

Constructing a Measure of Test - Wisdom Among Omani University Student's in Light of Item Response Theory

Mr. Suleiman obaid Alsulami*¹, Dr. Sharif Abdelrahman Alsoudi², Dr. Amjad Izat Joma³

1 Part-time Lecturer • Psychology, A'Sharqiyah University, Ibra, Oman.

2 Associate professor•Psychology, A'Sharqiyah University, Ibra, Oman.

3Associate professor•Psychology, A'Sharqiyah University, Ibra, Oman

Orcid No: 0009-0004-2299-8585

Orcid No: 0000-0002-2882-0225

Orcid No: 0000-0002-2605-1987

Email: suleiman-alsulami@moe.om

Email: sharif.alsoudi@asu.edu.om

Email: Amjad.joma@asu.edu.om

Received:

5/06/2024

Revised:

5/06/2024

Accepted:

31/08/2024

*Corresponding Author:
suleiman-alsulami@moe.om

Citation: Alsulami, S. obaid, Alsoudi, S. A., & Joma, A. I. Constructing a Measure of Test - Wisdom Among Omani University Student's in Light of Item Response Theory. Journal of Al-Quds Open University for Educational & Psychological Research & Studies. <https://doi.org/10.33977/1182-016-046-006>

2023@jrresstudy. Graduate Studies & Scientific Research/Al-Quds Open University, Palestine, all rights reserved.

Open Access



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract

Objectives: Constructing a scale of experimental wisdom among Omani university students, and verifying its psychometric properties in light of the item response theory according to the assumptions of Andrich's model.

Methods: The researcher relies on the descriptive approach, and a scale of experimental wisdom consisting of (60) paragraphs was constructed. It was applied to a sample of (719) Omani university students.

Results: The paragraphs of the test wisdom scale prepared in the study matched the assumptions of Andrich's model, which are the internal and external compatibility of the paragraphs, the correlation of the paragraphs with the scale, the differential performance of the paragraphs, unidimensionality, positional independence, and the equality of the discriminating ability of the paragraphs. With the exception of six paragraphs that did not achieve the specified correlation coefficient with the scale, and three paragraphs that did not achieve the assumption of the differential performance of the paragraphs. These paragraphs were deleted so that the scale in its final form consisted of (51) paragraphs. The values of the reliability coefficients of the scale reached (0.99, 0.92) for the paragraphs and individuals, respectively, while the values of the separation coefficients of the scale reached (11.86, 3.42) for the paragraphs and individuals, respectively, which are high reliability and separation coefficients. As for the results of the calibration analysis of the scale categories, the results indicated that the scale is effective in grading its paragraphs. Also, the measurement of the categories increases at a steady pace when moving from one category to another, and the distance between them fell within the range.

Conclusions: The study recommended using the experimental wisdom scale that was constructed and its psychometric properties verified in this study in other studies and research related to experimental wisdom.

Keywords: Test Wisdom, Item Response Theory, Andrich's Model.

بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة

أ. سليمان بن عبيد بن شامس السليمي*¹، د. شريف عبد الرحمن السعودي²، د. أمجد عزات جمعة³

¹محاضر غير متفرغ، قسم علم النفس، كلية الآداب والعلوم الإنسانية، جامعة الشرقية، ولاية إبراء، سلطنة عمان.

²استاذ مشارك، قسم علم النفس، كلية الآداب والعلوم الإنسانية، جامعة الشرقية، ولاية إبراء، سلطنة عمان.

³استاذ مشارك، قسم علم النفس، كلية الآداب والعلوم الإنسانية، جامعة الشرقية، ولاية إبراء، سلطنة عمان.

المخلص

الأهداف: بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية، والتحقق من خصائصه السيكومترية في ضوء نظرية الاستجابة للفقرة وفقاً لافتراضات نموذج أندريش.

المنهجية: اعتمدت الدراسة على المنهج الوصفي، كما تم بناء مقياس للحكمة الاختبارية مكون من (60) فقرة. وطبق على عينة بلغت (719) من طلبة الجامعات العمانية.

النتائج: طبقت فقرات مقياس الحكمة الاختبارية المعد في الدراسة لافتراضات نموذج أندريش والتمثلة في المطابقة الداخلية والخارجية لل فقرات، وارتباط الفقرات مع المقياس، والأداء التفاضلي لل فقرات، وأحادية البعد، والاستقلال الموضوعي، وتساوي القدرة التمييزية لل فقرات. باستثناء ست فقرات لم تحقق معامل الارتباط المحدد مع المقياس، وثلاث فقرات لم تحقق افتراض الأداء التفاضلي لل فقرات. وتم حذف هذه الفقرات ليصبح المقياس في صورته النهائية مكون من (51) فقرة. كما أن قيم معاملات الثبات للمقياس بلغت (0.99، 0.92) لل فقرات والأفراد على التوالي، في حين بلغت قيم معاملات الفصل للمقياس (11.86، 3.42) لل فقرات والأفراد على التوالي، وهي معاملات ثبات وفصل مرتفعة. أما نتائج تحليل معايرة فئات المقياس أشارت النتائج أن المقياس ذو فاعلية في تدرج فقراته. كما أن قياس الفئات يتزايد بخطى ثابتة عند الانتقال من فئة إلى أخرى والمسافة بينها وقعت ضمن المدى.

الخلاصة: أوصت الدراسة باستخدام مقياس الحكمة الاختبارية الذي تم بناؤه والتحقق من خصائصه السيكومترية في هذه الدراسة في دراسات وبحوث أخرى ترتبط بالحكمة الاختبارية.

الكلمات المفتاحية: الحكمة الاختبارية، نظرية الاستجابة للفقرة، نموذج أندريش.

المقدمة

تعدّ الاختبارات بأنواعها المختلفة من أهم أدوات القياس والتقييم المستخدمة في العديد من البحوث العلمية، وتتضمن تقديم عدد من المثيرات أو الأسئلة للمفحوص بهدف الحصول على عدد من الاستجابات، واعطاءه درجة تناسب المقدار الذي يملكه من السمة المرغوب قياسها، وبالتالي اتخاذ القرارات حوله. وتعتمد دقة هذه القرارات، وصحتها على خصائص هذه الاختبارات، حيث إنّ هناك العديد من الخصائص التي يجب أن تمتاز بها هذه الاختبارات، منها: الموضوعية التي تعني أن الاختبار لا يتأثر بالعوامل الذاتية للمصححين القائمين على ذلك الاختبار، والثبات الذي يعني أن يعطي الاختبار نتائج متسقة إذا ما أعيد تطبيقه على الأفراد أنفسهم في الظروف نفسها، والصدق، ويعتبر الصدق من أهم هذه الخصائص. ويتضمن مفهوم الصدق أن يقيس الاختبار ما وضع لأجله بشكل دقيق، وبالتالي لا بد عند بناء أي اختبار أن يحقق الصدق حتى يحقق الهدف المصمم لأجله، أي أن تكون السمة المراد قياسها هي التي يقيسها الاختبار فقط، دون تدخل عوامل أخرى معها (مجيد، 2014).

يتأثر صدق الاختبار بالعديد من العوامل المرتبطة بالاختبار نفسه، مثل: مدى وضوح فقرات الاختبار، بالإضافة لسهولة وصعوبتها. وأيضاً العوامل المرتبطة بتطبيق الاختبار، وتصحيحه، مثل: العوامل البيئية، وعدم وضوح التعليمات، وعدم تطبيق الاختبار على الفئة المراد قياس السمة فيها. والعوامل المرتبطة بالمفحوص، مثل: الاضطراب، والقلق، والغش، ومهارات التعامل مع الورقة الاختبارية أو ما يسمى بالحكمة الاختبارية (نوفل وأبو عواد، 2010).

عرّف ميلمان وآخرون (Millman, et al., 1965) الحكمة الاختبارية بأنها قدرة الفرد على الاستفادة من خصائص وطبيعة الاختبار دون معرفته بمحتوى فقرات الاختبار للحصول على درجة أعلى. وعرّفها كل من روجز وبيتنس (Rogers & Bateson, 1991) بأنها قدرات أو مهارات يستخدمها الأفراد لتعديل درجتهم في الاختبار بغض النظر عن محتوى الاختبار. كما عرّفها أورث (Orth, 2003) بأنها قدرة معرفية مكتسبة لمجموعة من المهارات التي يمكن للمفحوص أن يستخدمها في موقف الاختبار لتحسين درجته على الاختبار بغض النظر عن محتواه. نظراً لتأثير الحكمة الاختبارية في صدق الاختبار، فإنه من الضروري الكشف عن هذه السمة لدى الطلبة، من خلال بناء مقاييس عملية خاصة يتم من خلالها قياس هذه السمة بطرق علمية وموضوعية.

تقوم النظرية الكلاسيكية على مجموعة من الافتراضات البسيطة، مثل: أن الخطأ المعياري في القياس متساو لدى جميع المفحوصين، ودرجات الاختبار صممت حتى تخضع للتوزيع الاعتيادي على متصل القدرة (العنكبى، 2009). ولكنها عانت بعض أوجه القصور، مثل: انعدام خطية القياس، وأن الدرجة الكلية للاختبار ترتبط بفقرات الاختبار، وعدم وجود وحدة قياس ثابتة، والذي أثر على دقة القياس أو دقة الكشف عن الخصائص السيكومترية للاختبارات التي اعتمدت عليها (بخيت، 2013). وللتغلب على القصور الذي عانت منه النظرية الكلاسيكية؛ ظهرت النظرية الحديثة في القياس (Item Response Theory (IRT)، أو ما يعرف بنظرية الاستجابة للفقرة (زكري، 2009). تتميز نظرية الاستجابة للفقرة عن النظرية الكلاسيكية في القياس استقلالية القياس، أي إنّ تقدير معالم الفقرات يكون مستقلاً عن خصائص الأفراد الذين استخدموا في تقدير هذه المعالم، وكذلك أنّ تقدير قدرة الأفراد يكون مستقلاً عن عينة الفقرات التي تطبق عليهم (Susan & Steven, 2000). عرّف مولينيرس وهوجتتك (Molenaar & Hoijtink, 1996) نظرية الاستجابة للفقرة بأنها نظرية تتنبأ بأداء الأفراد وتفسر أدائهم في ضوء خاصية أو خصائص مميزة لذلك الأداء يسمى سمات. كما عرفها الحبيب (2018) بأنها نظرية تعمل على تحديد العلاقة بين استجابة المفحوص على فقرات اختبار ما، وبين السمة الكامنة التي تقف خلف هذا الأداء. تنبثق عن نظرية الاستجابة للفقرة مجموعة من النماذج الرياضية كل منها معتمد على معادلة رياضية تحدد علاقة أداء الفرد على فقرات المقياس بقدرته التي تكمن وراء هذا الأداء، وقدرة الفرد حسب موقعة على متصل القدرة، وهي: النموذج أحادي المعلمة أو ما يعرف بنموذج راش Rasch Model، والنموذج ثنائي المعلمة، والنموذج ثلاثي المعلمة (Baker, 2001؛ عليّات 2019).

يعتبر نموذج راش أحد أبسط نماذج نظرية الاستجابة للفقرة، وينطوي تحته أربعة نماذج اعتماداً على نمط الاستجابة للفقرات، وهي: نموذج راش ثنائي الاستجابة Rasch Dichotomous Model ويستخدم مع الفقرات التي تحتوي على استجابتين مثلاً: نعم، لا. ونموذج تقدير أندريش Andrich Rating Scale Model (ASCM) ويستخدم مع الفقرات ذات الاستجابة المتعددة التي لها الأوزان نفسها مثل مقياس ليكرت. ونموذج ماستر المتدرج Masters Partial Credit Model ويستخدم مع الفقرات ذات الاستجابة المتدرجة والتي لها أوزان مختلفة مثل الأسئلة الحسابية. وأخيراً النموذج المجمع Grouped Model ويستخدم مع الفقرات متعددة الاستجابة ذات الأوزان المختلفة (Alsoudi & Abu Shindi, 2023; Empreson & Reise, 2000; Thissen & Edwards, 2005). ونظراً لميزات هذه النظرية سيتم توظيفها في الدراسة الحالية لبناء مقياس الحكمة الاختبارية في ضوء نظرية الاستجابة للفقرة.

مشكلة الدراسة:

تؤدي نتائج الاختبارات دوراً مهماً في تقويم نتائج العملية التعليمية، حيث تعدُّ الأكثر استخداماً في قياس مستوى التحصيل لدى الطلبة، كما يسهم تحليل نتائج هذه الاختبارات في اتخاذ القرارات ذات الصلة بالبرامج التعليمية المختلفة (الزهراني، 2015). وعلى الرغم من أهمية هذه الاختبارات في العملية التعليمية باعتبارها الطريقة الأساسية في تقويم أداء الطلبة وتحصيلهم، إلا أنَّ الدرجات التي يحصل عليها الطلبة فيها قد لا ترجع إلى مستوى تحصيلهم فقط؛ فالتباين والفروق بين درجات الطلبة من القدرة نفسها قد يعزى إلى عوامل أخرى بعيداً عن معرفتهم بمحتوى الاختبار، ومنها مهارات التعامل مع الورقة الاختبارية أو ما يعرف بالحكمة الاختبارية (Harmon & Morse, 1996).

وحيث إنَّ من صفات الاختبار الجيد أن يمتلك خصائص سيكومترية جيدة، من حيث الصدق، والثبات، والموضوعية، وبالتالي تعد الحكمة الاختبارية من العوامل التي تؤثر على صدق الاختبار، ودقة نتائجه (نوفل وأبو عواد، 2010). وحسب اطلاع الباحثين على نتائج الدراسات السابقة، التي تناولت الحكمة الاختبارية وجود تفاوت في التعامل مع استراتيجيات الحكمة الاختبارية، وأيضاً لاحظ الباحثون كونهم مدرسين في الجامعات العمانية، أنَّ هناك فروقاً وتبايناً في درجات الطلبة من القدرة نفسها قد يرجع بعضها إلى مهارة التعامل مع الورقة الاختبارية. وكذلك أنَّ هناك العديد من الدراسات التي تناولت بناء مقياس الحكمة الاختبارية، والتحقق من خصائصه السيكومترية باستخدام النظرية الكلاسيكية، بينما هناك ندرة في الدراسات التي تناولت بناء مقياس الحكمة الاختبارية والتحقق من خصائصه باستخدام نظرية الاستجابة للفقرة.

لهذا هدفت الدراسة الحالية إلى بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات في سلطنة عمان، والتحقق من خصائصه السيكومترية باستخدام نظرية الاستجابة للفقرة وفق نموذج أندرس. وتتمثل مشكلة الدراسة في الإجابة على السؤال الرئيس الآتي: "هل يحقق مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية افتراضات نظرية الاستجابة للفقرة وفقاً لنموذج أندريش؟"

ويتفرع من هذا السؤال الأسئلة الآتية:

- السؤال الأول: ما مدى مطابقة فقرات مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية لافتراضات نموذج أندريش؟
- السؤال الثاني: ما معاملات الثبات والفصل لفقرات وأفراد مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش؟
- السؤال الثالث: هل يحقق تدرج فقرات مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية فاعلية وفقاً لنموذج أندريش؟

أهداف الدراسة:

سعت الدراسة إلى تحقيق الأهداف الآتية:

1. بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش.
2. التحقق من الخصائص السيكومترية لمقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش.

أهمية الدراسة:

تكمن أهمية الدراسة من خلال الآتي:

أولاً: الأهمية النظرية:

تكتسب الدراسة جانباً من أهميتها النظرية من خلال:

1. تستهدف موضوعاً حيويًا ذا تطبيقات عملية واسعة في ميدان علم النفس التربوي؛ إذ تناولت إحدى السمات الكامنة لدى الطالب وهي الحكمة الاختبارية، والتي يمكن الاستفادة منها من قبل العاملين والمهتمين في إثراء معارفهم حول تأثير الحكمة الاختبارية على صدق المقاييس ودقة النتائج، وإثراء المكتبات العلمية بها.
2. توفير مقياس للحكمة الاختبارية تم التحقق من خصائصه السيكومترية باستخدام نظرية الاستجابة للفقرة، يقيس مستوى الحكمة الاختبارية لدى طلبة الجامعات.
3. استخدام نموذج أندريش في تدرج مقياس الحكمة الاختبارية تبعاً لصعوبة مفرداته، يحقق أهداف القياس الموضوعي، وبالتالي التغلب على مشكلة اعتبار مفردات تصحيح المقياس ذات صعوبة واحدة.

ثانياً: الأهمية التطبيقية:

كما تكتسب الدراسة جانباً تطبيقياً من خلال:

1. التوصل إلى أداة على درجة من الموضوعية، لقياس الحكمة الاختبارية، من خلال بناء فقرات مقياس متحررة من خصائص الأفراد، وتحرر الأفراد من خصائص فقرات المقياس.
2. يستفيد منها التربويون العاملون في مجال القياس والتقويم ومجال علم النفس التربوي من معرفة استراتيجيات الحكمة الاختبارية التي يتقنها الطلبة، وبالتالي إعداد اختبارات يتجنب من خلالها هذه الثغرات؛ والتي تؤثر على صدق القياس.

حدود الدراسة:

اقتصرت الدراسة الحالية على الحدود الآتية:

- الحدود المكانية: الجامعات الحكومية، والخاصة في سلطنة عمان.
- الحدود البشرية: طلبة الجامعات العمانية في مختلف التخصصات، والمراحل الدراسية.
- الحدود الزمانية: تم تطبيق أدوات الدراسة خلال العام الدراسي 2023-2024.
- الحدود الموضوعية: بناء مقياس الحكمة الاختبارية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش.

مصطلحات الدراسة:**الحكمة الاختبارية Test-Wisness:**

تعرف بأنها قدرة معرفية أو مجموعة من مهارات مختلفة يستخدمها الفرد لتعديل درجته وتحسينها، ورفعها في الاختبار دون أن يكون لذلك علاقة بالمحتوى المقاس (Sarnacki, 1979). وتعرف إجرائياً: بأنها الدرجة التي يحصل عليها المفحوص من طلبة الجامعات العمانية على مقياس الحكمة الاختبارية المصمم في الدراسة.

نظرية الاستجابة للفقرة (IRT) Item Response Theory:

تعرف بأنها نظرية ذات اتجاه معاصر، وحديث في القياس النفسي والتربوي، تقوم على إمكانية التنبؤ بأداء الفرد على اختبار ما وفق خاصية أو سمة مميزة لهذا الأداء من خلال الاستجابة على مفردات الاختبار، وتحديد العلاقة بين سلوك الفرد الملاحظ على الاختبار، وبين السمة المطلوب قياسها (Sedoc & Unger, 2020).

نموذج أندريش Andrich model:

يعرف بأنه أحد النماذج المنبثقة عن نموذج راش أحادي المعلمة، والذي يركز على تدرج المقاييس ذات الاستجابات المتعددة متساوية التدرج، محددًا مجموعة الفقرات المشتركة في بنية التقدير (Thissen & Edwards, 2005). ويعرف إجرائياً: مدى قدرة النموذج على تحديد مجموعة الفقرات المشتركة في بنية تقدير مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية المصمم في الدراسة.

الإطار النظري**الحكمة الاختبارية**

تعد الحكمة الاختبارية من المجالات الخصبة للدراسة والبحث، حيث تعتبر مصدراً من مصادر تباين الخطأ المؤدي إلى الفروق بين الطلاب في التحصيل؛ والتي لا ترجع إلى معرفتهم بمحتوى الاختبار. وهي مهارة يمارسها الطالب للحصول على أعلى الدرجات؛ وبالتالي تعتبر من العوامل المضللة في القياس والتقويم (حسب الله، 2019). وتعددت المفاهيم حول الحكمة الاختبارية، حيث عرّفها جب (Gibb, 1964) بأنها قدرة الفرد على الإجابة الصحيحة، والتي تمكنه من الحصول على أعلى الدرجات. وعرّفها سارنكي (Sarnacki, 1979) بأنها مجموعة من المهارات المستخدمة من قبل المفحوص لرفع درجته في الاختبار، والتي ليس لها علاقة بالمحتوى المقاس. كما عرّفها زهران (2000) بأنها استخدام الفرد للمعلومات المتوفرة في الاختبار؛ والتي تمكنه من الإجابة الصحيحة والأداء السليم في الموقف الاختباري. وكذلك عرّفها العنزي (2014) بأنها مجموعة من المهارات النفسية والمعرفية؛ تمكن الطالب من الحصول على درجة مرتفعة في الاختبار بغض النظر عن محتوى الاختبار.

أهمية الحكمة الاختبارية

للحكمة الاختبارية دور مهم في حياة الطالب؛ وذلك لارتباطها بالعديد من المتغيرات؛ والتي تؤثر بشكل كبير في تنمية الجوانب العقلية والمعرفية لديه. حيث أشار دايموند وايفانس (Diamond & Evans, 1972) إلى وجود علاقة ارتباطية بين الحكمة الاختبارية وبعض مهارات الذكاء، ورفع مستوى الطالب التحصيلي. كما أشار ماك ودوفر (Mckey & Doverspike, 2001) إلى أن تمكن الطلبة من استخدام مهارات الحكمة الاختبارية في المواقف الاختبارية المتنوعة تغير من التوجه السلبي نحو بعض المواد، والاختبارات نحو الإيجابية. كما يرى دودين (2005) أن للحكمة الاختبارية الكثير من الأهمية منها حصول الطالب على درجة عالية في الاختبار مقارنة بطلاب لديه مستوى المعرفة نفسها، والتحصيل لكنة لا يمتلك هذه المهارات. وأيضاً تطوير اتجاهات إيجابية نحو الاختبارات بشكل عام، ونحو المواد الدراسية المرتبطة بها بشكل خاص.

العوامل المؤثرة على الحكمة الاختبارية

تعتبر الحكمة الاختبارية من المهارات التي يكتسبها الطالب عند مروره بمواقف اختبارية متعددة، يكتسب من خلالها القدرة على التعامل مع الورقة الاختبارية للحصول على درجات عالية. حيث تتأثر الحكمة الاختبارية بالعديد من العوامل منها كما أشار عليها (سعادة وعيسى، 2022) وهذه العوامل:

1. السمات الشخصية: وهي ما يتمتع به الطالب من خصائص شخصية تؤثر بصورة مباشرة على سلوكه ومهاراته عند ممارسة الحكمة الاختبارية، والتي تعتمد على ضبط النفس، والالتزان الانفعالي، والقدرة على التغلب على مشاعر القلق و تجنب العصبية، والتسرع، والحفاظ على مستوى عالي من الثقة بالنفس، والسعي لتخفيف الضغوط النفسية ومستوى الدافعية لديه.
2. الخبرة السابقة: تعرّض الطالب إلى المواقف الاختبارية المتنوعة والتي تتضمن أساليب مختلفة؛ تجعل الطالب يكتسب مهارات واستراتيجيات متعددة تساعده على الممارسة الجيدة في الاختبارات، حيث تتأثر الحكمة الاختبارية بالتدريب والممارسة.
3. القدرة العقلية: حيث ترتبط الحكمة الاختبارية بالقدرة العقلية وتؤثر عليها إيجاباً، فكلما كانت القدرة العقلية مرتفعة لدى الطالب؛ ساعد ذلك في امتلاكه مستوى عالياً من مهارات الحكمة الاختبارية من استدلال، واستنباط وغيرها من العمليات المعرفية المرتبطة بالموقف الاختباري.

استراتيجيات الحكمة الاختبارية

تعددت تصنيفات استراتيجيات الحكمة الاختبارية (طاحون، 2010)؛ حيث صنفها نكو (Nitko, 2001) إلى ثلاث استراتيجيات وهي: استخدام الوقت، وتجنب الخطأ، والتخمين. كما صنف كل من واطر وسبير (Watter & Sieber, 1990) استراتيجيات الحكمة الاختبارية إلى ثلاث استراتيجيات وهي: قبل البدء بالإجابة، وأثناء الإجابة عن الاختبار، وبعد الانتهاء من الإجابة على الاختبار. أما كوهين (Cohen, 2007) فقد صنف استراتيجيات الحكمة الاختبارية إلى ثلاث استراتيجيات وهي: استراتيجيات لغوية وتخص اختبار الفهم القرائي، واستراتيجيات إدارة الوقت أي اختبار البدائل من خلال استبعاد بدائل أخرى، واستراتيجيات الحكمة الاختبارية أي اختيار البديل الصحيح من خلال كلمة مفتاحية في النص. كما صنف سارناكي (Sarnakci, 1979) الحكمة الاختبارية إلى خمس استراتيجيات كما يلي:

1. استراتيجية استخدام الاختبار Test-using strategies: وتتمثل في الإدارة والتنظيم الجيد للوقت عند الإجابة عن فقرات الاختبار.
2. استراتيجية تجنب الخطأ Error-avoidance strategies: وتتمثل في التركيز وعدم التشتت الذهني عند الإجابة على الأسئلة الاختبارية.
3. استراتيجية التخمين Guessing strategies: وتتمثل في استخدام التخمين الذكي في اختيار الإجابة الصحيحة.
4. استراتيجية الاستنتاج المنطقي أو الاستنباطي (Deductive reasoning strategies): وتتمثل في الاستغلال الأمثل للمعلومات المتوافرة في أسئلة الاختبار للإجابة عن باقي الأسئلة.
5. استراتيجية استخدام المنبهات أو المؤشرات strategies (Intent consideration and cue-using): وتتمثل في القدرة على تمييز الإجابة الصحيحة عن الخاطئة باستخدام المؤشرات الموجودة في أسئلة الاختبار، من خلال إدراك مصمم الاختبار، والهدف من الاختبار.

حيث تم اعتماد هذه الاستراتيجيات في بناء مقياس الحكمة الاختبارية المعد في الدراسة الحالية، إذ كانت أكثر الاستراتيجيات شمولاً لمهارات الحكمة الاختبارية.

نموذج أندريش (Andrich Model)

يعدّ نموذج سلم التقدير لأندريش من أنسب نماذج نظرية الاستجابة للفقرة والمطور عن نموذج راش (نموذج أحادي المعلمة) لتدريج المقاييس النفسية والتربوية؛ حيث يستخدم لتدريج المقاييس ذات التدريج المتصل، والذي وضع بصفة خاصة نمط ليكرت لقياس سمة ما. كما يتم تقدير قيمة التدريج لكل فقرة لتحديد موقع الفقرة على متصل السمة المقاسة (فخرو وآخرون، 2009). ويقوم نموذج أندريش على فكرة أنّ كل فقرة من فقرات المقياس تحمل شحنة انفعالية إجمالية تعبر عن الاتجاه نحو العمل المهني لدى الفرد، وبما يتفق مع تقديره لتلك الفقرة، حيث يقوم نموذج أندريش (سلم التقدير) بتقدير هذه الشحنة وفق الدالة الرياضية الاحتمالية لهذا النموذج. وقد صمم نموذج أندريش خصيصاً للاستجابات المتعددة للبيانات من نوع ليكرت (Lustina, 2004). يعرف علم (2002) نموذج أندريش بأنه نموذج منبثق عن النموذج أحادي المعلمة أو ما يعرف بنموذج راش، والذي بياناته مستمدة من سلم التقدير، وتكون المسافات الفاصلة بين التدريجات متساوية، ومتعدد الاستجابات. كما أشار أبو جراد وحمودة (2019) إلى أنه نموذج تم تطويره عن نموذج راش ذي شكل متعدد الاستجابات، وتكون المسافات متساوية تفصل بين تدريجاتها. كذلك عرّف لانج وتاي (Lang & Tay, 2021) بأنه أبسط نماذج نظرية الاستجابة للفقرة أحادي البارامتر، يهدف إلى قياس السمة أو السلوك بشكل موضوعي، ومعالجة صعوبة الفقرات، ويفترض أنّ التمييز متساو بين كل فقرة، ويستخدم في الفقرات ذات الاستجابات الثنائية.

خصائص نموذج أندريش

يتميز نموذج أندريش بوجود عتبات تعبر عن الحدود بين الخطوات، وتكون ذات ثبات عبر المفردات (De & Ayula, 1993). كما يفترض هذا النموذج أنّ تكون قيم عينات الاستجابة متسقة عبر الفقرات على مقياس التقدير المحدد، وتكون المسافة بين صعوبة الخطوة في قسم من أقسام الاستجابة إلى القسم الذي يليه ثابتة عبر جميع الفقرات (الطراونة، 2020). ويفترض هذا النموذج أيضاً وجود تساوي لعدد القيم التي تأخذها الفقرات التي يحتويها المقياس. ويعدّ هذا النموذج حالة خاصة من نموذج التقدير الجزئي حيث تتجزأ درجة صعوبة الفاصل بين أي قيمتين متتاليتين للفقرة (التقي، 2009). في نموذج سلم التقدير لأندريش يتم تقدير قيمة واحدة لكل عنصر في المقياس نفسه، وفي الوقت نفسه يتم تقدير مجموعة من عتبات الاستجابة على نطاق كامل، بحيث يكون عدد فئات الاستجابة أكبر من عدد العينات التي تم تقديرها في المقياس. كما حدّد أندريش احتمال الاستجابة في فئة معينة وفقاً لما يلي $(x = 0, 1, \dots, m_i)$ لكل فقرة. ويتم حساب القيمة في نموذج سلم التقدير لأندريش وفق المعادلة الرياضية الآتية (Andrich, 1978):

$$P_{ix}(\theta) = \frac{\exp \sum_{j=0}^x [(\theta - (B_i + J_j))]}{\sum_{r=0}^{m_i} \exp \sum_{j=0}^r [(\theta - (B_i + J_j))]}$$

حيث إن: $\sum_{j=0}^0 [(\theta - (B_i + T_i))] = 0$

حيث من المحتمل أن تستجيب (Px) في فئة استجابة معينة (X) للفقرة، والتي تعتبر قيمة المقياس للعنصر (B_i) وهو مستوى السمة أو القدرة لفرد معين، ويكون (e) ، (i) هي القيمة الحدية للاستجابة المستهدفة. في حين (T_i) ، (i) ليس في صياغة المعادلة بل عبارة عن ثابت، وبالتالي يفترض أن يكون التمييز متسقاً في جميع العناصر (French & Dodd, 1999).

افتراضات نموذج أندريش

هناك مجموعة من الافتراضات لنموذج أندريش التي يتم التحقق منها عند بناء المقاييس المختلفة في ضوء نظرية الاستجابة للفقرة وهذه الافتراضات هي:

أولاً: مواعمة فقرات المقياس لافتراضات قياس راش

تشير مواعمة الفقرات إلى تحليل مواعمة فقرات المقياس لنموذج أندريش المنبثق عن نموذج قياس راش Rasch measurement models وتتمثل محكات المواعمة فيما يلي:

1. متوسط المربعات (MNSQ) Mean Square: تستخدم المطابقة الداخلية (INFIT) والخارجية (OUTFIT) لتحديد الفرق بين النموذج الإحصائي، والبيانات الملاحظة (Gustafson, 1980). وتتراوح قيم MNSQ من صفر إلى ما لانهاية، وتعتبر القيمة (1) مثالية، وتكون القيمة مقبولة ومناسبة إذا وقعت بين 0.5-1.5 (Linacre, 2012).
2. قيم زاي الإنتاجية The productive Zstd value: وتتراوح قيم زاي الإنتاجية بين (+2، -2) (Bond & Fox, 2007). ويمكن تجاهل هذه القيمة إذا تم قبول قيمة MNSQ (Linacre, 2005).
3. قيمة الخطأ المعياري The Standard Error Value: ويقصد به زيادة القدرة على تقدير أخطاء القياس أو تقدير الانحراف المعياري لتوزيع العينة المرتبط بطريقة التقدير، ويفضل أن لا تتجاوز قيمة الخطأ 0.15 (Brannick, 2003).
4. قيم الارتباط النقطي PTMEA CORR VALUE: ويشير قياس الارتباط النقطي (بيرسون) بين الاستجابات المسجلة Scored Responses، ومقياس القدرة Ability Measures لمعرفة مدى التوافق بين الاستجابات على الفقرة، وقدرات الأفراد ويتوقع أن يكون للفئة الأعلى ارتباط قوي بالقدرة وتعتبر القيمة 0.30 هي الحد الأدنى المقبول لهذا الارتباط (Linacre, 2012).

ثانياً: الأداء التفاضلي (DIF) Differential Item Functioning

يهدف هذا التحليل لتحديد فيما إذا كان هناك أداء تفاضلي على فقرات المقياس بين مجموعتين مختلفتين كالجنس مثلاً كمؤشر على صلاحية الفقرات. يفترض ألا يقل مستوى الدلالة الإحصائية عن مستوى الدلالة (0.05) كدليل إلى عدم وجود أداء تفاضلي (عدم وجود تحيز) بين الذكور والإناث على جميع فقرات المقياس، بالإضافة لذلك تم تحديد قيمة مؤشر الأداء التفاضلي لمقياس ليكرت، ومن المفترض ألا تزيد عن (0.50 Logits) وتعتبر القيم الأقل من ذلك غير معنوية (Linacre, 2012).

ثالثاً: أحادية البعد Unidimensionality

يشير مفهوم أحادية البعد إلى أن فقرات المقياس تقيس سمة أو بنية واحدة فقط، ويمثل مؤشراً على صدق بناء المقياس (Hambleton & Swaminathan, 1985). يمكن الكشف عن أحادية البعد باستخدام تحليل المكونات الرئيسي Principal Component Analysis (PCA).

رابعاً: الاستقلال المحلي Local Independence

يشير افتراض الاستقلال الموضوعي إلى أن استجابات المفحوصين على فقرات المقياس المختلفة مستقلة إحصائياً (Hambleton & Swaminathan, 1985). يتم التحقق من الاستقلال الموضوعي من خلال قيم ارتباط البواقي الملاحظة Observation Residual (Q3) Correlation Value بين أزواج الفقرات، ويجب ألا تتجاوز قيمته |0.30| (Christensen. et.al., 2017).

خامساً: تساوي القدرة التمييزية للفقرات Equal-Item-discriminations

يشير افتراض تساوي القدرة التمييزية للفقرات في RSM إلى أن لفقرات قيماً تمييزية متساوية، وقيمتها تساوي (1)، إلا أنها تجريبياً لا تساوي تماماً (Linacre, 2012).

الدراسات السابقة

هناك العديد من الدراسات التي استهدفت الحكمة الاختبارية، تضمن البعض منها بناء مقياس الحكمة الاختبارية، وتضمن البعض الآخر استخدام الحكمة الاختبارية وعلاقتها ببعض المتغيرات، وكذلك تناولت بعضها استخدام نموذج أندريش في تطوير المقاييس النفسية والتربوية، وفيما يلي عرض لهذه الدراسات مرتبة زمنياً من الأحدث إلى الأقدم.

دراسة السعودي وأبي شندي (2023) Alsoudi & Abu Shindi

الهدف: هدفت إلى تحليل نموذج راش (RSM) لفحص مجموعة من الخصائص السيكومترية للنسخة العربية من مقياس النزعة للتفكير الناقد. المنهجية: قام الباحثان بتطبيق النسخة العربية من مقياس النزعة للتفكير الناقد على عينة بلغت 251 (من أصل 324) من طلبة وطالبات جامعة الشرقية المسجلين في مساق القياس والتقييم الصفي من البكالوريوس، دبلوم التأهيل التربوي. النتائج: توصل الباحثان إلى النتائج التالية: أن جميع فقرات مقياس النزعة للتفكير الناقد باستثناء البند (11) حققت فرضيات نموذج راش، حيث أظهرت ملاءمة جيدة لتحليل الملاءمة، وأحادية البعد، والاستقلال المحلي، والمساواة في التمييز، والموثوقية، ومعاملات الفصل. كذلك تبين أن فئات المقياس تتزايد باطراد وبالتالي هذا يدل على أن النسخة العربية من المقياس تتمتع بخصائص سيكومترية جيدة.

دراسة الجهني (2022):

الهدف: هدفت إلى بناء مقياس الحكمة الاختبارية، والكشف عن مدى استخدام طالبات المرحلة المتوسطة بالمدينة المنورة لاستراتيجيات حكمة الاختبار. وكذلك التعرف إلى الفروق في استخدام استراتيجيات الحكمة الاختبارية وفقاً للمستوى التحصيلي، والمستوى الدراسي. المنهجية: قامت الباحثة بتطبيق مقياس الحكمة الاختبارية وفق تصنيف سارناكي مؤلف من (37) مفردة موزعة على خمس استراتيجيات للحكمة الاختبارية وهي: (استخدام الاختبار، تجنب الخطأ، التخمين، الاستنتاج المنطقي، استخدام المنبهات) على عينة مكونة من (474) طالبة اختبرت بصورة عشوائية عشوائية، وفقاً للمستوى التحصيلي، والمستوى الدراسي. النتائج: توصلت الباحثة في دراستها إلى النتائج التالية: أن درجة استخدام استراتيجيات الحكمة الاختبارية من قبل الطالبات كانت عالية في استراتيجية الاستنتاج المنطقي، واستراتيجية التخمين، وبدرجة متوسطة في استراتيجية استخدام الاختبار، واستراتيجية تجنب الخطأ، واستراتيجية استخدام المنبهات. كذلك توصلت الباحثة إلى عدم وجود فروق دالة إحصائية بين طالبات المرحلة المتوسطة وفقاً لمتغير التحصيل، بينما توجد فروق دالة إحصائية بين الطالبات تبعاً لمتغير المستوى الدراسي (أول، ثاني، ثالث متوسط) لصالح طالبات الصف الأول المتوسط في استراتيجية استخدام الاختبار، واستراتيجية التخمين.

دراسة قنديل وموسى (2022):

الهدف: هدفت إلى إيجاد طريقة موضوعية تقيس وجود الحكمة الاختبارية، ومستواها لدى طلبة كلية التربية الرياضية جامعة مدينة السادات، حيث تهدف إلى تصميم مقياس للحكمة الاختبارية لديهم. المنهجية: قام الباحثان ببناء مقياس للحكمة الاختبارية مكون من (77) مفردة موزعة على ستة محاور وهي: (التهيئة الذاتية، وإدارة الوقت، وصيغ الأسئلة، وخطة الحل، والتخمين، والمراجعة). النتائج: توصل الباحثان في دراستهما إلى النتائج التالية: أن غالبية طلبة كلية التربية بجامعة مدينة السادات يتميزون بدرجة عالية من الحكمة الاختبارية عند أداء الاختبارات.

دراسة النصاروين وآخرين (2022) Alnasraween et al.

الهدف: هدفت إلى الكشف عن مستوى حكمة الاختبار لدى الطلبة في جامعة عمان العربية بالأردن أثناء التعلم عن بعد. المنهجية: قام الباحثان باستخدام مقياس للحكمة الاختبارية الذي طوره الباحث حمادنة (2019) مكون من (28) فقرة، ويتضمن أربع استراتيجيات لحكمة الاختبار (تجنب الخطأ، التفكير الاستنباطي، التخمين، استغلال الوقت) طبق على عينة مكونة من (391) طالبا وطالبة من طلبة جامعة عمان العربية بالأردن. النتائج: توصل الباحثان في دراستهم إلى النتائج التالية: وجود مستوى عال من الحكمة الاختبارية لدى طلبة أفراد العينة. كذلك لا توجد فروق ذات دلالة إحصائية عند مستوى دلالة (0.05) بين المتوسطات الحسابية للدرجة الكلية لحكمة الاختبار تعزى لمتغيرات الدراسة، كذلك أظهرت النتائج وجود فروق دالة إحصائية عند مستوى دلالة (0.05) بين المتوسطات الحسابية لبعد تجنب الخطأ المنسوبة إلى أعضاء هيئة التدريس لصالح كليات العلوم الإنسانية. كما أظهرت النتائج وجود فروق ذات دلالة إحصائية عند مستوى دلالة 0.05 بين المتوسطات الحسابية لبعد التفكير الاستنتاجي المنسوبة للبرنامج الأكاديمي لصالح طلبة الماجستير، والبكالوريوس.

دراسة السعدي (2022):

الهدف: هدفت إلى الكشف عن الأداء على مقياس فرط الاستثارة المدرج وفق نموذج أندريش لدى عينة من الطالبات الموهوبات بالمرحلة الثانوية بمكة المكرمة. المنهجية: طبق الباحث مقياس فرط الاستثارة للباحثة السليمان 2016 والمكون من (50) مفردة تغطي 5 أبعاد على عينة مكونة من (575) طالبة موهوبة من المرحلة الثانوية في مدارس مدينة مكة المكرمة. النتائج: توصل الباحث في دراسته إلى النتائج التالية: تم تدرج مقياس فرط الاستثارة باستخدام نموذج أندريش على ميزان تدرج خطي واحد حيث بلغ عدد مفرداته (39) مفردة بعد حذف المفردات غير الملائمة لأسس القياس الموضوعي والبالغ عددها (11) مفردة.

دراسة السلمي وسان (2022):

الهدف: هدفت الدراسة الكشف عن درجة مطابقة الاستجابات لفرقات مقياس الذكاءات المتعددة مع نموذج أندريش، والقيم المتحررة لقدرات الأفراد، ولصعوبة الفقرات الناتجة عن تطبيق مقياس الذكاءات المتعددة على طلبة الجامعات السعودية، وتحديد الخصائص السيكو مترية للمقياس وفق نموذج سلم التقدير لأندريش. المنهجية: قام الباحثان باستخدام قائمة ماكينزي للذكاءات المتعددة تم تطبيقها على عينة بلغت (726) طالبا وطالبة من طلبة جامعة أم القرى، وجامعة الطائف في منطقة مكة المكرمة.

النتائج: توصل الباحثان إلى النتائج التالية: ملاءمة بيانات المقياس لنموذج أندريش المنبثق عن نموذج راش، والتحقق من درجة المطابقة للاستجابة عن مقياس الذكاءات المتعددة مع نموذج سلم التقدير لاندريش. كما تم التحقق من صدق المقياس، وثباته وكذلك تبين مدى اتساق نمط استجابات كل فرد مع استجابات معظم الأفراد.

دراسة البلوي (2022):

الهدف: هدفت إلى بناء مقياس للتمتمر الإلكتروني وفق نموذج أندريش المنبثق عن نموذج راش لدى طلبة جامعة طيبة بالعلا في المملكة العربية السعودية. المنهجية: قام الباحث ببناء مقياس للتمتمر الإلكتروني مكون من (39) فقرة موزعة على أربعة أبعاد للتمتمر الإلكتروني. طَبَّقَ الباحث الأداة على عينة مكونة من (650) طالبا وطالبة من طلبة جامعة طيبة بالعلا.

النتائج: توصل الباحث في دراسته إلى النتائج التالية: بعد جمع استجابات أفراد العينة وتحليلها وفق نموذج أندريش تم الإبقاء على (30) فقرة من فقرات المقياس موزعة على الأبعاد الأربعة. كذلك تمتع المقياس بدلالات صدق وثبات جيدة حيث بلغت قيمة معامل الثبات للمقياس (0.97).

دراسة يوسف (2022) Youssef:

الهدف: هدفت إلى إعداد أداة لقياس مستوى الكفاءة الأكاديمية لدى طلاب جامعة القاهرة بحيث يكون لها خصائص قياس مقبولة. المنهجية: قام الباحث بتطبيق المقياس المبني وفق تصنيف أندريش على عينة بلغت (577) طالبا وطالبة من طلاب جامعة القاهرة حيث أصبح العدد بعد حذف (100) طالب وطالبة لعدم استيفاء الشروط. النتائج: توصل الباحث في دراسته إلى النتائج التالية: بعد التأكد من صلاحية الأداة عن طريق طريقة القياس التقليدية، والحديثة تكونت الأداة من (48) مفردة تم توزيعها إلى (5) أبعاد.

دراسة روبيرسون (2020) Roberson:

الهدف: هدفت إلى معرفة المساهمة النسبية بين كل من الحكمة الاختبارية، والخلفية المعرفية في الأداء العالي على الاختبار. والمكون من سؤال يحتوي على بديل واحد، وأربعة بدائل تقليدية متعددة الاختيارات. المنهجية: طبق الباحث دراسته على عينة بلغت (183) من الطلبة الجامعيين المسجلين في الدورة التمهيدية لعلم النفس في جامعة ولاية الميسيسيبي، مع إزالة (8) طلاب بسبب أخطاء فنية أو عدم اكتمال المهام. النتائج: توصل الباحث في دراسته إلى النتائج التالية: بالنسبة لحالة البديل الفردي حصل المشاركون على دقة قدرها (42.2%) أي أعلى بنسبة (17.2%) من الفرصة الأساسية البالغة (25%) وبمتوسط دقة بلغت (40.75%) لحالة الاختيار من متعدد. كذلك أن المشاركين الذين ذكروا أنهم تعلموا المادة مسبقاً لم يظهروا أي اختلاف في الدقة مقارنة بأولئك الذين خمنوا، وأيضاً تشير النتائج إلى أن الاختبارات قد يكون تضخمت في الدرجات التي تعكس الحكمة الاختبارية، والمعرفة السابقة أكثر من التعلم التقليدي لمواد الاختبار.

التعليق على الدراسات السابقة

يتضح من خلال مراجعة الدراسات السابقة أنّ هناك بعض الدراسات اتفقت في الكشف عن مستوى حكمة الاختبار لدى الطلبة مثل: دراسة (Alnasraween et al., 2022). وهناك بعض الدراسات اتفقت في بناء مقاييس في الحكمة الاختبارية والتحقق من خصائصها بالطرق التقليدية مثل: دراسة (الجهني، 2022؛ قنديل وموسى، 2022)، لكن هذه الدراسات اختلفت مع دراسة الطراونة (2020) القائمة على تطوير مقياس الحكمة الاختبارية باستخدام النظرية الحديثة في القياس. هناك بعض الدراسات التي تناولت تصميم مقاييس مختلفة في ضوء نموذج أندريش مثل: دراسة السعيد (2022) التي ركزت على مقياس فرط الاستثارة مدرج وفق نموذج أندريش، ودراسة السلمي وسنان (2022)، والتي تناولت على مطابقة الاستجابات على فقرات مقياس الذكاءات المتعددة مع نموذج أندريش، ودراسة Yousf (2022) التي تناولت بناء مقياس مستوى الكفاءة الأكاديمية وفق تصنيف أندريش، ودراسة البلوي (2022) تناولت بناء مقياس للتمتمر الإلكتروني وفق نموذج أندريش، وأيضاً دراسة الطراونة (2020) التي ركزت على تطوير مقياس للحكمة الاختبارية مقنن على طلاب الجامعات في البيئة الأردنية وفق نظرية الاستجابة للفقرة.

منهج الدراسة:

اعتمدت الدراسة على المنهج الوصفي؛ والذي يعدّ من أكثر المناهج استخداماً في مثل هذه الدراسات. حيث يصف الظاهرة ويوضح خصائصها، ويقوم بجمع معلومات دقيقة عن الظاهرة وتصنيفها وتحليلها من خلال جمع معلومات وصفية، وكمية عنها،

والذي يلائم طبيعة الدراسة، وأهدافها من حيث بناء مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش.

مجتمع الدراسة:

تكوّن مجتمع الدراسة من جميع طلبة جامعات سلطنة عمان الحكومية، والخاصة والبالغ عددهم (113523) طالباً وطالبة، وفقاً لإحصائية وزارة التعليم العالي في سلطنة عمان خلال العام الأكاديمي 2023-2024.

عينة الدراسة:

تكوّنت عينة الدراسة من (769) طالباً وطالبة من طلبة الجامعات الحكومية والخاصة بمختلف التخصصات، والجنس، والجامعات. وتم استثناء منهم (50) طالباً وطالبة كانت استجاباتهم غير مكتملة على المقياس. وبالتالي يصبح العدد الكلي للعينة النهائية (719). تم اختيارهم بالطريقة العشوائية العنقودية، إذ تم بداية اختيار ثلاث جامعات، وهي: (جامعة الشرقية، وجامعة نزوى، وجامعة السلطان قابوس)، ثم تم اختيار مجموعة من شعب مساقات متطلبات الجامعة من كل جامعة؛ كون هذه الشعب تحتوي على طلبة من مختلف التخصصات، والسنوات الدراسية، حيث قام الباحث بتطبيق أداة الدراسة ورقياً على عينة الجامعات، ويوضح الجدول (1) خصائص توزيع أفراد عينة الدراسة.

جدول (1): خصائص عينة الدراسة

المتغير	الفئة	العدد	النسبة المئوية
الجنس	ذكر	255	35.47%
	أنثى	464	64.53%
التخصص	علمي	301	41.86%
	إنساني	418	58.14%
الجامعة	الشرقية	251	34.90%
	السلطان قابوس	238	33.12%
	نزوى	230	31.98%
المجموع		719	100%

أداة الدراسة:

تم بناء مقياس في الحكمة الاختبارية وفقاً للخطوات الآتية:

أولاً: مراجعة الأدب النظري، والدراسات السابقة، والشبكة المعلوماتية، والمجلات، والدوريات، ورسائل الماجستير، وأطروحات الدكتوراه مثل: (اسماعيل، 2009؛ أحمد، 2019؛ أبو هاشم، 2008؛ بني عواد، 2022؛ الجهني، 2022؛ حماد، 2010؛ الحسن، 2022؛ خليل، 2019؛ الخولي، 2018؛ الزهراني، 2015؛ Sarnacki, 1979). علماً بأن المقياس استند إلى استراتيجيات الحكمة الاختبارية وفقاً لتصنيف سارنكي (Sarnaki, 1979)، والمتضمن خمس استراتيجيات، وهي:

1. استراتيجية استخدام الاختبار: قدرة الطالب على إدارة الوقت وتنظيمه في الإجابة عن أسئلة الاختبار، بحيث يضمن الاستغلال الأمثل للزمن المحدد له، وأيضاً القدرة على التحكم في وقت الاختبار.
2. استراتيجية تجنب الخطأ: قدرة الطالب على تركيز الانتباه أثناء الإجابة عن أسئلة الاختبار، والابتعاد عن المشتتات التي تؤثر على التركيز، وبالتالي تعتبر مهارة يستطيع الطالب من خلالها تجنب الوقوع في الخطأ أو تدارك الخطأ أثناء الإجابة على الأسئلة.
3. استراتيجية التخمين: قدرة الطالب على استخدام التخمين الذكي في الإجابة عن أسئلة الاختبار، بما يضمن الحصول على أعلى الدرجات مع تجنب التخمين الخطر.
4. استراتيجية الاستنتاج المنطقي أو الاستنباطي: قدرة الطالب على الاستفادة من المعلومات المطروحة في الاختبار أو الأسئلة السابقة في الاختبار، أو أن يستدل على الإجابة أو يستنتجها بواسطة التحليل المنطقي الصادق، أو باستخدام معلومات معينة من فقرات أخرى.

5. استراتيجية استخدام المنبهات أو المؤثرات: قدرة الطالب على استخدام المنبهات أو المؤثرات التي تعمل على تمييز الإجابة الصحيحة من الإجابة الخاطئة، من خلال إدراك اهتمام مصمم الاختبار، والهدف من الاختبار.

ثانياً: بناء عدد كبير من الفقرات لمقياس الحكمة الاختبارية من خلال الاستعانة بمجموعة من المصادر مثل: مثل دراسات: (اسماعيل، 2009؛ أحمد، 2019؛ أبو هاشم، 2008؛ بني عواد، 2022؛ الجهني، 2022؛ حماد، 2010؛ الحسن، 2022؛ خليل، 2019؛ الخولي، 2018؛ الزهراني، 2015؛ سعادة وعيسى، 2022؛ سليمان، 2014؛ السلمي، 2018؛ الشمري، 2019؛ الشمري والسعدي، 2018؛ ضحا وآخرون، 2020؛ الطراونة، 2022؛ العبيدي، 2023؛ العرسان، 2022؛ القرشي، 2017؛ قنديل وموسى، 2022؛ المالكي، 2010؛ نصار، 2021؛ وادي، 2013؛ Sarnacki، 1979؛ Outum et al.، 2015؛ Alnasraween et al.، 2022). وتم تدرجها بأسلوب ليكرت، حيث بلغ عدد فقرات المقياس في صورته الأولية (80) فقرة، موزعة على خمس استراتيجيات للحكمة الاختبارية، بواقع (16) فقرة في كل استراتيجية من الاستراتيجيات السابقة.

ثالثاً: تم عرض فقرات المقياس على عدد من المحكمين الخبراء، والمختصين في القياس والتقويم، وعلم النفس التربوي من مختلف الجامعات في سلطنة عمان، والجامعات العربية يبلغ (11) محكم؛ للنظر في فقرات المقياس من حيث وضوحها، وسلامة التعبير فيها، وملائمتها لاستراتيجيات الحكمة الاختبارية. وبعد عملية التحكيم تم اتفاق أكثر من (8) محكمين على حذف (20) فقرة، وتعديل (7) فقرات ليصبح المقياس في صورته النهائية مكوناً من (60) فقرة موزعة على خمس استراتيجيات للحكمة الاختبارية بواقع (12) فقرة لكل استراتيجية.

رابعاً: تطبيق مقياس الحكمة الاختبارية على عينة استطلاعية للتحقق من وضوح الفقرات للطلبة، وكذلك لتقدير الزمن المناسب لتطبيق المقياس، حيث تم تطبيقه على عينة استطلاعية مكونة من (20) من طلبة جامعة الشرقية بمختلف التخصصات، والجنس، والسنة الدراسية تمثل مجتمع الدراسة، ومن خارج العينة النهائية؛ وذلك للتحقق من: مدى وضوح فقراته، وعدم وجود غموض فيها، وكذلك لتقدير الزمن المناسب لتطبيق المقياس، حيث تم تقدير الزمن المناسب للإجابة على جميع فقرات المقياس بواقع (15) دقيقة، وذلك من خلال إيجاد الفرق الزمني بين أول طالب ينهي الإجابة على جميع فقرات المقياس، وآخر طالب ينهي الإجابة عن جميع الفقرات. وأظهرت الدراسة الاستطلاعية وضوح الفقرات، وعدم وجود غموض فيها؛ لهذا تم الاحتفاظ بجميع فقرات المقياس.

المعالجة الإحصائية:

للولصول لنتائج الدراسة تم تحليل البيانات باستخدام برنامج Winsteps Rasch وباستخدام المعالجات الإحصائية الآتية:

1. إحصائيات متوسط المربعات (MNSQ)، وقيم زاي الإنتاجية (The Productive Zstd Value)، وقيم الخطأ المعياري (The Standard Error Value)، وقيم الارتباط النقطي (Point-Measure Correlation Value (PTMEA) للكشف عن مطابقة فقرات المقياس والأفراد الداخلية (INFIT) والخارجية (OUTFIT) لنموذج أندريش.
2. تحليل المكونات الرئيسية (PCA) Principal Component Analysis للكشف عن أحادية البعد في المقياس.
3. قيم ارتباط البواقي الملاحظة (Q3) Observation Residual Correlation Value للكشف عن الاستقلال الموضعي بين أزواج الفقرات.
4. اختبار فان دين ولينبرغ (Q1) of Parallel ICCs/IRFs Van Den Wollenberg Test للتحقق من افتراض تساوي القدرة التمييزية للفقرات.
5. اختبار مانتل هانزل (Mantel-Haenszel Method)، واختبار راش - ولش (الانحدار اللوجستي) (Rash-Welch (Logistic Regression) t-test للكشف عن الأداء التفاضلي في فقرات المقياس وفقاً لمتغير الجنس.
6. مؤشري الثبات والفصل Reliability & Separation Index للتحقق ثبات الأفراد والفقرات، وكذلك قدرة المقياس على الفصل بين الأفراد في البناء، والفصل بين مستويات صعوبة الفقرات.
7. معايرة مقياس التقدير Rating Scale Calibration للتحقق من فاعلية فئات الاستجابة على المقياس.

عرض النتائج ومناقشتها:

النتائج المتعلقة بالسؤال الأول ومناقشتها:

للإجابة عن السؤال الأول الذي نصه "ما مدى مطابقة فقرات مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية لافتراضات نموذج أندريش؟" تم تحليل البيانات باستخدام البرنامج الإحصائي (Winsteps)، وللتحقق من مطابقة مقياس الحكمة الاختبارية لافتراضات نموذج أندريش، تم إجراء التحليلات الإحصائية الآتية:

أولاً: مواعمة فقرات المقياس لافتراضات قياس راش

تم التحقق من مواعمة فقرات المقياس لافتراضات نموذج أندريش، ويوضح الجدول (2) نتائج مطابقة فقرات مقياس الحكمة الاختبارية لافتراضات نموذج أندريش.

جدول (2): نتائج مطابقة فقرات مقياس الحكمة الاختبارية لافتراضات نموذج أندريش

ارتباط الفقرة مع المقياس	المطابقة الخارجية OUTFIT		المطابقة الداخلية INFIT		الخطأ المعياري	الصعوبة	رقم الفقرة
	قيم زاي ZSTD	متوسط المربعات MNSQ	قيم زاي ZSTD	متوسط المربعات MNSQ			
.24	4.95	1.39	5.40	1.39	.05	-.39	1
.37	.96	1.06	.57	1.03	.05	.09-	2
.33	5.51	1.44	6.73	1.50	.05	.38-	3
.32	2.96	1.21	2.76	1.18	.05	.30-	4
.33	3.39	1.23	2.67	1.17	.05	.11-	5
.38	-.56	.97	-.88	.95	.04	.98	6
.36	2.74	1.15	2.70	1.15	.04	.59	7
.37	-.09	.99	-.44	.97	.05	.30-	8
.34	2.29	1.14	2.35	1.14	.04	.06	9
.35	.19	1.02	.64	1.06	.08	-1.38	10
.24	2.57	1.29	3.19	1.33	.07	1.30-	11
.32	-1.41	.91	-1.86	.89	.04	.03	12
.30	-.92	.93	-2.41	.85	.05	.47-	13
.34	3.68	1.22	3.60	1.21	.04	.28	14
.27	.28	1.03	1.56	1.18	.09	1.68-	15
.36	-1.13	.94	-1.56	.92	.04	.22	16
.32	3.93	1.27	4.28	1.27	.05	.09-	17
.39	-1.09	.92	-.99	.94	.05	.39-	18
.38	2.10	1.14	2.96	1.19	.05	.11-	19
.26	-.06	.99	.02	1.00	.06	.98-	20
.37	2.25	1.21	3.44	1.30	.06	.90-	21
.32	-.57	.94	.99	1.09	.07	1.29-	22
.32	-3.04	.82	-3.60	.80	.05	.11-	23
.36	.23	1.02	.10	1.01	.06	.84-	24
.30	-1.07	.92	-1.75	.88	.05	.58-	25
.25	-.06	.99	-.64	.96	.05	.22-	26
.26	3.89	1.23	3.98	1.23	.04	.32	27
.38	4.83	1.29	5.08	1.30	.04	.33	28
.32	1.82	1.12	2.24	1.14	.04	.03-	29
.33	2.90	1.19	3.15	1.19	.05	.05-	30
.45	2.73	1.16	2.77	1.15	.04	1.29	31

ارتباط الفقرة مع المقياس	المطابقة الخارجية OUTFIT		المطابقة الداخلية INFIT		الخطأ المعياري	الصعوبة	رقم الفقرة
	قيم زاي	متوسط المربعات	قيم زاي	متوسط المربعات			
	ZSTD	MNSQ	ZSTD	MNSQ			
.50	-3.01	.85	-2.86	.86	.04	.83	32
.41	3.79	1.22	3.69	1.21	.04	1.32	33
.47	-.21	.99	-.11	.99	.04	1.25	34
.40	.34	1.02	.49	1.02	.04	1.17	35
.40	2.80	1.16	3.18	1.18	.04	.35	36
.42	-3.39	.82	-3.91	.80	.04	.21	37
.41	-2.22	.86	-2.31	.86	.05	-.22	38
.48	-2.74	.85	-2.78	.85	.04	.22	39
.47	-5.75	.69	-6.32	.68	.04	.02-	40
.38	1.26	1.07	1.77	1.10	.04	.15	41
.48	-3.58	.81	-4.14	.79	.04	.23	42
.50	-3.12	.84	-3.19	.84	.04	.49	43
.48	-4.56	.76	-4.64	.77	.04	.31	44
.47	-1.44	.92	-1.40	.93	.04	.66	45
.47	-6.39	.69	-6.49	.70	.04	.57	46
.47	2.60	1.15	2.80	1.15	.04	.47	47
.48	-2.47	.85	-2.43	.86	.05	.07-	48
.48	-3.03	.84	-3.58	.82	.04	.44	49
.41	-1.55	.90	-1.93	.88	.05	.34-	50
.49	-5.19	.74	-5.46	.74	.04	.47	51
.51	-3.53	.82	-3.63	.82	.04	.36	52
.45	.05	1.00	-.02	1.00	.04	.43	53
.49	-3.72	.80	-3.96	.80	.04	.22	54
.43	-1.56	.90	-1.34	.92	.05	.32-	55
.42	.71	1.04	.70	1.04	.05	.08-	56
.49	-1.64	.90	-1.25	.93	.04	.01-	57
.42	-1.30	.92	-1.41	.92	.05	.15-	58
.42	-.52	.96	-.10	.99	.05	.49-	59
.37	3.66	1.29	5.23	1.39	.05	.47-	60
.39	-.02	1.01	.03	1.02	.05	.00	المتوسط
.08	2.84	.18	3.11	.19	.01	.63	الانحراف

ووفقاً للبيانات المستخدمة في الدراسة الحالية والموضحة في الجدول (2) بلغت قيمة المتوسط الحسابي للصعوبة (0.00)، وهي القيمة التي يفترضها النموذج البالغة (0)، وبلغت قيمة الانحراف المعياري (0.63)، وهي قريبة من القيمة التي يفترضها النموذج البالغة (1) (Linacre, 2012).

كما تراوحت جميع قيم MNSQ للفقرات بين (0.68 - 1.50) في المطابقة الداخلية INFIT، بينما تراوحت جميع قيم MNSQ للفقرات بين (0.69 - 1.44) في المطابقة الخارجية OUTFIT، كما تراوحت قيمة متوسط المربعات MNSQ الكلي للفقرات بين (1.01 ، 1.02) للمطابقة الداخلية والخارجية على التوالي، وتعتبر هذه القيم مقبولة ومناسبة لجميع الفقرات؛ حيث وقعت ضمن المدى (0.5 - 1.5) وفق محكات (Linacre 2012).

وتراوحت قيم زاي الإنتاجية Zstd للمطابقة الداخلية INFIT بين (6.73 - -6.49) بينما تراوحت قيم زاي الإنتاجية Zstd للمطابقة الخارجية OUTFIT بين (-6.39 - 5.51)، وقد تجاوزت مدى محك قيم زاي الإنتاجية الواقع بين (-2، +2)، حيث

كانت هناك مجموعة من الفقرات خارج المدى المقبول لقيم Zstd، إلا أنها حققت جميعها محك MNSQ، كما تراوحت قيم زاي الإنتاجية الكلي للفقرات بين (0.03 ، -0.02) للمطابقة الداخلية والخارجية على التوالي. وتراوحت قيمة الخطأ المعياري The Standard Error Value في الدراسة الحالية حسب الموضح في الجدول (3) بين (-0.04 - 0.09) مما يشير إلى عنصر الدقة في التقدير (Linacre, 2005)، كما اعتبر فيشر (2007) Fisher أنّ نطاق قيمة الخطأ ممتازة. وكانت جميع قيم PTMEA CORR كانت أكبر من الحد الأدنى المقبول وهو |0.30| (Wu & Adam, 2007). حيث تراوحت قيم الارتباط النقطي بين (0.30 و 0.51) باستثناء الفقرات (1, 11, 15, 20, 26, 27)، كانت قيمة PTMEA CORR أقل من الحد (0.30)، بالتالي لم تحقق محك مطابقة الفقرة؛ وعليه تم حذف (6) فقرات من المقياس.

ثانياً: الأداء التفاضلي للجنس (GDIF) Gender Differential Item Functioning

تم إجراء هذا التحليل لتحديد فيما إذا كان هناك أداءً تفاضلياً على فقرات المقياس بين الجنسين الذكور والإناث كمؤشر على صلاحية الفقرات. يستخدم Winsteps اختبار راش - ويلش (الانحدار اللوجستي) Rash- Welch (Logistic Regression) t-test مع مستوى الثقة 95%، وقيمة ت حرجة (2.00) لتحديد دلالة الفروق في صعوبة الفقرة بين مجموعتين، وكذلك طريقة مانتل هانزل Mantel-Haenszel Method باستخدام مربع كاي Chi-Square، كذلك يفترض ألا يقل (يساوي أو أكبر) مستوى الدلالة الإحصائية عن مستوى الدلالة (0.05) كدليل إلى عدم وجود أداء تفاضلي (عدم وجود تحيز) بين الذكور والإناث على جميع فقرات المقياس، بالإضافة لذلك تم تحديد قيمة مؤشر الأداء التفاضلي لمقياس ليكرت، ومن المفترض ألا تزيد عن (0.50) Logits وتعتبر القيم الأقل من ذلك غير معنوية.

وأشارت نتائج الاختبارين إلى أنّ قيم مستوى الدلالة للفروق بين مجموعتي الذكور والإناث كانت غير دالة على جميع الفقرات ماعدا الفقرات (47، 50، 59) كانت قيمة مستوى الدلالة الإحصائية أقل من مستوى الدلالة (0.05)، مما يدل على وجود أداء تفاضلي فيها. وعليه تم حذف الفقرات الثلاثة من المقياس.

ثالثاً: أحادية البعد Unidimensionality

يشير مفهوم أحادية البعد إلى أنّ فقرات المقياس تقيس سمة أو بنية واحدة فقط، ويمثل مؤشراً على صدق بناء المقياس (Hambleton & Swaminathan, 1985). تم الكشف عن أحادية البعد في الدراسة الحالية باستخدام تحليل المكونات الرئيسي Principal Component Analysis (PCA). كما هو موضح في الجدول (3).

جدول (3): تحليل المكونات الرئيسي (PCA) للكشف عن أحادية البعد في المقياس

القيمة المتوقعة للنموذج	التباين الملاحظ	الجذر الكامن	التباين الكلي
100.0%	100.00%	87.63	التباين الكلي
41.6%	41.8%	36.63	التباين المفسر بالمقياس
17.6%	17.7%	15.53	التباين المفسر بالأفراد
24.8%	24.7%	21.09	التباين المفسر بالفقرات
58.3%	58.2%	51.00	التباين غير المفسر الكلي
8.5%	5.7%	4.96	التباين غير المفسر للبناء الأول
7.5%	4.3%	3.80	التباين غير المفسر للبناء الثاني
5.1%	3.0%	2.59	التباين غير المفسر للبناء الثالث
3.8%	2.2%	1.92	التباين غير المفسر للبناء الرابع
3.3%	1.9%	1.68	التباين غير المفسر للبناء الخامس

يتضح من الجدول (3) أنّ نسبة التباين المفسر للمقياس بلغت (41.8%)، وهي أعلى من الحد الأدنى الذي حدده (Linacre, 2012) والبالغ (40%). بالإضافة إلى ذلك كانت قيمة التباين غير المفسر في البناء الأول First Contrast تساوي (5.7%)، وهي جيدة إذ وقعت ضمن المدى (5% - 10%)، في حين كانت قيم التباين غير المفسر من الثاني حتى الخامس ضمن المدى (3% - 5%)، وبالتالي هي جيدة جداً (Fisher, 2007).

كما كانت قيمة الجذر الكامن Eigenvalue للبناء الأول (4.96)، وهي أقل من القيمة التي افترضها (2005) Linacre والبالغة (5). مما يشير إلى عدم وجود بعد ثان في المقياس، كما أن نسبة التباين المفسر بواسطة الفقرات (24.7%) إلى التباين غير المفسر للبناء الأول (5.7%) كانت (4.33)، وهي أكبر ثلاث أضعاف على الأقل (Conrad. et al., 2015). وتشير هذه النتائج مجملًا إلى أن المقياس أجادي البعد.

رابعاً: الاستقلال المحلي Local Independence

يشير افتراض الاستقلال الموضوعي إلى أن استجابات المفحوصين على فقرات المقياس المختلفة مستقلة إحصائياً (Hambleton & Swaminathan, 1985). تم التحقق من الاستقلال الموضوعي من خلال قيم ارتباط البواقي الملاحظة Observation Residual Correlation Value (Q3) بين أزواج الفقرات، ويجب ألا تتجاوز قيمته |0.30| (Christensen. et.al., 2017). ويوضح الجدول (4) قيم ارتباط البواقي العظمى الملاحظة بين فقرات المقياس.

جدول (4): قيم ارتباط البواقي العظمى الملاحظة بين فقرات المقياس

رقم الفقرة	معامل الارتباط بين أزواج الفقرات	رقم الفقرة	معامل الارتباط بين أزواج الفقرات	رقم الفقرة	معامل الارتباط بين أزواج الفقرات
39	0.30	35	0.24	37	0.24
60	0.29	25	0.24	17	0.24
5	0.28	47	0.23	2	0.23
17	0.27	59	0.23	3	0.27
17	0.25	36	0.22	16	0.25
53	0.25	12	0.22	45	0.25
43	0.25	45	0.22	42	0.25
24	0.24	17	0.22	23	0.24
56	0.24	51	0.22	55	0.24
29	0.24	19	-0.22	26	0.24

يوضح الجدول (4) القيم العظمى لارتباطات البواقي الملاحظة في أزواج الفقرات. إذ تراوحت بين (0.22 - 0.30) وكانت جميعها لم تتجاوز (0.30)، مما يشير إلى تحقق افتراض الاستقلال الموضوعي لفقرات المقياس.

خامساً: تساوي القدرة التمييزية للفقرات Equal-Item-discriminations

يشير افتراض تساوي القدرة التمييزية للفقرات في RSM إلى أن للفقرات قيم تمييزية متساوية، وقيمتها تساوي (1)، إلا أنها تجريبياً لا تساوي تماماً (Linacre, 2012). تم تقدير تمييز الفقرات في المقياس الحالي والتي تراوحت بين (0.77 - 1.36)، إذ كان تمييز الفقرات يقترب من القيمة (1) لجميع الفقرات، وتم التحقق من تساوي القدرة التمييزية لهذه القيم باستخدام اختبار فان دين ويلنبرغ Van Den Wollenberg Test (Q1) of Parallel ICCS\IRFS وذلك لفحص الفرضية الصفرية التي تنص على أن تمييز الفقرات متساوي ولا يختلف عن الواحد، وكانت نتائجه كالتالي: (0.107, 50, 66.108, P=Q₁)، ويتضح من قيمة 0.10 < Q₁ < 0.05. وبذلك نقبل الفرضية الصفرية القائلة بأن تمييز الفقرات متساوي، ولا يختلف عن الواحد.

نلاحظ من نتائج التحليل الإحصائي للسؤال الأول أن فقرات مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية يتطابق مع افتراضات نموذج أندريش، ما عدا الفقرات التي تم حذفها عند فحص افتراض موافقة الفقرات؛ لعدم تحقيقها لمحكات الموافاة، وهذه الفقرات هي (1، 11، 15، 20، 26، 27). وكذلك الفقرات (47، 50، 59) تم حذفها من المقياس؛ لوجود تحيز فيها عند فحص افتراض كشف الأداء التفاضلي وفقاً لمتغير الجنس، وبالتالي يصبح المقياس في صورته النهائية مكون من (51) فقرة.

النتائج المتعلقة بالسؤال الثاني ومناقشتها

للإجابة عن السؤال الثاني الذي نصه "ما معاملات الثبات والفصل Reliability & Separation لفقرات وأفراد مقياس الحكمة الاختبارية لدى طلبة الجامعات العمانية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش؟" تم إجراء التحليل الإحصائي

باستخدام برنامج Winsteps؛ للتحقق من معاملات الثبات والفصل Reliability & Separation للفقرات، والأفراد من خلال إحصائيات الأفراد، وإحصائيات الفقرات على مقياس الحكمة الاختبارية، وذلك بعد حذف (9) فقرات من العدد الكلي لفقرات المقياس البالغ (60) فقرة من خلال بعض التحليلات الإحصائية في السؤال الأول. ويوضح الجدول (5) ملخص إحصائيات الأفراد والفقرات.

جدول (5): ملخص إحصائيات الأفراد على مقياس الحكمة الاختبارية

المصدر	الإحصائي	الدرجة الكلية	المعياري للمنموذج	المطابقة الداخلية INFIT		المطابقة الخارجية OUTFIT	
				متوسط المربعات	قيمة زاي	متوسط المربعات	قيمة زاي
الأفراد	الوسط الحسابي	19.0	15.	1.00	-0.07	1.01	ZSTD
	الانحراف المعياري	21.6	03.	.36	1.76	.35	MSNQ
الفقرات	الوسط الحسابي	2026.9	05.	1.02	-0.07	1.01	ZSTD
	الانحراف المعياري	289.4	01.	.19	2.95	.18	MSNQ
الأفراد	معامل الفصل	3.42	معامل الثبات	.92	.14		
الفقرات	معامل الفصل	11.86	معامل الثبات	.99	.05		

يلاحظ من الجدول (5) أنّ قيم متوسط المربعات للمطابقة الداخلية والخارجية للأفراد بلغت (1.00، 1.01) بانحراف معياري بلغ (0.36، 0.35) على التوالي، كما بلغت قيم متوسط المربعات للمطابقة الداخلية والخارجية للفقرات (1.02، 1.01) بانحراف معياري بلغ (0.19، 0.18) على التوالي، وتعدّ هذه القيمة مثالية (Linacre, 2012). في حين كانت قيم زاي للأفراد تساوي (-0.07، -0.15) بانحراف معياري (1.90، 1.76) للمطابقة الداخلية والخارجية على التوالي، وللفقرات بلغت قيم زاي (-0.04، -0.07) بانحراف معياري بلغ (3.23، 2.95) للمطابقة الداخلية والخارجية على التوالي، وتعدّ هذه القيم مثالية أيضاً حيث تراوحت بين (-2.00 - +2.00) (Bond & Fox, 2007). وبلغت قيمة معامل الثبات للأفراد (0.92)، و (0.99) للفقرات وتعد هذه القيم ممتازة وفقاً لمحكات (Fisher, 2007). كما كانت قيمة معامل الفصل للأفراد (3.42)، وللفقرات بلغت (11.86) وتشير هذه القيمة إلى قدرة المقياس على الفصل بين الأفراد، وكذلك تشير هذه القيمة إلى قدرة المقياس على فصل الأفراد، وقدرة الفقرات على التمييز بين الأفراد. وتعتبر قيمة مؤشر الفصل جيدة إذا تجاوزت 3.00 وممتازة إذا تجاوزت 6.00 (Fox & Jones, 1998). يتضح من نتائج التحليل الإحصائي للسؤال الثاني بالاعتماد على مؤشري الثبات والفصل للأفراد، والفقرات على مقياس الحكمة الاختبارية، إلى أنّ لفقرات والأفراد معاملات ثبات وفصل مرتفعة، وأنّ للمقياس القدرة على التمييز بين مستويات الصعوبة.

النتائج المتعلقة بالسؤال الثالث ومناقشتها:

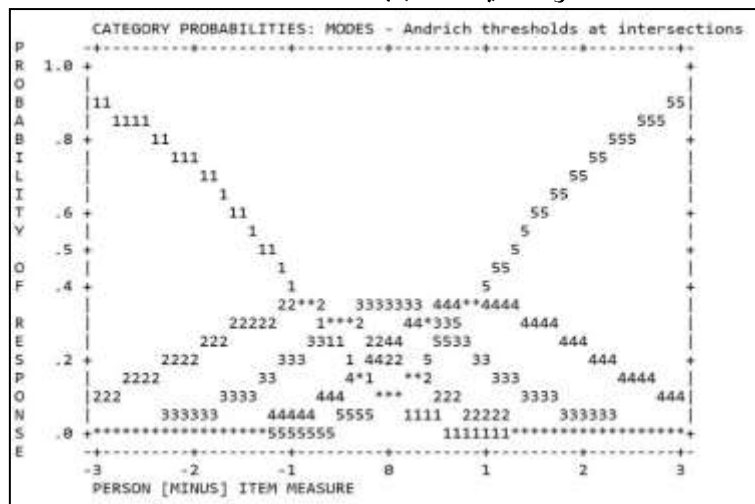
للإجابة عن السؤال الثالث الذي نصه "هل يحقق تدرّج فقرات مقياس الحكمة الاختبارية فاعلية وفقاً لنموذج أندريش؟" تم إجراء معايرة فئات المقياس لتحديد فاعلية المقياس لمعرفة فيما إذا كان يجب تضمين أحد الفئات في فئة أخرى. تم تدرّج مقياس الحكمة الاختبارية إلى خمس فئات هي: (دائماً، غالباً، أحياناً، نادراً، أبداً) ولتحديد فاعلية المقياس تم استخدام خمس محكات للحكم وفقاً للينكر (Linacre, 2002) وهي:

المحك الأول يتمثل في أنّ تحتوي كل فئة على (10) استجابات على الأقل. ويشير المحك الثاني إلى أنّ يتزايد متوسط القياس بخطى ثابتة Lockstep مع مستوى القياس. ويتضمن المحك الثالث قيم متوسطات المربعات MNSQ بحيث تكون ضمن المدى المقبول (0.50 - 1.50). ويشير المحك الرابع إلى أنّ كل فئة يجب أن تظهر ذروة منحنى الاحتمال A Probability Curve Peak، أي أنّ زيادة موقع الفرد على متصل السمة الكامنة (الحكمة الاختبارية) يرتبط باحتمالية متزايدة للتقديرات في الفئات العليا. وأخيراً المحك الخامس يحدد أن يكون الفرق بين عتبة كل فئة والفئة التي تليها ضمن المدى (1.00 - 5.00). ويوضح الجدول (6) ملخص نتائج فاعلية تدرّج مقياس الحكمة الاختبارية.

جدول (6): ملخص نتائج فاعلية تدرّج مقياس الحكمة الاختبارية

الفئات	الدرجة	عدد الأفراد الملاحظ	النسبة المئوية %	المطابقة الداخلية	المطابقة الخارجية	عتبة أندريش	فئة المقياس
				متوسط المربعات	متوسط المربعات		
أبداً	1	1445	5	1.04	1.06	0	-2.26
نادراً	2	2985	11	1.00	1.00	-1.80	-.90
أحياناً	3	6365	23	.99	1.00	-.48	-.01
غالباً	4	7879	28	1.00	.99	.61	.89
دائماً	5	9069	33	.99	1.00	1.87	2.30

يلاحظ من الجدول (6) وفيما يخص المحك الأول أقل عدد من الاستجابات في الفئة الأولى بلغ (1445)، وهو أعلى بكثير من العدد المحدد للمحك وهو (10) استجابات، وبالتالي تحقق المحك الأول. وفيما يخص المحك الثاني نلاحظ من الجدول أن هذا الشرط قد تحقق؛ إذ بلغت قيم قياس الفئات (2.30، .89، .01، -.90، -2.26) للفئات من الخامسة إلى الأولى على التوالي، مما يدل على أنماط استجابة متسقة ومتصاعدة. وبالنسبة للمحك الثالث نلاحظ من الجدول أن هذا الشرط تحقق أيضاً؛ إذ تراوحت قيم متوسط المربعات بين (.99 - 1.04) للمطابقة الداخلية، وبين (0.99 - 1.06) لفئات التدرج الخمس للمقياس أي أنها ضمن المدى المقبول (.50 - 1.50) وفق محكات (Linacre 2012). وفيما يخص المحك الرابع الذي يتضمن أن تظهر كل فئة ذروة منحنى احتمالية، أي أن زيادة موقع الفرد على متصل السمة الكامنة (الحكمة الاختبارية) يرتبط باحتمالية متزايدة للتقديرات في الفئات العليا، وتحقق هذا الشرط كما هو موضح في الشكل (1) الذي يظهر موقع الفرد على متصل السمة الكامنة، واحتمالية الاستجابة.



شكل (1): منحنى الاحتمالية لفئات مقياس الحكمة الاختبارية

يتضح من الشكل (1) الذي يوضح موقع الفرد على متصل السمة الكامنة، وتزايد الاحتمالية للتقديرات في الفئات العليا أنه كلما زادت قدرة الفرد الكامنة؛ يزداد احتمالية الاستجابة لدية على الفئات العليا من المقياس، ويظهر ذلك على يمين الشكل، بينما كلما اتجهت قدرة الفرد الكامنة للمستوى الأدنى نلاحظ احتمالية الاستجابة لدية على فئات المقياس الدنيا ويظهر ذلك يسار الشكل، بينما عندما تكون القدرة الكامنة في المنتصف نلاحظ تركيز الاستجابات على فئات المقياس المتوسطة أيضاً، ويظهر ذلك في منتصف الشكل، مما يحقق شرط المحك الرابع. وأخيراً فيما يخص المحك الخامس يحدد أن يكون الفرق بين عتبة كل فئة والفئة التي تليها ضمن المدى (1.00 - 5.00)، وللتحقق من ذلك تم حساب الفرق بين قيم عتبة أندريش بين كل فئتين متجاورتين في تدرج المقياس، كما هو موضح في الجدول (7).

جدول (7): الفروق بين عتبات فئات تدرج مقياس الحكمة الاختبارية

فئات المقياس	حساب الفجوات	مدى القبول	القرار
الفئة 1 - الفئة 2	(-1.80) - .00	1.00 < 1.80 < 5.00	مقبول

مقبول	1.00<1.32<5.00	- .48 – (-1.80)	الفئة 2 – الفئة 3
مقبول	1.00<1.09<5.00	.61 – (-.48)	الفئة 3 – الفئة 4
مقبول	1.00<1.26<5.00	1.87 – .61	الفئة 4 – الفئة 5

يلاحظ من الجدول أنّ الفرق بين عتبة كل فئة والفئة التي تليها في مقياس الحكمة الاختبارية كانت ضمن المدى المقبول.

خلاصة النتائج:

تم في هذه الدراسة بناء مقياس للحكمة الاختبارية لدى طلبة الجامعات العمانية مكون من (60) فقرة، وتم التحقق من خصائصه في ضوء نظرية الاستجابة للفقرة IRT وفقاً لنموذج أندريش Andrich model؛ وذلك عبر برنامج التحليل الإحصائي Winsteps. تم التحقق من خصائص الفقرات والمقياس من خلال التحليلات الآتية: مواعمة الفقرات، أحادية البعد، الاستقلال المحلي، تساوي القدرة التمييزية للفقرات، الأداء التفاضلي للجنس، مؤشري الثبات والفصل، ومعايرة تدرج المقياس، وفيما يلي ملخص نتائج الدراسة:

- أشارت نتائج مواعمة الفقرات إلى أن جميع الفقرات حققت شروط المطابقة والمتمثلة في: متوسط المربعات MNSQ، قيم زاي الإنتاجية Zstd، قيمة الخطأ المعياري Standard Error، قيم الارتباط النقطي PTMEA مما يدل على ملائمة مناسبة، وذات جودة لفقرات المقياس. باستثناء الفقرات (1، 11، 15، 20، 26، 27) حيث كانت قيمة ارتباطها أقل من |0.30|، عالية تم حذف هذه الفقرات الستة من المقياس.
- تم الكشف عما إذا كان هناك أداء تفاضلي بين الذكور والإناث على فقرات المقياس كمؤشر على صلاحية الفقرات، تم ذلك باستخدام قيم مستوى الدلالة الإحصائية، واستخدام طريقة الانحدار اللوجستي، وطريقة مانتل هانزل. أشارت النتائج إلى عدم وجود تحيز لدى فقرات المقياس باستثناء الفقرات (47، 50، 59)؛ اتفقت طرق الكشف الثلاث على وجود تحيز فيها؛ لهذا تم حذفها من المقياس.
- فيما يخص افتراض أحادية البعد واعتماداً على قيمة الجذر الكامن للبناء الأول، والذي بلغ (4.96) وهي أقل من القيمة التي افترضها (2005) Linacre التي تبلغ (5)، تم الكشف عن عامل واحد سائد فقط في المقياس فسر ما يقارب (41.6%) من التباين في الأداء على المقياس، ويعد ذلك مؤشراً على صدق بناء المقياس (Sumintono & Widhiarso, 2014).
- بالنسبة لافتراض الاستقلال الموضوعي أو المحلي لفقرات المقياس؛ تم استخدام قيم ارتباط البواقي الملاحظة Q3 بين أزواج الفقرات، وتبين أن الاستجابة عن أي فقرة من فقرات المقياس مستقلة عن بعضها البعض؛ مما حقق افتراض الاستقلال المحلي لفقرات المقياس.
- فيما يخص افتراض تساوي القدرة التمييزية للفقرات تم التحقق منه باستخدام اختبار فان دين ويلنبرغ Van Den Wollenberg Test (Q1) of Parallel ICCS\IRFS وأشارت النتائج إلى تساوي القدرة التمييزية للفقرات، مما يؤكد مطابقة الفقرات لنموذج أندريش.
- أشارت نتائج مؤشري الثبات والفصل، للفقرات والأفراد على مقياس الحكمة الاختبارية إلى أن للفقرات والأفراد معاملات ثبات مرتفعة بلغت (0.99، 0.92) على التوالي. كما بلغت قيمة مؤشر الفصل للفقرات والأفراد (11.86، 3.42) على التوالي وهي قيم جيدة للأفراد، وممتازة للفقرات.
- وأخيراً أظهرت نتائج تحليل معايرة فئات المقياس، ووفقاً لخمسة محكات مختلفة أشارت النتائج جميعها إلى أن المقياس ذو فاعلية في تدرج فئاته، وبالتالي إمكانية الاحتفاظ بجميع فئات المقياس، حيث كان قياس الفئات يتزايد بخطى ثابتة عند الانتقال من فئة إلى فئة أخرى. كما كانت المسافة بين فئات التدرج (العتبات) ضمن المدى المحدد (1.00 – 5.00).
- وأخيراً تشير جميع النتائج إلى أن لمقياس الحكمة الاختبارية خصائص سيكومترية جيدة بعد حذف 9 فقرات، وأصبح المقياس في صورته النهائية مكوناً من (51) فقرة موزعة على خمس استراتيجيات للحكمة الاختبارية.

التوصيات والمقترحات:

هدفت الدراسة الحالية إلى بناء مقياس الحكمة الاختبارية في ضوء نظرية الاستجابة للفقرة وفقاً لنموذج أندريش، وفي ضوء نتائج الدراسة تم تقديم التوصيات والمقترحات الآتية:

1. استخدام مقياس الحكمة الاختبارية الذي تم بناؤه والتحقق من خصائصه السيكومترية في هذه الدراسة في دراسات وبحوث أخرى ترتبط بالحكمة الاختبارية.
2. إجراء دراسات حول درجة ممارسة مهارات الحكمة الاختبارية لدى طلبة الجامعات، والمدارس الحكومية في سلطنة عمان، والكشف عن علاقتها بالمتغيرات النفسية والتربوية الأخرى، وكذلك المتغيرات الديمغرافية.
3. استخدام نماذج نظرية الاستجابة للفقرة المختلفة في التحقق من الخصائص السيكومترية عند بناء وتطوير المقياس النفسية والتربوية.

قائمة المراجع العربية باللغة العربية:

- أبو جراد، حمدي يونس وحمودة، الآء زياد. (2019). تطوير مقياس ما وراء المعرفي وفق نموذج اندريش دراسة ميدانية لدى عينة من طلبة الثانوية في محافظة غزة. مجلة العلوم التربوية والنفسية، 4(20)، 591-620.
- أيالا، آر وجاي، دي. (2017). النظرية والتطبيق في نظرية الاستجابة للفقرة (عبد الله زيد الكيلاني وإسماعيل سلامة البرهان، مترجم). دار جامعة الملك سعود للنشر. (العمل الأصلي نشر في 2008).
- اسماعيل، هشام ابراهيم. (2009). فاعلية برنامج تدريبي قائم على استراتيجيات حكمة الأداء على الاختبار على مستوى كل من فعالية الذات الاكاديمية والقلق الاختباري. المجلة المصرية للدراسات النفسية، 64(19)، 421-463.
- احمد، نرمين عوني محمد. (2019). البيضة العقلية والتشوهات المعرفية كمنبئين بالحكمة الاختبارية لدى طلاب كلية التربية جامعة الاسكندرية. مجلة كلية التربية، 10(35)، 1-60.
- أبو هاشم، السيد محمد. (2008). النموذج البنائي التنبؤي لمهارات الدراسة والحكمة الاختبارية والتحصيل الدراسي لدى طلاب المرحلة الثانوية. مجلة كلية التربية بالمنصورة، 68(1)، 210-270.
- أبو جراد، حمدي يونس. (2011). مدى التزام المعلمين بتحليل نتائج الاختبارات التحصيلية وعلاقته باتجاهاتهم نحوها. مجلة جامعة الأزهر بغزة، 2(13)، 89-106.
- البلوي، أحمد عبد الله عابد. (2022). بناء مقياس للتميز الإلكتروني وفق نموذج أندريش لدى طلبة جامعة طيبة بالمملكة العربية السعودية. مجلة كلية التربية، 114(19)، 510-536.
- بني عواد، علي محمد العرسان. (2022). مستوى حكمة الاختبار لدى طلاب جامعة الملك فيصل في ضوء متغيرات التخصص والنوع والمستوى الدراسي. مجلة البحوث في مجالات التربية النوعية، 42(8)، 797-829.
- بخيت، صلاح الدين فرح عطا الله. (2013). اعادة تدريج وتقنين قائمة تقديرات المعلم لصفات التلاميذ الموهوبين في مرحلة التعليم الاساسي باستخدام نموذج راش. مجلة الدراسات التربوية والنفسية، 3(7)، 275-288.
- الحسن، أسماء عدنان. (2022). البنية العاملية الاستكشافية والتوكيدية لمقياس الحكمة الاختبارية (دراسة ميدانية على عينة من طلبة جامعة حماة). مجلة طلبة التربية للبنات- جامعة بغداد، 4(33)، 53-69.
- حسب الله، عبد العزيز محمد. (2019). إدارة الوقت وحكمة الاختبار وبعض المتغيرات الديموغرافية كمنبئات بالتسويق الأكاديمي لدى طلاب كلية التربية. مجلة كلية التربية- جامعة اسيوط، 9(35)، 435-532.
- خليل، رانيا محمد توفيق أحمد وغشيم، محمد عبد السلام سالم. (2019). الفروق في مستوى الحكمة الاختبارية لدى طلاب المرحلة الاعدادية في ضوء متغيرات النوع والصف الدراسي. دراسات تربوية واجتماعية- جامعة حلوان، كلية التربية، 4(25)، 161-190.
- الخولي، هشام محمد. (2018). أثر برنامج تدريبي لمهارات الحكمة الاختبارية على الأداء والتخمين والمخاطرة والحذر في مواقف اختبارية معيارية المرجع ومحكية المرجع لدى طلاب الجامعة. مجلة كلية التربية، 115(29)، 282-346.
- دودين، حمزة. (2005). تدريس استراتيجيات تقديم الاختبارات. اللجنة الوطنية القطرية للتربية والثقافة والعلوم، العدد (152) مارس، مجلة التربية.
- الزهراني، محمد رزق الله. (2015). الخصائص السيكومترية لمقياس الحكمة الاختبارية لدى طلاب الجامعة في البيئة السعودية. المجلة العلمية لكلية التربية النوعية، 4(3)، 214-278.
- زكري، علي محمد عبد الله. (2009). الخصائص السيكومترية لاختبار (اوتيس-لينون) للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية. [رسالة دكتوراه غير منشورة]. جامعة أم القرى.

- زهران، محمد حامد. (2000). الإرشاد النفسي المصغّر للتعامل مع المشكلات الدراسية. القاهرة: عالم الكتب.
- السعيد، أحلام جميل. (2022). تدريج مقياس فرط الاستتارة وفق نموذج مقياس التقدير لاندريش على عينة من موهوبات المرحلة الثانوية بمدينة مكة المكرمة. الجمعية العربية للقياس والتقويم، 6(3)، 330-379.
- السعودي، شريف والشوا، هلا. (2017). مقارنة طرق التدريج العمودي المبنية على إجراءات النظرية الكلاسيكية في القياس بالطرق المبنية على نظرية الاستجابة للفقرة. مجلة دراسات العلوم التربوية، 44(4)، 55-71.
- سعادة، مروة صلاح ابراهيم وعيسى، أسماء عبد العزيز السيد. (2022). الاسهام النسبي للتحكم الانتباهي واليقظة العقلية في التنبؤ بمهارات الحكمة الاختبارية لدى الطالبات المعلمات بكلية الاقتصاد المنزلي جامعة المنوفية. مجلة بحوث التربية النوعية، 67(6)، 219-263.
- السلمي، طارق عبد العالي وسنان، هاني محمد عمر. (2022). استخدام نموذج اندريش في تدريج مقياس الذكاءات المتعددة لدى طلاب الجامعات السعودية. المجلة العلمية للعلوم التربوية والصحة النفسية، 5(4)، 51-101.
- السلمي، طارق بن عبد العال بن صمل. (2018). التفكير الايجابي والحكمة الاختبارية لدى طلبة المرحلة المتوسطة الموهوبين والعاديين في مدينة جدة. مجلة العلوم التربوية والنفسية، 19(1)، 331-365.
- سليمان، شاهر خالد. (2014). حكمة الاختبار لدى طلاب جامعة تبوك في ضوء متغيرات التخصص والمعدل التراكمي والمستوى الدراسي. مجلة العلوم التربوية والنفسية، 2(15)، 245-273.
- الشمري، صاحب أسعد ويس. (2019). الحكمة الاختبارية لدى طلبة الدراسات العليا في جامعة سامراء وعلاقتها بالانهاك بالتعلم والتوتر النفسي. المجلة العربية للتربية النوعية، 11(11)، 41-70.
- الشمري، صادق كاظم جريو، والسعدي، مروة عبدا عباس عصفور. (2018). الحكمة الاختبارية وعلاقتها بفاعلية الذات الابداعية لدى طلبة الجامعة. مجلة العلوم الانسانية، 2(25)، 171-199.
- ضحا، ايمان صلاح محمد؛ موسى، فاطمة عبد الرحمن عبد القادر وكاشف، أنعام أحمد عبد الحليم. (2020). الحكمة الاختبارية وتوجهات الهدف كعوامل منبئة بالأداء على الاختبارات المقالية والموضوعية لدى عينة من طالبات كلية التربية جامعة دمنهور. مجلة كلية التربية، 31(31)، 240-289.
- الطراونة، صبري حسن خليل. (2020). تطوير مقياس لحكمة الاختبار ومطابقة فقراته مع نموذج أندريش. مؤنة للبحوث والدراسات- سلسلة العلوم الانسانية والاجتماعية، 6(35)، 141-170.
- طاحون، حسين حسن حسين. (2010). فاعلية برنامج قائم على استراتيجيات الحكمة الاختبارية على كل من التحصيل الدراسي وقلق الاختبار لدى طلاب الدراسات العليا بكلية التربية. جامعة عين شمس، المجلة المصرية للدراسات النفسية، 69(20)، 79-126.
- العبيدي، شيماء صلاح حسين. (2023). حكمة الاختبار لدى الطلبة الموهوبين. مجلة جامعة الانبار للعلوم الانسانية، 1(19)، 622-652.
- عليمت، مهدي. (2019). بناء اختبار تحصيلي في الرياضيات وفق نظرية الاستجابة للفقرة (النموذج ثنائي المعلمة). مجلة جامعة النجاح للأبحاث- العلوم الإنسانية، 7(35)، 1079-1104.
- العنزي، سامي شطيبي عايض. (2014). علاقة الحكمة الاختبارية بتحصيل طلبة المرحلة الثانوية في دولة الكويت في اختبار تحصيلي مبني وفق نموذج راش. عالم التربية، 15(47)، 345-369.
- العلي، ديبالا. (2015). استخدام نموذج راش في تدريج مقياس سوء التكيف الدراسي من اختبار منيسوتا MMPI-2 لدراسة بعض العوامل المؤثرة على دقة القياس. مجلة جامعة تشرين للبحوث والدراسات العلمية، 3(37)، 277-291.
- العنكبي، حيدر جليل عباس. (2009). المفاضلة في تحديد درجات القطع لاختبار محكي المرجع. [اطروحة دكتوراه غير منشورة]. جامعة بغداد كلية التربية.
- علام، صلاح الدين محمود. (2002). القياس والتقويم التربوي والنفسي اساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي للطباعة والنشر.
- فخرو، حصة عبد الرحمن؛ عبد الرحيم، أنور رياض، ومحمد، محمد ابراهيم. (2009). فاعلية استخدام نموذج سلم التقدير في تحليل مفردات مقياس مداخل الدراسة لدى طلاب جامعتي قطر والمانيا. مجلة الدراسة في التربية وعلم النفس، 2(22)، 135-175.

- قنديل، وائل السيد وموسى، حنان ابراهيم. (2022). تصميم مقياس الحكمة الاختبارية لدى طلاب كلية التربية الرياضية جامعة مدينة السادات. مجلة نظريات وتطبيقات التربية البدنية وعلوم الرياضة، 1(37)، 132-170.
- القرشي، خديجة ضيف الله (2017). الحكمة الاختبارية وعلاقتها بأنماط التفكير وقلق الاختبار لدى طالبات كلية التربية بجامعة الطائف. المجلة التربوية، ج، 47، 240-289.
- مسعود، عبد الرحمن محمد عبد الرحمن. (2020). نمذجة العلاقات للعبء المعرفي والحكمة الاختبارية وقلق الاختبار لدى طلاب الجامعة. مجلة العلوم التربوية، 3(28)، 3، 43-110.
- مجيد، سوسن شاكر. (2014). اسس بناء الاختبارات والمقاييس النفسية والتربوية. (ط.3)، عمان: مركز دينو لتعليم التفكير.
- المالكي، نياض بن عايض و خليل، محمد المري محمد إسماعيل. (2010). علاقة قلق الاختبار بالحكمة الاختبارية لدى عينة من طلاب المرحلة الثانوية بمحافظة الليث التعليمية. [رسالة ماجستير غير منشورة]. جامعة أم القرى.
- نصار، علي ابراهيم علي. (2021). الحكمة الاختبارية لدى طلاب الصف الاول الثانوي في الاردن. مجلة كلية التربية-جامعة المنصورة، 2(113)، 668-700.
- نوفل، محمد بكر، وأبو عواد، فريال محمد. (2010). التفكير والبحث العلمي. (ط.1)، عمان: دار المسيرة.
- نجيب، حسين علي والرفاعي، غالب عوض صالح. (2006). تحليل ونمذجة البيانات باستخدام الحاسوب: تطبيق شامل للحزمة SPSS. عمان (الأردن): الاهلية للنشر والتوزيع.
- وادي، عفاف زياد. (2013). مهارات الحكمة الاختبارية وعلاقتها بالتحصيل الدراسي لدى كلية التربية ابن الهيثم للعلوم الصرفة. مجلة البحوث التربوية والنفسية، (37)، 295-322.

References

- Ahmed, Nermin Awni Muhammad. (2019). Mental alertness and cognitive distortions as predictors of experiential wisdom among students of the Faculty of Education, Alexandria University (In Arabic). *College of Education Journal*, 10(35), 1-60.
- Allam, Salah El-Din Mahmoud. (2002). *Educational and psychological measurement and evaluation*, its basics, applications and contemporary trends (In Arabic). Cairo: Dar Al-Fikr Al-Arabi for Printing and Publishing.
- Abu Jarad, Hamdi Younis and Hamouda, Alaa Ziyad (2019). Developing a metacognitive scale according to the Andrich model, a field study among a sample of high school students in the Gaza Governorate (In Arabic). *Journal of Educational and Psychological Sciences*, 4(20), 591-620.
- Abu Jarad, Hamdi Younis (2011). The extent of teachers' commitment to analyzing the results of achievement tests and its relationship to their attitudes towards them (In Arabic). *Al-Azhar University Journal in Gaza*, 2(13), 89-106.
- Abu Hashem, Sayed Muhammad. (2008). Predictive structural model of study skills, test wisdom, and academic achievement among secondary school students (In Arabic). *Mansoura College of Education Journal*, 68(1), 210-270.
- Ayala, R., and Jay, D. (2017). Theory and application in paragraph response theory (Abdullah Zaid Al-Kilani and Ismail Salama Al-Burhan, translator). King Saud University Publishing House (In Arabic). (Original work published in 2008).
- Al-Anzi, Sami Shatit Ayed. (2014). The relationship of test wisdom to the achievement of secondary school students in the State of Kuwait in an achievement test based on the Rasch model (In Arabic). *World Education*, 15(47), 345-369.
- Al-Ali, Diala. (2015). Using the Rasch model in grading the Academic Maladjustment Scale from the Minnesota MMPI-2 test to study some of the factors affecting the accuracy of the measurement (In Arabic). *Tishreen University Journal for Research and Scientific Studies*, 3(37), 277-291.
- Al-Ankabi, Haider Jalil Abbas. (2009). Trade-offs in determining cut-off scores for the reference-narrator test. [Unpublished doctoral dissertation] (In Arabic). University of Baghdad, College of Education.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-73.
- Al-Balawi, Ahmed Abdullah Ayed. (2022). Building a measure of cyberbullying according to the Andrich model among students at Taibah University in the Kingdom of Saudi Arabia (In Arabic). *College of Education Journal*, 114(19), 510-536.

- Bani Awad, Ali Muhammad Al-Arsan. (2022). The level of test wisdom among students at King Faisal University in light of the variables of specialization, gender, and academic level (In Arabic). *Journal of Research in Specific Education*, 42(8), 797-829.
- Bakhit, Salah al-Din Farah Atallah. (2013). Regrading and codifying the list of teacher ratings for the characteristics of gifted students in the basic education stage using the Rasch model (In Arabic). *Journal of Educational and Psychological Studies*, 3(7), 275-288.
- Baker, F., B. (2001). *The basics of item response theory*. ERIC clearinghouse on Assessment and Evaluation, USA.
- Brannick, M. (2003). "Basics of IRT one-linefile". //a:/item response theory.htm.
- Bond, T.G., & Fox, C.M. (2007). Applying the Rasch model: *Fundamental measurement in the human sciences* (2d ed.). Routledge.
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *Life Sciences Education*, 15(4), 14-21.
- Andrich, D. (2010). *Understanding the response structure and processing the polytomous Rasch model*. In M. L. Nering & R. Ostini (Eds.), *Handbook of Polytomous Item Response Theory Models* (pp.123-152). Routledge.
- Caudill, S. B., Franklin, G. & Mixon, F. G. (2023). Guess for Success? Application of a Mixture Model to Test-Wisness on Multiple-Choice Exame. *State*, 6, 734-739.
- Christensen, K, Makransky, G., & Horton, M. (2017). Critical values for Yens' q3: Identification of local dependence in the Rasch model using residual correlation. *Applied Psychological Measurement*, 41(3), 178-194.
- Conrad, K. M., Conrad, K.J., Funk, R. R., & DENNIS, M. L. (2015). Validation of the full and short-form self-help involvement scale against the Rasch measurement model. *Evaluation Review*, 39(4), 395-427.
- Cohen, A., D. (2007). *The coming of age of research on test-taking strategies*. *Language Assessment Quarterly*, 3(4), 307-330.
- Cunningham, G., K. (1986). *Educational And Psychological Measurement*. Macmillan Publishers Company, Net Work.
- De Ayula, R. (1993). *An introduction to polytomous item response theory models*. *Measurement & Evaluation in counseling & development*, 25(4), 172-189.
- Diamond, J. J., & Evans, W. (1972). An investigation of the cognitive correlates of test-wisness. *Journal of Educational Measurement*, 9(2), 145-150.
- Dodin, Hamza. (2005). *Teaching test-taking strategies* (In Arabic). Qatar National Committee for Education, Culture and Science, Issue (152) March, Education Magazine.
- Daha, Iman Salah Muhammad; Musa, Fatima Abdel Rahman Abdel Qader and Kashif, Inaam Ahmed Abdel Halim. (2020). Test wisdom and goal orientations as factors predicting performance on essay and objective tests among a sample of female students from the Faculty of Education, Damanhour University (In Arabic). *College of Education Journal*, (31), 240-289.
- Empreson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ, US: Lawrence Erlbaum Associates, INC, Publisher.
- Fakhro, Hessa Abdul Rahman; Abdel Rahim, Anwar Riad, and Mohamed, Mohamed Ibrahim (2009). The effectiveness of using the rating scale model in analyzing the vocabulary of the study approaches scale among students at Qatar and Minya universities (In Arabic). *Journal of Study in Education and Psychology*, 2(22), 135-175.
- Fisher, J. W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transaction*, 21(1), 1095.
- Field, A. (2000). *Discovering Statistics Using SPSS for Windows*. London- Thousand Oaks-New Delhi: Sage Publications.
- French, G. A., & Dodd, B. G. (1999). Parameter recovery for the rating scale model using PARSCALE. *Journal of outcome measurement*. 3(2), 176-199.
- Fox, C. M., & Jones, J.A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30-45.
- Gustafson, J. E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33(2), 205-233.
- Gibb, B. (1964). Test – Wisness of secondary cue response (Unpublished doctoral dissertation). Stanford University. Ann Arbor, MI.

- Al-Hassan, Asmaa Adnan. (2022). The exploratory and confirmatory factor structure of the experimental wisdom scale (a field study on a sample of Hama University students) (In Arabic). *Education Students Journal for Girls - University of Baghdad*, 4(33), 53-69.
- Hasaballah, Abdul Aziz Muhammad. (2019). Time management, test wisdom, and some demographic variables as predictors of academic procrastination among College of Education students(In Arabic). *Journal of the Faculty of Education - Assiut University*, 9(35), 435-532.
- Hayati, M & Ghojogh, N. (2013). In vestigating the influence of proficiency and gender on use of selected TEST-WISENESS strategies in Higher Education. *English Language Teaching*, 1(2), 169-181.
- Harmon, M., Morse, D. & Morse, L. (1996). Confirmatory Factor Analysis of The Gibb Experimental Test of Test-Wiseness. *Educational and Psychological Measurement*, 56(2), 276-386.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London, SAGE publications.
- Hambleton, R. K., & Swaminathan, H. (1989). *Item Response Theory Principles and Applications*. Kluwer – Nijh of Publishing, Boston, U.S.A.
- Hambleton, R. K., (1987). Determining Optimal Test Lengths with A Fixed Total Testing Time. *Educational and Psychological Measurement*, 2(47), 338-339.
- Hambleton, R., & Swaminathan, H. (1985). Item Response Theory: Multiple-Choice Item Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Hudia, S., Yusnida, D., Ismail, N. M., & Juliana, J. (2019). Test wiseness exploits by TOEFL test-takers: wisely or naively? English Education International Conference (EEIC) in conjunction with the 9th Annual International Conference. (AIC), *Universitas Syiah Kuala*, September 18-19, 2019, Banda Aceh, Indonesia.
- Ismail, Hisham Ibrahim. (2009). The effectiveness of a training program based on test performance wisdom strategies on the level of both academic self-efficacy and test anxiety (In Arabic). *Egyptian Journal of Psychological Studies*, 64(19), 421-463.
-
- Alimat, Mahdi. (2019). Building an achievement test in mathematics according to item response theory (two-parameter model) (In Arabic). *An-Najah University Journal of Research - Human Sciences*, 7(35), 1079-1104.
- Kim, S., & Kyllonen, P. C. (2006). *Rasch rating scale modeling of data from the standardized letter of recommendation*. ETS.
- Khalil, Rania Muhammad Tawfiq Ahmed and Ghasheem, Muhammad Abdel Salam Salem. (2019). *Differences in the level of test wisdom among middle school students in light of gender and academic grade variables* (In Arabic). *Educational and Social Studies - Helwan University, Faculty of Education*, 4(25), 161-190.
- Al-Khouly, Hisham Muhammad. (2018). *The effect of a training program for test wisdom skills on performance, guesswork, risk-taking, and caution in standard-referenced and standard-referenced test situations among university students* (In Arabic). *College of Education Journal*, 115(29), 282-346.
- Lang, J. W., & Tay, L. (2021). The science and practice of item response theory in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 8, 311-338.
- Linacre, J., M. (2002). *What do infit and outfit, mean-square and standardized mean?* *Rasch Measurement Transaction*, 16(2), 878.
- Linacre, J., M. (2005). *Winsteps Rasch measurement computer program*. MESA Press.
- Linacre, J., M. (2010). Predicting responses from Rasch measures. *Journal of Applied Measurement*, 11(1), 1.
- Linacre, J., M. (2012). *A users' guide to WINSTEPS: Rasch model computer program*. MESA Press.
- Lustina, M. (2004). *A comparison of Andrich's rating scale model and Rost's successive intervals model*.
- Molenaar, I. & Hoijtink, H. (1996). *Person-fit and the Rasch model*. *Education Resources Information*, 1-21.
- McKey, P. F., & Doverspike, D. (2001). African- Americans test-taking Attitudes and their effect on cognitive ability test performance. *Public Personal Management*, 30(1), 67-75.
- Magno, C. (2009). Demonstrating the difference between Classi-Cal test theory and item response theory using derived test data. *The International Journal of Education and Psychological Assessment*, 1(1), 1-11.
- Millman, J., Bishop, C., & Eble, R. (1965). *Analysis Of Test wiseness*. *Educational and Psychological Measurements*, New York, Vol. (25), 707-726.

- Nguyen, H. (2003). Constructing a new theoretical framework for test wiseness and developing the knowledge of test-taking strategies (KOTTS) measure. *Unpublished master thesis, Michigan State University, UMI No.1416088.*
- Nitko, J. (2001). *Educational assessment of students NEW YORK: Merrill Prentice Hall.*
- Otoum, A.; Khalaf, H. B.; Bajbeer, A., & Hamad, H. B. (2015). The Level of Test-Wiseness for students of Arts and Science Faculty at Sharourah and its Relationship with Some Variables. *Journal of Education and Practice, 6(29), 102-113.*
- Masoud, Abdul Rahman Muhammad Abdul Rahman. (2020). Relationship modeling of cognitive load, test wisdom, and test anxiety in university students (In Arabic). *Journal of Educational Sciences, 3(28), Part 3, 43-110.*
- Majeed, Sawsan Shaker. (2014). Foundations for constructing psychological and educational tests and standards (In Arabic). (3rd ed.), Amman: Debno Center for Teaching Thinking.
- Al-Maliki, Dhiab bin Ayed and Khalil, Muhammad Al-Marri Muhammad Ismail. (2010). The relationship of test anxiety to test wisdom among a sample of secondary school students in Al-Laith Educational Governorate (In Arabic). A magister message that is not published. Umm Al Qura University.
- Nassar, Ali Ibrahim Ali. (2021). Test wisdom among first year secondary school students in Jordan (In Arabic). *Journal of the Faculty of Education - Mansoura University, 2(113), 668-700.*
- Nofal, Muhammad Bakr, and Abu Awad, Faryal Muhammad. (2010). *Scientific thinking and research* (In Arabic). (1st edition), Amman: Dar Al Masirah.
- Najeeb, Hussein Ali and Al-Rifai, Ghaleb Awad Saleh. (2006). *Data analysis and modeling using the computer: a comprehensive application of the SPSS package* (In Arabic). Amman (Jordan): Al-Ahlia Publishing and Distribution.
- ALnasraween. M. S., ALsmadi. M. S., ALzboon. H.S & ALkurshe. T. O. (2022). The Level of Universities Students Test Wiseness in Jordan during Distance Learning in Light of Some Variables. *Education Research International, 1-10.*
- Orth, J. (1995). The Effect of Test-Taking Strategies for Multiple- Choice Test. *Unpublished doctoral dissertation, University of Pittsburgh, UMI No. 9614214.*
- Al-Obaidi, Shaima Salah Hussein. (2023). The wisdom of testing among gifted students (In Arabic). *Anbar University Journal of Human Sciences, 1(19), 622-652.*
- Qandil, Wael Al-Sayed and Musa, Hanan Ibrahim. (2022). Designing a wisdom test scale for students of the Faculty of Physical Education, Sadat City University (In Arabic). *Journal of Theories and Applications of Physical Education and Sports Sciences, 1(37), 132-170.*
- Al-Qurashi, Khadija Dhaif Allah. (2017). Test wisdom and its relationship to thinking patterns and test anxiety among female students of the College of Education at Taif University (In Arabic). *Educational Journal, vol. 47, 240-289.*
- Roberson, D. B. (2020). Test wiseness and background knowledge: Their relative contribution to high test performance. *Mississippi State University, Theses and Dissertation. 4253.*
- Rogers, W. T., & Bateson, D. J. (1991). Verification of a model of test-taking behavior of high school seniors. *Journal of Experimental Education, 59(4), 331-350.*
- Reckase, M., D. (1997). *The past and future of multidimensional item response theory.* Applied Psychological Measurement, 21(1), 25-36.
- Sarnacki, R. (1979). *An examination of test-wiseness in the cognitive test domain.* Review of Educational Research, 49(2), 252-279.
- Sedoc, J., & Ungar, L. (2020). *Item response for efficient human evaluation of chatbots.* Proceedings of the First Workshop on Evaluation and Comparison of NLP systems.
- Sick, J. R., (2009). Rasch measurement in language education Part3: The family of Rasch models. *SHIKEN, 13(1), 4-10.*
- Susan, E, & Steven, P. (2000). *Item Response Theory for Psychologist.* London, Lawrence Erlbaum Associates, Publishers Mahawah, New Jersey.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social disciplines].* Trim Komunikata Publishing House.

- Alsoudi, S. & Abu Shindi, Y. (2023). Rasch Rating Scale Modelling of the Arabic Version of the Critical Thinking Disposition Scale. *Journal of Educational and Psychological Studies-Sultan Qaboos University*, 17(4) 359-369.
- Al-Saidi, Ahlam Jameel. (2022). Grading the hyperexcitability scale according to the Landrich rating scale model on a sample of gifted secondary school students in the city of Mecca(In Arabic). *Arab Society for Measurement and Evaluation*, 6(3), 330-379.
- Al-Saudi, Sherif, and Al-Shawwa, Hala. (2017). Comparing vertical grading methods based on classical theoretical measurement procedures with methods based on item response theory(In Arabic). *Journal of Educational Science Studies*, 44(4), 55-71.
- Saada, Marwa Salah Ibrahim and Issa, Asma Abdel Aziz Al-Sayed. (2022). The relative contribution of attentional control and mental alertness in predicting test wisdom skills among female student teachers at the Faculty of Home Economics, Menoufia University (In Arabic). *Journal of Specific Education Research*, (67), 219-263.
- Al-Salami, Tariq Abdel-Aali and Sinan, Hani Muhammad Omar. (2022). Using the Andrich model in grading the multiple intelligence scale among Saudi university students (In Arabic). *Scientific Journal of Educational Sciences and Mental Health*, 5(4), 51-101.
- Al-Sulami, Tariq bin Abdel-Al bin Samal. (2018). Positive thinking and experiential wisdom among gifted and ordinary middle school students in Jeddah (In Arabic). *Journal of Educational and Psychological Sciences*, 1(19), 331-365.
- Suleiman, Shaher Khaled. (2014). The wisdom of the test among students at the University of Tabuk in light of the variables of specialization, cumulative average, and academic level (In Arabic). *Journal of Educational and Psychological Sciences*, 2(15), 245-273.
- Al-Shammari, owner of Asaad Wais. (2019). Experiential wisdom among graduate students at the University of Samarra and its relationship to engagement in learning and psychological stress (In Arabic). *Arab Journal of Specific Education*, (11), 41-70.
- Al-Shammari, Sadiq Kazem Jarrio, and Al-Saadi, Marwa Abda Abbas Asfour. (2018). Experiential wisdom and its relationship to creative self-efficacy among university students (In Arabic). *Journal of Human Sciences*, 2(25), 171-199.
- Al-Tarawneh, Sabri Hassan Khalil. (2020). Developing a measure of test wisdom and matching its items with Andrich's model (In Arabic). *Muta Research and Studies - Humanities and Social Sciences Series*.
- Tahoun, Hussein Hassan Hussein (2010). The effectiveness of a program based on test wisdom strategies on both academic achievement and test anxiety among graduate students at the Faculty of Education, Ain Shams University (In Arabic). *Egyptian Journal of Psychological Studies*, 69(20), 79-126.
- Thissen, D., & Edwards, M. (2005). Diagnostic Scores Augmented Using Multidimensional Item Response Theory: Preliminary Investigation of MCMC Strategies. *A presentation prepared for the symposium Enhancing the Diagnostic Value of Large Scale Achievement Test: Technical Developments and Applications* be presented at the Annual Meeting of the National Council on Measurement in Education in Montreal, PQ, Canada, April, 12-14.
- Uttaro, T., & Lehman, A. (1999). Graded response modeling of quality-of-life interview. *Evaluation and program planning*, 22(1), 41-52.
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Educational Measurement Solution.
- Watter, A., & Siebert, T. (1990). *Strategies to responding to the new TOEFL task*. *Language Assessment Quarterly*, 10(3), 20-41.
- Wadi, Afaf Ziad. (2013). Test wisdom skills and their relationship to academic achievement at the Ibn al-Haytham College of Education for Pure Sciences (In Arabic). *Journal of Educational and Psychological Research*, (37), 295-322.
- VonDavier, M. (2016). *Rasch Model*. In *Wimj. Vander Linden(ed.)*. *Handbook of Item Theory* (Boca Raton: CRC Press), Routledge Hand books.
- Alvarenga, M., Santos, T., & Andrade, D. (2020). Item Response Theory- based validation of a short form of the Disordered Eating Attitude Scale (DEAS-s) to Brazilian sample. *Cadernosde Saude Pública*.
- Yousf. O. M., E. (2022). Using Andrich Rating Scale Modeled Cairo Among Proficiency Academic for the Scale a build to (ARSM) Study Psychometric: Students University. *Egyptian Journal of Psychological Studies*, 116(32), 1-60.

- Al-Zahrani, Muhammad Rizkallah. (2015). Psychometric properties of the wisdom test scale among university students in the Saudi environment (In Arabic). *Scientific Journal of the College of Specific Education*, 4(3), 214-278.
- Zakri, Ali Muhammad Abdullah. (2009). *The sociometric properties of the Otis-Lennon test of mental ability assessed according to the classical measurement and the Rasch model among middle school students in Sabya Educational Governorate*. [Unpublished doctoral dissertation] (In Arabic). Umm Al Qura University.
- Zahran, Muhammad Hamed. (2000). *Mini-psychological counseling to deal with academic problems* (In Arabic). Cairo: World of Books.