تطوير مقياس لحكمة الاختبار ومطابقة فقراته مع نموذج اندريش

د.صبري حسن الطراونة *

ملخص

هدفت الدراسة إلى تطوير مقياس لحكمة الاختبار ومطابقة فقراته مع نموذج اندريش، تكونت عينة الدراسة من (2387) طالباً وطالبةً من طلبة الجامعات الأردنية، تم إتباع مجموعة من الإجراءات لإخراج المقياس بصورته النهائية، تم التحقق من مطابقة استجابات عينة الدراسة لافتراضات نموذج اندريش، ومطابقة فقرات المقياس لافتراضات نموذج اندريش، وتم التحقق أيضاً من افتراضات نظرية الاستجابة للفقرة لفقرات المقياس، وتكون المقياس في صورته النهائية من (59) فقرة، وأظهرت النتائج دلالات صدق وثبات مقبولة للمقياس.

الكلمات الدالة: طلبة الجامعات، نموذج اندريش، حكمة الاختبار.

* **كلية العلوم التربوية، جامعة مؤتة، الأردن.**

تاريخ تقديم البحث: 2018/5/13م م. تاريخ قبول البحث: 2018/11/12 م.

© جميع حقوق النشر محفوظة لجامعة مؤتة، الكرك، المملكة الأردنية الهاشمية، 2020م.

Developing a scale of Test Wiseness and its Fitness to Andrich Model

Sabri Hasan Al-Tarawneh

Abstract

The study aimed at developing a scale of test wiseness and its fitness to Andrich Model. The study sample consisted of (2387) students from Jordanian universities. Several procedures were followed to produce the final scale. The responses of the study sample to the fitness to Andrich model were verified. In addition to the scale's items fitness to Andrich model, the assumptions of item response theory to the scales items were verified. The final scale consisted of (59) item. The results showed acceptable indications of validity and stability.

Keywords: University Students, Andrich Model, Test Wiseness

مقدمة:

مع التطور الهائل الذي شهده مجال القياس النفسي والتقويم التربوي، إلا أنه لم يصل إلى الدرجة التي يمكن أن تمكن القائمين على العملية التعليمية من الحكم على مدى تمكن الطلبة وإتقانهم للمحتوى العلمي الذي درسوه، والذي تقيسه اختبارات التحصيل بشكل دقيق، لأن أداء المفحوصين على هذه الاختبارات لا يتوقف على مدى معرفتهم وإتقانهم للمحتوى الذي تعلموه فقط، بل تتدخل عوامل أخرى تؤثر على أدائهم عليها، ومن هذه العوامل استراتيجيات حكمة الاختبار.

وترجع أصول مفهوم حكمة الاختبار إلى عام (1946) عندما أشار له كرونباخ، وقال إن بعض المفحوصين ينجحون في تطوير أساليب منظمة للإجابة على الاختبارات، ويحصلون على درجات أعلى من الدرجات التي تمثل قدرتهم الحقيقية، وفي عام (1951) بين ثورندايك أن حكمة الاختبار والقدرة العامة التي يمتلكها المفحوص لفهم تعليمات الاختبار يوثران على درجته على هذه الاختبار (Nguyen, 2003).

وقد تعددت آراء الباحثين في ماهية حكمة الاختبار فقد عرفها روجرز وهارلي (Rogers & Harley, 1999) بأنها مهارة تتيح للمفحوص أن يستفيد من خصائص فقرات الاختبار في الحصول على درجة أعلى من درجته الحقيقية دون معرفته لمحتوى الاختبار، وعرفها عودة (Oudeh , 1989) بأنها نشاط عقلي يقوم به المفحوص لرفع درجته على الاختبار بشكل أكثر من درجته الحقيقية (التي تعكس معرفته الحقيقية بمحتوى الاختبار)، وعرفها اورث (Orth, 1995) بأنها قدرة معرفية مكتسبة لمجموعة من المهارات التي يمكن للمفحوص أن يستخدمها في موقف الاختبار لتحسين درجته على الاختبار بغض النظر عن محتواه, وتعرفها نغوين (Nguyen) بأنها بناء نفسي متعدد الأبعاد يتكون من العمليات العقلية السابقة والمتزامنة مع النشاطات العقلية للتعامل مع الاختبار (Nguyen, 2003) ، وهناك من قال بأنها قدرة معرفية مكتسبة لمجموعة من المهارات للاستفادة من خصائص موقف الاختبار التي يستخدمها المفحوص خلال إجابته على الفقرات لتحسين درجته على الاختبار (Hammad, 2010) .

وقد تعددت تصنيفات استراتيجيات حكمة الاختبار، فقد صنفت إلى: استراتيجيات استغلال الوقت، واستراتيجيات تجنب الأخطاء واستراتيجيات التخمين (Nitko, 2001)، وهناك تصنيف آخر لها، وفيه صنفت إلى استراتيجيات تستخدم قبل البدء بالإجابة، واستراتيجيات تستخدم خلال الإجابة

على الاختبار، واستراتيجيات تستخدم بعد الانتهاء من الإجابة على الاختبار (Wenden, 1991)، وصنفت كذلك إلى: استراتيجيات استخدام الوقت, واستراتيجية تجنب الخطأ، واستراتيجية التخمين، واستراتيجية الاستبدال الاستنباطي، واستراتيجية مراعاة النية أو القصد واستراتيجية استخدام دليل (Rogers and Yang, 1996) وصنفها بينج (Peng, 2005) إلى استراتيجية حذف البدائل الخاطئة، واستراتيجية استخدام معلومات من فقرات أخرى، واستراتيجية التخمين واستراتيجية تجنب الخطأ، أما سارناكي (Sarnacki, 1979) فقد صنفها إلى: استراتيجية استخدام الاختبار، واستراتيجية تجنب الخطأ، واستراتيجية التخمين، واستراتيجية الاستنتاج المنطقي أو الاستنباطي، واستراتجية استخدام دليل المنبهات والمؤشرات.

وفي دراسة قام بها (Abdel Wahab, 2007) أطلق على استراتيجيات حكمة الاختبار مهارات حكمة الاختبار، وصنفها إلى: مهارة التخمين الذكي أثناء الإجابة عن أسئلة الاختبار, ومهارة تنظيم وقت الاختبار ومهارة تجنب الأخطاء وتداركها أثناء الإجابة على الاختبار.

ويوجد مجموعة من العوامـــل التي تؤثر على حكمة الاختبار منها: الخبرة السابقة في تعامل المفحوص مع موقف الاختبار، والفترة الزمنية لآخر تعرض لموقف الاختبار، ونوع الاختبار المطبق، وغموض الصياغة النحوية واستخدام مصطلحـــات غير مألوفـــة في صياغة فقرات الاختبار، ومستوى ذكـــاء المفحوص وجودة صياغة فقرات الاختبار (Rogers and Bateson, 1991; Sarnacki, 1979)

وتعد حكمة الاختبار من أحد مصادر التباين والاختلاف في درجات الطلبة على الاختبارات التحصيلية، ولهذا فهي تحول دون تحقق الافتراض الذي يفترضه معدو الاختبارات بأن الفروق بين درجات المفحوصين على الاختبار تعزى إلى قدراتهم في محتوى المادة التي تعلموها وقدموا فيها الاختبار (Haladyna, Downing & Rodriguez, 2002).

ولا تعد حكمة الاختبار بديل عن تحصيل محتوى مادة الاختبار، وإدخال المحتوى إلى بناء المفحوص المعرفي من خلال دراسته له، فهي عامل مساعد للمفحوص ليحصل على درجة أعلى من درجته الحقيقية على الاختبار، وقد يفقد هذه الزيادة على درجته إذا لم يكن لديه هذه المهارة أو الاستراتجية، فهناك مفحوصون يبذلون جهداً كبيراً في الدراسة، لكنهم لم يحصلوا على درجات تناظر هذا الجهد، وقد يكون سبب ذلك، عدم تعلمهم لكيفية التعامل مع متغيرات موقف الاختبار، ويمكن

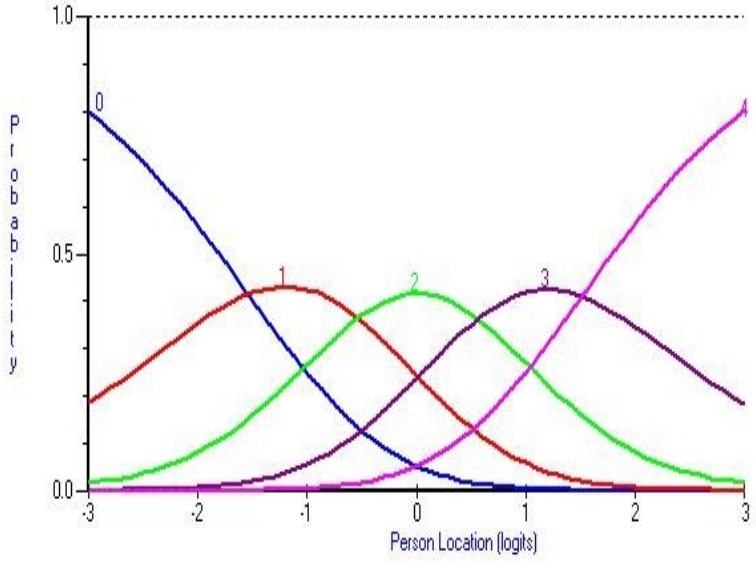
اكتساب مهارة أو استراتيجية حكمة الاختبار من خلال خبرة المفحوص في تقديمه للعديد من الاختبارات، أو من خلال تدريبه على هذه الاستراتيجيات (Nguyen, 2003)، والمفحوص الذين يمارس حكمة الاختبار يتميز عن غيره من الذين لا يمارسونها بأنه يحصل على درجات أعلى مما هو متوقع، ويشعر بالارتياح خلال تقديم الاختبار، ويكون لديه القدرة على إنهاء الإجابة على الاختبار في الوقت المحدد، ويستغل وقت الاختبار بشكل مناسب، ويكون لديه مستوى القلق منخفض ونادراً ما يتركون فقرات من دون الإجابة عنها (Roznowski & Bassett, 1992).

ولم يوجد أي مقياس لحكمة الاختبار طور باستخدام نظرية الاستجابة للفقرة، لأن أهم ما يميز نظرية الاستجابة للفقرة عن النظرية الكلاسيكية في القياس استقلالية القياس، أي أن تقدير معالم الفقرات يكون مستقلاً عن خصائص الأفراد الذين استخدموا في تقدير هذه المعالم، وكذلك أن تقدير قدرة الأفراد يكون مستقلاً عن عينة الفقرات التي تطبق عليهم (Susan & Steven, 2000).

وتستند نظرية الاستجابة للفقرة إلى عدة افتراضات، وهي: افتراض أحادية البعد، أي أن هناك قدرة أو سمة واحدة تفسر أداء المفحوص على المقياس، وافتراض الاستقلال الموضعي، أي أن استجابات المفحوصين من نفس القدرة على الفقرات المختلفة مستقلة عن بعضها البعض، فاستجابة المفحوص على فقرة ما لا تؤثر على أدائه على الفقرات الأخرى، وافتراض منحنى خصائص الفقرة، وهو عبارة عن اقتران يربط بين احتمال نجاح المفحوص في الإجابة على الفقرة والقدرة التي يقيسها المقياس، ويفترض أن يكون شكل المنحنى على شكل حرف (S)، وافتراض السرعة، أي أن إجابة المفحوص على الفقرة تعتمد على مقدار القدرة التي يقيسها المـقياس ولا تعتمد على السرعة في الإجابة (Hambleton & Swaminathan,1985).

وانبثق عن نظرية الاستجابة للفقرة مجموعة من النماذج الرياضية، يعتمد كل نموذج منها على معادلة رياضية تحدد علاقة أداء الفرد على فقرات المقياس بقدرته التي تكمن وراء هذا الأداء، وقدرة الفرد تفسر حسب موقعه على سلم القدرة، وتقسم نماذج نظرية الاستجابة للفقرة إلى قسمين: النماذج ثنائية التدريج، ومنها: النموذج أحادي المعلمة والنموذج ثنائي المعلمة والنموذج ثلاثي المعلمة والنموذج رباعي المعلمة، والنماذج متعددة التدريج، ومنها نموذج الاستجابة المتدرجة، وهو امتداد للنموذج ثنائي المعلمة، ونموذج الاستجابة الاسمية، ونموذج التقدير الجزئي ونموذج سلم التقدير (نموذج اندريش) و هما امتداد للنموذج أحادي المعلمة (Embreston & Reiase, 2000).

ويعتبر نموذج سلم التقدير لاندريش (Andrich) تطوير لنموذج راش، وحالة خاصة من نموذج التقدير الجزئي، طور ليناسب البيانات التي من نوع تدريج ليكرت، ويقوم على فكرة أن كل فقرة من فقرات المقياس تحمل شحنة انفعالية إجمالية، ويستخدم مع المقاييس متعددة التدريج، وفيه تجزأ درجه صعوبة الفاصل بين أي قيمتين متتاليتين للفقرة إلى جزأين، الأول يدل على صعوبة الفقرة، والثاني يدل على بعد الفاصل عن مستوى الصعوبة، ويفترض هذا النموذج تساوي عدد القيم التي تأخذها فقرات المقياس (Susan & Steven, 2000)، والشكل (1) يبن شكل إحدى الفقرات ذات التدريج الخماسي وتأخذ الإجابة عليها القيم (4,3,2,1,0) (Andrich, 1978).



شكل (1) رسم للإجابة على إحدى الفقرات حسب نموذج اندريش

يتبين من الشكل(1) وجود أربع عتبات للإجابة على الفقرة، الأولى تساوي حصول المفحوص على العلامة (صفر) أو العلامة (1)، والثانية تساوي حصول المفحوص على العلامة (1) أو العلامة (2)، والثالثة تساوي حصول المفحوص على العلامة (2) أو العلامة (3)، والرابعة تساوي حصول المفحوص على العلامة (3) أو العلامة (4)، وهذه العتبات تساوي

(−1.5, −0.5,0.5, 1.5) على الترتيب، ويكون عدد أقسام الاستجابة لهذه العتبات في نموذج سلم التقدير لاندريش (Andrich) داخل الفقرة متساوية لكل الفقرات، ويفترض هذا النموذج أن تكون قيم عتبات الاستجابة متسقة عبر الفقرات على مقياس التقدير المعطى، وتكون المسافة بين صعوبة الخطوة من قسم من أقسام الاستجابة إلى القسم الذي يليه ثابتة عبر جميع الفقرات.

وحول الكشف عن استخدام استراتيجيات حكمة الاختبار قام بينج (Peng, 2005) بدراسة هدفت إلى الكشف عن الاستراتيجيات التي يستخدمها الطلبة عند الإجابة على فقرات الاختبار، تكونت العينة من (429) طالباً وطالبةً من طلبة الصف الأول الثانوي من المدارس الصينية، أظهرت النتائج عدم وجود فروق تعزى لمستويات التحصيل في استراتيجيات التخمين، وتجنب الخطأ واستخدام معلومات من بنود أخرى، ووجود فروق في استراتيجية حذف البدائل الخاطئة فقط، وكانت الفروق لصالح الطلبة ذوي التحصيل المرتفع والمتوسط، وأظهرت النتائج كذلك أن ترتيب استراتيجيات حكمة الاختبار حسب درجة الاستخدام كان كالتالي: حذف البدائل الخاطئة، استخدام معلومات من بنود أخرى، التخمين، تجنب الخطأ.

وقام كل من محمد وجريجوري وأوستن (Mohamed, Gregory & Austin, 2006) بدراسة هدفت إلى المقارنة بين قدرات الاستعداد للاختبار وحكمة الاختبار، تكونت العينة من (69) طالباً وطالبةً من الطلبة الخريجين من طلبة الجامعات الكندية والطلبة خريجي الاتحاد العمالي العالمي للصيادلة، لتحقيق هدف الدراسة تم تطوير استبانة مكونة من (20) فقرة تقيس حكمة الاختبار، أظهرت النتائج أن استراتيجية استغلال الوقت كانت الأكثر استخداماً عند خريجي كلية الصيدلة، واستراتيجية تجنب الخطأ هي الأكثر استخداما عند خريجي الاتحاد العمالي العالمي للصيادلة.

وقام عبدالوهاب (Abdel Wahab, 2007) بدراسة أثر برنامج تدريبي لمهارة حكمة الاختبار على مستوى الأداء التحصيلي وقلق الاختبار، تكونت العينة من (146) طالباً وطالبةً من طلبة جامعة المنيا في مصر، تم بناء مقياس لمهارة حكمة الاختبار تكون من (36) فقرة، أظهرت النتائج وجود علاقة ارتباطية موجبة بين حكمة الاختبار والتحصيل الدراسي, ووجود علاقة ارتباطية سلبية بين حكمة الاختبار وقلق الاختبار، وأظهرت النتائج كذلك فعالية البرنامج التدريبي لمهارة حكمة الاختبار في إكساب الطلبة مهارات حكمة الاختبار.

وقام فاميولارو (Famularo, 2007) بدراسة هدفت إلى معرفة مدى استخدام الطلبة لاستراتيجيات حكمة الاختبار في الإجابة على فقرات الاختبار، تكونت العينة من (1164) طالباً وطالبةً من طلبة الصفوف: الأول والثاني والثالث المتوسط، أظهرت النتائج أن الاستراتيجيات المستخدمة هي: حذف البدائل الخاطئة، والتخمين وتجنب الخطأ.

وقامت حماد (Hammad, 2010) بدراسة هدفت إلى التعرف على العلاقة بين حكمة الاختبار وبين الأداء على اختبار تحصيلي مبني وفق نموذج راش، تكونت العينة من (211) طالبة من طالبات كلية التربية للبنات للأقسام الأدبية بجامعة أم القرى، تم استخدام مقياس لحكمة الاختبار مكون من (63) فقرة، أظهرت النتائج عدم وجود علاقة بين درجة حكمة الاختبار وبين التحصيل، وأظهرت النتائج كذلك عدم وجود فروق في درجة حكمة الاختبار تعزى لمتغيري التخصص ومستوى التحصيل.

كما قام حمادنه (Hamadnh, 2011) بدراسة هدفت إلى قياس درجة استخدام الطالب الجامعي لاستراتيجيات حكمة الاختبار، تم بناء استبانة تكونت من (15) فقرة تقيس استراتيجيات حكمة الاختبار، تكونت العينة من (244) طالباً وطالبةً من طلبة كلية العلوم التربوية في جامعة آل البيت، أظهرت النتائج أن جميع استراتيجيات حكمة الاختبار استخدمت بدرجة عالية، وقد جاءت الاستراتيجيات المستخدمة بعد الانتهاء من الإجابة على الاختبار في المرتبة الأولى، تلتها الاستراتيجيات المستخدمة خلال الإجابة، وفي المرتبة الأخيرة جاءت الاستراتيجيات المستخدمة قبل البدء بالإجابة، وأظهرت النتائج عدم وجود فروق في درجة استخدام استراتيجيات حكمة الاختبار تعزى إلى مستوى الطالب الجامعي.

وكما قام شاهر (Shaher, 2014) بدراسة هدفت إلى الكشف عن مدى امتلاك طلبة الجامعة لاستراتيجيات حكمة الاختبار، تكونت العينة من (672) طالباً من طلبة جامعة تبوك، تم تطوير أداة لقياس حكمة الاختبار مقسمة إلى خمس استراتيجيات، وأظهرت النتائج أن أكثر استراتيجية يمتلكها طلبة جامعة تبوك هي استراتيجية الاستنتاج المنطقي.

وحول بناء مقاييس باستخدام نموذج اندريش قام أبو جراد (Abu Grad, 2015) ببناء مقياس اتجاهات أعضاء هيئة التدريس نحو تقويم الطلبة للممارسات التربوية، تكونت العينة من (200) طالباً وطالبةً، تكون المقياس بصورته النهائية من (28) فقرة من نوع تدريج ليكرت الخماسي، وتحققت للمقياس دلالات صدق وثبات مقبولة.

كما قام الطراونة (Al-Trawneh, 2016) بدراسة هدفت إلى تطوير مقياس لمقاومة الإغراء لطلبة الجامعات الأردنية وفق نموذج اندريش في نظرية الاستجابة للفقرة، تكونت عينة الدراسة من (2454) طالباً وطالبةً من طلبة الجامعات الأردنية، تكون المقياس من (25) فقرة من نوع تدريج ليكرت الخماسي، وتحققت للمقياس دلالات صدق وثبات مقبولة.

ومن استعراض الدراسات السابقة يتبين أن بعضها بحث في درجة استخدام المفحوصين لاستراتيجيات حكمة الاختبار، وبعضها درس أثر برنامج تدريبي على مهارة حكمة الاختبار، وبعضها درس العلاقة بين حكمة الاختبار والتحصيل وبعض الدراسات هدفت إلى بناء مقاييس باستخدام نموذج اندريش، ولم توجد أي دراسة في حدود ومعرفة الباحث بحثت في تطوير مقياس لحكمة الاختبار باستخدام نموذج اندرش، لذلك سعت هذه الدراسة إلى تطوير مقياس لحكمة الاختبار باستخدام نموذج اندريش.

مشكلة الدراسة وأسئلتها:

تعتبر حكمة الاختبار من مهددات صدق الاختبارات وهي من مصادر التباين في الدرجات عليها، وبنفس الوقت هي قدرة معرفية يمتلكها المفحوص ويوظفها في مواقف الاختبار، مما قد يشكك في القرارات التربوية التي تبنى على نتائج الاختبارات، وهذا الواقع يجعل من الضروري توفير مقياس يتسم بالموضوعية يقيس درجة استخدام طلبة الجامعات لحكمة الاختبار، وتتمثل مشكلة الدراسة في الإجابة على الأسئلة التالية:

السؤال الأول: هل تتطابق الاستجابات على فقرات مقياس حكمة الاختبار لطلبة الجامعة مع افتراضات نموذج اندريش؟

السؤال الثاني: هل يحقق تدريج فقرات مقياس حكمة الاختبار لطلبة الجامعة بنموذج اندريش افتراضات نظرية الاستجابة للفقرة؟

السؤال الثالث: هل لمقياس حكمة الاختبار لطلبة الجامعة درجة مقبولة من الصدق والثبات؟

أهمية الدراسة:

تنبع الأهمية النظرية للدراسة من أهمية موضوعها فهي تهدف إلى تطوير مقياس لحكمة الاختبار لطلبة الجامعات باستخدام نموذج اندريش في نظرية الاستجابة للفقرة لتكون رائداً للمهتمين في دراسة هذا الموضوع، وتشكل إضافة جديدة إلى المكتبة العربية، مما يغطي النقص الذي تعاني منه في هذا الجانب.

وتنبع الأهمية التطبيقية للدراسة من أنها توفر مقياس لحكمة الاختبار يستخدم من قبل المدرسين الجامعيين والإدارات الجامعية لتمد القائمين على تصميم وإعداد الاختبارات التحصيلية بالمزيد من الوعي تجاه بعض أساليب وفنيات القياس، وهذا سيفيد في كثير من القرارات التربوية التي تتعلق بمستقبل الطلبة من خلال الوقوف على معرفة مدى استخدام الطلبة لاستراتيجية حكمة الاختبار، ومن جانب آخر يؤمل أن تفيد في خلق بيئة تعليمية تعمل على تحسين هذه المهارات عند الطلبة لتحسين مستواهم التحصيلي، وتوفير اتجاهات إيجابية نحو الاختبارات بشكل خاص والتعليم بشكل عام، وهذا سيكون له أثر إيجابي على العملية التعليمية.

مصطلحات الدراسة:

حكمة الاختبار: هي نشاط عقلي يقوم به المفحوص لرفع درجته على الاختبار بأكثر من درجته الحقيقية (Oudeh , 1989)، وتعرف إجرائياً في هذه الدراسة بالدرجة التي يحصل عليها المفحوص على المقياس المطور.

نموذج اندريش: هو أحد نماذج نظرية الاستجابة للفقرة المتعددة التدريج، طوره اندريش عن نموذج راش, وهو يستخدم مع الفقرات من نوع تدريج ليكرت ويسمى كذلك بنموذج سلم التقدير.

حدود الدراسة:

- الحدود البشرية: طلبة الجامعات الأردنية.
- الحدود الزمنية: تم إجراء الدراسة في الفصل الدراسي الثاني من العام الجامعي 2018/2017.

مجتمع الدراسة:

تكون مجتمع الدراسة من جميع طلبة الجامعات الأردنية المسجلين على الفصل الدراسي الثاني من العام الجامعي 2018/2017، والبالغ عددهم (248123) طالباً وطالبةً، وفق إحصائية هيئة اعتماد مؤسسات التعليم العالي في المملكة الأردنية الهاشمية.

عينة الدراسة:

تكونت عينة الدراسة الأولية من(2481) طالباً وطالبةً بنسبة (1%) من مجتمع الدراسة، تم اختيارهم بالطريقة العشوائية العنقودية، بعد استبعاد الأفراد الذين لم يطابقوا النموذج، وعددهم (53) طالباً وطالبةً، واستبعاد الأفراد الذين حصلوا على العلامة (صفر) على المقياس، وعددهم (12) طالباً وطالبةً أو العلامة الكاملة على المقياس، وعددهم (29) طالباً وطالبةً، وكان عدد الأفراد الذين دخلوا في التحليل النهائي (2387) طالباً وطالبةً.

إجراءات الدراسة:

تم إتباع الخطوات الآتية في تطوير المقياس:

1) طرح سؤال مفتوح نصه "ما هي استراتيجيات أو مهارات حكمة الاختبار التي قد يمتلكها طلبة الجامعة ويوظفوها عند إجابتهم على الاختبارات" على مجموعة من المختصين في القياس والتقويم من حملة درجة الدكتوراة والذين يدرسون في الجامعات الأردنية والعربية.

2) الرجوع إلى الأدب النظري والدراسات السابقة التي بحثت في موضوع حكمة الاختبار من خلال شبكة الانترنت، والكتب، والدوريات، ورسائل الماجستير، وأطروحات الدكتوراة، مثل (Abdel Wahab, 2007); Famularo, 2007 ; Hamadnh, 2011 ; Hammad, 2010; (Mohamed, Gregory & Austin, 2006 ; Peng, 2005; Nitko, 2001; Sarnacki, 1979;Shaher, 2014).

3) صياغة (75) فقرة لقياس استراتيجيات أو مهارات حكمة الاختبار، ودرجت بأسلوب ليكرت بثلاثة خيارات، وهي: دائماً، وأحياناً ونادراً، تقيس خمس استراتيجيات، وهي: استراتيجية استخدام الاختبار والاستعداد له، واستراتيجية الزمن، واستراتيجية التخمين، واستراتيجية تجنب الخطأ، واستراتيجية الاستفادة من خصائص الاختبار.

4) عرض المقياس على مجموعة من أساتذة القياس والتقويم في الجامعات الأردنية والعربية، بلغ عددهم (20) دكتور، وطلب منهم التعليق على فقرات المقياس من خلال وضوحها وسلامتها اللغوية، ومدى ملاءتها لقياس حكمة الاختبار، وبناءاً على الملاحظات التي قدمها المحكمين تم تعديل صياغة بعض الفقرات، وتم حذف بعض الفقرات وتم إضافة بعض الفقرات، وتكون المقياس من (72) فقرة.

5) تطبيق المقياس على عينة استطلاعية مكونة من (83) طالباً وطالبةً للتحقق من:

أ) وجود غموض أو عدم وضوح أو في الفقرات، وكانت جميع الفقرات واضحة لجميع أفراد العينة الاستطلاعية.

ب) تقدير الزمن المناسب لتطبيق المقياس: تم تقدير زمن تطبيق المقياس بـ (30) دقيقة، أي (25) ثانية لكل فقرة، وتم تحديد هذا الزمن من خلال حساب المعدل لزمن أول فرد أنهى الإجابة على جميع فقرات المقياس وآخر فرد أنهى الإجابة على جميع فقرات المقياس.

6) تطبيق المقياس على عينة الدراسة الكلية، وتم التطبيق على جميع أفراد العينة بحضور الباحث لتحري المزيد من الدقة في النتائج.

7) التحقق من مطابقة الأفراد لنموذج اندريش، لحذف الأفراد غير المطابقين.

8) التحقق من مطابقة الفقرات لنموذج اندريش لحذف الفقرات غير المطابقة.

9) تحليل البيانات والتوصل إلى النتائج.

10) لتفسير الدرجات الكلية على المقياس تم تصنيف درجات الأفراد على المقياس في ثلاث فئات، وعلى النحو الآتي: درجة استخدام منخفضة، ودرجة استخدام متوسطة ودرجة استخدام مرتفعة، وذلك بتقسيم مدى القدرة على متصل السمة إلى ثلاث مسافات متساوية، والجدول (1) يبين ذلك.

جدول (1) درجة الحكم على استخدام إستراتيجيات حكمة الاختبار لدى طلبة الجامعات

الاستراتيجية	درجة الاستخدام بالدرجات الخام			درجة الاستخدام باللوجت		
	منخفضة	متوسطة	مرتفعة	منخفضة	متوسطة	مرتفعة
استخدام المقياس والاستعداد له	أقل من 8	من 8 إلى 19	أكبر من 19	أقل من 1.21-	من 1.21- إلى 1.32	أكبر من 1.32
الزمن	أقل من 7	من 7 إلى 13	أكبر من 13	أقل من 0.93-	من 0.93- إلى1.13	أكبر من 1.13
التخمين	أقل من 2	من 2 إلى 4	أكبر من 4	أقل من - 1.43	من 1.43- إلى 1.52	أكبر من 1.52
تجنب الخطأ	أقل من 6	من 6 إلى 11	أكبر من 11	أقل من - 1.09	من 1.09- إلى 1.41	أكبر من 1.41
الاستفادة من خصائص المقياس	أقل من 15	من 15 إلى 28	أكبر من 28	أقل من - 0.88	من 0.88- إلى 1.29	أكبر من 1.29
الكلي	أقل من 57	من 57 إلى 114	أكبر من 114	أقل من - 0.73	من 0.73- إلى 1.38	أكبر من 1.38

النتائج:

نتائج السؤال الأول: هل تتطابق الاستجابات على فقرات مقياس حكمة الاختبار لطلبة الجامعة مع افتراضات نموذج أندريش؟

للإجابة عن السؤال تم تحليل البيانات باستخدام البرنامج الإحصائي (Bigsteps)، وتم إجراء الآتي:

1) يحذف البرنامج استجابات الأفراد الذين يحصلون على علامة كاملة على المقياس، وكذلك الأفراد الذين يحصلون على العلامة (صفر) على المقياس، ولقد حصل (29) طالباً وطالبةً على العلامة الكاملة، وحصل (12) طالباً وطالبةً على العلامة (صفر) من عينة الدراسة على المقياس، وبهذا حذف البرنامج استجابات (41) طالباً وطالبةً في التحليل الأولي.

2) يحذف البرنامج الفقرات التي تأخذ علامة كاملة من قبل جميع أفراد العينة، مثل أن يختار جميع أفراد العينة الخيار (دائماً) على الفقرة، ويحذف البرنامج كذلك الفقرات التي تأخذ العلامة (0) مثل أن يختار جميع الأفراد الخيار (نادراً) على الفقرة، ولم يجب جميع أفراد العينة على فقرة باختيار البديل (دائماً) أو البديل (نادراً)، وبهذا لم يحذف البرنامج أي فقرة من فقرات المقياس في التحليل الأولي.

3) التحقق من مطابقة الأفراد للنموذج لحذف الأفراد غير المطابقين لنموذج اندريش، باستخدام مؤشري المطابقة الداخلية والمطابقة الخارجية، وتم تقدير معلم القدرة لكل فرد، وتم حساب الإحصائي الداخلي للمطابقة (INFIT) والإحصائي الخارجي للمطابقة (OUTFIT) لكل تقدير، والجدول (2) يبين نتائج ذلك.

جدول (2) الوسط الحسابي والانحراف المعياري لمعلم القدرة وإحصائي المطابقة الداخلية وإحصائي المطابقة الخارجية

الإحصائي	القدرة	إحصائي المطابقة الداخلية (INFIT)		إحصائي المطابقة الخارجية (OUTFIT)	
		متوسط المربعات MSNQ	قيمة الإحصائي STD	متوسط المربعات MSNQ	قيمة الإحصائي STD
الوسط الحسابي	0.02	0.97	0.03	0.99	0.01
الانحراف المعياري	0.95	0.02	0.96	0.03	0.98

يتبين من الجدول (2) أن قيمة الوسط الحسابي للقدرة كان يساوي (0.02) وهي قريبة من (0)، وهي القيمة التي يفترضها النموذج، وأن قيمة الانحراف المعياري للقدرة كانت تساوي (0.95)، وهي قريبة من الـ (1) وهي القيمة التي يفترضها النموذج، وكانت قيمة الوسط الحسابي لمتوسط المربعات للمطابقة الداخلية (0.97) والانحراف المعياري لمتوسط المربعات للمطابقة الداخلية تساوي (0.02)، وكانت قيمة الوسط الحسابي لمتوسط المربعات للمطابقة الخارجية تساوي (0.99) والانحراف المعياري لمتوسط المربعات للمطابقة الخارجية تساوي (0.03)، وهذه القيم تقترب من القيمة التي يفترضها النموذج، وهي (1) للوسط الحسابي و(0) للانحراف المعياري، وكان الوسط الحسابي لإحصائي المطابقة الداخلية يساوي (0.03) والانحراف المعياري لإحصائي المطابقة الداخلية يساوي(0.96)، وكان الوسط الحسابي لإحصائي المطابقة الخارجية يساوي(0.01) والانحراف المعياري لإحصائي المطابقة الخارجية يساوي (0.98)، وهذه القيم قريبة من القيم التي يفترضها النموذج, وهي (0) للوسط الحسابي, و(1) للانحراف المعياري، وتم استبعاد (53) طالباً وطالبةً من ذوي القدرات المتدنية وذوي القدرات المرتفعة من أفراد العينة، لأن قيمة إحصائي المطابقة الداخلية أو قيمة إحصائي المطابقة الخارجية المقابلة لدرجاتهم على المقياس كانت أكبر من (2)، والنموذج يفترض أن استجابة الفرد الملاحظـــة تبتعد عن استجابته المتوقعة إذا زادت قيمة

إحصـــائي المطابقة الخارجية المقابلـــة أو قيمـــة إحصائـــي المطابقـــة الداخلـــية لدرجاتـــه عـــن (2)(Linacre & Wright, 2006).

4) التحقق من مطابقة الفقرات لنموذج اندريش لحذف الفقرات غير المطابقة: تم استخدام مؤشري المطابقة الداخلية والمطابقة الخارجية، ويبين الجدول(3) قيم المتوسطات الحسابية والانحرافات المعيارية لمعالم الصعوبة، وإحصائيات المطابقة الداخلية والخارجية.

جدول (3) قيم الوسط الحسابي والانحراف المعياري لمعالم الصعوبة، وإحصائيات المطابقة الداخلية والمطابقة الخارجية

الإحصائي	الصعوبة	إحصائي المطابقة الداخلية (INFIT)		إحصائي المطابقة الخارجية (OUTFIT)	
		متوسط المربعات MSNQ	قيمة الإحصائي STD	متوسط المربعات MSNQ	قيمة الإحصائي STD
الوسط الحسابي	0.01	0.98	0.03	0.97	0.02
الانحراف المعياري	0.99	0.03	0.97	0.04	0.95

يتبين من الجدول (3) أن قيمة الوســط الحســـابي للصعوبة كان يساوي (0.01) وهي قريبة من (0)، وهي القيمة التي يفترضها النموذج، وأن قيمة الانحراف المعياري للصـــعوبة كانت تســـاوي (0.99)، وهي قريبة من الـ (1) وهي القيمة التي يفترضها النموذج، وكانت قيمة الوسط الحسابي لمتوسط المربعات للمطابقة الداخلية (0.98) والانحراف المعياري لمتوسط المربعات للمطابقة الداخلية تساوي (0.03)، وكانت قيمة الوسط الحسابي لمتوسط المربعات للمطابقة الخارجية تساوي (0.97) والانحراف المعياري لمتوسط المربعات للمطابقة الخارجية تساوي (0.04)، وهذه القيم تقترب من القيمة التي يفترضها النموذج، وهي (1) للوسط الحسابي و(0) للانحراف المعياري، وكان الوسط الحسابي لإحصائي المطابقة الداخلية يساوي (0.03) والانحراف المعياري لإحصائي المطابقة الداخلية يساوي(0.97)، وكان الوسط الحسابي لإحصائي المطابقة الخارجية يساوي(0.02)

والانحراف المعياري لإحصائي المطابقة الخارجية يساوي(0.95)، وهذه القيم قريبة من القيم التي يفترضها النموذج، وهي (0) للوسط الحسابي، و(1) للانحراف المعياري، وتم استبعاد (6) فقرات.

نتائج السؤال الثاني: هل يحقق تدريج فقرات مقياس حكمة الاختبار لطلبة الجامعة بنموذج اندريش افتراضات نظرية الاستجابة للفقرة؟

تقوم نظرية الاستجابة للفقرة على أربع افتراضات، وهي:

أولا) افتراض أحادية البعد: تم التحقق من هذا الافتراض من خلال إجراء الآتي:

1) التحليل العاملي باستخدام البرنامج الإحصائي (SPSS) للتحقق من عدد العوامل التي يقيسها المقياس، ويبين الجدول (4) قيم الجذور الكامنة ونسبة التباين المفسر ونسبة التباين التراكمية لكل عامل.

الجدول (4) القيم المميزة و نسبة التباين المفسر للعوامل على المقياس

العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين المفسر التراكمية
1	18.423	37.372	37.372
2	7.022	10.037	47.409
3	4.748	7.080	54.489
4	3.590	5.316	59.805
5	2.050	3.417	63.222
6	1.616	2.193	65.415
7	1.306	1.676	67.091
8	1.085	1.141	69.232

يتبين من الجدول (4) وجود (8) عوامل كانت قيم الجذور الكامنة لها أكبر من (1)، وأن نسبة التباين التي فسرها العامل الأول كانت تساوي (37.372)، وهي أكبر من (20%)، وهذا يدل على أن المقياس أحادي البعد، أي يقيس سمة واحدة وهي حكمة الاختبار، ولوجود عدد قليل من الفقرات على بعض العوامل تم استخدام فرز العوامل واستخدام التدوير المائل، وتم استخدام (5) عوامل فسرت مجتمعة ما مقداره (63.222) من التباين، تم استخدام معيارين في اختيار الفقرات وتصنيفها على العوامل، الأول أن لا يقل تشبع الفقرة على العامل الذي تنتمي له عن (0.30)،

والثاني أن لا يزيد تشبع الفقرة على العوامل الأخرى عن (0.30)، وبهذا تم حذف (7) فقرات، وبقي من المقياس (59) فقرة مقسمة على خمس استراتيجيات، والملحق (1) يبين المقياس بصورته النهائية، والجدول (5) يبين درجة تشبع الفقرات على العوامل الخمسة المستخلصة للمقياس.

جدول (5) تشبع الفقرات على العوامل الخمسة للمقياس

الفقرة	تشبع الفقرات على العوامل					الفقرة	تشبع الفقرات على العوامل				
	1	2	3	4	5		1	2	3	4	5
1		0.563				31				0.419	
2		0.561				32				0.480	
3		0.563				33					0.353
4		0.554				34					0.535
5		0.731				35					0.504
6		0.375				36					0.476
7		0.476				37					0.451
8		0.687				38					0.480
9		0.531				39					0.381
10		0.675			0.721	40					
11		0.402			0.759	41					
12		0.436			0.369	42					
13		0.656			0.624	43					
14		0.631			0.735	44					
15			0.362		0.806	45					
16			0.608		0.762	46					
17			0.631		0.772	47					
18			0.745		0.748	48					
19			0.621		0.802	49					
20			0.515		0.716	50					

الفقرة	تشبع الفقرات على العوامل					الفقرة	تشبع الفقرات على العوامل				
	1	2	3	4	5		1	2	3	4	5
21			0.394			51	0.747				
22			0.494			52	0.666				
23			0.711			53	0.729				
24				0.437		54	0.723				
25				0.391		55	0.698				
26				0.575		56	0.707				
27				0.669		57	0.811				
28				0.607		58	0.812				
29				0.362		59	0.803				
30				0.352							

يتبين من الجدول (5) أن (20) فقرة تشبعت على العامل الأول، و(14) فقرة تشبعت على العامل الثاني، و(9) فقرات تشبعت على العامل الثالث، و(9) فقرات تشبعت على العامل الرابع و(7) فقرات تشبعت على العامل الخامس، والجدول (6) يبين توزيع فقرات المقياس بصورته النهائية المكون من (59) فقرة مقسمة على خمس استراتيجيات.

جدول (6) استراتيجيات حكمة الاختبار وأرقام فقرات كل استراتيجية

الرقم	الاستراتيجية	أرقام الفقرات	عدد الفقرات
1	استخدام المقياس والاستعداد له	1-14	14
2	الزمن	15-23	9
3	التخمين	24-32	9
4	تجنب الخطأ	33-39	7
5	الاستفادة من خصائص المقياس	40-59	20
الكلي		1-59	59

2) تم حساب معامل ارتباط بيرسون بين علامات العينة الكلية، على كل فقره من فقرات المقياس وعلاماتهم الكلية على كل استراتيجية تنتمي لها الفقرة بعد حذف علامة الفقرة، والجدول (7) يبين نتائج ذلك.

جدول (7) معامل ارتباط بيرسون بين علامات العينة الاستطلاعية على كل فقره وعلاماتهم الكلية على المقياس

رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط	رقم الفقرة	معامل الارتباط
1	0.639**	16	0.621**	31	0.660**	46	0.574**
2	0.690**	17	0.613**	32	0.646**	47	0.470**
3	0.621**	18	0.715**	33	0.736**	48	0.524**
4	0.563**	19	0.715**	34	0.542**	49	0.599**
5	0.670**	20	0.711**	35	0.527**	50	0.605**
6	0.562**	21	0.754**	36	0.659**	51	0.602**
7	0.668**	22	0.619**	37	0.675**	52	0.665**
8	0.684**	23	0.770**	38	0.493**	53	0.574**
9	0.682**	24	0.607**	39	0.440**	54	0.506**
10	0.647**	25	0.518**	40	0.598**	55	0.639**
11	0.572**	26	0.588**	41	0.715**	56	0.628**
12	0.591**	27	0.776**	42	0.542**	57	0.662**
13	0.606**	28	0.769**	43	0.508**	58	0.560**
14	0.496**	29	0.535**	44	0.668**	59	0.607**
15	0.648**	30	0.533**	45	0.709**		

يتبين من الجدول (7) أن جميع قيم معاملات الارتباط كانت دالة إحصائياً عند مستوى الدلالة (α =0.01)، وتراوحت بين (0.440) وبين (0.776)، وهذا يدل على أن فقرات المقياس تقيس نفس السمة، وهي حكمة الاختبار.

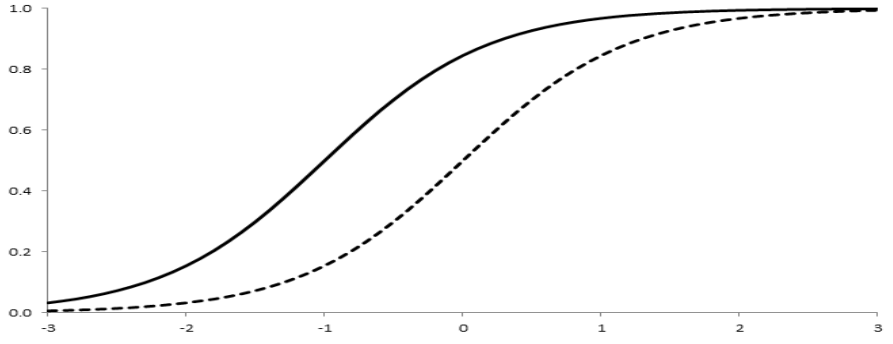
ثانياً) افتراض الاستقلال الموضعي: يتحقق هذا الافتراض إذا تحقق افتراض أحادية البعد، وبما أن افتراض أحادية البعد للمقياس تحقق، فبالتالي يتحقق افتراض الاستقلال الموضعي (Hambleton & Swaminathan, 1985)، ولتحري المزيد من الدقة تم حساب معاملات الارتباط بين فقرات المقياس على أداء جميع أفراد العينة، وكذلك أداء مجموعة الأفراد ذوي القدرات العليا، وهم يمثلون (27% من العينة الكلية)، الذين حصلوا على أعلى الدرجات على المقياس، وأداء مجموعة الأفراد ذوي القدرات الدنيا، وهم يمثلون (27% من العينة الكلية)، الذين حصلوا على أقل الدرجات على المقياس، وتم حساب متوسطات معاملات الارتباط في الحالات الثلاث، والجدول (8) يبين نتائج ذلك.

جدول(8) الوسط الحسابي لمعاملات الارتباط بين فقرات المقياس على العينة الكلية والمجموعتين العليا والدنيا

المجموعة	العدد	معاملات الارتباط
الكلية	2387	0.101
العليا	645	0.073
الدنيا	645	0.045

يتبين من الجدول (8) أن قيمة الوسط الحسابي لمعاملات الارتباط في حالة العينة ككل كانت أكبر من قيمة الوسط الحسابي لمعاملات الارتباط في حالة المجموعتين العليا والدنيا، وجميع قيم المتوسطات الحسابية لمعاملات الارتباط كانت قريبة من الصفر، وهذا يدل على أن فقرات المقياس لا يوجد من بينها فقرات تتأثر الإجابة عليها بفقرات أخرى، وهذا يدل على تحقق افتراض الاستقلال الموضعي (Onder, 2007).

ثالثاً) افتراض منحنى خصائص الفقرة: تم رسم العتبات ومنحنياتها لجميع فقرات المقياس، والشكل (2) يبين العتبتين ومنحنيات هاتين العتبتين للفقرة رقم (1)، العتبة الأولى تمتد من دائماً إلى أحياناً، والعتبة الثانية تمتد من أحياناً إلى نادراً.

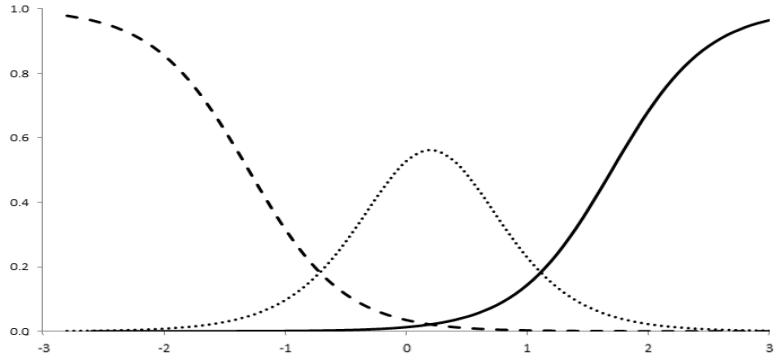


شكل (2) منحنيات عتبات الفقرة رقم (1)

P1:___ العتبة الأولى, وهي بين (0 و 1), P2 _ _: العتبة الثانية, وهي بين (1 و 2).

يتبين من الشكل (2) أن العتبتين كانتا متدرجتينن، وهذا يدل على تحقق افتراض منحنى خصائص الفقرة لهذه الفقرة، ولقد تحقق هذا الافتراض لجميع فقرات المقياس.

وتم عمل الرسم البياني للعلاقة بين مستويات التدريج، واحتمال القدرة للاستجابة على الفقرة لجميع الفقرات، والشكل (3) يوضح الرسم البياني للعلاقة بين مستويات التدريج واحتمال القدرة للاستجابة على الفقرة رقم (1).



_ _ _ (0)

…… (1)

____(2)

شكل (3) العلاقة بين مستويات التدريج واحتمال القدرة للاستجابة على الفقرة رقم (1)

يتبين من الشكل (3) وجود عتبتين للإجابة على الفقرة، الأولى تساوي حصول المفحوص على العلامة (صفر) أو العلامة (1), والثانية تساوي حصول المفحوص على العلامة (1) أو العلامة (2)، وهذه العتبات كانت تساوي (−0.4, 1.4) على الترتيب، ويتضح من الشكل (3) أن الخيارات متدرجة في الصعوبة، وبهذا يتحقق افتراض منحنى خاصية الفقرة للفقرة رقم (1)، ولقد تحقق هذا الافتراض لجميع فقرات المقياس.

رابعاً) التحرر من السرعة: تم التحقق منه من خلال حساب نسبة الأفراد الذين أكملوا الإجابة على المقياس، وكذلك حساب نسبة الفقرات التي لم يجيب عنها أفراد عينة الدراسة، وأجاب جميع أفراد العينة على فقرات المقياس، أي كانت نسبة الأفراد الذين أكملوا الإجابة على فقرات المقياس (100%)، وكذلك أجاب أفراد العينة على جميع الفقرات، أي كانت نسبة الفقرات التي تمت الإجابة عليها (100%)، ويتبين من هذه النسب أن المقياس يقيس القدرة ولا يقيس السرعة، وبالتالي تحقق فرض التحرر من السرعة، حيث أنه إذا كان نسبة الأفراد الذين

يكملوا الإجابة على المقياس تساوي (75%)، وتم الإجابة عن (80%) من فقرات المقياس، يتحقق افتراض التحرر من السرعة.

نتائج السؤال الثالث: هل لمقياس حكمة الاختبار لطلبة الجامعة باستخدام نموذج اندريش درجة مقبولة من الصدق والثبات؟

=للإجابة على السؤال تم إجراء ما يلي:

1) الصدق: يتحقق الصدق للمقياس في نظرية الاستجابة للفقرة إذا تحققت افتراضات النموذج المستخدم، وتحققت هذه الافتراضات كما تم توضيح ذلك في الإجابة عن السؤال الثاني وفي إجراءات الدراسة، أي تم التحقق من صدق المقياس من خلال الآتي:

أ) صدق المحتوى (المحكمين): وتم توضيح ذلك في الخطوة رقم (4) من إجراءات الدراسة.

ب) التحليل العاملي: وتم توضيح ذلك عند التحقق من افتراض أحادية البعد للمقياس في السؤال الثاني.

ت) الاتساق الداخلي: تم حساب معامل ارتباط بيرسون بين العلامات على كل فقره والعلامات الكلية على كل استراتيجية تنتمي لها الفقرة بعد حذف علامة الفقرة، وتم توضيح ذلك عند التحقق من افتراض أحادية البعد للمقياس في السؤال الثاني.

ث) الصدق التلازمي: تم تطبيق المقياس الذي طوره شاهر (Shaher, 2014) في دراسته التي هدفت إلى الكشف عن مدى امتلاك طلبة الجامعة لاستراتيجيات حكمة الاختبار على العينة الاستطلاعية، وتم حساب معامل ارتباط بيرسون لدرجات العينة الاستطلاعية بين كلا التطبيقين، وكان يساوي (0.89)، وهي قيمة مرتفعة، وهذا يدل على أن المقياس يتمتع بالصدق، وقد يعزى ذلك إلى الاستفادة من مراجعة الأدب النظري والدراسات السابقة في موضوع حكمة الاختبار، التي ساعدت الباحث في تعريف حكمة الاختبار، وحصر جميع الاستراتيجيات التي قد يستخدمها المفحوص عند الإجابة على فقرات الاختبار.

2) الثبات: يشير مفهوم الثبات في نظرية الاستجابة للفقرة إلى دقة تقدير مواقع الأفراد والفقرات على متصل السمة، وتحدد هذه الدقة بحساب معامل الفصل بين الفقرات (Gi) ومعامل الفصل بين الأفراد (Gp)، وللتحقق من ثبات المقياس تم إجراء ما يلي:

أ) تم حساب قيم معامل الفصل بين الفقرات وقيم معامل الفصل بين الأفراد، للمقياس بصورته النهائية بعدد فقراته (59) فقرة، وتم حساب معامل الثبات للفقرات وللأفراد باستخدام المعادلة(1)، والجدول (9) يبين قيم معاملات الثبات للفقرات وللأفراد.

$$R = G^2 / (1+ G^2) \qquad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

حيث: G هي معامل الفصل

جدول (9) قيم معاملات الثبات للفقرات وللأفراد

الاستراتيجية	معامل الفصل بين الفقرات	معاملات الثبات للفقرات	معامل الفصل بين الأفراد	معاملات الثبات للأفراد
استخدام المقياس والاستعداد له	8.04	0.984	2.79	0.886
الزمن	8.87	0.987	2.91	0.894
التخمين	7.75	0.983	3.55	0.926
تجنب الخطأ	5.44	0.967	3.09	0.905
الاستفادة من خصائص المقياس	7.97	0.984	3.14	0.908
الكلي	10.7	0.991	3.90	0.938

يتبين من الجدول (9) أن قيمة معامل الثبات للأفراد مرتفعة، وهذا يدل على كفاية عينة الأفراد على الفصل بين فقرات المقياس، ويتبين من الجدول كذلك أن قيمة معامل الثبات للفقرات مرتفعة، وهذا يدل على كفاية فقرات المقياس على الفصل بين الأفراد، ويتبين كذلك من الجدول أن جميع قيم معاملات الفصل بين الفقرات كانت أكبر من (2) وبهذا تكون الفقرات كافية للفصل بين الأفراد،

وكذلك جميع قيم معاملات الفصل بين الأفراد كانت أكبر من (2) وبهذا تكون عينة الأفراد كافية للفصل بين الفقرات.

ت) تم حساب معامل الثبات للمقياس باستخدام معادلة كرونباخ الفا، والجدول (10) يبين نتائج ذلك.

جدول (10) قيم معاملات الثبات باستخدام معادلة كرونباخ الفا

الاستراتيجية	قيم معاملات الثبات
استخدام المقياس والاستعداد له	0.901
الزمن	0.898
التخمين	0.913
تجنب الخطأ	0.883
الاستفادة من خصائص المقياس	0.931
الكلي	0.951

يتبين من الجدول (10) أن قيم معاملات الثبات باستخدام معادلة كرونباخ الفا تراوحت بين (0.883) وبين (0.931) للاستراتيجيات وللكلي كانت تساوي (0.951)، وهي قيم مرتفعة، وقد يعزى ذلك إلى طبيعة عينة الدراسة، فهم طلبة جامعات وعلى درجة عالية من الوعي في الالتزام بتعليمات تطبيق المقياس، وقد يعزى ذلك إلى دقة الإجراءات التي اتبعت في صياغة وكتابة وتطوير فقرات المقياس.

التوصيات:

توصي الدراسة بما يلي:

- إجراء العديد من الدراسات لبناء مقاييس تقيس حكمة الاختبار باستخدام نماذج أخرى من نماذج نظرية الاستجابة للفقرة.

- عمل دورات لأعضاء هيئة التدريس لتوجيه أنظارهم إلى استراتيجيات حكمة الاختبار التي قد يستخدمها الطلبة عند إجاباتهم على فقرات الاختبارات، للأخذ بها عند بناء اختباراتهم لقياس القدرة الحقيقية للطلبة.
- قياس مدى استخدام طلبة الجامعات الأردنية لاستراتيجيات حكمة الاختبار.
- تدريب الطلبة على استخدام استراتيجيات حكمة الاختبار.

Reference:

Abu Grad, H. (2015). Using Andrich Model in Constructing the Scale of the Attitudes of Faculty Staff Members Towards Students' Evaluation of Educational Practices at University Level, Journal of AL-Quds open University for Educational and Psychological and Studies, 4(15),13-44.

Abdel Wahab, M. (2007). The Effect of a Training Program of Test Wiseness Skill on the Achievement Performance and Test Anxiety among a sample of MInia Students Faculty Education. Unpublished Master Thesis, Minia University, Egypt.

Al-Trawneh, S. (2016). Developing a scale of resistance of temptation Students of Jordanian universities using Andrich model in latent traits theory. Journal of Mutah Lil-Buhuth Wad-Dirasat, 1 (4), 297-320.

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43 (4), 561-73.

Embretson, S. & Reiase, S. (2000). Item Response Theory for Psychologists. New jersey: Lawrence Erlbaum Associates.

Famularo, L. (2007). The Effect of Response Format and Test Taking Strategies on Item Difficulty: A Comparison of Stem-Equivalent Multiple-Choice and Constructed-Response Test Items. Unpublished doctoral dissertation, Boston College, UMI No. 3283877.

Haladyna, M., Downing. M. & Rodriguez, C. (2002): A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. Applied Measurement in Education, 15(3), 309–334.

Hambleton, R. & Swaminathan, H. (1985). Item Response Theory: Principles and Applications, Boston, Kluwer: Nijhoff Publishing.

Hamadnh, I. (2011). The use of University Students of Test wiseness Strategies. Journal of Al-Manara,17(1),391-312.

Hammad, D. (2010). Relationship between test-wiseness and performance outcomes using a multiple-choice achievement test constructed according to Rasch model and applied to female students at the College of Education (literary sections) in Umm Al-Qura University, Unpublished doctoral dissertation, Umm Al-Qura University.

Linacre, J. & Wright, B (2006). A user guide to Bigsteps (Rasch: Model Computer Program), Retrieved from www.winsteps.com.

Mohamed, A, Gregory, P, & Austin, Z. (2006). Test wiseness among International Pharmacy Graduates and Canadian Senior Pharmacy Students. American Journal of Pharmaceutical Education, 70(6),131-146.

Nguyen, D. (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.

Onder, I. (2007). An Investigation of Goodness of Model Data Fit Model Veri Uyumunun Araştirilmasi, Hacettepe Universities Egitim Fakultesi Dergisi. (32)P210-220.

Orth, J. (1995). The Effect of Test-Taking Strategies for Multiple-Choice Tests. Unpublished doctoral dissertation, University of Pittsburgh, UMI No. 9614214

Oudeh, A .(1989). The effect of teaching guessing strategy on the university acquisition for of test wiseness- experimental study, Yarmouk Research: series of human and social sciences.6(22),123-138.

Peng, Y. (2005). Test Preparation Strategies and Test Taking Strategies Use in Chinese High School Students. Unpublished doctoral dissertation, University of Nevada, Las Vegas, UMI No. 1435632.

Rogers, T. & Bateson, J. (1991). The influence of test-wiseness on performance of high school seniors on school leaving examinations. Applied Measurement in Education, 4 (2) ,159 – 183.

Rogers, T. & Harley, D. (1999). Empirical Comparison of Three and Four-Choice Items and Tests: Susceptibility to Test wiseness and Internal Consistency Reliability. Educational and Psychological Measurement, 59(2), 216-234.

Rogers, T. & Yang, P. (1996). Test - wiseness: Its nature and application. European Journal of Psychological Assessment, 12(3), 247-259.

Roznowski, M. & Bassett J. (1992). Training Test-wiseness and Flawed Item Types. Applied Measurement in Education, 5(1), 14-35.

Sarnacki, S. (1979). An examination of Test wiseness in the cognitive domain. Review of Educational research. 49(2),252-279.

Shaher, S. (2014). The tes twiseness among Tabouk University students in light of variables of specialization, the cumulative rate and academic level. Journal of Educational and Psychological Science,15(2),245-273.

Susan. E. & Steven, P. (2000). Item Response Theory for Psychologists, London, Lawrence Erlbaum Associates, publishers Mahwah, New Jersey.

Wenden, A. (1991). Learner Strategies for Learner Autonomy. Englewood Cliffs. NJ: Prentice-Hall.