مجلة الآداب
للدراسات النفسية والتربوية

OPEN ACCESS
تاريخ الاستلام: 28/ 09/ 2025م
تاريخ القبـول: 25/ 10/ 2025م

# استخدام أنموذج سلم التقدير لأندريش المنبثق عن أنموذج راش وفق نظرية الاستجابة للفقرة لبناء مقياس تقويم جودة البرامج الأكاديمية من قبل الطلبة في جامعة القصيم

د. خالد محمد الحسن الحياصات *
km.alheasat@qu.edu.sa

**الملخص:**

هدفت الدراسة إلى بناء مقياس تقويم جودة البرامج الأكاديمية من قبل طلبة الجامعة، ولتحقيق هدف الدراسة تم بناء الصورة الأولية للمقياس من 46 فقرة وفق سلم ليكرت الخماسي، وبعد تطبيق المقياس على عينة استطلاعية، وتحكيمه أصبح عدد فقرات المقياس 40 فقرة، وقد طبق المقياس على عينة مكونة من (5835) طالباً وطالبة، أشارت النتائج إلى مطابقة 35 فقرة لافتراضات نموذج سلم التقدير لأندريش المنبثق عن نموذج راش، ذي المعلمة الواحدة، كأحد نماذج نظرية الاستجابة للفقرة الملائمة للمقياس المستخدم، وإن المقياس يتصف بأحادية البعد، كما تمتع المقياس بصورته النهائية 35 فقرة بخصائص سيكومترية مناسبة، إذ بلغت قيمة معامل الثبات لفقرات المقياس 0.99، ومعامل الثبات للأفراد 0.96، كما تمتع المقياس بدلالات متعددة للصدق.

**الكلمات المفتاحية:** الجودة، تقويم البرنامج، أنموذج راش، أنموذج أندريش، نظرية الاستجابة للفقرة.

* استاذ القياس والتقويم المساعد، قسم علم النفس، كلية اللغات والعلوم الإنسانية، جامعة القصيم، السعودية

**للاقتباس:** الحياصات، خالد محمد الحسن. (2025). استخدام أنموذج سلم التقدير لأندريش المنبثق عن أنموذج راش وفق نظرية الاستجابة للفقرة لبناء مقياس تقويم جودة البرامج الأكاديمية من قبل الطلبة في جامعة القصيم ، *مجلة الآداب للدراسات النفسية والتربوية*، 7(4)، 216-246.

© 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
https://doi.org/10.53285/artsep.v7i4.2970

OPEN ACCESS
Received: 28-09-2025
Accepted: 25-10-2025

مجلة الآداب
للدراسات النفسية والتربوية

# Using Andrich's Rating Scale Model Derived from the Rasch Model within Item Response Theory to Develop a Scale for Evaluating the Quality of Academic Programs by Students at Qassim University

**Khaled Mohammed Alhasan Alheasat** *
km.alheasat@qu.edu.sa

**Abstract:**

This study aimed to develop a scale for evaluating the quality of academic programs from the perspective of university students. For the purpose of the study, an initial version of the scale consisting of 46 items was constructed using a five-point Likert scale. After administering the scale to a pilot sample and subjecting it to expert review, the number of items was reduced to 40. The scale was then administered to a sample of 5,835 male and female students. The results indicated that 35 items conformed to the assumptions of Andrich's Rating Scale Model, which is derived from the one-parameter Rasch model, one of the Item Response Theory models appropriate for the scale used. The scale demonstrated unidimensionality, and its final version of 35 items showed suitable psychometric properties. The item reliability coefficient reached 0.99, and the person reliability coefficient reached 0.96. The scale also demonstrated multiple forms of validity.

**Keywords:** Quality, Program Evaluation, Rasch Model, Andrich Model, Item Response Theory.

* Assistant Professor of Measurement and Evaluation, Department of Psychology, College of Languages and Humanities, Qassim University, Saudi Arabia.

**Cite this article as:**: Alheasat, Khaled Mohammed Alhasan. (2025). Using Andrich's Rating Scale Model Derived from the Rasch Model within Item Response Theory to Develop a Scale for Evaluating the Quality of Academic Programs by Students at Qassim University. *Journal of Arts for Psychological & Educational Studies* 7(4) 216-246

© 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:**

تُعد جودة البرامج الأكاديمية من الركائز الأساسية لتحقيق التميز في مؤسسات التعليم العالي، إذ تمثل مؤشراً جوهرياً على كفاءة التعليم وملاءمته لمتطلبات سوق العمل والمعايير العالمية، وتنعكس هذه الجودة بشكل مباشر على مخرجات التعليم ومدى توافقها مع متطلبات سوق العمل والمجتمع (هيئة تقويم التعليم والتدريب، 2018؛ Harvey & Green, 1993). وفي ضوء هذا التوجه، أولت جامعة القصيم اهتماماً بالغاً بجودة برامجها الأكاديمية، سعياً منها لتحقيق التميز الأكاديمي وضمان مخرجات تعليمية قادرة على المنافسة محلياً ودولياً، وذلك ضمن رؤيتها الرامية إلى تحقيق الريادة في مجالات التعليم، والبحث العلمي، وخدمة المجتمع.

تكمن أهمية ضمان جودة البرامج الأكاديمية في كونها تُسهم في تحسين الأداء المؤسسي، ورفع كفاءة الخريجين، وتدعيم مكانة الجامعة في التصنيفات الأكاديمية، كما تلعب دوراً جوهرياً في عمليات الاعتماد الأكاديمي، سواء على المستوى المحلي عبر هيئة تقويم التعليم والتدريب في السعودية، أو على المستوى الدولي عبر هيئات الاعتماد المتخصصة مثل ABET أو AACSB (NCAAA, 2022).

وفي إطار عمليات التقويم المستمر والتحسين المستدام، يبرز دور الطالب كعنصر محوري في منظومة الجودة، إذ يُمثل المستفيد الأول من الخدمات التعليمية، والأقدر على تقديم تقييم واقعي وشامل لتجربته التعليمية، لذلك، أصبح تضمين آراء الطلبة وملاحظاتهم في تقويم البرامج من الممارسات المعتمدة عالمياً لضمان الشفافية والمصداقية في عمليات التحسين الأكاديمي (Zhao & Kuh, 2004).

إن إشراك الطالب في تقييم جودة البرامج الأكاديمية لا يُعزز فقط من جودة مخرجات التعليم، بل يُنمّي لديه - أيضاً - روح المسؤولية والمشاركة الفاعلة، ويعكس مدى نضج المؤسسة التعليمية في تبني نهج تشاركي يقوم على الاستماع لكافة أصحاب العلاقة (Biggs & Tang, 2011). وتشير الدراسات إلى أن مشاركة الطلبة في تقييم البرامج تُعزز من الشفافية داخل المؤسسات

الأكاديمية، وتسهم في توجيه عمليات التطوير واتخاذ القرار، وتمنح الطلبة دوراً فاعلاً في تحسين البيئة التعليمية. ومن هنا تبرز الحاجة إلى أدوات علمية دقيقة تعتمد على نماذج قياس متقدمة لضمان جمع بيانات موثوقة ومفيدة في تقويم هذه الجودة، بما ينعكس إيجاباً على تطوير السياسات والبرامج الأكاديمية.

ومع تزايد التوجه نحو تحسين مخرجات التعليم وربطها باحتياجات سوق العمل، أصبحت الحاجة ملحة لتبني أدوات دقيقة وموثوقة في تقييم هذه الجودة، خاصة من منظور المستفيد الأول من العملية التعليمية، وهو الطالب. وبالرغم من الجهود المبذولة على المستوى المؤسسي والوطني، فإن الأدوات المتوفرة في هذا المجال ما تزال محدودة، إذ لا تتوفر مقاييس معيارية موحدة على المستوى المحلي، باستثناء تلك التي طورتها هيئة تقويم التعليم والتدريب في المملكة العربية السعودية. وقد عملت الهيئة على تطوير مجموعة من الأدوات المعتمدة – في صورة استبانات – منها استبانة جودة التعليم العالي، التي تهدف إلى قياس رضا الطلبة عن عدة أبعاد، تشمل: التهيئة الطلابية، جودة العملية التعليمية، جاهزية البيئة التعليمية، والبنية التحتية. ومع ذلك، فإن هذه الأدوات لم تخضع للتحقق الكافي من خصائصها السيكومترية بالشكل الذي يؤهلها لتكون أدوات قياس مقننة ذات موثوقية وصدق مرتفعين.

وقد تم بناء فقرات هذه الأدوات استناداً إلى معايير مشتقة من النظرية الكلاسيكية في القياس. التي تعرضت لانتقادات عدة، أبرزها: اعتماد معالم الفقرة مثل الصعوبة والتمييز والتخمين على عينة الأفراد التي طبق عليها المقياس، واعتماد قدرة الفرد على عينة الفقرات التي استجاب لها المقياس (Hambleton, Swaminathan & Rogers, 1991)، وتساوي أخطاء القياس لجميع الأفراد الذين طبق عليهم المقياس (علام، 2000). نتيجةً لذلك، برزت نظرية الاستجابة للفقرة ( Item Response Theory - IRT) أو ما تسمى نظرية السمات الكامنة (Latent Trait Theory – LTT)، إذ قدمت حلولاً لمعظم الإشكالات التي عجزت النظرية الكلاسيكية عن معالجتها. فهي تهدف إلى تحديد العلاقة بين أداء الفرد في اختبار ما وبين السمات أو القدرات الكامنة وراء هذا الأداء، وبالتالي تستجيب لمتطلبات القياس الموضوعي المتمثل في تحرير تدريج أدوات القياس من خصائص الأفراد،

وتحرير تقديرات قدرات الأفراد من خصائص الفقرات التي استخدمت في هذا القياس ( Hambleton & Swaminathan, 1985)، وبالتالي الحصول على مؤشرات إحصائية للفقرات لا تتأثر بخصائص المفحوصين، كما أن تقديرات قدرات المفحوصين لا تعتمد على مستوى صعوبة الفقرات في المقياس (Crocker & Algina, 1986). ومن هذا المنطلق، تبرز أهمية خاصية اللاتغيّر (Invariance)، التي تتيح تقدير معالم الفقرات بثبات عبر مجموعات مختلفة من الأفراد، وتُعد هذه الخاصية من أبرز وأهم خصائص نظرية الاستجابة للفقرة (Lord, 1980). وقد أصبحت هذه النظرية تمثل الاتجاه السائد حالياً والمستقبلي في مجال اختيار الفقرات وتطوير أدوات القياس (Anastasi, 1982).

وتقوم هذه النظرية على مجموعة من الافتراضات الأساسية: أحادية البعد (Unidimensionality)، والاستقلال الموضعي (Local Independent)، بالإضافة إلى وتَرية منحنى خصائص الفقرة (Item Characteristic Curve)، وقد نتج عن هذه النظرية مجموعة من النماذج التي تُستخدم في تطوير الاختبارات وبناء المقاييس النفسية والتربوية، وكان من أهمها وأكثرها استخداماً نموذج راش (Rasch) ذي المعلمة الواحدة، الذي يلائم تدريج البيانات ثنائية الاستجابة (Dichotomous)، التي تُستخدم عندما تأخذ إجابة الفرد عن الفقرة قيماً ثنائية مثل (0 أو 1)، كما هو الحال في اختبارات الصواب والخطأ. وقد طُوّر من هذا النموذج نماذج متعددة منها: نموذج التقدير الجزئي (Partial Credit)، الذي يستخدم عندما تكون الإجابة عن الفقرة من عدة خطوات (Wright & Masters, 1982)، ونموذج مقياس التقدير (Rating scale Model) الذي طُوّر من قبل أندريش (Andrich, 1978) ليلائم البيانات التي تكون الاستجابة عن فقراتها متدرجة، مثل بيانات مقياس ليكرت، الذي تتراوح الاستجابة عنه بين "موافق بشدة" و "غير مواق بشدة" (Embretson & Reise, 2000) كالبيانات المأخوذة في هذه الدراسة، ويشير (Reeve، 2004) إلى أن نموذج سلم التقدير المنبثق من نموذج "راش" هو حالة خاصة من نموذج التقدير الجزئي ولكنه يختلف عنه في أن المسافة بين صعوبة الخطوة من قسم من أقسام الاستجابة إلى القسم الذي يليه ثابتة عبر جميع الفقرات (علام، 2005؛ De Gruijter & Van der 2008)، بينما في نموذج التقدير الجزئي تكون صعوبة الخطوة من قسم من أقسام الاستجابة إلى القسم الذي يليه غير متساوية عبر الفقرات

(Bjorner,et al., 2004). وللكشف عن مدى ملاءمة البيانات المأخوذة من المقاييس مع نموذج سلم التقدير، يوفر برنامج WINSTEPS، إحصائي المطابقة الكلي، الذي له مؤشران: مؤشر المطابقة الداخلية (INFIT ZSTD) الذي يعبر عنه بمتوسط المربعات التقاربي (INFIT MNSQ)، وهو مؤشر إحصائي حساس للسلوكيات غير المتوقعة التي تؤثر في الاستجابات عن الفقرات التي تكون قريبة من مستوى قدرة الفرد، والمؤشر الآخر هو مؤشر المطابقة الخارجية (Outfit ZSTD) ويعبر عنه بمتوسط المربعات التباعدي (Outfit MNSQ)، وهو مؤشر إحصائي أكثر حساسية للسلوكيات غير المتوقعة التي تؤثر في الاستجابات عن الفقرات التي تكون بعيدة من مستوى قدرة الفرد، وذلك لكل تقدير من تقديرات قدرات الأفراد. كما يمكن التوصل إلى تفسيرات مشابهة لإحصائيات مطابقة الفقرات. وللحكم على مدى مطابقة بيانات الأفراد والفقرات مع نموذج سلم التقدير، يفضل أن تكون قيمة المتوسط الحسابي لمتوسطات المربعات (MNSQ) الداخلية والخارجية قريبة من (1)، والمتوسط الحسابي لمتوسط قيم إحصائي المطابقة الداخلية والخارجية (ZSTD) قريب من (0) (دي أيالا، 2017/2009).

وقد شهدت الدراسات - في السنوات الأخيرة - توسعاً ملحوظاً في استخدام نماذج نظرية الاستجابة للفقرة، خاصة نموذج أندريش (Andrich Rating Scale Model)، نظراً لقدرته على التعامل مع بيانات التقدير المتدرج وخصائصه السيكومترية الدقيقة. ففي الدراسة التي أجراها كل من Oliveira, Fernandez, Sisto (2014) التي هدفت إلى تقييم مفردات قائمة قلق المدرسة، تم استخدام نموذج أندريش في تحليل استجابات (253) طالباً وطالبة تراوحت أعمارهم بين (9 – 12) عاماً، وبينت النتائج ملاءمة فقرات القائمة للنموذج، وعدم وجود فقرات متحيزة للجنس، وتمتعت القائمة بدرجة مرتفعة من الثبات، إذ بلغت قيمته للفقرات (0.98).

وهدفت دراسة الطراونة (2018) إلى تطوير مقياس التدين لدى طلبة الجامعة باستخدام نموذج سلم التقدير لأندريش في نظرية استجابة الفقرة، تكونت عينة الدراسة من (2404) طلاب من الجنسين من طلبة الجامعات الأردنية، وتم التحقق من مطابقة أفراد عينة الدراسة وفقرات المقياس

لافتراضات نموذج سلم التقدير، وقد أظهرت النتائج دلالات صدق وثبات وفق النظرية الحديثة في القياس مقبولة للمقياس.

وأجرى كل من You, Kim, Black, & Min (2018) دراسة هدفت إلى التعرف على مستوى الدافعية نحو العلوم لدى طلبة الجامعات، كما هدفت الدراسة إلى التحقق من صدق وثبات مقياس الدافعية الثاني نحو العلوم باستخدام نموذج أندريش، وتكون المقياس من (25) فقرة موزعة على خمسة عوامل، وتم تطبيق المقياس على (431) من طلبة إحدى الجامعات البحثية غرب الولايات المتحدة الأمريكية، وتوصلت الدراسة إلى أن المقياس استمتع بمعامل ثبات مرتفع لكل من الأفراد والفقرات وهي (0.92) (0.98) على التوالي.

وهدفت دراسة Aiouadj & Ziad, (2023) إلى معايرة فقرات مقياس الاكتئاب والقلق والضغوط (DASS42) باستخدام نموذج مقياس تصنيف أندريش المنبثق من نموذج راش. وتكونت عينة الدراسة من (404) طلاب من طلبة المرحلة الثانوية، وتم تقدير معالم الفقرات والأفراد والخطأ المعياري للقياس لكليهما، إذ أظهرت النتائج وقوع جميع فقرات المقياس (30 فقرة) بعد المعايرة ضمن نطاق مقبول وفق قيم مؤشر المطابقة الداخلية (MNSQ).

ودراسة Cortés-Rodríguez, et al., (2023) التي هدفت إلى بناء مقياس لقياس القدرة الريادية لصغار المزارعين ذوي النطاق الصغير في بلدة سان بابلو هويكستيبك (أواكساكا، المكسيك) باستخدام مقياس تصنيف راش-أندريش، واستخدم مقياس مكون من (17) فقرة، وتم التحقق من افتراضات نظرية الاستجابة للفقرة وحذف الفقرات التي لا تحقق ذلك، بالإضافة إلى تحقق افتراضات وعتبات النموذج، إذ أظهرت النتائج ملاءمة المقياس للنموذج المستخدم.

وقامت الخريف وتعلب (2023) بدراسة هدفت إلى توظيف نموذج سلم التقدير في تقنين مقياس كاليفورنيا للدافعية العقلية من خلال التحقق من تحقيق المقياس لافتراضات نظرية استجابة الفقرة، والتحقق من مطابقة فقرات المقياس لنموذج سلم التقدير، وأظهرت النتائج تحقق افتراضات نظرية استجابة الفقرة في استجابات أفراد العينة، وعدم ملاءمة (3) فقرات من المقياس لنموذج سلم التقدير.

كما أجرى العنزي (2024) دراسة هدفت إلى استخدام نموذج أندريش للاستجابة للمفردة لتدريج مفردات مقياس تقدير الذات، وبينت النتائج تحقق افتراضات النموذج، وتمتع المقياس بمعامل ثبات مرتفع.

وقام السليمي وآخرون (2025) بدراسة هدفت إلى بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية، والتحقق من خصائصه السيكومترية في ضوء نظرية الاستجابة للفقرة وفقاً لافتراضات نموذج أندريش، وقد أظهرت النتائج مطابقة فقرات المقياس لافتراضات نموذج أندريش، كما بينت النتائج أن قيم معاملات الثبات للفقرات والأفراد وكذلك معاملات الفصل للمقياس كانت مرتفعة، وأن المقياس ذو فعالية في تدريج فقراته.

وبما أن الباحث لم يجد أية دراسة لبناء مقياس لتقييم البرامج الأكاديمية من قبل الطلبة باستخدام نظرية الاستجابة للفقرة، فقد جاءت الدراسة الحالية لتستفيد من هذه الجهود العلمية في توظيف نموذج أندريش لتطوير مقياس لتقييم جودة البرامج الأكاديمية من قبل الطلبة في جامعة القصيم، مستندة إلى الأسس المنهجية التي أظهرتها الدراسات السابقة في التحقق من الخصائص السيكومترية، وضمان موضوعية المقاييس التعليمية الحديثة ودقتها.

**مشكلة الدراسة:**

تتمثل مشكلة هذه الدراسة بالحاجة إلى أداة قياس دقيقة لقياس جودة البرامج الأكاديمية لدى طلبة الجامعات، تتفق ومعايير القياس الموضوعي، فضلا عن أن أدوات القياس التي طُوِّرت لتقويم جودة البرامج الأكاديمية لم تخضع لإجراءات كافية للتحقق من خصائصها السيكومترية، إضافة إلى أن ما قُدِّم من خصائص كان مبنياً على أسس النظرية الكلاسيكية في القياس (CTT) التي عجزت عن تفسير العديد من القضايا الجوهرية في القياس النفسي ( Hambleton & Swaminathan, 1985)، وكذلك لم تحظَ نظرية الاستجابة للفقرة بالاهتمام الكافي في الدراسات العربية عند بناء المقاييس عامة، رغم ما تتمتع به من مزايا موضوعية في بناء أدوات القياس، فضلاً عن افتقار الأدب النظري في الدراسات العربية إلى مقياس خاص بقياس جودة البرامج الأكاديمية

لدى طلبة الجامعات، وبالتالي دعم جهود الجامعات نحو تحسين برامجها وتحقيق الاعتماد الأكاديمي لها. لذا تهدف هذه الدراسة إلى تطوير مقياس تقويم جودة البرامج الأكاديمية من قبل طلبة الجامعة وفق أنموذج أندريش أو ما يسمى بأنموذج سلم التقدير Rating Scale Model كأحد النماذج المنبثقة عن نموذج راش Rasch وفق نظرية الاستجابة للفقرة Item Response Theory التي عدّتها أنستازي (Anastasi, 1982) أنها تشكل الإطار الحالي والمستقبلي لتطوير المقاييس، وقد سعت الدراسة الحالية إلى الإجابة عن الأسئلة الآتية:

**السؤال الأول:** ما مدى مطابقة البيانات لاستجابات الأفراد عن فقرات مقياس تقويم جودة البرامج الأكاديمية لنموذج سلم التقدير لأندريش؟

**السؤال الثاني:** هل يحقق تدريج مقياس تقويم جودة البرامج الأكاديمية بنموذج سلم التقدير افتراضات النموذج؟

**السؤال الثالث:** ما دلالات الصدق والثبات لمقياس تقويم جودة البرامج الأكاديمية بعد تدريجه باستخدام نموذج سلم التقدير لأندريش؟

## أهداف الدراسة:

تهدف هذه الدراسة إلى تحقيق مجموعة من الأهداف العلمية والتطبيقية، من أبرزها:

1. تطوير مقياس علمي يعتمد على أنموذج أندريش لقياس جودة البرامج الأكاديمية من وجهة نظر طلبة جامعة القصيم.
2. تحليل خصائص فقرات المقياس باستخدام نظرية استجابة الفقرة، بما يضمن صدق الأداة وثباتها.

## أهمية الدراسة:

تتجلى أهمية هذه الدراسة من جانبين: الأول نظري والثاني تطبيقي؛ فعلى الصعيد النظري تُسهم في إضافة مقياس جديد إلى المكتبة العربية، يقيس تقويم الطلبة لجودة البرامج الأكاديمية؛ يمتاز بالدقة والموضوعية في قياس السمة من خلال إيجاد فقرات متحررة من خصائص الأفراد، وذلك باستخدام أنموذج أندريش (RSM). أما من الناحية التطبيقية، فتبرز أهمية الدراسة في توفير

أداة مقننة وموثوقة يمكن للجامعات وصانعي القرار الاعتماد عليها في تشخيص جوانب القوة والضعف في البرامج الأكاديمية من منظور الطلبة، بما يدعم خطط التحسين المستمر ويعزز من فرص الحصول على الاعتماد الأكاديمي محلياً ودولياً استناداً إلى بيانات علمية دقيقة.

**محددات الدراسة:**

**الحدود الموضوعية:** اقتصرت الدراسة الحالية على استخدامها نموذج أندريش في اشتقاق الخصائص السيكومترية للمقياس.

**الحدود المكانية:** تُجرى هذه الدراسة في جامعة القصيم بالمملكة العربية السعودية، وتشمل عدداً من كلياتها الأكاديمية التي يتم اختيارها لأغراض الدراسة.

**الحدود الزمانية:** تم تطبيق أداة الدراسة خلال الفصل الدراسي الثاني من العام الجامعي 1446هـ.

**الحدود البشرية:** تستهدف الدراسة طلبة جامعة القصيم في مرحلة البكالوريوس، وذلك باعتبارهم الفئة الرئيسة المستهدفة في تقييم البرامج الأكاديمية.

**مصطلحات الدراسة:**

**الجودة (Quality):** مدى تحقيق البرنامج الأكاديمي لأهدافه المعلنة، ووفائه بمعايير الاعتماد المحددة، بما يضمن تحسين مخرجاته التعليمية، ورفع كفاءة أدائه، وتعزيز قدرته التنافسية محلياً ودولياً (هيئة تقويم التعليم والتدريب، 2018).

**تقويم البرنامج (Program Evaluation):** الأساليب المستخدمة للحصول على آراء المستفيدين من البرنامج، وأيضاً، مؤشرات الأداء، وتشمل: الطلاب، وأعضاء هيئة التدريس بالكلية، والخريجين، وسوق العمل؛ وذلك بهدف تحسين البرنامج التعليمي وتطويره ليتجاوب مع التقدم الذي يطرأ على محتوى المادة واحتياجاتها (أبا بطين وسمور، 2019).

**أنموذج راش (Rasch model):** هو أحد نماذج النظرية الحديثة في القياس، وهو نموذج لوجستي ذو معلمة واحدة، ويعمل على تقدير احتمالية إجابة الفرد عن فقرة ما إجابة صحيحة، بدلالة قدرته ($\theta$) ، ومعامل صعوبة الفقرة (bi)، بغض النظر عن عدد الفقرات وحجم العينة. وقد اقترحه جورج راش، ويطلق عليه النموذج أحادي المعلمة (one-parameter model) ويهتم بتحديد موقع

الفقرة الاختبارية على سلم صعوبة لجميع الفقرات التي تشكل الاختبار، كما يهتم بتدريج مستويات قدرة الفرد باختبار معين على نفس متصل الفقرات (Allen & Yen, 2002).

**أنموذج أندريش: نموذج سلم التقدير (Model Rating Scale):** هو أحد نماذج نظرية الاستجابة للفقرة أحادية البعد، الذي يستخدم لتحليل الاستجابات التي تتكون من سلسلة من الفئات الترتيبية، التي يفصلها عدد من العتبات (Reeve, 2004).

**ويُعرّفه الباحث إجرائياً:** نموذج رياضي من نماذج نظرية الاستجابة للفقرة المنبثق من نموذج راش (نموذج المعلمة الواحدة)، ويأخذ شكل استجابات متعددة بتدريجات تفصل بينها مسافات متساوية لمقياس تقييم جودة البرنامج الأكاديمي.

## منهج الدراسة وإجراءاتها

### أولاً: منهج الدراسة

بناءً على متغيرات الدراسة وتحقيقاً لأهدافها؛ تم استخدام المنهج الوصفي.

### ثانياً: مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلاب وطالبات جامعة القصيم ممن هم في مرحلة البكالوريوس المنتظمين في الدراسة في الفصل الدراسي الثاني للعام الجامعي 1446هـ والبالغ عددهم (68440) طالباً وطالبة، منهم (30536) طالباً، و (37904) طالبة، وذلك وفق إحصائية مجلس شؤون الجامعات (https://www.cua.gov.sa/).

### ثالثاً: عينة الدراسة

تكونت عينة الدراسة من (5835) طالباً وطالبة، وبنسبة (9%) من مجتمع الدراسة.

### أداة الدراسة:

### مقياس تقويم الطلبة لجودة البرامج الأكاديمية

### خطوات تطوير المقياس:

لأغراض الدراسة الحالية تم بناء مقياس تقويم الطلبة لجودة البرامج الأكاديمية وفق نظرية الاستجابة للفقرة، تبعاً للخطوات التي بينها Huling, Drasgow & Parsons (1983) التي تتلخص بالآتي:

**الخطوة الأولى:** جرى الاطّلاع على الأدبيات النظرية والدراسات السابقة ذات الصلة بتقييم البرامج الأكاديمية (المفيز وآخرون، 2021؛ الشمري، 2019؛ العصيل، 2019؛ الحياصات وبني عامر، 2016؛ أبو هاشم، 2016)، أيضاً تم الاستناد إلى معايير الاعتماد الأكاديمي الصادرة عن المركز الوطني للتقويم والاعتماد الأكاديمي في المملكة العربية السعودية، إضافةً إلى الاستعانة بالمقاييس المعتمدة في استطلاعات الرأي المتعلقة بجودة التعليم الجامعي وبرامجه من قبل هيئة تقويم التعليم والتدريب، وكذلك مقاييس تقويم جودة البرامج الأكاديمية المُعَدّة من قِبل عمادة التطوير والجودة في جامعة القصيم.

**الخطوة الثانية:** تم صياغة (46) فقرة، وتم اعتماد مقياس ليكرت (Likert) في تدريج الفقرات على متصل خماسي يتضمن الفئات: "موافق بشدة (5)، موافق (4)، محايد (3)، غير موافق (2)، غير موافق بشدة (1)".

وتم عرض فقرات المقياس - بصورتها الأولية - البالغة (46) فقرة على (10) محكمين من المتخصصين في القياس والتقويم والجودة والاعتماد الأكاديمي بهدف إبداء آرائهم وملاحظاتهم في ملاءمة الفقرات من حيث صياغتها ومناسبتها لموضوع المقياس، ومدى ملاءمتها للمستجيب، وأسفرت النتائج عن بعض التعديلات في صياغة بعض الفقرات، مع الإبقاء على مضمونها، وحذف الفقرات التي كان هناك توافقاً بنسبة 30% بين المحكمين على عدم مناسبتها، أو تكرارها، وبناء على ذلك فقد تم حذف 6 فقرات ليصبح عدد فقرات المقياس 40 فقرة، شكلت الصورة الأولية للمقياس.

وتم تطبيق المقياس بصورته الأولية على إحدى الشعب الدراسية في الجامعة وتكونت من (62) طالباً وطالبة بمستوى البكالوريوس، ومن غير عينة الدراسة، حيث بين الباحث مدى أهمية الإجابة بجدية، وكذلك طلب منهم إبداء الرأي حول فقرات المقياس من حيث الصياغة ومدى الوضوح، وجمعت استجاباتهم وملاحظاتهم، وتم الاحتفاظ بجميع الفقرات، وتم تقدير معامل ثبات الاتساق الداخلي باستخدام معادلة كرونباخ ألفا، وقد بلغت قيمته (0.89)، ويُعد ذلك مؤشراً مقبولاً للسير بإجراءات بناء المقياس.

**الخطوة الثالثة:** التحقق من افتراض أحادية البعد للبيانات المتحققة بالمقياس: تم تطبيق المقياس المتضمن (40) فقرة من خلال توزيع الاستبانة إلكترونياً عن طريق عمادة التطوير والجودة في الجامعة، إذ بلغ عدد الردود على المقياس (5835) طالباً وطالبة من طلبة الجامعة في مرحلة البكالوريوس.

وللتحقق من افتراض أحادية البعد تم إدخال البيانات في ذاكرة الحاسوب بهدف إجراء تحليل لتلك البيانات، إذ تفترض نظرية الاستجابة للفقرة وجود قدرة واحدة تفسر أداء الفرد في المقياس، وللتحقق من هذا الافتراض استخدم الباحث أسلوب التحليل العاملي (FA) بطريقة المكونات الرئيسية (PCA)، وذلك لاستجابات الأفراد البالغ عددهم (5835) طالباً وطالبة على فقرات مقياس تقييم جودة البرنامج الأكاديمي بصورته الأولية، إذ تم حساب قيمة الجذر الكامن Eigen Value ونسبة التباين المفسّر Variance Explained لكل عامل من العوامل باستخدام البرنامج الإحصائي SPSS v27، وجرى التدوير باستخدام طريقة التدوير المتعامد Varimax Rotation للعوامل، التي كانت قيمة الجذر الكامن لها أكبر من واحد، ويبين الجدول (1) نتائج هذا التحليل.

**جدول (1)**

***نتائج تحليل المكونات الأساسية لمقياس تقييم جودة البرامج الأكاديمية بالصورة الأولية***

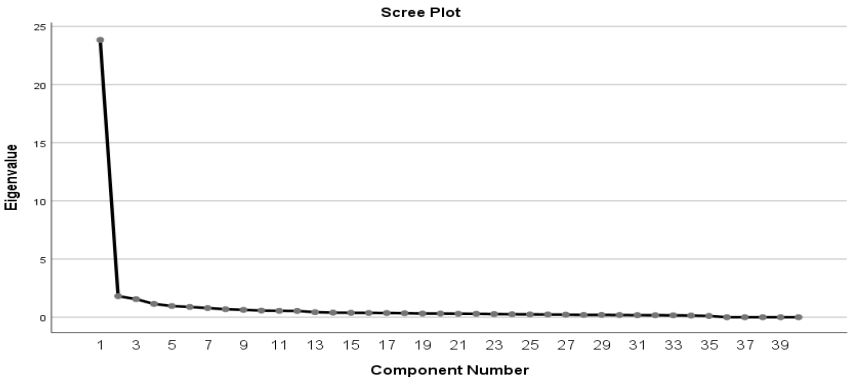
| العامل | الجذر الكامن | نسبة التباين المفسر % | نسبة التباين التراكمية % |
|---|---|---|---|
| **الأول** | 23.84 | 59.6 | 59.158 |
| **الثاني** | 1.810 | 4.525 | 63.629 |
| **الثالث** | 1.556 | 3.890 | 67.732 |
| **الرابع** | 1.142 | 2.855 | 70.870 |

يلاحظ من الجدول رقم (1) أن هناك (4) عوامل زادت قيمة الجذر الكامن لها عن (1)، وهي تفسر ما نسبته 70.87% من تباين درجات المقياس بصورته الأولية وذلك قبل حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج المستخدم، وتبين النتائج أن نسبة التباين المفسر للعامل الأول كانت مرتفعة، إذ بلغت قيمتها (59.6%) وهي أكبر من (20%) مما يدل على تحقق افتراض أحادية البعد في بيانات المقياس (Hattie, 1985)، وباعتماد محك الجذر الكامن بوصفه مؤشراً على

أحادية البعد، فقد بلغت نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر بكثير من 2، وهذا مؤشر على أحادية البعد (Hambleton & Swaminthan, 1985; Hattie, 1985)، ويمكن تأكيد ذلك من خلال ملاحظة انفراد العامل الأول بجذر كامن كبير مقارنة بالعوامل الأخرى الموضحة في التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للمقياس الكون من (40) فقرة أو ما يعرف بمخطط سكري (Scree Plot) (أنظر الشكر (1)).

**الشكل (1)**

***التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة لمقياس تقويم جودة البرامج الأكاديمية***




أيضاً تم تقدير ثبات الاتساق الداخلي باستخدام معادلة كرونباخ ألفا (Cronbach's Alpha) على عينة الدراسة البالغ عددها (5835) طالباً وطالبة، وبلغت قيمته (0.98)، وهو مؤشر قوي على أحادية البعد (Hattie, 1985).

**الخطوة الرابعة: اختيار النموذج الملائم:** يُعد اختيار النموذج الملائم خطوة محورية في بناء المقاييس، إذ أشار كل من Huling, Drasgow & Parsons (1983) إلى أن هذه الخطوة هي من أهم وأدق المراحل بعد التحقق من افتراض أحادية البعد، والنموذج الذي تم اختياره ليناسب هذه البيانات هو أنموذج أندريش المنبثق من أنموذج راش، وتحديداً نموذج سلم التقدير (Rating Scale Model)، وذلك لعدة اعتبارات؛ من أبرزها: كونه أحد النماذج اللوجستية التي تتميز بسهولتها الرياضية، وعدم تأثره بالاستجابات غير الجادة من قبل المستجيبين، بالإضافة إلى ملاءمته لبيانات المقياس المبنية وفق تدريج ليكرت الخماسي، فضلاً عن توفر برامج إحصائية متخصصة تدعم تحليله مثل برنامج (Winsteps v 5.8.5.0).

**الخطوة الخامسة:** تحليل بيانات المقياس: بعد اختيار النموذج الملائم لبيانات مقياس تقويم جودة البرامج الأكاديمية وهو نموذج سلم التقدير لأندرش، تم إدخال استجابات (5835) طالباً وطالبة على فقرات المقياس البالغ عددها (40) فقرة في ذاكرة الحاسوب، وأخضعت البيانات للتحليل باستخدام البرنامج الإحصائي (Winsteps v5.8.5.0) بهدف التحقق من درجة مطابقة الاستجابات على فقرات المقياس.

**الأساليب الإحصائية**

1. تحليل البيانات باستخدام نموذج سلم التقدير (RSM)
2. تم التحقق من افتراضات نموذج سلم التقدير باستخدام برنامج Winsteps v5.8.5.0 من خلال طريقة المكونات الرئيسة للبواقي المعيارية Rasch Principal Components Analysis of Item Residuals (Rasch PCAR)، وأيضاً حساب ارتباطات البواقي بين أزواج الفقرات.
3. تم حساب معامل الثبات للمقياس من خلال حساب معامل الفصل بين الفقرات، أيضاً معامل الفصل بين الأفراد.

**نتائج الدراسة وتفسيراتها**

للإجابة عن السؤال الأول الذي ينص على: "ما مدى مطابقة البيانات لاستجابات الأفراد عن فقرات مقياس تقويم جودة البرامج الأكاديمية لنموذج سلم التقدير لأندريش؟"، قام الباحث بتحليل البيانات باستخدام البرنامج الإحصائي (Winsteps v5.8.5.0)، وذلك للتحقق والحكم على مدى مطابقة الاستجابات على فقرات المقياس، وتم التحقق من عدم وجود استجابات ذات نمطية واحدة للأفراد، وكذلك عدم وجود فقرات تتشابه الاستجابات عليها، وبالتالي لم يتم اسقاط أي من الاستبانات، أو حذف فقرات، كما تم تصحيح استجابات الطلبة في المقياس وفق تدريج ليكرت الخماسي، إذ أعطيت الاستجابات للفقرات الدرجات 5 (موافق بشدة)، 4 (موافق)، 3 (محايد)، 2 (غير موافق)، 1 (غير موافق بشدة)، وقد بلغت الاستجابات المدخلة (5835) طالباً وطالبة، وبلغ عدد الفقرات (40) فقرة، وتم التحقق من مطابقة الأفراد والفقرات للنموذج على النحو الآتي:

- تم التحقق من مطابقة الأفراد للنموذج لحذف الأفراد غير المطابقين لنموذج سلم التقدير، وتم استخدام مؤشرات المطابقة الداخلية (INFIT)، التي تتمثل بمتوسط المربعات

(MNSQ) وإحصائي المطابقة المعيارية (ZSTD)، وهو مؤشر إحصائي للسلوكات غير المتوقعة التي تؤثر في استجابات الأفراد على الفقرات ذات مستوى صعوبة قريب من مستوى قدرتهم، وكذلك إحصائيا المطابقة الخارجية (OUTFIT) وهي: (MNSQ, ZSTD)، ويعد هذا المؤشر بديلاً عن السابق ولكن له صفات قريبة ومشابهة، ولكنه أكثر حساسية للسلوكات غير المتوقعة من الأفراد عن الفقرات التي تبعد عن مستوى قدرتهم، والجدول (2) يبين ملخصاً لتقديرات القدرة للأفراد.

**جدول (2)**

***الوسط الحسابي والانحراف المعياري لمعلم القدرة وإحصاءات المطابقة الداخلية والخارجية لاستجابات الأفراد على المقياس***

| الإحصائي | القدرة | إحصائي المطابقة الداخلية (INFIT) | | إحصائي المطابقة الخارجية (OUTFIT) | |
|---|---|---|---|---|---|
| | | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD |
| الوسط الحسابي | -0.13 | 1.01 | 0.04 | 1.02 | 0.02 |
| الانحراف المعياري | 0.73 | 0.29 | 1.22 | 0.28 | 1.20 |

يتضح من الجدول (2) أن الوسط الحسابي لأوساط المربعات (MNSQ) الداخلية والخارجية كانت قيمها 1.01، 1.02 على الترتيب وهي قريبة جداً من الواحد، وهو الوضع المثالي، أما المتوسط الحسابي لقيم إحصائي المطابقة الداخلية والخارجية (ZSTD) فهي تقترب من الصفر فقد بلغت قيمها 0.04 و 0.02 على التوالي، والانحراف المعياري لها؛ بعيد إلى حد ما عن الواحد، لأن القيم التي يفترضها النموذج (0، 1)، وسبب ابتعاد هذه القيم عن الوضع المثالي هو الأفراد غير المطابقين للنموذج، وبالرجوع إلى تقديرات القدرة للأقراد، ولدى تفحص إحصائيات المطابقة الداخلية والخارجية للأفراد، إذ يشيران فيما إذا كانت قدرة الفرد متطابقة مع قدرات مجموعة الأفراد التي ينتمي إليها في قياس السمة التي يقيسها المقياس، إذ يشترط أن تبلغ قيمها في المدى (+2 إلى -2)، (Julin, 1988; Alastair & Hutchinson, 1987)، وهي القيم التي يتوقعها النموذج، كما تشير القيم خارج ±1.96 إلى أنه $\alpha < 0.05$، أي وجود دلالة، بالتالي رفض الفرضية الصفرية لملاءمة البيانات مع نموذج سلم التقدير (Linacre, 2014; Oktaviyanthi et al,2024).

وبناء على تلك النتائج تبين أن هناك (582) فرداً تبتعد استجاباتهم الملحوظة عن الاستجابات المتوقعة تبعاً لقدراتهم، بمعنى أن قيم إحصائي المطابقة الخارجية المقابلة لدرجاتهم تزيد على (+2) أو تقل عن (-2)، أو قيم متوسطات المربعات المناظرة لهذه الدرجات تزيد على 1، وهي القيم التي يتوقعها النموذج (Wright & Ston, 1979) ويعد هؤلاء الأفراد غير مطابقين، لأن استجاباتهم تبتعد عن توقعات النموذج.

بعد أن تم استبعاد الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج، تم إعادة التحليل للتحقق من مدى مطابقة الفقرات للنموذج، فقد تم تقدير معلمة الصعوبة لكل فقرة، والخطأ المعياري في تقديرها، وقيم إحصائيات كل من المطابقة الخارجية والداخلية لكل معلم من معالم الصعوبة للفقرات، ويبين الجدول (3) ملخصاً لتلك النتائج.

**جدول (3)**

***الوسط الحسابي والانحراف المعياري لمعالم الصعوبة وإحصاءات المطابقة الداخلية والخارجية لمطابقة الفقرات للنموذج***

| الإحصائي | القدرة | إحصائي المطابقة الداخلية (INFIT) | | إحصائي المطابقة الخارجية (OUTFIT) | |
|---|---|---|---|---|---|
| | | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD |
| الوسط الحسابي | 0.0 | 1.00 | -7.4 | 1.02 | -7.4 |
| الانحراف المعياري | 0.32 | 0.50 | 6.5 | 0.56 | 6.5 |

يتبين من الجدول رقم (3) اقتراب المتوسط الحسابي لمتوسط المربعات الداخلية والخارجية من الوضع المثالي المتوقع من نموذج سلم التقدير وهو 1، إذ بلغت قيم متوسطات المربعات الداخلية والخارجية (MSNQ) (1، 1.02) بانحراف معياري (0.50، 0.56) على التوالي، وكذلك بلغ المتوسط الحسابي لقيم إحصائيات المطابقة الداخلية والخارجية (ZSTD) (-7.4، -7.4) بانحراف معياري (6.5، 6.5)، ويلاحظ عدم اقتراب هذه الإحصائيات من الوضع المثالي الذي يتوقعه النموذج، وهو اقترابه من (0،1).

ولدى تفحص قيم إحصائي المطابقة الداخلية والخارجية للفقرات، الذي يشير إلى مدى استقرار مستوى الصعوبة النسبي للفقرات عبر مستويات القدرة المختلفة، فقد تبين وجود (5) فقرات غير مطابقة للنموذج وبعيدة إلى حد ما عن توقعات النموذج.

بعد حذف الأفراد غير المطابقين للنموذج، وإعادة التحليل للكشف عن الفقرات غير المطابقة للنموذج تم حذف الفقرات غير المطابقة للنموذج، وأعيد التحليل للمرة الثالثة بعد حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج، وذلك للتحقق من موضوعية تفسير نتائج المقياس بصورته النهائية (35) فقرة، وتحقيقاً لمدى صدق النموذج في تحقيقه لموضوعية القياس، والجدول (4) يلخص نتائج التحليل للقيم المتحررة لقدرات الأفراد.

**جدول (4)**

*نتائج التحليل للقيم المتحررة لقدرات الأفراد (عدد الأفراد = 5253) (عدد الفقرات = 35)*

| الإحصائي | القدرة | إحصائي المطابقة الداخلية (INFIT) | | إحصائي المطابقة الخارجية (OUTFIT) | |
|---|---|---|---|---|---|
| | | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD |
| الوسط الحسابي | -0.16 | 0.99 | 0.0 | 0.99 | 0.0 |
| الانحراف المعياري | 1.03 | 0.02 | 1.0 | 0.02 | 1.0 |
| الخطا المعياري في التقدير | 0.01 | 0.0 | 0.01 | 0.0 | 0.01 |

يتضح من نتائج الجدول (4) أن متوسط توزيع القدرة بلغ (0.16-) لوجيت، والانحراف المعياري (1.03) لوجيت، في حين بلغ الخطأ المعياري للوسط الحسابي لتقديرات القدرة (0.01)، وهي قيم قريبة من الوضع المثالي الذي يفترضه النموذج، وهذا يشير إلى دقة عالية في تحديد مواقع الأفراد على متصل السمة. كما أظهرت النتائج أن قيمة الوسط الحسابي والانحراف المعياري لمتوسط المربعات للمطابقة الداخلية والخارجية قد بلغا (0.99) و(0.02) على التوالي، وهذه القيم تقترب من القيم التي يفترضها النموذج (1 للوسط الحسابي و 0 للانحراف المعياري).

وأما فيما يتعلق بالتقديرات النهائية المتحررة لصعوبة الفقرات، فالجدول (5) يلخص نتائج التحليل للقيم التدريجية (Scale Values) المتحررة لصعوبة الفقرات.

جدول (5)

*نتائج التحليل للقيم المتحررة لصعوبة الفقرات (عدد الفقرات = 35، عدد الأفراد = 5253)*

| الإحصائي | القدرة | إحصائي المطابقة الداخلية (INFIT) | | إحصائي المطابقة الخارجية (OUTFIT) | |
|---|---|---|---|---|---|
| | | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD | متوسط المربعات MSNQ | قيمة الإحصائي ZSTD |
| الوسط الحسابي | 0.0 | 0.99 | -0.03 | 0.99 | -0.03 |
| الانحراف المعياري | 0.43 | 0.01 | 0.8 | 0.02 | 1.0 |
| الخطا المعياري في التقدير | 0.07 | 0.0 | 0.13 | 0.0 | 0.16 |

يلاحظ من الجدول (5) أن قيم التقديرات المتحررة لصعوبة الفقرات تتوزع بمتوسط قدره (0) لوجيت، وانحراف معياري (0.43) لوجيت، وقد بلغ الخطأ المعياري للوسط الحسابي لتقديرات الصعوبة (0.07) وهي قيمة متدنية، وهذا يشير إلى دقة تقديرات الصعوبة للفقرات. كما أظهرت النتائج أن قيمة الوسط الحسابي والانحراف المعياري لمتوسط المربعات للمطابقة الداخلية قد بلغ (0.99) و(0.01) على التوالي، في حين بلغت قيمة الوسط الحسابي والانحراف المعياري لمتوسط المربعات للمطابقة الخارجية يساوي (0.99) و(0.02) على التوالي، وهذه القيم تقترب من القيم التي يفترضها النموذج (1 للوسط الحسابي و 0 للانحراف المعياري).

**للإجابة عن السؤال الثاني الذي ينص على: "هل يحقق تدريج مقياس تقويم جودة البرامج الأكاديمية بنموذج سلم التقدير افتراضات نظرية الاستجابة للفقرة؟**

**أولاً: افتراض أحادية البعد Unidimensionality**

تم التحقق من افتراض أحادية البعد باستخدام برنامج WINSTEPS 5.8.5.0 عن طريق المكونات الرئيسية على البواقي المعيارية (PCAR)، إذ أظهرت النتائج أن نسبة التباين المفسر بواسطة المقياس Raw variance explained by measures بلغ 48.8% ، أي أنه تجاوز الحد الأدنى المحدد لذلك وهو 40%، كما أن الجذر الكامن للتباين غير المفسر في عامل البواقي الأول Unexplned variance in 1st contrast يقل عن 3 إذ بلغت قيمته (1.78) وحدة بنسبة (1.7%) من التباين الكلي، إذ تدل هذه القيمة المنخفضة على استقرار النموذج وملاءمته في قياس السمة المستهدفة، وبالتالي فإن شرط أحادية البعد قد تحقق (Linacre,2014; Holster & Lake, 2016).

**ثانياً: التحقق من افتراض الاستقلال الموضعي Local Independence**

تم التحقق من هذا الافتراض من خلال التحقق من مؤشر $Q_3$ في برنامج WINSTEPS 5.8.5.0، من خلال حساب معامل الارتباط بين بواقي أزواج فقرات المقياس، وحصلت جميع نتائج القيم العظمى لمعامل الارتباط بين البواقي المعيارية لأزواج فقرات المقياس قيماً أقل من معيار عدم تحقق شرط الاستقلال الموضعي من خلال مؤشر Q3< 0.3 (Christensen et al, 2017).

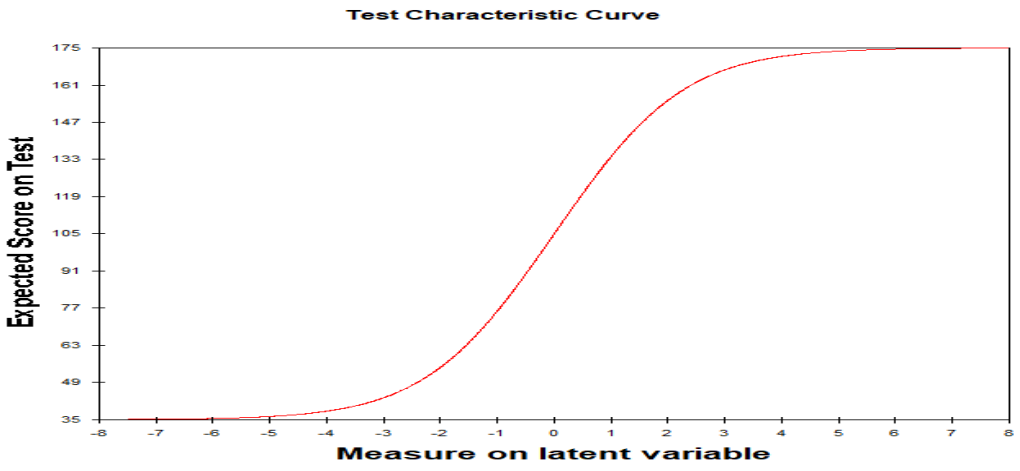
كما يذكر Hambleton & Swaminatan (1985) أن افتراض الاستقلال الموضعي يكافئ افتراض أُحادية البعد، ويعني ذلك أنه إذا تحقق افتراض أُحادية البعد في المقياس فإن المقياس يحقق افتراض الاستقلال الموضعي.

**ثالثاً: وتَرية منحنى خصائص الفقرة Item Characteristic Curve**

إن ما يميز نماذج الاستجابة للفقرة الاختبارية أحادية البعد هو وجود دالة مميزة خاصة لكل فقرة تأخذ الشكل اللوغاريتمي الاحتمالي (Rasch, 1960)، وللتحقق من هذا الافتراض تم فحص منحنى دالة المعلومات للمقياس كما في الشكل (2)، إذ يُظهر أنه كلما زادت قدرة الفرد زادت معها احتمالية الاستجابة الصحيحة، ثم إن خصائص الفقرة التي تؤثر في احتمالية الاستجابة الصحيحة عبارة عن معلم الصعوبة، على افتراض تساوي تمييز الفقرات، وإنعدام تأثير التخمين.

**شكل (2):**

***منحنى دالة المعلومات للمقياس Item Characteristic Curve (ICC)***

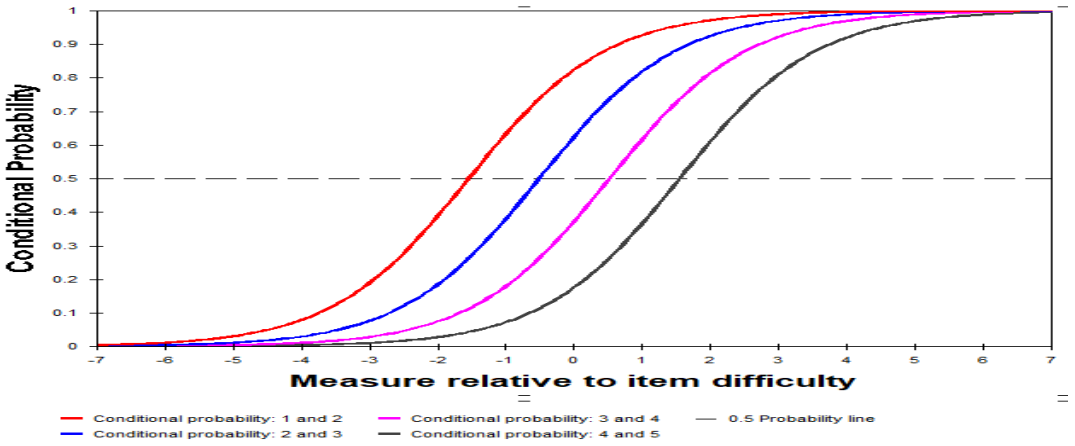


يتضح من الشكل (2) أنه كلما زاد مقدار ما يمتلكه الفرد من السمة المتمثلة بمفردات المقياس، كلما زادت قدرته على اختيار البديل المناسب لما يمتلكه من هذه السمة وارتفعت معدلاته على الفقرة، ويتضح أيضاً أن جميع القيم تتمركز حول خط منحنى الفقرات واقتربت جميعها من خط الترجيح اللوغاريتمي، مما يشير إلى وجود مطابقة للأفراد والفقرات.

ويوضح الشكل (3) العتبات الأربع (العتبة الأولى: تمتد من غير موافق بشدة إلى غير موافق، والعتبة الثانية: تمتد من غير موافق إلى محايد، والعتبة الثالثة: تمتد من محايد إلى موافق، والعتبة الرابعة: تمتد من موافق إلى موافق بشدة)

**الشكل (3):**

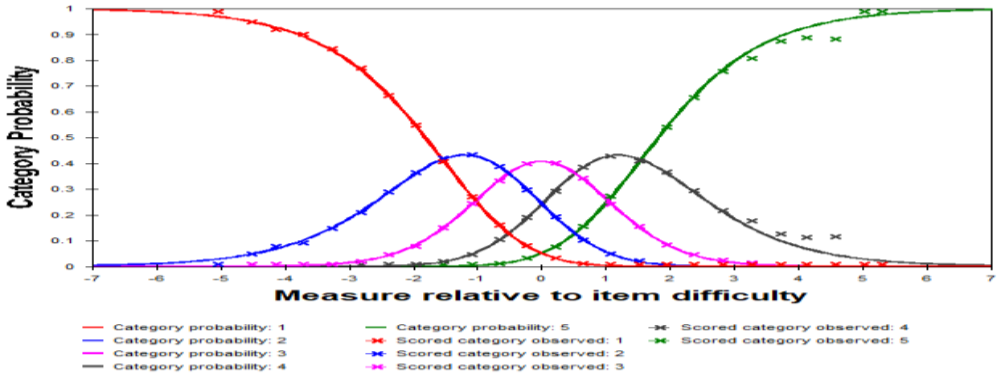
***منحنيات عتبات مقياس تقويم جودة البرامج الأكاديمية***




يتبين من الشكل (3) أن العتبات الأربع كانت متدرجة، وهذا يدل على تحقق فرض نموذج سلم التقدير، الذي ينص على أن معامل الصعوبة هو المعلم المؤثر، فكلما زاد معامل الصعوبة، قل احتمال أن يجيب الفرد إجابة صحيحة عن الفقرات، كما أن معالم التمييز والتخمين متساويات لجميع الفقرات، وقد تحقق هذا الفرض لجميع فقرات المقياس. ويتضح ذلك من خلال الشكل (4) الذي يوضح العلاقة بين مستويات الصعوبة للبدائل الخمس، واحتمال الاستجابة على الفقرة.

شكل (4):

*العلاقة بين مستويات الصعوبة للبدائل الخمس واحتمالية الإجابة الصحيحة لفقرات المقياس*




يتضح من الشكل (4) أن البدائل متدرجة في الصعوبة، وأن البديل (5) هو الأفضل، وبذلك يكون تحقق افتراض وترية منحنى خصائص الفقرة.

**رابعاً: افتراض التحرر من السرعة في الأداء (speededness)**

تفترض نظرية الاستجابة للفقرة تحرر الاختبار من عامل السرعة، بمعنى أن يكون الوقت مناسباً وكافياً لجميع أفراد العينة، ونظراً لأن المقياس من نوع الأداء العادي، بمعنى أن عامل السرعة لا يؤدي دوراً مهماً في الاستجابة على فقرات المقياس، فإن الافتراض الخاص بعامل السرعة محقق.

**للإجابة عن السؤال الثالث الذي ينص على: "ما دلالات الصدق والثبات لمقياس تقويم جودة البرامج الأكاديمية بعد تدريجه باستخدام نموذج سلم التقدير لأندريش؟**

**دلالات الصدق**

عند استخدام أحد نماذج نظرية الاستجابة للفقرة في تطوير وبناء المقاييس لا بد من توفير أدلة صدق المقياس، إذ لا يعني مطابقة البيانات للنموذج المستخدم أنها تقدم دليلاً على صدقه، وإنما يشير ذلك إلى أن فقرات المقياس تقيس سمة عامة (Hambleton & Swaminathan, 1985). لذا تم التحقق من صدق مقياس تقويم الطلبة لجودة البرامج الأكاديمية بأكثر من طريقة، وهي:

أولاً: الصدق المنطقي (Logical Validity) من خلال التحليل النظري لمفهوم جودة البرامج الأكاديمية، وكتابة فقرات تمثل مكوناته، وطريقة صوغها وتحكيمه من قبل مجموعة من

المتخصصين، إذ تم تعديل الفقرات اعتماداً على نتائج التحكيم، إذ تُعد طريقة التحليل المنطقي إحدى الطرائق والإجراءات التي تستخدم للتحقق من صدق البناء لأدوات القياس (Construct Validity) (Cronbach, 1971).

**ثانياً: الأساليب الارتباطية**

تم إيجاد معاملات الارتباط بين الدرجة على الفقرة والدرجة على المقياس الكلي بعد حذف الفقرة، وذلك بعد حذف الفقرات غير المطابقة والأفراد غير المطابقين لنموذج سلم التقدير، بغرض معرفة مدى إسهام كل فقرة بما يقيسه المقياس الكلي، والجدول (6) يبين ذلك.

**جدول (6)**

***معاملات الارتباط بين الدرجة على الفقرة والدرجة على المقياس الكلي (عدد الفقرات = 35)***

| الفقرة | معامل الارتباط | الفقرة | معامل الارتباط | الفقرة | معامل الارتباط |
|---|---|---|---|---|---|
| i1 | 0.639 | i15 | 0.624 | i29 | 0.628 |
| i2 | 0.644 | i16 | 0.645 | i30 | 0.639 |
| i4 | 0.644 | i18 | 0.651 | i31 | 0.650 |
| i5 | 0.642 | i19 | 0.652 | i33 | 0.646 |
| i6 | 0.641 | i20 | 0.648 | i34 | 0.650 |
| i7 | 0.646 | i21 | 0.646 | i35 | 0.655 |
| i8 | 0.644 | i22 | 0.651 | i36 | 0.646 |
| i9 | 0.637 | i23 | 0.632 | i37 | 0.644 |
| i10 | 0.660 | i25 | 0.648 | i38 | 0.615 |
| i11 | 0.651 | i26 | 0.634 | i39 | 0.638 |
| i13 | 0.641 | i27 | 0.646 | i40 | 0.635 |
| i14 | 0.646 | i28 | 0.634 | | |

يلاحظ من خلال الجدول (6) أن جميع معاملات الارتباط بين درجات الأفراد على الفقرات ودرجاتهم على المقياس كانت عالية، وموجبة، ودالة إحصائياً عند مستوى دلالة (0.01) ما يوفر دليلاً على فاعلية فقرات المقياس، إذ تقيس هذه الفقرات ما يقيسه المقياس الكلي.

**ثالثاً: صدق التدرج**

تعد (كاظم، 1994) أن الفقرات غير الملائمة للنموذج هي تلك التي لا تتسق في تدريجها مع تدريج باقي الفقرات على المتصل موضوع القياس، وغالباً ما تتضمن هذه الفقرات قياساً لصفة أخرى، غير التي هي موضوع القياس، أو تكون الفقرات غامضة، أو هناك عيب في صياغتها.

وبناءً على ذلك فإن استبعاد الفقرات غير الصادقة، وإبقاء الفقرات الصادقة في تعريفها للمتغير موضوع القياس يؤكد صدق فقرات الاختبار في قياسها للسمة المراد قياسها، وهذا ما تم إجراؤه من خلال استبعاد الفقرات غير الملائمة للنموذج، إذ تم استبعاد (5) فقرات من فقرات المقياس، ومما تجدر الإشارة إليه أن استبعاد الفقرات جاء بعد استبعاد الأفراد غير الملائمين للنموذج، وحذف الأفراد غير الملائمين معناه حذف الأفراد غير الصادقين في استجاباتهم، وهم الذين يختلف نمط استجاباتهم عن معظم الأفراد، وبالتالي فإنه بحذف الفقرات والأفراد غير الملائمين يضمن أمرين أساسيين: أولاً، صدق فقرات الاختبار في قياس السمة المراد قياسها، وثانياً، صدق تدرج قدرات الأفراد على متصل هذا المتغير.

**دلالات الثبات**

قدرت معاملات الثبات للمقياس بطريقتين: أولاهما باستخدام الطرق التقليدية في القياس والثانية باستخدام نظرية الاستجابة للفقرة. أما بالنسبة للطرق التقليدية، فقد تم استخراج معامل ثبات الاتساق الداخلي باستخدام معادلة كرونباخ ألفا، وقد بلغت قيمته باستخدام هذه الطريقة (0.98) وهي قيمة عالية تشير إلى أن المقياس يتمتع بثبات عال يتيح إمكانية استخدامه في مواقف ذات علاقة بمتغير بالإتجاهات نحو العمل المهني وتمتع المقياس بدلالات اتساق داخلي كمؤشر على الثبات عالية.

أما باستخدام نظرية الاستجابة للفقرة، وبعد الحصول على القيم المتحررة لكل من صعوبة الفقرات وقدرات الأفراد، تم الحصول على نوعين من المعاملات: معامل الثبات الخاص بالأفراد (Person Reliability)، ومعامل الثبات الخاص بالفقرات (Item Reliability)، فالثبات وفق نظرية

الاستجابة للفقرة يعني الدقة في تقدير موقع كل من الأفراد والفقرات على متصل السمة، ويمكن تحديد مدى دقة الفقرات في تعريف هذا المتصل بحساب معامل الفصل بين الفقرات (Item separation Index) (Gi)، وهو النسبة بين الانحراف المعياري للقيم التقديرية لصعوبة الفقرات، ومتوسط الخطأ المعياري لصعوبة الفقرات (Wright & Masters, 1982). وقد بلغت قيمة معامل الفصل لفقرات المقياس بصورته النهائية (26.8). وهذه القيمة تزيد على 2، ولذلك تُعد هذه الفقرات كافية لتعريف متصل السمة الذي تقوم بقياسه. وكذلك بلغت قيمة معامل الفصل للأفراد (Person Separation Index) (Gp) (4.82). وهذه القيمة تزيد على 2، ولذلك فإن عينة الأفراد تُعد أيضاً كافية للفصل بين الفقرات (Yen & Fizpatrick, 2006).

ومن خلال هذه المعاملات Gi، Gp تم حساب معامل الثبات لكل من الفقرات والأفراد، وفق الصيغة الرياضية الآتية ($R=G^2/1+G^2$)، إذ ترمز G إلى معامل الفصل و R إلى معامل الثبات، وقد بلغت قيم معاملات الثبات لكل من الفقرات والأفراد 0.99 و 0.96 على التوالي، وهما قيمتان مرتفعتان، إذ تدل الأولى على كفاية عينة الفقرات في الفصل بين مستويات الأداء للأفراد، في حين تشير الثانية إلى كفاية عينة الأفراد في الفصل بين فقرات المقياس.

## التوصيات

في ضوء ما أسفرت عنه نتائج الدراسة الحالية، يوصي الباحث بالآتي:

- تعميم تطبيق المقياس على مختلف الكليات والبرامج الأكاديمية بالجامعة، بما يتيح مؤشرات موضوعية تسهم في تحسين جودة البرامج وتطويرها.
- تشجيع الباحثين على استخدام نماذج الاستجابة للفقرة (IRT) ولا سيما أنموذج راش وأندريش في تطوير مقاييس تقويم مشابهة في الجامعات السعودية والعربية، بما يسهم في تعزيز دقة القياس التربوي.
- توسيع نطاق استخدام أنموذج أندريش في مجالات تقييم أخرى مثل تقويم المقررات أو أداء أعضاء هيئة التدريس.

- إجراء دراسات مقارنة بين المقياس المطوَّر والمقاييس التقليدية للكشف عن مدى تفوق النماذج الحديثة في توفير بيانات أكثر دقة وموضوعية.

## قائمة المراجع

### اولاً: المراجع العربية

أبا بطين، عبدالحكيم وسمور، رضا. (2019). *الجودة الشاملة والاعتماد الأكاديمي بالتعليم العالي*. الرياض، عمادة البحث العلمي، جامعة الملك سعود.

أبو هاشم، السيد محمد. (2016). تقويم برنامج ماجستير علم النفس بكلية التربية جامعة الملك سعود من وجهة نظر الطلبة في ضوء معايير الاعتماد الأكاديمي. *المجلة العربية لضمان جودة التعليم الجامعي*، 9(24)، 201 – 219.

الحياصات، خالد محمد وبني عامر، أيمن. (2016). تقويم برامج جامعة القصيم للدراسات العليا في ضوء معايير الهيئة الوطنية للتقويم والاعتماد الأكاديمي (NCAAA). مجلة جامعة طيبة للعلوم التربوية، 11(3)، 445 – 464.

الخريف، جيهان عزيز، وتعلب، صبرين صلاح. (2023). توظيف نموذج سلم التقدير في تقنين مقياس كاليفورنيا للدافعية العقلية لدى طلاب المرحلة الثانوية في بريدة. *مجلة الآداب للدراسات النفسية والتربوية*، *5* (2)، 202 – 244.

دي أيالا، ر. ج. (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة*. (الكيلاني، عبدالله زيد والبرصان، إسماعيل سلامة، مترجم)، الرياض: دار جامعة الملك سعود. (العمل الأصلي نشر في 2009).

السليمي، سليمان بن عبيد، السعودي، شريف عبدالرحمن، وجمعة، أمجد. (2025). بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة. مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية، 16(46)، 28 – 107.

الشمري، سلمان حديد. (2019). درجة توافر متطلبات معيار التعليم والتعلم في برنامج الرياضيات بجامعة شقراء في ضوء معايير المركز الوطني للتقويم والاعتماد الأكاديمي (NCAAA). *مجلة جامعة أم القرى للعلوم التربوية والنفسية*، 10(2)، 431 – 455.

الطراونة، صبري. (2018). تطوير مقياس التدين لدى طلبة الجامعة باستخدام نموذج سلم التقدير في نظرية الاستجابة للفقرة. *مجلة العلوم التربوية والنفسية*، جامعة البحرين، 19(3)، 87 – 108.

العصيل، عبدالعزيز بن فالح. (2019). تقويم إعداد معلم العلوم الشرعية في كلية الشريعة والدراسات الإسلامية بالأحساء في ضوء معايير الاعتماد الأكاديمي. *مجلة العلوم التربوية*، 32 (2)، الرياض، 273 – 298.

علام، صلاح الدين محمود. (2000). *القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة: دار الفكر العربي.

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

العنزي، عبدالله قريطان. (2024). تدريج مقياس تقدير الذات في البيئة السعودية على عينة من طلبة الجامعة وفق نموذج أندريش. مجلة الإرشاد النفسي، 80، 189 – 220.

كاظم، أمينة محمد. (1994). *تدرج ومعايرة المقاييس: معايير نمو طفل ما قبل المدرسة*. ضمن: محمد عماد الدين إسماعيل، ناهد رمزي، ليلى كرم، أمينة كاظم، هدى ناشف. القاهرة: المجلس القومي للطفولة والأمومة، 114–232.

مجلس شؤون الجامعات. (2025). إحصاءات الجامعات والطلاب. مجلس شؤون الجامعات. https://www.cua.gov.sa/

المفيز، خولة، العيفان، مي والحمدان، أمل. (2021). جودة التعليم والتعلم في الجامعات وفق معايير (NCAAA) لاعتماد برامج الدراسات العليا. *المجلة العلمية لجامعة الملك فيصل*، 22(1)، 246 - 253.

هيئة تقويم التعليم والتدريب. (2018). **دليل معايير الاعتماد البرامجي**. الرياض: المركز الوطني للتقويم والاعتماد الأكاديمي.

## Arabic References (Romanized)

Aba Butain, Abdulhakeem, & Samour, Reda. (2019). *Al-jawda al-shamila wa al-i'timad al-akademi bi al-ta'lim al-'ali* [Total quality and academic accreditation in higher education]. Riyadh: Deanship of Scientific Research, King Saud University.

Abu Hashem, Al-Sayyid Mohammed. (2016). *Taqwim barnamaj majisteer 'ilm al-nafs bi kulliyat al-tarbiyah jami'at al-malik saud min wijhat nazar al-tullab fi daw' ma'ayir al-i'timad al-akademi* [Evaluation of the Master's program in Psychology at King Saud University in light of academic accreditation standards]. *The Arab Journal of Quality Assurance in Higher Education*, 9(24), 201–219.

Al-Hayassat, Khaled Mohammed, & Bani Amer, Ayman. (2016). *Taqwim baramij jami'at al-qasim li al-dirasat al-'ulya fi daw' ma'ayir al-hay'a al-wataniyya li al-taqwim wa al-i'timad al-akademi (NCAAA)*

[Evaluation of graduate programs at Qassim University in light of national accreditation standards (NCAAA)]. *Journal of Taibah University for Educational Sciences*, 11(3), 445–464.

Al-Kharif, Jehan Aziz, & Ta'lab, Sabrin Salah. (2023). *Tawzif namudhaj sullam al-taqdir fi taqneen miqyās California li al-dafi'iyyah al-'aqliyyah lada tullab al-marhalah al-thanawiyyah fi Buraydah* [Using the rating scale model to standardize the California Measure of Mental Motivation among high school students in Buraidah]. *Journal of Arts for Psychological and Educational Studies*, 5(2), 202–244.

De Ayala, Ronald J. (2017). *Al-nazariyyah wa al-tatbiq fi nazariyyat al-istijabah lil-faqrah* [Theory and application of item response theory] (Al-Kilani, Abdullah Zaid, & Al-Barsan, Ismail Salama, Trans.). Riyadh: King Saud University Press. (Original work published 2009)

Al-Sulaimi, Sulaiman Bin Ubaid, Al-Saudi, Sharif Abdulrahman, & Jum'ah, Amjad. (2025). *Bina miqyās al-hikmah al-ikhtibariyyah lada tullab al-jami'at al-'umaniyyah fi daw' nazariyyat al-istijabah lil-faqrah* [Developing an experimental wisdom scale among Omani university students in light of item response theory]. *Al-Quds Open University Journal for Educational and Psychological Research and Studies*, 16(46), 28–107.

Al-Shammari, Salman Hadid. (2019). *Darajat tawaffur mutalabat ma'yar al-ta'lim wa al-ta'allum fi barnamaj al-riyadiyyat bi jami'at Shaqra fi daw' ma'ayir al-markaz al-watani li al-taqwim wa al-i'timad al-akademi (NCAAA)* [Availability of teaching and learning standard requirements in the mathematics program at Shaqra University in light of NCAAA standards]. *Umm Al-Qura University Journal of Educational and Psychological Sciences*, 10(2), 431–455.

Al-Tarawneh, Sabri. (2018). *Tatwir miqyās al-tadayyun lada tullab al-jami'ah bi istikhdam namudhaj sullam al-taqdir fi nazariyyat al-istijabah lil-faqrah* [Developing a religiosity scale for university students using the rating scale model of item response theory]. *Journal of Educational and Psychological Sciences*, University of Bahrain, 19(3), 87–108.

Al-Aseel, Abdulaziz Bin Faleh. (2019). *Taqwim i'dad mu'allim al-'ulum al-shar'iyyah fi kulliyat al-shari'ah wa al-dirasat al-islamiyyah bi al-Ahsa fi daw' ma'ayir al-i'timad al-akademi* [Evaluation of the preparation of religious sciences teachers at the College of Sharia and Islamic Studies in Al-Ahsa in light of accreditation standards]. *Journal of Educational Sciences*, 32(2), 273–298.

Allam, Salah El-Din Mahmoud. (2000). *Al-qiyas wa al-taqwim al-tarbawi wa al-nafsi: Asasiyyatuhu wa tatbiqatuhu wa tawajjuhatuhu al-mu'asirah* [Educational and psychological measurement and evaluation: Fundamentals, applications, and contemporary trends]. Cairo: Dar Al-Fikr Al-Arabi.

Allam, Salah El-Din Mahmoud. (2005). *Namadhij al-istijabah lil-mufradah ahadiyyat al-bu'd wa muta'addidat al-ab'ad fi al-qiyas al-nafsi wa al-tarbawi* [Unidimensional and multidimensional item response models in psychological and educational measurement]. Cairo: Dar Al-Fikr Al-Arabi.

Al-Anzi, Abdullah Quraitan. (2024). *Tadreej miqyās taqdeer al-dhāt fi al-bī'ah al-sa'udiyyah 'ala 'ayyinah min tullab al-jami'ah wifq namudhaj Andrich* [Calibration of a self-esteem scale among Saudi university students according to Andrich's model]. *Journal of Psychological Counseling*, 80, 189–220.

Kazem, Amina Mohammed. (1994). *Tadaruj wa mu'ayarat al-maqayis: Ma'ayir numu tifl ma qabl al-madrasah* [Scaling and calibration of measures: Developmental norms of preschool children]. In Ismail, Mohammed Emad El-Din, Ramzi, Nahid, Karam, Laila, Kazem, Amina, & Nashif, Huda (Eds.), *The National Council for Childhood and Motherhood*, 114–232. Cairo: The National Council for Childhood and Motherhood.

Majlis Shu'oon Al-Jami'at [Council of University Affairs]. (2025). *Ihsa'at al-jami'at wa al-tullab* [Universities and students statistics]. Riyadh: Council of University Affairs. https://www.cua.gov.sa/

Al-Mufayz, Khawla, Al-Ifan, Mai, & Al-Hamdan, Amal. (2021). *Jawdat al-ta'lim wa al-ta'allum fi al-jami'at wifq ma'ayir (NCAAA) li i'timad baramij al-dirasat al-'ulya* [Quality of teaching and learning in universities according to NCAAA standards for graduate program accreditation]. *Scientific Journal of King Faisal University*, 22(1), 246–253.

Education and Training Evaluation Commission. (2018). *Dalil ma'ayir al-i'timad al-barnamaji* [Program accreditation standards guide]. Riyadh: National Center for Assessment and Academic Accreditation.

**ثانياً: المراجع الانجليزية**

Aiuadj, S., Ziad, R. (2023). Use of Rasch-Andrich's Rating scale Model in the Calibrating of the Depression, Anxiety and Stress Scale (DASS-42) on high school students. *Journal of Psychological and Educational Sciences*, 9(2), 133- 147.

ALastair , P. & Hutchinson , G. (1987). Calibrating graded assessment: Rasch partial credit analysis of performance in writing. *Language Testing*, (4) ,73- 92

Allen, M. J., & Yen, W., (2002*). Introduction to measurement theory*, Long Grove, IL. Waveland Press, Inc.

Anastasi, A. (1982). *Psychological Testing (5th ed.)*. Macmillan, New York.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573.

Biggs, J., & Tang, C. (2011). Teaching for Quality Learning at University. McGraw-Hill Education.

Bjorner, J.; Petersen, M. ;Groenvold, M.; Aaronson, N.; Ahlner-Elmqvist,M.; Arraras, J.; Bredart,A.; Fayers, P.; Jordhoy.M; Sprangers,M.;Watson.M& Young,T.(2004): Use Of Item Response Theory to Develop A

Shortened Version of The EORTC QLQ-C30 Emotional Functioning Scale. *Quality of Life Research* ,13,1683–1697.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *Applied psychological measurement*, 41(3), 178-194. DOI: 10.1177/0146621616677520.

Cortés-Rodríguez, C. A., Martínez-Gómez, G., Romo-Lozano, J. L., Arvizu-Barrón, E. (2023). Evaluation of the Entrepreneurial Ability of Small-Scale Farmers through the Rasch–Andrich Model. *Journal Agriculture*, 13(3), 721. https://doi.org/10.3390/agriculture13030721.

Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory. Harcourt*, New York.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC:American Council on Education.

De Gruijter, N. M. & Van der kamp, J. TH. (2008). *Statistical Test Theory for the Behavioral Sciences*. Taylor & Francis Group, Boca Raton, London New York.

Embretson, S. & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahawah, NJ: Erlbaum.

Hambleton, R. k, & Swaminthan, H. (1985*). Item Response Theory: Principles and applications*, Boston, Kluwer - Nijhoff publishing.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage, Newbury Park, CA.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, 18(1), 9–34.

Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and Items. *Applied Psychological Measurement*, 9, 139-164.

Holster, T. A., & Lake, J. (2016). Guessing and the Rasch model. *Language Assessment Quarterly*, 13(2), 124–141. https://doi.org/10.1080/15434303.2016.1160096

Huling, C., Drasgow, F. & Parsons, K. (1983). *Item Response Theory: Application to Psychological Measurement (1st ed)*. Illinois: Dow Jones-Irwin.

Julian , E.R. (1988). Using computerized patient simulations to measure the clinical Competence of physicians. *Applied Measurement in Education* , (4) , 299-318

Linacre J. M. (2014). *A user's guide to WINSTEPS 5.8.5.0 Mini step Rasch-model computer programs*. Program Manual 4.6.1., Beaverton, Oregon. http://WINSTEPS 5.8.5.0 .com

Lord, F . M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.

National Center for Academic Accreditation and Evaluation (NCAAA). (2022). *Program accreditation standards and guidelines*. Riyadh: Education & Training Evaluation Commission.

Oktaviyanthi, R., Agus, R. N., Garcia, M. L. B., & Lertdechapat, K. (2024). Cognitive load scale in learning formal definition of limit: A Rasch model approach. Infinity. *Journal of Mathematics Education*, 13(1), 99-118. https://doi.org/10.22460/infinity.v13i1

Oliveira, S., Fernandes, D. & Sisto, F. (2014). Analysis of the school anxiety inventory in Brazil using the Rasch rating scale. *Psychological Reports*, 115(1), 165- 178. Retrieved on March 7, 2015, from EBSCO: Academic Search premier.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmark's Paedagogiske Institute.

Reeve, B. (2004). *An Introduction to Modern Measurement Theory*. Division of Cancer Control and Population Science, National Cancer Institute.

Reeve, B. (2004). *Applications of Item Response Theory (IRT) Modeling for Building and Evaluating Questionnaires Measuring Patient-Reported Outcomes*. Retrieved from http://www.researchgate.net

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

Wright, B. & Ston, M. (1979). *Best Test design a handbook for Rasch measurement (1st ed.)*. Chicago: MESA press.

Yen, W. M., & Fitzpatrick, R. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111–153). Westport, CT: American Council on Education and Praeger Publishers.

You. h., Kim, k., Black, k. & Min, k. (2018). Assessing science motivation for college students: validation of the science motivation questionnaire ii using the rasch-andrich rating scale model. EURASIA Journal of Mathematics, *Science and Technology Education*, 14(4), 1161-1173. DOI: 10.29333/ejmste/81821.

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, 45(2), 115–138.