



كلية التربية

كلية معتمدة من الهيئة القومية لضمان جودة التعليم

إدارة: البحوث والنشر العلمي (المجلة العلمية)

=====

الأخطاء الشائعة في إجراءات التحقق من ثبات وصدق أدوات القياس المستخدمة في البحوث التربوية العربية

إعداد

د/ سومية شكري محمد محمود

مدرس بقسم علم النفس التربوي

كلية التربية بجامعة المنيا (مصر)، وجامعة الإمام عبد الرحمن بن فيصل (السعودية)

﴿ المجلد الخامس والثلاثون - العدد السابع - يوليو ٢٠١٩ م ﴾

http://www.aun.edu.eg/faculty_education/arabic

المخلص:

هدفت الدراسة الحالية إلى تحديد الأخطاء الشائعة في إجراءات التحقق من ثبات وصدق أدوات القياس المستخدمة في البحوث التربوية العربية، حيث تم حصر طرق حساب ثبات وصدق أدوات القياس المستخدمة في عينة من ٧٢ بحثاً منشوراً في مجلة الجامعة الإسلامية للدراسات التربوية والنفسية، خلال عام ٢٠١٢ وعام ٢٠١٦، وبلغ عدد أدوات القياس المستخدمة في هذه البحوث ٩٢ أداة، كان ٨٥% منها من إعداد الباحثين، وقد تم رصد ١٢ خطأ متكرراً في إجراءات التحقق من الثبات، و ٩ أخطاء متكررة في إجراءات حساب الصدق، كما اعتمد الباحثون بشكل أساسي على طريقة الاتساق الداخلي لحساب ثبات أدوات القياس، فبلغ مجموع نسبة استخدام التجزئة النصفية أو معادلة ألفا أو كيودر ريتشاردسون منفردة أو مع طرق أخرى ٨٧% تقريباً، وبلغت نسبة استخدام طريقة إعادة التطبيق منفردة أو مع طرق أخرى ٣٢%، كما كانت طريقة الصدق المرتبط بالمحتوى هي الأكثر استخداماً، إذ بلغت نسبة استخدامها منفردة أو مع طرق أخرى ٧٧%، دون الاعتماد على أي طريقة كمية للتقدير، كما استُخدمت طريقة الاتساق الداخلي كطريقة لحساب الصدق في أكثر من نصف البحوث، حيث بلغت نسبة استخدام هذه الطريقة منفردة أو مع طرق أخرى ٥٢%، وقد وضّحت الدراسة الفرق بين استخدام طريقة الاتساق الداخلي لحساب الصدق، واستخدامها لحساب الثبات. وأوصت الدراسة بإعداد مقياس علمي لتقدير دقة إجراءات التحقق من صدق وثبات أدوات القياس في البحوث التربوية.

الكلمات المفتاحية: الثبات - ثبات الاتساق الداخلي - الصدق - صدق التكوين الفرضي.

Abstract:

This study aimed at identifying the Common mistakes in reliability and validity procedures of the measuring tools used in the Arabic educational research. The procedures of reliability and validity of the measuring tools used in a sample of 72 papers published in 2012 and 2016 in the Journal of the Islamic University for Educational and Psychological Studies were evaluated. The number of measuring tools used in these researches was 92 tools 85% were created by the researchers. Twelve frequent mistakes were detected in reliability procedures and nine repeated mistakes were detected in validity procedures. The most frequent method of reliability was the internal consistency including split-half method, Alpha coefficient and Kudar – Richardson. The most frequent method of validity was the content related validity with the percentage of use alone or with other methods 77% without applying any quantitative method of estimation the content validity index. More than half of the researches used the method of "internal consistency" as a method of measuring validity. The study clarified the difference between using the "internal consistency" method to detect validity and using it to detect reliability. The study recommended creating an analytical rubric to assess the accuracy of reliability and validity procedures.

Key words: Reliability – Internal consistency – Validity – construct validity.

مقدمة:

يمثل البحث التربوي وسيلة مهمة للنهوض بالمجتمعات؛ إذ يُسهم في حل المشكلات المتعلقة بتربية الإنسان وبنائه، وفي ضوء نتائجه يتم اتخاذ القرارات التي تؤثر في عملية التعليم والتعلم. وتتوقف دقة نتائج البحوث التربوية على عدة عوامل، منها التحقق من ثبات وصدق أدوات القياس المستخدمة، والمطلع على البحوث التربوية والنفسية يجد أن معظم الباحثين يفضلون إعداد أدوات قياس خاصة ببحثهم، بغض النظر عن مدى توفر هذه الأدوات في الأدب النفسي والتربوي، وذلك نتيجة الاعتقاد السائد بأن قيام الباحث بإعداد أدواته الخاصة يدل على مهاراته البحثية، مع عدم الوعي الحقيقي بصعوبة الوصول إلى المقاييس ذات المستوى المقبول من الصدق والثبات، التي يتطلب إعدادها فريقاً مديراً من الباحثين المتخصصين، وعينات كبيرة من المفحوصين، كما يتطلب جمع عدة أدلة للتحقق من ثباتها وصدقها.

مما يثير التساؤل حول مدى دقة الإجراءات التي يتبعها الباحثون للتحقق من صدق وثبات أدوات القياس المستخدمة في البحوث التربوية، فاستخدام أدوات غير مناسبة يؤدي إلى جمع بيانات خاطئة مما يقود إلى نتائج مضللة، وقد أشار عزت (٢٠١٦: ٨-٩) إلى أن هناك العديد من الأخطاء التي يقع فيها الباحثون والمناقشون تتعلق بالخصائص السيكمترية لأدوات البحث. وأشار باهي (٢٠١١: ١٤٠٦) إلى أن هناك أخطاء كثيرة في استخدام المصطلحات العلمية الخاصة بكل من الثبات والصدق، حيث تم استتساخ العديد من المصطلحات التي لا أساس لها من الصحة، مما يبرز الحاجة إلى محاولة وصف هذه الحالات، وتحديد مدى انتشارها، وهل تمثل ظاهرة تتطلب المواجهة أم أنها مجرد حالات فردية؟.

مشكلة الدراسة:

هناك مجموعة من طرق حساب ثبات أدوات القياس وصدقها، محددة ومعروفة في الأدب السيكمترية، وبمطالعة البحوث التربوية العربية المنشورة في الدوريات العلمية المحكمة، نجد أن هناك خلطاً لدى العديد من الباحثين بين مفهوم الثبات والصدق، فنجدهم يخلطون بين إجراءات التحقق من ثبات الاتساق الداخلي وإجراءات التحقق من صدق التكوين الفرضي لأدوات القياس المستخدمة في البحث، كما أن هناك أخطاء متعلقة بدقة إجراءات تقدير كل من صدق المحتوى والصدق التمييزي، وكذلك هناك ندرة في البحوث التي تعتمد على مقاييس مقيّنة، فالجميع - بغض النظر عن تخصصه - يقوم بإعداد أدوات قياس خاصة ببحثه. وهكذا تتعدد الأخطاء في بحوثنا التربوية، مما يلفت الانتباه إلى رصد حجمها، وهل هي أخطاء نادرة أم أنها تمثل ظاهرة تتطلب المواجهة والعلاج؟ وبناءً على ما سبق يمكن صياغة مشكلة الدراسة في التساؤلات التالية:

١. ما الطرق الأكثر استخداماً للتحقق من ثبات أدوات القياس في البحوث التربوية؟
٢. ما الطرق الأكثر استخداماً للتحقق من صدق أدوات القياس في البحوث التربوية؟
٣. ما نسبة البحوث التي تعتمد على أدوات قياس مقننة أو منشورة؟
٤. ما الأخطاء الشائعة في إجراءات حساب ثبات أدوات القياس في البحوث التربوية؟
٥. ما الأخطاء الشائعة في إجراءات التحقق من صدق أدوات القياس في البحوث التربوية؟

أهداف الدراسة:

١. تحديد وتصنيف الأخطاء الشائعة في إجراءات حساب ثبات وصدق أدوات القياس في البحوث التربوية.
٢. تحديد نسبة الأخطاء الشائعة في إجراءات حساب ثبات وصدق أدوات القياس في البحوث التربوية.
٣. مناقشة المفاهيم المرتبطة بطرق حساب ثبات وصدق أدوات القياس في البحث التربوي، والتي تتداخل لدى طلاب الدراسات العليا والباحثين.

أهمية الدراسة:

الأهمية النظرية:

تتناقش الدراسة بعض المفاهيم المهمة في القياس النفسي والتربوي، وبخاصة تلك المتعلقة بثبات الاتساق الداخلي وصدق التكوين الفرضي، وهما مفهومان كثيراً ما يحدث بينهما تداخل وخطأ لدى العديد من طلاب الدراسات العليا والباحثين.

الأهمية التطبيقية:

١. تُسهم الدراسة في تحسين جودة البحث التربوي العربي من خلال تحسين مستوى ثبات وصدق أدوات القياس المستخدمة فيه.
٢. تُسهم الدراسة في زيادة وعي الباحثين بالإجراءات الدقيقة لحساب ثبات وصدق أدوات القياس، والتي يتوقف عليها الثقة في نتائج البحوث.

حدود الدراسة:

الحد الموضوعي: تتحدد نتائج الدراسة بطرق حساب ثبات وصدق أدوات القياس في ضوء النظرية الكلاسيكية في القياس.

الحد المكاني والزمني: تتحدد نتائج الدراسة بالمجلة المختصة التي تم اختيار عينة البحوث منها، وهي مجلة الجامعة الإسلامية للدراسات التربوية والنفسية، خلال عامي: ٢٠١٢م، ٢٠١٦م.

التعريف الإجرائي لمتغيرات الدراسة:

١. ثبات أداة القياس: يُقصد به ثبات درجات أداة القياس، أي: مدى خلو تباين درجات عينة محددة من المفحوصين عليها من تباين الخطأ العشوائي.
٢. صدق أداة القياس: هو صفة فريدة unitary لا تتعلق بالأداة في حد ذاتها، وإنما تتعلق بتفسير الدرجات المستمدة منها، وتعني مدى صحة القرارات التي يتم اتخاذها في ضوء نتائج الأداة.

الإطار النظري والدراسات السابقة:

يُعرف البحث التربوي بأنه تطبيق المنهج العلمي في دراسة المشكلات التربوية، ومن ثم فهناك مجموعة من المعايير والإجراءات التي يحددها المنهج العلمي حتى يحقق البحث التربوي الهدف منه (Oluwataya, 2012: 391)، وتتعدد أدوات جمع المعلومات في البحث التربوي بتنوع أهدافه، فمنها الاختبارات والمقاييس النفسية، واختبارات التحصيل، والاستبانة، والمقابلة، والملاحظة، واستمارة تحليل المحتوى، وأياً كانت الأداة التي يستخدمها الباحث فعليه أن يتحقق من معايير بنائها وصلاحياتها، ومن أهم هذه المعايير ثبات الأداة وصدقها (عطية، ٢٠٠٩: ١٠٧-١٠٨، أبو علام، ٢٠٠٦: ٤٤٧)؛ وذلك لضمان صحة ودقة البيانات التي يتم جمعها، فالقصور في التحقق من ثبات وصدق الأدوات يؤدي إلى جمع بيانات خاطئة، ومن ثم الوصول إلى نتائج مضللة. وقد بينت دراسة (Fong, 1994) أن العديد من البحوث التي أجريت في مجال التربية وعلم النفس تفقد الواقعية، ويصعب تعميم نتائجها، وتنفذ إلى نتائج مضللة؛ بسبب الأخطاء المنهجية في تصميمها وإجراءها، وتكمن خطورة ذلك في طبيعة المشكلات التي يتناولها البحث التربوي، الذي يُعنى بالإنسان بالدرجة الأولى، سواء أكان معلماً أم متعلماً، فقد تؤدي النتائج المضللة إلى توجيه العملية التربوية والتعليمية في اتجاهات خاطئة، لذا بات من الضروري التحقق من دقة إجراءات ثبات وصدق أدوات القياس المستخدمة في البحوث التربوية؛ لما له من دور في تحسين جودة هذه البحوث وزيادة الثقة في نتائجها.

ويشير مفهوم الثبات إلى مدى اتساق نتائج المقياس، أي: مدى إمكانية الحصول على نفس النتائج إذا قمنا بتطبيق المقياس نفسه عدة مرات على نفس المفحوصين (أبو علام، ٢٠٠٦: ٤٤٨، ميخائيل، ٢٠١٥: ٩٥)، ويقابله مفهوم الدقة Precise (عبيدات، ٢٠١٠: ٣٣٠).

ويُعرف ثبات المقياس إحصائياً بأنه نسبة تباين الدرجات الحقيقية إلى تباين الدرجات الملاحظة، حيث يتضمن تباين الدرجات الملاحظة في القياس النفسي عادة مقداراً من تباين الخطأ، وتتعدد مصادر أخطاء القياس، فمنها ما يتعلق بالمقياس نفسه، كالخطأ في أسلوب معاينة الفقرات للنطاق السلوكي المستهدف بالقياس، ومنها ما يتعلق بإجراءات التطبيق والتصحيح، كالأخطاء في تعليمات الاختبار الشفهية، أو أخطاء الطباعة، أو عدم مناسبة البيئة الفيزيائية أو النفسية للتطبيق، ومنها مصادر تتعلق بالمختبرين، كميلهم للتخمين أو العش، بالإضافة لتأثير المرغوبية الاجتماعية على استجاباتهم (علام، ٢٠١١: ١٣١-١٤٣).

وتشير أدبيات القياس والتقويم التربوي إلى طرق متعددة لتقدير قيم معامل الثبات، كإعادة التطبيق، والصور المتكافئة، وإعادة التطبيق بصور متكافئة، والاتساق الداخلي، وثبات تقديرات المحكمين. ولكل طريقة مجموعة من الإجراءات والضوابط يضيق المقام بسردها هنا، والذي ينبغي التركيز عليه أن هذه الطرق تختلف باختلاف مصدر الأخطاء العشوائية التي نود تقدير تأثيرها في الدرجات الملاحظة، وهذا يعتمد على طبيعة الاختبار، وفيما تستخدم درجاته؟، لذلك لابد أن يكون هناك مبرر منطقي لاختيار طريقة حساب الثبات، بالإضافة لضرورة الالتزام بالإجراءات العلمية لاستخدام الطريقة المختارة (أبو علام، ٢٠٠٦: ٤٦٨، علام، ٢٠١١: ١٤٤-١٦٨، أبو دنيا، ٢٠١٨: ٢٥٢-٢٥٧، ملحم، ٢٠١٨: ٣٠٢-٣١٣).

أما الصدق فيُعد أهم خاصية من خصائص القياس، ويشير إلى صحة الاستدلالات التي نتوصل إليها من درجات المقاييس، من حيث فائدتها ومعناها (أبو علام، ٢٠٠٦: ٤٤٧)، ويكون المقياس صادقاً عندما يقيس ما أُعد لقياسه، ويقابله مفهوم الصحة *Accurate* (عودة، ٢٠١٠: ٣٣٠)، فهو يعني مدى تلبية المقياس للأغراض والاستعمالات الخاصة التي صُمم من أجلها. ويُعد الصدق صفة نسبية؛ إذ لا يوجد مقياس صالح في كل المواقف، إنما يتوقف صدق المقياس على وظيفته، والغرض منه، فالاختبار الذي يتم إعداده بغرض التنبؤ قد لا يعمل بنفس درجة الصدق في التشخيص (ميخائيل، ٢٠١٥: ٨٦-٨٧).

وعلى الرغم من ظهور مصطلحات عديدة للتعبير عن طرق حساب الصدق، والتي أُصطلح على تسميتها بـ"أنواع الصدق"، فإنه من الواجب هنا التأكيد على أنه لا توجد أنواع مختلفة للصدق؛ فالصدق مفهوم مجرد، وفريد، ومتعدد الأبعاد، لذا تتعدد طرق التحقق منه، ولا تُعد هذه الطرق بدائل لبعضها البعض، لكن كلاً منها يهتم بجمع نوع معين من الأدلة يبرر جانباً من جوانب الصدق المختلفة (AERA, APA & NCME, 2014: 11).

وعلى الرغم من التصنيفات العديدة لطرق حساب الصدق، فإن التصنيف الأكثر شيوعاً في أدبيات ومراجع القياس والتقويم العربية يضعها في ثلاث فئات رئيسية، هي:

١. ما يرتبط بالمحتوى *Content Validity*، ويُقصد به مدى تمثيل فقرات المقياس تمثيلاً سليماً للمجال المستهدف بالقياس، ولذلك فهو يتطلب تحديد هذا المجال تحديداً واضحاً، وتحديد عناصره، وبناء مجموعة ممثلة من الفقرات، بحيث يتم مقارنة الفقرات بالمواصفات التي تحدد المجال. ولصدق المحتوى بُعدان: الأول هو الصدق الظاهري، الذي يرتبط بمدى وضوح المحتوى بالنسبة للمفحوصين، من حيث: سهولة القراءة، واتساق الأسلوب، والتنسيق، ووضوح اللغة المستخدمة، وإلى أي مدى تبدو الفقرات مرتبطة بهدف الأداة من وجهة نظر المفحوصين. أما البعد الثاني فهو الصدق العيني، الذي يُعنى بتقدير مدى تمثيل الفقرات للمجال المحدد بجدول المواصفات من وجهة نظر مجموعة من الخبراء.

٢. ما يرتبط بمحك Criterion-Related Validity ، ويتم من خلال إيجاد معامل الارتباط بين درجات المقياس ودرجات مقياس آخر يقيس السمة نفسها وتم التحقق من صدقه، ويُعرف الأخير بالمحك، وقد يكون المحك أداءً مستقبلياً للمفحوصين، فحينئذ يُسمى الصدق بالصدق التنبؤي، وقد يتم تطبيق المقياس المحك بالتزامن مع المقياس محل الدراسة، وحينئذ يُسمى الصدق بالصدق التلازمي.

٣. ما يرتبط بالبناء أو التكوين Construct Validity ، وتُستخدم هذه الطريقة عندما يتم تصميم المقياس في ضوء نظرية علمية، أو افتراضات محددة، ويكون الهدف تحديد ما إذا كان المقياس يمثل النظرية أم لا؟، وتعتمد هذه الطريقة على تحديد مدى الارتباط بين الجوانب التي يقيسها المقياس وبين النظرية أو الافتراضات. ويُعد صدق البناء مفهوماً شاملاً يتضمن سائر أنواع الصدق، ويتطلب الاستنتاجات المنطقية والوسائل التجريبية والإحصائية. وللكشف عن صدق البناء هناك ثلاث خطوات، وأولها: تعريف الإطار النظري للسمة المستهدفة بالمقياس. وثانيها: اشتقاق فروض حول نتائج المقياس في ضوء الإطار النظري المحدد. وثالثها: اختبار هذه الفروض منطقياً أو تجريبياً. أما إجراءات التحقق من صدق البناء فتتمثل في الآتي:

أ- دراسة العلاقة بالمقاييس الأخرى، وذلك من خلال دراسة الارتباط بين نتائج المقياس ونتائج مقاييس أخرى تقيس مكوناً يرتبط نظرياً مع المكون المسهدف بالمقياس، أو بدراسة مدى تمايز نتائج المقياس محل الدراسة عن نتائج مقياس آخر يقيس مكوناً لا يرتبط مع المكون المسهدف بالمقياس، أو بإجراء التحليل العاملي لنتائج المقياس؛ لتحديد نسبة التباين التي تفسرها العوامل المكونة للمقياس بالنسبة للتباين الكلي، وفحص محتوى العوامل والاستدلال من خلالها على طبيعة المكون الفرضي.

ب- الدراسات التجريبية، وتتم باختبار تغير درجات المقياس عند تقديم أنواع معينة من المعالجات التجريبية.

ج- مقارنة درجات مجموعات متميزة في المكون المسهدف بالمقياس.

د- التحليل الداخلي للاختبار، ويتم من خلال جمع معلومات عن محتوى الاختبار نفسه، والعمليات المستخدمة في الاستجابة على فقراته، والارتباطات بين فقرات الاختبار، ودراسة تجانس المحتوى، والتأكد من أنه يقيس سمة أحادية البعد، ولذلك يمكن استخدام معامل ألفا أو كيوودر - ريتشاردسون كمؤشر للتجانس أو إحادية البعد.

(أبو علام، ٢٠٠٦: ٤٥١-٤٦٠، عبيدات، ٢٠١٠: ٣٦٧-٣٨٦، علام، ٢٠١١: ١٨٩-

٢٢٨، ميخائيل، ٢٠١٥: ٨٧-٩٣، ملحم، ٢٠١٨: ٣١٨-٣٢١).

وهناك العديد من الأساليب أو المؤشرات الإحصائية والمنطقية التي تُستخدم في الكشف عن صدق البناء، وقد يُستخدم أكثر من مؤشر للمقياس الواحد؛ نظراً لتعدد الفرضيات التي يفرضها الإطار النظري، ومنها بالإضافة للتحليل العاملي، مؤشرات الثبات، سواء ثبات الاتساق الداخلي باستخدام التجزئة النصفية وحساب معامل ألفا أو كيوذر - ريتشاردسون، أو ثبات الإعادة للكشف عن ثبات السمة وعلاقته بثبات المقياس، حيث تُعرف السمة السيكلوجية بأنها مجموعة من السلوكيات التي تميل إلى الحدوث معاً، وبالتالي فإن قوة الارتباط بين الفقرات المعدة لقياس السمة تُعد مؤشراً أولياً لصدق البناء، وكذلك عندما يكون هناك افتراض بأن السمة المقاسة تتغير مع مرور الزمن، فإن معامل ثبات التجزئة النصفية ومعامل ثبات الإعادة يُستخدمان معاً كدليل على صدق البناء، فإذا كان معامل ثبات التجزئة النصفية مرتفعاً، وكان معامل ثبات الإعادة منخفضاً، فإن ذلك يعني بدرجة مقبولة من الثقة أن المقياس صادق في قياس هذه السمة (عبيدات، ٢٠١٠: ٣٥٦، ٣٨١)

ومن ثم يُعد الاتساق الداخلي مؤشراً ضرورياً، لكنه غير كاف لصدق البناء؛ لأنه يعكس مدى خلو المقياس من أخطاء المعاينة، فإذا افترضنا سمة أحادية البعد، فهناك عدد كبير وغير محدود من المؤشرات السلوكية الدالة على هذه السمة، وتمثل الفقرات الموجودة في المقياس عينة منها، وعندما نثبت أن هناك تجانساً بين هذه الفقرات، فهذا يدل على أننا لم ندرج في المقياس مؤشرات تعبر عن سمة أخرى، غير أن ذلك لا يثبت بأي شكل أن عينة الفقرات كافية، فقد تكون الفقرات متجانسة، ولكن هناك العديد من المؤشرات الأساسية الدالة على السمة خارج المقياس، لذا فالاتساق الداخلي يظل مؤشراً أولياً يحتاج لمزيد من الأدلة لإثبات صدق البناء عندما يكون هناك افتراض بأن السمة المستهدفة بالمقياس سمة أحادية البعد.

وتُعد أحادية البعد صفة مرغوبة في القياس النفسي؛ نظراً لصعوبة تفسير الدرجة الكلية على المقاييس متعددة الأبعاد، لذا يتم تحليل السمة متعددة الأبعاد إلى أبعاد أحادية؛ ليكون هناك معنى للدرجة على كل بعد، فالقياس النفسي الجيد هو الذي يسمح بتقدير قدرة واحدة للممتحن في كل مرة، ولا يسمح بإسهام مقدرتين في درجة قياس واحدة، ففي المسائل اللفظية في الرياضيات يجب أن يبحث معد الفقرات عن طرق بديلة لتقديم الفقرات، بحيث لا يستمر الأفراد مرتفعو القدرة اللفظية متميزين في الحل (Sijtsma & Molenaar, 2002: 18-19, Bond & Fox, 2007: 34).

ومن الضروري أن يهدف الاختبار إلى قياس قدرة واحدة عندما تُستعمل الدرجة الكلية وحدها محكاً للتقييم (Sheng & Wikle, 2007: 900)، وجدير بالذكر أنه من الصعب أن يتحقق افتراض أحادية البعد عملياً بشكل تام (Harris & Kolen, 1986: 37, Wainer & Mislevy, 2000: 91)؛ نظراً لتعدد العوامل المعرفية والشخصية، وتداخلها مع تلك المتعلقة بإجراءات تطبيق الاختبار، التي دائماً ما تؤثر على الأداء في الاختبار، أو في بعض نطاقاته على الأقل (Hambleton et al., 1991: 9)، وتلك العوامل تكون غير معروفة، ويصعب التحكم فيها (Osterlind, 1992: 47)، لذلك فمن الأفضل اعتبار أحادية البعد متغيراً متصلاً، وعدم اعتبارها متغيراً ثنائياً، إما أن يتحقق أو لا (Blais & Laurier, 1995: 88).

ومن الجدير بالذكر أننا عندما نستخدم الاتساق الداخلي كمؤشر أولي على صدق البناء، يجب أن نستخدم معامل ألفا أو معامل كيودر- ريتشاردسون كإحصاءة تعبر عن متوسط معاملات الارتباط بين درجات جميع الفقرات، حيث يعد الصدق صفة لنتائج المقياس ككل، في حين أن الاعتماد على الارتباط بين درجة الفقرة ودرجة البعد - وهو الإجراء السائد في البحوث التربوية العربية - يقدم معلومات خاصة بالفقرة، وهو معامل تمييز الفقرة، حيث إن معامل تمييز الفقرة هو معامل الارتباط بين درجة الفقرة والدرجة الكلية على المقياس بعد تحرير الدرجة الكلية من درجة الفقرة (عبيدات، ٢٠١٠: ٢٧٩-٢٨٠، علام، ٢٠١١: ٢٨٠، Jeeva, 2016: 51)، ويُستخدم معامل تمييز الفقرة كمؤشر لصدقها، ومن ثم يتم استبعاد الفقرات منخفضة التمييز (Thorndike, 1982: 53).

ويلاحظ مما سبق أن هناك تداخلاً بين إجراءات طرق التحقق من صدق البناء والطرق الأخرى من الصدق، لذا فقد قدمت معايير القياس النفسي - التي أعدتها الجمعية الأمريكية لعلم النفس ورابطة البحث التربوي الأمريكية والمجلس الوطني للقياس في التربية - تصنيفاً لطرق التحقق من صدق المقياس يعالج هذا التداخل، فأشارت إلى أن هناك خمسة أنواع من الأدلة تدعم جوانب الصدق المختلفة، وهي كالتالي:

١. الأدلة القائمة على المحتوى، وتعتمد على التحليلات المنطقية، وتقييم الخبراء لمحتوى المقياس، الذي يتضمن: الفقرات، والمهام، والصياغة، والإخراج، والحكم على كفاية الفقرات، ووضوحها، والعلاقة بينها وبين السمة المستهدفة بالقياس، كما يمكن أن تتضمن مراجعة الخبراء فحص التحيز للنوع أو الثقافة أو العمر.
٢. الأدلة القائمة على عمليات الاستجابة، وتعني تحديد مدى تلاؤم طريقة الاستجابات المطلوبة من المفحوصين مع السمة المستهدفة بالقياس، وتتضمن ملاحظة المفحوصين أثناء الأداء على المقياس، وإجراء مقابلات معهم لتحديد أسباب إعطاء إجابات محددة للأسئلة، وكذلك فحص الطرق التي يستخدمها المحكمون أو المصححون لأعطاء الدرجات في ضوء المحكات المحددة للتقدير أو التصحيح، من أجل التأكد من أنهم يستخدمون المحك كما هو مخطط له، وأن تقديراتهم لا تتأثر بأي عوامل خارجية ليس لها علاقة بالسمة المستهدفة بالقياس.
٣. أدلة تعتمد على البنية الداخلية للمقياس، وهنا يكون الهدف اختبار مدى اتفاق المكونات الداخلية للمقياس مع السمة المعرفة نظرياً، ويتم ذلك باستخدام التحليل العاملي التوكيدي ودالة الأداء التفاضلي للفقرة (DIF) Differential item function؛ وذلك لتحديد الفقرات المتحيزة، أي الفقرات التي يتباين احتمال نجاح الأفراد من نفس مستوى القدرة في الإجابة عليها لأنهم ينتمون إلى مجموعات مختلفة، وهناك العديد من طرق حساب دالة الأداء التفاضلي للفقرة، بعضها يعتمد على النظرية الكلاسيكية في القياس، وبعضها يعتمد على نظرية الاستجابة للفقرة.

٤. الأدلة القائمة على العلاقات مع متغيرات أخرى، وتشمل الصدق المرتبط بمحك، بنوعيه التنبؤي والتلازمي، وكذلك الصدق النقاري، والصدق التمييزي.
٥. الدليل القائم على نتائج الاختبار، ويتعلق بالنتائج الإيجابية والسلبية المتوقعة لاستخدام المقياس، فهو يعكس مدى فائدة المقياس.
- (Goodwin & Leech, 2003: 183-184)

وهناك العديد من التهديدات التي تواجه جودة البحث التربوي، مثل أخطاء تصميم البحث، وطرق اختيار العينات، واستخدام الأساليب الإحصائية غير المناسبة، واستخدام أدوات غير صالحة لجمع البيانات (Oluwatayo, 2012: 398)، ومن ثم فقد اهتم عدد من الباحثين بجودة البحث التربوي، وظهر ذلك من خلال الدراسات التي اهتمت بمراجعة دقة الأساليب الإحصائية المستخدمة أو تحديد الأخطاء الشائعة لدى الباحثين.

فهدفت دراسة (Baumberger & Bangert (1996 إلى مراجعة البحوث التربوية المنشورة في مجلة صعوبات التعلم Learning disabilities ، خلال الفترة من ١٩٨٩م إلى ١٩٩٣م ، وتوصلت إلى أن ٨٠% من إجراءات تلك البحوث كانت غير ملائمة.

وهدف دراسة عفانة (٢٠١١) إلى معرفة الأخطاء الشائعة لدى طلبة الدراسات العليا في تصميم البحوث التربوية في جامعات قطاع غزة، ولتحقيق هذا الهدف تم وضع بطاقة ملاحظة يتم تطبيقها أثناء المناقشات العلمية؛ لتحديد أهم الأخطاء الشائعة في الهياكل الأساسية للبحث، وتوصلت الدراسة إلى أن هناك ثلاثة أخطاء في رسائل الماجستير المقدمة في الجامعة الإسلامية تتعلق بتساؤلات البحث، والدراسات السابقة، وعينة البحث، بينما كان هناك أربعة عشر خطأً رئيسياً في رسائل الماجستير في جامعة الأزهر بغزة، تتعلق بعنوان الرسالة، والمقدمة، والمشكلة، والتساؤلات، وصياغة الفروض، وأهمية البحث، وأهدافه، والدراسات السابقة، ومنهجية البحث، وأدواته، وخطوات البحث، والأساليب الإحصائية، والمراجع، والطباعة، بينما كان هناك خمسة أخطاء رئيسية في رسائل الماجستير المقدمة في جامعة الأقصى بغزة، تتعلق بصياغة العنوان، وكتابة المقدمة، وصياغة الفروض، والدراسات السابقة، والأخطاء المطبعية.

وأشار عفانه (٢٠١١: ٣٢٧) إلى أن طلاب الدراسات العليا بجامعات غزة يجدون صعوبة في إعداد أدوات البحث، وبخاصة أدوات تحليل المضمون والمقابلة والملاحظة، كما أن معظم الباحثين الذين يجرون بحثاً تجريبية لا يقومون بالتحقق من الخصائص السيكومترية للأدوات؛ نظراً لعدم توفر العينات.

وهدف دراسة (علي، وعبد الحليم، وحسن، ٢٠١٣) إلى تحديد مدى مناسبة الأساليب الإحصائية المستخدمة في حساب كفاءة الأدوات في بعض الرسائل العلمية بالأقسام التربوية بكلية التربية - جامعة أسيوط، خلال الفترة من ٢٠٠٤م إلى ٢٠٠٩م، في ضوء مستويات القياس، ومن أهم نتائج هذه الدراسة ارتفاع نسبة الأساليب الإحصائية غير المناسبة عند حساب

الصدق، حيث تراوحت هذه النسبة بين ٣١% إلى ٧٨.٥%، ومن هذه الأخطاء قيام الباحثين عند استخدام أسلوب المقارنة الطرفية للتحقق من صدق الاختبار بحساب دلالة الفروق بين الأقياء والضعفاء على الاختبار محل الدراسة، وليس على المحك. وكان من أهم نتائجها أيضاً ارتفاع نسبة الأساليب الإحصائية غير المناسبة عند حساب الثبات، حيث تراوحت هذه النسبة بين ٤٦% و ٧١%، ومن هذه الأخطاء الخلط بين استخدام معادلة كيودر-ريتشاردسون ومعادلة ألفا، واستخدام طريقة إعادة التطبيق مع اختبارات الذكاء والذاكرة والتحصيل، وعدم الالتزام بشروط الفاصل الزمني بين التطبيقين.

كما أشار ذبيحي وشويار (٢٠١٧: ٢٠) إلى الأخطاء الشائعة المتعلقة بالتحقق من الخصائص السيكومترية للمقاييس، ومنها: الخلط بين الاتساق الداخلي والصدق، وحساب دلالة الفروق بين درجات الربع الأعلى والربع الأدنى على المقياس نفسه، دون الرجوع إلى محك خارجي في طريقة المقارنة الطرفية، وكذلك استخدام طريقة التجزئة النصفية في حساب ثبات كل المقاييس دون التحقق من شروط استخدامها.

وقد حدد (عزت، ٢٠١٦: ٨-٩) الأخطاء التي يقع فيها الباحثون والمناقشون المتعلقة بالخصائص السيكومترية لأدوات البحث، ومن أبرزها: حساب الصدق قبل حساب الثبات، رغم أن الثبات شرط ضروري للصدق؛ لأن البدء بحساب الثبات يمثل عملية فلترة أولية لأداة القياس من البنود غير الثابتة، ثم يأتي الصدق لتنقية الأداة من العبارات غير الصادقة، ومن تلك الأخطاء أيضاً استخدام مصطلح "صدق الاتساق الداخلي"، والصواب هو "ثبات الاتساق الداخلي"، ومنها كذلك عدم الاهتمام بتحليل الفقرات، والاكتفاء بالتحليل الكلي للمقياس.

وترى الباحثة أن إجراءات التحقق من الخصائص السيكومترية للمقاييس يجب أن تسير بتسلسل منطقي، فإذا اهتم الباحث بالتحقق من صدق محتوى الأداة، فهنا يجب أن يبدأ بدراسة صدق المحتوى قبل الثبات، ولا يُعد ذلك كافياً لحساب الثبات؛ لأنه لا يمثل صدقاً مطلقاً، لكنه صدق لجانب واحد هو المحتوى؛ إذ إن الصدق مفهوم متعدد الأبعاد، فقد تمثل عينة الفقرات المجال المستهدف بالمقياس تمثيلاً جيداً، ولكن يفشل المقياس في تقديم أي معلومات عن السمة المستهدفة بالمقياس، وذلك عندما لا يستطيع التمييز بين المفحوصين، ثم يأتي دور التحليل التفصيلي للفقرات، فيتم حساب معاملات صعوبة الفقرات ومعاملات تمييزها، وذلك بهدف استبعاد الفقرات متطرفة الصعوبة، التي تُعد أحد مصادر أخطاء القياس، حيث تؤدي هذه الفقرات إلى انخفاض الثباين، ولا تقدم معلومات عن المفحوصين، فعندما تكون الفقرات سهلة جداً ينجح جميع المفحوصين في الإجابة عنها، وعندما تكون الفقرات صعبة جداً يفشل جميعهم في الإجابة عنها، فيظهرون كعينة متجانسة (عودة، ٢٠١٠: ٣٣٩)، ومن ناحية أخرى تؤدي الفقرات متطرفة الصعوبة إلى انخفاض دافعية المفحوصين في الاستجابة على المقياس أو قد تسبب شعورهم بالملل، كما أن الفقرات الصعبة تدفع المفحوصين للتخمين في الاختبارات المعرفية.

وفي التحليل التفصيلي للمقياس يتم حساب معاملات تمييز الفقرات؛ بهدف استبعاد الفقرات منخفضة التمييز، التي لا تقدم معلومات عن تباين المفحوصين، وأخيراً يأتي التحليل الكلي للمقياس، ويفضّل البدء بحساب الثبات، فإذا لم يتحقق شرط الثبات يجب التوقف وإعادة مراجعة الجوانب المختلفة للمقياس، أما إذا تم التحقق من ثبات المقياس فيمكن الاستمرار وجمع أدلة مختلفة للصدق طبقاً للهدف من المقياس.

ومن ناحية أخرى فقد لاحظت الباحثة أن جميع الأبحاث التي حاولت التحقق من صدق المحتوى لم تعتمد على أي أسلوب كمي لمعالجة استجابات الخبراء أو المحكمين، رغم وجود العديد من الأساليب الكمية للتحقق من صدق المحتوى، وقد أوردها الضوي (٢٠١٥) في دراسة بعنوان استخدام الأساليب الكمية في التحقق من صدق محتوى الاختبار النفسي.

الطريقة والإجراءات:

أولاً - منهج الدراسة:

تم اتباع المنهج الوصفي للإجابة عن تساؤلات الدراسة، وهو المنهج الذي يصف الواقع وصفاً دقيقاً من خلال التعبير الكمي أو الكيفي عن الظواهر العلمية، ويهدف إلى مساعدة الباحث للوصول إلى استنتاجات وتعميمات تساعد في تطوير الواقع محل الدراسة، من خلال تنظيم المعلومات وتصنيفها وتحليلها (عبيدات، وعبد الحق، وعدس، ٢٠٠٣: ١٩١).

ثانياً - عينة الدراسة:

شملت العينة ٧٢ بحثاً علمياً محكماً منشوراً في مجلة الجامعة الإسلامية للدراسات التربوية والنفسية، خلال عامي: ٢٠١٢م، و ٢٠١٦م.

ثالثاً - الخطوات الإجرائية للدراسة:

١- اختيار وعاء للنشر مصنّف دولياً، ويتيح المحتوى المنشور مجاناً على شبكة الإنترنت، مما يمكن من التحقق من نتائج الدراسة الحالية كأحد شروط البحث العلمي، ومن ناحية أخرى فإن الإتاحة المجانية تجعل المحتوى يصل إلى أكبر عدد من الباحثين، مما يزيد من خطورة الأخطاء، ويجعلها قابلة للانتشار والتكرار بين الباحثين، لذلك وقع الاختيار على مجلة الجامعة الإسلامية للدراسات التربوية والنفسية. وهي مجلة علمية محكمة، تصدر كل شهرين عن عمادة البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة، وبدأ إصدارها عام ١٩٩٣م لكل التخصصات العلمية، ثم تخصصت في العلوم التربوية والنفسية عام ٢٠١٢م، وتقوم بنشر البحوث العلمية الأصيلة في المجالات المختلفة للتربية وعلم النفس، باللغتين العربية والإنجليزية، من نتاج الباحثين في داخل الجامعة الإسلامية وخارجها، والتي لم يسبق نشرها من قبل، وتحمل الرقم المعياري الدولي: ISSN: 2410-3152، وتتيح جميع الأبحاث المنشورة منذ عام ٢٠١٢م حتى الآن، على الرابط الآتي:

<https://journals.iugaza.edu.ps/index.php/IUGJEPS/issue/archive>

٢- حصر الأبحاث المنشورة منذ ٢٠١٢م حتى الآن، ويوضحها جدول (١).

جدول (١): عدد الأبحاث المنشورة بمجلة الجامعة الإسلامية للدراسات التربوية والنفسية

السنة	٢٠١٢	٢٠١٣	٢٠١٤	٢٠١٥	٢٠١٦	٢٠١٧	٢٠١٨	٢٠١٩	مجموع
عدد المجلدات	٢	٤	٤	٤	٤	٤	٦	٤	
عدد الأبحاث	٤٢	٧١	٣٣	٣٣	٣٠	٧٧	١٦٩	١٥٤	٦٠٩

٣- اختيار أعداد محددة بهدف مراجعة كل الأبحاث المتضمنة بها وتحليل إجراءاتها؛ للإجابة عن تساؤلات البحث، وقد تم اختيار الأعداد المنشورة في عامي: ٢٠١٢م، و ٢٠١٦م، أي بمعدل زمني كل ٤ سنوات، ويرجع اختيار العام (٢٠١٢) إلى أنه البداية الفعلية لتخصص المجلة في العلوم التربوية والنفسية، والبداية الفعلية لإتاحة المحتوى المنشور مجاناً على شبكة الإنترنت، أما اختيار العام (٢٠١٦) فيرجع إلى تضاعف أعداد الأبحاث المنشورة من بعده تضاعفاً كبيراً. وقد ضم هذان العامان ٧٢ بحثاً، بما يعادل ١٢% تقريباً من الأبحاث المنشورة بالمجلة.

٤- إعداد جدول لتسجيل البيانات اللازمة؛ للإجابة عن تساؤلات الدراسة، بحيث يحتوي الجدول على رقم البحث وعنوانه ورقم المجلد والعدد، وعدد الأدوات المستخدمة به، وتصنيفها ما بين (مقننة، أو من إعداد الباحث)، وتسجيل طرق حساب الصدق وطرق حساب الثبات المستخدمة في البحث، ورصد الأخطاء في إجراءات كل منهما إن وجدت. ويوضح جدول (٢) شكل البطاقة المستخدمة.

جدول (٢): جدول تسجيل البيانات المطلوبة للتحليل

مسلسل البحث ضمن مجموعة البحوث المستهدفة، وعنوانه، وتاريخ النشر، ورقم المجلد، ورقم العدد.		
عدد الأدوات المستخدمة في البحث:		
نوع الأدوات	مقننة	من إعداد الباحث
طرق حساب الصدق		
الأخطاء إن وجدت		
طرق حساب الثبات		
الأخطاء إن وجدت		

٥- جمع البيانات اللازمة للتحليل باستخدام الجدول السابق، حيث تم جمع بيانات ٧٢ بحثاً.
٦- تفرغ البيانات التي تم جمعها في صورة جدول تكراري يوضح عدد الأدوات المستخدمة، ونسبة الأدوات المقننة منها، وتكرارات الطرق المستخدمة في حساب الثبات، وتكرارات الطرق المستخدمة في حساب الصدق، ونوعية الأخطاء، وتكراراتها.

نتائج الدراسة ومناقشتها:

الإجابة عن التساؤل الأول: ما الطرق الأكثر استخداماً للتحقق من ثبات أدوات القياس في البحوث التربوية؟

للإجابة عن هذا التساؤل تم حصر طرق حساب الثبات في البحوث التي تم استقرارها، وكان إجمالي عدد الأدوات المستخدمة في البحوث محل الدراسة ٩٢ أداة. ويوضح جدول (٣) طرق حساب الثبات وتكراراتها ونسبتها المئوية بالنسبة لعدد الأدوات.

جدول (٣) طرق حساب الثبات وتكراراتها ونسبتها المئوية

م	طريقة / طرق حساب ثبات المقياس	التكرار	النسبة
١	إعادة التطبيق	١٢	١٣ %
٢	الاتساق الداخلي باستخدام التجزئة النصفية	١٠	١١ %
٣	الاتساق الداخلي باستخدام معامل كيودر - ريتشاردسون	٤	٤.٣ %
٤	الاتساق الداخلي باستخدام معامل ألفا	٢٣	٢٥ %
٥	إعادة التطبيق + الاتساق الداخلي باستخدام تجزئة نصفية	١	١ %
٦	إعادة التطبيق + الاتساق الداخلي باستخدام معامل كيودر - ريتشاردسون	٥	٥.٤ %
٧	إعادة التطبيق + الاتساق الداخلي باستخدام معامل ألفا	١١	١٢ %
٨	الاتساق الداخلي باستخدام تجزئة نصفية + معامل ألفا	١٥	١٦.٣ %
٩	الاتساق الداخلي باستخدام تجزئة نصفية + معامل كيودر - ريتشاردسون	١	١ %
١٠	إعادة التطبيق + الاتساق الداخلي باستخدام تجزئة نصفية + معامل ألفا	١	١ %
١١	ثبات التصحيح	١	١ %
١٢	لم تقم بحساب الثبات للاعتماد على مقاييس مقننة	٨	٩ %
	المجموع	٩٢	١٠٠ %

يتضح من الجدول السابق اعتماد الباحثين بشكل أساسي على طريقة الاتساق الداخلي لحساب ثبات أدوات المقياس؛ حيث بلغت نسبة استخدام طريقة التجزئة النصفية أو معامل ألفا أو معامل كيودر - ريتشاردسون منفردة أو مع طريقة أخرى ٨٧% تقريباً، وبلغت نسبة استخدام طريق إعادة التطبيق منفردة أو مع طرق أخرى ٣٢%، مما يبين اتجاه الباحثين نحو الاعتماد على طريقة الاتساق الداخلي لحساب الثبات؛ نظراً لأنها تختصر الوقت، ولا تتطلب التطبيق إلا مرة واحدة. ويمكن أن نستدل من النتائج السابقة على أن الباحثين يتجهون لاختيار الطريقة العملية لحساب الثبات، بغض النظر عن طبيعة مصادر الأخطاء غير المنتظمة التي تهدد القياس.

الإجابة عن التساؤل الثاني: ما الطرق الأكثر استخداماً للتحقق من صدق أدوات القياس في البحوث التربوية؟

للإجابة عن هذا التساؤل تم حصر طرق حساب الصدق في البحوث التي تم استقراؤها مع الالتزام بنفس المسميات الواردة في تلك البحوث، وإن كانت مسميات خاطئة، على أن يتم توضيح ذلك عند مناقشة النتائج، وكان إجمالي عدد المقاييس المستخدمة في البحوث محل الدراسة ٩٢ أداة، ويوضح جدول (٤) طرق حساب الصدق وتكراراتها ونسبتها المئوية بالنسبة للعدد الكلي للأدوات.

جدول (٤) طرق حساب الصدق كما وردت بالأبحاث وتكراراتها ونسبتها المئوية

م	طريقة / طرق حساب صدق أدوات القياس	التكرار	النسبة
١	صدق المحكمين أو صدق المحتوى فقط	٣٢	٣٥ %
٢	الاتساق الداخلي فقط	١٢	١٣ %
٣	الصدق التمييزي فقط	١	١ %
٤	الصدق المرتبط بالمحتوى + الاتساق الداخلي	٣٥	٣٨ %
٥	الصدق المرتبط بالمحتوى + التحليل العملي	١	١ %
٦	الصدق المرتبط بالمحتوى + الصدق التمييزي	٢	٢ %
٧	الصدق المرتبط بالمحتوى + الاتساق الداخلي + الصدق التمييزي	١	١ %
٨	لم تقم بحساب الصدق للاعتماد على مقاييس مقننة أو منشورة	٨	٩ %
	المجموع	٩٢	١٠٠ %

يتضح مما سبق أن الباحثين يعتمدون بشكل أساسي على طريقة الصدق المرتبط بالمحتوى؛ حيث بلغت نسبة استخدام هذه الطريقة منفردة أو مع طرق أخرى ٧٧%، تليها طريقة الاتساق الداخلي، حيث بلغت نسبة استخدام هذه الطريقة منفردة أو مع طرق أخرى ٥٢%. وبمراجعة إجراءات هذه الطرق في البحوث محل الدراسة وجد أن هناك الكثير من الأخطاء العلمية والإجرائية، ومنها إطلاق العديد من الباحثين مسمى "صدق المحكمين" على طريقة الصدق المرتبط بالمحتوى، كما لو كان الهدف هو التأكد من صلاحية المحكمين لقياس السمة، كما أفتقدت تعليمات المحكمين إلى المعايير الإجرائية المستخدمة في تقدير صلاحية فقرات المقياس، ولم تستخدم أي دراسة أي طريقة كمية لمعالجة استجابات المحكمين، وفي بعض الدراسات لم يتم وصف عينة المحكمين على الإطلاق.

ويمكن أن نستنتج من نسبة استخدام الصدق المرتبط بالمحتوى فقط، التي بلغت ٣٥%، قصور إدراك الباحثين لمفهوم الصدق كمفهوم متعدد الأبعاد، والاكتفاء ببعد واحد فقط، هو الصدق المرتبط بالمحتوى، فرغم أهمية هذا الجانب، فإنه غير كافٍ، وبخاصة عند استخدام أساليب إحصائية تشترط اعتدالية توزيع الدرجات؛ لأنه من الممكن أن تمثل الفقرات السمة تمثيلاً جيداً وتنتمي جميعها للمجال المستهدف بالقياس، لكنها تفشل في تقديم أي معلومات عن تباين مستوى السمة لدى المفحوصين؛ نظراً لتطرف معاملات الصعوبة، وضعف معاملات التمييز، فيحصل الجميع على درجات متشابهة، ويظهرون كعينة متجانسة، لذلك يجب الاهتمام بقدرة أداة القياس على التمييز بالإضافة لصدق محتواه.

وهناك إجراءان للتحقق من قدرة المقياس على التمييز بين المستويات المختلفة للسمة المقاسة، أولهما: تحليل داخلي وعلى مستوى الفقرات، حيث يتم حساب معامل تمييز الفقرات بحساب معامل الارتباط بين درجة الفقرة والدرجة الكلية على المقياس عندما يقيس سمة أحادية البعد بعد تحرير الدرجة الكلية من درجة الفقرة، أو حساب معامل الارتباط بين درجة الفقرة ودرجة البعد بعد تحرير درجة البعد من درجة الفقرة، وذلك عندما يتكون المقياس من عدة أبعاد، ثم استبعاد الفقرات التي يقل معامل تمييزها عن ٠,٣، وسُمي هذا التحليل بأنه تحليل داخلي لأننا اعتمدنا على الدرجة الكلية للمقياس أو البعد كمحك لقياس قدرة الفقرة على التمييز.

أما الإجراء الثاني لتقدير الصدق التمييزي للمقياس فهو التحليل الخارجي وعلى مستوى الدرجة الكلية للمقياس، حيث يتم تصنيف المفحوصين بناءً على درجاتهم الكلية إلى مجموعتين: واحدة مرتفعة السمة، وأخرى منخفضة السمة، ثم اختيار مقياس خارجي يقيس السمة نفسها كمحك، وحساب دلالة الفرق بين متوسطي درجات هاتين المجموعتين على الاختبار المحك، وتعد دلالة هذا الفرق دليلاً على قدرة المقياس على التمييز، وهو ما يُعرف بالصدق التمييزي. ويشكل آخر يمكن اختيار محك خارجي مناسب، وتصنيف المفحوصين في ضوءه إلى مجموعتين أو مجموعات متميزة في مستوى السمة المستهدفة بالقياس، ثم حساب دلالة الفروق بين متوسطات درجات هذه المجموعات على المقياس، وتعد دلالة هذه الفروق دليلاً على قدرة المقياس على التمييز.

وبمراجعة إجراءات الطريقة التي أطلق عليها الباحثون "صدق الاتساق الداخلي" نجد أن ذلك المسمى الخطأ يتضمن إجراءً صحيحاً، حيث يتم في هذه الطريقة حساب معامل الارتباط بين درجة الفقرة والدرجة الكلية للبعد، أو حساب معامل الارتباط بين درجة الفقرة والدرجة الكلية للمقياس، أي حساب معاملات تمييز الفقرات، الذي يسهم بشكل جيد في التحقق من صدق الفقرات، التي تمثل بنية المقياس.

وهنا يجب أن نميز بين هذا "الاتساق الداخلي" المزعم من قبل الباحثين، الذي يجب أن يعاد تسميته بـ"التحقق من قدرة الفقرات على التمييز"، وبين الاتساق الداخلي كأحد طرق حساب الثبات، الذي يعتمد على إيجاد متوسط معاملات الارتباطات بين كل فقرات المقياس، وذلك باستخدام معامل ألفا أو كيودر- ريتشاردسون، والذي يمكن أيضاً أن يُستخدم كمؤشر أولي لصدق البناء، لكنه غير كافٍ؛ لأنه يدل على تجانس الفقرات ولا يدل على كفايتها في تمثيل المجال المستهدف، فقد تكون جميع الفقرات متجانسة وتمثل بعداً واحداً، غير أنها تقيس سمة أخرى مختلفة عن السمة المستهدفة بالقياس.

الإجابة عن التساؤل الثالث: ما نسبة البحوث التي تعتمد على أدوات قياس مقننة أو منشورة؟

للإجابة عن هذا التساؤل تم حصر أدوات القياس التي أعدها الباحثون والأدوات المقننة أو المنشورة في البحوث التي تم استقراؤها، حيث ظهر أن الباحثين قاموا بإعداد ٧٨ مقياساً من بين ٩٢ مقياساً، أي ما يعادل ٨٥ % من المقاييس، وهي نسبة كبيرة، وتعد مؤشراً على عدم جودة البحث التربوي، وبخاصة مع وجود أخطاء في إجراءات التحقق من ثبات هذه الأدوات وصدقها، فلن يتم الوثوق في نتائج البحوث لآبد من الاعتماد على أدوات قياس ذات خصائص سيكومترية جيدة، وهنا يجب الإشارة إلى أن إعداد أدوات القياس ليس هدفاً في حد ذاته، لكنه وسيلة لجمع البيانات التي تسهم في الإجابة عن تساؤلات البحث بأعلى قدر من الصحة والدقة. وقد لاحظت الباحثة أن بعض الباحثين يعيدون حساب صدق وثبات الأدوات المقننة أو المنشورة، وبعضهم لا يقوم بأي إجراء للتحقق من ثبات هذه الأدوات، وكان ينبغي على الباحث الذي تبني مقياساً مقنناً أن يقوم بإعادة حساب ثبات درجاته فقط على عينة مماثلة لعينة بحثه؛ لأن ثبات درجات المقياس مرتبط بعينة البحث في النظرية الكلاسيكية للقياس.

كما أن استخدام مقاييس غير مقننة يضعف فرصة الاستفادة من نتائج البحث التربوي، والقدرة على تعميمها؛ لأن نتائج البحث تصبح محددة بالأدوات، وكذلك لا يمكن مقارنة نتائج الدراسات المختلفة.

الإجابة عن التساؤل الرابع: ما الأخطاء الشائعة في إجراءات التحقق من ثبات أدوات القياس في البحوث التربوية؟

للإجابة عن هذا التساؤل تم تصنيف الأخطاء الشائعة في إجراءات حساب ثبات أدوات القياس في البحوث التربوية، ويمكن عرضها فيما يلي:

١. في المقاييس المكونة من عدة أبعاد، تم حساب معامل ألفا لدرجة كل بعد دون حسابه للدرجة الكلية على المقياس، وفي دراسات أخرى تم حساب معامل ألفا للدرجة الكلية على المقياس دون حسابه لدرجة كل بعد.

٢. عند استخدام إعادة التطبيق لم يتم وصف عينة حساب الثبات، أو الفاصل الزمني، كما تم اختيار الفاصل الزمني دون ذكر لمعيار اختيار هذا الفاصل، حيث تراوح الفاصل الزمني بين التطبيقين بين أسبوعين وثلاثة أسابيع، بغض النظر عن طبيعة العينة أو طبيعة السمة المستهدفة بالقياس، وكذلك تم استخدام على معامل ارتباط بيرسون في حساب الارتباط بين درجات التطبيقين دون التحقق من شرط اعتدالية توزيع الدرجات. وتتسق هذه النتيجة مع ما أشارت إليه دراسة الحازمي (٢٠٠٣)، التي هدفت إلى تقييم الأساليب الإحصائية الارتباطية المستخدمة في رسائل الماجستير المقدمة لكلية التربية بجامعة أم القرى بمكة المكرمة، حيث لم يجد الباحث أي رسالة حاولت التأكد من توافر افتراضات وشروط مقاييس العلاقة، مثل: مستوى القياس، وحجم العينة، وتجانسها، وخطية العلاقة، واعتدالية التوزيع.
٣. في الدراسات النادرة التي أجرت التحليل التفصيلي لل فقرات تم دمج إجراءات حساب معاملات الصعوبة ومعاملات التمييز مع إجراءات حساب الثبات.
٤. إجراء التحليل التفصيلي لل فقرات بعد حساب الثبات.
٥. في بعض الدراسات استخدم الباحث عدة أدوات، ووصف الطريقة المستخدمة لحساب الثبات مكثفياً بكتابة معامل ثبات واحد فقط.
٦. الدمج بين ثبات إعادة وثبات الاتساق الداخلي، ففي عدة أبحاث تم إعادة تطبيق الأداة على عينة صغيرة من المفحوصين، ثم تم دمج درجات التطبيقين معاً لحساب ثبات الاتساق الداخلي باستخدام معامل ألفا.
٧. عند استخدام المقاييس المقننة تم الاكتفاء بسرد معلومات الصدق والثبات المنشورة مع المقاييس، دون إعادة حساب الثبات، على الرغم من أن الثبات مرتبط بالعينة طبقاً للنظرية الكلاسيكية في القياس، ومن ثم يجب إعادة حساب الثبات على عينة استطلاعية مشابهة لعينة البحث الأساسية.
٨. عند استخدام الاختبارات الإنشائية تم إهمال ثبات تقديرات المحكمين أو المصححين، وتم الاكتفاء بثبات الاتساق الداخلي.
٩. الاعتماد على عينات صغيرة لحساب الثبات، حيث تراوح حجم العينة في بعض البحوث ما بين ٨ و ١٥ مفحوصاً فقط.
١٠. حساب الثبات باستخدام معادلة سبيرمان براون للتجزئة النصفية للمقاييس متعددة الأبعاد، رغم أن استخدام التجزئة النصفية يفترض أن الاختبار مكون من نصفين متكافئين.
١١. إعداد استبانة تضم مجموعتين من الفقرات كل منهما تقيس متغيراً مختلفاً، ثم استخدام الدرجة الكلية للاستبيان (مجموع درجتي المتغيرين) في حساب الثبات.
١٢. نسب بعض الباحثين الثبات إلى العينة، فقال: "وكان معامل ثبات العينة يساوي"، رغم أن الثبات صفة من صفات درجات الاختبار.

الإجابة عن التساؤل الخامس: ما الأخطاء الشائعة في إجراءات التحقق من صدق أدوات القياس المستخدمة في البحوث التربوية؟

للإجابة عن هذا التساؤل تم تصنيف الأخطاء الشائعة في إجراءات حساب صدق أدوات القياس، ويمكن عرضها فيما يلي:

- ١- استخدام الاتساق الداخلي كطريقة لحساب الصدق، رغم أن الاتساق الداخلي هو أحد طرق حساب الثبات، كما أن إجراءات حسابه تمثلت في حساب معامل الارتباط بين درجة الفقرة ودرجة البعد أو حساب معامل الارتباط بين درجة الفقرة والدرجة الكلية، وهو ما يشير إلى معامل تمييز الفقرة.
- ٢- بالنسبة لطريقة الصدق المرتبط بالمحتوى، كان هناك قصور في وصف عينة المحكمين من حيث عددهم وتخصصهم، وعدم ذكر معايير إجرائية للتحكيم، وعدم توضيح طريقة معالجة استجابات المحكمين، أو استخدام أي طريقة كمية للتعبير عن الصدق المرتبط بالمحتوى.
- ٣- إطلاق مصطلح "صدق المحكمين" على طريقة الصدق المرتبط بالمحتوى، وأحياناً لا يتم ذكر اسم الطريقة، ويكتفى بشرح إجراءاتها.
- ٤- عند التحقق من الصدق التمييزي للمقياس، يتم حساب الفروق بين الفئة الأعلى والفئة الأدنى على نفس المقياس، دون الاستناد لمحك خارجي.
- ٥- عند التحقق من الصدق التمييزي للمقياس، لا يتم توضيح الاختبار الإحصائي المستخدم في مقارنة متوسطات المجموعات.
- ٦- القصور في استخدام التحليل التفصيلي للفقرات، وعند إجرائه لا يتم إعادة التحليل بعد حذف الفقرات ضعيفة التمييز.
- ٧- القصور في استخدام التحليل العاملي، وفي الدراسة الوحيدة التي استخدمته لم يتم عرض نتائجها، ولم يتم التحقق من شرط مناسبة العينة للتحليل، فقد تم استخدام عينة من ٨٩ طالباً فقط لاجراء تحليل عاملي لمقياس مكون من ٥٢ فقرة.
- ٨- ذكر أكثر من طريقة لحساب الصدق، ثم استخدام أسلوب واحد وبطريقة خاطئة، كأن يذكر الباحث أنه تم التحقق من الصدق الظاهري وصدق الاتساق الداخلي والصدق البنائي بحساب معاملات الارتباط بين درجة كل مجال والدرجة الكلية.
- ٩- عدم التحقق من الجوانب المختلفة لصدق الأدوات، والاكتفاء في معظم الأحيان بالصدق المرتبط بالمحتوى أو الطريقة المسماة خطأً "الاتساق الداخلي".

وفي ضوء ما ورد في الإجابة عن التساولين الأخيرين يتضح أن نتائج الدراسة الحالية تتسق مع نتائج دراسة باهي (٢٠١١)، التي توصلت إلى أن هناك أخطاء كثيرة في استخدام المصطلحات العلمية الخاصة بكل من الثبات والصدق، إذ تم استتساخ العديد من المصطلحات التي لا أساس لها من الصحة.

توصيات الدراسة:

في ضوء نتائج الدراسة يمكن صياغة التوصيات الآتية:

- (١) إعادة النظر في برامج الدراسات العليا وتضمينها ساعات كافية وعملية تهدف إلى إكساب الدارسين المهارات الضرورية والمعلومات الأساسية التي تجنبهم الوقوع في أخطاء إجراءات التحقق من صدق وثبات المقاييس المستخدمة في البحث التربوي.
- (٢) إعداد دليل مبسط لطرق التحقق من صدق وثبات الأدوات المستخدمة في البحث التربوي ونشره على مواقع المجالات العربية.
- (٣) إعداد مقاييس تقدير تفصيلية لتحكيم الأبحاث التربوية قبل قبولها للنشر، في ضوء الخطوات العلمية لاختيار أدوات القياس المناسبة وإعدادها.
- (٤) إعداد قاعدة بيانات موحدة تحتوي على أدوات القياس المقننة والمنشورة، واتخاذ الإجراءات المناسبة لتشجيع الباحثين على الاعتماد عليها، بما يضمن تطور مستوى البحث التربوي، ويتيح الفرصة لمقارنة نتائج الدراسات المختلفة، والوصول إلى استنتاجات عامة من خلال منهج التحليل البعدي.
- (٥) المراجعة الدورية للأبحاث المنشورة والوقوف على الأخطاء المنهجية التي يقع فيها الباحثون، والمتعلقة باختيار الأدوات المناسبة، وإجراءات التحقق من صدق وثبات المقاييس، واختيار الأساليب الإحصائية المناسبة، والتحقق من شروط استخدامها، والتفسير التصحيح لنتائجها.

قائمة المراجع:

- باهي، مصطفى حسين إبراهيم. (٢٠١١). بعض القضايا الهامة للتحليل الإحصائي في البحوث النفسية والتربوية. المؤتمر العلمي السنوي العربي السادس - الدولي الثالث - تطوير برامج التعليم العالي النوعي في مصر والوطن العربي في ضوء متطلبات عصر المعرفة: جامعة المنصورة - كلية التربية النوعية، مج ٣، ص ١٣٨٨-١٤١٢.
- الحازمي، عبدالعزيز بن إبراهيم بن خليل. (٢٠٠٣). معاملات الارتباط وعلاقتها بمستويات القياس: دراسة تقويمية على رسائل الماجستير المقدمة لكليات التربية جامعة أم القرى بمكة المكرمة، (رسالة ماجستير غير منشورة)، جامعة أم القرى، مكة المكرمة.
- حسن، عزت عبد الحميد محمد. (٢٠١٦). أخطاء الباحثين والمناقشين الشائعة في البحوث النفسية والتربوية. أعمال مؤتمر: تكنولوجيا التربية والتحديات العالمية للتعليم: الجمعية العربية لتكنولوجيات التربية، القاهرة، ٦٧ - ٨٥.
- أبو دنيا، نادية عبده. (٢٠١٨). القياس والتقويم النفسي والتربوي في العملية التعليمية، الدمام: مكتبة المنتبي.
- نبيحي، لحسن، وشوبار، لياس. (٢٠١٧). أخطاء شائعة في البحوث العلمية. مجلة جيل العلوم الإنسانية والاجتماعية: مركز جيل البحث العلمي، (٢٨)، ص ١١-٢٣.
- عبيدات، ذوقان، وعبد الحق، كايد، وعدس، عبد الرحمن. (٢٠٠٣). البحث العلمي "مفهومه أدواته أساليبه"، القاهرة: دار الفكر.
- الضوى، محسوب عبد القادر. (٢٠١٥). استخدام الأساليب الكمية في التحقق من صدق محتوى الاختبار النفسي، مجلة كلية التربية - جامعة الاسكندرية، ٢٥ (٢)، ص ٢١-١١٤.
- عطية، محسن علي. (٢٠٠٩). البحث العلمي في التربية: مناهجه، أدواته، وسائله الإحصائية، عمان: دار المناهج للنشر والتوزيع.
- عفانة، عزو إسماعيل سالم. (٢٠١١). أخطاء شائعة في تصاميم البحوث التربوية لدى طلبة الدراسات العليا في الجامعات الفلسطينية. أعمال مؤتمر البحث العلمي: مفاهيمه، أخلاقياته، توظيفه: الجامعة الإسلامية بغزة، ص ٣٠٥-٣٣٦.
- علام، صلاح الدين محمود. (٢٠١١). القياس والتقويم التربوي والنفسي، أساسياته وتطبيقاته وتوجهاته المعاصرة، ط٥، القاهرة: دار الفكر العربي.

-
- أبو علام، رجاء محمود (٢٠٠٦). *مناهج البحث في العلوم النفسية والتربوية*، ط ٥ ، القاهرة: دار النشر للجامعات.
 - علي، عماد أحمد حسين، وعبد الحليم، محمد رياض أحمد، وحسن، علي صلاح عبدالمحسن. (٢٠١٣). *الممارسات الإحصائية الخاطئة في حساب صدق وثبات الأدوات في الرسائل العلمية، مجلة كلية التربية: جامعة أسيوط*، ٢٩ (٣)، ص ٣٢٤-٣٤٨.
 - عودة، أحمد سليمان. (٢٠١٠). *القياس والتقويم في العملية التدريسية*. الأردن: دار إريد، ط ٤.
 - ملحم، سامي محمد. (٢٠١٥). *القياس والتقويم في التربية وعلم النفس*، ط ٧، عمان: دار المسيرة.
 - ميخائيل، أمطانيوس نايف. (٢٠١٥). *القياس والتقويم التربوي للأسوياء وذوي الاحتياجات الخاصة*، ط ١، عمان: دار الإعصار العلمي للنشر والتوزيع.

-
- American Educational Research Association, American Psychological Association & National Council of Measurement in Education .(1999). Standards for Educational and Psychological testing. Washinton: American Educational Research Association.
 - Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. Measurement and Evaluation in Counseling and Development, 36(3), 181-191.
 - FONG, M. L., & MALONE, C. M. (1994). Defeating ourselves: Common errors in counseling research. Counselor Education and Supervision, 33(4), 356-362.
 - Oluwatayo, J. A. (2012). Validity and Reliability Issues in Educational Research, Journal of Educational and Social Research, 2 (2), 391-400.
 - Baumberger JP1, Bangert AW. (1996) . Research designs and statistical techniques used in the Journal of Learning Disabilities, 1989-1993, retrived from: <https://www.ncbi.nlm.nih.gov/pubmed/8732893>.
 - Bond, T. & Fox, C. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
 - Sijtsma, K. & Molenaar, I. (2002). Introduction to nonparametric item response theory. London: Sage Publications.
 - Sheng, Y. & Wikle, C. (2007). Comparing multiunidimens-ional and unidimensional item response theory models. Educational and Psychological Measurement, 67(6), 899-919.
 - Harris, D. & Kolen, M. (1986). Effect of examinee group on equating relationships. Applied Psychological Measurement, 10, 35-43.

- Wainer, H. & Mislevy, R. (2000). Item response theory, item calibration and proficiency estimation. In: H. Wainer, N. Dorans, R. Flaugher, B. Green, R. Mislevy, L. Steinberg & D. Thissen (Eds.), Computerized adaptive testing: A primer (2nd ed., pp. 61-100). London: Lawrence Erlbaum Associates.
- Hambleton R., Swaminathan H. & Rogers, H. (1991). Fundamentals of item response theory. London: Sage Publications.
- Osterlind, S. (1992). Constructing test items. London: Kluwer Academic Publishers.
- Blais, J. & Laurier, M. (1995). The dimensionality of a placement test from several analytical perspectives. Language Testing, 12(1), 72-98.
- Thorndike, R. (1982). Applied psychometrics. London: Houghton Mifflin Company.
- Jeeva ,R. K.V.(2016). CONTEMPORARY ISSUES IN TEACHER EDUCATION, retrieved from: <https://books.google.com.sa/books?id=g5XDAAAQBAJ&printsec=frontcover&hl=ar#v=onepage&q&f=false>.