



التكامل بين الدلالة العملية والدلالة الإحصائية لاختبارات "ت وتحليل التباين ومعامل الارتباط"
كميل عزمي، ايمان خلف، محمود الأمين*

قسم علم النفس التربوي، كلية التربية، جامعة سوهاج، سوهاج 82524، مصر
*الباحث المسؤول: mahmoud.185983@azhar.moe.edu.eg

ملخص البحث

سعت الدراسة إلى التكامل بين الدلالة العملية والدلالة الإحصائية في التحليلات الكمية للبحوث التربوية. واستخدمت منهج تحليل المحتوي. وركزت على تقديم تفاصيل حول مقاييس حجم الأثر التي تقترن باختبارات الدلالة الإحصائية الأكثر شيوعاً. واستخدمت الدراسة استمارة لجمع البيانات من البحوث لتحليلها وفقاً لمفهوم الدلالة الإحصائية والعملية وقد طبقت الدراسة على عينة من رسائل الماجستير بلغ عددها (67) رسالة بالأقسام التربوية بكلية التربية جامعة سوهاج في الفترة من 2000-2020م. وخرجت الدراسة بمجموعة من النتائج منها عدد الرسائل التي تم فيها حساب الدلالة العملية إلى جانب الدلالة الإحصائية (17) رسائل من أصل (67) بنسبة (25.37%) وتوزعت على قسمي علم النفس التربوي بعدد (8) رسائل وقسم مناهج وطرق تدريس اللغة العربية بعدد (9) رسائل، بلغ عدد الاختبارات الإحصائية التي حسبت الدلالة العملية جنباً إلى جنب مع الدلالة الإحصائية (301) اختبار من إجمالي (1714) اختبار بنسبة (17.6%)، وبلغ عدد الاختبارات الدالة إحصائياً (1225) بنسبة (71.47%) لاختبارات الفروض الإحصائية (ت وتحليل التباين ومعامل الارتباط لبيرسون) والغير دالة (489) بنسبة (28.53%)، وعدد الفرضيات الإحصائية ذات الدلالة العملية الضعيفة (385) بنسبة (22.46%) والفرضيات الإحصائية ذات الدلالة العملية المتوسطة (328) بنسبة (19.14%) والفرضيات الإحصائية ذات الدلالة العملية القوية (520) بنسبة (30.34%). وقدمت الدراسة توصيات مقترحة لتحقيق التكامل بين الدالتين العملية والإحصائية في تحليلات البحوث التربوية.

الكلمات المفتاحية: الدلالة العملية، الدلالة الإحصائية، حجم الأثر

بيانات المقال

الاستشهاد المرجعي: كميل عزمي، ايمان خلف، محمود الأمين. (2023). التكامل بين الدلالة العملية والدلالة الإحصائية لاختبارات "ت وتحليل التباين ومعامل الارتباط، مجلة سوهاج لشباب الباحثين، مجلد 3 (4)، 15-29.

تاريخ استلام البحث: 2022/12/22

تاريخ قبول البحث: 2023/02/15

تاريخ نشر البحث: 2023/03/02

<https://doi.org/10.21608/sjyr.2023.288320>

Publisher's Note: SJYR stays neutral regarding jurisdictional claims in published maps and institutional affiliations

1. المقدمة

إن مفهوم الدلالة الإحصائية للنتائج يركز على مدي الثقة التي نوليها لنتائج الفروق أو العلاقات بصرف النظر عن حجم هذه الفروق أو تلك العلاقات، فإذا كان الفرق دال إحصائياً عند مستوي (0.05) بين مجموعتين، فذلك يشير إلى أن الفرق بين المجموعتين حقيقي وأن مجتمع العينة الأولي يختلف عن مجتمع المجموعة الثانية، وأنا نثق في هذا الحكم بنسبة 95%، وهنا نكون في حاجة إلى ما يسمى بالدلالة العملية للتعبير عن مقدار هذا الفرق من حيث كونه صغيراً أو متوسطاً أو كبيراً، لذا فالدلالة العملية هي الجانب المكمل لمفهوم الدلالة الإحصائية، حيث تشير الدلالة العملية إلى الفرق أو حجم الارتباط بصرف النظر عن مدي الثقة التي نضعها في النتائج، وأن كلاً منهما مكمل للآخر ويعوض النقص الكامن فيه، فهما كوجهي عملة واحدة يؤدي استخدامهما معاً إلى إثراء نتائج البحوث النفسية والتربوية وتوسعي الدراسة الحالية لتقديم مقاييس حجم التأثير وطرق حسابها وكيفية تفسيرها من خلال السياق البحثي للتحقق من الدلالة العملية والتي تؤكد على قيمة النتائج في العالم الحقيقي، ومن هذا المنطلق أحس الباحث بمشكلة الدراسة وبضرورة الحاجة إلى التكامل بين الدلالة العملية والإحصائية في البحوث التربوية والنفسية، حيث لا توجد أي قيمة للدلالة الإحصائية إذ لم تقرن بالدلالة العملية لذا على الباحثين في الوقت الراهن التوجه نحو التحقق من فرضية ما إذا كان هناك للتأثير دلالة عملية وعدم الاكتفاء بالتحقق من فرضية هل التأثير دال إحصائياً أم لا؟

1.1. مشكلة الدراسة

إن القصور الواضح في استخدام الدلالة العملية في البحوث التربوية والنفسية من حيث عدم التزام الباحثين باستخراج الدلالة العملية لنتائج أبحاثهم وكذلك ضعف قدرة الباحث على اختيار وحساب وتفسير مؤشرات حجم الأثر المقابلة لفروض

الاختبارات الإحصائية الخاصة بالدراسة، وتركيز اهتمام الباحثين على عرض نتائج الدلالة الإحصائية فقط، والاعتماد عليها اعتماداً كلياً في صياغة وتفسير نتائجهم، وقل أن تجد من بين هؤلاء الباحثين من يركز في نقاش دراسته على معطيات الدلالة العملية ومنطقية النتائج التي توصل إليها في المجال التطبيقي، فبالرغم من أهمية تقديم أحجام التأثير كأحد أهم المقاييس الإحصائية للتحقق من دلالة النتائج العملية وقيمتها في العالم الواقعي إلا أن واقع البحث النفسي والتربوي يفرض عدد من المعوقات التي تحول دون ذلك منها عدم تأهيل الباحثين من خلال المقررات الإحصائية المقدمة. وعدم وجود قواعد وخطوات واضحة المعالم لمقاييس حجم الأثر لكيفية استخدامها وطريقة الاختيار للمقاييس الأمثل المقابل للأسلوب الإحصائي. وكيفية تفسير قيم هذه المؤشرات كما لوحظ فشل برامج الحزم الإحصائية الشائعة الاستخدام في توفير بعض مقاييس حجم الأثر المناسبة لتقييم النتائج. وعدم إلزام الباحثين أكاديمياً وتحملهم مسؤولية تقديم معلومات كاملة فيما يتعلق بالنتائج التي توصل إليها البحث من خلال إظهار الدلالة العملية للنتائج. كما أن التحدي الذي يطرحه هذا العصر من خلال توفير كم هائل وضخم من البيانات في شتى المجالات وكيفية التعامل معها يحتاج إلى استخدام أساليب تكشف عن المعنى الحقيقي لهذه البيانات. كما أشارت دراسة [1] إلى أنه لطالما خدمت الدلالة الإحصائية كطريقة سائدة للاستدلال الإحصائي إلا أن سوء التفسير وسوء الاستخدام على نطاق واسع أدى إلى التشكيك في فائدتها في تعميم النتائج، وإثارة المخاوف بشأن الآثار بعيدة المدى لهذه الممارسات على التقدم العلمي خصوصاً في مجال البحث التربوي والنفسي، مما دفعنا إلى استخدام منهج مكمل يتضمن مقاييس حجم الأثر التي تلعب دوراً رئيسياً في تعميم الأهمية العملية، كما أشارت دراسة [2] إلى قلة الثقة لدى متخذي القرار التربوي في مصداقية نتائج الدراسات التربوية، مما أدى إلى وجود فجوة بين كثرة الدراسات والأبحاث العلمية وقلة تطبيق نتائجها في الميدان التربوي، ومن خلال تتبع الأسباب التي أدت إلى هذا الأمر؛ نجد أن الكثير من الباحثين لا يدركون المفاهيم الصحيحة للأساليب الإحصائية التي يستخدمونها في دراساتهم، والتي ينبون عليها قراراتهم في تفسير نتائج هذه الدراسات، حيث أنهم يكتفون بنتائج الدلالة الإحصائية فقط في اختبار فرضياتهم البحثية، ولا يدركون أن إيجاد الدلالة الإحصائية ما هي إلا خطوة أولى في تحليل النتائج، وتتبعها خطوات أخرى أكثر أهمية من الدلالة الإحصائية، ويجب أن تركز عليها القرارات العلمية في تفسير النتائج.

وأن الباحثين التربويين لا يولون اهتماماً مناسباً لاستخدام مقاييس إحصائية بخلاف مقياس الدلالة الإحصائية، كما أكدت دراسة [3] على أن الاعتماد على مؤشر إحصائي وهو الدلالة الإحصائية والتي تقدم بيانات غير كافية مما يمنع الاتساق بين تحليل البيانات وتفسير النتائج، كما أشارت الدراسة بأنه يجب على باحثي العلوم التربوية والنفسية تحديد الافتراضات الأساسية للبحث وكذلك خصائص التصميم والقياس والتحليل بحيث يمكن وضع هذه التقديرات في سياقها. وفي ضوء العرض السابق تتحدد مشكلة الدراسة الحالية في تضمين جوانب قصور الاعتماد على أسلوب الدلالة الإحصائية وبالتالي الحاجة إلى معالجة هذا القصور من خلال إجراء عملية التكامل بين الدلتين الإحصائية والعملية لاختبارات الفروض الإحصائية المقابلة من حيث طرق اختيار المقاييس الأمثل لحجم الأثر وكيفية حسابة وتفسيره من خلال السياق البحثي ودرجة الاتساق بين نتائج الدلالة الإحصائية والعملية، ومنه معالجة خطوة الاعتماد على نتائج الدلالة الإحصائية فقط في تفسير نتائج الدراسات التربوية.

1.2. أسئلة الدراسة

يمكن تحديد مشكلة الدراسة في الأسئلة التالية:

- 1- ما واقع الدلالة العملية والدلالة الإحصائية للاختبارات الإحصائية (ت وتحليل التباين ومعامل ارتباط بيرسون) لرسائل الماجستير في الأقسام التربوية (علم النفس التربوي ومناهج وطرق تدريس اللغة العربية والصحة النفسية) في كلية التربية بجامعة سوهاج في الفترة من 2000-2020؟
- 2- كيف يتم التكامل بين الدلالة الإحصائية والدلالة العملية؟

1.3. أهداف الدراسة

تهدف الدراسة إلى ما يأتي:

- 1- التعرف على توزيع حجم الأثر المقترن بالفرضيات الصفرية التي تم اختبارها حسب المستويات التي اقترحها كوهين لحجم الأثر.
- 2- التعرف على مدى التزام الباحثين بتزويد نتائجهم بجميع المتطلبات اللازمة لحساب حجم الأثر واستخدامها في تفسير النتائج أو تقديم التوصيات.
- 3- التعرف على نسبة الفرضيات التي اتفقت فيها نتائج الدلالة الإحصائية مع قيمة حجم الأثر (الدلالة العملية) في رسائل الماجستير والتي استخدمت الاختبارات الإحصائية (ت أو تحليل التباين أو معامل الارتباط).

1.4. أهمية الدراسة

تتضح أهمية الدراسة النظرية والتطبيقية في النقاط التالية:

- 1- تحسين الممارسات البحثية من خلال تقديم المساعدة للباحثين في مجال العلوم النفسية والتربوية، وتقديم الآلية الصحيحة لاستخدام هذه الأساليب والمؤشرات.
- 2- توظيف مؤشرات حجم الأثر في البحوث التربوية وتحديد مؤشر حجم الأثر المقابل لكل اختبار إحصائي والاستفادة منها في التفسيرات والقرارات المتعلقة بفروض أبحاثهم.

2. أدوات وطرق البحث

2.1. منهج الدراسة

تستخدم الدراسة منهج تحليل المحتوى من المنهج الوصفي القائم على تحليل الوثائق واستنتاج المعلومات، حيث يستطيع منهج تحليل المحتوى الوصفي أن يصف ظروفًا وممارسات معينة توجد في الواقع، وأن يبرز الاتجاهات، وأن يكشف نواحي الضعف، وأن يتتبع تطور أعمال ومفاهيم معينة، وذلك بناءً على قائمة محددة البنية تم إعدادها لذلك.

2.2. حدود الدراسة وأداتها

تمثلت حدود الدراسة الحالية على تقييم وحساب حجم الأثر الناتج من استخدام الأساليب الإحصائية لوصف وتحليل بيانات رسائل الماجستير بالأقسام التربوية (علم النفس وعددها 23 رسالة - قسم الصحة النفسية وعددها 25 رسالة - مناهج وطرق تدريس اللغة العربية وعددها 19 رسالة) في كلية التربية بجامعة سوهاج في الفترة من 2000 وحتى 2020 م.

2.3. إجراءات الدراسة

جاءت إجراءات الدراسة على النحو التالي: الخطوة الأولى: تقديم الإطار العام للدراسة من خلال (مقدمة الدراسة، والمشكلة، والتساؤلات، والأهداف، والأهمية، والمنهج، وحدود الدراسة وأداتها، وإجراءات السير فيها). والخطوة الثانية: تقديم الإطار النظري للدراسة الذي يركز على الإحصاء وتطور العلوم التربوية والنفسية وطرق استخدام وحساب وتفسير مؤشرات مقاييس حجم الأثر؛ وقد تضمن على ما يأتي: مفهوم الدلالة الإحصائية ومكوناتها والاستخدام والتفسير الأمثل لها وتقديم مقاييس حجم الأثر كوجه مكمّل لها. والخطوة الثالثة: تقديم إجراءات الدراسة ونتائجها حول واقع الدلالة الإحصائية لفروض الاختبارات الإحصائية في البحوث التربوية والنفسية في ضوء دلالاتها العملية. والخطوة الرابعة: عرض توصيات الدراسة المقترحة لتحقيق التكامل بين الدلالة الإحصائية والدلالة العملية، وانتهت الدراسة بقائمة المراجع المستخدمة في الدراسة.

2.4. الإطار النظري للبحث

نتناول في هذا الجزء من البحث المفهوم الصحيح للدلالة الإحصائية ومكوناتها والدلالة العملية كوجه مكمّل في البحوث التربوية والنفسية.

2.4.1. مفهوم الدلالة الإحصائية Statistical Significance

تُشير الدلالة الإحصائية إلى الادعاء بأن النتيجة من البيانات التي تم إنشاؤها عن طريق الاختبار أو التجريب من غير المحتمل أن تحدث بشكل عشوائي أو عن طريق الصدفة، ولكن من المحتمل بدلاً من ذلك أن تُعزى إلى سبب محدد ومعين وأن العلاقة بين المتغيرات ناتجة عن شيء آخر غير الصدفة، حيث تعتبر الدلالة الإحصائية هي الطريقة لإثبات أن إحصائية معينة موثوقة رياضياً عندما تتخذ قرارات بناءً على نتائج التجارب. التي نجرها كما أنها تمثل درجة اقتراب قيمة مقاييس العينات من مقاييس المجتمع الأصل، ويمكن اكتشافها بواسطة الاختبار الإحصائي وهو مجموعة من القواعد تمكن الباحث من رفض أو قبول الفرض الإحصائي وبموجبه يمكن الحكم على الفرض الإحصائي. كما تُشير الدلالة الإحصائية إلى الاحتمال الشرطي للبيانات حيث يبدأ الباحث بافتراض عدم وجود تأثيرات أو علاقة أو اختلاف فيما يعرف باسم الفرضية الصفرية H_0 بعد ذلك يقوم الباحث بتقييم احتمالية البيانات بالنظر إلى H_0 ثم تحديد البيانات المتطرفة أو الأكثر تطرفاً بافتراض صحة H_0 ومع استخدام مؤشر القيمة الاحتمالية (p-value) لتحديد ما إذا كانت النتيجة ذات دلالة أم لا؟ بدءاً من افتراض أن التأثير الحقيقي هو صفر إلى نسبة p-value، وتشير قيمة إحصائيات الاختبار المحسوبة من عينات عشوائية افتراضية والتي تكون متطرفة أو أكثر تطرفاً من إحصائية الاختبار فعندما تكون p-value منخفضة فهذا يشير إلى أنه عندما تكون H_0 صحيحة فإن عدداً صغيراً من النتائج سيكون متطرفاً أو أكثر تطرفاً من النتيجة الحالية، بالنظر إلى أنه يبدو أن هناك احتمالاً منخفضاً للحصول على نتائج متطرفة أو أكثر تطرفاً مما لوحظ عليه عند افتراض أن التأثير الفعلي هو صفر.

2.4.2. المشكلات المصاحبة للدلالة الإحصائية

اعتقاد الباحث بأن الدلالة الإحصائية تتضمن أهمية نظرية أو عملية حيث إنه من الخطأ في تفسير الدلالة الإحصائية هو الإشارة إلى أنها تتضمن أو تعني أهمية نظرية أو عملية وذلك أن هذا المفهوم الخاطئ يتضمن تفسير مهم إحصائياً للفرق على أنه اختلاف له آثار عملية على الرغم من عدم وجود شيء في تعريف الدلالة الإحصائية يشير إلى أن النتيجة الدالة مهمة من الناحية العملية. إن الدلالة الإحصائية لن تكون أبداً بديلاً عن المنطق العلمي ولا بد من اعتماد صحة الاستنتاجات العلمية على الأساليب الإحصائية والتقنيات المختارة بشكل مناسب والتحليلات التي أجريت بشكل صحيح والتفسير الصحيح للنتائج الإحصائية ممثلة في مقاييس حجم الأثر.

2.4.3. القيمة الاحتمالية Probability Value

تشير القيمة الاحتمالية على أنها احتمال الحصول على نتيجة متطرفة أو أكثر تطرفاً من النتيجة التي تم الحصول عليها بالفعل بافتراض أن الفرضية H_0 صحيحة، كما أنه في العديد من الحالات تعتمد القيمة الاحتمالية على الخطأ المعياري الذي يتأثر بدوره بحجم العينة فمع زيادة حجم العينة يصبح الخطأ المعياري أقل وتميل القيمة p-value إلى الانخفاض عن المستوي الدلالي α - level مما يؤدي إلى رفض الفرضية H_0 وبذلك تشير النتيجة ذات الدلالة الإحصائية بشكل أساسي إلى أنه تم استخدام حجم عينته كبير. وأن القيمة الاحتمالية p-value ليست فرصة منخفضة للإنذار الخاطئ أو الصدفة، ولكنها فرصة منخفضة للتأثير، وأن التأثير الناتج من تفسير القيمة الاحتمالية هو (ليس صفرًا) ويضرب لذلك مثلاً وهو بعد أن يخضع متدرب لتدريب تسأل ما هو مقدار تحسنك؟ فيجيب "ليس صفرًا" تحسنت بعض الشيء، وهذا التحسن الضئيل هو القيمة

الاحتمالية المنخفضة، وأن التأثير أو الاختلاف أو الارتباط المُعبر عنه بالقيمة الاحتمالية الضئيلة $P < 0.05$ هو ليس صفراً وأنه اختلاف قليل أو نادر وهذا النوع من الإجابات هو ما نحصل عليه من اختبارات الدلالة الإحصائية.

2.4.4. مستوى الدلالة أو المعنوية

يشير مفهوم مستوى الدلالة إلى المستوى الذي عنده يتم رفض الفرضية الصفرية، وهو الحد الأقصى لاحتمال الوقوع في خطأ من النوع الأول وإن اختيار 0.05 & 0.01 . كمستويات للدلالة كان تاريخياً اختياراً تعسفياً من قبل (Fisher) ولكن كان معقول في معظم الظروف، من الممارسات الشائعة الحكم على نتيجة بأنها دالة إذا كان لها من تأثير الصدفة ما لا يتكرر أكثر من مرة كل عشرين مرة ($0.05 = 1:20$)، إن كان هذا أمراً اختيارياً ولكنه مناسب للدلالة، وهذا لا يعني أننا نخذع مرة كل عشرين مرة (0.05)، ولكن بما يمكن تجاهله من الفشل في إعطاء نتيجة مهمة إذا كان (0.05) لا يبدو عالياً بما فيه الكفاية فيمكننا أن نأخذ واحد من مائة (0.01) كما يُفضل المعيار المنخفض للدلالة وتجاهل جميع النتائج التي تفشل في الوصول إلى هذا المستوى ويجب اعتبار الحقيقة على أنها مثبتة تجريبياً فقط إذا تم تصميمها بشكل صحيح ونادراً ما تفشل التجربة في إعطاء هذا المستوى من الأهمية، كما تستخدم مستويات الدلالة كمؤشر للحكم على قوة الأدلة في الدراسات الفردية المستخدمة في تأسيس الاستنتاجات العلمية والبحثية.

كما يوصي [4] الباحث بالآتي:

- 1- لا تبني استنتاجاتك فقط على ما إذا كان التأثير أو الارتباط تجاوز مستوى الدلالة.
- 2- لا تعتقد أن التأثير أو الارتباط موجود لمجرد أنه كان ذا دلالة إحصائية.
- 3- لا تعتقد أن التأثير أو الارتباط غير موجود لمجرد أنه لم يكن ذا دلالة إحصائية.

2.4.5. التشجيع على الأسلوب الإحصائي المكمل

ويشير [5] إلى أنه إذا كانت اختبارات الدلالة الإحصائية مفيدة في تحديد الأهمية الإحصائية فإن احجام التأثير مفيدة في تحديد الأهمية العلمية كلاً منهما مفيد بمفرده ولكنهما معاً يمكنهما فعل أشياء لا يستطيع أي منهما القيام بها بمفرده، مع اختبارات الدلالة الإحصائية وأحجام التأثير يجب أن يشعر الباحث بحرية استخدام أي تقنية إحصائية من شأنها إلقاء الضوء على الجوانب المثيرة للاهتمام من بياناتهم.

كما اهتم علماء الاجتماع بالإجابة عن ثلاثة أسئلة وذلك عند فحص العلاقات بين المتغيرات:-

- 1- هل الأثر الملحوظ حقيقي يعزو إلى الصدفة وهي (الدلالة الإحصائية)؟
- 2- إذا كان هناك أثراً حقيقياً فكم هو حجمه وهو (حجم الأثر)؟
- 3- هل الأثر كبير بما يكفي ليكون نافعا وهو (الدلالة العملية)؟

كما عبر [6] بأنه يعتبر اختبار الدلالة الإحصائية هو حجر الزاوية في البحث الكمي لكن الدراسات التي تفشل في الإبلاغ عن مقاييس حجم الأثر من المحتمل أن تفقد جزءاً قوياً من التحليل حيث يركز البحث الكمي في التربية وعلوم النفس بشكل أساسي على وصف العلاقات بين المتغيرات أو الفروق بينها، وغالباً ما يعتمد الباحث بشكل كبير على التحليلات التي تحدد ما إذا كان التأثير الملحوظ حقيقياً أم ينسب للصدفة دون النظر بشكل كامل إلى قوة العلاقة بين تلك المتغيرات، بينما يتفق معظم الباحثين على أهمية تحديد الدلالة العملية لنتائجهم، حيث أن اختبار الدلالة الإحصائية وحده لا يوفر جميع المعلومات حول حجم التأثير أو ما إذا كانت العلاقات بين المتغيرات أو الفروق بينها ذات مغزى.

حجم التأثير

وعرفه [7] على أنه " درجة تواجد ظاهرة في مجتمع ما أو فروق بين متوسطات أو علاقة بين المتغيرات " وهذا يعني أنه عندما تكون الفرضية الصفرية خاطئة فإن هذا يعني بالضرورة وجود فرق كبير وواضح بين المتوسطات، حيث يعتبر حجم الأثر توصيف لدرجة اختلاف نتائج العينة عن الفرضية الصفرية، ومن هنا كلما كبرت قيمة حجم الأثر زادت احتمالية تواجد الظاهرة قيد الدراسة في ذلك المجتمع، ومن ذلك نجد ارتباط حجم التأثير بالأهمية العملية لنتائج الدراسة وأن الفروق أو العلاقة هي حقيقية وليست نتيجة الصدفة. لذلك يحتاج الباحثون لسؤال أنفسهم ما الشكل الذي ستكون عليه النتيجة حتى تكون مهمة؟ أو ما أقل فرق بين المتوسطات يجب أن يظهر حتى تكون النتائج ذات دلالة عملية جوهرية؟ يحدث ذلك عندما تكون درجة الاختلاف أو الارتباط كبيراً بما يكفي لتكون النتائج ذات أهمية عملية، وذلك باستخدام مقياس حجم الأثر.

كما يعتبر مقياس حجم التأثير مؤشراً إحصائياً كمياً يمكن حسابه، ويمكن أن يعطي معناً كيفياً يعتمد أول ما يعتمد على مجال الدراسة والفائدة المتوقعة من إجرائها، أي أنه مؤشر لمدي قدرتنا على استخدام النتائج تفسيراً أو تطبيقاً، أو هو كم التباين الذي أمكن تفسيره للمتغير التابع حينما اعتبرنا متغيراً مستقلاً في علاقة معه أو مؤثر عليه.

لذا فكما أشار [8] على أن الدلالة الإحصائية ليست كافية لصناعة قرار صائب إذ هي مؤشر إحصائي لإظهار الفرق بين مجموعتين أو أكثر، أو أن العلاقة بين متغيرين أو أكثر راجعة إلى فروق أو علاقة حقيقية، وليست نتيجة المصادفة ولكن هذا لا يعني الاعتماد على هذا الفرق واتخاذ القرارات بناءً عليه. كذلك إن زيادة حجم العينة قد يؤدي في بعض الأحيان إلى رفض الفرضية الصفرية مما يعني وجود دلالة إحصائية، وهنا يختلط الأمر على الباحث: هل الدلالة الإحصائية المتحصل عليها تعني أن العامل أو المعالجة لها تأثير على الناتج أم هذه الدلالة الإحصائية جاءت نتيجة لكبر حجم العينة، ومن ثمّ فالدلالة المتحصل عليها توقع الباحث في حيرة من أمره لأنه يلاحظ أنه على الرغم من رفض الفرضية الصفرية H_0 ، فإن هذه النتيجة ليس لها أهمية

عملية تُذكر، لذا فالكفاية تتحقق إذا ما حسبنا الدلالة العملية، وهذه الدلالة تعد مؤشراً إحصائياً لمدي القدرة علي استخدام النتائج تفسيرياً أو تطبيقياً، كما يتيح لنا معرفة حجم التأثير التأكد من الأهمية العملية للدلالة الإحصائية. أولاً- تصنيفات أحجام التأثير

لقد صيغت مؤشرات إحصائية خاصة بغرض حساب الدلالة العملية وكذلك تعزيز القرار الإحصائي وتحسين جودة البحوث في البحث التربوي والنفسي، حيث تخطي عدد هذه المؤشرات لأكثر من 70 مؤشراً للدلالة العملية وتم تصنيفها كالتالي:

1- المؤشرات التي تدل على مقدار التأثير في الدراسات التي تبحث الفروق وتهتم هذه المؤشرات بالتأثير الذي يحدثه المتغير المستقل في المتغير التابع، وتستخدم الاختبارات الإحصائية التي تبحث بدورها الفروق بين متوسطات درجات المجموعات.

2- المؤشرات التي تدل على مقدار التأثير في الدراسات الارتباطية والتنبؤية على الرغم من وجود أنواع مختلفة من معاملات الارتباط نتيجة اختلاف نوع المعامل المستخدم تبعاً لطبيعة البيانات، ألا إن معامل ارتباط بيرسون هو الأكثر استخداماً في الدراسات التربوية والنفسية، وهذه المؤشرات يشار إليها أحياناً بمصطلح مقدرات نسبة التباين.

ثانياً- تفسير أحجام التأثير

يجب علي الباحث أن يضع في اعتباره عند حساب مقياس حجم الأثر:

أنّ الحساب الإحصائي هو عملية رياضية في حين تفسير أحجام التأثير هو فعل من أفعال البشر حيث لا توجد نظرية إحصائية يمكنها إصدار هذه الأحكام (كبير - متوسط - صغير) فهي معايير نسبية وعادةً ما يتم اتخاذ مثل هذه الأحكام من خلال معرفة وفهم الدراسة وسياق البحث وربط أحجام التأثير بالنتائج العملية ودرجة عالية من الحكم الذاتي والموضوعية وأيضاً من خلال مقارنة هذه الأحجام مع أحجام تأثير أخرى صراحةً أو ضمناً.

تفسير أحجام التأثير من خلال مقارنة التأثير الناتج من عينة الدراسة مع النطاق الكلي للتباين في المجتمع حيث من المهم إدراك أن التباين الكلي يختلف عند مستويات متعددة للمجتمع فقد يبدو الاختلاف أو التأثير الذي يبدو كبيراً عند مقارنته بالمجتمع الأكبر (التباين الكلي) صغيراً مقارنةً بالتوزيع الإجمالي.

إنّ المعايير العالمية لتفسير أحجام التأثير لا تساعد الباحث في تفسير نتائج بحثه حيث اقترح Cohen بشأن المعايير التي وضعها لتفسير أحجام التأثير (كبير - متوسط - صغير) بأنه ينبغي تطبيقها فقط عندما لا يتوفر إطار مرجعي أفضل لتقييم الأهمية العملية، ومن المهم أن يدرك الباحث أن كوهين اقترح هذه المعايير لتحليل القوة وليس لتفسير النتائج ولكن لسوء الحظ تجاهل الباحثون إلي حد كبير حجج كوهين التحذيرية وقاموا بتوحيد مبادئه التوجيهية في فولكلور البحث النفسي والاجتماعي وأظهر الإدراك المتأخر أنه إذا تم تقديم مثل هذه العتبات فسيتم تبنيها كوسيلة للراحة وغالباً دون تفكير كبير في كيفية تطبيقها علي المجال قيد البحث والدراسة وأنها معايير نسبية وليست مطلقة فعلي الرغم من أن معايير Cohen قد اكتسبت بعض القبول في مجال البحوث التربوية والنفسية وغيرها إلا أنه لا ينبغي تطبيقها بشكل أعمي، حيث يعتمد تفسير الدلالة العملية علي المجال الذي تتم دراسته ويمثل في النهاية حكماً قيمياً من جانب الباحث، فمع استمرار تطور طرق التفسير سيحتاج الباحثون إلي تقديم تفسيرات منطقية لاختباراتهم وتفسيرات واضحة لنتائجهم مما له الأثر علي تجاوز الدلالة الإحصائية.

ثالثاً: - مقياس حجم الأثر للاختبارات المعملية (البارامترية):

(1) مقياس حجم الأثر لاختبار "ت" لعينة واحدة:

$$d = \frac{\bar{x} - t}{\sigma} \quad (1)$$

\bar{x} هو متوسط العينة و t قيمة إحصاء اختبار (ت) و σ الانحراف المعياري وعند مقارنة متوسط عينة بمتوسط مجتمع تستخدم المعادلة (2) أو (3):

$$d = \frac{\bar{x} - \mu}{\sigma_x} \quad (2)$$

وبالمثل:

$$\text{Hedges' } g = \frac{\bar{x} - \mu}{s_x} \quad (3)$$

وفي حالة عدم معرفة الدرجات الخام أو المتوسطات الحسابية أو الانحرافات المعيارية لنتائج بحث ما فإنه يمكن حساب حجم الأثر من خلال المعادلة (4):

$$d = \frac{t}{\sqrt{df}} \quad (4)$$

t قيمة إحصائية اختبار "ت" لعينة واحدة ودرجة الحرية هي $df = n - 1$ يُفسر d في ضوء المحكات التالية:

- 1- حجم تأثير ضعيف إذا كانت $0.2 \leq d < 0.5$
- 2- حجم تأثير متوسط إذا كانت $0.5 \leq d < 0.8$
- 3- حجم تأثير كبير إذا كانت $0.8 \leq d$

وأشار كلاً من [9] بأنه في حالة العينة الواحدة للاختبار الثنائي ذو الحدين (Dichotomous Data) يمكننا معرفة ما إذا كانت النسبة المئوية في المجتمع ستكون مختلفة بشكل كبير عن نسبة 50% والتي قدمها كوهين من خلال المعادلة (5) وأشار لها بما يسمى Cohen's g:

$$\text{Cohen's } g = P_{\text{observed}} - .50 \quad (5)$$

وقدم كوهين تفسيراً للمؤشر g في ضوء المحكات التالية:

$$1- \text{ حجم تأثير ضعيف جداً إذا كانت } 0.00 \leq g < 0.05$$

$$2- \text{ حجم تأثير ضعيف إذا كانت } 0.05 \leq g < 0.15$$

$$3- \text{ حجم تأثير متوسط إذا كانت } 0.15 \leq g < 0.25$$

$$4- \text{ حجم تأثير كبير إذا كانت } 0.25 \leq g$$

(2) مقياس حجم الأثر لاختبار "ت" لعينتين مستقلتين:

في حالة التصميم التجريبي الأكثر شيوعاً المكون من عينتين مستقلتين نستخدم مقياس Hedges' g في حالة عينتين متساويتين $n_1 = n_2$

$$\text{Hedges' } g = \frac{\bar{x}_1 - \bar{x}_2}{s_{1/2}} \quad (6)$$

وهو الفرق المعياري بين متوسطي درجات المجموعتين (\bar{x}_1 و \bar{x}_2) مقسوماً على الانحراف المعياري لأي من المجموعتين $s_{1/2}$ شريطة التجانس، ويمكن حساب s_{pooled} من العلاقة:

$$s_{\text{pooled}} = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

في حالة عينتين غير متساويتين $n_1 \neq n_2$

$$\text{ES} = g = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}} \quad (7)$$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

بالمثل يمكن استخدام Cohen's d حيث

في حالة عينتين متساويتين $n_1 = n_2$

$$\text{Cohen's } d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{1/2}} \quad (8)$$

ويمكن حساب σ_{pooled} من العلاقة:

$$\sigma_{\text{pooled}} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

في حالة عينتين غير متساويتين $n_1 \neq n_2$

$$\text{Cohen's } d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\text{pooled}}} \quad (9)$$

$$\sigma_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$$

وتظهر العلاقة بين انحراف العينة والمجتمع المجمعين s_{pooled} ، σ_{pooled} من خلال العلاقة:

$$\sigma_{\text{pooled}} = s_{\text{pooled}} \sqrt{\frac{df}{N}}$$

ويختلف مؤشر Cohen's d في أنه مقياس وصفي وفي حالة العينات الكبيرة بينما مؤشر Hedges' g مقياساً استدلالياً وفي حالة العينات الصغيرة نسبياً.

كما قدم [10] المؤشر g^* كعامل تصحيح لمؤشر Cohen's d وأقل تحيزاً من خلال المعادلة (10):

$$g^* = d \cdot \left(1 - \frac{3}{4(df) - 1}\right) \quad (10)$$

في العينات المرتبطة، $df = n - 1$ في العينات المستقلة $df = n_1 + n_2 - 2$

كما أوضح Hedges' g بأن المؤشرات Cohen's d، Hedges' g تقريباً متطابقة إذا كانت قيمة حجم العينة $N > 20$ ويمكن تفسيرها جميعاً بنفس الطريقة لجميع الاهداف العملية ما لم تكن تتعامل مع عينات صغيرة جداً. وإذا لم يتحقق تجانس تباين العينتين يتم استخدام معامل دلتا جلاس Glass's Δ

$$\Delta = \frac{\bar{X}_1 - \bar{X}_2}{s_c} \quad (11)$$

s_c هو الانحراف المعياري للعينة الضابطة لأن قيم العينة الضابطة لم تتأثر بالمتغير المستقل تحت شرط حجم العينة. وفي حالة معرفة قيمة النسبة t ودرجات الحرية لنتائج بحث ما فإنه يمكن حساب حجم الأثر d بدلالة t من خلال المعادلة (12)، (13).

في حالة $n_1 = n_2$

$$d = \frac{2t}{\sqrt{df}} \quad (12) \quad (df = n - 1)$$

وفي حالة $n_1 \neq n_2$

$$d = \frac{t(n_1 + n_2)}{\sqrt{n_1 n_2 (n_1 + n_2 - 2)}} \quad (13)$$

من خلال النسبة t يمكن حساب مؤشر حجم الأثر ل Hedges' g وذلك من خلال المعادلة (12) في حالة $n_1 = n_2$ نستخدم المعادلة (14):

$$g = \frac{2t}{\sqrt{N}} \quad (14)$$

وفي حالة $n_1 \neq n_2$ نستخدم المعادلة (15):

$$g = t \sqrt{\frac{(n_1 + n_2)}{n_1 n_2}} = t \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (15)$$

كما أشار [11] إلى أنه توجد معادلات (12,13,14,15) تستخدم في حالة العينات المستقلة أو غير المستقلة وهما مقياسي Cohen's d، Hedges' g. وفي حالة معرفة معامل الارتباط r للعلاقة بين مجموعتين (معامل الارتباط بين درجات القياسين) من القيم فإنه يمكن حساب حجم التأثير وفقاً للمعادلة (16) باستخدام المؤشر d والمعادلة (17) باستخدام المؤشر g كالتالي:

$$d = \frac{2r}{\sqrt{1 - r^2}} \quad (16)$$

$$g = \frac{r}{\sqrt{1 - r^2}} \sqrt{\frac{n_1 n_2}{df(n_1 + n_2)}} \quad (17)$$

ويسمى حجم التأثير في هذه الحالة حجم التأثير لمعامل الارتباط Effect size of correlation، ولا يتم حساب حجم الأثر باستخدام قيمة النسبة t إلا إذا كانت النتائج التي أسفر عنها البحث دالة إحصائياً عند مستوي 0.05 أو أقل، ويدل حجم التأثير في هذه الحالة علي مدى الانتماء للمتغير التابع موضع الاهتمام بالبحث وهو مقياس لما يسمى بالدلالة العملية للنتائج التي أسفر عنها البحث.

كما يمكن التعبير عن r باستخدام g من خلال المعادلة (18):

$$r = \frac{g}{\sqrt{g^2 + 4\left(\frac{df}{N}\right)}} \quad (18)$$

كما أوضح [12] بأنه يمكن التعبير عن الفرق بين مجموعتين مستقلتين عن طريق إيجاد نسبة التباين المشترك بينهما باستخدام المؤشر ω^2 من خلال المعادلة (19):

$$\omega^2 = \frac{t^2 - 1}{t^2 + n_1 + n_2 - 1} \quad (19)$$

وتظهر العلاقة بين مؤشري حجم الأثر d لكوهين و g لهيدجز من خلال المعادلتين (20) و (21) حيث يمكن حساب d من خلال g والعكس

$$d = \frac{g}{\sqrt{\frac{N}{df}}} \quad (20)$$

$$g = \frac{d}{\sqrt{\frac{N}{df}}} \quad (21)$$

ويمكن تفسير قيم حجم التأثير للمؤشرات g و Cohen's d وفق المحك التالي الذي فسره [13] حجم تأثير صغير جداً إذا كانت $d, g < 0.20$

- 1- حجم تأثير صغير إذا كانت $0.20 \leq d, g < 0.50$
- 2- حجم تأثير متوسط إذا كانت $0.50 \leq d, g < 0.80$
- 3- حجم تأثير كبير إذا كانت $0.80 \leq d, g < 1.10$
- 4- حجم تأثير كبير جداً إذا كانت $1.10 \leq d, g < 1.50$
- 5- حجم تأثير ضخم إذا كانت $1.50 \leq d, g$

(3) مقياس حجم الأثر لاختبار "ت" لعينتين مرتبطتين:

وفي حالة معرفة القيمة t, r معامل الارتباط بين القياس القبلي والبعدي، n حجم العينة تستخدم المعادلة (22):

$$ES = t \sqrt{\frac{2(1-r)}{n}} \quad (22)$$

ومن طرق حساب حجم التأثير للمتغير المستقل على المتغير التابع، أيضاً في حالة معرفة قيمة النسبة t طريقة تعتمد على حساب قوة العلاقة بين المتغيرين هي مقياس مربع إيتا η^2 وهي الدليل القوي على الأثر الفعلي للمعالجة التجريبية على نتائج البحث حيث تتراوح قيمة η^2 بين (0, 1) لذا يتم ضرب القيمة المتحصل عليها في 100% حيث يتم تفسيرها بالنسبة المئوية والتي تعكس نسبة التباين الكلي لدرجات المتغير التابع من تأثير المتغير المستقل، ثم يتم حساب حجم التأثير من معادلة كوهين من خلال المعادلتين (23)، (24) [14].

$$\eta^2 = \frac{t^2}{t^2 + df} \quad (23)$$

$df = n - 1$ في العينات المرتبطة، $df = n_1 + n_2 - 2$ في العينات المستقلة

وبعد حساب نسبة التباين من خلال قيم مربع إيتا العادية يتم حساب حجم الأثر بدلالة مربع إيتا من المعادلة (24):

$$d = 2 \times \sqrt{\frac{\eta^2}{1 - \eta^2}} \quad (24)$$

وقد أشار [15] إلى أن مربع إيتا يسمي أحياناً نسبة الارتباط، ويعتبر مقياساً وصفيّاً للترابط بين العينات موضوع البحث في حالة المجموعات المستقلة أو غير المستقلة وهو من مقاييس حجم الأثر القائمة على حساب التباين المشترك. ومن خلال المعادلة (25) يمكن اشتقاق علاقة مربع إيتا η^2 بدلالة حجم التأثير d وهي كالتالي:

$$\eta^2 = \frac{d^2}{d^2 + 4} \quad (25)$$

يستخدم مربع إيتا η^2 ومؤشر كوهين d في إيجاد حجم التأثير ومستوياته وذلك طبقاً لنوعية التصنيف في تحليل التباين وفيما يلي معادلات حجم التأثير في حالة استخدام هذا الأسلوب في البحوث التربوية والنفسية:

(4) مقياس حجم الأثر لاختبار تحليل التباين الاحادي والثنائي:

أشار [16] إلى أنه في حالة استخدام تحليل التباين لمتغير واحد له ثلاثة مستويات مستقلة، حيث يتم تفسير درجات المتغير التابع نتيجة تأثير المتغير المستقل باستخدام مربع إيتا:

$$\eta^2 = \frac{SS_{\text{effect(between)}}}{SS_{\text{total}}} \quad (26)$$

حيث $SS_{\text{effect(between)}}$: مجموع المربعات بين المجموعات، SS_{total} المجموع الكلي للمربعات ثم حساب حجم التأثير للمتغير المستقل على التابع وفق المعادلة (27):

$$d = 2 \times \sqrt{\frac{\eta^2}{1 - \eta^2}} \quad (27)$$

في حالة تحليل التباين الثنائي تستخدم المعادلة (28):

$$\eta^2 = \frac{SS_{\text{eff}(1)}}{SS_{\text{total}}} \quad (28)$$

حيث $SS_{\text{eff}(1)}$: مجموع المربعات للمتغير المستقل الأول، SS_{total} المجموع الكلي للمربعات ثم تطبيق معادلة حجم الأثر (27) وبالمثل في حالة المتغير الثاني والثالث، وكذا التفاعل بين المتغيرات. ويمكن استخدام المؤشر مربع إيتا الجزئي (η_p^2 : partial eta squared) وذلك لتصحيح اخطاء مربع إيتا الذي يتأثر بالعوامل الخارجية حيث ينصح [17] في حالة تحليل التباين الثنائي باستخدام مربع إيتا الجزئي عند المقارنة بين حجم أثر متغير معين في مجموعة من الدراسات التي استخدمت تصميم تحليل التباين كامل العشوائية متعدد المتغيرات، لأن عدد وطبيعة المتغيرات الأخرى قد يختلف من دراسة لأخرى وأن كل متغير من المتغيرات وكذلك التفاعلات بينهم يساهم في مجموع التباين الكلي لمجموعة البيانات حيث تعمل الإضافة التي تمت في المقام بعزل أثر العوامل الأخرى غير الخطأ، ويتم حساب مربع إيتا الجزئي η_p^2 من المعادلات (29) ، (30) ، (31) ، (32):

$$\eta_p^2 = \frac{SS_{\text{effect(between)}}}{SS_{\text{error}} + SS_{\text{effect(between)}}} \quad (29)$$

$$\eta_p^2 = \frac{SS_A}{SS_{\text{error}} + SS_A} \quad (30)$$

$$\eta_p^2 = \frac{SS_B}{SS_{\text{error}} + SS_B} \quad (31)$$

$$\eta_p^2 = \frac{SS_{A*B}}{SS_{\text{error}} + SS_{A*B}} \quad (32)$$

حيث:

SS_{error} مجموع مربعات الخطأ أو (مجموع المربعات بين المجموعات).

SS_A مجموع المربعات بين مجموعات المتغير المستقل A أو (أثر A).

SS_B مجموع المربعات بين مجموعات المتغير المستقل B أو (أثر B).

SS_{A*B} مجموع المربعات بين مجموعات التفاعل A * B أو (أثر التفاعل).

ويلاحظ من خلال معادلات η_p^2 و η^2 تستخدم مربع إيتا العادية η^2 (في حالة تحليل التباين الأحادي) وتستخدم مربع إيتا الجزئية η_p^2 (في حالة تحليل التباين في اتجاهين فأكثر) ثم بعد ذلك حساب حجم الأثر من خلال المعادلة (33):

$$d = 2 \times \sqrt{\frac{\eta_p^2}{1 - \eta_p^2}} \quad (33)$$

حيث يمثل η^2 مقدار التباين الذي يمثله المتغير المستقل ويُفسر بين (0,1) فتشير إحصائية حجم التأثير المحسوب بالتباين إلى النسبة المئوية لإجمالي تباين المتغير التابع بواسطة المتغير المستقل.

ويمكن حساب مربع إيتا الجزئي η_p^2 من خلال النسبة الفائية F لتحليل التباين باستخدام المعادلة (34):

$$\eta_p^2 = \frac{F df_{\text{effect}}}{F df_{\text{effect}} + df_{\text{error}}} \quad (34)$$

حيث: df_{effect} درجات حرية الأثر وتساوي (عدد المجموعات الفرعية للمتغير -1)

حيث: F النسبة الفائية المحسوبة لتحليل التباين

حيث: df_{error} درجة حرية الخطأ

كذلك قدم [18] المؤشر الغير متحيز: (ϵ^2 Epsilon squared) لتصحيح اخطاء مربع إيتا الذي يتأثر بالعوامل الخارجية من المعادلة (35):

$$\epsilon^2 = \frac{SS_{\text{(between)}} - (k - 1)MS_{\text{error}}}{SS_{\text{total}}} \quad (35)$$

وقد أشار [19] إلى أنه قد تكون قيمة مربع إيتا (η^2) متحيزة في التقدير للقيمة الحقيقية للمجتمع، فقد تكون متزايدة عندما يتم حسابها من إحصاءات العينة ولتصحيح قيم مربع إيتا من التحيز يستخدم المؤشر (ω^2 : Omega squared) وهو مقياس لحجم الأثر يعتمد على فحص نسبة التباين المفسر إلى التباين غير المفسر وهو أقل تحيزاً من η^2 من المعادلة (36):

$$\omega^2 = \frac{SS_{\text{(between)}} - (k - 1)MS_{\text{within}}}{SS_{\text{total}} + MS_{\text{within}}} \quad (36)$$

حيث $SS_{(between)}$: مجموع المربعات بين المجموعات، SS_{total} المجموع الكلي للمربعات MS_{within} متوسط المربعات داخل المجموعات وهو يساوي (مجموع المربعات داخل المجموعات مقسوماً علي درجات الحرية داخل المجموعات) k ، عدد المجموعات أو عدد المتوسطات. كما يمكن حساب حجم تأثير تحليل التباين للقياسات المتكررة repeated measures أيضاً باستخدام المقدر غير المتحيز ω^2 من المعادلة (37).

$$\omega^2 = \frac{df_{ef} (ms_{ef} - ms_{er})}{SS_{total} + ms_{er}} \quad (37)$$

حيث: df_{effect} درجات حرية الأثر وتساوي (عدد المجموعات الفرعية للمتغير -1).

حيث SS_{total} المجموع الكلي للمربعات

حيث MS_{error} متوسط مربعات الخطأ أو (متوسط المربعات داخل المجموعات)

حيث ms_{ef} متوسط مربعات المتغير المستقل.

إن افتراض تساوي مجموعتين وتجانس التباين مهم عند حساب قيم مربع أوميغا ω^2 إلا أنه في حالة تحليل التباين في اتجاه واحد one way ANOVA قد يواجه مربع أوميغا ω^2 مشكلة انتهاك افتراض تساوي المجموعات مع تحقق افتراض تجانس التباين أو قد يواجه العكس انتهاك افتراض تجانس التباين مع تحقق افتراض تساوي المجموعات، وفي حالة انتهاك الافتراضين معاً (أي في حالة عدم تساوي المجموعات، وعدم تجانس التباين) سيؤدي ذلك إلي تقدير مبالغ فيه لحجم الأثر Effect Size overestimation عندئذ يفضل استخدام مقاييس حجم الأثر المصححة، لأنها تعزل إحصائياً التحيز الصاعد في تقدير القيمة الحقيقية للمجتمع من بيانات العينة، والسبب في ذلك أنها تأخذ في الحسبان العوامل التي تؤدي إلي أخطاء المعاينة، مثل كثرة عدد متغيرات الدراسة فكلما زاد عدد المتغيرات في الدراسة ارتفع خطأ المعاينة وكذلك صغر حجم العينة فكلما صغر حجم العينة ارتفع خطأ المعاينة. والمعادلة رقم (36) تستخدم في حالة تحليل التباين الأحادي أما في حالة التباين في اتجاهين أو أكثر فيتم استخدام معادلة مربع أوميغا الجزئية (38)[20].

$$\omega_p^2 = \frac{SS_{effect} - (df_{effect})MS_{error}}{SS_{effect} + (N - df_{effect})MS_{error}} \quad (38)$$

حيث: ω_p^2 : partial omega squared وهو مربع أوميغا الجزئي.

حيث: SS_{effect} مجموع المربعات بين المجموعات، N العدد الكلي لأفراد العينة

MS_{error} متوسط مربعات الخطأ أو (متوسط المربعات داخل المجموعات)

df_{effect} درجات حرية الأثر وتساوي (عدد المجموعات الفرعية للمتغير -1).

حيث أنه يمكن أن تتأثر جميع أحجام التأثيرات الأربعة (ω_p^2 ، η_p^2 ، ω^2 ، η^2) بتصميم البحث الذي يقوم عليه نموذج ANOVA حيث كلما زاد عدد المتغيرات فإن النسب التي يشرحها أي متغير ستخفض تلقائياً مما يؤدي إلي صعوبة المقارنة لذا اقترح كلاً من [21] مقياس لحجم الأثر η_G^2 مربع إيتا المعمم (Generalized Eta Squared)، مربع أوميغا المعمم ω_G^2 (Generalized Omega Squared)، والذي لا تتأثر قيمتهما بتصميم الدراسة ويتم تقدير مربع إيتا المعمم η_G^2 من خلال المعادلة (39):

$$\eta_G^2 = \frac{SS_{effect}}{\delta(SS_{effect}) + \sum SS_{Measured} + \sum SS_{Subjects}} \quad (39)$$

حيث SS_{effect} مجموع مربعات التأثير، $\delta = 1$ في حالة التحيز، $\delta = 0$ في حالة عدم التحيز، SS_{Meas} هو مجموع المربعات لجميع مصادر التباين التي تتضمن عوامل مقاسة بخلاف تلك التي تتضمن الموضوعات، $SS_{subjects}$ مجموع المربعات لمصادر التباين التي تتضمن الموضوعات.

كما يمكن تقديم المؤشر ω_G^2 مربع أوميغا المعمم (Generalized Omega Squared) من خلال المعادلة (40):

$$\omega_G^2 = \frac{SS_{effect} - (df_{effect})(MS_{error})}{\delta[(SS_{eff}) - (df_{eff})(MS_{err})] + \sum [SS_{meas} - (df_{meas})(MS_{err})] + (N)(MS_{s_{Cells}})} \quad (40)$$

2.5. إجراءات الدراسة

وصف وبناء أداة الدراسة: تمثلت أداة الدراسة في استمارة خاصة لجمع البيانات من عينة مجتمع الدراسة، وقد اعتمد الباحث لتصميم الأداة بعد قراءة أحدث ما كتب في أدبيات الموضوع الخطوات التالية:

- 1- تحديد أهم مشكلات الدلالة الإحصائية وتفسيراتها في البحث التربوي والنفسي.
- 2- تفسير القيمة الاحتمالية P-value والمستوي الدلالي الخاص بها.

- 3- تحديد أهم مقاييس حجم الأثر المقابلة للاختبارات الإحصائية والتي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي.
- 4- تم عرض الأداة على مجموعة من أعضاء هيئة التدريس بعدد من الأقسام التربوية المختلفة بجامعة (سوهاج - قنا - أسيوط) وقد شمل التعديل تقديم المحكات والمعايير المفسرة لحجم الأثر وشرح بعض معادلات حجم الأثر.
- 5- قام الباحث بعد عرض الأداة على بعض السادة المحكمين بإعادة بناء الأداة بشكلها النهائي.

3. النتائج وتفسيرها

بعد تفرغ استمارات جمع البيانات من عينة الدراسة، وإجراء المعالجة الإحصائية اللازمة جاءت نتائجها وتفسيرها كما يلي:

نتائج المحور الأول: حجم الأثر مقارنة بالفرضية الصفرية H_0 .

جدول رقم 1. يوضح عدد الفرضيات الإحصائية الدالة وغير الدالة موزعة حسب الاختبار الإحصائي وحجم الأثر

حجم الأثر	دال		غير دال		المجموع		المجموع الكلي	
	F	t	r	t	F	t	r	t
أقل من 0.2	55	23	51	140	178	163	34.1%	481
-0.2	95	62	110	63	116	125	22.2%	385
0.49	61	100	148	4	71	104	13.6%	328
-0.5	157	288	75	-	157	288	30.1%	520
0.79	368	473	384	207	522	680	100%	1714
0.8 فما فوق								
المجموع	368	473	384	207	522	680	100%	1714

نلاحظ من الجدول 1. أنه من الملفت للنظر أن عدد الفرضيات التي احتوت على حجم أثر أقل من الصغير بلغت 129 دالة إحصائياً وعدد 352 فرضية غير دالة إحصائياً بنسبة بلغت 28.06% من المجموع الكلي للفرضيات 1714 المقابلة للأساليب الإحصائية الثلاث وت تحليل التباين ومعامل ارتباط بيرسون وتتفق هذه النتيجة مع دراسة [22] والتي أشارت إلى أن السبب الذي جعل من حجم الأثر أقل من الحد الأدنى لحجم الأثر الصغير حسب معايير كوهين هو حجم العينة حيث أن زيادة عدد أفراد العينة يساعد على رفض الفرضية الصفرية H_0 حتى لو كان حجم الأثر صغيراً ومن هنا تم تصنيف الدلالة العملية إلى أربع مستويات (دلالة عملية كبيرة -متوسطة - صغيرة - صغيرة جداً) حيث تم إضافة مستوي رابع لمستويات كوهين الثلاثة يحتوي على مقادير حجم الأثر التي تقل عن الحد الأدنى لحجم الأثر الصغير والدالة إحصائياً في ذات الوقت حيث تبين وجود حجم أثر أقل من 0.2 لذا تم إضافة المستوي الرابع (أقل من 0.2) لمستويات كوهين الثلاثة، كما تم تحديد نتيجة الدلالة الإحصائية (دال-غير دال) لجميع الاختبارات الإحصائية.

كما نجد أن الاختبارات الدالة إحصائياً بلغ عددها 1225 اختباراً مقابل لمختلف مستويات الدلالة العملية وبنسبة بلغت 71% من إجمالي 1714 اختبار والاختبارات الغير دالة إحصائياً عددها 489 اختبار بنسبة بلغت 29%، وأن أكثر الفرضيات الدالة إحصائياً لها مستوي دلالة عملية 0.8 فما فوق أي أن لها دلالة عملية كبيرة وبلغ عددها 520 فرضية.

ومن خلال الجدول وبنسبة لاختبار ت (t) فإن عدد الفرضيات التي اقترنت بحجم أثر أقل من الصغير بلغت (163) من أصل (680) فرضية بنسبة تعادل (24%)، كما بلغ عدد الفرضيات التي اقترنت بحجم أثر صغير (125) بنسبة تعادل (18.4%)، بينما بلغ عدد الفرضيات التي اقترنت بحجم أثر متوسط (104) بنسبة تعادل (15.3%) من النسبة الكلية للفرضيات، أما بالنسبة للفرضيات التي اقترنت بحجم أثر كبير فما فوق فقد بلغ عددها (288) فرضية أي ما تعادل نسبته (42.4%) من العدد الكلي للفرضيات الخاصة باختبارات، كما يُلاحظ من الجدول (13) أن النصيب الأكبر للاختبارات الإحصائية الغير دالة كان من نصيب اختبارات (t)، والعكس في ذلك لاختبارات معامل الارتباط بعدد 128 اختباراً غير دال إحصائياً بنسبة 25% من عدد الاختبارات الدالة وغير الدالة لاختبارات معامل ارتباط بيرسون.

كما نلاحظ من خلال الجدول رقم (1) أن عدد الفرضيات لاختبار تحليل التباين التي احتوت على حجم أثر أقل من الصغير (178) فرضية من أصل (522) فرضية بنسبة (34.1%) من الفرضيات، كما بلغ عدد الفرضيات التي اقترنت بحجم أثر صغير (116) فرضية بنسبة (22.2%) و (71) فرضية اقترنت بحجم أثر متوسط أي ما يعادل نسبته (13.6%) من الفرضيات لاختبار تحليل التباين (F)، (157) من الفرضيات اقترنت بحجم أثر كبير أي ما تعادل نسبته (30.1%).

أما بالنسبة لاختبار معامل الارتباط البسيط ليرسون (r) نلاحظ أن عدد الفرضيات التي اقترنت بحجم أثر أقل من الصغير بلغت (140) من أصل (522) فرضية للاختبار (r) بنسبة بلغت (27.3%)، كما بلغ عدد الفرضيات الإحصائية التي احتوت على حجم أثر صغير (144) بنسبة بلغت (28.2%) من العدد الكلي للفرضيات الإحصائية لاختبار (r)، بينما بلغ عدد الفرضيات التي اقترنت بحجم أثر متوسط (153) فرضية إحصائية بنسبة بلغت (29.9%) من العدد الكلي للفرضيات أما بالنسبة للفرضيات التي اقترنت بحجم أثر كبير فقد بلغ عددها (75) فرضية من النسبة الكلية للفرضيات بنسبة تعادل (14.6%) من عدد الفرضيات الكلي لاختبار (r)، ويُلاحظ من خلال الجدول (13) نجد أن ثلاثة أرباع الاختبارات المستخدمة في رسائل ماجستير

بالأقسام التربوية الثلاث (علم النفس التربوي- طرق تدريس لغة عربية- صحة نفسية) كانت ذات دلالة إحصائية وأن أقل هذه الاختبارات الدالة عدداً يقابل دلالة عملية ضعيفة جداً، وأكثرها عدداً يقابل دلالة عملية متوسطة، أما بالنسبة للاختبارات الغير دالة إحصائياً فقد كان أكثرها يقابل دلالة عملية ضعيفة جداً.

نتائج المحور الثاني: التكامل بين الدلالة العملية والإحصائية

أولاً: تطبيق التكامل بين الدالتين من خلال فروض اختبار (t):

الفرضية الأولى: "توجد فروق ذات دلالة إحصائية في مقياس عمليات تحمل الضغوط في بُعد "إعادة التفسير" تبعاً لاختلاف النوع"، كما قدم الباحث بخصوص الفرضية الأولى المخرجات الإحصائية في الجدول التالي:

جدول رقم 2. يوضح المخرجات الإحصائية في مقياس عمليات تحمل الضغوط في بُعد (إعادة التفسير) تبعاً لاختلاف النوع الخاصة بإحدى فرضيات رسائل الماجستير لعينة الدراسة

البُعد	n ₁	\bar{x}_1	σ_1	n ₂	\bar{x}_2	σ_2	t _{cal}	t _{tab}	α
إعادة التفسير	84	13	2.4	64	12.4	3	2.2	1.9	0.05

التعليق على الجدول 2. وفق الطريقة الكلاسيكية والتي تغلب على البحوث النفسية والتربوية: هناك فروق ذات دلالة إحصائية بين الذكور والإناث في مقياس عمليات تحمل الضغوط في بُعد (إعادة التفسير) وتكون لصالح الذكور. نقد للتعليقات وكتابة جدول المخرجات: بالرغم من أن هناك فروق ذات دلالة إحصائية أي أنه يوجد أثر للبرنامج أو المتغير المستقل ألا أننا لا نعلم مقدار حجم هذا الأثر هل هو (صغير- متوسط - كبير) أي أنه لا توجد دلالة عملية تؤيد الدلالة الإحصائية، كما أن الباحث في تعليقه على الجدول (34) لم يقدم (ت المحسوبة أو ت الجدولية أو درجة الحرية أو المقارنة بين المحسوبة والجدولية عند مستوي الدلالة المذكور) لذا سوف نعيد كتابة الجدول (35) بطريقة نوضح فيها الدلالة الإحصائية وحجم الأثر (الدلالة العملية) ليكون التعليق أكثر وضوحاً:

جدول رقم 3. يوضح تعديل لجدول (2) المخرجات الإحصائية في مقياس عمليات تحمل الضغوط في بُعد (إعادة التفسير) تبعاً لاختلاف النوع الخاصة بإحدى فرضيات رسائل الماجستير لعينة الدراسة

البُعد	n ₁	\bar{x}_1	σ_1	n ₂	\bar{x}_2	σ_2	t _{cal}	t _{tab}	α	الدلالة	Cohen's d
إعادة التفسير	84	13	2.4	64	12.4	3	2.2	1.9	0.05	دال	0.224

حيث تم حساب حجم الأثر باستخدام المعادلة (9) Cohen's d كالتالي:

$$\text{Cohen's } d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\text{pooled}}} \quad (9)$$

$$\sigma_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$$

بالتعويض في المعادلة (9) من معطيات الجدول (2):

$$\text{Cohen's } d = \frac{13 - 12.4}{\sqrt{\frac{(84 - 1)(2.4)^2 + (64 - 1)(3)^2}{84 + 64 - 2}}} = 0.224$$

التعليق على الجدول 3. وفق الطريقة الحديثة التي تبرز التكامل بين الدالتين الإحصائية والعملية: هناك فروق دالة إحصائية في مقياس عمليات تحمل الضغوط في بُعد (إعادة التفسير) تبعاً لاختلاف النوع وهذا لأن قيمة ت المحسوبة تساوي 2.2 وهي أكبر من قيمة ت الجدولية أو الدرجة والتي تساوي 1.9 عند درجة حرية تساوي 146 وهذا يدل على أن هناك تأثير للبرنامج أو المتغير المستقل في مقياس عمليات تحمل الضغوط في بُعد (إعادة التفسير) تبعاً لاختلاف النوع لصالح الذكور ويقدر حجم هذا الأثر وفق معامل Cohen's d بـ 0.224 وهو أثر صغير.

4. الاستنتاجات

- 1- بلغ عدد الرسائل الخاصة بقسمي علم النفس التربوي ومناهج وطرق تدريس اللغة العربية والصحة النفسية في كلية التربية بجامعة سوهاج في الفترة (2000-2020) (67) رسالة ماجستير.
- 2- عدد اختبارات الفروض المعتمدة في الدراسة (1714) فرضاً لاختبارات الفروض الإحصائية (ت وتحليل التباين ومعامل ارتباط بيرسون) لرسائل الماجستير.
- 3- عدد الرسائل التي تم فيها حساب الدلالة العملية إلى جانب الدلالة الإحصائية (17) رسائل من أصل (67) بنسبة (25.37%).

- 4- عدد الرسائل التي حسبت حجم الأثر في قسم مناهج وطرق تدريس لغة عربية 9 رسائل بنسبة 53% من إجمالي عدد الرسائل التي حسبت حجم الأثر وتأتي في المرتبة الأولى للرسائل التي حسبت حجم الأثر.
- 5- يأتي اختبار تحليل التباين كأعلى عدد اختبارات الفروض التي حسبت الدلالة العملية بعدد (84) اختبار بنسبة مئوية (44%) من مجموع اختبارات الفروض التي حسبت حجم الأثر لقسم علم النفس التربوي.
- 6- اكتفاء الكثير من الباحثين بالدلالة الإحصائية بأنه يوجد أثر، وعدم وجود الدلالة يعني غياب الأثر، ولكنه ليس من الضروري أن غياب الدلالة الإحصائية يعني غياب الدلالة العملية، فلم توجد رسالة واحدة حسبت حجم الأثر لنتيجة غير دالة إحصائياً

5. التوصيات والدراسات المستقبلية

من خلال النتائج التي أسفرت عنها الدراسة الحالية خرج الباحث بمجموعة من التوصيات:

- 1- نشر الثقافة الإحصائية في أوساط الباحثين وطلبة الدراسات العليا من خلال دورات إلزامية في مجال التحليلات الإحصائية وتدريب مادة الإحصاء من ذوي الخبرة والمعرفة بالأساليب الإحصائية المتقدمة وكيفية الاستفادة العملية من هذه الأساليب.
- 2- تخصيص وحدة في المؤسسات الأكاديمية لتقديم الاستشارات الإحصائية والبحثية للباحثين وأعضاء هيئة التدريس.
- 3- إدراك المفهوم الحقيقي للدلالة الإحصائية، وعدم التحيز لتحليلات إحصائية دون غيرها وتقديم النتائج حتى ولو كانت غير دالة إحصائياً.
- 4- ضرورة تضمين تدريس طرق حساب حجم الأثر في المقررات الدراسية وخاصةً لطلاب الدراسات العليا.
- 5- حث الباحثين على ضرورة حساب حجم الأثر الخاصة باختبارات الدلالة الإحصائية لفروضهم العلمية وعدم الاكتفاء فقط بالدلالة الإحصائية وتضمين ذلك في ملخصات دراساتهم.
- لا بد من إعطاء معلومات شاملة ودقيقة ووافية عن أهمية المشكلات المدروسة وعن تصميماتها البحثية وتقديم معلومات ومؤشرات قوية عن الخصائص السيكومترية للمقاييس المستخدمة.

6. قائمة المراجع

أولاً: المراجع العربية

- [2] على الثبيتي (2006). تصاميم البحوث العلمية ودورها في صدق نتائج الدراسات التربوية، مجلة رسالة الخليج العربي، 29 (108)، 60-13.
- [11] عزت عبد الحميد (2011). الإحصاء النفسي والتربوي تطبيقات باستخدام برنامج SPSS18. القاهرة: دار الفكر العربي.
- [15] فؤاد أبو حطب، آمال صادق (1991). مناهج البحث وطرق التحليل الإحصائي في العلوم النفسية والتربوية والاجتماعية، القاهرة: الأنجلو المصرية.
- [22] حمدي أبو جراد (2013). قوة الاختبارات الإحصائية وحجم الأثر في البحوث التربوية المنشورة في مجلة القدس المفتوحة للأبحاث والدراسات. مجلة العلوم التربوية والنفسية، 2، (14)، 368-349.
- ثانياً: المراجع الأجنبية

- [1] Morris, S. & Shokri, A. (2020). Effect size and Effect Size Uncertainty in Organizational Research Methods. doi:10.1093/acre/fore/9780190224851.013.238.
- [3] Heyvaert, M. & Onghena, P. (2014). Analysis of single-case: Randomization tests for measures of effect size. <https://doi.org/10.1080/09602011.2013.818564>.
- [4] Wasserstein, R. L., and Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose," The American Statistician, 70, 129–133.
- [5] Frick, R. W. (1996). The appropriate use of null hypothesis testing. Psychological Methods, 1, 379–390.
- [6] Vaske, J. J., Gliner, J. A., & Morgan, G. A. (2002). Communicating judgments about practical significance: effect size, confidence intervals and odds ratios. Human Dimensions of Wildlife, 7, 287-300.
- [7] Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. Ed.). New York: Academic Press.
- [8] Kirk, R. E (1996). Practical Significance: A Concept whose time has come. Educational and Psychological Measurement, 56(5), 746-759.

-
- [9] Rosnow, R. & Rosenthal, R. (2003). Effect sizes for experimental psychologists. *Canadian Journal of Experimental Psychology*, 57(3): 221-237.
- [10] Hedges, L.V., Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- [12] Cooper, H., & Hedges, L. (1994). *The Handbook of Research Synthesis*. New York: Russel Sage Foundation.
- [13] Sun, S. (2008). A Comprehensive Review of Effect Size reporting and Interpreting Practices in Academic Journal in Education and Psychology. *College of education, University of Cincinnati*.
- [14] Fritz, c. o, Morris, P.E., & Richler, J. J. (2012). Effect size estimates: current use, calculation, and interpretation. *Journal of Experimental psychology: General*.2012, 141(1):2-18.
- [16] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative sciences: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*. 4:863.
- [17] Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.
- [18] Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences*, 21, 554-559.
- [19] Volker, M. (2006). Reporting effect size estimates in school psychology research. *Psychology in the schools*, 43, 6,653-672.
- [20] Steyn, H. S. & Ellis, S. M. (2009). Estimating an Effect Size in one-way multivariate analysis of variance (MANOVA). *Multivariate behavioral research*, 44(1), 106-129.
- [21] Olejnik, S. & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.

English abstract

*Article***Integration between the Practical significance and Statistical significance of the tests "t and analysis of variance and correlation coefficient"**

Kamiel Azmy, Eman Khalaf, Mahmoud Alamin

Department of Educational Psychology, Faculty of Education, Sohag University, Sohag 82786, Egypt

*Correspondence Author: mahmoud.185983@azhar.moe.edu.eg**Abstract**

The study sought to integrate the practical significance and statistical significance in the quantitative analysis of Educational Research. The study used a content analysis approach, focused on providing details about the measures of effect size that are associated with the most common statistical significance tests. The study used a form to collect data from research for analysis according to the concepts of statistical and practical significance. The study was applied to a sample of (67) master's theses in the educational departments at the Faculty of Education, Sohag University in the period from 2000 to 2020. The study came out with a set of results, including the number of messages in which the practical significance was calculated along with the statistical significance (17) messages out of (67) by (%25.37), the number of statistical tests that calculated the practical significance along with the statistical significance (301) tests out of a total of (1714) tests by (17.6%), and the number of statistically significant tests (1225) by (%71.47) for statistical hypothesis tests (t, Analysis of variance and correlation coefficient the number of statistical hypotheses with a weak practical significance is 489 (28.53%), and the number of statistical hypotheses with a weak practical significance is 385 (22.46%). The study provided proposed recommendations for the integration of practical and statistical significance in the analysis of Educational Research.

Keywords: Practical Significance, Statistical Significance, Effect Size