ISSN:2536-9245

مؤشرات التحليل البعدي للعوامل المؤثرة في مواصفات الاختبار الجيد

د/ وجدان وديع محمد خياط*

الملخص:

هدفت الدراسة الحالية إلى التعرف على مؤشرات التحليل البعدي للعوامل المؤثرة في مواصفات الاختبار الجيد، في ضوء اختلاف نوع الاختبار (تحصيلي – نفسي)، ولتحقيق أهداف الدراسة تم استخدام المنهج التحليلي البعدي، حيث تمت مراجعة دراسات سابقة في خصائص الاختبار الجيد، وخلصت الباحثة إلى الإبقاء على (٧٩) أداة قياس اشتملت على بيانات كافية لحساب حجم التأثير في متغيرات الدراسة، منها: (١٦) اختبار تحصيلي و(٦٣) اختبار نفسي. وتم الحصول على (٢٥٠) حجم تأثير باستخدام معادلة كوهين، منها: (٤٥) حجم تأثير للاختبارات التحصيلية، و (٢٠٥) حجم تأثير للاختبارات النفسية. وأظهرت النتائج ما يلي: بلغ مستوى حجم تأثير أساليب الصدق وفق محك كوهين كالتالي: حجم التأثير الكلي لصدق المقارنة الطرفية بلغ (٤,٠٨٧)، وهو حجم تأثير مرتفع، يليه أسلوب الصدق الذاتي، حيث بلغ حجم التأثير الكلي لها (٠,٩٣٠)، وهو حجم تأثير مرتفع. ثم أسلوب الصدق التلازمي حيث بلغ حجم التأثير الكلي (٠,٦٩٦)، وهو حجم تأثير متوسط. وأخيراً اسلوب الصدق البنائي، حيث بلغ حجم التأثير الكلي (٠,٦٩٢)، وهو حجم تأثير متوسط. وتوجد فروق ذات دلالة إحصائية بين متوسطي حجم التأثير لأساليب: الصدق البنائي، المقارنة الطرفية، والصدق التلازمي لصالح الاختبارات التحصيلية. وكذلك توجد فروق ذات دلالة إحصائية بين متوسطي حجم التأثير لأسلوب الصدق الذاتي لصالح الاختبارات النفسية. ووجد أنه بلغ مستوى حجم تأثير أساليب الثبات كالتالي: حجم التأثير الكلي لأسلوب إعادة التطبيق بلغ (٠,٨٦٧)، وهو حجم تأثير مرتفع، يليه اسلوب الفا كرونباخ، حيث بلغ حجم التأثير الكلي (٠,٨٥١)، وهو حجم تأثير مرتفع. وأخيراً اسلوب التجزئة النصفية، حيث بلغ حجم التأثير الكلي (٠,٨٥١)، وهو أيضاً حجم تأثير مرتفع. وأخيراً توجد فروق ذات دلالة إحصائية بين متوسطي حجم التأثير في أساليب الثبات (الفا كرونباخ– التجزئة النصفية– إعادة التطبيق) لصالح الاختبارات التحصيلية.

الكلمات المفتاحية: التحليل البعدي، حجم التأثير، الدلالة العملية، الاختبار الجيد، الخصائص السيكومترية، الثبات، الصدق، الاختبار التحصيلي والنفسي

* أستاذ القياس والتقويم المساعد بقسم علم النفس بكلية التربية بجامعة أم القرى

ISSN:2536-9245

Indicators of The Meta Analysis of The Factors Affecting The Specifications of A Good Test

Abstract

The current study aimed to identify the indicators of the Meta Analysis of the factors affecting the specifications of a good test, in light of the type of test (achievement - psychological). To achieve the objectives of the study, the Meta Analysis approach was used. And the researcher reviewed the previous studies in the characteristics of a good test. Afterward, the researcher concluded by maintaining (79) measurement tools that included sufficient data to calculate the size of the effect on the study variables, including: (16) achievement test and (63) psychological test. (250) effect sizes were obtained, using Cohen's equation, including: (45) effect sizes for achievement tests, and (205) effect sizes for psychological tests. The results of the study are as follows: The level of effect size of the validation methods according to Cohen's criterion was as follows: the total effect size of the validity of the discriminative comparison amounted to (4.087), which is a high effect size, followed by the intrinsic-validity method, as the total effect size of it reached (0.930), which is an effect size High. Then the construct validity method, where the total effect size was (0.696), which is a medium effect size. Finally, the constructive validity method, where the total effect size was (0.692), which is a medium effect size. - There are statistically significant differences between the mean effect size of the following methods: construct validity, discriminative comparison, and concurrent validity in favor of achievement tests. - There are statistically significant differences between the mean effect size of the intrinsic-validity method in favor of psychological tests.The effect size of the stability methods was as follows: the total effect size of test -retest method was (0.867), which is a high effect size, followed by Cronbach's alpha method, where the total effect size was (0.851), which is a high effect size. Finally, the split-half method, where the total effect size was (0.851), which is also a high effect size. andThere are statistically significant differences between the mean effect size in the stability methods (Cronbach's alpha / split-half / test -retest) in favor of the achievement tests.

Keywords: Meta Analysis, effect size, practical significance, good test, psychometric properties, reliability, validity, Achievement and Psychological Test

المقدمة:

يشهد مجال البحوث في الآونة الأخيرة ثورة معلوماتية هائلة في مختلف العلوم بكافة المجالات ومنها مجال القياس النفسي والتربوي، وحتى يتم الاستفادة من نتائج هذه الأبحاث وتحقق أهدافها المرجوة لابد أن تقوم منهجية البحث العلمي على آلية تتضمن عدد من الخطوات ومنها استخدام الأدوات لجمع البيانات والتحقق من الفروض البحثية، حيث يُعد اختيار الأدوات وتطبيقها من أهم الخطوات لأنه عن طريقها يتم جمع بيانات الدراسة، لذلك على الباحث أن يختار الأداة التي تتناسب مع دراسته وأن يتحقق من خصائصها السيكومترية والتي يأتي في مقدمتها الصدق والثبات.

ونرى في مجال البحوث أنه قد أدى تطور وتعدد النظريات في مجال القياس الى التنوع في الأبحاث العلمية، وترتب على هذا التنوع ظهور الكثير من الأبحاث والرسائل والأوراق العلمية في الموضوعات نفسها. والتي أُجريت بشكل مستقل عن بعضها البعض، وتوصلت بعضها إلى نتائج متشابهة، والبعض الآخر قد أظهرت نتائج مختلفة أو نتائج متضاربة.

ومن هذه الموضوعات تقنين الأدوات والتحقق من خصائصها السيكومترية سواء النفسية أو التحصيلية، حيث يُعد بناء الأدوات والاختبارات من الموضوعات المهمة التي تهم الباحثين والمهتمين في مجال البحث التربوي عامة ومجال القياس والتقويم خاصةً، لأنها تزودنا ببيانات يُستند إليها في اتخاذ العديد من القرارات المهمة التي تخص الفرد والمجتمع.

ولذلك يجب أن تحقق الاختبارات –بمختلف أنواعها– أهداف القياس النفسي والتقويم وأن تتمتع بخصائص ومواصفات معينة، منها: الموضوعية والثبات والصدق ووجود معايير وسهولة التصحيح والتفسير (علام، ٢٠١٥؛ حسن ، ٢٠١٨). وذلك لضمان صحة ودقة البيانات التي يتم جمعها، فالقصور في التحقق من ثبات وصدق الأدوات يؤدي للوصول إلى نتائج مضللة (محمود، ٢٠١٩). وكذلك يصعب اتخاذ قرارات صائبة تتعلق بالأفراد (علام، ٢٠١٥)، لهذا ينبغي أن تكون أدوات القياس على قدر كبير من الثبات والصدق وأن تكون خالية من التحيز والتشويه (باهي، ٢٠١٤).

ونظراً لتعدد الدراسات في مجال تقنين الاختبارات، فهذه بعض الدراسات على سبيل المثال لا الحصر، دراسات هدفت إلى التعرف على خصائص الاختبار الجيد ومدى توافرها في الاختبارات والتحقق من الخصائص السيكومترية (مثل: الحربي، ٢٠٢١؛ ضيف، ٢٠١٦)، وهدفت دراسة حسن (٢٠١٨) إلى تحديد الأخطاء الشائعة في بناء الاختبارات الموضوعية وتأثيرها على الخصائص السيكومترية للاختبارات التحصيلية.

وإزاء هذا الكم من الدراسات والابحاث في مجال الاختبارات والتي تحققت من خصائص الاختبار الجيد سواء كانت اختبارات تحصيلية أو نفسية. يأتي دور البحث العلمي من جديد، إذ لابد من تحليل هذه البحوث والدراسات للتعرف على سماتها، ولتحديد خصائص الاختبار الجيد الأكثر استخداماً ومدى تأثر هذه الخصائص ببعض العوامل التي

تؤثر عليها، وكذلك تحديد المتغيرات التي تؤثر على خصائص الاختبار منها المرحلة الدراسية ونوع الاختبار.

حيث يرى ولسون Wilson (١٩٧٧) أن الفهم الكامل لأي ظاهرة يتطلب دراسة تحليلية جنباً إلى جنب مع دراسة عامة في سياق متكامل أي دراسة توليفيه أو تركيبية. ويتفق ييجر Yager (١٩٨٠) مع ولسون في ذلك ويضيف أن التحليل والتوليف وجهان لعملة واحدة في البحث العلمي، وإذا ما أحسن استخدامهما معاً فإنهما يشكلان أبعاداً أساسية للبحث ككل (في: عبدالحميد، ١٩٨٧). وقد أشارت محاسنة (٢٠٢٠) أن مسألة الجمع بين الدراسات المستقلة وتفسيرها من المسائل الأساسية والمهمة في جميع العلوم، ومن الصعب إجراء دراسة على عينات كبيرة بسبب وجود عدد من العقبات، منها: الوقت والتكلفة، وتوافر الباحثين، بالإضافة إلى ذلك فإن مناقشة نتائج دراسة واحدة لا يكفي، ما يستلزم توليف نتائج العديد من الدراسات لاتخاذ قرار حولها.

ويعد التوليف أو التركيب حديثاً نسبياً في مجال بحوث التربية العلمية. وأحد مداخل توليف الأبحاث يطلق عليه اسم التحليل البعدي Meta – Analysis وقد قدمه لأول مرة "جين جلاس" Gene Glass في عام ١٩٧٦ (عبدالحميد، ١٩٨٧).

ويشير منهج التحليل البعدي، إلى إعادة تحليل للتحليل الثانوي من مجموع الدراسات الفردية، حيث يتميز عن كل من التحليل الأولي للبيانات Primary Analysis والذي يشير للتحليل الأصلي لبيانات البحث، والتحليل الثانوي Secondary Analysis (re–analysis) والذي يشير إلى إعادة تحليل البيانات، بهدف الإجابة عن تساؤلات البحث باستخدام أفضل وأنسب الطرق الإحصائية، أو الإجابة عن أسئلة جديدة باستخدام بيانات قديمة (سكران، ٢٠٠٦).

ويُعرف التحليل البعدي بإنه تحليل إحصائي لمجموعة كبيرة من نتائج الدراسات السابقة للوصول إلى التكامل فيما بينها. وتكمن قيمته في إضفاء معنى على النتائج التي تتوصل إليها الدراسات في العلوم الإنسانية؛ إذ تفتقد معظم تلك الدراسات إلى التراكم؛ أي أن تُبنى نتائج بعضها على بعض، ولذا يجد الباحثون أنفسهم في معظم الأحيان أمام كم كبير من الدراسات التي لا تجمعها نتيجة أو نتائج عامة، مما يقلل من فرص الاستفادة منها (الجهني، ٢٠١٧).

وتهدف بحوث التحليل البعدي إلى تقويم دقيق للدراسات السابقة، وذلك من خلال مجموعة من الإجراءات الإحصائية التي تم تصميمها لتجميع النتائج من خلال الدراسات المستقلة والتي لها نفس التساؤلات تقريباً، حيث يستخدم "التحليل البعدي" ملخص الإحصائيات من الدراسات الفردية كمنطلق للبيانات. فالافتراض الأساسي لهذا التحليل هو أن كل دراسة تمدنا بتقدير مختلف للنتائج داخل العينة. وبتجميع النتائج عبر الدراسات، يمكن تمثيل العلاقات بصورة أكثر وضوحاً من التي نحصل عليها من تقدير الدراسات الفردية، والطريقة الأكثر شيوعاً واستخداماً لدى الباحثين في مجال التربية وعلم النفس هي أنه تُجمع الدراسات المتعلقة بموضوع البحث ثم تُصنف إلى دراسات دالة إحصائياً؛

ودراسات غير دالة إحصائياً. ويركز التحليل البعدي على التقييم وليس على ما إذا كانت بعض الفرضيات مقبولة أو مرفوضة، بل كيف تختلف المخرجات وأحجام التأثير من دراسة إلى أخرى (الجهني، ٢٠١٧؛ سكران، ٢٠٠٦ ؛ محمد، ٢٠٢٠؛ Glass,2000).

وتعتبر مقاييس حجم التأثير Effect Size العملة المشتركة لدراسات "التحليل البعدي"، والتي تُلخص النتائج التي نحصل عليها، وحجم التأثير هو مسمى يُطلق على مجموعة من المؤشرات التي تقيس مقدار تأثير المعالجات والتي تختلف عن اختبارات الدلالة في أنها مستقلة عن حجم العينة (سكران، ٢٠٠٦). وتكشف قيمة حجم التأثير عن المدى الذي يؤثر فيه المتغير المستقل على المتغير التابع سلباً أو إيجاباً في الدراسة (Yildirim et al. 2022) ويرتبط حجم التأثير بالأهمية العملية لنتائج الدراسة وأن الفروق أو العلاقة هي حقيقية وليست نتيجة الصدفة (عبدالفتاح، ٢٠١١).

وكمحاولة من علماء القياس النفسي لتجميع النتائج الامبريقية من الدراسات الفردية بهدف تكامل تلك النتائج وتوليفها وفهمها، فقد استخدم الباحثين التحليل البعدي حيث كان في الآونة الأخيرة احتجاج لإنهاء الاعتماد على التقييمات من خلال الاختبارات في نظام التعليم العام في الولايات المتحدة حيث يُنظر إليها أنها غير عادلة، فكان من الضروري على المختصين في القياس النفسي إثبات إن الاختبارات عادلة وموثوقة وقد استخدموا لإثبات ذلك نتائج الدراسات السابقة التي تمت في مجال الاختبارات، وهذا يؤدي إلى تعزيز وترسيخ الدراسات القائمة على الأدلة في القياس التعليمي، مما يؤدي إلى الابداع والتحسين التدريجي (Rios et al.2020).

ويشير جلاس (Glass, 2000) أن حوالي ٢٥٪ من المقالات في العلوم النفسية كانت في مصطلح عنوان "التحليل البعدي"، حيث انه يتم الاستفادة من التحليل البعدي لتحسين التعليم وكذلك يعتبر من أكثر الطرق فائدة ودقة لتجميع نتائج البحوث. أما في مجال القياس النفسي فقد توصلت دراسة ريوس وأخرون (Rios et al.2020) من خلال مسح الدراسات في مجال القياس النفسي والتربوي ان التحليل البعدي ليس مستخدماً بشكل كبير حيث تم نشر أقل من بحث في التحليل البعدي في المتوسط العام، وقد تم نشر (٢٨ تحليل بعدي بين ١٩٨٦ إلى ٢٠١٦) في هذا المجال، وقد استخدم الباحثون التحليل البعدي في الثلاثة المجالات التالية: تأثير (تصميم- شكل) الاختبار، إعداد الاختبارات- صدق الاختبارات.

وبما انه تعتبر الاختبارات الأداة الأكثر استعمالاً وتداولاً في النظام التربوي وبعض المجالات النفسية أو تكاد تكون الوحيدة، لذا أي قصور في بناء وإعداد هذه الاختبارات سيؤدي حتماً إلى خلل في عملية التقويم التربوي واتخاذ القرارات فيؤثر بدوره على العملية التعليمية، وعليه من الضروري أن تتوفر في هذه الأداة الخصائص والمواصفات التي تجعل منها وسيلة صالحة للتقويم ومن هذه المواصفات: الموضوعية، الثبات، الصدق، الشمول، معامل التمييز، معامل السهولة والصعوبة (اسعادي وَعزي، ٢٠١٥).

مشكلة الدراسة وتساؤلاتها:

هناك العديد من التهديدات التي تواجه جودة البحث التربوي، ومن بينها استخدام أدوات غير صالحة لجمع البيانات (محمود، ٢٠١٩)؛ حيث تعتبر أدوات جمع البيانات ومنها الاختبارات والمقاييس من أهم أدوات القياس في العلوم النفسية والتربوية، والعناية باختيار هذه الأدوات يعد أمراً مهماً لأنه بناء على جودة أداة الدراسة يمكن للباحث أن يتخذ قراراً؛ ويشير (عبدالوهاب، ٢٠٢١) أن استخدام أدوات قياس غير صالحة لجمع البيانات يعد أكثر التهديدات التي تواجه جودة البحث التربوي، مما يتطلب ضرورة مراجعة الإجراءات السيكومترية قبل تعميم نتائج البحوث التربوية؛ حيث أن هناك كثير من الأبحاث في مجال تقنين الاختبارات في مختلف التخصصات سواء النفسية أو التربوية، وكذلك دراسة خصائص الاختبار الجيد وتأثير مختلف المتغيرات عليها.

ويرجع أهمية التحقق من أدوات القياس في العلوم النفسية والتربوية مقارنة بالعلوم الطبيعية، لأنه في العلوم الطبيعية يكون هناك شبه إجماع على نتائج البحوث، وتستخدم مصطلحات محددة التعريف، ووسائل وطرق معيارية محددة، وهذا يقود إلى فهم علمي. بينما هذا ليس متوافراً في العلوم الإنسانية والتربوية، حيث إن السلوك الإنساني يكون دائماً أكثر صعوبة وتعقيداً عند تفسيره، كما أنه لا يوجد اتفاق مشترك على تعريف المتغيرات، كما أن خصائص العينات متباينة بين الدراسات (سكران، ٢٠٠٦).

وهذا من خصائص البحوث في العلوم الإنسانية والتربوية عدم وصولها إلى نتائج متماثلة، ومعنى ذلك أن البحوث التي تُجرى حول موضوع واحد قد لا يدعم بعضها بعضاً. ولعل أكثر من يعانون من هذه المشكلة المسؤولون عن وضع السياسات واتخاذ القرارات العملية حين يريدون الاستناد إلى نتائج هذه البحوث، فيجدون أنفسهم حائرين في طوفان من النتائج المتعارضة (أبو هاشم، ٢٠٠٥). وقد وضحت محمود (٢٠١٩) أن العديد من البحوث التي أُجريت في مجال التربية وعلم النفس تفتقد إلى الواقعية، ويصعب تعميم نتائجها، وتقود إلى نتائج مضللة؛ بسبب الأخطاء المنهجية في تصميمها وإجراءاتها والتحقق من أدواتها.

وكما يرى عيشوش (٢٠١٧) أنه في مجال العلوم الإنسانية والتربوية بشكل عام يعاني الباحثون من وجود كم كبير من الدراسات والبحوث والتي يستعصي أحيانا تجميعها واستنباط النتائج النهائية منها، إذ إننا نحتاج بعد كل مرحلة من الإنتاج البحثي إلى تقييم أثر التراكم المعرفي الحاصل وربطه ببعضه البعض لاستنتاج التوجه العام له وأوجه القصور فيه وفهم الظواهر بشكل أكثر وضوح.

ومن هنا نشأت الحاجة إلى البحوث حول تكامل البحث Research Integration وهي جهود يبذلها فريق من الباحثين يسعون بها إلى إحداث التكامل بين نتائج الدراسات المنفصلة والوصول من ذلك إلى استنتاجات تستوعبها ككل (أبو حطب و صادق، ٢٠١٠). وأحد مداخل البحوث التكاملية هو التحليل البعدي Meta- Analysis

ويعد أسلوباً إحصائياً يطبق على نتائج البحوث والدراسات الكمية أو الرقمية بهدف إحداث تكامل بين نتائجها (أبو هاشم، ٢٠٠٥).

ومن المجالات التي وُجد فيها فيها عدد كبير من الدراسات "المنفصلة" المستقلة، هي الدراسات التي تهتم بتقنين الاختبارات وتحديد أساليب التحقق من خصائص الاختبار الجيد حيث وُجد كثير من الدراسات في هذا المجال إلا أن هناك عدداً من المشكلات التي تواجه الباحثين في التحقق من موثوقية ومصداقية النتائج المستمدة من أدوات القياس، وعلى كثرة الدراسات والأبحاث نجد ان هناك اختلافات في تحديد أفضل أساليب التحقق من خصائص الاختبار الجيد والعوامل المؤثرة عليها. وهنا تأتي أهمية دور بحوث "التحليل البعدي" فتقوم منهجية هذه البحوث على تجمع النتائج والدراسات التي اهتمت بتقنين أدوات القياس ودراسة أفضل أساليب التحقق من خصائص الاختبار الجيد ومختلف العوامل المؤثرة في صحة الاختبارات.

ومن هذه الدراسات دراسة عفانة (٢٠١١) التي هدفت إلى تحديد أخطاء شائعة في تصاميم البحوث التربوية لدى طلبة الدراسات العليا في الجامعات الفلسطينية إلى أنه ليس من السهل على الباحث تحديد نوع الأسلوب الإحصائي الملائم لإيجاد معاملي صدق وثبات أدوات البحث، وخاصة عندما تتعدد تلك الأدوات، إذ أن هناك فرقاً بين الاختبارات والمقاييس، ولكل نوع طرق معينة لإيجاد صدقه وثباته إحصائياً، إذ أن الصدق والثبات يتطلبان شروطاً معينة في ضوء أنواع القياسات المختلفة. وتشير بشتة وَبوعموشة (٢٠٢٠) في دراستهما التي هدفت إلى دراسة الصدق والثبات في البحوث الاجتماعية الى أن البحث يعتمد دائماً على القياس، وأيا كانت الأداة أو الأدوات التي يختارها الباحث عليه أن يراعي المعايير اللازمة لبنائها وصلاحيتها للقياس، إذ هناك خاصيتان مهمتان ينبغي أن تتوفر في أداة القياس وهما الصدق والثبات. فالصدق يعد أهم وأكثر الخصائص السيكومترية أهمية، والذي يشير إلى المدى الذي تقيس فيه الأداة ما يُراد قياسه. أما الثبات فيشير إلى مدى الدقة والاستقرار والاتساق في نتائج القياس، فهو المدى الذي تظل فيه أداة القياس ثابتة في قياس ما يراد قياسه، لو طبقت أداة القياس مرتين فأكثر على نفس الأفراد أو الأشياء لقياس الخاصية في مناسبات مختلفة.

وكما هو مُلاحظ من خلال الاطلاع على الدراسات أن هناك ندرة في البحوث التي تعتمد على مقاييس مقنّنة، فالجميع –بغض النظر عن تخصصه– يقوم بإعداد أدوات قياس خاصة ببحثه (محمود، ٢٠١٩).

ويشير ذبيحي وَشوبار (٢٠١٧) في دراستهما التي هدفت إلى تحديد أخطاء شائعة في البحوث العلمية أن العديد من الباحثين عند التحقق من الخصائص السيكومترية لديهم خلط بين مفهومي الصدق والاتساق الداخلي، كذلك عند التحقق من صدق المقارنة الطرفية يلجأ بعض الباحثين إلى تعيين صدق المقياس من خلال حساب الفروق بين درجات الربع الأعلى من أفراد العينة على المقياس ودرجات الربع الأدنى من أفراد العينة على المقياس دون الرجوع إلى محك خارجي.

وبسبب وفرة الأبحاث المستقلة في مجال تقنين الاختبارات في العلوم الإنسانية والتربوية، والأبحاث التي حددت الطرق المستخدمة للتحقق من خصائص الاختبار الجيد خاصةً، وكما ويرى شاكار (shachar,2008) أنه غالباً ما ينتج عن البحث في مجال العلوم الإنسانية والتربوية نتائج متناقضة بسبب الاختلافات بين الدراسات في خصائص العينة وأدوات القياس وطرق البحث، مما يؤدي إلى نقطة يصعب فيها مقارنة نتائج البحث وهنا تظهر أهمية التحليل البعدي حيث تحاول الباحثة القيام بتجميع النتائج الفردية لهذه الدراسات والمقارنة بينها، وبناء عليه سيتم تحديد أكثر الطرق استخداماً وأفضل أساليب التحقق من خصائص الاختبار الجيد اللازم توافرها في أداة القياس النفسي والتربوي واي من هذه الأساليب أكثر فعالية على حسب نوع الاختبار والمرحلة الدراسية، من خلال الإجابة عن تساؤلات الدراسة، **وتتحدد مشكلة الدراسة في السؤال الرئيس التالي:**

ما مؤشرات التحليل البعدي للعوامل المؤثرة في مواصفات الاختبار الجيد

ويندرج تحت السؤال الرئيس الأسئلة الفرعية التالية:

١. ما مستوى حجم أثر أساليب الصدق (البنائي –التلازمي– الذاتي– المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة؟
٢. هل توجد فروق ذات دلالة إحصائية بين متوسطي حجم الأثر وفقًا لنوع الاختبار (تحصيلي – نفسي) في أساليب الصدق (البنائي – التلازمي– الذاتي– المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة؟
٣. ما مستوى حجم أثر أساليب الثبات (ألفا كرونباخ – التجزئة النصفية– إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة؟
٤. هل توجد فروق ذات دلالة إحصائية بين متوسطي حجم الأثر وفقًا لنوع الاختبار (تحصيلي – نفسي) في أساليب الثبات (الفا كرونباخ– التجزئة النصفية– إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة؟

أهداف الدراسة:

١. تحديد مستوى حجم أثر أساليب الصدق (البنائي –التلازمي– الذاتي– المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة.
٢. التعرف على مستوى الدلالة الإحصائية للفروق بين متوسطي حجم الأثر وفقًا لنوع الاختبار (تحصيلي – نفسي) في أساليب الصدق (البنائي – التلازمي– الذاتي– المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة.

٣.تحديد مستوى حجم أثر أساليب الثبات (ألفا كرونباخ – التجزئة النصفية– إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة.

٤.التعرف على مستوى الدلالة الإحصائية للفروق بين متوسطي حجم الأثر وفقًا لنوع الاختبار (تحصيلي – نفسي) في أساليب الثبات (الفا كرونباخ– التجزئة النصفية– إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة.

أهمية الدراسة:

١–قد تساهم في تدعيم وتكامل البحوث التي أُجريت في تقنين الاختبارات باستخدام أسلوب التحليل البعدي.

٢–من خلال التحليل لنتائج الدراسات السابقة التي أُجريت في هذا المجال والتي اختلفت في نتائجها فقد تساهم هذه الدراسة في تحديد أفضل أساليب التحقق من خصائص الاختبار الجيد على حسب نوع الاختبار (تحصيلي – نفسي) والمرحلة الدراسية (ابتدائي– متوسط– ثانوي– جامعي).

٣–قد تسهم في توثيق حركة البحث العلمي في تقنين الاختبارات خاصة في ضوء كثرة عدد الأبحاث التي قامت بتقنين الاختبارات وتحديد أفضل أساليب التحقق من الصدق والثبات.

٤–قد تسهم في توعية الباحثين إلى ضرورة تضمين الدراسات والبحوث مؤشرات عن حجم التأثير وفقاً لتعليمات الجمعية الأمريكية APA

مصطلحات الدراسة:

التحليل البعدي: يعرفه جلاس (Glass) بأنه الأسلوب الكمي المنظم للخروج بالمعلومات وعرضها من بين نتائج البحوث العديدة السابقة في مجال ما وبطريقة موضوعية، بما يساعد في الحكم على فاعلية نتائج هذه البحوث (محاسنة و الشريفين، ٢٠٢٠) . **ويعرف إجرائياً بأنه:** أسلوب كمي إحصائي منظم لتنظيم وتلخيص واستخراج المعلومات من كم هائل من البيانات التي توصلت إليها مجموعة البحوث التي تناولت تقنين الاختبارات وخصائص الاختبار الجيد، بهدف التوصل إلى تحديد خصائص الاختبار الجيد.

خصائص الأداة الجيدة: إجراءات التحقق من الأدوات وتوفر الخصائص السيكومترية مما يجعلها أداة صالحة للقياس. **تُعرف إجرائياً بأنها:** مجموعة أساليب التحقق من خصائص الاختبار الجيد والتي سيتم تحديدها من الأبحاث عينة الدراسة والتي سيتم ترميزها وتفريغها في النموذج التي أعدته الباحثة وسيقتصر على الطرق التالية: الصدق البنائي– صدق المقارنة الطرفية– الصدق التلازمي– الصدق الذاتي– ثبات الفا كرونباخ– الثبات بطريقة التجزئة النصفية– الثبات بطريقة إعادة التطبيق.

الدلالة العملية Practical Significance: مقدار التباين في المتغير التابع الذي يمكن تفسيره من خلال متغير مستقل يرتبط به أو يؤثر فيه (الشريفين، ٢٠١٧). **ويعرف إجرائياً:**

بإنها قيمة الدلالة العملية (حجم الأثر) الذي توصلت إليه الدراسات عينة البحث ويُحسب عن طريق معادلة كوهين.

حدود الدراسة:

اقتصرت الدراسة على:

- الرسائل والأبحاث في البيئة العربية التي تمت خلال الفترة من ١٩٩٠- ٢٠٢١ .
- الأبحاث التي استخدمت الأدوات السيكومترية (الاختبارات- الاستبيانات) لقياس المتغيرات التربوية، مع استبعاد الأدوات الأخرى كالبرامج التدريبية أو التعليمية أو العلاجية أو الإرشادية.
- الأبحاث التي استخدمت الأدوات السيكومترية التي اعتمد بناؤها وتقنينها على مبادئ نظرية القياس التقليدية في تحليل الأدوات والمفردات، وذلك وفق مدخل القياس معياري المرجع.
- أساليب التحقق من خصائص الاختبار الجيد الصدق (الصدق البنائي – الصدق التلازمي- الصدق الذاتي- صدق المقارنة الطرفية)، والثبات (الفا كرونباخ – التجزئة النصفية- إعادة التطبيق)
- الدراسات التي تشتمل على البيانات الإحصائية التالية: (متغيرات الدراسة، عدد عينة الدراسة- المتوسط الحسابي- الانحراف المعياري- قيمة الاختبارات الإحصائية) واستبعدت الدراسة الحالية الدراسات والبحوث التي لم تقدم البيانات الإحصائية اللازمة لحساب حجم الأثر وللتحليل.

الإطار النظري:

أولاً: التحليل البعدي Meta Analysis:

كلمة Meta في أصلها الاغريقي تعني: ما وراء، بعد. ومعنى تقنية تحليل الميتا أو كما يسميه البعض بالتحليل التلوي، أو البعدي أو التجميعي أو تحليل التحليلات هو: تقنية إحصائية للجمع بين نتائج مجموعة من الدراسات التي تعالج الفرضيات البحثية المشتركة أي يشير إلى التحليل الإحصائي لمجموعة عريضة من نتائج التحليلات الفردية لغرض دمج النتائج، حيث تصبح وحدة التحليل هنا هي الدراسات نفسها (عيشوش، ٢٠١٧).

ويهدف أسلوب التحليل البعدي إلى استقراء تعميمات دقيقة من البيانات والنتائج، عن طريق التقييم الدقيق للدراسات الأولية والتي تشتمل بياناتها على نتائج كمية، واستخدام الأساليب الإحصائية المناسبة التي تعمل على تجميع النتائج المشتركة فيما بينها وتلخيصها لاستنتاج نتائج كلية (مهدي، ٢٠١٩).

وتساعد نتائج البحوث المجمعة على تطوير وبناء النظريات باعتباره أحد الأهداف المهمة للعلم في جميع المجالات. إذ يحاول الباحثون أولاً العثور على العلاقات الأمبيريقية التي يمكن الكشف عنها في هذه الدراسات، حتى يمكن أن يأخذوها في الاعتبار عند بناء النظرية (أبو علام، ٢٠١١).

والاستعمال الأكثر شيوعاً للتحليل البعدي يتمثل في المراجعات الأدبية الكمية. حيث يختار الباحث التأثير أو المستكشف البحثي تحت عدد كبير من الظروف المختلفة، وهنا يستخدم التحليل البعدي لوصف القوة العامة للتأثير، وتحت أية ظروف يقوى أو يضعف هذا التأثير (سكران، ٢٠٠٦).

وكون أسلوب التحليل البعدي أسلوباً كمياً، لذلك يعتمد اعتماداً كبيراً على الإعداد والطرق الإحصائية في شكل عملي خاص لتنظيم واستخراج المعلومات من كم هائل من البيانات والنتائج والتي يصعب استخراجها بأساليب أخرى. ويشترك أسلوب التحليل البعدي مع الدراسات المسحية Survey Studies في كثير من خصائصها، فهو عملية لمسح وتحليل التجمع الهائل لنتائج الدراسات في مجال معين بطريقة كمية ويختلف عنها في كون الدراسات المسحية تفتقر لإمكانية التوصل للعلاقة السببية بين المتغيرات المستقلة والتابعة (عبدالحميد، ١٩٨٧).

ويمكن القول بأن أسلوب التحليل البعدي نشأ كرد فعل لحاجة بحوث التربية بوجه عام وبحوث التربية العلمية بوجه خاص إلى طريقة مقننة لتحليل التحليل الذي أدت إليه هذه البحوث بحيث يؤدي ذلك إلى التعرف على فاعلية هذه البحوث وبالتالي إمكانية التوصل إلى اتخاذ قرار معين في تبني نتائج هذه البحوث من عدمه (عبدالحميد، ١٩٨٧). كما يهدف أيضاً لتحديد حجم التأثير وهو مقياس لقوة العلاقة بين المتغيرات واتجاهها من خلال تجميع البيانات من دراسات متعددة (مهدي، ٢٠١٩).

طرق التحليل البعدي:

هناك العديد من الطرق التي استخدمها الباحثون للتحليل البعدي، كما ذكرها محمد (٢٠٢٠) من أهمها:

١- **طريقة جلاس Glass:** وفي هذه الطريقة يتم جمع كل الدراسات المرتبطة بموضوع أو مشكلة محددة، ثم يتم تحويل نتائج هذه الدراسات إلى وحدة قياس مشتركة لكي يُمكن إجراء مقارنات بين هذه الدراسات بعد تقسيمها وبالتالي فإنه من الممكن استخراج أكثر من نتيجة من كل دراسة. فمثلاً عندما يستخدم باحث في إحدى الدراسات أربعة مقاييس لقياس المتغير، فهنا تم اعتبارها أربع نتائج ويتم التعامل معها على هذا الأساس، ويشير Guzzo et. al (١٩٨٧) انه يمكن إيجاد حجم التأثير لهذه الطريقة عن طريق المعادلة التالية:

$$\text{حجم التأثير (d)} = \frac{\text{متوسط المجموعة الضابطة} - \text{متوسط المجموعة التجريبية}}{\text{الإنحراف المعياري للمجموعة الضابطة}}$$

٢- **طريقة مانسفيلد وبوس Mansfield and Busse:** وتسمى بطريقة التحليل البعدي المبني على أثر الدراسة وليس على نتائجها، حيث كان من الانتقادات التي وجهت لطريقة جلاس أن الدراسة الواحدة تُمثل أكثر من مرة وهكذا يبدو أن عدد الدراسات كبير جداً مع أنه في الواقع أقل بكثير، وتتضح هذه المشكلة بشكل أكبر إذا

ما قسمت الدراسات بناء على تصنيف معين، وأُدرج تحت أحد التصنيفات عدد من النتائج الخاصة بدراسة أو دراستين فقط حيث يكون من الصعب تعميم نتائج التحليل البعدي. وتحاول هذه الطريقة البديلة التغلب على هذه المشكلة من خلال حساب حجم أثر واحد فقط لكل دراسة، وفي حالة وجود أكثر من مقياس للمتغير التابع يقوم الباحث إما بجمع أحجام التأثير واستخراج المتوسط الوزني لهذه الأحجام إذا كانت المقاييس تقيس متغيراً واحد، ولكن إذا كانت تقيس متغيرات مختلفة يتم إجراء أكثر من تحليل بعدي وفقاً لعدد المتغيرات المقاسة.

٣- **طريقة ستوفر Stouffer:** وتعتمد هذه الطريقة على جمع مستويات الدلالة لكل دراسة. حيث يقوم الباحث بتسجيل كل مستوى دلالة خاص بفرضية معينة، ثم يستخرج Z Score المقابلة لكل مستوى دلالة، ثم يجمع درجات Z ويقسم الناتج على الجذر التربيعي لعدد الدراسات المستخدمة في التحليل، وأخيراً يستخرج مستوى الدلالة المقابل لدرجة Z. ويستخدم ذلك كمستوى دلالة لكل الدراسات التي شملها التحليل.

٤-**طريقة هدجز وأولكن Hedges and Olkin:** وفي هذه الطريقة يقوم الباحث بحساب حجم التأثير لكل دراسة كما في الطرق السابقة، ولكن أهم ما تتميز به هذه الطريقة عن الطرق الأخرى هو تأكيدها على ضرورة بحث التجانس بين المتغيرات المستخدمة في التحليل البعدي قبل جمعها، والخطوة الأولى في جمع أحجام التأثير المستخرجة من عدد من الدراسات هي تحديد ما إذا كانت أحجام التأثير هذه متجانسة أم لا، فإذا ما تبين للباحث أن أحجام التأثير غير متجانسة قام بتقسيمها إلى مجموعات بناء على أسس نظرية أو عملية، ثم يقيس التجانس داخل هذه المجموعات، وقد يضطر أيضاً إلى تقسيم هذه المجموعات إلى مجموعات أصغر حتى يحقق التجانس.

٥-**طريقة هنتر وشميدت Hunter and Schmidt:** وفي هذه الطريقة يتم تحويل نتائج كل دراسة إلى أحجام أثر ثم يجمع أحجام التأثير، ثم يحاول التخلص من الأخطاء المرتبطة بالعينة أو أسلوب القياس المستخدم فإذا ما اتضح بعد عملية التصحيح أن حجم التباين المتبقي كبير نسبياً اعتبر ذلك مؤشراً على احتمال وجود متغيرات وسيطة قد تفسر وجود التباين.

خطوات أسلوب التحليل البعدي

هناك عدد من الخطوات تتطلبها منهجية التحليل البعدي وهي كما ذكرها عبدالحميد، فرماوي وَ محاسنة (١٩٨٧؛ ٢٠٠٣؛ ٢٠٢٠):

١-**تحديد بؤرة الاهتمام والتركيز على الهدف:** يبدأ الباحث باختيار الموضوع الذي يود استعراض الدراسات والبحوث التي تمت حوله، ويسمي هذه المرحلة ريوس واخرون (Rios et al.2020) بمرحلة صياغة المشكلة حيث تعد أهم مرحلة في التحليل

البعدي لأنها تؤثر على العديد من القرارات مثل: استراتيجيات البحث، معايير تضمين/ استبعاد الدراسة، واختيار المتغيرات لاستخراج البيانات

٢-**البدء بجمع البحوث والدراسات السابقة التي لها علاقة ببؤرة الاهتمام:** في ضوء تحديد الموضوع يتم اختيار الدراسات التي تتفق مع هذا الموضوع وذلك من خلال العديد من المصادر مثل أبحاث الدوريات والرسائل العلمية وابحاث المؤتمرات والبحوث المنشورة بالكتب التجميعية.

٣-**تحديد معايير تضمين واستبعاد الدراسات:** وفي هذه المرحلة يتم تحديد المعايير التي على أساسها سيتم اختيار الدراسات التي سوف يتضمنها التحليل وتعتمد هذه المعايير على بعض العوامل ومنها طبيعة الأسئلة البحثية والتعريف الإجرائي الذي تبناه الباحث للمصطلحات الأساسية بالدراسة ومتطلبات مدخل التحليل البعدي الذي سوف يستخدمه.

٤-**فحص الدراسات والبحوث:** في هذه الخطوة يضع الباحث تعريف إجرائي لمجال التركيز وفي ضوئه يطابق مدى علاقة كل دراسة بمجال الاهتمام.

٥-**تحديد خصائص الدراسات:** حيث يتم تحديد المظاهر الرئيسية للدراسة وفقاً لبعض العوامل ومنها: القياسات التي استخدمت- نوع التصميم البحثي للدراسة- التعريفات الإجرائية لمتغيرات الدراسة- الجودة المنهجية للدراسة- المتغيرات التابعة- عدد أفراد العينة وأعمارهم وجنسهم- التحليلات الإحصائية المستخدمة. وهناك العديد من العوامل التي يمكن تصنيف الدراسات في ضوئها.

٦-**إعداد ملف لتفريغ البيانات وتبويبها:** حيث يقوم الباحث بجدولة البيانات السابقة ثم يقوم بتبويبها وفقاً لتلك المتغيرات. ويرى ريوس واخرون (Rios et al.2020) أهمية هذه المرحلة في تصميم استمارة بها معلومات تصف معايير الاهلية لأدراج الدراسات ويجب أن يتم تناول المتطلبات التالية: (أ) تحديد خصائص الدراسات، (ب) تصاميم البحث المؤهلة، (ج) خصائص العينة، (د) البيانات الإحصائية المطلوبة، (هـ) الحدود الجغرافية والزمنية. حيث تساعد تحديد المعلومات المطلوبة في التقليل من الذاتية وزيادة صدق البيانات وتصنيف الدراسات

٧-**الترميز:** يتم الترميز في الاستمارة التي أعدها الباحث ويتطلب الترميز دقة شديدة من الباحث وأن يهتم بالتفاصيل ويجب أن يكون الترميز موثوقاً لضمان صحة النتائج (Pigott & Polanin, 2020)

٨-**إيجاد حجم التأثير:** تختلف مقاييس حجم التأثير باختلاف تصميم الدراسة ونوع البيانات، ويعد كوهين (Cohen) أول من حسب حجم التأثير، وكان ذلك في عام ١٩٦٧م، ثم عدّل عليه جلاس (Glass) عام ١٩٧٦م. وتعد الصيغة التي اقترحها

هيدجز (Hedges) أكثر استخداماً لا يجاد متوسط الفروق المعيارية للتخلص من التحيز الناجم عن حجم العينة

٩-**تطبيق اختبار التجانس:** من الضروري إجراء اختبار التجانس لقيم حجوم التأثير، فهل الدراسات متشابهة أو متجانسة في نطاق الموضوع أو في النتائج؟ وبالتالي يجري تحديد النموذج المراد تطبيقه وفقاً للدراسة، وتوفر البرامج الإحصائية نوعين من النماذج هما:

- **نموذج التأثيرات الثابتة:** يفترض أن جميع الدراسات لها حجم أثر واحد فقط، وأن الاختلافات ناجمة عن الخطأ في اختيار عينة الدراسة، أو قد يعود ذلك لأن عدد الدراسات قليل.
- **نموذج التأثيرات العشوائية:** ويفترض هذا النموذج أن حجم التأثير الفعلي قد يختلف من دراسة إلى أخرى بسبب متغيرات مثل: العمر، والمستوى التعليمي، وحجم العينة خصوصاً عندما تكون عينة الدراسة أقل من (٢٠) فرداً. ويجري تحديد النموذج المستخدم وفقاً لقيمة اختبار كوكران (Q)، ومقارنته بقيمة اختبار كاي تربيع بدرجات حرية (df= k−1)، وتمثل (k) عدد الدراسات المتضمنة في التحليل.

١٠- **مقارنة حجوم التأثير** : لبناء تصور عام عن الدراسات المشمولة فإنه تتمثل الخطوة المهمة في وصف توزيع أحجام التأثير من الدراسات المشمولة عن طريق الرسوم البيانية، الذي يعرض أحجام التأثير بثقة ٩٥٪ من كل دراسة Pigott & (Polanin, 2020)

ثانياً: الدراسات السابقة

تعد دراسات التحليل البعدي في مجال تقنين الاختبارات على حد علم الباحثة قليلة نسبياً سواء على المستوى العربي والأجنبي، وتكاد تكون معدومة على المستوى العربي، حيث وُجدت دراسة واحدة فقط استخدمت التحليل البعدي لمعاملات ثبات درجات اختبار المصفوفات المتتابعة أما باقي الدراسات اقُتصرت على الدراسات التي تتعلق بدراسة الخصائص السيكومترية وأكثر الطرق تكراراً في الأبحاث، والدراسات التي تناولت الأخطاء الشائعة في بناء الاختبار ويُستفاد منها في معرفة أهم الأخطاء التي تكررت في بناء الاختبارات والتي من الممكن أن تؤثر على نتائج صدق وثبات الاختبارات.

- **دراسات في التحليل البعدي**

قام الهنائي (٢٠١١) بدراسة هدفت إلى التحليل البعدي لمعاملات ثبات درجات اختبار المصفوفات المتتابعة لريفن والاخطاء المعيارية للقياس المناظرة لها، حيث استخدم الباحث (٥٦) دراسة من الدراسات التي تناولت اختبارات المصفوفات المتتابعة لريفن، استخرج منها (١٤٣) معامل ثبات، منها (١١٧) معامل ثبات بطريقة الاتساق الداخلي

و(٢٦) معامل ثبات بطريقة إعادة التطبيق، كما استخرج منها (١٢٦) خطأً معيارياً للقياس. وباستخدام طريقة فاشا- هاس للتحليل البعدي للثبات، توصلت نتائج الدراسة إلى أن القيمة التقديرية لمعامل ثبات درجات اختبارات المصفوفات المتتابعة لريفن بلغت (٠,٨٥) بطريقة الاتساق الداخلي، و (٠,٧٦) بطريقة إعادة التطبيق، في حين بلغت القيمة التقديرية للخطأ المعياري للقياس (٢,٥١) في طريقة الاتساق الداخلي، و (٣,٤٣) في طريقة إعادة التطبيق. وقد بينت النتائج أنه لا توجد فروق ذات دلالة إحصائية عند مستوى أقل من (٠,٠٥) في معاملات ثبات الاتساق الداخلي والأخطاء المعيارية للقياس تعزى إلى نوع العينة، بينما وجدت فروق ذات دلالة إحصائية عند مستوى أقل من (٠,٠٥) في معاملات ثبات الاتساق الداخلي والأخطاء المعيارية للقياس تعزى إلى بيئة التطبيق، والفئة العمرية، ونوع الاختبار المستخدم في الدراسات التي تناولت اختبارات المصفوفات المتتابعة لريفن.

- **دراسات في الخصائص السيكومترية للأدوات**

سعى باهي وَفياض (٢٠٠٨) إلى التعرف على أكثر طرق التحقق من صدق وثبات الاختبارات والمقاييس، ولتحقيق هدف الدراسة استخداما (٨٤٥) بحثاً علمياً، بحيث تعتمد هذه البحوث على أدوات قياس يلزم التحقق من صدقها وثباتها، وقد توصلا إلى وجود أخطاء كثيرة في استخدام في استخدام المصطلحات العملية المتعلقة بالصدق والثبات، كما توصّلا إلى أن أكثر طرق الصدق استخداماً هي طريقة الاتساق الداخلي بنسبة (٢٥.١٦٪)، وأقلها استخداماً طريقة الصدق المتعلق بمحك بنسبة (٠،٢٧٪) كما كانت أكثر طرق الثبات استخداماً طريقة التطبيق وإعادة التطبيق بنسبة (٨٣.٢٦٪) وأقلها استخداماً طريقة التجانس الداخلي بنسبة (٠،٢٧٪).

وبعد إجراءات التطبيق حسبت قيم معاملات صعوبة الفقرات وتمييزها في نموذجي الاختبار، وكذلك معاملات صدقها وثباتها. فأظهرت النتائج أن الفرق كان دالاً إحصائياً بين صعوبة وتمييز الفقرات التي لم تحتو المخالفات في قواعد الصياغة، أو التي احتوتها، وبينت النتائج أن الفرق بين قيم معاملي ثبات الاتساق الداخلي في نموذجي الاختبار كان دالاً إحصائياً، وكان الفرق بين قيم معاملي صدقها دالاً أيضاً.

قام حسن (٢٠١٤) بدراسة هدفت إلى إلقاء الضوء على أهم مفهومين من مفاهيم القياس وهما: الثبات والصدق بهدف التعرف على الطرق الشائعة لحساب كل منهما، وكيفية تحسين كل من الثبات والصدق وقد توصلت نتائج الدراسة انه يجب استخدام أكثر من طريقة لحساب ثبات أدوات القياس وعدم الاكتفاء بطريقة واحدة؛ لأن طرق حساب الثبات تختلف في الكشف عن مصدر الخطأ في الاختبار. كما يجب حساب الصدق بطرق تغطي جميع مكونات الصدق كلما كان ذلك ممكناً، وكذلك أشار الباحث انه على الباحثين أن يكونوا على وعي تام بأن استخدام أداة قياس غير صادقة يعني لا قياس.

في حين قامت محمود (٢٠١٩) بدراسة هدفت 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 الفترة ٢٠١٢-٢٠١٦، وبلغ عدد أدوات القياس المستخدمة في هذه البحوث ٩٢ أداة، كان ٨٥٪ منها من إعداد الباحثين، وقد تم رصد ١٢ خطأً متكرراً في إجراءات التحقق من الثبات، و٩ 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٪. وأوصت الدراسة بإعداد مقياس علمي لتقدير دقة إجراءات التحقق من صدق وثبات أدوات القياس في البحوث التربوية.

وسعى عبدالوهاب (٢٠٢١) في دراسته إلى تقييم إجراءات القياس النفسي في عينة من تقارير البحوث المنشورة في المجلات التربوية العربية المحكمة، ولتحقيق هذا الهدف تم بناء بطاقة تقييم تكونت من (٩٥) مفردة، تمثل كل مفردة منها إجراءً سيكومترياً وتم التحقق من صدقها وثباتها بالطرق المناسبة، ثم استخدامها في تقييم (١٤٧) أداة بحث في (٩٥) بحثاً منشوراً في أربع مجلات تربوية في عام ٢٠١٩م، وأسفرت النتائج على أن أكثر الطرق استخداماً في التحقق من صدق أدوات القياس هي الصدق القائم على المحكمين، وأقلها استخداماً طريقة الصدق المرتبط بمحك وطريقة التمايز العمري، كما كانت أكثر الطرق استخداماً في التحقق من ثبات الأدوات طريقة معامل ألفا كرونباخ، وأقلها استخداماً طريقة الصور المتكافئة، وتراوحت نسب تحقق الإجراءات السيكومترية في بطاقة التقييم بشكل عام بين (صفر-١٠٠٪)، وكانت نسبة الأخطاء الشائعة في تطبيق الإجراءات السيكومترية في تقارير البحوث التربوية بشكل عام (٦٢.١١٪)

كما قام طباع وَ خنيش (٢٠٢١) بدراسة هدفت إلى تقديم نظرة عن بعض الأخطاء الناجمة عن جمع أدلة الصدق والثبات في البحوث التربوية والنفسية، ومن الأخطاء الناجمة عن جمع أدلة الصدق منها صدق الاتساق الداخلي يتم حسابه مرتين مره تحت مسمى الصدق وهو تقدير الارتباط بين العبارات ببعدها والارتباط بين البعد والاختبار ككل، ومرة أخرى تحت مسمى الثبات باستخدام معادلة ألفا وهذا يعتبر تكرار لقياس الثبات مرتين، وهناك ممارسة أخرى شائعة بين الباحثين في استخدام طريقة الصدق الداخلي،

عندما يتم تقدير معاملات ارتباط البنود بالبُعد أو بالدرجة الكلية للأداة، ويتم الحكم على صدق البنود على الدلالة الإحصائية باستبعاد كل عبارة غير دالة إحصائياً فهذا يعتبر تهديداً للصدق البنائي الذي يؤدي إلى عدم تغطية أبعاد السّمه المقُاسة. ومن الأخطاء الناجمة عن جمع أدلة الثبات انتهاك افتراضات استخدامات الفا كرونباخ، فمعامل ألفا يتطلب أن تكون العبارات متجانسة الأبعاد (أي تتضمن بعدًا أو عاملًا واحدً)، ويكون تباين الدرجات الحقيقية للعبارات متساوية، حيث لا تختلف إلاّ بمقدار ثابت، ويطغى على تقديرات الثبات في البحوث النفسية والتربوية معامل ألفا وطرق التجزئة النصفية، والصيغ المتكافئة وهذا يعود إلى سهولة استخدامها من جهة وإمكانية تقديرها باستخدام البرامج الإحصائية من جهة أخرى ومن الأخطاء الشائعة عدم مراعاة الفترة الفاصلة بين الاختبارين خاصة إذا كانت السّمة المقاسة متغيرة نسبياً، فاختبارات التحصيل مثلاً تتطلب أن لا تكون الفترة الفاصلة طويلة حتى لا يحدث نضجاً لدى الأفراد المختبرين، ولا تكون قصيرة فيحدث تذكّر للإجابات المقدمة في الفترة الأولى، في حين أن مقاييس الدافعية واستراتيجيات التنظيم الذاتي والاتجاهات على سبيل المثال سمات ثابتة نسبياً يمكن أن تكون الفترة الفاصلة بين التطبيقين أكبر.

التعليق على الدراسات السابقة وأوجه الاستفادة منها:

يتضح من الدراسات السابقة انها تناولت بشكل عام أكثر الطرق استخداما في التحقق من الصدق والثبات، وكذلك بعض الأخطاء الشائعة التي تؤثر على خصائص الأداة السيكومترية، وبعض الأخطاء الشائعة في بناء الاختبارات التي تؤثر في خصائص الاختبار الجيد وسيتم مناقشة نتائج الدراسات السابقة على حسب ما يفيد الباحثة في تفسير تساؤلات الدراسة، حيث استخلص من نتائجها:

١-أظهرت دراسة باهي وَفياض؛ ومحمود (٢٠٠٨؛ ٢٠١٩) أن أكثر طرق التحقق من الصدق استخداماً هي طريقة الاتساق الداخلي

٢- أظهرت دراسة باهي وَفياض؛ وَ عبدالوهاب (٢٠٠٨؛ ٢٠٢١) أن أقل طرق التحقق من الصدق استخداماً هي طريقة الصدق المرتبط بالمحك

٣- أظهرت الدراسات محمود؛ عبدالوهاب وَ طباع وخنيش (٢٠١٩؛ ٢٠٢١؛ ٢٠٢١) أكثر طرق الثبات تكراراً هي طريقة ألفا كرونباخ (الاتساق الداخلي).

٤- توصلت دراسة باهي وَفياض (٢٠٠٨) إلى وجود أخطاء كثيرة في استخدام المصطلحات المتعلقة بالصدق والثبات

٥- وضحت دراسة طباع وَخنيش (٢٠٢١) الأخطاء الناجمة عن جمع أدلة الصدق وتتمثل في:

- الاتساق الداخلي يتم حسابه مرتين، مرة تحت مسمى الصدق وهو تقدير الارتباط بين العبارات ببعدها والارتباط بين البعد والاختبار ككل، ومرة أخرى تحت مسمى الثبات باستخدام معادلة ألفا وهذا يعتبر تكرار لقياس الثبات مرتين
- في استخدام طريقة الاتساق الداخلي، عندما يتم تقدير معاملات ارتباط البنود بالبُعد أو بالدرجة الكلية للأداة، ويتم الحكم على صدق البنود على الدلالة الإحصائية باستبعاد كل عبارة غير دالة إحصائياً فهذا يعتبر تهديداً للصدق البنائي الذي يؤدي إلى عدم تغطية أبعاد السّمه المقُاسة

٦- وضحت دراسة طباع وَخنيش (٢٠٢١) الأخطاء الناجمة عن جمع أدلة الثبات وتتمثل في:

- انتهاك افتراضات استخدامات الفا كرونباخ، فمعامل ألفا يتطلب أن تكون العبارات متجانسة الأبعاد (أي تتضمن بعدًا أو عاملًا واحدً)، ويكون تباين الدرجات الحقيقية للعبارات متساوية، حيث لا تختلف إلاّ بمقدار ثابت
- عدم مراعاة الفترة الفاصلة بين الاختبارين خاصة إذا كانت السّمة المقاسة متغيرة نسبياً

وبناء على ما سبق، يتضح أن الدراسات في التحليل البعدي شبه معدومة، واستعاضت الباحثة بالدراسات التي حددت معايير الاختبار الجيد وكذلك الدراسات التي تناولت الأخطاء الشائعة في التحقق من الخصائص السيكومترية، لذلك فإن الدراسة الحالية ستقوم بالتحليل البعدي لعينة الدراسة ومن ثم حساب حجم تأثير كل نوع ولعدم وجود أبحاث مطابقة لشغل الدراسة فإنه سوف يُستفاد من النتائج السابقة في التعليق على النتائج وتفسير حجم التأثير لكل نوع.

منهج وإجراءات الدراسة:

منهج الدراسة:

استخدمت الباحثة منهج التحليل البعدي Meta Analysis طريقة جلاس وهو منهج وصفي تحليلي يهدف إلى استخلاص النتائج الكامنة وراء عدة نتائج مستمدة من دراسات فردية ذات خصائص محددة.

مجتمع الدراسة:

يتكون مجتمع الدراسة الحالية من الدراسات العربية التي قامت بتقنين الاختبارات في المجال النفسي والتربوي، والمنشورة خلال الفترة الممتدة بين (١٩٩٠- ٢٠٢١ م)

عينة الدراسة:

تكونت عينة الدراسة من (١١٥) دراسة في خصائص الاختبار الجيد تم اختيارها بطريقة عشوائية بسيطة، منها (٣٥) اختباراً تحصيلياً وَ(٨٠) اختباراً نفسياً، ووفقاً لمعايير القبول والاستبعاد التي وضعتها الباحثة تم استبعاد (٣٦) دراسة بعد ترميز البيانات، وبالتالي تكونت العينة النهائية للدراسة من (٧٩) دراسة في خصائص الاختبار الجيد منها (١٦)

اختباراً تحصيلياً وَ(٦٣) اختباراً نفسياً، والتي احتوت على بيانات كافية لحساب حجم التأثير للأساليب الإحصائية (الصدق البنائي، الصدق التلازمي، الصدق الذاتي، صدق المقارنة الطرفية، ثبات ألفا كرونباخ، التجزئة النصفية، إعادة التطبيق)، وقد بلغ عدد حجوم التأثير للأساليب الإحصائية في عينة الدراسة (٢٥٠) حجم أثر، كما هو موضح في جدول(١)

جدول (١) توزيع الأساليب الإحصائية المشمولة في العينة على حسب نوع الاختبار (تحصيلي – نفسي)

خصائص الاختبار الجيد	الأساليب الإحصائية	عدد حجوم التأثير حسب نوع الاختبار		مجموع حجوم التأثير لكل متغير
		تحصيلي	نفسي	
الصدق	الصدق البنائي	١٠	٤٥	٥٥
	الصدق التلازمي	٤	١٨	٢٢
	الصدق الذاتي	١	٥	٦
	صدق المقارنة الطرفية	٣	٢٨	٣١
الثبات	ألفا كرونباخ	١٠	٤٩	٥٩
	التجزئة النصفية	٨	٣٦	٤٤
	إعادة التطبيق	٩	٢٤	٣٣
المجموع		٤٥	٢٠٥	٢٥٠

معايير قبول واستبعاد الدراسات:

١-**<u>زمن إجراء البحث أو الدراسة:</u>** عدم ارتباط تجميع الدراسات بسنوات محددة للنشر، بل حرصت الباحثة على تجميع أكبر قدر ممكن من الدراسات في مجال تقنين الاختبارات وتحديد خصائص الاختبار الجيد، وتراوحت تلك الدراسات التي تم تجميعها في الفترة ١٩٩٠– ٢٠٢١ م في البيئة العربية.

٢-**<u>النطاق الموضوعي:</u>** اهتمت الدراسة الحالية بتحليل البحوث والدراسات التي تناولت تقنين الاختبارات في المجال النفسي والتربوي والتي تناولت الأساليب الإحصائية (الصدق البنائي، الصدق التلازمي، الصدق الذاتي، صدق المقارنة الطرفية، ثبات ألفا كرونباخ، التجزئة النصفية، إعادة التطبيق)

٣-**<u>اكتمال البيانات اللازمة للتحليل:</u>** أن تشتمل الدراسات على البيانات الإحصائية التالية: (متغيرات الدراسة، عدد عينة الدراسة– المتوسط الحسابي– الانحراف المعياري– قيمة الاختبارات الإحصائية) استبعدت الدراسة الحالية الدراسات والبحوث التي لم تقدم البيانات الإحصائية اللازمة للتحليل.

الأساليب الإحصائية:

١. استخدام برنامج اكسل Excel لتبويب البيانات
٢. التكرارات والنسب المئوية بواسطة برنامج SPPS

٣. اختبار Z للفروق بين المتوسطات

٤. استخدام معادلات تحويل الأساليب الإحصائية إلى حجم التأثير

(Rosenthal et al. 1994 ; Thalheimer & Cook .2002; Nakagawa & Cuthill. 2007; Ferguson. 2016)

جدول (٢) معادلات تحويل الأساليب الإحصائية إلى حجم التأثير

Hedges's g		Cohen's d		
from r	from t	from r	from t	
$g = \frac{r}{\sqrt{1-r^2}}\sqrt{\frac{df(n_1+n_2)}{n_1 n_2}}$	$g = \frac{t\sqrt{n_1+n_2}}{\sqrt{n_1 n_2}}$	$d = \frac{2r}{\sqrt{1-r^2}}$	$d = \frac{t(n_1+n_2)}{(\sqrt{df})(\sqrt{n_1 n_2})}$	مجموعتين مستقلتين
$g = \frac{r}{\sqrt{1-r^2}}\sqrt{\frac{df}{n}}$	$g = \frac{2t}{\sqrt{N}}$ $N = (n_1+n_2)$		$\frac{2t}{\sqrt{df}} = d$	مجموعتين مرتبطتين

٥. استخدام نموذج التأثير العشوائي (Random–Effect Models) كأحد النماذج الإحصائية للتحليل البعدي

٦. حساب حجم التأثير باستخدام برنامج Comprehensive Meta Analysis V.2

٧. تفسير قيم حجم التأثير كما يرى كوهن في (Decoster,2009) أنه إذا كانت القيمة المحسوبة لحجم تأثير المتغير المستقل على المتغير التابع = ٠.٢ فأن حجم التأثير يكون ضعيفاً أو صغيراً، أما إذا كانت = ٠.٥ فتدل على حجم تأثير متوسط ، وإذا كانت = ٠.٨ فتدل على حجم تأثير مرتفع.

نتائج الدراسة :

<u>التساؤل الأول:</u> " ما مستوى حجم تأثير أساليب الصدق (البنائي –التلازمي– الذاتي– المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة "؟

للإجابة عن هذا التساؤل استخدمت الباحثة نموذج التأثير العشوائي (Random–Effect Models) كأحد النماذج الإحصائية للتحليل البعدي عن طريق استخدام البرنامج الإحصائي (Comprehensive Meta – Analysis 2)، وجاءت النتائج كما يوضحها جدول (٣)

جدول (٣) متوسطات أحجام أثر أساليب الصدق (البنائي- التلازمي- الذاتي- المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة

المتغير نوع الصدق	عدد حجوم التأثير	عدد المفحوصين	Q	I^2	فترة الثقة	حجم التأثير الكلي d	مستوى حجم التأثير
البنائي	٥٥	٢٤٣٠٧	٢٠٨٤,٥	٩٧,٤	٠,٦٤٧- ٠,٧٣٣	٠,٦٩٢	متوسط
التلازمي	٢٢	١٠٩٥٠	١٢١٧,٣	٩٨,٣	٠,٥٩٨- ٠,٧٧٣	٠,٦٩٦	متوسط
الذاتي	٦	٢٣٤٠	١٥٧,٩	٩٦,٨	٠,٨٨٥- ٠,٩٥٨	٠,٩٣٠	مرتفع
المقارنة الطرفية	٣١	٤٠٨٢	٢٩٥٢,٧	٩٨,٩	٣,١٣٩- ٥,٠٣٤	٤,٠٨٧	مرتفع

يتضح من نتائج جدول (٣) أن:

✓**الصدق البنائي**

بلغ عدد حجوم التأثير للصدق البنائي (٥٥) حجم أثر من أصل (١١٤) حجم أثر للأساليب الإحصائية للتحقق من الصدق، وطبقت على عينة قوامها (٢٤٣٠٧) مفحوصين، وأسفرت النتائج عن وجود تجانس منخفض[١] بين الدراسات التي أجري لها تحليل بعدي؛ حيث بلغت قيم مؤشرات التجانس (Q = ٢٠٨٤,٥ & I^2 =٩٧,٤)، كما بلغ متوسط حجم التأثير للصدق البنائي كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٠.٦٩٢) وهو حجم أثر **متوسط** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير.

✓**الصدق التلازمي**

بلغ عدد حجوم التأثير للصدق التلازمي (٢٢) حجم أثر من أصل (١١٤) حجم أثر للأساليب الإحصائية للتحقق من الصدق، وطبقت على عينة قوامها (١٠٩٥٠) مفحوصاً، وأسفرت النتائج عن وجود تجانس منخفض بين الدراسات التي أجري لها تحليل بعدي؛ حيث بلغت قيم مؤشرات التجانس (Q = ١٢١٧,٣ & I^2 =٩٨,٣)، كما بلغ متوسط حجم التأثير للصدق التلازمي كأحد مؤشرات تقنين الاختبارات للعوامل المؤثرة في مواصفات الاختبار الجيد (٠.٦٩٦) وهو حجم أثر **متوسط** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير

✓**الصدق الذاتي**

بلغ عدد حجوم التأثير للصدق الذاتي (٦) حجوم أثر من أصل (١١٤) حجم

[١] يشير مدينا وآخرون (Medina et al, 2006,194) إلى أن I^2 يُعد مؤشراً على التجانس بين الدراسات الداخلة فى ما وراء التحليل، ويُعد مؤشر I^2 مكملاً لإختبار Q. ويشير فولار(Fowler, 2016,38) إلى أنI^2 (Heterogeneity) تعطينا مؤشر محدد لعدم التجانس عن طريق مقارنة التغير في حجوم التأثير للدراسات بالنسبة لتغير حجوم التأثير الكلي وقيمها = (25%, 50% , 75%) تكون على التوالي (منخفضًا، متوسطًا، مرتفعًا). مما يعنى أنه إذا كان I^2 أكبر من أو يساوى ٧٥٪ فإن التجانس يُعد منخفضًا، وإذا تراوح I^2 بين ٢٥ % و٧٥ % فإن التجانس يُعد متوسطًا، وإذا كان I^2 من أقل من أو يساوي ٢٥ % فإن التجانس يُعد مرتفعًا.

أثر للأساليب الإحصائية للتحقق من الصدق، وطبقت على عينة قوامها (٢٣٤٠) مفحوص، وأسفرت النتائج عن وجود تجانس منخفض بين الدراسات التي أجري لها تحليل بعدي؛ حيث بلغت قيم مؤشرات التجانس (Q = ١٥٧,٩ & I^2 =٩٦,٨)، كما بلغ متوسط حجم التأثير للصدق الذاتي كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٠.٩٣٠) وهو حجم أثر **مرتفع** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير.

✓**صدق المقارنة الطرفية:**

بلغ عدد حجوم التأثير لصدق المقارنة الطرفية (٣١) حجم أثر من أصل (١١٤) حجم أثر للأساليب الإحصائية للتحقق من الصدق، وطبقت على عينة قوامها (٤٠٨٢) مفحوص، وأسفرت النتائج عن وجود تجانس منخفض بين الدراسات التي أجري لها تحليل بعدي؛ حيث بلغت قيم مؤشرات التجانس (Q = ٢٩٥٢,٧ & I^2 =٩٨,٩)، كما بلغ متوسط حجم التأثير لصدق المقارنة الطرفية كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٤.٠٨٧) وهو حجم أثر **مرتفع** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير.

يتضح من جدول (٣) أن أكثر الأساليب الإحصائية فعالية للتحقق من الصدق هو صدق المقارنة الطرفية حيث بلغ حجم التأثير الكلي (٤.٠٨٧)، ولكنه يأتي في المرتبة الثانية من حيث الطرق الأكثر تكراراً، ثم يليه الأسلوب الإحصائي الأكثر فعالية هو الصدق الذاتي حيث بلغ حجم التأثير الكلي (٠.٩٣٠) ولكنه يأتي في المرتبة الأخيرة من حيث الطرق الأكثر تكراراً ويرجع ذلك إلى سبب كون الطريقة غير شائعة الاستخدام؛ حيث انه من النادر نجد أن الباحثين يستخدمون الصدق الذاتي للتحقق من الصدق، ثم يليه الأسلوب الإحصائي الصدق التلازمي وهو أحد أنواع الصدق المحكي حيث بلغ حجم التأثير (٠.٦٩٦) وهو حجم تأثير متوسط أي أن فعاليته في التحقق من صدق الأداة متوسطة ويقع في المرتبة الثالثة من حيث تكرار الاستخدام، واخيراً في المرتبة الرابعة من حيث فعالية الأسلوب الاحصائي وهو الصدق البنائي حيث بلغ حجم التأثير (٠.٦٩٢) وهو حجم تأثير متوسط ولكنه من حيث التكرار هو الأكثر تكراراً حيث يقع في المرتبة الأولى حيث بلغت عدد الأساليب الإحصائية للصدق البنائي (٥٥) من أصل (١١٤).

في البحث الحالي يتضح من الجدول (٣) أن أكثر الطرق تكراراً هي طريقة الصدق البنائي حيث تكرر استخدامه (٥٥) مرة من أصل (١١٤) وهذا يتفق مع دراسة باهي وفياض؛ ومحمود (٢٠٠٨؛ ٢٠١٩) أن الباحثين يعتمدون بشكل أساسي على الصدق البنائي وأنها أكثر الطرق استخداماً ودراسة عبدالوهاب (٢٠٢١) أن الصدق البنائي أكثر الطرق استخداماً بعد صدق المحكمين، ويرجع زيادة الاعتماد على طريقة الاتساق الداخلي (الصدق البنائي) في التحقق من الصدق كما يشير عبدالوهاب (٢٠٢١) إلى سهولة هذا الإجراء حيث لا يتطلب وجود محك خارجي، وسهولة المعالجة الإحصائية في هذه الطريقة عند استخدام البرامج الإحصائية

وأقل الطرق استخداماً بعد الصدق الذاتي هو الصدق التلازمي حيث تكرر الأسلوب الإحصائي (٢٢) مرة من أصل (١٤٤) بالرغم من أن فعاليته متوسطة إلا انه يستخدم قليلاً وهذا يتفق مع دراسة باهي وَفياض؛ عبدالوهاب (٢٠٠٨؛ ٢٠٢١) انه أقل الطرق استخداماً للتحقق من صدق أدوات القياس هي طريقة الصدق المرتبط بالمحك، ويرجع قلة استخدام الطرق الأخرى إلى عدة عوامل كما يشير عبدالوهاب (٢٠٢١) منها عدم قدرة الباحثين على الحصول على محك خارجي يتم قياس المتغير من خلاله أو عدم رغبة الباحثين في تطبيق أكثر من أداة على المفحوصين في المتغير الواحد، أو قد يعود إلى عدم معرفة بعضهم بخطوات إجراء الطريقة نفسها إحصائية

ومع اختلاف فعالية كل طريقة في التحقق من الصدق إلا أن جميع الطرق تعتبر من الأدلة الجيدة حيث ان فعاليتها تراوحت بين المرتفع والمتوسط، ولابد من التحقق من صدق أدوات القياس بأكثر من طريقة، كما يوصي حسن (٢٠١٤) في دراسته بانه يجب حساب الصدق بطرق تغطي جميع مكونات الصدق كلما كان ذلك ممكناً، وكذلك أشار الباحث انه على الباحثين أن يكونوا على وعي تام بأن استخدام أداة قياس غير صادقة يعني لا قياس.

وتأتي أهمية صدق النتائج واختيار الأساليب الإحصائية المناسبة للصدق حيث أشار حسن (٢٠١٤) على أهمية صدق النتائج في عملية تقويم نوعية البحوث التربوية والنفسية المنشورة في بعض دوريات التربية وعلم النفس وذلك لتحقيق بعض الأهداف، والتي من أهمها تجنب الاستعمالات غير الملائمة للأساليب الإحصائية؛ لما لها من أثر واضح في تهديد صدق نتائج البحوث، وعلاقة ذلك بقواعد اتخاذ القرارات الإحصائية، وهناك عدد من النقاط التي لابد للباحث أن يتحقق منها عند اختيار طريقة التحقق من الصدق:

- بما أن المقارنة الطرفية هي أعلى حجم تأثير إي أنها الأكثر فعالية في التحقق من أن أداة القياس تقيس فعلاً ما وضعت لقياسه، لذلك لابد عند استخدام هذه الطريقة أن يقسم الباحث المجموعة العليا والمجموعة الدنيا على ميزان أو محك خارجي وليس على الأداة المراد التحقق منها فإذا كان الافراد مرتفعين في الميزان أو المحك وايضاً مرتفعين في السمة، والمنخفضين في الميزان أو المحك هما منخفضان في المقياس فهذا يدل على صدق المقياس، إلا أن هناك خطأ شائعاً يقع فيه كثير من الباحثين كما تشير دراسة ذبيحي وَشوبار؛ عبدالوهاب (٢٠١٧؛ ٢٠٢١) هو تقسيم الأفراد على المقياس ذاته الى مرتفعي الدرجات ومنخفضي الدرجات بناء على الدرجة الكلية للمقياس دون الرجوع إلى محك خارجي وهذه من الأخطاء التي يقع فيها الباحثين ولابد من مراعاة شروط الاختبارات الإحصائية بين الأساليب البارامترية واللابارامترية وتحقق شروطها، ولعل ارتفاع حجم التأثير للمقارنة الطرفية جدول (٣) قد يعود إلى خطأ تقسيم أفراد المجموعتين العليا والدنيا.
- ومن الأخطاء التي تكون قد أدت إلى ارتفاع حجم التأثير في الصدق البنائي كما ذكر عبدالوهاب (٢٠٢١) هو الاعتماد على الدلالة الإحصائية لمعامل الارتباط المصحح عند قبول المفردة وفق الصدق البنائي، حيث أن أغلب الباحثين اللذين اعتمدوا على

هذه الطريقة ربطوا قبول المفردة بمستوى دلالة معامل الارتباط بين المفردات وما تقيسه، ونسبة كبيرة منهم اكتفوا فقط بالدلالة الإحصائية للارتباط دون النظر إلى قيمة الارتباط ومدى تجاوزها الحد الأدنى المقبول

- ومن الأخطاء في الصدق ما توصلت إليه محمود؛ عبدالوهاب (٢٠١٩؛ ٢٠٢١) عدم إعادة التحليل بعد حذف المفردات الضعيفة سيكومترياً

التساؤل الثاني: " **هل توجد فروق ذات دلالة إحصائية بين متوسطي حجم التأثير وفقًا لنوع الاختبار (تحصيلي – نفسي) في أساليب الصدق (البنائي – التلازمي– الذاتي– المقارنة الطرفية) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة"؟**

للإجابة على هذا التساؤل استخدمت الباحثة نموذج التأثير العشوائي (Random–Effect Models) كأحد النماذج الإحصائية للتحليل البعدي، وجاءت النتائج كما يوضحها جدول (٤) على النحو الآتي:

جدول (٤) قيم Z في نموذج التأثير العشوائي (Random-Effect Models) للفروق بين متوسطي حجم التأثير في أساليب الصدق (البنائي- التلازمي- الذاتي- المقارنة الطرفية) لدى عينة الدراسة وفقًا لنوع الاختبار (تحصيلي – نفسي)

المتغير نوع الصدق	نوع الاختبار	عدد حجوم التأثير	حجم التأثير الكلي d	مستوى حجم التأثير	التباين	الخطأ المعياري	قيمة Z	df	قيمة p
البنائي	تحصيلي	١٠	٠,٧٠٢	متوسط	٠,٠٠١	٠,٠٣٥	٢١,١	٥٤	٠,٠٠٠
	نفسي	٤٥	٠,٦٨٩	متوسط	٠,٠٠١	٠,٠٣٦			
	المجموع	٥٥	٠,٦٩٣	متوسط	٠,٠٠١	٠,٠٣٣			
التلازمي	تحصيلي	٤	٠,٧٧٠	متوسط	٠,٠٣٧	٠,١٩٣	١٢,١	٢١	٠,٠٠٠
	نفسي	١٨	٠,٦٧٦	متوسط	٠,٠٠٤	٠,٠٦١			
	المجموع	٢٢	٠,٦٨٦	متوسط	٠,٠١٠	٠,٠٩٩			
الذاتي	تحصيلي	١	٠,٩١١	مرتفع	٠,٠٠٠	٠,٠٠٠	١٥,٥	٥	٠,٠٠٠
	نفسي	٥	٠,٩٣٣	مرتفع	٠,٠٠٩	٠,٠٩٧			
	المجموع	٦	٠,٩٢٣	مرتفع	٠,٠٠٨	٠,٠٩١			
المقارنة الطرفية	تحصيلي	٣	٥,٨٨٩	مرتفع	٠,٠٣٧	٠,١٩٣	٨,٩	٣٠	٠,٠٠٠
	نفسي	٢٨	٣,٨٨٧	مرتفع	٠,٠٠٤	٠,٠٦١			
	المجموع	٣١	٤,١٢٨	مرتفع	٠,٠١٠	٠,٠٩٩			

يتضح من نتائج جدول (٤) ما يأتي:

✓**الصدق البنائي:**

- بلغ عدد حجوم التأثير للأسلوب الاحصائي الصدق البنائي **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (١٠) حجوم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٠.٧٠٢) وهو حجم أثر **متوسط**، في حين بلغ عدد حجوم التأثير للأسلوب الاحصائي الصدق البنائي **للاختبارات النفسية**

كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٤٥) حجم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٠.٦٨٩) وهو حجم أثر **متوسط**.

- وجود فروق ذات دلالة إحصائية بين متوسطي حجم التأثير في الصدق البنائي كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في عينة الدراسة وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح حجم التأثير في **مجموعة الدراسات التي تناولت الاختبارات التحصيلية** (حجم التأثير الأعلى = ٠.٧٠٢)، حيث جاءت قيمة (Z= ٢١.١) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٥٤).

✓**الصدق التلازمي:**

- بلغ عدد حجوم التأثير للأسلوب الاحصائي الصدق التلازمي **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٤) حجوم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٠.٧٧٠) وهو حجم أثر **متوسط**، في حين بلغ عدد حجوم التأثير للأسلوب الاحصائي الصدق التلازمي **للاختبارات النفسية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (١٨) حجم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٠.٦٧٦) وهو حجم أثر **متوسط**.
- وجود فروق ذات دلالة إحصائية بين متوسطي حجم التأثير في الصدق التلازمي كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في الدراسات التي يشملها البحث الحالي وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح حجم التأثير في **مجموعة الدراسات التي تناولت الاختبارات التحصيلية** (حجم التأثير الأعلى = ٠.٧٧٠)، حيث جاءت قيمة (Z= ١٢.١) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٢١).

✓**الصدق الذاتي:**

- بلغ عدد حجوم التأثير للأسلوب الإحصائي الصدق الذاتي **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (١) حجم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسة (٠.٩١١) وهو حجم أثر **مرتفع**، في حين بلغ عدد حجوم التأثير للأسلوب الاحصائي الصدق الذاتي **للاختبارات النفسية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٥) حجوم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٠.٩٣٣) وهو حجم أثر **مرتفع**.
- وجود فروق ذات دلالة إحصائية بين متوسطي حجم التأثير في الصدق الذاتي كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في الدراسات التي يشملها البحث الحالي وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح حجم التأثير في **مجموعة الدراسات التي تناولت الاختبارات النفسية** (حجم التأثير الأعلى = ٠.٩٣٣)، حيث جاءت قيمة (Z= ١٥.٥) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٥).

✓**صدق المقارنة الطرفية:**

- بلغ عدد حجوم التأثير للأسلوب الاحصائي صدق المقارنة الطرفية **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٣) حجوم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٥.٨٨٩) وهو حجم أثر **مرتفع**، في حين بلغ عدد حجوم التأثير للأسلوب الاحصائي صدق المقارنة الطرفية **للاختبارات النفسية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٢٨) حجم أثر، كما بلغ متوسط حجوم التأثير في هذه الدراسات (٣.٨٨٧) وهو حجم أثر **مرتفع**.
- وجود فروق ذات دلالة إحصائية بين متوسطي حجم التأثير في صدق المقارنة الطرفية كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في الدراسات التي يشملها البحث الحالي وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح حجم التأثير في **مجموعة الدراسات التي تناولت الاختبارات التحصيلية** (حجم التأثير الأعلى = ٥.٨٨٩)، حيث جاءت قيمة (Z= ٨.٩) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٣٠).

وبمقارنة الأساليب الإحصائية الأربعة للتحقق من الصدق من جدول (٤) نجد أنه توجد فروق ذات دلالة إحصائية بين متوسطي حجم التأثير للأساليب الإحصائية الصدق (البنائي، التلازمي، المقارنة الطرفية) وفقاً لنوع الاختبار (تحصيلي – نفسي) لصالح الاختبارات التحصيلية، بينما الفروق بين متوسطي حجم التأثير للصدق الذاتي الفروق لصالح الاختبارات النفسية

وبمكن أن يرجع أن الأساليب الإحصائية أكثر فعالية في التحقق من صدق الأداة إي أن الأداة تقيس فعلاً ما وضعت لقياسه في الاختبارات التحصيلية منها في الاختبارات النفسية، حيث أن الاختبارات التحصيلية يتم بنائها وفق خطوات أكثر دقة من الاختبارات النفسية، حيث يتم بناء فقراته في ضوء جدول المواصفات مما يضمن ان فقرات الاختبار تكون شاملة لجميع

الأهداف التي يقيسها الاختبار وهذا كما يشير عبدالرحمن (٢٠٠٨) إلى أن تكون أداة القياس في العلوم النفسية والتربوية (الاختبار التحصيلي) على درجة كبيرة من الحساسية لشدة العلاقة بين القدرة والأداء حتى يتم قياس الأداء وإرجاعه إلى الخاصية الواحدة.

<u>التساؤل الثالث</u>: " ما مستوى حجم التأثير لأساليب الثبات (ألفا كرونباخ – التجزئة النصفية– إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في عينة الدراسة؟".

للإجابة عن هذا التساؤل استخدمت الباحثة نموذج التأثير العشوائي (Random–Effect Models) كأحد النماذج الإحصائية للتحليل البعدي عن طريق استخدام البرنامج

الإحصائي (Comprehensive Meta – Analysis 2)، وجاءت النتائج كما يوضحها جدول (٥)

جدول (٥) متوسطات أحجام أثر أساليب الثبات بطريقة (الفا كرونباخ – التجزئة النصفية- إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة

المتغير نوع الثبات	عدد حجوم التأثير	عدد المفحوصين	Q	I^2	فترة الثقة	حجم التأثير الكلي d	مستوى حجم التأثير
الفا كرونباخ	٥٩	٢٠٢٣٠	٣٠٤٨,٩	٩٨,١	٠,٨١٩ -٠,٨٧٨	٠,٨٥١	مرتفع
التجزئة النصفية	٤٤	١٦٥٨٩	٨٨٠,٧	٩٥,١	٠,٨٠٩ -٠,٨٥٦	٠,٨٣٤	مرتفع
إعادة التطبيق	٣٣	٤٥٥٩	٧٢٥,٥	٩٥,٦	٠,٨٢٦ -٠,٨٩٩	٠,٨٦٧	مرتفع

يتضح من نتائج جدول (٥) ما يأتي:

✓ألفا كرونباخ

بلغ عدد حجوم التأثير للأسلوب الاحصائي الفا كرونباخ (٥٩) حجم أثر، طبقت على عينة قوامها (٢٠٢٣٠) مفحوص، وأسفرت النتائج عن وجود تجانس منخفض بين الدراسات التي أجري لها تحليل بعدي؛ حيث بلغت قيم مؤشرات التجانس (Q = ٣٠٤٨.٩ & I^2 =٩٨.١)، كما بلغ متوسط حجم التأثير للثبات بطريقة ألفا كرونباخ كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٠.٨٥١) وهو حجم أثر **مرتفع** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير.

✓التجزئة النصفية:

بلغ عدد حجوم التأثير للأسلوب الاحصائي التجزئة النصفية (٤٤) حجم أثر، طبقت على عينة قوامها (١٦٥٨٩) مفحوص، وأسفرت النتائج عن وجود تجانس منخفض بين الدراسات التي أُجري لها تحليل بعدي؛ حيث بلغت قيم مؤشرات التجانس (Q = ٨٨.٠٧ & I^2 =٩٥.١)، كما بلغ متوسط حجم التأثير للثبات بطريقة التجزئة النصفية كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٠.٨٣٤) وهو حجم أثر **مرتفع** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير

✓إعادة التطبيق:

بلغ عدد حجوم التأثير للأسلوب الإحصائي إعادة التطبيق (٣٣) حجم أثر، طبقت على عينة قوامها (٤٥٥٩) مفحوص، وأسفرت النتائج عن وجود تجانس منخفض بين الدراسات الداخلة في التحليل؛ حيث بلغت قيم مؤشرات التجانس (Q = ٧٢٥.٥ & I^2 =٩٥.٦) ، كما بلغ متوسط حجم التأثير للثبات بطريقة التجزئة النصفية كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٠.٨٦٧) وهو حجم أثر **مرتفع** وفقًا للمدى المرجعي لكوهين في تفسير حجم التأثير.

يتضح من جدول (٥) أن أكثر الأساليب الإحصائية فعالية للتحقق من مدى اتساق النتائج هي طريقة إعادة التطبيق حيث بلغ حجم التأثير الكلي (٠،٨٦٧) وهو حجم تأثير مرتفع، ولكنه أقل الطرق استخداماً في عينة الدراسة حيث تكرر استخدامه (٣٣) مرة

من أصل (١٣٦)، ثم يليه الأسلوب الاحصائي ألفا كرونباخ حيث بلغ حجم التأثير الكلي (٠.٨٥١) وهو حجم تأثير مرتفع، ويأتي في المرتبة الأولى من حيث الطرق الأكثر تكراراً حيث تكرر (٥٩) مرة من أصل (١٣٦)، وأخيراً في المرتبة الثالثة من حيث الأسلوب الأكثر فعالية في تقدير الثبات هي طريقة التجزئة النصفية حيث بلغ حجم التأثير الكلي (٠.٨٣٤) وهو حجم تأثير مرتفع ولكنه يأتي في المرتبة الثانية من حيث الطرق الأكثر تكراراً حيث تكرر (٤٤) مرة من أصل (١٣٦). ومما سبق نلخص نتائج التساؤل الثالث حيث نجد أن جميع الأساليب الإحصائية لتقدير الثبات كانت فعاليتها مرتفعة ولكن الفارق بينها جداً بسيط وهي عالترتيب إعادة التطبيق، ألفا كرونباخ والتجزئة النصفية، أما ترتيبها من حيث الطرق الأكثر استخداماً في تقدير ثبات أدوات البحث كالتالي: ألفا كرونباخ ثم التجزئة النصفية ثم إعادة التطبيق وأن اختيار أي من هذه الطرق يعتمد على إمكانيات الباحث وقدرته على تطبيق الاختبار مرتين، أو مرة واحدة

ونتيجة الدراسة الحالية أن أكثر الطرق تكراراً في عينة الدراسة هي طريقة الفا كرونباخ هذه النتيجة تتفق مع دراسة الهنائي؛ محمود وَعبدالوهاب (٢٠١١، ٢٠١٩، ٢٠٢١) بإن الفا كرونباخ اكثر طرق تقدير الثبات استخداماً للتحقق من أدوات الدراسة، ولا تتفق مع دراسة باهي وِفياض (٢٠٠٨) حيث توصلت دراستهما أن الفا كرونباخ أقل الطرق استخداماً. أي أن الطرق الأكثر تكراراً كالتالي: الفا كرونباخ ثم إعادة التطبيق ثم التجزئة النصفية والنتيجة تتنفق مع عبدالوهاب (٢٠٢١) أما ترتيب طرق تقدير الثبات حسب الأكثر فعالية هي كالتالي: طريقة إعادة التطبيق ثم ألفا كرونباخ، ثم التجزئة النصفية

ومن نتيجة الدراسة نجد أن طريقة الفا كرونباخ تأتي في المرتبة الثانية من حيث الطرق الأكثر فعالية في التحقق من قياس مدى اتساق واستقرار الدرجات على أداة الاختبار، ولكنها في المرتبة الأولى من حيث الطرق الأكثر استخداماً ويمكن أن يرجع كثرة الدراسات التي تستخدم ألفا كرونباخ وارتفاع حجم التأثير لمعامل ألفا كرونباخ كما يشير يانج وجرين (Yang & Green, 2011) للأسباب التالية:

١. سهولة تفسير معامل ألفا، وأن قيمته التي تقترب من (١.٠) يعتبر مؤشراً للاتساق الداخلي. مع التنبيه إلى عدم انتهاك المسلمات التي يستند إليها معامل ألفا كرونباخ وهي كما ذكرها حسن (٢٠١٤)
   - أن يكون المقياس أحادي البعد، لذلك لابد من فهم البنية الداخلية للمقياس والتأكد من أنه أحادي البعد
   - كتابة البنود بشكل واضح، وكتابة تعليمات الاختبار بصورة يسهل فهمها.
٢. يتميز معامل ألفا كرونباخ عن غيره من معاملات التحقق من الثبات بالموضوعية وعدم تدخل الباحث، فمثلاً في طريقة التجزئة النصفية يتم التدخل في تقسيم بنود الأداة، وفي طريقة إعادة التطبيق يتدخل الباحث في تحديد الفترة الزمنية بين التطبيقين
٣. يوجد إطار مرجعي لتفسير قيم معامل الثبات بطريقة ألفا كرونباخ على أنها صغيرة أو متوسطة أو كبيرة.

وكذلك يشير عبدالوهاب (٢٠٢١) الى كثرة استخدام طريقة معامل ألفا كرونباخ إلى سهولة إجرائه باستخدام البرامج الإحصائية، وعدم حاجة الباحث فيه إلى تطبيق الأداة أكثر من مرة والانتظار فترة زمنية بين التطبيقين، كما أنه لا يحتاج الباحث إلى إعداد صور متكافئة للأداة (مكافئة للأداة)، ويعود صلاحية استخدام معامل ألفا كرونباخ في حال كون مفردات الأداة ثنائية أو متعددة الاستجابات، وصلاحية استخدامه في حالة كون الظاهرة سريعة أو بطيئة التغير عبر الزمن

وأقل الطرق استخداماً في نتيجة الدراسة هي طريقة إعادة التطبيق حيث بلغت عدد مرات تكراراها (٣٣) مرة من (١٣٦) وتتفق مع نتيجة (الهنائي،٢٠١١) بإنها أقل الطرق تكراراً، ونتيجة (عبدالوهاب، ٢٠٢١) كانت في المرتبة الثانية من حيث درجة استخدامها، وتتعارض مع نتيجة (باهي وَفياض، ٢٠٠٨) بإنها أكثر الطرق استخداماً.

مع أنه توصلت نتيجة الدراسة الحالية أن إعادة التطبيق هي الأكثر فعالية في تقدير ثبات الاختبار وقد يعود ذلك إلى أنها الطريقة الأكثر ارتباطاً بالمفهوم السائد للثبات لدى كثير من الباحثين بأنه الحصول على نفس الدرجات تقريباً مهما أُعيد تطبيق الأداء ، وكذلك يشيران (باهي وَفياض، ٢٠٠٨) بإنه هو الاتفاق على أدوات القياس بمرور الوقت، وكذلك ما يميزها عن الطرق الأخرى ان يتم استخدام الأداة ذاتها ولكن بعد فترة زمنية مقارنة بالطرق الأخرى التي تتطلب صورتين متكافئتين للأداة أو بناء أداة من نصفين كل نصف يكافئ الآخر

واخيراً وبعد نتيجة هذا التساؤل نتوصل إلى أنه مع اختلاف فعالية كل طريقة في التحقق من تقدير الثبات، إلا أن كل طريقة لها جوانب قوة وقصور لذلك على الباحث أن يختار الطريقة التي تلائم السمة الذي يريد قياسها وكذلك مناسبتها لأفراد العينة وهذا ما أشار اليه (حسن، ٢٠١٤) انه يجب على الباحث استخدام أكثر من طريقة لحساب ثبات أدوات القياس وعدم الاكتفاء بطريقة واحدة؛ لأن طرق حساب الثبات تختلف في الكشف عن مصدر الخطأ في الاختبار وتوصلت نتائج دراسة (حسن، ٢٠٢٢) إلى انه عدم دقة استخدام الأساليب الإحصائية، وعدم تمشيها مع تصميم البحث، يؤدي إلى نتائج لا يمكن الثقة بها أو الاعتماد عليها وكذلك يؤدي إلى تضخيم خطأ التباين

<u>التساؤل الرابع:</u> " هل توجد فروق ذات دلالة إحصائية بين متوسطي حجم الأثر وفقًا لنوع الاختبار (تحصيلي – نفسي) في أساليب الثبات (الفا كرونباخ– التجزئة النصفية– إعادة التطبيق) كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد لدى عينة الدراسة؟"

للإجابة على هذا التساؤل استخدمت الباحثة نموذج التأثير العشوائي (Random–Effect Models) كأحد النماذج الإحصائية للتحليل البعدي، وجاءت النتائج كما يوضحها جدول (٦) على النحو الآتي:

جدول (٦) قيم Z في نموذج التأثير العشوائي (Random-Effect Models) للفروق بين متوسطي حجم الأثر في أساليب الثبات (الفا كرونباخ- التجزئة النصفية- إعادة التطبيق) لدى عينة الدراسة وفقًا لنوع الاختبار (تحصيلي – نفسي)

المتغير نوع الثبات	نوع الاختبار	عدد حجوم التأثير	حجم التأثير الكلي d	مستوى حجم التأثير	التباين	الخطأ المعياري	قيمة Z	df	قيمة p
الفا كرونباخ	تحصيلي	١٠	٠,٨٥٩	مرتفع	٠,٠٠٢	٠,٠٤٠	٢٥,٤	٥٨	٠,٠٠٠
	نفسي	٤٩	٠,٨٤٩	مرتفع	٠,٠٠٤	٠,٠٦٦			
	المجموع	٥٩	٠,٨٥٢	مرتفع	٠,٠٠٤	٠,٠٦٢			
التجزئة النصفية	تحصيلي	٨	٠,٨٧٢	مرتفع	٠,٠٠٢	٠,٠٣٩	٣٠,٦	٤٣	٠,٠٠٠
	نفسي	٣٦	٠,٨٢٦	مرتفع	٠,٠٠١	٠,٠٣٠			
	المجموع	٤٤	٠,٨٣٥	مرتفع	٠,٠٠١	٠,٠٢٨			
إعادة التطبيق	تحصيلي	٩	٠,٩١١	مرتفع	٠,٠٣٢	٠,١٧٩	١٨,٣	٣٢	٠,٠٠٠
	نفسي	٢٤	٠,٨٤٦	مرتفع	٠,٠٠٣	٠,٠٥٧			
	المجموع	٣٣	٠,٨٥٨	مرتفع	٠,٠٠٤	٠,٠٦٣			

يتضح من نتائج جدول (٦) ما يأتي:

✓ **ألفا كرونباخ:**

- بلغ عدد حجوم التأثير للأسلوب الاحصائي ألفا كرونباخ **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (١٠) حجوم أثر، كما بلغ متوسط حجوم الأثر في هذه الدراسات (٠.٨٥٩) وهو حجم أثر **مرتفع**، في حين بلغ عدد حجوم التأثير **للاختبارات النفسية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٤٩) حجم أثر، كما بلغ متوسط حجوم الأثر في هذه الدراسات (٠.٨٤٩) وهو حجم أثر **مرتفع**.
- وجود فروق ذات دلالة إحصائية بين متوسطي حجم الأثر في الثبات بطريقة ألفا كرونباخ كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في الدراسات التي يشملها البحث الحالي وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح حجم الأثر في **مجموعة الدراسات التي تناولت الاختبارات التحصيلية** (حجم الأثر الأعلى = ٠.٨٥٩)، حيث جاءت قيمة (٨.٩ =Z) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٥٨).

✓**التجزئة النصفية:**

- بلغ عدد حجوم الأثر للأسلوب الاحصائي التجزئة النصفية **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٨) حجوم أثر، كما بلغ متوسط حجوم الأثر في هذه الدراسات (٠.٨٧٢) وهو حجم أثر **مرتفع**، في حين بلغ عدد حجوم الأثر للأسلوب الاحصائي التجزئة النصفية **للاختبارات النفسية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٣٦) حجم أثر، كما بلغ متوسط حجوم الأثر في هذه الدراسات (٠.٨٢٦) وهو حجم أثر **مرتفع**.
- وجود فروق ذات دلالة إحصائية بين متوسطي حجم الأثر في الثبات بطريقة التجزئة النصفية كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في الدراسات التي يشملها البحث الحالي وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح

حجم الأثر في **مجموعة الدراسات التي تناولت الاختبارات التحصيلية** (حجم الأثر الأعلى = ٠.٨٧٢)، حيث جاءت قيمة (Z= ٣٠.٦) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٤٣).

✓**إعادة التطبيق:**

- بلغ عدد حجوم الأثر للأسلوب الإحصائي إعادة التطبيق **للاختبارات التحصيلية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٩) حجوم أثر، كما بلغ متوسط حجوم الأثر في هذه الدراسات (٠.٩١١) وهو حجم أثر **مرتفع**، في حين بلغ عدد حجوم الأثر للأسلوب الاحصائي إعادة التطبيق **للاختبارات النفسية** كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد (٢٤) حجم أثر، كما بلغ متوسط حجوم الأثر في هذه الدراسات (٠.٨٤٦) وهو حجم أثر **مرتفع**.
- وجود فروق ذات دلالة إحصائية بين متوسطي حجم الأثر في الثبات بطريقة إعادة التطبيق كأحد مؤشرات تقنين الاختبارات المؤثرة في مواصفات الاختبار الجيد في الدراسات التي يشملها البحث الحالي وفقًا لنوع الاختبار (تحصيلي – نفسي) لصالح حجم الأثر في **مجموعة الدراسات التي تناولت الاختبارات التحصيلية** (حجم الأثر الأعلى = ٠.٩١١)، حيث جاءت قيمة (Z= ١٨.٣) دالة إحصائياً عند مستوى دلالة (٠.٠١) ودرجة حرية (٣٢).

بمقارنة طرق التحقق من الثبات في جدول (٦) نجد أنه توجد فروق ذات دلالة إحصائية بين متوسطي حجم التأثير للأساليب الإحصائية (الفا كرونباخ، التجزئة النصفية، إعادة التطبيق) وفقاً لنوع الاختبار (تحصيلي– نفسي) لصالح الاختبارات التحصيلية، إي ان هذه الطرق للتحقق من الثبات هيا أكثر فعالية في الاختبارات التحصيلية من الاختبارات النفسية وهذه النتيجة تتفق مع دراسة (الهنائي، ٢٠١١) في انه توجد فروق ذات دلالة إحصائية في معاملات الثبات تُعزى إلى نوع الاختبار. ويمكن أن يرجع ذلك إلى عدة أسباب منها تأثير الاخطاء العشوائية في موقف الاختبار حيث أنه تعليمات الاختبار التحصيلي أكثر وضوحاً من الاختبارات النفسية، إجابة الفرد على الاختبار التحصيلي تقيس مدى تُمكن الطالب من المادة العلمية بينما في الاختبارات النفسية ممكن أن يكون فيها عشوائية أو تخمين في اختيار البدائل لإنه الاختبارات النفسية لا توجد فيها إجابة صحيحة أو خاطئة مما يؤدي إلى عشوائية الإجابة وايضاً من الأسباب مثلاً اذا كانت البدائل على مقياس ليكرت (موافق بشدة– موافق– محايد– غير موافق– غير موافق بشدة) فإنه لا يوجد تقدير كمي يعبر بدقة عن درجة امتلاك الفرد أو مدى انطباق البديل على المفحوص

واحياناً تكون تعليمات الاختبار غير واضحة، وكما نجد أنه أوقات تطبيق الاختبارات النفسية وما تشمله من استبيانات يمكن أنه لا يناسب العينة، كذلك الاعتماد على المقاييس الالكترونية ونشرها في مواقع التواصل الاجتماعي ويكون نشرها بشكل عشوائي مما يؤدي وصولها الى أفراد ليسوا من ضمن العينة المستهدفة، بل الاختبارات التحصيلية يتم تطبيقها على العينة المستهدفة، كذلك عدم فهم المفحوص للمصطلح المراد

قياسه، عدم فهم المفحوص للعبارات هذه كلها من الأخطاء العشوائية التي يمكن أن توثر في ثبات الاختبارات النفسية أكثر من التحصيلية.

ومن الأخطاء العشوائية التي يمكن أن تؤثر في طريقة إعادة التطبيق وتؤدي إلى ارتفاع زائف في معامل الثبات هو عامل التذكر، حيث إن الاختبارات التحصيلية يكون التذكر فيها أكثر من الاختبارات النفسية، أضافةً أن مهارات المفحوص مع كثرة المران ممكن أنها تتطور وتصبح أفضل في التطبيق الثاني وهذا ما يفسر ارتفاع حجم الأثر في الاختبارات التحصيلية بطريقة إعادة التطبيق عن الطريقتين الأخرى ألفا كرونباخ والتجزئة النصفية

ومن العوامل التي تؤثر في إعادة التطبيق كما يشير (عبدالرحمن، ٢٠٠٨) إن طريقة إعادة التطبيق لتعيين معامل ثبات الاختبارات التحصيلية، أو حتى اختبارات القدرات العقلية تحتاج إلى حذر وحيطة، وبالذات في تقدير الفترة الزمنية بين التطبيقين، وهذا التقدير يعتمد في غالبة على نوعية الاختبار والقدرة التي يقيسها.

ومما يفسر ان الفروق بين متوسطي الأسلوب الاحصائي (طريقة الثبات) ونوع الاختبار لصالح الاختبارات التحصيلية ان الاختبارات التحصيلية هي ان الاختبارات مصيرية تحدد نجاح أو رسوب الطالب فهنا يقل عامل التخمين والعشوائية في الإجابة عنها عن الاختبارات النفسية التي يمكن أن يكثر فيها التخمين او العشوائية كذلك ان الاختبارات النفسية تعتمد على حالة الفرد ويمكن أن تؤثر حالته في الإجابة عن الاختبار ويغير اجابته من مره إلى أخرى. ومما يفسر ارتفاع ثبات الاختبارات التحصيلية عن النفسية ان الثبات يتأثر بحالة الفرد كما يشير السيد (١٩٧١) أي انه في الاختبارات التحصيلية يكون الطالب اكثر استعداداً للاختبار وتوجد مادة علمية يستذكرها الطالب وكذلك كثرة المران والاستعداد للاختبارات التحصيلية يكون أكثر منها في الاختبارات النفسية.

توصيات الدراسة:

١. بناء على نتائج الدراسة فإن أكثر طرق التحقق من الصدق التي تناسب الاختبارات النفسية هي طريقة الصدق الذاتي بينما الاختبارات التحصيلية تلائمها طريقة الاتساق الداخلي والمقارنة الطرفية والتلازمي.
٢. بناء على نتائج الدراسة أن طرق التحقق من الخصائص السيكومترية للأداة ذات تأثير عالي في الاختبارات التحصيلية مقارنة بالاختبارات النفسية.

المقترحات:

١. إجراء دراسة للتحليل البعدي لأنواع الثبات وتأثرها بطول المقياس
٢. إجراء دراسة للتحليل البعدي باستخدام طرق نظرية الاستجابة للفقرة وإمكانية التعميم.
٣. إجراء دراسة للتحليل البعدي ودراسة العوامل المؤثرة في خصائص الاختبار الجيد بتحديد مقياس معين.
٤. إجراء دراسة للتحليل البعدي للعوامل المؤثرة في خصائص الاختبار الجيد للدراسات في البيئة الأجنبية.

المراجع

أبو حطب، فؤاد؛ وَصادق، آمال. (٢٠١٠). *مناهج البحث وطرق التحليل الإحصائي في العلوم النفسية والتربوية والاجتماعية*. مكتبة الأنجلو المصرية: مصر

أبو علام، رجاء محمود. (٢٠١١). *مناهج البحث في العلوم النفسية والتربوية*. دار النشر للجامعات: القاهرة

أبو هاشم، السيد محمد. (٢٠٠٥). *مؤشرات التحليل البعدي Meta- Analysis لبحوث فعالية الذات في ضوء نظرية باندورا*. عمادة البحث العلمي بجامعة الملك سعود.

أبو هاشم، السيد محمد. (٢٠٢٠). معامل ألفا للتحقق من ثبات درجات أدوات القياس بين الحقائق والمعتقدات الخاطئة لدى الباحثين. *مجلة البحث العلمي في التربية*، (٢١)، ١٧٩ – ٢١٠.

باهي، مصطفى حسين إبراهيم. (٢٠١٠). العلاقة بين الدلالة الإحصائية وحجم التأثير في البحوث التربوية والنفسية. *مجلة مستقبل إعداد المعلم في كليات التربية وجهود الجمعيات العلمية في عمليات التطوير بالعالم العربي*،٢ ، ٤١٥ – ٤٤٤

باهي، مصطفى حسين وَفياض، ناهد خيري. (٢٠٠٨). التحليل السيكومتري لأدوات القياس الثبات والصدق: دراسة تحليلية. *المجلة العلمية للتربية البدنية وعلوم الرياضة*، (٥٤)، ٤٠٩– ٤٢٥

بشتة، حنان وَبوعموشة، نعيم. (٢٠٢٠). الصدق والثبات في البحوث الاجتماعية. *مجلة دراسات في علوم الانسان والمجتمع- جامعة جيجل*، ٣ (٢)، ١١٧ – ١٣٣

الجهني، ليلى سعيد. (٢٠١٧). كفاءة التعليم الإلكتروني في ضوء التحليل البعدي لنتائج الدراسات المنشورة في بعض الدوريات العربية خلال ٢٠٠٥ إلى ٢٠١٥. *المجلة الدولية التربوية المتخصصة*، ٦ (٧)، ١٧– ٣٣

الحربي، ثناء أحمد. (٢٠٢١). معايير الجودة في إعداد الاختبارات الإلكترونية عن بعد لمادة العلوم من وجهة نظر معلمات ومشرفات المرحلة المتوسطة. *مجلة شباب الباحثين بكلية التربية جامعة سوهاج*، (٩)، ٣٤ – ١٠٥

حسن، عزت عبد الحميد محمد. (٢٠١٤). الثبات أم الصدق أولاً؟. *المجلة المصرية للدراسات النفسية*، ٢٤ (٨٣)، ١٠٥– ١٢٨.

حسن، علي صلاح عبد المحسن. (٢٠١٨). الأخطاء الشائعة في بناء الاختبارات الموضوعية وتأثيرها على الخصائص السيكومترية للاختبارات التحصيلية لدى طلاب كلية التربية بجامعة أسيوط. *إدارة البحوث والنشر العلمي (المجلة العلمية)*، ٣٤ (٢)، ٣٦٦– ٤٠٦.

حسن، علي صلاح عبد المحسن. (٢٠٢٢). الممارسات الإحصائية لحجوم تأثير البرامج بالدراسات التربوية: (دراسة تقييميه في ضوء الإحصاء البارامتري واللابارامتري). *مجلة كلية التربية– عين شمس*، (٤٦)، ٤٤٣– ٤٧٦.

ذبيحي، لحسن وَشوبار، لياس. (٢٠١٧). أخطاء شائعة في البحوث العلمية. مركز جيل البحث العلمي، *مجلة جيل العلوم الإنسانية والاجتماعية*، (٢٨)، ١١–٢٣.

سكران، السيد عبد الدايم. (٢٠٠٦). التحليل البعدي "META – ANALYSIS" كمنهج وصفي تحليلي لتجميع نتائج البحوث وتكاملها في مجال التربية وعلم النفس. *مجلة كلية التربية بالزقازيق*، ١ – ٣٨.

السيد، فؤاد البهي. (١٩٧١). *علم النفس الاحصائي وقياس العقل البشري*. دار الفكر العربي: القاهرة

الشربيني، زكريا أحمد؛ إبراهيم، رزق سند؛ دمنهوري، رشاد صالح وَمطحنة، السيد خالد. (٢٠١٠). *القياس النفسي*. مكتبة الشقري: الرياض

الشريفين، نضال كمال. (٢٠١٧). التحليل البعدي للأبحاث المنشورة في المجلة الأردنية في العلوم التربوية: الدلالة العملية وقوة الاختبار. *مجلة اتحاد الجامعات العربية للتربية وعلم النفس*، ٣ ، ١٣٠- ١٧٠.

ضيف، حسن. (٢٠١٦). *بناء اختبار تحصيلي في مادة الرياضيات لتلاميذ السنة الرابعة متوسط*. [رسالة ماجستير غير منشورة]. جامعة قاصدي مرباح، كلية العلوم الإنسانية والاجتماعية. قسم علم النفس وعلوم التربية

طباع، فاروق وَخنيش، يوسف. (٢٠٢١). بعض الأخطاء الناجمة عن جمع أدلة الصدق والثبات في البحوث التربوية والنفسية. *مجلة العلوم الاجتماعية والإنسانية*، ٢٢ (١)، ٢٤٣ – ٢٥٨

عبدالحميد، محمد جمال الدين. (١٩٨٧). أسلوب التحليل البعدي لنتائج البحوث والدراسات السابقة، *جامعة قطر- كلية التربية*، ٥ (٥)، ٣١٧ – ٣٥٧

عبدالرحمن، سعد. (٢٠٠٨). *القياس النفسي النظرية والتطبيق*. هبة النيل العربية للنشر والتوزيع: مصر

عبدالفتاح، فيصل أحمد. (٢٠١١). قوة الاختبارات الإحصائية وحجم الأثر في البحوث التربوية والنفسية، *مجلة جامعة الملك سعود- العلوم التربوية والدراسات الإسلامية*، ٢٣ (٢)، ١٩٥- ٢١٤

عبدالوهاب، محمد محمود. (٢٠٢١). تقييم إجراءات القياس النفسي في تقارير البحوث التربوية العربية. *مجلة العلوم التربوية والنفسية، جامعة البحرين- مركز النشر العلمي*، ٢٢ (١)، ١١٩ – ١٦٢

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 [عرض ورقة]، الجامعة السعودية بغزة، ٣٠٥- ٣٣٦

علام، صلاح الدين محمود. (٢٠١٥). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي

عيشوش، رياض. (٢٠١٧). استخدام تحليل الميتا في البحوث الإدارية: المقاربات النظرية والتطبيقات العملية. *مجلة البحوث الاقتصادية والمالية*، ٤(٢)، ٢٠٩ – ٢٣٤

اسعادي، فارس وَ عزي، ايمان. (٢٠١٥). مواصفات الاختبار التحصيلي الجيد. *مجلة الدراسات والبحوث الاجتماعية*، (١١)، ١٨٩ – ٢٠٠.

فرماوي، أحمد محمود. (٢٠٠٣). *تطبيقات التحليل البعدي في مجال التربية الخاصة*. في المركز المصري لمصادر معلومات العلوم الاجتماعية (إكسير) www.gulfkids.com

محاسنة، نور مفلح . (٢٠٢٠). التحليل البعدي لنتائج الرسائل الجامعية التي تناولت فاعلية المنحى البنائي الاجتماعي في تنمية المهارات اللغوية في الأردن خلال الفترة من ٢٠١٠- ٢٠١٧. *مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية*، ٣٢(١١)، ١٥١- ١٦١.

محاسنة، نور مفلح؛ والشريفين، كمال نضال. (٢٠٢٠). التحليل البعدي لنتائج الرسائل الجامعية التي تناولت فاعلية نموذج التعلم البنائي في الأردن خلال الفترة من ٢٠١٠ – ٢٠١٧. *مجلة الجامعة الإسلامية للدراسات التربوية والنفسية*، ٥ ، ٥٨٨ – ٦٠٩.

محمد، خلف الديب عثمان. (٢٠٢٠). التحليل البعدي لبحوث برامج واستراتيجيات علاج صعوبات القرأءة بالمرحلة الابتدائية. *مجلة التربية- جامعة الأزهر*، ٢ (١٨٨)، ٢-٣٢

محمود، سوميه شكري محمد. (٢٠١٩). الأخطاء الشائعة في إجراءات التحقق من ثبات وصدق أدوات القياس المستخدمة في البحوث التربوية العربية. *مجلة كلية التربية جامعة أسيوط- كلية التربية*، ٣٥ (٧)، ٦٧١ – ٦٩٥

مهدي، رشا أحمد. (٢٠١٩). كفاءة استراتيجيات ما وراء المعرفة في ضوء التحليل البعدي لنتائج الدراسات المنشورة في بعض الدوريات العربية خلال الفترة من ٢٠٠٠ إلى ٢٠١٧. *مجلة كلية التربية. جامعة أسيوط– كلية التربية. ٣٥* (٢)، ١– ٣٣

الهنائي، سيف سعيد عبدالله. (٢٠١١). *تحليل بعدي لثبات درجات اختبارات المصفوفات المتتابعة لرفن.* [رسالة ماجستير منشورة، كلية التربية، جامعة السلطان قابوس]. قاعدة معلومات دار المنظومة.

DeCoster, J. (2009). Meta-Analysis Notes.

Ferguson, C. J. (2016). An effect size primer: a guide for clinicians and researchers.

Fowler, M. (2016). *The effect of ADHD on academic achievement: A meta-analytic review*. St. John's University (New York).

Glass, G. V. (2000). Meta-analysis at 25. www.gvglass.info/papers/m] eta25.html

Guzzo, R. A., Jackson, S. E., & Katzell, R. A. (1987). Meta-analysis analysis. *Research in organizational behavior*, *9*(1), 407-442.

Medina, T., Meca, J., Martinez, F & Botella,J. (2006). Assessing Heterogeneity In Meta-Analysis: Q Statistic or I^2 Index?. *Psychological Methods*, 11 (2), 193- 206

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews*, *82*(4), 591-605.

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research*, *90*(1), 24-46.

Rios, J. A., Ihlenfeldt, S. D., Dosedel, M., & Riegelman, A. (2020). A Topical and Methodological Systematic Review of Meta-Analyses Published in the Educational Measurement Literature. *Educational Measurement: Issues and Practice*, *39*(1), 71-81.

Rosenthal, R., Cooper, H., & Hedges, L. (1994). Parametric measures of effect size. *The handbook of research synthesis*, *621*(2), 231-244.

Shachar, M. (2008). Meta-Analysis: The preferred method of choice for the assessment of distance learning quality factors. *The International Review of Research in Open and Distributed Learning*, *9*(3), 1-15

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, *1*(9), 1-9

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century?. Journal of psychoeducational assessment, 29(4), 377-392.

Yildirim, E., Tikman, F., & Senturk, M. (2022). Meta-analysis on the effectiveness of using the drama approach in social studies course. International Journal on Social and Education Sciences, 4(1), 17-33.