مجلة الجامعة الإسلامية للدراسات التربوية والنفسية، المجلد الحادي والعشرون، العدد الأول، ص407– ص434 يناير 2013
ISSN 1726-6807 http://www.iugaza.edu.ps/ar/periodical/

استخدام نموذج المحاولات ذي الحدين في فحص تقديرات المحكمين لمقاييس الاتجاهات

د. ماجد محمود الجودة

استاذ مساعد في القياس والتقويم التربوي،كلية التربية

جامعة الأميرة نورة بنت عبد الرحمن،الرياض

ملخص: هدفت هذه الدراسة إلى استخدام أسلوب نموذج المحاولات ذي الحدين لفحص تقديرات المحكمين لملاءمة فقرات مقاييس الاتجاهات، وذلك بالاعتماد على النظرية الحديثة في القياس.

وقد اشتملت عينة الدراسة على (150) محكماً من درجات علمية مختلفة: معلمي رياضيات، ومشرفي رياضيات، وأعضاء هيئة تدريس جامعات حكومية وخاصة في الرياضيات، والقياس والتقويم، والمناهج، وعلم النفس، وطلبة الدراسات العليا في القياس والتقويم، وأساليب تدريس الرياضيات في الأردن. ولأغراض الدراسة تم استخدام مقياس اتجاهات نحو الرياضيات يتألف من(72) فقرة وقد تمّ تطبيقه على (413) طالباً وطالبة من طلبة الصف العاشر الأساسي من مدارس مديرية تربية إربد الأولى.

وقد عُرضت فقرات المقياس على عينة المحكمين (n=150) من أجل تقييمها بأسلوب المحاولات ذي الحدين، بالاعتماد على معيارين (نسبة الاتفاق بين المحكمين لا تقل عن 70%، ونسبة الاتفاق بين المحكمين لا تقل عن80%)، وقد تم دراسة الخصائص السيكومترية للفقرات المعتمدة، بالاعتماد على استجابات أفراد عينة الدراسة، ثم تمّ فحص تقديرات المحكمين باستخدام نموذج المحاولات ذي الحدين، ثم أعيد تقدير الخصائص السيكومترية للفقرات التي تمّ اعتمادها بالمعيارين بعد استبعاد المحكمين غير المطابقين.

وأظهرت نتائج مقارنة الخصائص السيكومترية للمقياس- قبل استبعاد المحكمين غير المطابقين للنموذج وبعد استبعادهم- وجود فروق ذات دلالة إحصائية في معاملات الصدق والثبات، حيث ظهر تحسن إيجابي على هذه المعاملات. وتوصي الدراسة باستخدام هذا الأسلوب في إعداد مقاييس الاتجاهات وتحكيمها.

الكلمات المفتاحية: المحاولات، ذي الحدين، فحص تقديرات، المحكمين

Using Binomial Trials Model for Examining Referees' Rates of Attitude Scales

Abstract: The current study aimed at using method of Binomial Trials Model for examining judges' rates of the appropriateness of Attitude scale's items depending on modern measurement theory. The sample of the Study consisted of (150) Judges from different categories: mathematics teachers, mathematics supervisors, public and private university faculty members teaching mathematics, measurement and evaluation, curriculum, and psychology, and graduate students of measurement and evaluation, and mathematics teaching methods in Jordan. The study administered a

mathematics attitude scale of (72) items to 413 male and female 10th Graders in Irbid first Educational directorate.
The scale's items were given to a sample of referees (n=150) to evaluate by the said model, based on two criteria (referees' agreement rate should not be less than 70%, referees' agreement rate should not be less than 80%).
The Psychometric properties of the items adopted in the two criteria were studied depending on the responses of the subjects. Moreover, the referees' rates were examined by binomial trials model. The referees who did not fit the model were suspended, then the agreement rates were calculated again and the psychometric properties were estimated again in the model.
The Psychometric comparison results of the properties before suspending referees who did not fit the model and afterwards showed that there are significant statistical differences in reliability and validity coefficients, revealing positive improvements in them. The study recommended using this model in designing attitude scales.

مقدمة:

تعد مقاييس الاتجاهات من المقاييس المهمة التي يستخدمها الباحثون، وقد كانت بداياتها الأولى عام 1925، عندما قام بوجاردوس (Bogardus) ببناء مقياس أطلق عليه مقياس البعد الاجتماعي، وأراد من خلاله التعرف على مدى تقبل الأمريكيين أو نفورهم من أبناء القوميات الأخرى. ثم طور ثيرستون (Thurstone) عام 1929 مقياساً للاتجاهات عرف فيما بعد باسمه، حيث اقترح ثيرستون إعطاء أهمية لكل فقرة بالقدر الذي تستحقه، والتمييز بين هذه الفقرات باستخدام وسيطات قيم عددية يتم تقديرها نتيجة تحكيم الفقرات من قبل محكمين أكفاء وذوي خبرة على متصل متساوي المسافات. ثم جاء ليكرت (Likert) عام 1932 ليضع مقياساً كمياً يكشف درجة القبول أو الرفض لدى المفحوصين، وليسد الثغرة الرئيسة في طريقة ثيرستون المعتمدة على المحكمين. أما جتمان (Guttman) فقد انشأ عام 1947 مقياساً متجمعاً متدرجاً، وحقّق فيه شرطاً هاماً هو الحصول على مقياس يقيس سمة أو اتجاهاً من بعد واحد(One-dimensional)؛ ذلك لأن جتمان يعتبر الميدان خاضعاً للقياس المتدرج التجميعي، إذا أمكن ترتيب الاستجابات بطريقة معينة بحيث تجعل من يجيب على إحدى الفقرات بالقبول أعلى مرتبة من الذي يجيب عنها بالرفض، وبذلك يتسنى معرفة نمط إجابته عليها (Shaw & Wright, 1967).

إنّ أي مقياس يتم تصميمه يتكون في نهاية الأمر من مجموعة من الفقرات، والفقرة كما هو معلوم، هي وحدة بناء المقياس؛ لذا يجب اختيارها بناءً على دراسات نظرية وميدانية وتجريبية وإحصائية للتثبت من صلاحيتها للقياس المقصود. وإنّ هذه الدراسات التي يطلق عليها اسم تحليل الفقرات Item Analysis ضرورية للحصول على فقرة مناسبة، وصالحة من حيث شكلها، وتكوينها، وصعوبتها، وصدقها وترتيبها فعلى سبيل المثال فإن الفقرة "315 + 678 + 32.16 "

غير صالحة إذا كانت في مقياس لذكاء الأطفال في سن ما قبل المدرسة؛ لأنها شديدة الصعوبة ولا تميز بينهم، ولو وضعنا في نفس المقياس فقرة كهذه " هل تقلق كثيراً أثناء نومك بالليل؟" فإنها لا تصلح لأنها ليست صادقة في قياس الذكاء وإن كانت صادقة في قياس الشخصية، والدراسات الميدانية والتجريبية هي التي سوف تفصل في كل هذا، وبناءً عليها سوف نحذف الفقرة، أو نضعها كما هي في المقياس النهائي، بل ويتحدد ترتيبها أيضاً في المقياس وفقا لصعوبتها (طه، 2006).

ومن الأمور التي يجب مراعاتها عند تطوير أدوات القياس درجة صدق وثبات تلك الأدوات. ويرى العديد من الباحثين في مجال قياس الاتجاهات وبناء مقاييسها أنّ ضعف معاملات الصدق والثبات يُعدّ من الأسباب الهامة في ظهور التناقض في نتائج البحوث والدراسات (عودة، 1990). وأنّ العديد من الدراسات والبحوث التربوية توصي بضرورة التشدد في معاملات الصدق والثبات في مقاييس الاتجاهات (الصمادي والسرطاوي، 1996). ودعا شبيكي (Schibeci, 1984) نتيجةً لمراجعته أكثر من 200 دراسة تناولت قياس الاتجاهات نحو العلوم لضرورة التشديد على جوانب الصدق والثبات في الأبحاث المنشورة في تلك المجلات.

وغالباً ما يعتمد الباحثون على الأحكام الذاتية لمجموعة من الخبراء والمختصين في التحقق من صدق أدوات القياس، وخصوصاً في مقاييس الاتجاهات، حيث تُعرض فقرات المقياس على مجموعة من الخبراء ويطلب منهم إبداء رأيهم فيما يتعلق بصدق هذه الفقرات؛ أي مدى انسجامها وقدرتها على قياس الاتجاه، وكذلك اُستخدمت أحكام المختصين في بناء الاختبارات محكية المرجع. وتستخدم أحكام المختصين كذلك في تدريج فقرات مقاييس الاتجاهات لثيرستون، كطريقة الفترات المتساوية البعد، بحيث تقدم مجموعة كبيرة من الفقرات لعدد من المحكمين، ويطلب من كل محكم أن يقدر على تدريج من إحدى عشرة نقطة أهمية ظهور أو عدم ظهور كل فقرة في القائمة النهائية، أو مقدار مساهمتها في قياس السمة المرغوب قياسها فإذا كان الغرض هو اختيار الفقرات الأنسب، فيكون الاختيار على أساس التقديرات العالية أولاً مع وجود الاتفاق بين هذه التقديرات ثانياً، أما إذا لم يكن هناك مجالٌ للاختيار في الفقرات فإن وزن الفقرة أو أهميتها النسبية هو وسيط التقديرات(عودة، 1993).

وقد وُجهّتْ إلى أحكام المختصين بشكل عام في مختلف المجالات انتقادات شديدة من قبل العديد من الباحثين حيث يرى لوكاس (Lucas, 1975) أن الإجماع بين المحكمين لا يضمن صحة ملاءمة الأداة لقياس الاتجاهات العلمية حيث إن هذا الأسلوب عُرْفٌ انبثق من المبدأ السياسي الذي يأخذ برأي الأغلبية والإجماع بينهم، وهذا أمر خاطئ.

ويرى شريغلي وكوبالا (Shrigley&Koballa,1984) أنه من الضروري استخدام تحليل فقرات المقياس وعدم اللجوء إلى آراء المحكمين، واتفق الخليلي (1989) مع ذلك عندما قام بتحليل فقرات مقياس الاتجاهات نحو الفيزياء للتحقق من صدقه، ولم يستند إلى آراء المحكمين في الحكم على ذلك، إلا أنه استرشد بآرائهم في كتابة الفقرات.

ويتفق منبي (Munby, 1982) مع لوكاس ويضيف جوانب أخرى من نقاط الضعف في استخدام الأحكام الذاتية للمحكمين؛ فهو يرى أن سياق فهم المحكم يختلف عن سياق فهم من يطبق عليه المقياس ولا يكون المقياس صادقاً إلا إذا تطابق السياقان، لكنه يرى أن السياق يختلف من شخص لآخر، وقد يختلف لنفس الشخص إذا أجري قياس اتجاهاته قبل وبعد تنفيذ معالجة تجريبية عليه، ويقترح أن تصاغ الفقرات بشكل أسئلة تطرح على عينة من مجتمع الدراسة المعني بالمقياس لمعرفة السياق الذي يهم كل فقرة.

ويتفق برات (Bratt, 1984) مع الكثير من الانتقادات الموجهة إلى الأحكام الذاتية ويعتبر استخدام الأحكام مصدراً للخطأ في صدق الأدوات ناتجاً عن العامل البشري، ويجب تناول البعد الانفعالي في إجراءات التحقق من صدق الأدوات، ويقترح أن يتم كتابة العديد من الصيغ لنفس الفقرة، ثم يجرّب المقياس ويستخدم الحاسوب لإجراء تحليل عنقودي cluster analysis، ثم معايرة متعددة الأبعاد multidimensional scaling.

ويرى ألدرسون (Alderson, 1993) أنه لا يوجد معيار للمفاضلة بين أحكام المختصين؛ وذلك نظراً لصعوبة الاتفاق بينهم فاختلاف المختصين يعني اختلاف الأحكام.

ويرى براون (Brown, 1996) أن أحكام المختصين هي نتاج خلفيات نظرية وعملية فردية، وأن مطوري الاختبارات يستخدمونها لمرة واحدة، ويعتمدونها في جميع مراحل بناء الاختبار، وهذا يجعلها عرضة للأخطاء.

ويرى حاجيبور (Hajipour, 2003) أن أحكام المختصين تتأثر بشكل مباشر بمستويات المعرفة عندهم. وعند استحالة التحقق من درجة الصدق والموضوعية في أحكامهم، فإن الاعتماد عليها يزيد من احتمالية الوقوع في الأخطاء، و دعا إلى التقليل من الاعتماد على أحكام المختصين في تطوير اختبارات اللغات.

ويرى كرونباخ (Cronbach,1995) أن خصائص المحكم مثل الشدة والتساهل والتفاعل بين المحكم والسمة المقيمة تعمل كمصدر للتباين بين تقييمات المحكمين، مع أنها ليست ذات صلة بعملية التقييم ومعاييره وتظهر على شكل تباين منتظم يهدد صدق وإجراءات التقييم.

استخدام نموذج المحاولات ذي الحدين في فحص تقديرات المحكمين

حدثت في غضون العقود الماضية تطورات جوهرية وسريعة في منهجيات القياس النفسي وطرق تصميم الاختبارات والمقاييس، فعلى الرغم من انتشار نظرية القياس الكلاسيكية في بناء الاختبارات وتحليل البيانات المستمدة منها، إلا أن هذه النظرية لم تستطع التغلب على كثير من المشكلات السيكومترية المعاصرة، أو تناول العديد من قضايا القياس المنهجية المهمة؛ وأدى ذلك إلى محدودية وقلة فاعلية هذه النظرية في كثير من التطبيقات العملية، لذلك قام علماء القياس بجهود بحثية مبتكرة لتطوير نظرية سيكومترية معاصرة يمكن استخدامها في التغلب على الكثير من المشكلات، أطلق عليها نظرية استجابة الفقرة (Item Response Theory) (IRT)، وهذه النظرية بصورة كبيرة استندت إلى التقنيات المتقدمة في الحاسوب (علام، 2005).

يتضح من العرض السابق أن استخدام الأحكام الذاتية يعاني بشكل عام من مشاكل وانتقادات مختلفة من قبل الباحثين، إلا أن اللجوء إليها أحياناً يكون ضرورةً ملحةً لا يستغنى عنها وخصوصاً في مقاييس الاتجاهات ومن هنا تظهر الحاجة لتطويرها وتحسينها فتأتي هذه الدراسة لاستخدام أسلوب يستند إلى النظرية الحديثة في القياس في فحص تقديرات المحكمين الثنائية لملاءمة فقرات مقاييس الاتجاهات بالاعتماد على نموذج المحاولات ذي الحدين في فحص تلك الأحكام وتالياً الشرح النظري لهذا الأسلوب .

يفضل بعض المحكمين استخدام الأسلوب غير المباشر(الإسقاطي) في عملية التحكيم، حيث يمكن له أن يسقط رأيه بمدى ملاءمة الفقرة من وجهة نظر خبير متخصص بتلك الفقرة؛ أي بدلاً من أن يطلب من المحكم أن يحدد ما إذا كان يوافق على الفقرة أم لا، يطلب منه أن يحـــدد احتمـــال أن يوافق خبير متخصص عليها، وفي هذه الحالة يمكن استخدام نموذج المحاولات ذي الحدين الذي يستند إلى نظرية ذات الحدين، وتقوم هذه النظرية على توزيع ذي الحدين ويسمى توزيـــع ذي الحدين أحيانا باسم توزيع برنولي Bernoulli Distribution، ويكون فيه احتمال النجـــاح ثابتـــاً ومستقلاً عند تكرار التجربة العشوائية أكثر من مرة، وتسمى كل تجربة عشوائية بمحاولة برنولي (Bernoulli Trial) .

ويصف توزيع برنولي التجارب العشوائية التي يكون نواتجها ناتجين فقط، الناتج الأول يـــسمى نجاحاً ويرمز له بالواحد(1)، والثاني فشلاً ويرمز له بالصفر (0)، ويحقق المتغير العشوائي (x) الذي له توزيع برنولي الخصائص التالية:-

1- نتيجة المحاولة الواحدة إما نجاحاً (success) وإما فشلاً (failure).

2- المحاولات مستقلة .

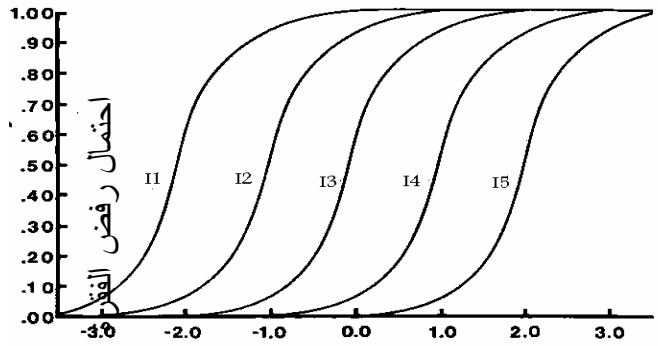
3- احتمال النجاح p نفسه لكل المحاولات.

لذا فإن احتمال ظهور المشاهدة (x) من بين عدد المحاولات (m) يعطى بالعلاقة التالية: -

$$\Pr(X = x|m, p) = \pi_x = \binom{m}{x} p^x (1-p)^{m-x}.$$

(Neter, Wasserman & Whitmore, 1982)

والنموذج النظري هنا يقوم على أن المحكم يحدد موقع الخبير المتخصص على متصل السمة (التخصصية) ويحدد موقع الفقرة على نفس المتصل، وفي ضوء هذا النموذج تزداد احتمالية رفض الفقرة كلما زادت التخصصية لدى الخبير، إذ يفترض أن يزداد عدد المعايير التي يتم الحكم على الفقرة في ضوئها؛ أي يتوقع أن تكون فرص رفض الفقرة من قبل الخبير قليلة عندما يكون موقعه على متصل السمة إلى يسار موقع الفقرة وعالية عندما يكون موقعه إلى يمينها، فموقع الخبير المتخصص يشير إلى الحد الأدنى من المعايير التي يطبقها على الفقرة، وموقع الفقرة يشير إلى عدد المعايير المتحققة في الفقرة، وفي ضوء النظرية الحديثة للقياس يمكن تقدير كل من موقع الخبير المتخصص qj وموقع الفقرة bi على متصل السمة إذا عرف احتمال الرفض(أو القبول) للفقرة من قبل الخبير، والشكل (1) يوضح منحنى احتمالات رفض الفقرة بدلالة موقع الخبير المتخصص وموقع الفقرة على متصل السمة.



الشكل(1): منحنى احتمالات رفض الفقرة التخصصية وعدد المعايير

يُظهرْ الشكل(1) منحنى احتمالات رفض الفقرة لخمس فقرات، لكل فقرة المنحنى الخاص بها (I5, I4, I3, I2, I1)ويلاحظ من الشكل أن احتمال رفض الفقرة يزداد بزيادة درجة التخصصية لدى المحكم (زيادة عدد المعايير المطبقة). ولمعرفة احتمال رفض الفقرة سيتم عرض فقرات المقياس على مجموعة من المحكمين، ويطلب من كل محكم أن يقدر احتمال رفض الفقرة من قبل خبير يمتلك الحد الأدنى من التخصصية من خلال الإجابة عن السؤال الآتي : " إذا أعطيت هذه

الفقرة لمئة خبير يمتلك كل منهم الحد الأدنى من التخصصية، ما عـدد الخبـراء مـنهم الـذين سيعتبرون هذه الفقرة غير ملائمة لقياس السمة ؟ "، وفي هذه الحالـة يمكـن تطبيـق نمـوذج المحاولات ذي الحدين حيث يصبح العدد (100) هو عدد المحاولات (m)، والمطلوب من المحكم هو تقدير عدد الخبراء الذين يطبقون الحد الأدنى من المعايير لرفض الفقرة، وهذا العدد هو (x)، وبالتالي يمكن تقدير احتمال رفض الفقرة بالاعتماد على رؤية المحكم للحد الأدنى للمعايير التـي يملكها الخبير المختص لرفض الفقرة والتي يعبر عنها هنا بدرجة التخصص التي تزداد احتماليـة رفض الفقرة بزيادتها وعلى عدد المعايير المتحققة في الفقرة. وبالتالي يمكن اسـتخدام نمـوذج راش متعدد الأوجه Many-Faceted Rasch Model من خلال برنامج FACETS الذي قـام بتطويره لينكر Linacer بهدف تحليل البيانات المتعلقة بالمفحوصين والناتجة من أوجـه عـدة على سبيل المثال الفقرة، والمفحوص، والمحكم وهكـذا، (Zhu, Ennis & Chen , 1998).

ويعد هذا النموذج موسعاً لنموذج راش، وهو يستند إلى المعادلة آلاتية.

$$\Pr(X = xji / m, qj, bi) = \binom{m}{xji} \frac{\exp[xji\,(qj - bi)]}{[1 + \exp(qj - bi)]^m}$$

(Engelhard & Anderson, 1998)

حيث تتراوح x_{ji} من صفر إلى m وتكون بالتالي نتيجة المعادلة رقم 2 هي احتمال الحصول على العدد x_{ji} من m من الخبراء الذين سيرفضون الفقرة i التي درجة صعوبتها b_i (درجة المعايير المتحققة فيها)، و qj تمثل لنا عدد المعايير التي يطبقها الخبير المختص (درجة التخصص للخبير) لرفض الفقرة، والتي تقابل قدرة المفحوص في نموذج راش. وبالتالي يمكن فحص تقديرات المحكمين من خلال مقارنة عدد المشاهدات الملاحظة x_{ni} بالمشاهدات المتوقعة E_{ni} بالاعتماد على النموذج. ويمكن فحص الباقيات المعيارية وفق المعادلة التالية:-

$$z_{ni} = \frac{x_{ni} - E_{ni}}{\left[\sum_{k=0}^{m} (k - E_{ni})^2 \pi_{nik}\right]^{1/2}}$$

$$E_{ni} = \sum_{k=0}^{m} k\pi_{nik}$$

حيث أن :

x_{ni} : العدد الملاحظ للخبراء الذين يرفضون الفقرة i والمقدر من قبل المحكم n.

E_{ni} : العدد المتوقع للخبراء الذين يرفضون الفقرة i والناتج من النموذج.

π_{nik} : احتمال الحصول على العدد x_{ni} من m من الخبراء الذين يملكون الحد الأدنى من المعايير من قبل المحكم n على الفقرة I، (Engelhard & Anderson, 1998)

وتجدر الإشارة كذلك بان قيم هذه الباقيات المعيارية يمكن تلخيصها على شكل وسط مربعات الأخطاء (Mean Square Error) كإحصائي يستخدم للحكم على مطابقة نموذج المحاولات ذي الحدين للبيانات المولدة من المحكمين والفقرات ويحسب وسط مربعات الخطأ من المعادلة التالية: -

$$MSE_n = \sum_{i=1}^{I} z_{ni}^2 / (I).$$

حيث إن (I) تعني عدد الفقرات فعندما تطابق البيانات النموذج فإن القيمة المتوقعة لوسط مربعات الخطأ تساوي واحداً (MSE=1)، والأساس المنطقي لهذه القيمة المتوقعة مبنية على مجموع مربعات الفروق المعيارية Z^2 وعدد الفقرات وهي مشابهة إحصائياً لتوزيع مربع كاي بدرجات حرية (I). وأن القيمة المطلوبة لمطابقة البيانات تتراوح بين (0.6) و (1.5) (Engelhard & Anderson, 1998). وبالتالي يمكن معرفة المحكمين الذين لا تتطابق تقديراتهم مع النموذج ليتم استبعادهم. ولتوضيح نموذج المحاولات ذي الحدين نفترض أنه تم عرض خمس فقرات أعدت لقياس اتجاه ما على لجنة من المحكمين مؤلفة من عشرة محكمين، وتم الطلب من كل محكم الإجابة عن السؤال التالي: من بين مئة محكم يملك كل منهم الحد الأدنى من التخصصية قدّر عدد الذين سيعتبرون هذه الفقرة غير ملائمة لقياس السمة منهم ؟ و في هذه الحالة يصبح عدد المحاولات (M=100) وعدد الفقرات (I=5) وعدد المحكمين (N=10)، وتم تشكيل مصفوفة البيانات التي تم الحصول عليها من المحكمين، ولتكن كما في الجدول 1.

الجدول 1

مصفوفة البيانات (المحكم x الفقرة) في الأسلوب الثاني

الفقرة	المحكم									
	J1	**J2**	**J3**	**J4**	**J5**	**J6**	**J7**	**J8**	**J9**	**J10**
1	36	37	37	36	70	37	38	40	39	42
2	40	41	42	45	47	47	46	45	46	50
3	60	61	62	65	65	65	68	67	68	74
4	61	62	64	65	66	65	64	65	65	70
5	20	21	22	23	23	25	26	26	27	60

وباستخدام برنامج FACETS للينكر و رايت (Linacre& Wright ,1994) والذي يقوم على نموذج راش المتعدد الأوجه. وبتطبيق نموذج المحاولات ذي الحدين تم إيجاد الحسابات المتعلقة بالنموذج والتي تمثل qj، وهو الحد الأدنى للمعايير المطبقة (التخصصية) المقدر من قبل المحكم بوحدات الترجيح اللوغارتيمي لوجيت، وb_i والتي تمثل معامل الصعوبة للفقرة i (درجة المعايير المتحققة فيها) المقدر من قبل المحكم بوحدات الترجيح اللوغارتيمي لوجيت ، بالإضافة إلى المشاهدات المتوقعة الناتجة من النموذج E_{ni}، وقيم الوسط الحسابي لمربعات الأخطاءMSE_n.

ويلخص الجدول (2) مصفوفة البيانات وقيم b_i ، qj ، E_{ni} ، Z_{ni} ، MSE_n.

الجدول 2

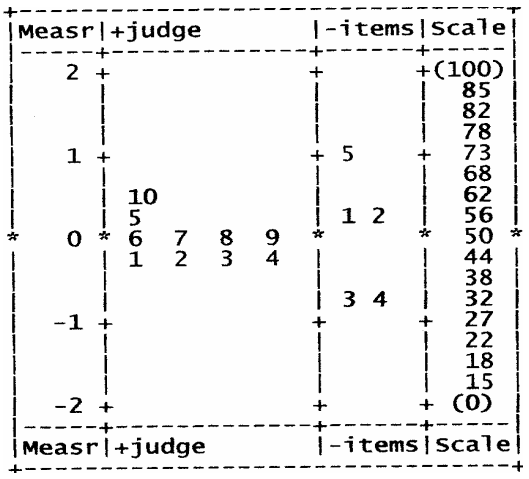
القيم الملاحظة والقيم المتوقعة وقيم الباقيات المعيارية ومواقع المحكمين والفقرات ومتوسط الرافضين للفقرة ومعدل رفض المحكم ومتوسطات مربعات الأخطاء

الفقرة	المحكم										معدل الرافضين للفقرة	Bi	MSE
	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10			
E_{ni} 1 Z_{ni}	6 35.6 0.08	3 7 36.7 0.07	3 7 37.7 -0.14	3 6 39.1 -0.64	7 0 47.04 4.6	3 7 40.18 0.65	3 8 40.81 -0.57	4 0 41.02 -0.12	3 9 41.44 -0.49	4 2 52.6 -2.12	41.2	0.3	2.71
E_{ni} 2 Z_{ni}	4 0 39.2 0.16	4 1 40.3 0.15	4 2 41.3 0.14	4 5 42.8 0.44	4 7 50.85 -0.77	4 7 43.9 0.63	4 6 44.54 0.29	4 5 44.75 0.05	4 6 45.18 0.16	5 0 56.4 -1.28	44.9	0.15	0.3
3 E_{ni} Z_{ni}	6 0 60.2 -0.04	6 1 61.3 -0.05	6 2 62.3 -0.06	6 5 63.7 0.26	6 5 70.83 -1.28	6 5 64.73 0.054	6 8 65.33 0.56	6 7 65.53 0.31	6 8 65.92 0.44	7 4 75.2 - 0.28	65.5	-0.71	0.24
4 E_{ni} Zn_i	6 1 59.4 0.33	6 2 60.4 0.32	6 4 61.5 0.52	6 5 62.9 0.43	6 6 70.1 -0.89	6 5 63.9 0.22	6 4 64.53 -0.11	6 5 64.72 0.06	6 5 65.12 -0.02	7 0 74.5 -1.04	64.7	-0.67	0.26

5 E_{ni} Z_{ni}	2 0 22.8 -0.66	2 1 23.6 -0.61	2 2 24.4 -0.55	2 3 25.5 -0.58	2 3 32.14 -1.96	2 5 26.4 -0.31	2 6 26.88 -0.2	2 6 27.05 -0.24	2 7 27.4 -0.09	6 0 37.2 4.73	27.3	0.93	2.78
معدل الرفض للمحكم	4 3.4	4 4.4	4 5.4	4 6.8	4 9	4 7.8	4 8.4	4 8.6	4 9	5 9.2			
qj	- 0.92	- 0.52	- 0.02	- 0.41	0 .18	-0.01	- 0.70	- 0.6	- 0.50	0 .4			
MSE	0.12	0.1	0.12	0.24	5.6	0.19	0.16	0.04	0.09	5.93			

يتضح من الجدول 2 أن تقديرات المحكمين لعدد الخبراء الذين يستطيعون رفض الفقرة هي بمثابة تقدير لمعايير الفقرة، فمثلاص بلغ الوسط الحسابي لتقدير المحكمين لعدد الخبراء الذين يستطيعون رفض الفقرة الأولى 41.2 وبقسمة الوسط الحسابي على 100 تصبح قيمة معيار الفقرة المقدر من قبل المحكمين تساوي 0.412 ولذا تصبح قيم المعايير للفقرات المقدرة من قبل المحكمين متراوحة ما بين 0.273 ولغاية 0.655 . أما معاملات الصعوبة (مواقع الفقرات) bi بوحدات التدريج اللوغاريتمي لوجيت الناتجة من نموذج المحاولات ذي الحدين BTM، فقد تراوحت من0.71- ولغاية 0.93 لوجيت، وكذلك يوضح الجدول القيم المولدة من النموذج لمواقع المحكمين حيث تراوحت قيم مواقع المحكمين من 0.92- إلى 0.4. وبالاعتماد على نتائج مطابقة البيانات لنموذج المحاولات ذي الحدين يتضح من الجدول 2 أن هناك ثلاث مشاهدات من أصل خمسين مشاهدة لم تطابق البيانات، حيث كانت قيمة الباقيات المعيارية Standardized Residuals لها أكبر من 2. فمثلاً القيمة المتوقعة لعدد الخبراء الذين يرفضون الفقرة 1 بلغت 47.04 للمحكم (J5) أقل من القيمة الملاحظة 70 وهي ذات دلالة إحصائية، والقيم الأخرى تعود للمحكم (J10) على الفقرة 1 والفقرة 5 وتلخصت البواقي المعيارية على شكل أوساط مربعات الخطأ (MSE_n) وتراوحت قيم MSE_n بالنسبة للمحكمين من 0.04 ولغاية 5.93 وللفقرات من 0.24 إلى 2.78 .

وتجدر الإشارة كذلك إلى أن هذا النموذج يمكن من خلاله معايرة تقديرات المحكمين والفقرات على تدريج مشترك Common Scale وذلك من خلال خريطة المتغيرMap of Variable التي تهدف إلى تحديد مواضع كل من المحكمين و الفقرات على نفس التدريج كما في الشكل 2.



الشكل: (2) خريطة المتغير للمحكمين والفقرات

يتضح من الشكل 2 بأن المحكم رقم 10 كان الأكثر تشدداً في قبوله للفقرات، ويتضح كذلك أن الفقرة رقم 5 حصلت على أعلى تقدير من قبل المحكمين.

وفي ضوء العرض التوضيحي السابق يتبيّن الأساس النظري والمنطقي الذي ستقوم عليه هذه الدراسة في استخدام أسلوب نموذج المحاولات ذي الحدين.

مشكلة الدراسة:

اعتاد الباحثون عند تطوير أدوات القياس وخصوصاً في مقاييس الاتجاهات، عرض فقرات المقياس على مجموعة من المحكمين الذين يدعي الباحث أنهم من ذوي الاختصاص، ولاحظنا في ضوء ما تقدم أن أحكام المختصين تعاني وتواجه كثيراً من المشاكل و الانتقادات، إلا أن استخدامها يكون أحياناً ضرورةً ملحةً ولا نستطيع تجنبها في كثير من المواقف، فهي تستخدم كثيراً في تطوير أدوات القياس والحكم على مصداقيتها.

وتُردّ الانتقادات الموجهة لأدوات القياس إلى أن الأشخاص الذين يقومون بتطويرها غير مؤهلين، ولا توجد معايير للأشخاص الذين يقع عليهم الاختيار في الحكم عليها؛ فقد يلجأ بعض المحكمين إلى الحكم على مدى ملاءمة فقرات المقياس لقياس اتجاه ما بشكل عشوائي، وقد يتشدد بعضهم في قبول فقرات الاستبانة، في حين تجد تساهلاً من الآخرين من خلال عملية التسرع في إبداء الأحكام، وقد تعتمد تقديراتهم على انفعالاتهم النفسية والظروف المحيطة بهم مما يجعل مصداقية الأداة عرضةً للخطأ ومجالاً للشك. ومن هنا تظهر الحاجة لفحص تقييمات هؤلاء المحكمين بهدف الوقوف على جوانب مهمة فيها والعمل على تقييمها وإمكانية تحسين جودتها، فتأتي هذه الدراسة بهدف استخدام أسلوب نموذج المحاولات ذي الحدين لفحص تقديرات المحكمين في ضوء

النماذج الحديثة لنظرية استجابة الفقرة والتي من خلالها يمكن تحسين تلك التقييمات، والكشف عن المحكمين الذين تكون تقديراتهم غير ملائمة مع مجموعة المحكمين، ليتم استبعادهم من عملية التحكيم ودراسة أثر ذلك على صدق وثبات الأداة عند التطبيق، وذلك بعد عملية التحليل لتقديرات المحكمين. وتتحدد مشكلة الدراسة بالإجابة عن الأسئلة الآتية:

1- هل توجد فروق ذات دلالة إحصائية ($a = 0.05$) في ثبات المقياس قبل فحص تقييمات المحكمين، وبعدها باستخدام نموذج المحاولات ذي الحدين؟

2- هل توجد فروق ذات دلالة إحصائية ($a = 0.05$) في صدق المقياس قبل فحص تقييمات المحكمين، وبعدها باستخدام نموذج المحاولات ذي الحدين؟

اهمية الدراسة:

يُعدّ صدق الأدوات التي يستخدمها الباحثون من الأمور المهمة جداً في البحث التربوي، وخصوصاً عندما يرغبون بتعميم نتائج دراساتهم، وكثيراً ما يلجأ الباحثون إلى تطوير الاستبانات واستخدامها لجمع بياناتهم والتي تتطلب مهارةً وجهداً في تصميمها. ويعتمد الباحثون على مساعدة المختصين في التحقق من صدق هذه الأدوات إلا أن اختيارهم للمحكمين قد لا يتم وفق المعايير؛ ممّا يجعلها عرضةً للشك والأخطاء وضعف الصدق، ممّا يؤدي إلى عدم سلامة نتائج الدراسة بأكملها وعدم صحة وسلامة النتائج تجعل البحث بلا قيمة ولا يعدو أن يكون مضيعة للوقت والجهد والمال للباحث أو المستفيد من البحث وكل هذا يحصل لأن بعض المحكمين لا يملكون الخبرة، أو لا يبدون الاهتمام اللازم. والمشكلة التي تواجه هؤلاء الباحثين تتمثل في كيفية استبعاد هؤلاء المحكمين بطريقة علمية، بل كيفية تخليص الباحث لدراسته من جوانب الضعف التي من الممكن أن تعتري دراسته.

إن هذه الدراسة تطمح إلى تحسين وتطوير آلية التحكيم، لفحص تقديرات المحكمين وتقديم وسيلة يستفيد منها الباحثون في التحقق من صدق أدواتهم، وذلك من خلال توظيف أحد نماذج النظرية الحديثة في القياس لتطوير أسلوب لفحص تقديرات المحكمين الثنائية لمدى ملاءمة فقرات مقاييس الاتجاهات والعمل على ضبط تقديراتهم، وبذلك توفر الحلول المناسبة للمشاكل والانتقادات الموجهة للطرق المتبعة في الحكم على مدى ملاءمة فقرات مقاييس الاتجاهات، والمتمثلة بالأحكام الذاتية لمجموعة من الخبراء، وهذا سيؤدي إلى تطوير الأدوات المستخدمة من قبل الباحثين في بحوثهم. إن مجمل الهدف الأساسي من هذه الدراسة هو تحسين جودة حكم الخبراء في مقاييس الاتجاهات بواسطة تلافي عشوائية التحكيم، وإيجاد آلية للتحقق من صدق التحكيم، وذلك عن

طريق استبعاد المحكمين غير المطابقين للنموذج المستخدم في هذه الدراسة، والتي تقوم أساساً على النظرية الحديثة في القياس.

أدبيات الدراسة:

لقد حظي موضوع فحص تقديرات المحكمين باهتمام بعض الباحثين، وعند الرجوع الى الأدب النظري والدراسات السابقة تبين أن اهتمام الباحثين في هذا المجال كان في جوانب متعددة منها: فحص أثر خصائص المحكمين على دقة التحكيم، أو مقارنة أحكام المختصين بغير المختصين، أو فحص أثر الخبرة والتدريب على جودة ودقة التحكيم بالطرق الإحصائية التقليدية. وبعض الدراسات استخدمت النماذج الحديثة في تحليل تقديرات المحكمين باستخدام نموذج المحاولات ذي الحدين في فحص جودة تقديرات المحكمين في تقدير علامة القطع على الاختبارات محكية المرجع، ففي مجال فحص أثر خصائص المحكم على دقة التحكيم، قامت وكسلي ويوتز (Wexely&Youtz,1985) بدراسة هدفت إلى فحص العلاقة بين خصائص المحكم ودقته، حيث تم دراسة بعض الخصائص الشخصية والخصائص المعرفية والمتغيرات المتعلقة بالعمل، واعتقادات المحكم عن الطبيعة البشرية، واشتملت عينة الدراسة على 24 ممرضة خبيرة، متوسط أعمارهن 44.5 سنة، وبمتوسط خبرة 7.3 سنة، قمن بالإشراف على 82 ممرضة مساعدة، متوسط أعمارهن 38.4 سنة، بمتوسط خبرة 4.7 سنة، بحيث أشرفت كل ممرضة خبيرة على 6 ممرضات مساعدات بالمتوسط ولمدة عام كامل، وقامت المشرفات بتقييم الممرضات المساعدات في نهاية العام، وطبقت الباحثتان خمسة مقاييس نفسية من أجل دراسة الخصائص المتعلقة بالممرضات المشرفات، وحسبت المتوسطات والانحرافات المعيارية على المقاييس النفسية، وكذلك الخصائص السيكومترية لفقرات المقاييس، ومعاملات الارتباط. وأظهرت النتائج أن هناك علاقة ارتباطيه بين خصائص المحكم ودقة التحكيم، حيث بيّنت النتائج أن اعتقادات المحكم عن الآخرين مهمة في دقة التقييم، فالمحكمون المتساهلون ينظرون نظرة إيجابية للآخرين والطبيعة البشرية، حيث ينظرون إليهم على أنهم محبون لغيرهم، ويستحقون الثقة، ومستقلون، لذلك كانت تقديراتهم عالية وقليلة التباين وقليلة الدقة، في حين أن المحكمين المتشددين والذين يعتقدون بتغير الطبيعة البشرية يوزعون تقديراتهم وينشرونها، فلذلك كانت تقديراتهم عالية التباين وأكثر دقة.

أما دراسة المقرن والورع (2003) فقد هدفت إلى توضيح العلاقة بين درجات مجموعة من المحكمين، من خلال استكشاف مدى تجانس أو تباعد درجاتهم المعطاة لمشروعات طلاب في مقرر التصميم المعماري، في قسم العمارة وعلوم البناء في كلية العمارة والتخطيط بجامعة الملك

سعود في الرياض، وشارك في عملية التقييم ثلاث فئات مختلفة هي:أساتذة المقرر وعددهم سبعة، ومحكمون خارجيون لم يسبق لهم أن حضروا مناقشات مشاريع الطلبة وعددهم خمسة، والطلبة أنفسهم وعددهم 48 طالباً. وبلغ عدد المشاريع في هذه الدراسة 28 مشروعاً. واختبرت الدراسة ثلاثة فرضيات إحصائية تتمحور حول: عدم اختلاف درجات المحكمين المعطاة لمشروعات التصميم المعماري المقدمة من الطلبة، واستخدم اختبار (ت) للمقارنة بين درجات المحكمين، وخلصت الدراسة إلى وجود فروق دالة إحصائياً بين متوسط الدرجات التي وضعها المحكمون لمشروعات الطلاب، واتضح أن مصدر هذه الفروق نتج من ابتعاد درجات المحكمين الزائرين عن الدرجات الأخرى من المحكمين، وتساهم نتائج هذه الدراسة في بيان أسباب التقارب أو التعارض المحتملة في درجات المحكمين، وتشجع على المساهمة في مناقشة مداخل فلسفية جديدة للنقد المعماري، يمكن استعمالها في تحكيم مشروعات الطلاب.

وفي مجال استخدام نموذج المحاولات ذي الحدين في تقييم تقديرات المحكمين قام انجلهارد و اندرسون (Engelhard & Anderson, 1998) بدراسة بعنوان: "استخدام نموذج المحاولات ذي الحدين لفحص تقديرات المحكمين في تقدير مستويات الأداء"، وهدفت هذه الدراسة لوصف أسلوب جديد لفحص جودة الأحكام القيمية للمحكمين التي يتم الحصول عليها من تقديرات التحكيم، باستخدام نموذج ذي الحدين. وتكونت عينة الدراسة من 26 محكماً، طلب منهم تقدير درجة القطع لاختبار رياضيات مكون من 60 فقرة من نوع الاختيار من متعدد للصف الحادي عشر في ثلاث جولات تحكيمية متعاقبة باستخدام أسلوب أنجوف المعدل، في الجولة الأولى قدر المحكمون درجة القطع للاختبار دون تزويدهم بمعلومات عن الفقرات، وفي الجولة الثانية تم تزويد المحكمين بمعلومات عن تقديراتهم في الجولة الأولى وتمثلت بالوسط الحسابي، أعلى تقدير وأدنى تقدير وطلب منهم تقدير درجة القطع في ضوء المعلومات المتوفرة، وفي الجولة الأخيرة تم تزويد المحكمين بمعاملات الصعوبة الفعلية للفقرات والتي تم الحصول عليها تجريبياً، ومعلومات عن تقديراتهم في الجولة الثانية، وطلب منهم تقدير درجة القطع للاختبار. واستخدم برنامج FACETS لتحليل البيانات المولدة من الجولات الثلاث لفحص جودة تقديرات المحكمين باستخدام نموذج ذي الحدين. وأظهرت النتائج وجود اختلافات واضحة في معاملات الصعوبة المقدرة من المحكمين في كل جولة تحكيمية، وأظهرت استقراراً في معاملات الصعوبة المقدرة من المحكمين مع الجولات، وبلغت قيمة معامل الثبات 0.92 وكانت أعلى قيمة، وأشارت النتائج إلى وجود اختلافات واضحة بين المحكمين في رؤيتهم للحد الأدنى للكفاية، وكانت هذه الفروق تزيد مع الجولات.

وفي الأردن قام بني عطا (2005) باستخدام نموذج المحاولات ذي الحدين؛ لفحص جودة تقديرات المحكمين لدرجة القطع لاختبار محكي المرجع بدلالة عدد بدائل الفقرة في الرياضيات. وتكون الاختبار من 30 فقرة من نوع الاختيار من متعدد، بثلاثة نماذج تختلف قي عدد البدائل فقط في ثلاث جولات تحكيمية. وتكونت عينة الدراسة من 343 طالباً وطالبة من الصف العاشر الأساسي، و45 محكماً من المعلمين، وتم توزيع المحكمين على ثلاث مجموعات متساوية. واستخدم الباحث برنامج FACETS لفحص جودة تقديرات المحكمين من خلال نموذج ذي الحدين، وأظهرت النتائج وجود فروق جوهرية ذات دلالة إحصائية بين تقديرات المحكمين تعزى لصالح الجولة التحكيمية، وعدم وجود فروق ذات دلالة تعزى لعدد بدائل الفقرة، وكشفت النتائج عن أن عدد المحكمين الذين لم تتوافق تقديراتهم مع توقعات النموذج كان متقارباً في الجولتين الأولى والثانية لجميع نماذج الاختبار، في حين كان هناك تحسن ملحوظ على عدد المحكمين الذين توافقت تقديراتهم مع النموذج، وبينت النتائج أن هناك تبايناً في تقديراتهم للحد الأدنى من الكفاية لفقرات نماذج الاختبار الثلاث من خلال تحليل البواقي المعيارية لتقديراتهم. كما أظهرت النتائج أن تقديرات المحكمين لدرجات القطع لنماذج الاختبار اتسقت في الجولة الثانية، وأن التباين بين المحكمين قد انخفض في هذه الجولة، وبعد استبعاد المحكمين الذين لم تتطابق تقديراتهم مع النموذج كشفت النتائج بأن علامات القطع المقدرة لنماذج الإختبار كانت متقاربة في جميع الجولات، وأن ثبات تصنيف الطلبة استقر في الجولات التحكيمية خاصة في الجولة الثانية، وكشفت أيضا عن ارتفاع معاملات دقة القرار لتصنيف الطلبة لجميع نماذج الاختبار على مدار الجولات الثلاث.

نلاحظ من العرض السابق لهذه الدراسات أن اهتمامات الباحثين قد ركزت على مقارنة أحكام المختصين مع غير المختصين، ودراسة الخصائص المتعلقة بالمحكم، وأثر ذلك على دقة وأخطاء التحكيم، حيث أظهرت نتائج هذه الدراسات أنه ليس من السهولة استبدال أحكام المختصين بغير المختصين، وتبقى مسألة اختيار المحكم المختص مسألة صعبة ومهمة، وإن تدريب المحكم أمر مهم في تحسين أحكامه وبيّنت الدراسات أن الخصائص المتعلقة بالمحكم تلعب دوراً مهماً في دقة التحكيم.ولقد أظهرت بعض الدراسات الآثار السلبية لخصائص المحكمين مثل الشدة واللين والنزعة المركزية وضيق المدى وتفاعل المحكم وغيرها على دقة تقييماتهم.

وأيدت نتائج الدراسات أهمية استخدام النماذج الحديثة إلى جانب النظرية التقليدية في القياس في دراسة الخصائص السيكومترية لأدوات القياس، ونظراً لما تلعبه الأحكام الذاتية في التحقق من صدق مقاييس الاتجاهات فإن نتائج هذه الدراسات تبرر أهمية استخدام هذه النماذج في التحقق من

ذلك، والذي يتمثل في هذه الدراسة، كونها ستعيد النظر في طريقة التحقق من صدق مقاييس الاتجاهات من خلال استخدام نموذج من نماذج النظرية الحديثة في القياس لفحص تقييمات المحكمين .

الطريقة والإجراءات

مجتمعات الدراسة وعيناتها:

اولاً:الطلبة

تكوّن مجتمع الدراسة من جميع طلبة الصف العاشر في مدارس مديرية تربية إربد الأولى، والبالغ عددهم 7957 طالباً وطالبة (3964 طالب،3993 طالبة)، والتي تم الحصول عليها من قسم التخطيط في مديرية تربية إربد الأولى، وتكونت عينة الدراسة من 9 مدارس حيث تم اختيار شعبة، أو شعبتين عشوائياً من كل مدرسة.

ثانيا: المحكمون.

تكونت عينة المحكمين من 150 محكماً من معلمي الرياضيات، ومشرفي الرياضيات وأعضاء هيئة تدريس جامعات حكومية وخاصة في الرياضيات والقياس والتقويم والمناهج وعلم النفس، ومن طلبة الدراسات العليا في القياس والتقويم وأساليب تدريس الرياضيات في الأردن. ويوضح الجدول 3 أعداد المحكمين موزعين حسب المؤهل العلمي.

الجدول 3

أعداد المحكمين ومؤهلاتهم العلمية

الدرجة العلمية	العدد الكلي
دكتوراه	31
ماجستير	55
بكالوريوس	64
المجموع	150

أداة الدراسة:

تمثلت أداة الدراسة في مقياس الاتجاهات نحو الرياضيات الذي طوّره أبوزينة والكيلاني (1980)، ويتكون من 60 فقرة موزعة بالتساوي على ستة مقاييس فرعية، ولأغراض الدراسة الحالية تم إضافة 12 فقرة أخرى قدّر الباحث أن غالبيتها العظمى ليس لها علاقة بقياس

الاتجاهات نحو الرياضيات، وذلك لغاية الاستفادة منها في الكشف عن جدية التحكيم، وبواقع فقرتين في كل بعد، وبالتالي أصبح المقياس مكوناً من 72 فقرة بواقع 12 فقرة لكل بعد

صدق المقياس: تم التحقق من الدلالات التالية عن صدق المقياس:

الأولى: مستخلصة من التحليل النظري الذي سبق إعداد المقياس الذي تضمن تعريف المفهوم المقاس وتحديد أبعاده وتعريفها إجرائيا وصياغة الفقرات وفق العلاقة المنطقية بين مضمون الفقرة والتعريف الإجرائي للبعد المقاس.

الثانية: تعبر عن نوع من صدق المفهوم المتحقق في أحكام المختصين الذين عرض عليهم المقياس، وأبدوا أحكامهم حول ارتباط الفقرات بالبعد المقاس والتعديلات المترتبة على هذه الأحكام.

الثالثة: تم حساب معاملات الارتباط بين المقاييس على الأبعاد الفرعية، وذلك من خلال التحليلات على عينة التجريب وأيدت نتائج صدق المقياس من حيث أنها توضح التكوين العاملي المتوقع.

ثبات المقياس: استخرج أبو زينة والكيلاني (1980) معاملات الثبات النصفية لأبعاد المقياس من استجابات عينة تجريبية مكونة من 84 فرداً، وتراوحت القيم الناتجة بعد تصحيحها بمعادلة سبيرمان- براون بين 0.63 الى 0.88 على الأبعاد المختلفة للمقياس. وقد عُدّت قيم معاملات الثبات مقبولة لأغراض تطبيق المقياس في الدراسة.

الخصائص السيكومترية للمقياس في الدراسة الحالية:

أولاً: صدق المقياس

تم التحقق من صدق المقياس من خلال الدلالات الآتية:

الأولى : حساب معاملات الارتباط بين المقاييس على الأبعاد الفرعية، وذلك من خلال التحليلات على عينة الدراسة ولوحظ أن معاملات الارتباط فيما بين الأبعاد والأداة الكلية أقل من نظيرتها في المقياس الأصلي، وربما يعود ذلك إلى تجانس أفراد الدراسة الحالية، حيث جميعهم في صف واحد، أما أفراد الدراسة الأصلية، فكانوا من فئات مختلفة في الصفوف والمستوى التعليمي.

الثانية: حساب معامل الارتباط بين علامة الطالب على مقياس الاتجاهات وبين علامته في التحصيل الدراسي في مادة الرياضيات للفصل الدراسي الأول لعام 2008-2009، وذلك من خلال عملية التحليل على استجابات عينة الطلبة المكونة من 413 طالباً وطالبة، حيث بلغ معامل الارتباط 0.54 وهو ذو دلالة إحصائية عند مستوى دلالة 0.01.

ثانياً: ثبات المقياس في هذه الدراسة

استخرجت معاملات الثبات النصفية من استجابات عينة الطلبة والمكونة من 413 طالباًوطالبة والتي أجريت عليها عملية تحليل الفقرات في هذه الدراسة، وتراوحت القيم الناتجة بعد تصحيحها بمعادلة سبيرمان- براون بين 0.79 الى 0.90 وتُعدّ هذه القيم كافية لأغراض استخدام المقياس في الدراسة الحالية.

إجراءات الدراسة:

فيما يلي عرض للخطوات التي اتبعها الباحث للوصول إلى نتائج الدراسة.

أولاً: تحكيم المقياس من المحكمين.

بعد تصميم المقياس بصورته النهائية، والمكون من 72 فقرة على الأبعاد الستة، تم عرضه على عينة المحكمين، لإجراء عملية التحكيم بالأسلوب المقترح، حيث طُلب من كل محكم أن يقوم بتقدير عدد الخبراء الذين سيعتبرون الفقرة ملائمة لقياس الاتجاه فيماً لو أعطيت لمئة خبير متخصص في تلك الفقرة، وأن يقوم بوضع ذلك العدد في العمود المخصص، وقد شُرحت هذه التعليمات شفوياً للمحكمين، وتم الحصول على موافقة وزارة التربية والتعليم لعرض المقياس على عينة من معلمي الرياضيات في المدارس التابعة لمديرية تربية اربد الأولى .

ثانياً:تطبيق مقياس الاتجاه نحو الرياضيات على الطلبة.

طُبّق المقياس المكوّن من 72 فقرة بتدريج خماسي (ليكرت)، على عينة الطلبة، بعد الحصول على موافقة وزارة التربية والتعليم لتطبيق هذا المقياس على مدارس مديرية تربية إربد الأولى، وتم توضيح طريقة الاستجابة عليه.

ثالثا:جمع البيانات وتخزينها.

جُمعت الاستبانات التي تم الحصول عليها من خلال التطبيق على عينة الطلبة وعينة المحكمين، وتم حصر أعدادها، وترقيمها وحفظها في ملفات خاصة، كما تم الحصول على علامات الطلبة في التحصيل الدراسي في مادة الرياضيات للفصل الاول لعام 2008-2009 م وذلك من خلال علاماتهم في دفاتر المعلمين في المدارس المختلفة وبعد ذلك تم إدخال استجابات المحكمين والطلبة، وعلاماتهم في التحصيل الدراسي الى برنامج التحليل الإحصائي SPSS في ملفات خاصة بالمحكمين وبالطلبة.

رابعاً:تحليل البيانات ومعالجتها إحصائياً

مرّت عملية تحليل البيانات بالخطوات الآتية:

1- من خلال برنامج التحليل الإحصائي SPSS تم تحليل استجابات المحكمين على ملفات المحكمين، وتم استخراج المتوسطات الحسابية والنسب المئوية للفقرات، وتم تحديد الفقرات التي أجمع عليها ما نسبته على الأقل 70% أو 80% من المحكمين.

2- حُلّلت استجابات الطلبة على الفقرات التي تم الإجماع عليها من قبل المحكمين، كما ظهرت من الخطوة (1.) ومن خلال برنامج التحليل الإحصائي SPSS حُسب معامل الارتباط بين علامة الطالب على مقياس الاتجاهات وبين علامته في التحصيل الدراسي في مادة الرياضيات للفصل الدراسي الأول لعام 2008-2009 كمؤشر على صدق مقياس الاتجاهات، واُستخرج معامل ثبات الاتساق الداخلي للفقرات وفق معادلة كرنباخ ألفا، ومؤشرات التمييز للفقرات.

3- اُستخدم برنامج التحليل (FACETS) لتحليل استجابات المحكمين التي تم الحصول عليها من نموذج المحاولات ذي الحدين، وتم تفحص الأشخاص الذين لا تتطابق بياناتهم الإحصائية مع هذا النموذج، وحُصرت أعدادهم وأسقطت استجاباتهم من ملف البيانات الخاص بنموذج المحاولات ذي الحدين.

5- بعد إسقاط استجابات المحكمين الذين لا تطابق بياناتهم الإحصائية النموذج في هذا الأسلوب، أعيد حساب نسب الاتفاق على الفقرات، وتم تحديد الفقرات التي أجمع عليها ما نسبته على الأقل 70% أو 80% من المحكمين .

6- أعيد تحليل استجابات الطلبة على الفقرات التي تم الإجماع عليها من قبل المحكمين ، بعد إسقاط استجابات المحكمين الذين لا تتطابق بياناتهم الإحصائية مع النموذج في هذا الأسلوب، وحسب معامل الارتباط بين علامة الطالب على مقياس الاتجاهات وعلامته في مبحث الرياضيات للفصل الدراسي الأول لعام 2008-2009 م، كمؤشر على صدق مقياس الاتجاهات، واُستخرج معامل ثبات الاتساق الداخلي للفقرات وفق معادلة كرنباخ ألفا، ومؤشرات التمييز للفقرات.

7- لمقارنة معاملات الصدق اُستخدم اختبار مقارنة معاملي الارتباط للبيانات المرتبطة، وذلك بعد تعديل معامل الصدق الظاهر بعد تطبيق النموذج في هذا الأسلوب ،وذلك نظراً لاختلاف التباين في استجابات الطلبة على الأداة قبل وبعد تطبيق النموذج، وذلك من خلال معادلة التعديل الآتية :

$$r_{YX} = \frac{r_{y'x'}}{\sqrt{r^2_{y'x'} + \frac{s^2_{x'}}{s^2_x}(1 - r^2_{y'x'})}}$$

حيث:

r_{YX} : معامل الصدق بعد التعديل.

$r_{y'x'}$: معامل الصدق قبل التعديل.

$s^2_{x'}$: تباين المجموعة قبل التعديل .

s^2_x : تباين المجموعة بعد التعديل.

(Allen & Yen, 1979)

وتمت مقارنة معاملات الثبات للأداة قبل إسقاط استجابات المحكمين غير المطابقين للنموذج وبعده لهذا النموذج، أي مقارنة معاملات الثبات التي ظهرت في الخطوتين 6,5 باستخدام اختبار معاملي الثبات المعدّلين بمعادلة سبيرمان براون للبيانات المرتبطة (W) (Alsawalmeh&Feldt,2000).

8- تم مقارنة معاملات الثبات من خلال استخدام اختبار مقارنة معاملات الثبات لاختبارين طبقا على نفس العينة. (Alsawalmeh&Feldt,1994)

النتائـــج ومناقشتها

للإجابة عن أسئلة الدراسة المتعلقة بمقارنة الخصائص السيكومترية (الصدق، الثبات) للمقياس قبل فحص تقييمات المحكمين وبعدها باستخدام نموذج المحاولات ذي الحدين، جرى تحليل تقييمات المحكمين على فقرات المقياس المكون من 72 فقرة باستخدام هذا الأسلوب في التحكيم، وحُسبت نسب الاتفاق على الفقرات بين المحكمين البالغ عددهم 150 محكماً. وتبين أن 38 فقرة قد حققت معيار 70%، و أن 26 فقرة قد حققت معيار 80% . وقد حُسب مؤشر التمييز (معامل ارتباط الفقرة بالأداة)، لكل فقرة من الفقرات المعتمد وفق معيار 70%، وتجدر الإشارة إلى أن الفقرات المعتمدة وفق معيار 80% متضمنة في تلك الفقرات، وتبين أن معاملات التمييز تراوحت بين 0.29 و 0.63 بوسيط يساوي 0.47، وتتضمن هذه الفقرات فقرتين من الفقرات المضافة للمقياس بنسبة 16%، أما الفقرات الأصلية في المقياس فتبلغ نسبتها 60%.

وقد جرى تقدير معامل الصدق وفق معامل الارتباط بين علامة الطالب على الأداة كمقياس للاتجاهات نحو الرياضيات، وعلامته في مبحث الرياضيات للفصل الدراسي الأول لعام 2008-2009 م، في كلتا الأداتين. وتم تقدير الثبات للأداتين وفق معامل كرونباخ ألفا، حيث تتكون الأولى من 38 فقرة معتمدة وفق معيار 70%، وتتكون الثانية من 26 فقرة معتمدة وفق معيار

80%، وتبين أن معامل الارتباط بين علامة الطالب على هذه الفقرات كمقياس للاتجاهات نحو الرياضيات، وعلامته في مبحث الرياضيات للفصل الدراسي الأول لعام 2008-2009 م كمؤشر على صدق المقياس قد بلغ 0.55 وهو ذو دلالة إحصائية عند مستوى دلالة 0.01 .

وتبين أن معامل الثبات وفق معادلة كرونباخ ألفا للأداة المكونة من 38 فقرة التي تم اعتمادها وفق معيار 70% قد بلغ 0.88 تقريباً.

وتبين أن معامل الارتباط بين علامة الطالب على هذه الفقرات كمقياس للاتجاهات نحو الرياضيات، وعلامته في مبحث الرياضيات للفصل الدراسي الأول لعام 2008-2009 م كمؤشر على صدق المقياس قد بلغ 0.565 وهو ذو دلالة إحصائية عند مستوى دلالة 0.01 وتبين معامل الثبات وفق معادلة كرونباخ ألفا للأداة المكونة من 26 فقرة التي تم اعتمادها وفق معيار 80% قد بلغ 0.866،.

وتم فحص تقييمات المحكمين للفقرات وفق أسلوب التحكيم باستخدام نموذج المحاولات ذي الحدين، حيث أخضعت تقييمات 150 محكماً لفقرات المقياس البالغ عددها 72 فقرة إلى ذاكرة الحاسوب، وجرى تحليلها باستخدام برنامج التحليل (FACETS) المصمم لتحليل البيانات وفق نموذج راش متعدد الأوجه، ويقوم هذا البرنامج بحذف المحكمين الذين لا تتلاءم استجاباتهم مع جميع المحكمين وهم الذين لا تتطابق بياناتهم الإحصائية مع النموذج.

ويوضح الجدول 4 أعداد ونسب المحكمين المطابقين وغير المطابقين لنموذج المحاولات ذي الحدين موزعة حسب المؤهل العلمي للمحكم.

الجدول رقم 4

أعداد ونسب المحكمين المطابقين وغير المطابقين لنموذج المحاولات ذي الحدين موزعة حسب المؤهل العلمي للمحكم

الدرجة العلمية	غير المطابقين لنموذج ذي الحدين	النسبة المئوية	المطابقون لنموذج ذي الحدين	النسبة المئوية	المجموع
الدكتوراه	25	%80	6	%20	31
الماجستير	43	%78	12	%22	55
البكالوريوس	46	%72	18	%28	64
المجموع	114	%76	36	%24	150

يتضح من الجدول (4)أن أعداد المحكمين غير المطابقين للنموذج الثاني كانت عاليةً جداً في مختلف الدرجات العلمية، وكانت أعلاها نسبياً في درجة الماجستير، حيث ما يقارب 78٪ ممن

يحملون درجة الماجستير، لم تطابق بياناتهم الإحصائية هذا النموذج، و76% من جميع المحكمين لم تتطابق بياناتهم الإحصائية مع هذا النموذج. وتراوحت الدرجات اللوجستية للمحكمين في هذا النموذج والتي تمثل معيار المحكم في الأساس النظري الذي يقوم عليه النموذج بين (0.68 – إلـــى 0.12) ووُجد أن المتوسط الحسابي لـــهذه الدرجـــات (0.22 -)، وحسب الأساس النظري لهذا النموذج فإنه كلما قل معيار المحكم، فإن المحكم يعتبر متساهلاً في قبوله للفقرات، ووجد الباحث أن 144 محكماً كانت درجاتهم اللوجستية سالبة؛ أي أن غالبية المحكمين اتصفوا بصفة التساهل في قبول الفقرات في هذا الأسلوب من التحكيم. وجرى إسقاط تقييمات المحكمين غير المطابقين لنموذج المحاولات ذي الحدين والبالغ عددهم 114 محكماً من ملف بيانات المحكمين، وأعيد حساب نسب الاتفاق على الفقرات من خلال تقييمات المحكمين الذين تطابقت بياناتهم الإحصائية مع هذا النموذج، وأعيد تطبيق معيار 70٪ و 80% في اعتماد الفقرات، وكذلك تقدير الخصائص السيكومترية على مقياس الاتجاهات لاستجابات الطلبة البالغ عددهم 413 طالباً وطالبة. وتبين أن 27 فقرة قد حققت معيار 70%، وأن 10 فقرات قد حققت معيار 80%.

وقد حُسب مؤشر التمييز (معامل ارتباط الفقرة بالأداة)، لكل فقرة من الفقرات المعتمد وفق معيار 70%، وتجدر الإشارة إلى أن الفقرات المعتمدة وفق معيار 80% متضمنة في تلك الفقرات ولوحظ ارتفاع معاملات التمييز في معظم الفقرات عما كانت عليه قبل إسقاط تقييمات المحكمين غير المطابقين للنموذج الثاني، وأن معاملات التمييز تتراوح بين 0.42 و 0.63 بوسيط يساوي 0.54، وتتضمن هذه الفقرات فقرة واحدة من الفقرات المضافة للمقياس بنسبة 8%، أما الفقرات الأصلية في المقياس فتبلغ نسبتها 43%. وتبين أن معامل الارتباط بين علامة الطالب على هذه الفقرات كمقياس للاتجاهات نحو الرياضيات، وعلامته في مبحث الرياضيات للفصل الدراسي الأول لعام 2008-2009 م كمؤشر على صدق المقياس بلغ 0.58 وهو ذو دلالة إحصائية عند مستوى دلالة 0.01، ويدلّ على وجود ارتباط متوسط القوة بينهما. وتبين أن معامل الثبات وفق معادلة كرونباخ ألفا= 0.87 تقريباً للأداة المكونة من 27 فقرة والتي تم اعتمادها وفق معيار 70%، وفي معيار 80% تبين أن معامل الارتباط بين علامة الطالب على هذه الفقرات كمقياس للاتجاهات نحو الرياضيات وعلامته في مبحث الرياضيات للفصل الدراسي الأول لعام 2008-2009 م كمؤشر على صدق المقياس بلغ 0.57، ويدلّ على وجود ارتباط متوسط القوة بينهما، ومعامل الثبات للأداة المكونة من 10 فقرات قد بلغ 0.80 .

ولاختبار معاملي الصدق قبل تطبيق النموذج، وبعده اُستخدم اختبار معاملي الارتباط للبيانات المرتبطة، وذلك بعد تعديل معامل الصدق الذي ظهر بعد تطبيق النموذج وإسقاط تقييمات

المحكمين غير المطابقين لنموذج المحاولات ذي الحدين، حيث بلغ معامل الصدق بعد تعديله لاختلاف التباين في استجابات الطلبة على الأداة قبل تطبيق النموذج وبعده، ما يقارب 0.69، وبلغت قيمة المختبر الإحصائي ت= 7 وبدرجات حرية 410 وهذه القيمة ذات دلالة إحصائية عند مستوى دلالة إحصائية أقل من 0.01 .

ويلخص الجدول 5 الخصائص السيكومترية المقدّرة للأداة قبل إسقاط المحكمين غير المطابقين للنموذج الثاني، وبعده في كلا المعيارين 70% و 80% .

الجدول 5

الخصائص السيكومترية للأداة قبل إسقاط تقييمات المحكمين غير المطابقين لنموذج المحاولات ذي الحدين وبعده في كلا المعيارين 70% و 80% .

	معيار 70%		معيار 80%	
	قبل	بعد	قبل	بعد
الصدق	0.55	0.58	0.565	0.57
الثبات	0.88	0.87	0.866	0.81
الوسط الحسابي	3.87	4.01	4.04	4.1
التباين	262	144	136.5	53.29
عدد فقرات الأداة	38	27	26	10

ولاختبار معاملي الصدق قبل تطبيق النموذج وبعده في معيار 80% استخدم اختبار معاملي الارتباط للبيانات المرتبطة، وذلك بعد تعديل معامل الصدق الذي ظهر بعد تطبيق النموذج وإسقاط تقييمات المحكمين غير المطابقين لنموذج المحاولات ذي الحدين، حيث بلغ معامل الصدق بعد تعديله نظراً لاختلاف التباين في استجابات الطلبة على الأداة قبل تطبيق النموذج الثاني وبعده، ما يقارب 0.74، وبلغت قيمة المختبر الإحصائي ت= 7.9 بدرجات حرية 410 وهذه القيم ذات دلالة إحصائية عند مستوى دلالة إحصائية أقل من 0.01.

وفي معيار 80% تم تعديل معامل الثبات الذي تم الحصول عليه بعد تطبيق النموذج بمعادلة سبيرمان براون حيث بلغ معامل الثبات بعد تعديله 0.92 تقريباً وبلغت قيمة المختبر الإحصائي للفرق بين معاملي الثبات (W= 1.7)، ووُجد أن هذه القيمة دالة إحصائياً عند مستوى دلالة إحصائية 0.01 حيث تجاوزت درجات الحرية 1000 لكلا البسط والمقام.

وقد عُدلّت قيم معامل الثبات الذي تم الحصول عليها بعد إسقاط تقييمات المحكمين غير المطابقين لنموذج المحاولات ذي الحدين النموذج بمعادلة سبيرمان براون (Spearman-Brown Formula) وذلك لاختلاف عدد الفقرات بين المقياس قبل تطبيق النموذج وبعد تطبيقه، وكذلك من أجل مقارنة هذا المعامل بمعامل الثبات الذي تم الحصول عليه قبل تطبيق نموذج المحاولات ذي الحدين في كلا المعيارين 70% ومعيار 80%، وبلغ معامل الثبات بعد تعديله 0.91 تقريباً، وتم اختبار هذا الفرق من خلال اختبار معاملي الثبات للبيانات المرتبطة W، حيث بلغت قيمة المختبر الإحصائي للفرق بين معاملي الثبات (W= 1.4)، ووُجد أن هذه القيمة دالة إحصائياً عند مستوى دلالة إحصائية 0.01 حيث تجاوزت درجات الحرية 1000 لكلا البسط والمقام .

أظهرت نتائج تحليل نسب الاتفاق ،على الفقرات قبل تحليل تقديرات المحكمين بهذا الأسلوب أن 38 فقرة فقط قد حققت معيار 70٪ ، و26 فقرة في معيار 80%، وهو عدد قليل إذا تمّ مقارنته بعدد فقرات المقياس الأصلي، وربما يعود ذلك إلى أسباب منها : وجود ضعف في فقرات المقياس في مدى ملاءمتها لقياس الاتجاه نحو الرياضيات، أو وجود ضعف في مضمونها، أو في صياغتها اللغوية، أو بسبب عدم وجود معايير واضحة ومحددة لدى المحكمين في تقييم فقرات المقياس، وقد يستخدم بعضهم معايير مناسبة ولكن دون إدراكهم لبناء التقييم، أو عدم إدراكهم لعملية تقييم الفقرات وغرض تقييمها، أو عدم إدراكهم للفقرات نفسها وأبعادها، مّما أدى إلى استجابة بعض المحكمين بعدم الموافقة على الفقرات التي لم يستطيعوا فهمها، وربما لجأ بعض المحكمين إلى عشوائية التحكيم لبعض الفقرات التي لم يفهموها، وهذا قلل من عدد الفقرات التي تحقق المعيارين .

وأظهرت نتائج تحليل تقييمات المحكمين باستخدام نموذج المحاولات ذي الحدين أن 114 محكماً لم تطابق بياناتهم الإحصائية هذا النموذج، وهذه نسبة عالية جداً وإن دلّت على شيء فإنها تدلّ على وجود الأخطاء والعشوائية العالية في عملية تقييم الفقرات من قبل المحكمين بهذا الأسلوب، فربما لم يدرك الغالبية من المحكمين عملية التحكيم بهذا الأسلوب فظهرت العشوائية في إعطاء الأرقام، وعدم وضوح المعايير في أذهانهم واختلافها من محكم إلى آخر، وهذا ما أشار إليه منبي (Munby, 1982) في قضية اختلاف فهم سياق التحكيم من شخص لآخر. وربما يكون التسرع في عملية التحكيم قد ساهم في زيادة التقييمات غير الملائمة وعدد المحكمين غير المطابقين لهذا النموذج، وهذه النتيجة تعطي مؤشراً إلى أنه لا ينصح باستخدام هذا الأسلوب من التحكيم عندما يكون عدد المحكمين قليلاً.

وأظهرت نتائج تفحص تدريجات المحكمين الناتجة من هذا النموذج أن 144 محكماً كانت درجاتهم اللوجستية سالبة؛ أي أن غالبية المحكمين اتصفوا بصفة التساهل في قبول الفقرات في هذا الأسلوب من التحكيم، ربما يعود ذلك إلى اختلاف فهم المحكم للأسلوبين، أو إلى أن رأي المحكم غير مباشر في الأسلوب الثاني وهذه النتيجة تفسر الفرق في عدد الفقرات التي تم اعتمادها بين الأسلوبين، وهذا يعطي مؤشراً إلى استخدام هذا الأسلوب عندما يرغب الباحث في زيادة عدد فقرات المقياس التي يتفق المحكمون عليها.

وعندما أسقطت تقييمات المحكمين غير المطابقين للنموذج، وأعيد تطبيق المعيارين على نسب الاتفاق على الفقرات، تبين أن التقييمات غير الملائمة والتي تم إسقاطها قد قلّلت من نسب الاتفاق على الفقرات التي تم اعتمادها، وزادت في بعضها، حيث تقلص عدد الفقرات إلى 27 في معيار 70% و10 فقرات في معيار 80% ؛ أي أن 11 فقرة انخفضت نسب الاتفاق عليها إلى أقل من 70٪، و16 فقرة إلى دون 80%، وربما يعود ذلك إلى وجود بعض المحكمين الذين وافقوا على فقرات ليست على درجة عالية من الجودة في قياس الاتجاه نتيجة لغموض فيها، أو التسرع في الحكم عليها، أو التقييم العشوائي لها.

وأظهرت عملية إسقاط التقييمات غير الملائمة للمحكمين غير المطابقين للنموذج الثاني إلى ارتفاع معاملات تمييز الفقرات التي تم اعتمادها بالمعيارين، وخرجت بعض الفقرات التي كانت معاملات تمييزها قليلة مقارنةً ببقية الفقرات، وأظهرت أيضاً وجود تحسن ذي دلالة إحصائية في الخصائص السيكومترية للمقياس على الفقرات التي تم اعتمادها في كلا المعيارين، وبسبب وجود حجم عينة كبير، وارتفاع معامل الارتباط بين علامة الطالب على المقياس قبل عملية إسقاط المحكمين غير المطابقين للنموذج وبعدها؛ كان وراء كشف تلك الفروق، وهذه النتيجة تؤكد أيضاً على أهمية استخدام نماذج النظرية الحديثة في القياس في تحسين عملية التقييم للمحكمين على مقاييس الاتجاهات، والتي أشارت إليه نتائج الدراسات السابقة كدراسة انجلهارد و اندرسون (Engelhard & Anderson, 1998) ودراسة بني عطا (2005) التي استخدمت هذا الأسلوب في تحسين جودة التحكيم.

وفي الخلاصة فإن هذه النتائج تؤكد على أهمية استخدام هذا الأسلوب في البحوث التربوية، وفي مجال التحقق من صدق المقاييس باستخدام تقييمات المحكمين، حيث تعمل على تحسين جودة تقديرات المحكمين والخصائص السيكومترية للمقاييس.

التوصيات: نظراً لما أظهرته نتائج هذه الدراسة من تحسن في الخصائص السيكومترية لفقرات مقياس الاتجاهات، جراء فحص تقييمات المحكمين واستبعاد غير المطابقين لنموذج المحاولات

ذي الحدين، فإن الباحث يوصي باستخدامه لفحص تقييمات المحكمين مستقبلاً في مقاييس الاتجاهات. ويوصي الباحث بالبحث في امكانية تعميم استخدامه على مقاييس أخرى غير مقاييس الاتجاهات.

المراجع

المراجع العربية:

1. أبو زينة، فريد وكيلاني، عبدالله. (1980). أثر التخصص والمستوى التعليمي على الاتجاهات نحو الرياضيات عند فئات من المعلمين والطلبة في الأردن. دراسات:العلوم الإنسانية، 7(1) .
2. بني عطا، زايد. (2005). استخدام نموذج ذي الحدين لفحص تقديرات المحكمين لدرجة القطع لاختبار بدلالة عدد بدائل الفقرة. أطروحة دكتوراه غير منشورة، جامعة اليرموك.
3. الخليلي، خليل. (1989). الاتجاهات نحو الفيزياء بنيتها وقياسها.مجلة أبحاث اليرموك"سلسلة العلوم الإنسانية والاجتماعية،"5 (1).
4. الصمادي، احمد والسرطاوي، عبد العزيز. (1996). تكوين مقياس اتجاهات طلاب الجامعة نحو المعاقين.مجلة العلوم الاجتماعية،24 (2)
5. المقرن، عبد العزيز والورع، مأمون.(2003).آلية تحكيم مشروعات التصميم المعماري "تقويم درجات المحكمين-دراسة مقارنة".دراسات،العلوم الهندسية،30 (2).
6. طه،فرج. (2006). أصول علم النفس الحديث. الرياض: دار الزهراء للنشر والتوزيع.
7. علام، صلاح الدين. (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.
8. عودة، احمد .(1993). القياس والتقويم في العملية التدريسية. إربد، الأردن: دار الأمل.
9. عودة، احمد. (1990). تطوير مقياس اتجاهات المعلمين نحو الامتحانات المدرسية (نموذج مقترح) .مجلة كلية التربية – جامعة الإمارات ،53,8-71

المراجع الأجنبية:

1. Alderson, J.C. (1993). Judgments in language testing. In D. Douglas & C. Chapelle (Eds.). A new decade of language testing. Alexandria, VA, USA: TESOL.
2. Allen, M. &Yen, W.(1979). Introduction to Measurement Theory. California :Brooks/Cole Publishing Company Monterey.
3. Alsawalmeh,Y.M.,&Feldt,L.S.(1994). A Modification of Feldt's Test of the equality of tow Dependent Alpha Coefficints. Psychometrika . 59(1), 49-57.
4. Alsawalmeh,Y.M.,&Feldt,L.S.(2000). ATest the Equality of two Related Alpha Coefficients Adjusted by Spearman-Brown Formula.Applied Psychological Measurement, 24 (2), 163-172.
5. Bratt, M. (1984). Further comment on the validity studies of attitude measures in science education. Journal of Research in Science Teaching,21(9), 951-952
6. Brown, J. D. (1996). Testing in Language Programs. Prentice Hall Regents. Prentice-Hall, Inc.
7. Cronbach L. J. (1995). Giving method variance its due. In P. E. Shrout&S. T. Fiske (Eds.), Personality research, methods, and theory: A Festschrift in honor of Donald Fiske (145–157). Hillsdale, NJ:Lawrence Erlbaum Associates, Inc.
8. Draney K. & Wilson M.(2008). A LLTM approach to the examination of teachers' ratings of classroom assessment tasks. Psychology Science Quarterly, 50 (3), 417-432
9. Engelhard,J.&Anderson,D.(1998).A binomial Trials Model for Examining the ratings of standard-setting Judjes.Applied Measurement in Education,.11 (2), 209-230.
10. Hajipour, J.(2003). An approach to the validation of judgments in language testing.Kyoto, Japan: Kyoto Institute of Technology.
11. Linacer,J. M & Wright,B.D.(1994).Auser's guid to FACETES: Rash measurement computer program.Chicago:MESA Press
12. Lucas, A.H. (1975). Hidden assumption in measures of knowledge about science and scientists. Science Education, 59 (4), 481-485.
13. Munby, H. (1982). The impropriety of Panel of judges validation in science attitude scale: a research comment. Journal of Research in Science Teaching, 19 (7), 617-619
14. Neter,J.,Wasserman,W.&Whitmore,G.A.(1982).Applied Statistics.(2nd ed).Bosten:allyn and Bacon,INC.
15. Schibeci, R.A. (1984). Attitudes to science: An update. Studies in Science Education, 11(1), 26-59

16. Shaw,M.E and Wright,J.M.(1967). scales for Measurement of Attitudes**.**New York, Mc :Graw Hill.
17. Shrigley, R.L and Koballa, T.R.Jr. (1984). Attitude measurement: judging the emotional entinisity of likert-type science attitude statement. Journal of Research in Science Teaching, 21(2), 111-118
18. Wexely, N.,& Youtz ,A.(1985).Rater beliefs about others:their effects on rating and rater accuracy.Journal of occupational psychology ,58 (1),256-275
19. Zhu,w.;Ennis,C.D.&Chen,A.(1998). Many-Faceted Rash Modeling Expert Judgment in Test Development. Measurement in Physical Education and Exercise Science,22(2),, 21-3