيع عينة الدراسة حسب الجنس والكلية

الكلية	ذكور	إناث	الإجمالي
الطب والعلوم الصحية	109	70	179
التمريض	66	86	152
الإجمالي	175	156	331

أداة الدراسة

اعتمدت الدراسة الحالية على استبانة قياس آراء الطلبة حول المنهج السريري، وهي أداة جاهزة أعدها مركز التطوير الأكاديمي وضمان الجودة بجامعة حضرموت عام 2025م ضمن منظومة تقويم البرامج الأكاديمية وتحسين جودة العملية التعليمية في الكليات الصحية، ومنها كلية الطب والعلوم الصحية وكلية التمريض. وقد أُعدت الاستبانة بهدف التعرف إلى آراء الطلبة حول مختلف جوانب تجربتهم في المنهج السريري وتقويم مستوى رضاهم عنه.

وقد خضعت الاستبانة للتحكيم عند إعدادها من قبل الجهة المعدّة لها، إلا أن بيانات المحكمين ونتائج إجراءات التحكيم الأصلية لم تكن متاحة للباحثين وقت إعداد الدراسة، كما لم تتوفر نتائج إحصائية موثقة بصورة كافية للتحقق من الخصائص السيكومترية الأصلية للأداة. وبناءً عليه، لم يُعد الباحثان بناء الأداة أو تطوير فقراتها، وإنما اقتصرت الإجراءات في الدراسة الحالية على التحقق من اتساقها الداخلي وثباتها في عينة الدراسة الحالية قبل استخدامها في التطبيق النهائي.

تتكون الاستبانة في صورتها المستخدمة في الدراسة الحالية من (24)فقرة موزعة على خمسة أبعاد رئيسة، هي :محتوى المنهج السريري (6 فقرات)، وطرق التدريس والتعليم (6 فقرات)، والبيئة والموارد السريرية (4 فقرات)، والتقييم والامتحانات (4 فقرات)، والرضا العام (4 فقرات). واعتمدت الاستبانة مقياس ليكرت خماسي الاستجابة، تراوحت بدائله بين «أوافق بشدة» و«لا أوافق بشدة»، وأُعطيت الدرجات (5، 4، 3، 2، 1)، على التوالي. وجميع الفقرات مصاغة في الاتجاه الموجب؛ ومن ثم لا تتطلب أي منها عكس الدرجات.

التحقق من الخصائص السيكومترية للأداة في الدراسة الحالية

للتحقق من ملاءمة الأداة للتطبيق في الدراسة الحالية، طُبقت الاستبانة على عينة استطلاعية قوامها (65) طالبًا وطالبة من خارج عينة الدراسة الأساسية .واستخدمت بيانات هذه العينة للتحقق من الاتساق الداخلي والثبات قبل التطبيق النهائي للأداة.

أولًا: الاتساق الداخلي

تم التحقق من الاتساق الداخلي من خلال حساب معاملات ارتباط بيرسون بين درجة كل فقرة والدرجة الكلية للبُعد الذي تنتمي إليه، وكذلك حساب معاملات الارتباط بين الدرجة الفرعية لكل بُعد والدرجة الكلية للاستبانة. والجدولين (2، 3) يوضحان ذلك.

جدول 2

معاملات ارتباط بيرسون بين درجة كل فقرات والدرجة الكلية للبُعد الذي تنتمي إليه

محتوى المنهج السريري		طرق التدريس والتعليم		البيئة والموارد السريرية		التقييم والامتحانات		الرضا العام	
م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط
1	.622**	**7**	.760**	13	.803**	17	.826**	21	.824**
2	.683**	**8**	.731**	14	.783**	18	.784**	22	.874**
3	.727**	**9**	.696**	15	.785**	19	.718**	23	.781**
4	.827**	10	.853**	16	.829**	20	.791**	24	.520**
5	.821**	11	.810**						
6	.708**	12	.835**						

** الارتباط دال إحصائياً عند مستوى دلالة α=0.01.

يتضح من الجدول (2) أن جميع فقرات الاستبانة حققت معاملات ارتباط موجبة ودالة إحصائياً مع الدرجة الكلية للأبعاد التي تنتمي إليها عند مستوى الدلالة ($\alpha \geq 0.01$)، إذ بلغت قيمة الدلالة الإحصائية لجميع الفقرات (0.000)، وهو ما يشير إلى وجود علاقة ارتباط معنوية وقوية بين كل فقرة والبُعد الذي تقيسه. وتُعد هذه النتيجة مؤشراً مهماً على تمتع الأداة بدرجة مرتفعة من الاتساق الداخلي.

كما تم حساب معاملات ارتباط بيرسون بين الدرجة الفرعية لكل بُعد مع الدرجة الكلية لأداة الدراسة (Nunnally & Bernstein, 1994)، وكانت النتائج كما يوضحها الجدول رقم (3)

جدول 3

معاملات ارتباط بيرسون بين الدرجة الفرعية لكل بُعد والدرجة الكلية لأداة الدراسة

م	أبعاد أداة الدراسة	معامل ارتباط
1.	محتوى المنهج السريري	.895**
2.	طرق التدريس والتعليم	.926**
3.	البيئة والموارد السريرية	.828**
4.	التقييم والامتحانات	.706**
5.	**الرضا العام**	**.848****

** الارتباط دال إحصائياً عند مستوى دلالة α=0.01.

يتضح من الجدول (3) وجود ارتباطات موجبة ودالة إحصائياً بين الدرجة الفرعية لكل بُعد والدرجة الكلية لأداة الدراسة عند مستوى الدلالة ($\alpha \geq 0.01$)، إذ تراوحت معاملات الارتباط بين (0.706) و(0.926)، وجميعها ذات دلالة إحصائية عالية.

ثالثاً: ثبات أداة الدراسة

تم استخدام معامل أوميغا (Omega) ومعامل ألفا كرونباخ (Cronbach's Alpha) معًا للتحقق من ثبات أداة الدراسة، كما اقترحه (McDonald, 1978, 1999)، والجدول (4) يوضح النتائج:

جدول 4

معامل أوميغا (Omega) ومعامل ألفا كرونباخ (Cronbach's Alpha)

م	الأبعاد	عدد الفقرات	معامل الفا كرونباخ	معامل اوميجا
1	محتوى المنهج السريري	6	0.82	0.83
2	طرق التدريس والتعليم	6	0.87	0.87

3	البيئة والموارد السريرية	4	0.81	0.81
4	التقييم والامتحانات	4	0.78	0.77
5	الرضا العام	4	0.74	0.78
استبانة قياس رأي الطلبة حول المناهج السريرية		24	0.94	0.94

يتضح من جدول (4) أن معاملات الثبات لأبعاد استبانة الدراسة جاءت مرتفعة باستخدام كل من معامل ألفا كرونباخ ومعامل أوميغا، مما يدل على تمتع الأداة بدرجة جيدة من الاتساق الداخلي والثبات. فقد كان أعلى معامل ألفا كرونباخ لمجال طرق التدريس (0.87)، في حين كان أعلى معامل أوميغا للمجال نفسه (0.87)، وهي قيم مرتفعة تدل على ثبات جيد لفقرات هذا البُعد. كما بلغ أقل معامل ألفا كرونباخ لمجال الرضا العام (0.74)، في حين كان أقل معامل أوميغا لمجال التقييم والامتحانات (0.77)، وهي قيم تجاوزت الحد المقبول للثبات (0.70) وفقًا لـ(Nunnally & Bernstein (1994، وعلى مستوى الأداة ككل، بلغ معامل ألفا كرونباخ (0.94)، وكذلك معامل أوميغا (0.94)، وهي قيم مرتفعة جداً تدل على تمتع الأداة ككل بدرجة عالية من الثبات والاتساق الداخلي.

وعليه، تشير معاملات الثبات إلى تمتع الأداة بمستوى مقبول من الاتساق الداخلي يدعم استخدامها في جمع بيانات الدراسة.

التطبيق النهائي للاستبانة

بعد التحقق من الخصائص السيكومترية للاستبانة من نتائج التطبيق على العينة الاستطلاعية؛ تكونت الاستبانة النهائية من (24) فقرة، تم تطبيقها على عينة الدراسة الأساسية التي بلغت (331) طالباً وطالبة، ثم تم إدخال البيانات إلى ذاكرة الحاسوب لبدء التحليلات الإحصائية باستخدام برنامج SPSS V25، وبرنامج باراسكال (PARSCALE 4.1).

الأساليب الإحصائية:

تم استخدام الأساليب الإحصائية الآتية:

1. معامل ارتباط بيرسون لحساب الاتساق الداخلي للمقياس.
2. معامل ألفا كرونباخ، ومعامل أوميغا لحساب الثبات.
3. التحليل العاملي الاستكشافي لمعرفة تشبع الفقرات على العوامل.
4. دالة المعلومات (Information Function) لحساب دقة القياس (الثبات) وفقاً لمعالم نموذج التقدير الجزئي عبر برنامج باراسكال.
5. تحليل البيانات باستخدام نموذج التقدير الجزئي باستخدام برنامج باراسكال (PARSCALE 4.1).

النتائج ومناقشتها:

الإجابة عن السؤال الأول الذي ينص على:" هل تحقق استبانة قياس رأي الطلبة حول المناهج السريرية افتراضات نموذج التقدير الجزئي؟"

أولاً: افتراض أحادية البعد(Unidimensionality)

للتحقق من البنية العاملية للأداة ومدى ملاءمتها لافتراض أحادية البعد، أُجري التحليل العاملي الاستكشافي. وقبل ذلك، تم التحقق من ملاءمة مصفوفة الارتباطات للتحليل العاملي من خلال اختبار كايزر- ماير-أولكن (KMO)، إذ بلغت قيمته (KMO=0.933)، وهي أعلى من الحد الأدنى المطلوب الذي يجب ألا يقل عن (0.6) حسب محك كايزر(1974) Kaiser. (2) اختبار بارتليت لاختبار افتراض الكروية بدلالة قيمة كا2، إذ كانت (χ^2 = 3976.77)، بدرجة حرية=276، وبمستوى دلالة (p < 0.0001)، وهي دالة إحصائياً وفقاً لمحك (Guilford & Fruchter, 1973)، أي أن مصفوفة معاملات الارتباط في كل اختبار ليست مصفوفة الوحدة، وهذا يؤكد أن مصفوفة الارتباطات ليست الوحدة. وهذان المؤشران يدعمان استخدام البيانات في التحليل العاملي الاستكشافي.

وبناءً على ذلك، تم إجراء التحليل العاملي الاستكشافي باستخدام طريقة تحليل العوامل الأساسية (PAF) نظرًا لملاءمتها لهدف التحليل المتمثل في الكشف عن العوامل الكامنة المشتركة التي تفسر الارتباطات بين فقرات الأداة، بدلًا من استخدام تحليل المكونات الرئيسة (PCA) الذي يركز أساسًا على اختزال التباين الكلي للمتغيرات (Fabrigar et al., 1999) ، كما تم استخدام التدوير المائل (Promax) بدلاً من التدوير المتعامد (Varimax)، وذلك استناداً إلى الافتراض النظري بأن أبعاد الاستبانة مترابطة، كونها تمثل جوانب مختلفة لمفهوم عام يتمثل في جودة تجربة التعلّم السريري. ويعرض الجدول (5) قيم الجذور الكامنة للعوامل المستخرجة.

جدول 5

قيم الجذر الكامن (Eigenvalues) ونسبة التباين المفسر للتحليل العاملي الاستكشافي

العامل	الجذر الكامن	نسبة التباين ٪	التباين التراكمي ٪
1	9.822	40.926	40.926
2	1.649	6.871	47.797
3	1.172	4.884	52.681
4	1.121	4.671	57.353

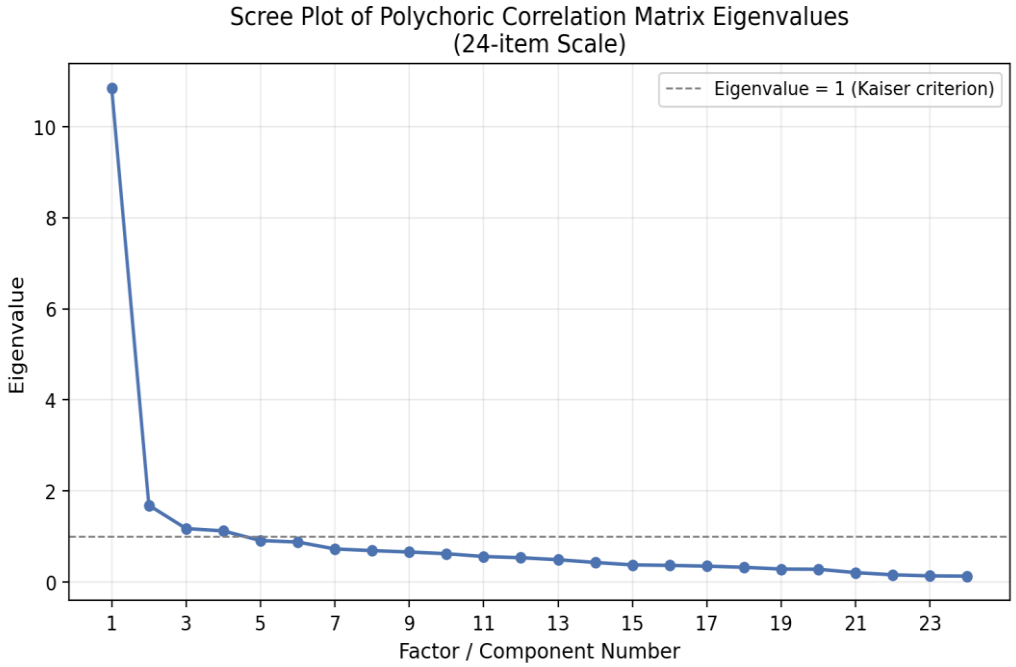
يتضح من الجدول (5) أن قيمة الجذر الكامن للعامل الأول بلغ (9.82)، الذي يفسر بمفرده (40.9%) من التباين في الدرجة بالنسبة للتباين الكلي الذي يرجع إلى جميع العوامل، وتشير إلى هيمنة عامل عام أول في البنية العاملية للأداة،

إذ أشار Reckase (1979) إلى أن أحادية البعد تتحقق بصورة مقبولة إذا استطاع العامل الأول أن يفسر ما نسبته (20%) على الأقل من التباين الكلي

كما يُظهر مخطط الركام (Scree Plot) في الشكل (1) تغيرًا حادًا في ميل المنحنى بعد العامل الأول، يعقبه انخفاض واستقرار نسبي في قيم الجذور الكامنة للعوامل اللاحقة، مما يشير إلى هيمنة العامل الأول في البنية العاملية للأداة، وفق محك كاتل (Cattell, 1966). وبالاستناد إلى تكامل هذه النتيجة مع نسبة التباين التي يفسرها العامل الأول، تم الاحتفاظ بالعامل الأول بوصفه العامل العام المهيمن، مما يدعم افتراض أحادية البعد للأداة.

شكل 1

التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للمقياس (Scree Plot)



يتضح من الشكل (1) أن هناك فرقاً كبيراً بين جذر العامل الأول وجذر العامل الثاني لذا تم الإبقاء على العامل الأول فقط، ورفض بقية العوامل المستخلصة من التحليل، وهذا ما يوضحه الجدول (6).

جدول 6

معاملات تشبع فقرات استبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض بالعامل العام

م	الفقرة	معامل التشبع
1	محتوى المنهج السريري وثيق الصلة بالاحتياجات الفعلية للممارسة المهنية.	.536
2	عدد الساعات المخصصة للتدريب العملي السريري كافٍ	.592
3	تنوّع الحالات السريرية التي اطلعتُ عليها ملائمة	.562
4	هناك تكامل واضح بين المعرفة النظرية والمهارات السريرية في المنهج.	.700
5	المنهج منظم بشكل منطقي ومتسلسل، مما يسهل عملية التعلم.	.670
6	المنهج يغطي مجموعة كافية ومتنوعة من الحالات السريرية.	.643
7	طرق التدريس السريري المستخدمة (مثل العروض، المناقشات) فعالة ومحفزة.	.641
8	يتم استخدام المحاكاة السريرية والتدريب العملي بشكل كافٍ ومناسب.	.730
9	يتم تشجيع التعلم الذاتي وحل المشكلات (Problem-Based Learning) في المنهج.	.596
10	الوقت المتاح في تجربة التعلّم كافٍ لدمج المحتوى بطرق مكّنتني من تطبيقه على تعلّمي	.737
11	أُتيحت لي الفرصة لمناقشة أهداف تجربة التعلّم مع عضو هيئة التدريس.	.662
12	يتلقى الطلبة تغذية راجعة بناءة ومفيدة حول أدائهم السريري.	.744
13	الإشراف السريري المقدم من أعضاء هيئة التدريس فعال وداعم.	.688
14	تتوفر فرص كافية للتفاعل المباشر مع المرضى والحالات السريرية المتنوعة.	.595
15	المرافق والمعدات السريرية المتاحة للتدريب حديثة وكافية.	.623
16	البيئة السريرية آمنة وداعمة للتعلم دون خوف من الأخطاء.	.624
17	أساليب التقييم السريري (مثل OSCE، الامتحانات العملية) عادلة وموضوعية.	.556
18	معايير التقييم واضحة ومفهومة للطلبة.	.582
19	التقييم يعكس فعليًا مهاراتي السريرية ويقيسها بدقة.	.597
20	يتم استخدام أساليب تقييم متعددة لقياس مهاراتي ومعرفتي خلال التجربة السريرية.	.559
21	أنا راضٍ بشكل عام عن تجربتي في المنهج السريري.	.783
22	أنا راضٍ عن مدى مراعاة تجربة التعلّم لاحتياجاتي الفردية.	.776
23	أنا راضٍ عن وضوح أهداف التدريب وعن تشجيع عضو هيئة التدريس لمشاركتي.	.687
24	المنهج السريري أعدني بشكل جيد للمرحلة المهنية القادمة.	0.335

يضح من الجدول (6) أن معاملات تشبع الفقرات بالعامل العام تراوحت بين (0.335–0.783)، وجاءت جميعها أعلى من (0.30)، وهو الحد الأدنى الذي يمكن الاسترشاد به لقبول التشبع العاملي في

التحليل العاملي الاستكشافي وفق .Hair et al (2019) ، وبذلك أسهمت جميع الفقرات في قياس العامل العام، مع ملاحظة أن الفقرة (24) سجلت أدنى تشبع عاملي (0.335)، وهو ما يشير إلى إسهامها النسبي الأقل في العامل مقارنة ببقية الفقرات، دون أن يستدعي ذلك استبعادها استنادًا إلى معيار التشبع المعتمد.

كما تم التحقق من أحادية البعد في إطار نموذج راش من خلال تحليل المكونات الرئيسة للبواقي المعيارية(Principal Components Analysis of Standardized Residuals: PCA) ، بهدف التحقق من خلو البواقي من بنية عاملية ثانوية جوهرية بعد تفسير البعد الرئيس، وهو إجراء شائع في فحص افتراض أحادية البعد في نماذج راش .(Linacre, 1998) وتعرض نتائج هذا التحليل في الجدول.(7)

جدول7

نتائج تحليل المكونات الرئيسة للبواقي

المؤشر	القيمة
التباين المفسر بواسطة القياس	41.9%
القيمة الكامنة لأول تباين متبقٍ	2.7
نسبة التباين لأول تباين متبقٍ	6.4%
الحكم النهائي	أحادية البعد متحققة

يتضح من الجدول (7) أن النموذج فسر (41.9%) من التباين الكلي، وهي نسبة تتجاوز الحد (20%) الذي يعد مقبولاً في الدراسات التربوية والنفسية (Bond & Fox, 2015). كما بلغت القيمة الكامنة لأول تباين متبقي (2.7)، وهي أقل من القيمة (3.0) التي تشير إلى وجود عامل ثانوي قوي وفق المعايير التي اقترحها(2012) Linacre ، مما يدعم تحقق افتراض أحادية البعد.

وعليه يمكن اعتبار الأداة أحادية البعد بدرجة مقبولة، مما يدعم استخدام نموذج التقدير الجزئي في تحليل بياناتها.

ثانياً افتراض الاستقلال الموضعي: Local Independence

يقصد بالاستقلال الموضعي الاستقلال الإحصائي لأداء المفحوصين على فقرات الاختبار، أي: إن أداء المفحوص على فقرة لا يتأثر بأدائه على فقرة أخرى (Lawson & Brailovsky, 2006)، ويرى (Hambleton & Swaminathan 1985) أنَّ مفهوم الاستقلال الموضعي وأحادية البعد متكافئان؛ وبالتالي، فتحقق أحدها يضمن تحقق الآخر، وهو ما يؤكده lord (1980) بأنَّ الاستقلال الموضعي يمكن الاستدلال عليه من خلال أحادية البعد، فتحقيقه يعني تحقق افتراض الاستقلال الموضعي.

كما تم أيضا التحقق من افتراض الاستقلال الموضعي من خلال تحليل مصفوفة ارتباطات البواقي المعيارية بين الفقرات، إذ بلغ متوسط الارتباطات المتبقية (0.042-)، وهو قريب من الصفر، مما يشير إلى

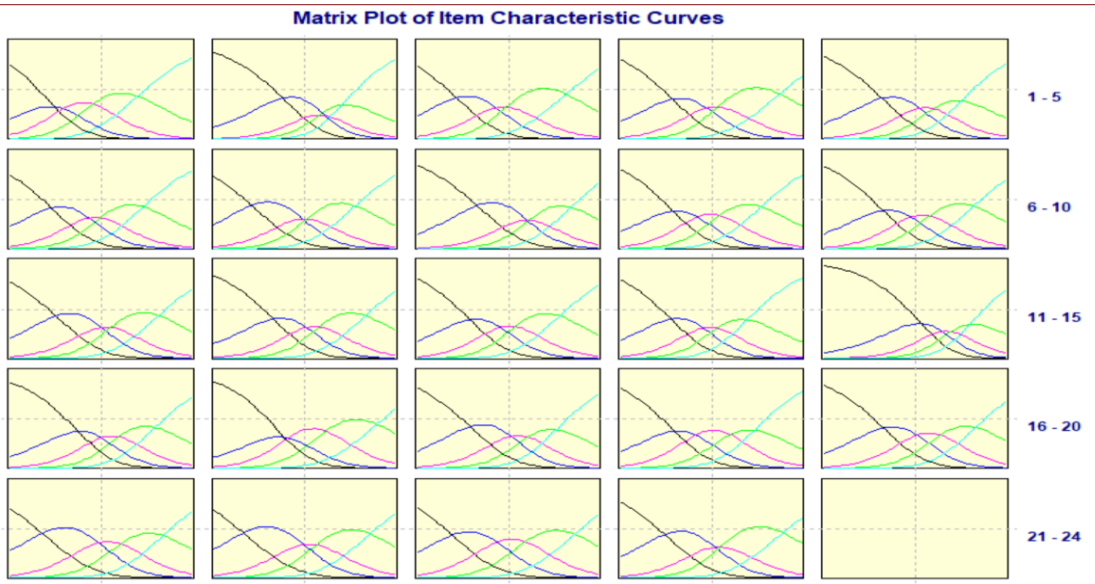
محدودية العلاقات المتبقية بين الفقرات بعد التحكم في أثر البعد الكامن المقاس، ويدعم تحقق افتراض الاستقلال الموضعي بصورة عامة.(Linacre, 2024; Smith, 2002)

ثالثًا: افتراض المنحنيات المميزة للفئات الاستجابية:

تُستخدم المنحنيات المميزة لفئات الاستجابة لفحص العلاقة بين مستوى السمة الكامنة واحتمال اختيار كل فئة من فئات الاستجابة على الفقرة. ويُتوقع أن تنتقل الاحتمالية تدريجيًا من الفئات الأدنى إلى الفئات الأعلى مع ارتفاع مستوى السمة، بحيث تمثل كل فئة نطاقًا من مستويات السمة تكون فيه الأكثر احتمالًا للاختيار. وفي نموذج التقدير الجزئي، يُفترض ثبات معامل التمييز بين الفقرات، بينما تتحدد مواقع المنحنيات وفق صعوبة الفقرة ومعلمات الانتقال بين فئات الاستجابة. وقد استُخدم برنامج PARSCALE لفحص المنحنيات المميزة لفئات الاستجابة للفقرات، كما يوضحها الشكل الآتي.

الشكل 2

منحنيات احتمال فئات الاستجابة لفقرات المقياس



الافتراض الرابع: التحرر من عامل السرعة.

تم مراعات هذا الافتراض عند تطبيق أداة الدراسة، فلم يقرن الإجابة على الاستبانة بزمن محدد، وترك الحرية للطالب في الإجابة دون ضغط عامل الزمن، كما لم يلاحظ الباحثان أي شكوى تجاه عنصر الزمن من الطلاب، ولم يلاحظا أي حالة ترك فقرة بدون استجابة. ومن ثم يمكننا القول بتحقق افتراض التحرر من عامل السرعة تجريباً أثناء عملية الاستبانة.

الإجابة عن السؤال الثاني الذي ينص على:" ما مدى جودة مطابقة فقرات الاستبانة لنموذج التقدير الجزئي؟"

للإجابة عن هذا السؤال، اتبع الباحثان الخطوات الآتية:

تم أولاً فحص ملاءمة الأفراد لنموذج التقدير الجزئي باستخدام برنامج PARSCALE، إذ يعتمد البرنامج على تقدير معاملات الأفراد واستبعاد الحالات التي تظهر أخطاء معيارية مرتفعة أو استجابات غير مستقرة قد تؤثر في دقة التقدير. وقد أسفرت نتائج التحليل الأولي عن وجود اثنين من أفراد العينة كانت استجاباتهم غير ملائمتين (رقم 27 و110)، مما استدعى استبعادهما من التحليل لضمان استقرار التقديرات.

وبعد ذلك، أُعيد تحليل البيانات للتحقق من ملاءمة فقرات الاستبانة للنموذج(Item Fit) ، باستخدام إحصاء مربع كاي (Chi-Square) المتاح ضمن مؤشرات المطابقة في برنامج .PARSCALE ويقيس هذا المؤشر مدى توافق الاستجابات الملاحظة مع الاستجابات المتوقعة وفق نموذج التقدير الجزئي؛ وتشير قيمة الدلالة الإحصائية غير الدالة إلى عدم توفر دليل إحصائي كافٍ على سوء مطابقة الفقرة للنموذج (DeMars, 2005)، ويعرض الجدول (8) مؤشرات ملاءمة فقرات الاستبانة وفق نموذج التقدير الجزئي بعد استبعاد الحالات غير الملائمة.

جدول 8

إحصاءات كاي تربيع والقيم الاحتمالية لفحص مطابقة فقرات الاستبانة لنموذج التقدير الجزئي

م	رمز الفقرة	مربع كاي	درجة الحرية	القيم الاحتمالية	م	رمز الفقرة	مربع كاي	درجة الحرية	القيم الاحتمالية
0001	C1	22.442	15	0.096	0013	L1	15.576	16	0.483
0002	C2	17.568	17	0.417	0014	L2	19.055	16	0.265
0003	C3	14.001	16	0.599	0015	L3	5.658	15	0.985
0004	C4	18.252	15	0.249	0016	L4	15.799	16	0.467
0005	C5	10.894	15	0.761	0017	A1	11.384	15	0.726
0006	C6	14.731	16	0.545	0018	A2	9.606	15	0.844
0007	T1	10.818	16	0.821	0019	A3	9.254	15	0.864
0008	T2	21.923	17	0.187	0020	A4	15.135	15	0.442
0009	T3	15.217	16	0.509	0021	S1	50.516	15	0.000
0010	T4	21.975	15	0.108	0022	S2	36.363	15	0.002
0011	T5	13.056	15	0.599	0023	S3	17.503	15	0.289
0012	T6	24.723	15	0.054	0024	S4	23.042	15	0.083

يتضح من الجدول (8) أن القيمة الاحتمالية (p-value) المصاحبة لإحصاء كاي تربيع استُخدمت معيارًا لفحص مطابقة الفقرات لنموذج التقدير الجزئي عند مستوى الدلالة .($\alpha = 0.05$) وتشير القيمة الاحتمالية التي تقل عن أو تساوي (0.05) إلى وجود دليل إحصائي على سوء مطابقة الفقرة للنموذج، في حين تشير القيمة الاحتمالية الأكبر من (0.05) إلى عدم توفر دليل إحصائي كافٍ على سوء المطابقة وفق هذا المؤشر .(DeMars, 2005) وبناءً على هذا المعيار، ظهرت الفقرتان (S1=21) و (S2=22)فقط بقيم دالة إحصائيًا، إذ بلغت القيمة الاحتمالية ($p < 0.001$) للفقرة(S1) ، و ($p = 0.002$)للفقرة(S2) ، في حين كانت القيم الاحتمالية لبقية الفقرات أكبر من (0.05). وبناءً عليه، استُبعدت الفقرتان (S1) و(S2)، اللتان تمثلان نسبة (8.3%) من إجمالي فقرات الاستبانة، ليصبح عدد الفقرات في صورتها النهائية (22) فقرة.

وقد اتفقت نتائج الدراسة مع دراسة كل من (العناتي، 2017؛ Boyd et al., 2014؛ Liu et al., 2023) في أن استبعاد عدد محدود من الفقرات غير الملائمة يعد إجراء منهجيا في نموذج التقدير الجزئي وغيره من النماذج الحديثة، الغرض منه تحسين المطابقة وتحقق القياس الموضوعي للحصول على أداة تتمتع بخصائص سيكومترية مناسبة.

الإجابة عن السؤال الثالث الذي ينص على: " ما مستوى دقة القياس الذي توفره الاستبانة وفقاً لدالة المعلومات في نموذج التقدير الجزئي؟"

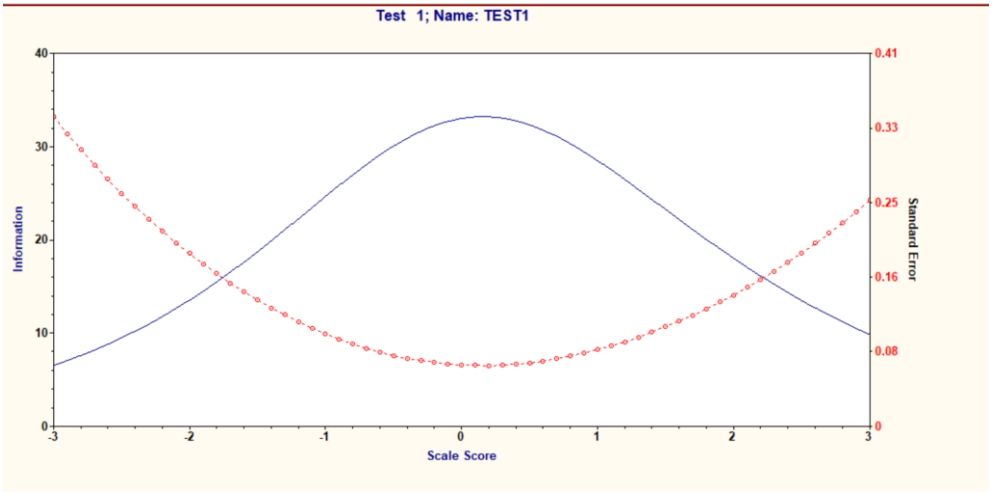
للإجابة عن هذا السؤال، تم استخراج دالة معلومات الاختبار الكلية (Function Test Information) ومنحنى الخطأ المعياري الشرطي للقياس (Conditional Standard Error of Measurement). باستخدام برنامج PARSCALE 4.1، وذلك بعد تقدير معالم فقرات استبانة قياس رأي الطلبة حول المناهج السريرية وفق نموذج التقدير الجزئي.

وتُحسب دالة معلومات الاختبار عند كل مستوى من مستويات السمة الكامنة (θ) من خلال تجميع معلومات الفقرات المكونة للاختبار، إذ تمثل معلومة الفقرة مقياساً لعكس تباين الخطأ المعياري لتقدير القدرة عند ذلك المستوى. وبناءً عليه، فإن ارتفاع قيمة دالة المعلومات عند مستوى معين من السمة الكامنة يشير إلى انخفاض الخطأ المعياري للقياس عند ذلك المستوى، وبالتالي ارتفاع دقة القياس في التمييز بين الأفراد عنده.

ويُوضح الشكل (3) منحنى دالة المعلومات (الخط المتصل) ومنحنى الخطأ المعياري الشرطي للقياس (الخط المتقطع) لاستبانة الدراسة، كما تم توليده بواسطة برنامج PARSCALE.

شكل3

منحنى دالة معلومات الاختبار الكلية والخطأ المعياري الشرطي للقياس وفق نموذج التقدير الجزئي



ويوضح الجدول (9) القيم التقريبية لدالة المعلومات والخطأ المعياري الشرطي عند مستويات مختارة من السمة الكامنة (θ)، كما يبين المنحنى أعلاه:

جدول 9

قيم دالة المعلومات والخطأ المعياري الشرطي عند مستويات مختارة من θ

θ (الدرجة على المتّصلة)	دالة المعلومات I(θ)	الخطأ المعياري الشرطي SE(θ)	ملاحظة
-3.00	≈9.0	≈0.33	
-2.00	≈16.0	≈0.25	
-1.00	≈27.0	≈0.19	
0.00	≈32.0	≈0.18	
+0.15	≈ 33.0 (الأعلى)	≈ 0.17 (الأدنى)	نقطة القمة
+1.00	≈29.0	≈0.19	
+2.00	≈16.5	≈0.25	
+3.00	≈9.5	≈0.32	

يتضح من الشكل (3) والجدول (9) أن دالة المعلومات بلغت أعلى قيمة لها ($I(\theta) \approx 33$) عند مستوى السمة الكامنة $\theta \approx 0.15$ ، في حين بلغ الخطأ المعياري للقياس عند هذا المستوى أدنى قيمة له تقريبًا. ($SE(\theta) \approx 0.17$) وتشير هذه النتيجة إلى أن دقة قياس الاستبانة تكون أعلى عند مستويات السمة

التي تتوفر عندها أكبر كمية من المعلومات، بينما تتناقص دقة القياس كلما انخفضت كمية المعلومات المتاحة عند مستويات أخرى من السمة. ويُعد هذا من الخصائص الأساسية لدالة معلومات الاختبار في نظرية الاستجابة للفقرة، إذ ترتبط معلومات الاختبار عكسيًا بمربع الخطأ المعياري للقياس (Hambleton et al., 1991; Embretson & Reise, 2000).

وبلغ معامل الثبات التقديري المستند إلى دالة المعلومات نحو (0.96)، وهي قيمة مرتفعة، كما تقاربت مع قيمتي ألفا كرونباخ وأوميغا للأداة ككل (0.94). ويشير هذا التقارب إلى اتساق النتائج المتعلقة بدقة القياس بين المؤشرات المستندة إلى النظرية الكلاسيكية للقياس والمؤشر المستند إلى نموذج الاستجابة للفقرة، مع التأكيد على أن دالة المعلومات تقدم ميزة إضافية تتمثل في بيان **تغير دقة القياس باختلاف مستوى السمة الكامنة**، بدلًا من تقديم تقدير واحد للثبات على مستوى الأداة ككل (Baker, 2001; Embretson & Reise, 2000).

وقد اتفقت نتيجة هذا السؤال في قيمة الثبات الهامشي أنه مرتفع (≈0.96) مع دراسة Na et al., 2015 التي أظهرت أن الثبات مرتفع وبلغ (0.85).

الإجابة عن السؤال الرابع الذي ينص على: " هل توجد فقرات ذات أداء تفاضلي في الاستبانة تبعاً لمتغير الجنس (ذكور/إناث)؟"

وللإجابة عن هذا السؤال تم الكشف عن الأداء التفاضلي للفقرات باستخدام برنامج PARSCALE 4.1 باستخدام نموذج التقدير الجزئي، وذلك بحساب الفرق بين معاملي الصعوبة لمجموعتي (ذكور/ إناث)، على كل فقرة من الفقرات. ووفقا لما ذكره Keeves & Alagumalai (1999) فإن الفقرة تُعدّ ذات أداء تفاضلي؛ إذا كان الفرق المطلق بين قيمتي معلم الصعوبة للمجموعتين، أكبر من أو يساوي (0.5)؛ أي $|\Delta b| \geq 0.50$، ويوضح جدول (10) نتائج الأداء التفاضلي لاستبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض تبعاً لمتغير الجنس (إناث/ ذكور). وقد عُدت مجموعة الذكور هي المجموعة المرجعية، ومجموعة الإناث المجموعة المستهدفة.

جدول 10

نتائج الأداء التفاضلي لاستبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض (إناث/ ذكور)

رقم الفقرة	رمز الفقرة	الصعوبة b ذكور	الصعوبة b إناث	Δb = b ذكور−b إناث	$\|\Delta b\|$	وجود DIF ($\|\Delta b\| \geq 0.50$)
0001	C1	-0.606	-0.434	0.172	0.172	لا يوجد DIF
0002	C2	0.868	0.405	-0.463	0.463	لا يوجد DIF

0003	C3	-0.335	0.169	0.504	0.504	نعم - يوجد DIF
0004	C4	0.289	0.140	-0.149	0.149	لا يوجد DIF
0005	C5	0.297	0.510	0.213	0.213	لا يوجد DIF
0006	C6	0.087	-0.343	-0.430	0.430	لا يوجد DIF
0007	T1	-0.107	0.184	0.291	0.291	لا يوجد DIF
0008	T2	0.877	0.548	-0.329	0.329	لا يوجد DIF
0009	T3	0.095	0.000	-0.095	0.095	لا يوجد DIF
0010	T4	0.224	0.405	0.181	0.181	لا يوجد DIF
0011	T5	0.281	0.184	-0.097	0.097	لا يوجد DIF
0012	T6	0.403	0.405	0.002	0.002	لا يوجد DIF
0013	L1	0.395	-0.133	-0.528	0.528	نعم - يوجد DIF
0014	L2	0.151	-0.140	-0.291	0.291	لا يوجد DIF
0015	L3	1.712	1.035	-0.677	0.677	نعم - يوجد DIF
0016	L4	0.621	0.316	-0.305	0.305	لا يوجد DIF
0017	A1	0.095	0.646	0.551	0.551	نعم - يوجد DIF
0018	A2	0.240	0.495	0.255	0.255	لا يوجد DIF
0019	A3	-0.099	0.316	0.415	0.415	لا يوجد DIF
0020	A4	0.371	0.646	0.275	0.275	لا يوجد DIF
0023	S3	-0.018	0.316	0.334	0.334	لا يوجد DIF
0024	S4	0.216	0.383	0.167	0.167	لا يوجد DIF

يتضح من الجدول (10) أن أربع فقرات من أصل (22) فقرة تجاوزت محك فرق معلم الصعوبة البالغ (0.50) لوجيت، وهي الفقرات (3، 13، 15، 17)، بنسبة بلغت (18.18%). وقد تراوحت قيم الفرق المطلق في هذه الفقرات بين (0.504–0.677) لوجيت، في حين جاءت القيم المطلقة لبقية الفقرات دون هذا المحك. وتشير هذه النتيجة إلى وجود فروق في موقع هذه الفقرات على متصل السمة بين مجموعتي الذكور والإناث وفق معيار الكشف المستخدم.

ولا ينبغي تفسير هذه الفروق بوصفها دليلًا مباشرًا على تحيز الفقرات أو إرجاعها إلى سبب محدد؛ إذ يكشف تحليل الأداء التفاضلي عن اختلاف في أداء الفقرة بين مجموعتي المقارنة، لكنه لا يحدد بمفرده

مصدر هذا الاختلاف. ومن ثم، يمكن أن ترتبط الفروق بالسياق التعليمي أو السريري أو بخصائص الخبرة التي تقيسها الفقرة، إلا أن إثبات أي من هذه التفسيرات يتطلب أدلة مستقلة.

وبناءً على ذلك، لم تُستبعد الفقرات الأربع تلقائيًا لمجرد تجاوزها معيار الكشف عن الأداء التفاضلي، إذ إن ظهور DIF يمثل مؤشرًا إحصائيًا يستدعي فحص الفقرة وتفسير أدائها في ضوء مضمونها وسياق تطبيق الأداة، ولا يُعد بمفرده دليلًا كافيًا على تحيز الفقرة أو عدم صلاحيتها للقياس (Camilli & Shepard, 1994) وبالنظر إلى ارتباط الفقرات الأربع بالمحتوى المقاس وأهميتها في تمثيل الأبعاد التي صُممت الأداة لقياسها، وإلى أن التحليل الحالي كشف اختلافًا في أدائها دون أن يحدد السبب المباشر لهذا الاختلاف، فقد أُبقيت ضمن الأداة النهائية، مع تسجيل ظهور الأداء التفاضلي فيها بوصفه نتيجة سيكومترية تستدعي المراجعة والتحقق في دراسات لاحقة.

وقد اتفقت هذه النتيجة مع دراسة كل من) العناتي، 2014؛2017 ,.Boyd et al ؛ (Liu et al., 2023 التي أشارت إلى اقتصار الأداء التفاضلي على عدد محدود من الفقرات وفق الجنس، في حين اختلفت مع دراسة الوابل وتعلب (2023) التي أظهرت عددًا أكبر من الفقرات ذات الأداء التفاضلي حسب الجنس. كما اختلفت مع دراسة Na et al. (2015) التي أظهرت أداءً تفاضليًا في عدد كبير من الفقرات. وقد يُعزى اختلاف النتائج بين الدراسات إلى اختلاف خصائص الأدوات والعينات ومناهج الكشف والمعايير المستخدمة في تحديد الأداء التفاضلي، ولذلك ينبغي تفسير المقارنات بين هذه الدراسات في ضوء تلك الاختلافات المنهجية.

الإجابة عن السؤال الخامس: الذي ينص على: "هل توجد فقرات ذات أداء تفاضلي في الاستبانة تبعاً لمتغير الكلية (الطب والعلوم الصحية/التمريض)؟"

وللإجابة عن هذا السؤال تم الكشف عن الأداء التفاضلي للفقرات (DIF)، باستخدام برنامج PARSCALE 4.1، وذلك باستخدام الفروق في معاملي صعوبة الفقرات لمجموعتي (الطب والعلوم الصحية/ والتمريض) المستمدة من نموذج التقدير الجزئي، على كل فقرة من الفقرات. بالاعتماد على المحك الذي ذكره Keeves & Alagumalai (1999) في أن الفقرة تُعد ذات أداء تفاضلي؛ إذا كان الفرق المطلق لفارق صعوبة الفقرة للمجموعتين، تساوي أو أكبر من (0.5) لوجيت؛ أي $|\Delta b| \geq 0.50$، ويوضح جدول (11) نتائج الأداء التفاضلي لاستبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض تبعاً لمتغير الكلية (الطب والعلوم الصحية/ والتمريض). وقد عُدت مجموعة كلية الطب والعلوم الصحية هي المجموعة المرجعية، ومجموعة كلية التمريض المجموعة المستهدفة.

جدول 11

نتائج الأداء التفاضلي لاستبانة قياس رأي الطلبة حول المناهج السريرية في كليتي الطب والعلوم الصحية والتمريض تبعاً لمتغير الكلية (الطب والعلوم الصحية/ والتمريض)

رقم الفقرة	رمز الفقرة	الصعوبة b طب	الصعوبة b تمريض	Δb = طب b−تمريض b	|Δb|	الحكم
0001	C1	-0.539	-0.693	-0.154	0.154	لا يوجد DIF
0002	C2	0.438	0.580	0.142	0.142	لا يوجد DIF
0003	C3	-0.130	-0.223	-0.093	0.093	لا يوجد DIF
0004	C4	0.117	0.086	-0.031	0.031	لا يوجد DIF
0005	C5	0.306	0.287	-0.019	0.019	لا يوجد DIF
0006	C6	-0.343	-0.128	0.215	0.215	لا يوجد DIF
0007	T1	0.001	-0.128	-0.129	0.129	لا يوجد DIF
0008	T2	0.565	0.606	0.041	0.041	لا يوجد DIF
0009	T3	-0.050	-0.073	-0.023	0.023	لا يوجد DIF
0010	T4	0.131	0.296	0.165	0.165	لا يوجد DIF
0011	T5	0.073	0.174	0.101	0.101	لا يوجد DIF
0012	T6	0.233	0.361	0.128	0.128	لا يوجد DIF
0013	L1	-0.064	0.094	0.158	0.158	لا يوجد DIF
0014	L2	-0.240	0.038	0.278	0.278	لا يوجد DIF
0015	L3	1.167	1.237	0.070	0.070	لا يوجد DIF
0016	L4	0.364	0.328	-0.036	0.036	لا يوجد DIF
0017	A1	0.430	0.094	-0.336	0.336	لا يوجد DIF
0018	A2	0.350	0.166	-0.184	0.184	لا يوجد DIF
0019	A3	0.037	-0.017	-0.054	0.054	لا يوجد DIF
0020	A4	0.482	0.312	-0.170	0.170	لا يوجد DIF
0023	S3	0.015	0.086	0.071	0.071	لا يوجد DIF
0024	S4	0.262	0.118	-0.144	0.144	لا يوجد DIF

يتضح من الجدول (11) أن جميع فقرات الاستبانة لم تتجاوز قيمتها المطلقة للفرق بين معلمي الصعوبة قيمة المحك (0.50) التي حددها(1999) Keeves & Alagumalai للحكم على وجود الأداء التفاضلي للفقرة (DIF)، تبعاً لمتغير الكلية (الطب والعلوم الصحية/ والتمريض)، إذ تراوحت قيم الفرق المطلق بين أقل قيمة (0.018) للفقرة (C=5)، وأعلى قيمة (0.336) للفقرة (A1=17)، بمتوسط عام (0.125).

وقد اختلفت نتائج هذه الدراسة مع دراسة (المسكري وآخرون، 2021؛ الوابل وتعلب، 2023) التي خلصت إلى وجود أداء تفاضلي تبعاً للتخصص الأكاديمي، وهو متغير يكافئ متغير الكلية في الدراسة الحالية، كما اختلفت تبعاً لاختلاف السياق التعليمي.

ويعزو الباحثان هذه النتيجة إلى تقارب محتوى برنامجي كليتي الطب والعلوم الصحية والتمريض، وتشابه البيئة التعليمية، والخبرات الأكاديمية والسريرية.

وتدل هذه النتيجة على أن فقرات الاستبانة تعمل بصورة متكافئة بين طلبة كليتي الطب والعلوم الصحية والتمريض؛ بمعنى أن الفرق في مستوى السمة الكامنة (الرضا عن المناهج السريرية) بين الطلبة في الكليتين هو المفسر الوحيد لاختلاف استجاباتهم على الفقرات؛ وليس الفروق في استجاباتهم، مع تحقق تكافؤ أداء الفقرات عبر مجموعتي المقارنة، وبذلك يتحقق افتراض التكافؤ القياسي بين المجموعتين، بما يشير إلى أن الفقرات لا تظهر أداءً تفاضليًا لصالح إحدى الكليتين، ويتسق ذلك مع طبيعة فقرات الاستبانة التي تتناول جوانب متعددة عامة في التعلم تتمثل في (محتوى المنهج، طرق التدريس، البيئة السريرية، التقييم والامتحانات، الرضا العام)، التي لا يفترض نظرياً أن تتأثر بنوع الكلية (الطب والعلوم الصحية/ والتمريض) للطلبة بقدر ما تتأثر بمستوى رضاه الفعلي، خاصة أن الكليتين تهدفان إلى خدمة المرضى.

التوصيات

1. اعتماد استبانة قياس رأي الطلبة حول المناهج السريرية بصورتها الحالية المكوّنة من (22) فقرة، في ضوء ما أظهرته نتائج التحليل السيكومتري من مؤشرات ملاءمة للنموذج.
2. الابقاء على الفقرات ذات الأداء التفاضلي تبعاً لمتغير الجنس (C3، L1، L3، A1) في صورتها الحالية، نظراً لاحتمال إسهامها في تمثيل البناء المقاس، ولما قد يترتب على حذفها من تأثير على شمولية القياس (Messick, 1989)، مع مراعاة تفسير نتائجها عند المقارنة بين المجموعات.
3. إجراء دراسات لاحقة للتحقق من طبيعة الفصل الفعلي بين الجنسين في بيئة التدريب السريري بالمؤسسة محل الدراسة، من خلال تضمين مؤشرات تصنيفية إضافية أو الاستناد إلى البيانات والسجلات الإدارية ذات الصلة.

البحوث المقترحة

استكمالاً لما توصلت إليه الدراسة الحالية، ومعالجةً لبعض حدودها المنهجية، تقترح الدراسة إجراء البحوث المستقبلية الآتية:

1. إجراء دراسة تعتمد تصميماً ذا أربع مجموعات متقاطعة (ذكور–طب وعلوم صحية، إناث–طب وعلوم صحية، ذكور–تمريض، إناث–تمريض)، بهدف الفصل الكمي بين أثر متغير الجنس وأثر متغير الكلية في تفسير الأداء التفاضلي للفقرات.
2. إجراء دراسة تتضمن أداة مستقلة لقياس الكفاءة الذاتية السريرية أو القلق الأدائي (Clinical Self-Efficacy)، تُطبّق بالتوازي مع استبانة قياس الرأي، بهدف اختبار العلاقة المحتملة بين تقديرات الطلبة لذواتهم واتجاهاتهم نحو بيئة التدريب السريري، وفحص الفروق المحتملة بين الجنسين.
3. تكرار الدراسة الحالية في مؤسسات تعليم صحي أخرى تختلف عن جامعة حضرموت من حيث تنظيم بيئة التدريب السريري ودرجة الفصل بين الجنسين، بهدف تعزيز إمكانية تعميم النتائج.
4. إجراء دراسات مقارنة تطبّق نماذج أخرى من نظرية الاستجابة للفقرة على البيانات نفسها، مثل نموذج الاستجابة المتدرجة أو نموذج التقدير الجزئي المعمّم، بهدف مقارنة دقة التقدير وكفاءة النماذج.
5. دراسة الأداء التفاضلي لفقرات الاستبانة في ضوء متغيرات تصنيفية إضافية لم تتناولها الدراسة الحالية، مثل المستوى الدراسي، ونوع المستشفى التدريبي وموقعه، والمعدل التراكمي للطلبة.

قائمة المراجع

اولاً: المراجع العربية

الثبيتي، ساعد بن سعيد. (2016). تدريج فقرات مقياس اتجاهات طلبة جامعة أم القرى نحو الإشراف الأكاديمي وفق نظرية الاستجابة للمفردة. *مجلة البحث العلمي في التربية، 5*(17)، 43-63. https://doi.org/10.21608/jsre.2017.10380

حمدان، عادل سمير محمد. (2025). مقارنة بين نموذجي «التقدير الجزئي» و«سلم التقدير» لمقياس الكفاءة الذاتية في الإدارة الصفية للطالب المعلم. *مجلة العلوم التربوية، جامعة قطر، 25*(1)، 151-176. https://doi.org/10.29117/jes.2025.0213

الدوسري، راشد حماد. (2018). استخدام طريقتي مانتل-هانزل والانحدار اللوجستي للكشف عن الأداء التفاضلي لفقرات اختبار في الرياضيات. *مجلة كلية التربية، 70*(2)، 241-264. https://doi.org/10.21608/MKMGT.2018.132907

علام، صلاح الدين محمود. (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. دار الفكر العربي.

العناتي، جهاد محمد. (2017). الأداء التفاضلي لفقرات اختبار الذكاء غير اللفظي TONI-4 وفقا لجنس الطالب ودخل الأسرة والمستوى التعليمي للأم. *مجلة دراسات تربوية*، 60، 60-72. https://doi.org/10.34118/0136-000-060-005

القرشي، خديجة ضيف الله. (2016). استخدام نموذج التقدير الجزئي في انتقاء فقرات مقياس اتجاهات طلبة الدراسات العليا نحو برنامج SPSS في منطقة مكة المكرمة. *المجلة الدولية التربوية المتخصصة، 5*(9)، 529-552. https://doi.org/10.12816/0036004

المسكري، هلال؛ المحرزي، راشد؛ وحسن، عبد الحميد. (2021). الأداء التفاضلي لمفردات اختبار القدرة اللفظية في مقياس الخليج للقدرات العقلية المتعددة لطلبة دول مجلس التعاون الخليجي تبعًا لمتغيري الجنس والدولة. *مجلة الدراسات النفسية والتربوية، 15*(1)، 120-137. https://doi.org/10.53543/jeps.vol15iss1pp120-137

الوابل، وديان؛ وتعلب، صبرين صلاح. (2023). الأداء التفاضلي لفقرات اختبار الذكاء الثلاثي تبعًا لمتغيري الجنس والتخصص لدى طلاب جامعة القصيم. *مجلة الآداب للدراسات النفسية والتربوية، 5*(1)، 59-102. https://doi.org/10.53285/artsep.v5i1.1378

Arabic References

Al-Thubaytī, Sāʿid ibn Saʿīd. (2016). Tadrīj faqārāt miqyās ittijāhāt ṭalabat Jāmiʿat Umm al-Qurá naḥw al-ishrāf al-akādīmī wafqan li-naẓariyyat al-istijābah lil-mufradah. *Majallat al-Baḥth al-ʿilmī fī al-Tarbiyah, 17*(5), 43–63. https://doi.org/10.21608/jsre.2017.10380

Ḥamdān, ʿĀdil Samīr Muḥammad. (2025). Muqāranah bayna namūdhajay «al-taqdīr al-juzʾī» wa«sullam al-taqdīr» li-miqyās al-kafāʾah al-dhātiyyah fī al-idārah al-ṣaffiyyah lil-ṭālib al-muʿallim. *Majallat al-ʿUlūm al-Tarbawiyyah, Jāmiʿat Qaṭar, 25*(1), 151–176. https://doi.org/10.29117/jes.2025.0213

Al-Dawsarī, Rāshid Ḥammād. (2018). Istikhdām ṭarīqatay Mantel-Haenszel wa-al-inḥidār al-lūjistī li-l-kashf ʿan al-adāʾ al-tafāḍulī li-faqārāt ikhtibār fī al-riyāḍiyyāt. *Majallat Kulliyyat al-Tarbiyah, 70*(2), 241–264. https://doi.org/10.21608/MKMGT.2018.132907

ʿAllām, Ṣalāḥ al-Dīn Maḥmūd. (2005). *Namādhij al-istijābah lil-mufradah al-ikhtibāriyyah uḥādiyyat al-buʿd wa-mutaʿaddidat al-abʿād wa-taṭbīqātuhā fī al-qiyās al-nafsī wa-al-tarbawī*. Dār al-Fikr al-ʿArabī.

Al-ʿAnātī, Jihād Muḥammad. (2017). Al-adāʾ al-tafāḍulī li-faqārāt ikhtibār al-dhakāʾ ghayr al-lafẓī 4-TONI wafqan li-jins al-ṭālib wa-dakhl al-usrah wa-al-mustawá al-taʿlīmī lil-umm. *Majallat Dirāsāt Tarbawiyyah*, 60, 60–72. https://doi.org/10.34118/0136-000-060-005

Al-Qurashī, Khadījah Ḍayf Allāh. (2016). Istikhdām namūdhaj al-taqdīr al-juzʾī fī intiqāʾ faqārāt miqyās ittijāhāt ṭalabat al-dirāsāt al-ʿulyā naḥw barnāmaj SPSS fī minṭaqat Makkah al-Mukarramah. *Al-Majallah al-Dawliyyah al-Tarbawiyyah al-Mutakhaṣṣiṣah, 5*(9), 529–552. https://doi.org/10.12816/0036004

Al-Maskarī, Hilāl; al-Muḥrizī, Rāshid; wa-Ḥasan, ʿAbd al-Ḥamīd. (2021). Al-adāʾ al-tafāḍulī li-mufradāt ikhtibār al-qudrah al-lafẓiyyah fī miqyās al-Khalīj lil-qudarāt al-ʿaqliyyah al-mutaʿaddidah li-ṭalabat Duwal Majlis al-Taʿāwun al-Khalījī tabaʿan li-mutaghayyiray al-jins wa-al-dawlah. *Majallat al-Dirāsāt al-Nafsiyyah wa-al-Tarbawiyyah, 15*(1), 120–137. https://doi.org/10.53543/jeps.vol15iss1pp120-137

Al-Wābil, Widyān; wa-Taʿlab, Ṣabrīn Ṣalāḥ. (2023). Al-adāʾ al-tafāḍulī li-faqārāt ikhtibār al-dhakāʾ al-thulāthī tabaʿan li-mutaghayyiray al-jins wa-al-takhaṣṣuṣ ladá ṭullāb Jāmiʿat al-Qaṣīm. *Majallat al-Ādāb lil-Dirāsāt al-Nafsiyyah wa-al-Tarbawiyyah, 5*(1), 59–102. https://doi.org/10.53285/artsep.v5i1.1378

ثانياً: المراجع الإنجليزية

Alanazi, N. H. (2025). Clinical Learning Environment, Supervision, and Nurse Teacher Evaluation (CLES + T) Scale: Saudi Arabic version—A cross-sectional validation study of nursing interns and students. *Journal of Nursing Management*, 2025, Article 4483497. https://doi.org/10.1155/jonm/4483497

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573. https://doi.org/10.1007/BF02293814

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). Marcel Dekker.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Boyd, M. P., Callaghan, J. E., Knight, C. J., & Crocker, P. R. E. (2014). Identifying differential item functioning of rating scale items with the Rasch model: An introduction and an application. *Measurement in Physical Education and Exercise Science, 18*(2), 100-115. https://doi.org/10.1080/1091367X.2014.893089

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage Publications.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chan, D. S. K. (2002). Associations between student learning outcomes from their clinical placement and their perceptions of the social climate of the clinical learning environment. *International Journal of Nursing Studies, 39*(5), 517–524. https://doi.org/10.1016/S0020-7489(01)00057-8

Chang, H.-H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement, 33*(3), 333–353. https://doi.org/10.1111/j.1745-3984.1996.tb00496.x

DeMars, C. E. (2005). Type I error rates for PARSCALE's fit index. *Educational and Psychological Measurement, 65*(1), 42–50. https://doi.org/10.1177/0013164404264849

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fan, C.-W., Chang, K.-C., Lee, K.-Y., Yang, W.-C., Pakpour, A. H., Potenza, M. N., & Lin, C.-Y. (2022). Rasch modeling and differential item functioning of the self-stigma scale-short version among people with three different psychiatric disorders. *International Journal of Environmental Research and Public Health, 19*(14), Article 8843. https://doi.org/10.3390/ijerph19148843

Guilford, J. P., & Fruchter, B. (1973). Fundamental statistics in psychology and education (5th ed.). McGraw-Hill.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer-Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Lawrence Erlbaum Associates.

Hur, Y., & Seo, D. G. (2023). Development of a character qualities test for medical students in Korea using polytomous item response theory and factor analysis: A preliminary scale development study. *Journal of Educational Evaluation for Health Professions, 20*, 20. https://doi.org/10.3352/jeehp.2023.20.20

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36. https://doi.org/10.1007/BF02291575

Keeves, J. P., & Alagumalai, S. (1999). New approaches to measurement. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 23–42). Pergamon.

Kim, S.-H., & Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's area measures, and the likelihood ratio test on detection of differential item functioning. *Applied Measurement in Education, 8*(4), 291–312. https://doi.org/10.1207/s15324818ame0804_2

Kim, S.-H., & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement, 22*(4), 345–355. https://doi.org/10.1177/014662169802200403

Lawson, D. M., & Brailovsky, C. (2006). The presence and impact of local item dependence on objective structured clinical examinations scores and the potential use of the polytomous, many-facet Rasch model. *Journal of Manipulative and Physiological Therapeutics, 29*(8), 651–657. https://doi.org/10.1016/j.jmpt.2006.08.002

Leis, M., Schmidt, K., & Kaufman, S. (2015). Using the partial credit model to evaluate the student engagement in mathematics scale. *Journal of Applied Measurement, 16*(3), 251-267.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement, 2*(3), 266-283.

Linacre, J. M. (2012). *A user's guide to Winsteps Rasch-model computer programs*. Winsteps.com.

Linacre, J. M. (2024). *Winsteps® Rasch measurement computer program user's guide*. Winsteps.com.

Liu, Y., Lin, Z., & Wang, X. (2023). Gender-based differential item function for the positive and negative semantic dimensions of the relationship satisfaction scale with item response theory. *Behavioral Sciences, 13*(10), 825. https://doi.org/10.3390/bs13100825

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. https://doi.org/10.1007/BF02296272

McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability theory. Educational and Psychological Measurement, 38(1), 75–79. https://doi.org/10.1177/001316447803800111

McDonald, R. P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). American Council on Education/Macmillan.

Na, M., Gross, A. L., & West, K. P., Jr. (2015). Validation of the food access survey tool to assess household food insecurity in rural Bangladesh. *BMC Public Health, 15*, 863. https://doi.org/10.1186/s12889-015-2215-z

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19*(4), 353–368. https://doi.org/10.1177/014662169501900405

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4(3), 207–230. https://doi.org/10.3102/10769986004003207

Roff, S., McAleer, S., Harden, R. M., Al-Qahtani, M., Ahmed, A. U., Deza, H., Groenen, G., & Primparyon, P. (1997). Development and validation of the Dundee Ready Education Environment Measure (DREEM). *Medical Teacher, 19*(4), 295-299. https://doi.org/10.3109/01421599709034208

Saarikoski, M., & Leino-Kilpi, H. (2002). The clinical learning environment and supervision by staff nurses: Developing the instrument. *International Journal of Nursing Studies, 39*(3), 259–267. https://doi.org/10.1016/S0020-7489(01)00031-1

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231.

Strand, P., Sjöborg, K., Stalmeijer, R., & Wichmann-Hansen, G. (2013). Development and psychometric evaluation of the Undergraduate Clinical Education Environment Measure (UCEEM). *BMC Medical Education, 13*, 117. https://doi.org/10.1186/1472-6920-13-117

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147–169). Lawrence Erlbaum Associates.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense (Canada).

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*(2), 223-233. https://doi.org/10.1080/15434300701375832