

ما وراء التحليل للأبحاث المنشورة في المجلة الأردنية في العلوم التربوية: الدلالة العملية وقوة الاختبار.

د. نضال كمال محمد الشريفيين*

الملخص

هدفت هذه الدراسة إلى تقويم الدلالة الإحصائية والدلالة العملية وقوة الاختبارات الإحصائية لجميع الدراسات المنشورة في المجلة الأردنية في العلوم التربوية التي استخدمت التصاميم الكمية، وتكونت عينة الدراسة من (1363) اختبار الإحصائي من نوع $(t, F_1, r, \chi^2, R^2, Z, F_2)$. وتم حساب كل من $(d, f, r^2, C, F^2, q, T^2)$ كمؤشرات للدلالة العملية (حجم الأثر)، ومن ثم استخراج قوة الاختبارات الإحصائية من جداول كوهن. أشارت نتائج الدراسة أن ما نسبته (75.79%) من الفرضيات اقترنت بدلالة عملية صغيرة، و(10.86%) بدلالة عملية متوسطة، كما وجد أن ما نسبته 54.86% من الفرضيات التي تم اختبارها قد اقترنت بمستوى قوة اختبار متدنية أو متوسطة. وبيّنت النتائج عدم وجود استقلالية للدلالة الإحصائية عن كل من: الدلالة العملية وقوة الاختبار الإحصائي، وأوصت الدراسة بضرورة أن يتبع حساب الدلالة الإحصائية حساب الدلالة العملية، وقوة الاختبار الإحصائي لإعطاء تفسيرات أفضل للنتائج.

الكلمات المفتاحية:

الدلالة العملية، الدلالة الإحصائية، حجم الأثر، قوة الاختبار الإحصائي، ما وراء التحليل.

* قسم علم النفس الإرشادي والتربوي - جامعة اليرموك - كلية التربية، الأردن.

1- المقدمة وخلفية الدراسة

يُعد البحث عمليةً منظمةً للتوصل إلى حلول لمشكلات، أو إجابات عن تساؤلات، تستخدم فيها أساليب في الاستقصاء والملاحظة مقبولة ومتعارف عليها بين الباحثين في مجال معين، ويمكن أن تؤدي إلى معرفة جديدة (الكيلاني والشريفين، 2011). وإن إجراء عملية البحث ليس أمراً سيبيراً، وإنما يتطلب ذلك الملاحظة والوصف والتحليل والتفسير المناسب، ويستخدم الباحثون فيه عادةً أدوات قياس عديدة وأشكال وصف دقيقة، كما يقومون باختيار الوسائل المناسبة لتجميع البيانات، والقيام بإجراءات ملائمة في تحليلها وتفسيرها، وذلك بهدف الوصول إلى قرارات قابلة للتعميم. وإن ما يجب التنويه إليه هنا أن جمع البيانات في حد ذاته ليس غاية يُسعى إليها، بل هي وسيلة لتحقيق أهداف معينة؛ أي أن عمل الباحث لا ينتهي عند جمع البيانات، بل على العكس من ذلك، حيث إن عمله يتبدى بها، ولكن هذه البيانات هي مادة خام لا تعطي ثمارها، ما لم يوجد الباحث والمحلل الجيد الذي يستطيع استثمارها على الوجه الأمثل. ولكن تحليل البيانات وتفسيرها يتأثر بعدة عوامل، تتعلق بمدى معرفة الباحث بالأساليب الإحصائية الملائمة لتحليل بيانات بحثه، ودرجة فهم الباحث للإحصاءات بوصفها رموزاً لظواهر تكمن وراءها. فالإحصاء وسيلة من وسائل البحث، إذا أحسن استخدامها فإنها تُعطي نتائج يعتمد عليها، أما إذا أُسيء استخدامها فإنها توصل إلى نتائج مضللة (النجار، 2005).

وغالباً ما يستخدم الباحثون مبادئ الإحصاء الاستدلالي (Inferential Statistics)، التي تركز على الوصول إلى استنتاجات حول خصائص المجتمع من خلال استخدام المعلومات المتوفرة عن العينة، من أجل فحص فرضياتهم المتعلقة بالظواهر التي يدرسونها، وذلك من أجل اتخاذ قرارات معينة. يمر الباحث في عملية فحصه للفرضية بعدة خطوات تبدأ بصياغة الفرضيات، حيث يصوغ نوعين منها هما:

- **الفرضية الصفرية (Null Hypothesis):** وهي التي تخضع للاختبار الإحصائي التي ينبغي أن تكون مبنية على مفاهيم معرفة تعريفاً إجرائياً، بحيث يمكن جمع البيانات حولها عن طريق الملاحظة والتجربة، ويتم إجراء الاختبار الإحصائي على هذه البيانات، ويُرمز لها (H_0) . ومن أمثلتها (لا توجد فروق في نسب النجاح في التوجيهي تُعزى لجنس الطالب).

- الفرضية البديلة (**Alternative Hypothesis**): وهي الفرضية التي يمكن أن تُقبل كبديل للفرضية الصفرية، عندما يتم رفض الفرضية الصفرية، ولا تخضع للاختبار الإحصائي، وقد تكون متجهة أو غير متجهة، فإذا كانت متجهة (**Directional Hypothesis**) عندها ينبغي أن يكون الاختبار بديل واحد، أي أن منطقة الرفض في جهة واحدة، ففي بعض المواقف ربما يُفضل الباحث الإشارة إلى وجود فرق لصالح جهة دون أخرى، أو تحديد اتجاه العلاقة أو اتجاه الفرق بين متوسطين ... الخ، أما إذا كانت غير متجهة (**Non-Directional Hypothesis**) عندها ينبغي أن يكون الاختبار بديلين أي أن منطقة الرفض في الجهتين، وهنا يشير الباحث إلى وجود فرق بين مجموعتين أو أكثر، ولكن لا يحدد لصالح من هذا الفرق (عودة والخليلي، 1998).

وبعد أن يتم إخضاع الفرضية الصفرية للاختبار يُتخذ قرار بحق الفرضية الصفرية، وهذا القرار يكون إما برفضها (**Reject**) أو بعدم رفضها (**Not-Reject**)، ويتم اتخاذ القرار في ضوء الاختبار الإحصائي المستخدم، باعتباره قانوناً أو قاعدة تربط إحصائي العينة (**Statistic**) مع معلمة المجتمع (**Parameter**)، ومن أمثلة الاختبارات الإحصائية المستخدمة اختبار Z (**Z-test**)، اختبار t (**t-test**)، اختبار χ^2 (**Chi-square test**) ... الخ، وفي الواقع لا يستطيع الباحث التأكد فيما إذا كان القرار الذي اتخذه صائباً أم غير صائب، وكل ما يعرفه هو احتمال صواب هذا القرار، فإذا كان الفرض صحيحاً، ولم يتمكن من رفضه (مستوى الثقة) أو كان خاطئاً بالفعل، ورفضه الباحث (قوة الاختبار)، فإنه في كلتا الحالتين يكون قد اتخذ قراراً صائباً، وإذا رفض القرار وهو في الواقع صحيح (الخطأ من النوع الأول) (**Type I Error**)، أو لم يتمكن من رفضه، في حين أنه في الواقع خاطئ (الخطأ من النوع الثاني) (**Type II Error**)، عندئذٍ يكون الباحث قد وقع في خطأ، لذا فإنه أياً كان القرار هناك درجة من الخطأ، وينبغي ألا ينظر إلى اختبار الفرضيات باعتباره برهنة لها (الكيلاني والشريفين، 2011).

وعادة ما يكون الباحث مدفوعاً برغبة قوية إلى رفض H_0 ، لأن الفشل في رفض H_0 قد يُعد مؤشراً لفشل الباحث في برهنة ما يريد إثباته، لذلك فهو يُفضل رفض الفرضية الصفرية H_0 على عدم رفضها (جعفر، 1998). وقد أشار ولكنسون (Wilkinson, 1999) إلى أن لدى رؤساء تحرير المجلات العلمية تمييزاً ملموساً وغير مبرر تجاه الدراسات التي تُفضي نتائجها إلى وجود دلالة إحصائية،

مقارنة بالدراسات التي تُفضي نتائجها لعدم وجود دلالة إحصائية، متجاهلين بأن تلك النتائج سواء التي أشارت لوجود دلالة أم التي لم تشر لوجود دلالة، كلاهما مهمة من وجهة نظر علمية.

فاحتمال رفض الفرضية الصفرية وهي صحيحة، يسمى بالخطأ من النوع الأول، وأقصى احتمال لوقوع الباحث في هذا النوع من الخطأ، هو ما يُسمى بمستوى الدلالة الإحصائية، ويُرمز لها بالرمز (α) ، ومعنى الدلالة الإحصائية أن الفرق بين القيمة النظرية لمعلمة المجتمع والقيمة الناتجة من العينة، هو فرق حقيقي، ولا يعزى إلى الصدفة. أما احتمال عدم رفض الفرضية الصفرية وهي صحيحة، يُسمى بمستوى الثقة $(1 - \alpha)$. بينما احتمال رفض الفرضية الصفرية وهي في الواقع خاطئة، يُسمى بقوة الاختبار الإحصائي (Power of Statistical test)، ويُرمز له بالرمز $(1 - \beta)$ ، أما احتمال عدم رفض الفرضية الصفرية وهي في الواقع خاطئة، يُسمى بالخطأ من النوع الثاني، واحتمال وقوع الباحث في هذا النوع من الخطأ يُرمز له بالرمز β .

والباحث عادةً يرغب بتصميم دراسة تتسم بدرجة عالية من القوة، بمعنى أنه عندما يختبر الفرضية الصفرية، فإنه يحاول دائماً اتخاذ القرار الصائب، فهو يرغب في رفض الفرضية الصفرية الخاطئة، وقبول الفرضية الصفرية الصحيحة؛ بمعنى أنه يرغب بالتقليل من نوعي الخطأ، وهذا أمر غير ممكن. ولكن الباحثين يتخذون قراراتهم استناداً إلى مستوى الدلالة α ، حتى إن بعض الباحثين - في بعض الدراسات - يضعون $(\alpha = 0.00)$ ، ولكنهم لا يضعون في حساباتهم ما يترتب على ذلك من زيادة فرصة الوقوع في الخطأ من النوع الثاني. لذا ينصح الإحصائيون بأن يُراعي الباحث عند اختياره لمستوى الدلالة طبيعة الدراسة وفيما تستخدم نتائجها، فإذا كانت الدراسة من النوع الذي يؤدي فيه رفض الفرضية الصفرية إلى مخاطر أو إنفاق أموال طائلة أو إهدار للوقت والجهد، فإنه ينبغي الحذر من هذا النوع من الخطأ، بأن يضحى الباحث إلى حد ما بقوة الاختبار من أجل تقليل هذا الخطأ، كأن يجعل $(\alpha = 0.01)$ أو ربما $(\alpha = 0.001)$ ، وإذا كانت الدراسة استطلاعية، ويحاول الباحث التأكد إن كانت هناك حاجة لمزيد من الدراسات حول الموضوع أو الظاهرة، عندئذٍ فإن زيادة قوة الاختبار يعتبر أمراً مهماً، بينما الخطأ من النوع الأول أقل أهمية (علام، 1993).

فعندما يتخذ الباحث قراراً برفض الفرضية الصفرية، فهذا يعني أن هناك أثراً للمتغير المستقل على المتغير التابع، أو أن الفروق على المتغير التابع التي تعود إلى المتغير المستقل ليست صفراً ولا تعود للصدفة،

ويكون اتجاه هذا الأثر وفق ما حدد في الفرضية البديلة. وهناك عوامل كثيرة تُساهم في رفض الفرضية الصفرية أو عدم رفضها منها: مستوى الدلالة الإحصائية (α)، وحجم العينة، وحجم الأثر (Effect Size)، وقوة الاختبار الإحصائي (Test Power) ($1 - \beta$) (Cohen, 1988). فرفض الفرضية الصفرية لا يُقدم الكثير من المعلومات، فما يقدمه هو أن أثر المتغير المستقل الذي أُطلق عليه كوهن (Cohen) حجم الأثر، ليس صفرًا، ولا يعود إلى الصدفة (Konstantopoulos, 2008).

اهتم علماء الاجتماع بالإجابة عن ثلاثة أسئلة أساسية عندما يفحصون العلاقات بين المتغيرات (Kirk).

- هل الأثر الملحوظ حقيقي أم يجب عزوه للصدفة (الدلالة الإحصائية)؟
- إذا كان الأثر حقيقياً فكَم هو حجمه (حجم الأثر)؟
- هل الأثر كبير بما يكفي ليكون نافعاً (الدلالة أو الأهمية العملية)؟

فقد تُشير نتائج اختبار الفرضية الصفرية إلى وجود فروق ذات دلالة إحصائية بين متوسطات المعالجة مثلاً. ولا تُعزى هذه الفروق إلى الصدفة أو خطأ المعاينة (Sampling Error)، أي أن المتغير المستقل له أثر في المتغير التابع، ولكن يبقى السؤال، ما حجم هذا الأثر؟ هل هو تأثير قوي أو صغير؟ فأحياناً إذا كانت قيمة الاختبار الإحصائي ذات دلالة عند احتمال (0.05) توصف بأنها ذات دلالة فقط، ولكن إذا كانت ذات دلالة عند احتمال (0.01) توصف بأنها ذات دلالة قوية، وقد يوحي ذلك بأن تأثير المعالجة أكبر. إلا أن هذا التفسير غير صحيح، فاختبار الدلالة الإحصائية (Statistical Significance) ليس مقياساً ملائماً لقياس تأثير المعالجة، لأن الدلالة الإحصائية لها علاقة بحجم العينة، إذ إنه بزيادة حجم العينة بشكل كبير تصبح إمكانية رفض الفرضية الصفرية أكبر، لذا فإنه ومن أجل إيجاد قوة العلاقة بين المتغير المستقل والمتغير التابع أو بمعنى إيجاد حجم الفروق في المتغير التابع الناتجة عن اختلاف مستويات المتغير المستقل، أو مدى مساهمة المتغير المستقل في التباين في المتغير التابع، يتم اللجوء إلى ما يُسمى بالدلالة العملية (Practical Significance).

الدلالة العملية تعني: مقدار التباين في المتغير التابع الذي يمكن تفسيره من خلال متغير مستقل يرتبط به أو يؤثر فيه (Hasse, 2001). لذا فالدلالة 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 (يعني وجود دلالة)، 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 لكبير حجم العينة (Kirk, 1996)، ومن ثم فالدلالة المتحصل عليها توقع الباحث في حيرة مع نفسه، لأنه يُلاحظ أنه على الرغم من رفض الفرضية H_0 ، فإن هذه النتيجة ليس لها أهمية عملية تُذكر، لذا فالكفاية تتحقق إذا ما حسبنا الدلالة العملية للاختبار، وهذه الدلالة تُعد مؤشراً إحصائياً لمدى القدرة على استخدام النتائج تفسيرياً أو تطبيقياً لاسيما وأن ويلكنسون (Wilkinson, 1999) أشار إلى عدم تأثر حجم الأثر بحجم العينة وهذا ما يميزه عن الدلالة الإحصائية.

وقد لخص هيوستن (Huston, 1993) مزايا مقاييس فحص الدلالة العملية في الآتي:

- إنها تُعطي تقديراً لمستوى وجود الظاهرة في المجتمع، وذلك من خلال قيمة على مقياس متصل (Continuous Scale) يبدأ من الصفر (عندما تكون الفرضية H_0 صحيحة)، وقيمة أكبر من الصفر (إذا كانت الفرضية H_0 خاطئة)، وبالتالي فهو مؤشر لمدى الانحراف عن الفرضية الصفرية.

- إنها تُعطي قيمة محددة تُشكل محكاً لوصف حجم الأثر (Effect Size) بأنه صغير أو متوسط أو كبير.

- يمكن استخدامها نتائجها كمقياس كمي للمقارنة بين دراستين.

- يمكن استخدامها في تحليل القوة لتقدير حجم العينة اللازمة لدراسة ما.

فالدلالة العملية تعني تلك القيمة الرقمية للفرق بين معالم المجتمع الحقيقية وغير المعروفة (H_s) والقيم المحددة من خلال الفرضية الصفرية (H_0)، هذا الفرق ($\mu_s - \mu_0$) يمثل أحد أهم عناصر ما يُسمى بمؤشر الدلالة العملية أو حجم الأثر، وهذا المؤشر يرتبط ارتباطاً وثيقاً بمؤشر قوة الاختبار، فكلما زاد حجم الأثر زادت قوة الاختبار. كما أن الدلالة العملية مرتبطة بدلالة الفروق الإحصائية، ولها أهميتها العملية والتطبيقية للقرار المتخذ (النجار، 2005). فقد أشار ثالهيمر وكوك (Thalheimer & Cook, 2002) أنه بينما تقدم اختبارات الدلالة الإحصائية احتمالية أن تكون النتائج التجريبية مختلفة عن النتائج التي يتم الوصول إليها بمحض الصدفة، تقدم مقاييس حجم الأثر التقدير النسبي لأثر المعالجة

التجريبية، كما تسمح أحجام التأثير بمقارنة أثر المعالجة التجريبية من تجربة لأخرى. فالدلالة العملية تُمثل الوجه الآخر للدلالة الإحصائية، لأنها تتعلق بالمزايا الفعلية الناتجة من تقييم الفروق الإحصائية المحسوبة في اختبار دلالة الفروق، فهي مؤشر لمدى القدرة على استخدام النتائج تفسيراً أو تطبيقاً (الصياد، 1988)، وهي نسبة التباين المفسر للمتغيرات موضع الدراسة (Plucker, 1997). لذا فإن الدلالة الإحصائية لا تكفي وحدها لاتخاذ قرار سليم؛ فهي ليست بديلاً عن الدلالة الإحصائية، ولكنها مكملة لها، وتعد مقياساً آخر إذ يجب على الباحث الاستعانة به بعد الاستنتاج الإحصائي لاختبار دلالة الفروق لاتخاذ القرار السليم (Park, 2008)، مما يتطلب حساب الدلالة العملية، وذلك بحساب قوة العلاقة بين المتغير المستقل والمتغير التابع للبيانات موضع البحث. فالدلالة العملية تتضمن حكم قيمة يعطيه الباحث والمستخدم للبحث حول المضامين العملية لنتائج الدراسة (Vaske, Gliner & Morgan, 2002).

ويمكن بيان أهمية عمل الدلالة العملية من خلال النقاط الآتية:

- مقياس يوضح قوة العلاقة بين متغيرين أو أكثر، وذلك من خلال قيمة واحدة ومحددة تتبع الأسلوب الإحصائي المستخدم والبال إحصائياً (Rea & Parker, 1997).
- مقياس يبين فيما إذا كان المتغير التابع مُفسراً ومتأثراً بالمتغير المستقل (Mahadevan, 2000).
- مقياس لمؤشر الفروق الحقيقية أو المعيارية بين متوسطين أو أكثر (Nix & Barnette, 1998).
- مقياس للدلالة العملية ومدى أهمية النتائج المتحصل عليها، وكونها لا ترجع للمصادفة (Mclean & Ernest, 1998).

فقد وُجِعت مجموعة من الانتقادات لاختبارات الفرضية الصفرية (Null Hypothesis Statistical Testing) وقد تركزت تلك الانتقادات حول ثلاثة محاور أساسية هي:

المحور الأول: حول مدى منطقية اختبارات الفرضية الصفرية، وتمثل الجدل وفق ما أشار إليه نيكس وبارنيت (Nix & Barnette, 1998) في "أن اختبارات الفروق في اختبارات الفرضية الصفرية تفترض مواقف غير ممكنة الحدوث، وبخاصة عندما تفترض أن الفروق في الفرضية الصفرية تساوي صفراً". فعندما يتم رفض الفرضية الصفرية في دراسة ما، فإن الباحث يكون أمام خيارين هما: إما الأخذ بما

افترضه نيكس وبارنيت بخصوص اختبار الفرضية الصفرية. وإما العمل على اختبار الأخطاء في الفرضيات الصفرية (قياس خطأ العينات).

ونتيجة لذلك تبين لكثير من الباحثين أنه بزيادة حجم العينة يمكنهم الحصول على نتائج إيجابية لدراساتهم وذات دلالة إحصائية، مما يُدعم عملية نشر هذه الدراسات في المجلات العلمية، وأنه عندما يعتمد الباحث إلى زيادة قوة الاختبار $(1 - \beta)$ يكون من الصعب الحصول على دلالة إحصائية. ومن هذا المنطلق نجد الباحث في هذه الحالة يعمل على زيادة حجم العينة للوصول إلى قوة اختبار عالية (High Level of Statistical Power)، وبالتالي فإن أية فروق بين المتغيرات - ولو كانت صغيرة - دالة إحصائياً، فالعينات الكبيرة تجعل أي فرق بسيط بين الإحصائي والمعلم المناظر له فرقاً ذا دلالة إحصائية، وبذلك يصبح الاختبار الإحصائي الأكثر حساسية، لأنه يصل إلى مستويات الدلالة الإحصائية بعينة أصغر من غيره.

فاعتماد الباحثين فقط على الدلالة الإحصائية من خلال استخدام الاختبارات الإحصائية (F، t ، χ^2 ... الخ) أو مؤشر الدلالة (p-value)، (Sig. value) فإنهم يتعدون عن معنى الأثر (Effect) (McLean & Ernest, 1998)، وكذلك استخدام كلمة دلالة (Significance) للإشارة إلى الدلالة الإحصائية فإنه يؤدي إلى سوء فهم لهذه الكلمة (Thompson & Snyder, 1997)، فبعض الباحثين يشيرون إلى الدلالة بمعنى رفض H_0 ، ولو كانت بحجم أثر صغير (Small Effect Size) فلا يتم التركيز على الدلالة العملية بقدر التركيز على الدلالة الإحصائية. لذا ينبغي على الباحثين حساب الدلالة العملية.

وأما المحور الثاني: فهو كيفية تفسير الفرضية الصفرية (Interpretation of Null Hypothesis Statistical Testing) فقد صور بلوكر (Plucker, 1997) أن سوء التفسير للفرضية الصفرية بمأثل الوقوف على حافة فجوة (يعني p-level)، والباحث عندما يرغب في عبور هذه الفجوة ينبغي عليه إيجاد الحجم المناسب لهذه الفجوة (أي حجم الأثر المناسب) قبل العبور. والإشارة إلى العلاقة بين المتغيرات المستقلة والتابعة يتيح الفرصة لحساب الدلالة العملية وهو ما يُعِينهم على تخطي تلك الفجوة التي أمامهم (Daniell, 1997; Fan, 1999).

وأخيراً المحور الثالث: حول استخدام الطرق والأساليب الإحصائية البديلة حيث لا يتحقق معظم الباحثين من افتراضات الاختبارات الإحصائية المستخدمة (Assumption of Statistical Tests)، مما يجعل البحث غامضاً وغير دقيق ويؤثر سلباً على النتائج، وإن معظم الباحثين يتوقفون عند حصولهم على نتائج ذات دلالة إحصائية، ولا يواصلون تحليلاتهم لتحديد فيما إذا كانت هذه النتائج تعود لخطأ المعاينة (Sampling Error)، أو تعود لحجم الأثر (Nix & Barnette, 1998).

فالدلالة الإحصائية تبين فيما إذا كانت الفروق في النتائج تعود إلى المصادفة أو لخطأ في القياس، أما الدلالة العملية فهي تتعلق بمعرفة المزايا الفعلية والحقيقية الناتجة عن استخدام معالجة معينة، والقدرة على تفسير حجم الاختلاف الناتج. فالدلالة الإحصائية شرط مسبق للدلالة العملية، فقد أشار ليفين (Levin, 1998) إلى أن الباحثين عندما يقومون بمناقشة الدلالة العملية في نتائج أبحاثهم، فإن ذلك سيؤثر بشكل إيجابي عند عرضهم لنتائج أبحاثهم. لذا ينبغي على الباحث أن يقدر الدلالة العملية من خلال المؤشرات الخاصة بها، ويقرر مستوى الدلالة المطلوب α ، وكذلك قوة الاختبار الإحصائي $(1 - \beta)$ ، وهذا سيساعد على تحديد حجم العينة المناسب لتحليل بيانات البحث.

وقد صنف كيرك (Kirk, 1996) مؤشرات الدلالة العملية في ثلاث فئات هي: مؤشرات قوة العلاقة (Measures of Strength of Association)، ومؤشرات حجم الأثر (Measures of Effect Size) ومؤشرات أخرى (Other Measures). أما أرون و أرون (Aron & Aron, 1997) فقد صنفاها في فئتين هما: مقياس حجم الأثر، ومقياس قوة العلاقة.

1-1 مقياس حجم الأثر

تختلف مقياس حجم الأثر باختلاف الاختبار الإحصائي المستخدم لاختبار الفرضية، وفيما يأتي أبرز هذه المقياس المستخدمة كمؤشرات لحجم الأثر في هذه الدراسة:

• في حالة استخدام اختبار t

أ- لاختبار الفرضية المتعلقة بالفرق بين وسطي مجتمعين مستقلين فإن:

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{S_c}$$
$$S_c = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

حيث تشير d : إلى حجم الأثر (Effect Size).

\bar{X}_1 : وسط العينة الأولى.

\bar{X}_2 : وسط العينة الثانية.

S_c : الانحراف المعياري المشترك للعينتين.

ب- لاختبار الفرضية الإحصائية المتعلقة بوسطي مجتمعين مرتبطين.

$$S_{x_a-x_b} = \sqrt{S_{x_a}^2 + S_{x_b}^2 - 2S_{x_a}S_{x_b}} \quad , \quad d = \frac{\bar{X}_{x_a-x_b}}{S_{x_a-x_b}}$$

حيث تشير d : إلى حجم الأثر (Effect Size).

$\bar{X}_{x_a-x_b}$: متوسط الفروق بين قيم x_a وقيم x_b المتناظرتين أو بين الاختبار القبلي والاختبار

البعدي.

S_{a-b} : الانحراف المعياري للفروق.

ج- لاختبار الفرضية المتعلقة بمتوسط حسابي لمجتمع واحد.

$$d = \frac{|\bar{X} - \mu|}{\sigma}$$

حيث \bar{X} : وسط العينة.

μ : وسط المجتمع.

σ : الانحراف المعياري للمجتمع.

وقد اعتبر كوهن (Cohen, 1988) أن حجم الأثر يكون صغيراً عندما

$$0.20 \leq d \leq 0.49 \text{ ، ومتوسطاً عندما } 0.5 \leq d \leq 0.79 \text{ ، وكبيراً عندما } d \geq 0.80$$

• في حالة استخدام اختبار فرضية تتعلق بمعامل ارتباط واحد (r)

عندما يرغب باحث بمعرفة فيما إذا كان معامل الارتباط بين متغيرين ذا دلالة إحصائية، بمعنى أن

معامل الارتباط بين المتغيرين في المجتمع الأصلي يختلف عن الصفر اختلافاً جوهرياً، أي أن المتغيرين في

الأصل مستقلان على أساس افتراضات الصدفة والعشوائية، ولم يكن سببه ناشئاً عن خطأ المعاينة، فإنه

يمكن التعبير عن حجم الأثر بمربع معامل الارتباط ($d = r^2$) (أي معامل التحديد (Coefficient of Determination)) ووفق معيار كوهن (Cohen, 1988) فإن حجم الأثر يكون صغيراً عندما $0.10 \leq r^2 \leq 0.29$ ، ومتوسطاً عندما $0.30 \leq r^2 \leq 0.49$ ، وكبيراً عندما $r^2 \geq 0.50$.

• في حالة استخدام أسلوب تحليل الانحدار المتعدد (Multiple Regression)

ويستخدم هذا الأسلوب عندما يكون الغرض التنبؤ بقيمة متغير تابع بمعلومية قيم المتغيرات المستقلة (متغيرين مستقلين أو أكثر). ويتم حساب حجم الأثر (F^2) باستخدام المعادلة الآتية:

$$F^2 = \frac{R^2}{1 - R^2}$$

حيث R^2 : مربع معامل الارتباط المتعدد.

ووفق معيار كوهن (Cohen, 1988) فإن F^2 (حجم الأثر) يُعد صغيراً عندما تكون: $0 \leq F^2 \leq 0.14$ ، ومتوسطاً عندما $0.15 \leq F^2 \leq 0.34$ ، وكبيراً عندما $F^2 \geq 0.35$.

• في حالة استخدام اختبار فرضية الفرق بين معاملي ارتباط مستقلين

لاختبار الفرضية المتعلقة بمعاملي ارتباط مستقلين نستخدم اختبار Z ، أما مؤشر حجم الأثر فهو q ويعبر عنه بالعلاقة: $q = Z_1 - Z_2$.

حيث Z_1 : علامة فشر المقابلة لمعامل الارتباط في العينة الأولى.

Z_2 : علامة فشر المقابلة لمعامل الارتباط في العينة الثانية.

ووفق معيار كوهن فإن حجم الأثر q يُعد صغيراً عندما $q \leq 0.29$ ، ومتوسطاً عندما $0.30 \leq q \leq 0.49$ ، وكبيراً عندما $q \geq 0.50$.

• في حالة استخدام اختبار F_1 وذلك عند استخدام أسلوب تحليل التباين

يمكن استخدام أسلوب تحليل التباين مهما كان عدد المتغيرات المستقلة، وهناك ميزة لأسلوب تحليل التباين تلخص بإمكانية اختبار أثر التفاعل. لذلك يعتبر تحليل التباين من أهم الأساليب المستخدمة

في البحوث في مختلف مجالات العلوم التربوية والاجتماعية. وحجم الأثر المناسب لهذا الأسلوب هو f

$$\eta^2 = \frac{SS_a}{SS_t} = \frac{SS_a}{SS_a + SS_e} . \eta^2 : \text{مربع ايتا} .$$

SS_a : مجموع المربعات للمتغير المستقل a .

SS_t : مجموع المربعات الكلي .

SS_e : مجموع المربعات للخطأ .

ومن مزايا مربع ايتا: سهولة حسابه، أما من أبرز سلبياته تأثره بكل من: مستويات المتغير

المستقل وحجم العينة (Huston, 1993).

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}} .$$

وأما حجم الأثر f فيمكن حسابه من خلال المعادلة الآتية:

ووفق معيار كوهن يُعد حجم الأثر صغيراً إذا كانت قيمة: $0 \leq f \leq 0.24$ ، ومتوسطاً عندما

$0.25 \leq f \leq 0.39$ ، وكبيراً $0.40 \leq f \leq 1$.

• في حالة استخدام اختبار F_2 عند استخدام أسلوب تحليل التباين المتعدد

ويستخدم أسلوب تحليل التباين المتعدد عندما يكون هناك أكثر من متغير تابع بغض النظر عن عدد المتغيرات المستقلة، ولفحص دلالة الفرق بين المتوسطات لكل متغير مستقل على كل متغير من المتغيرات التابعة، نستخدم إما اختبار هوتلنج (Hotlling Trace)، أو اختبار ولكس لامدا (Willks Lambda)، ومؤشر حجم الأثر في حالة استخدام One way MANOVA (أي

$$D^2 = \frac{NT^2}{n_1 n_2} : \text{المتغير المستقل بمستويين، يعطي بالعلاقة الآتية (Stevens, 1980)} .$$

حيث N : العدد الكلي للأفراد في المجموعتين $(n_1 + n_2)$ ،

n_1 : عدد أفراد المجموعة الأولى

n_2 : عدد أفراد المجموعة الثانية

$$T^2 = \frac{(N-2)PF}{N-P-1} .$$

وتحسب T^2 من خلال المعادلة الآتية:

حيث P: عدد المتغيرات التابعة، F: قيمة الإحصائي F (هوتلنج).

ويُعد وفق رأي ستيفنز (Stevens, 1980) حجم الأثر للاختبار تحليل التباين المتعدد الأحادي MANOVA One-way صغيراً عندما: $D^2 \leq 0.49$ ، ومتوسطاً عندما $0.5 \leq D^2 \leq 0.95$ ، وكبيراً عندما $D^2 \geq 1$.

هذا ولم يتمكن الباحث من إيجاد الجداول الإحصائية الخاصة بإيجاد قوة الاختبار الخاصة باختبار ولعكس لامدا، كما أن عدد الاختبارات المنشورة في المجلة الأردنية في البحوث التربوية كان عددها محدوداً جداً، لذلك اكتفى الباحث باختبار هوتلنج.

• في حالة استخدام اختبار χ^2 (Chi-Square)

يُستخدم اختبار χ^2 (كـ χ^2) لتوضيح - وليس لإثبات - أهمية الفروق المعنوية بين التكرارات الملاحظة (Observed Frequencies)، والتكرارات المتوقعة (Expected Frequencies) لعينة واحدة، أو لتوضيح أهمية الفروق بين عينتين بالنسبة لظاهرة معينة أو أكثر، وهناك مجموعة من المؤشرات التي تُستخدم لحساب الدلالة العملية، منها مؤشر حجم الأثر في حالة استخدام جدول التوافق

(2×2)، يتم استخدام المعادلة الآتية (Cohen, 1988, 223): $\omega = \sqrt{\frac{\chi^2}{N}}$ حيث ω تعني مؤشر حجم الأثر، وN: حجم العينة، وفي حالة استخدام جدول التوافق لأكثر من (2×2)، يتم

$$\omega = \sqrt{\frac{C^2}{1-C^2}} \text{ : استخدام المعادلة الآتية.}$$

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \text{ ويتم حساب قيمة C من خلال المعادلة الآتية:}$$

حيث C: معامل الاقتان/التوافق (Contingency Coefficient).

ووفق معيار كوهن يُعد حجم الأثر صغيراً عندما تكون قيمة: $0 \leq \omega \leq 0.29$ ، ومتوسطاً عندما

$$0.30 \leq \omega \leq 0.49 \text{، وكبيراً عندما } 0.49 \leq \omega \leq 1.$$

ومن خلال استخدام هذه المؤشرات يمكن للباحث حساب الدلالة العملية باستخدام مؤشر حجم الأثر، وبالتالي التحقق فيما إذا كان للمتغير المستقل تأثيراً على المتغير التابع، وذلك بقياس العلاقة بين المتغيرين حسب نوع ووحدة القياس والاختبار الإحصائي المستخدم. وفي هذا الصدد اعتبر كوهن (Cohen, 1990) أن حجم الأثر من أكثر المفاهيم التي تُساهم في تفسير النتائج، وقد حظي استخدام الدلالة العملية باهتمام الباحثين في تفسير نتائج الدراسات والبحوث الأجنبية المنشورة.

1 - 2 قوة الاختبار الإحصائي

هناك ما يشبه الإجماع بين الباحثين على أهمية المعلومات وجوهريتها الناتجة عن قوة الاختبارات الإحصائية التي يستخدمونها، هذه المعلومات التي يطلب كوهن من الباحثين تحديدها قبل إتمام دراساتهم، وإعطاء معلومات عنها عند نشر تلك الدراسات لاحقاً. وتأتي قيمة المعلومات عن قوة الاختبار الإحصائي من تدليلها على دقة قرار الباحث وصدقه برفض الفرضية الصفرية الخاطئة من جهة، وتدليلها على احتمال الخطأ من النوع الثاني من جهة أخرى. واقترح كوهن بألا تقل قوة الاختبار الإحصائي المستخدم لاختبار الفرضية الصفرية عن (0.80) (Park, 2008).

وتتأثر قوة الاختبار بعوامل عديدة من أهمها:

- حجم الأثر (Effect Size)، فكلما زادت قيمة حجم الأثر زادت قيمة التباين المفسّر، وبالتالي تزداد قوة الاختبار.
- حجم العينة إذ أنه بزيادة حجم العينة يزداد تمثيل العينة للمجتمع، وبالتالي تزداد دقة التصميم، وهذا يؤدي بدوره إلى زيادة قوة الاختبار في الكشف عن الفروق. وللحصول على نتيجة دالة في ظل قوة متدنية لا بد من وجود حجم أثر كبير، أو استخدام حجم عينة كبير.
- مستوى الدلالة الإحصائية المستخدم لاختبار الفرضيات الإحصائية حيث بزيادة قيمة α تزداد قوة الاختبار، وذلك لأن الخطأ من النوع الثاني يقل، وفيما إذا كان الاختبار بذييل واحد أو بذييلين، وشكل التوزيع، والتصميم الإحصائي المستخدم... الخ. (Aron & Aron, 1997).

وهناك أسباب عديدة تجعل الباحثين يهتمون بقوة الاختبار منها:

- يوجه حساب قوة الاختبار الباحثين إلى اتخاذ قرارات بشأن زيادة حجم العينة، ودرجة تمثيل العينة لمجتمع الدراسة.
 - يقدم تفسيراً لرفض الفرضية الصفرية، فعند الفشل في رفض H_0 يكون الهدف التأكد من أن H_0 صحيحة، فإن ذلك يزيد الثقة بصحة الفرضية.
 - إذا كانت نتائج الاختبار ذات دلالة إحصائية، وكان حجم الأثر كبيراً فهناك احتمال كبير في الحصول على نتائج ذات دلالة عملية، أما إذا كان حجم الأثر صغيراً فهذا يعني أنها ذات دلالة إحصائية بسبب عوامل أخرى.
 - إذا كانت نتائج الاختبار غير دالة إحصائياً وقوة الاختبار صغيرة فهذا يعني إما أن فرضية البحث غير صحيحة، أو أن ضعف قوة الاختبار راجع لصغر حجم العينة، أما إذا كانت قوة الاختبار عالية فهذا يعني أن هناك احتمال قوي لأن تكون الفرضية غير صحيحة (Huston, 1993).
- ويمكن حساب قوة الاختبار الإحصائي باستخدام جداول كوهن، وحسب نوع الاختبار الإحصائي المستخدم، فالنتائج غير الدالة إحصائياً إذا تمتعت بقوة اختبار كبيرة وتصميم جيد، فإن نتائجها ستكون مفيدة من الناحية العملية، كونها تشير إلى عدم وجود تأثير عملي كبير على مستوى المجتمع. أما قوة الاختبار المتدنية فإنها تنقص من احتمالية الرفض للفرضية الصفرية، وقد تحول عندئذٍ دون نشر كثير من البحوث غير ذات الدلالة الإحصائية (Park, 2008).

وفي ضوء أهمية كل من الدلالة الإحصائية والدلالة العملية وقوة الاختبار الإحصائي، حظي موضوع الدلالة العملية وقوة الاختبار الإحصائي والدلالة الإحصائية باهتمام الكثير من الباحثين، فقد تناول ثالهيمر وكوك (Thalheimer & Cook, 2002) في مقالة لهما حول كيفية حساب أحجام الأثر للدراسات المنشورة التي تستخدم اختبارات t بمنهجية مبسطة، واقترحا طرقاً جديدة لحساب حجم الأثر d لكوهن، وذلك في حالة عدم تضمين الانحرافات المعيارية أو الأخطاء المعيارية في البحوث المنشورة، وكذلك في حالة وجود الأخطاء المعيارية وعدم الانحرافات المعيارية.

ومن الدراسات التي بحثت في حجم الأثر وقوة الاختبار:

دراسة دانييل (Daniel, 1997) التي أشارت أن متوسط قوة الاختبار عندما يكون حجم الأثر صغيراً ومتوسطاً وكبيراً قد بلغ (0.13)، (0.64)، (0.97) على الترتيب، وكذلك بأن حجم العينة المستخدمة في فحص الفرضيات كان صغيراً، فقد بلغ متوسط حجم العينة للاختبارات المستخدمة (99). كما أشارت أن معظم الدراسات التي تم مراجعتها لم تحدد مسبقاً قيمة مستوى الدلالة α ، كما لم يتم تحديد حجم الأثر لأي من هذه الدراسات، واتفقت هذه النتيجة مع نتيجة دراستين سابقتين.

الأولى: قام بها ولي وداوسون (Wooley & Dawson, 1983) لحساب قوة الاختبارات الإحصائية لـ 3556 اختباراً إحصائياً مستخدماً في مجلة تدريس العلوم، حيث أشارت نتائج التحليل أن قوة الاختبارات الإحصائية عندما كان حجم الأثر صغيراً ومتوسطاً وكبيراً قد بلغت 0.29، 0.63، 0.85 على الترتيب.

والثانية قام بها برور (Brewer, 1972) حول قوة الاختبارات الإحصائية للأبحاث المنشورة في مجلة البحث التربوي وجد أن متوسط القوة هو 0.23، 0.56، 0.79. لحجم الأثر الصغير والمتوسط والكبير على التوالي.

ومن الدراسات العربية حول قوة الاختبار الإحصائي وحجم الأثر وحجم العينة:

الدراسات المنشورة في مجلة أبحاث اليرموك-سلسلة الدراسات الإنسانية والاجتماعية دراسة جرادات وجودة (2005)، والتي خلصت إلى الاستنتاج بأن العامل الرئيس وراء النسبة العالية (72.5%) من الفرضيات الصفرية المرفوضة والنسبة العالية (52%) من الاختبارات الإحصائية التي حققت الحد الأدنى للقوة (0.8) الذي اقترحه كوهن قد يكون الحجم الكبير لعينات الدراسة، وليس حجم الأثر المقترن بالفرضيات الصفرية التي جرى اختبارها، إذ لم تزد نسبة الفرضيات الصفرية التي اختبرت، واقتزنت بحجم أثر كبير عن (26.5%).

كما أشارت دراسة النجار (2005) التي هدفت إلى تقويم استخدام الباحثين للدلالة العملية في اختبار χ^2 لعينة من الأبحاث المنشورة في مجلة جامعة الملك سعود-العلوم الإدارية إلى أن الباحثين لم

يوفقوا في الحصول على نتائج ذي دلالة إحصائية كبيرة لاختبار χ^2 . بالإضافة إلى ذلك أظهرت النتائج علاقة طردية بين قوة الاختبار χ^2 والدلالة العملية، وأن العلاقة بينهما قوية، ويمكن الاعتماد عليها في اتخاذ قرارات.

كما بينت دراسة الدرايع (2003) التي تم فيها إجراء مراجعة للدراسات المنشورة في مجلة جامعة مؤتة للبحوث والدراسات لسلسلة العلوم الإنسانية والاجتماعية للفترة بين 1996-2000 إلى الاستنتاج أن هذه الدراسات تتمتع بقوة اختبار عالية ودلالة عملية مقبولة... ويمكن أن تؤخذ نتائجها على المستوى العملي شريطة أن تكون نتائج هذه الدراسات ذات دلالة إحصائية.

وكذلك بينت دراسة محمود (2003) التي تناولت رسائل الماجستير في كلية العلوم التربوية في جامعة مؤتة أن (45.2%) من الاختبارات الواردة فيها كانت ذات دلالة إحصائية، تمتع (17.8%) بحجم أثر كبير و(11.17%) بحجم أثر متوسط و(71.03%) بحجم أثر صغير، وأن (85.98%) من الاختبارات الدالة إحصائياً كانت ذات قوة إحصائية عالية، بينما (8.82%) ذات قوة إحصائية متوسطة و(5.3%) قوة إحصائية منخفضة.

بينما وجد إبراهيم (2000) أن الأبحاث المنشورة في مجلة دراسات- للفترة 1997-1998 ذات دلالة عملية متوسطة وصغيرة، في حين كانت قوة الاختبارات الإحصائية المستخدمة على الأغلب عالية. كما أشارت دراسة جعفر (1997) حول واقع الدلالة الإحصائية والعملية وقوة الاختبار الإحصائي في رسائل الماجستير في مجال علم النفس التربوي في الجامعة الأردنية أن 30.4% من الفرضيات الدالة إحصائياً كانت ذات دلالة عملية معدومة، 30.9% ذات دلالة عملية صغيرة، و11.8% منها ذات دلالة عملية متوسطة وعالية، وتمتعت 56% من الفرضيات غير الدالة بقوة اختبار صغيرة، مما يدل حسب رأي الباحثة على ضعف التصاميم الإحصائية المستخدمة في الدراسات.

بينما أشارت نتائج دراسة قام بها النجار (1991) حول واقع الدلالة العملية للاختبارات الإحصائية الشائعة في رسائل الماجستير في كليتي التربية بجامعة أم القرى وجامعة الملك سعود إلى ارتفاع نسبة حجم الأثر الصغير لجميع الاختبارات، ما عدا اختبار t في قسم المناهج والتدريس، حيث كانت نسبة حجم الأثر مرتفعة في ذلك القسم.

واتفقت بذلك مع دراسة الحجماات وعليان (1995) التي تناولت الاختبارات الإحصائية الواردة في رسائل الماجستير في مجال الإرشاد والتوجيه في الجامعة الأردنية أن الدلالة العملية (حجم الأثر) لحوالي 76% من الاختبارات الإحصائية كانت صغيرة، وتراوح بين (0-30)، بينما أظهرت النتائج أن (53%) من الاختبارات المستخدمة كانت قوتها الإحصائية عالية.

2 - مشكلة الدراسة

تأتي هذه الدراسة استكمالاً لمجهودات سابقة، حاولت أن تربط بين حجم الأثر كمؤشر للدلالة العملية، وقوة الاختبارات الإحصائية للأبحاث المنشورة في المجلة الأردنية في العلوم التربوية، وهي مجلة فصلية علمية محكمة، أسستها اللجنة العليا للبحث العلمي في وزارة التعليم العالي والبحث العلمي في الأردن، وتصدر عن عمادة البحث العلمي والدراسات العليا في جامعة اليرموك، وكون هذه المجلة قد صدر منها (20) عدداً لغاية إجراء هذه الدراسة في خمسة مجلدات؛ كان لا بُد من إجراء دراسة تقييمية للكشف عن مدى استفادة الباحثين في مجال العلوم الإنسانية والاجتماعية من نتائج الدراسات السابقة والدراسات العالمية في إيجاد مؤشرات للدلالة العملية، وقوة الاختبارات الإحصائية التي يستخدمونها في اختبار فرضياتهم الصفرية بدلاً من الاعتماد فقط على الدلالة الإحصائية لاتخاذ قرارات حول تلك الفرضيات، والتي اعتقد كارفر (Carver, 1993) بأن الاعتماد عليها يُعد مفسدة للطريقة العلمية. ووصفها كيرك (Kirk, 1996) بالتمرينات الساذجة التي تجيب عن أسئلة تختلف عن أسئلة العلم.

وكما أوضح ويلكنسون (Wilkinson, 1992) فقد واجهت الدلالة الإحصائية عدة انتقادات منها: أن جهود الباحثين منصبية بالدرجة الأولى على الحصول على نتائج ذات دلالة إحصائية، وكأن هناك قراراً فاصلاً ما بين الدلالة المعنوية وغير المعنوية، كما أن معظم الدراسات المنشورة ذات الدلالة الإحصائية ذات مستوى قوة متدنٍ، ولم يأخذ الباحثون اعتبار أن بعض الدراسات قد تكون دالة إحصائياً، ولكنها في الوقت نفسه ليست ذات أهمية؛ بمعنى أنها لا تقدم معلومات هامة ومفيدة وذات طابع تطبيقي في المجتمع. ناهية عن أن بعض الباحثين يعطي الدراسات الدالة عند مستوى الدلالة ($\alpha = 0.01$) أهمية أكبر من الدراسات الدالة عند مستوى الدلالة ($\alpha = 0.05$) ويتجاهلون الحديث عن حجم العينة.

حيث يركز الباحثون على تطبيق الاختبارات الإحصائية المتعددة بحثاً عن الدلالة الإحصائية، دون اعتبار لدلالة الفرق أو العلاقة في المجال العملي، ومدى منطقية نتيجة الدلالة الإحصائية كمعيار

لاتخاذ قرار دقيق حيال الظواهر التي يدرسونها. وبسبب خضوع الدلالة الإحصائية لعوامل عديدة مثل: حجم العينة، ومستوى الدلالة الإحصائية، وحجم الأثر، وقوة الاختبار المستخدم، فقد يتوصل الباحث إلى الدلالة الإحصائية عن طريق زيادة حجم العينة أو زيادة مستوى الدلالة (α)، وقد يفشل في رفض الفرضية الصفرية، ليس بسبب عدم وجود أثر للمعالجة، بل لضعف قوة الاختبار الإحصائي في الكشف عن الفروق، ومن النادر جداً أن تجد بحثاً يأخذ بالدلالة العملية وقوة الاختبار الإحصائي المستخدم لاتخاذ قرار حيال الظاهرة المدروسة.

في ضوء ما تقدم جاءت هذه الدراسة للتعرف على الدلالة العملية وقوة الاختبارات الإحصائية المستخدمة في فحص الفرضيات الإحصائية المستخدمة في البحوث المنشورة في المجلة الأردنية في العلوم التربوية، وبالتحديد فإن هذه الدراسة سعت للإجابة عن الأسئلة الآتية:

- السؤال الأول: ما نسبة الفرضيات الإحصائية التي اقترنت فيها الدلالة الإحصائية بدلالة عملية صغيرة ومتوسطة وكبيرة للبحوث المنشورة في المجلة الأردنية في العلوم التربوية؟
- السؤال الثاني: ما نسبة الفرضيات الإحصائية التي اقترنت فيها الدلالة الإحصائية بقوة الاختبار الإحصائي متدنية ومتوسطة وعالية للبحوث المنشورة في المجلة الأردنية في العلوم التربوية؟
- السؤال الثالث: ما واقع الدلالة العملية وقوة الاختبارات الإحصائية المصاحبة للاختبارات الإحصائية المستخدمة في فحص الفرضيات للبحوث المنشورة في المجلة الأردنية في العلوم التربوية؟

3 - منهجية الدراسة

استخدم الباحث أسلوب ما وراء التحليل (Meta- Analysis)، وهو أسلوب إحصائي منظم يتم خلاله جمع نتائج عدد كبير من الدراسات الكمية المختلفة، والتي تتناول مجموعة من الفرضيات البحثية، وتحليلها للحصول على مقياس مشترك مثل حجم الأثر (Effect Size)، لتقويمها والوصول إلى استنتاجات أقوى وذات معنى من تلك التي يمكن الحصول عليها من دراسة منفردة، ويعد أحد الجوانب التطبيقية لتقويم الدراسات والبحوث (Patty, Qi, Frank & Ronald, 2010)

4 - مجتمع الدراسة وعينتها

تكون مجتمع الدراسة وعينتها من جميع البحوث المنشورة في المجلة الأردنية في البحوث التربوية منذ صدورها في العام 2005 ولغاية المجلد الخامس، العدد الرابع، كانون الأول 2009. واشتملت على

خمسة مجلدات تضمنت 20 عدداً، وقد بلغ عددها (121) بحثاً، وقد تم استثناء الأبحاث النوعية، وكذلك بعض أجزاء من الدراسات المحدودة جداً في بعض الدراسات لافتقارها للبيانات الأساسية كحجم العينة، والتباين، وقيمة الاختبار الإحصائي، ومجموع المربعات لمصادر التباين ... الخ، حيث تضمنت الدراسات التي استخدمت تصاميم كمية (1363) اختباراً إحصائياً من نوع (F_1 , t) (تحليل التباين)، r معامل الارتباط البسيط، χ^2 ، R^2 (تحليل الانحدار المتعدد)، Z (الفرق بين معاملي ارتباط)، F_2 (تحليل التباين المتعدد). والجدول (1) يبين أعداد الفرضيات الإحصائية التي تم حساب الدلالة العملية وقوة الاختبار الإحصائي لها، ونسبها المئوية تبعاً لنوع الاختبار المستخدم.

الجدول 1. عدد الفرضيات الإحصائية تبعاً لنوع الاختبار المستخدم

نوع الاختبار	عدد الفرضيات	النسبة المئوية
T	319	23.4
F_1	472	34.6
R	382	28.1
χ^2	104	7.6
R^2	25	1.8
Z	3	0.2
F_2	58	4.3
الكلي	1363	100

5 - إجراءات الدراسة

من أجل الإجابة عن أسئلة الدراسة حول الدلالة الإحصائية والدلالة العملية وقوة الاختبارات الإحصائية للبحوث المنشورة في المجلة الأردنية في العلوم التربوية، تم ذلك وفق الخطوات الآتية:

1-5- حصر جميع الأبحاث المنشورة في المجلة الأردنية في العلوم التربوية منذ إصدارها في عام 2005 ولغاية المجلد الخامس، العدد الرابع، كانون الأول 2009. وتصنيف تلك البحوث إلى كمية وأخرى نوعية.

5-2- تحليل الأبحاث والدراسات الكمية والمستخدمه للاختبارات الإحصائية للإجابة عن أسئلتها أو فرضياتها.

5-3- تصميم أداة لجمع البيانات تحوي المعلومات الأساسية والضرورية واللازمة للإجابة عن أسئلة الدراسة.

5-4- تم حساب حجم الأثر لكل اختبار إحصائي مستخدم.

5-5- تم إيجاد قوة الاختبار الإحصائي لكل اختبار من هذه الاختبارات بالاستعانة بجدول كوهن.

6 - نتائج الدراسة ومناقشتها

6 - 1 للإجابة عن السؤال الأول حول نسبة الفرضيات الإحصائية التي اقترنت فيها الدلالة

الإحصائية بالدلالة العملية للبحوث المنشورة في المجلة الأردنية في العلوم التربوية التي شكلت عينة الدراسة، فقد حسبت أحجام التأثير كمؤشرات على الدلالة العملية للفرضيات المفحوصة باستخدام الاختبارات الإحصائية $(t, F_1, r, \chi^2, Z, R^2, F_2)$ اختبار هوتلنج) وهذه المؤشرات كانت على النحو الآتي: $(d, f, r^2, C, F^2, q, T^2)$ على الترتيب، وتم تصنيفها وفق الفئات التي اقترحها كوهن (Cohen, 1988)، وتم التحقق من نتائج اختبار الفرضيات الدالة منها وغير الدالة، والجدول (2) يبين نتائج التحليل.

جدول 2. التكرارات والنسب المئوية لكل فئة من فئات الدلالة العملية (حجم الأثر) والدلالة

الإحصائية للاختبارات المستخدمة في فحص الفرضيات

الاختبار المستخدم	فئة الدلالة العملية (حجم الأثر)	الدلالة الإحصائية			
		غير دال		دال	
		النسبة المئوية	التكرار	النسبة المئوية	التكرار
t للفرق بين وسطين	صغير 0 - 0.49	84.43%	141	39.47%	60
	متوسط - 0.50 0.79	9.58%	16	21.71%	33
	كبير 1 - 0.8	5.99%	10	38.82%	59
	الكلي	52.35%	167	47.65%	152
F لتحليل التباين	صغير 0 - 0.24	89.7%	220	53.33%	120
	متوسط - 0.25	8.07%	5	20.44%	46

						0.39	
%17.16	81	%8.91	22	%26.23	59	كبير 1 - 0.4	
%100	472	%52.33	247	%47.67	225	الكلي	
%90.58	346	%96.72	118	%87.69	228	صغير 0 - 0.29	r معامل ارتباط البيسط
%6.28	24	%2.46	3	%8.08	21	متوسط - 0.3 0.49	
%3.14	12	%0.82	1	%4.23	11	كبير 1 - 0.5	
%100	382	%31.94	122	%68.06	260	الكلي	χ ²
%87.50	91	%89.61	69	%81.48	22	صغير 0 - 0.29	
%10.58	11	%9.09	7	%14.82	4	متوسط - 0.3 0.49	
%1.92	2	%1.3	1	%3.7	1	كبير 1 - 0.5	تحليل الانحدار المتعدد
%100	104	%74.04	77	%25.96	27	الكلي	
%24	6	%100	1	%20.83	5	صغير 0 - 0.14	
%20	5	%0	0	%20.83	5	متوسط - 0.15 0.34	الفرق بين معاملي ارتباط
%56	14	%0	0	%58.34	14	كبير 1 - 0.35	
%100	25	%4	1	%96	24	الكلي	
%66.67	2	%0	0	%100	2	صغير 0 - 0.29	تحليل التباين المتعدد
%33.33	0	%0	1	%0	0	متوسط - 0.3 0.49	
%0	0	%0	0	%0	0	كبير 1 - 0.5	
%100	3	%33.33	1	%66.67	2	الكلي	
%81.03	47	%100	15	%74.42	32	صغير 0 - 0.49	
%12.07	7	%0	0	%16.28	7	متوسط - 0.5 0.95	
%6.9	4	%0	0	%9.3	4	كبير 1 - 0.96	الكلي
%100	58	%2.36	15	%5.86	43		
%100	1363	%46.22	630	%53.78	733		

تشير البيانات في الجدول (2) إلى أن عدد الفرضيات الإحصائية التي تم اختبارها قد بلغ (1363) فرضيةً إحصائيةً، منها (23.4%) استخدمت اختبار t لدلالة الفرق بين وسطين حسابيين، و(34.6%) استخدمت أسلوب تحليل التباين، و(28%) استخدمت اختبار فرضية تتعلق بمعامل ارتباط واحد، و(7.6%) استخدمت اختبار χ^2 ، و(1.8%) استخدمت اختبار تحليل الانحدار المتعدد (اختبار هوتلنج)، و(0.22%) استخدمت اختبار Z لفحص دلالة الفرق بين معاملي ارتباط مستقلين، و(4.26%) استخدمت أسلوب تحليل التباين المتعدد.

وقد بلغ عدد الفرضيات الدالة إحصائياً (733) فرضية أي ما نسبته (53.78%) من مجموع الفرضيات الصفرية، ولم يتم حساب الدلالة العملية لأي منها، في حين بلغ عدد الفرضيات غير الدالة إحصائياً (630) فرضية، أي ما نسبته (46.22%) من مجموع الفرضيات التي تم اختبارها، وتبين أن الباحثين قد اكتفوا بالنتائج التي تم الحصول عليها من الدلالة الإحصائية للحكم على أثر المتغير أو المتغيرات المستقلة، ولم يتم البحث في قوة هذا الأثر؛ علماً أنه يجب عدم الاكتفاء بالدلالة الإحصائية فقط؛ إذ تعتبر الخطوة الأولى التي لا بُد من اقتراحها ومؤشرات للدلالة العملية، لاسيما وأن ما نسبته (75.79%) من الدراسات كانت ذات دلالة عملية صغيرة، حيث بلغ عددها (1033) فرضية من أصل (1363)، وتعتبر هذه مشكلة جديّة من الناحية التطبيقية، فكيف سيتم الأخذ بتوصيات هذه البحوث من قبل المتخصصين في المجالات التربوية التي تناولتها الأبحاث المنشورة، رغم افتقارها لمعلومات أساسية ومهمة، تساعد في عملية تفسير تلك النتائج. وربما يكون أحد الأسباب في الحصول على نتائج ذات دلالة إحصائية أحجام العينات الكبيرة، واختلفت بذلك مع ما أشارت إليه دراسة فاجلي (Carver, 1993) التي أشارت بأن معظم الدراسات التي يتم نشرها هي التي تكون ذات دلالة إحصائية، في حين اتفقت مع دراسة حجماث وعليان (1997) والصيد (1988)، والتي بيّنت أن نشر الدراسات في المجالات لا يكون متحيزاً تبعاً لدالاتها الإحصائية.

وفيما يتعلق باقتران الدلالة الإحصائية بالدلالة العملية عند استخدام اختبار t ، فقد بينت النتائج أن ما نسبته (39.47%) من الفرضيات الدالة إحصائياً اقتترنت بدلالة عملية صغيرة، وأن (21.71%) اقتترنت بدلالة عملية متوسطة، في حين وجد أن ما نسبته (38.82%) اقتترنت بدلالة عملية كبيرة. أما الفرضيات غير الدالة إحصائياً (من المفروض حساب الدلالة العملية فقط للاختبارات

ذات الدلالة الإحصائية، ولكن لغرض البحث تم حساب الدلالة العملية للبيانات غير الدالة إحصائياً، وذلك لإبراز تأثير حجم العينة على عوامل أخرى مثل: المؤشر الإحصائي)، فإن معظمها قد اقترنت بدلالة عملية صغيرة، وشكلت ما نسبته (84.4%)، وأن ما نسبته (9.5%) قد اقترنت بدلالة عملية متوسطة، في حين أن ما نسبته (6%) اقترنت بدلالة عملية كبيرة. وبشكل عام يلاحظ من الجدول (2) أن الباحثين الذين استخدموا اختبار t لم يوفقوا في الحصول على نتائج ذات دلالة عملية كبيرة، فقد بلغت نسبة الفرضيات ذات الدلالة العملية الكبيرة (21.6%)، في حين بلغت نسبة الفرضيات ذات الدلالة الصغيرة (63%)، أما المتوسطة فقد بلغت نسبتها (15.4%). وبذلك تتفق هذه النتيجة مع دراسة الدرايع (2003) والصيد (1988)، وجعفر (1998) التي أشارت أن اختبار t في البحوث العربية كانت قيمة الدلالة العملية منخفضة. وتختلف مع دراسة النجار (1991) التي أشارت إلى أن حجم الأثر كان مرتفعاً للإحصائي t .

أما فيما يتعلق باقتران الدلالة الإحصائية بالدلالة العملية عند استخدام اختبار F ، فقد أشارت النتائج أن ما نسبته (53.3%) من الفرضيات الدالة إحصائياً اقترنت بدلالة عملية صغيرة، وما نسبته (20.4%) اقترنت بدلالة عملية متوسطة، وما نسبته (26.3%) اقترنت بدلالة عملية كبيرة، وكذلك فإن الفرضيات الإحصائية غير الدالة إحصائياً التي اقترنت بدلالة عملية صغيرة بلغت نسبتها (89%)، وما نسبته (2%) اقترنت بدلالة عملية متوسطة، في حين أن ما نسبته (9%) اقترنت بدلالة عملية كبيرة.

مرة أخرى تبين من الجدول (2) أن الباحثين الذين استخدموا أسلوب تحليل التباين (الاختبار الإحصائي F_1)، لم يوفقوا في الحصول على نتائج ذات دلالة عملية كبيرة، فقد بلغت نسبة الفرضيات ذات الدلالة العملية الصغيرة (72.03%)، ونسبة الفرضيات ذات الدلالة العملية المتوسطة (10.81%)، وذات الدلالة العملية الكبيرة بلغت نسبتها (17.16%). وبذلك تتفق هذه النتيجة مع ما أشارت إليه دراسة إبراهيم (2000)، والدرايع (2003) وحجمات وعلبان (1997) ... واختلفت مع دراسة جعفر (1998) والنجار (1991).

وفيما يتعلق باقتران الدلالة الإحصائية بالدلالة العملية عند فحص الدلالة الإحصائية الخاصة بمعامل ارتباط واحد. فقد بينت النتائج أن ما نسبته (87.7%) من الفرضيات الدالة إحصائياً اقترنت بدلالة عملية صغيرة، وأن (8%) اقترنت بدلالة عملية متوسطة، وما نسبته (4.3%) اقترنت بدلالة

عملية كبيرة. وكذلك فإن الفرضيات غير الدالة إحصائياً، قد اقترنت بما نسبته (96.72%) بدلالة عملية صغيرة، وما نسبته (2.45%) بدلالة عملية متوسطة، وما نسبته (0.83%) بدلالة عملية كبيرة. وأيضاً لم يوفق الباحثون في الحصول على نتائج ذات دلالة عملية كبيرة فقد بلغت نسبة الفرضيات ذات الدلالة العملية الكبيرة فقط (3.14%)، والفرضيات ذات الدلالة العملية المتوسطة (6.28%)، في حين بلغت نسبة الفرضيات ذات الدلالة الصغيرة (90.58%)، وتتفق هذه النتيجة مع العديد من الدراسات مثل دراسة الدرايع (2003) وإبراهيم (2000) ومحمود (2003) وجرادات ووحدة (2005)، وحجمات وعليان (1997) وجعفر (1998)، وكذلك مع دراسة النجار (1991)، الذين أجمعوا على أن الدلالة العملية لدلالة معامل الارتباط البسيط صغيرة، وما ظهور الدلالة الإحصائية إلا نتيجة أحجام العينات الكبيرة، إذ أن دلالة معامل الارتباط تتأثر بدرجة كبيرة بحجم العينة (Carver, 1993)، ولم تختلف نتيجة هذه الدراسة مع أية دراسة أخرى.

أما فيما يتعلق باقتران الدلالة الإحصائية بالدلالة العملية عند استخدام اختبار χ^2 ، فقد وجد بأن ما نسبته (81.48%) من الفرضيات ذات الدلالة الإحصائية قد اقترنت بدلالة عملية صغيرة، في حين أن ما نسبته (14.82%) قد اقترنت بدلالة عملية متوسطة، وما نسبته (3.7%) اقترنت بدلالة عملية كبيرة. وأما الفرضيات التي ليس لها دلالة فقد اقترن ما نسبته (89.61%) منها بدلالة عملية صغيرة، وما نسبته (9.09%) بدلالة عملية متوسطة وحوالي (1.3%) بدلالة عملية كبيرة. وبشكل عام فقد كانت نسبة الفرضيات التي استخدمت اختبار χ^2 التي اقترنت بدلالة عملية صغيرة (87.5%)، وما نسبته (10.58%) بدلالة عملية متوسطة، وحوالي (1.92%) اقترنت بدلالة عملية كبيرة، وهذا يشير بطبيعة الحال إلى أن واقع الدلالة العملية المصاحب للدلالة الإحصائية لاختبار χ^2 في عينة البحث الحالية كانت عند حجم الأثر المنخفض، وهذه النتيجة تتفق مع ما توصلت إليه دراسة النجار (2005) والصيد (1988) ومحمود (2003) من أن الدلالة العملية لاختبار χ^2 في البحوث العربية كانت منخفضة.

كما يتضح من الجدول أن عدد الفرضيات التي استخدمت أسلوب تحليل الانحدار المتعدد للكشف عن دلالة معامل الارتباط المتعدد (R^2) بلغ (25) فرضية، كانت (24) منها ذات دلالة إحصائية وفرضية واحدة ليست ذات دلالة إحصائية، وصُنفت بأنها ذات دلالة عملية صغيرة، وقد بلغت نسبة الفرضيات التي اقترنت بدلالة عملية صغيرة (20.83%)، وما نسبته (20.83%) اقترنت بدلالة

عملية متوسطة، وحوالي (58.34%) اقترنت بدلالة عملية كبيرة، وتُعد هذه النسبة أفضل من النسب السابقة، رغم أنها ما زالت دون المستوى حتى تكون إمكانية الأخذ بتوصيات هذه الدراسات مقبولة. وربما يعود السبب في ذلك أنه في الواقع لا تعمل المتغيرات بشكل منفرد، فمعامل الارتباط المتعدد يُعبر عن قوة العلاقة بين متغير تابع واحد من جهة ومجموعة من المتغيرات الأخرى -المستقلة إجرائياً- عندما تؤخذ مجتمعة من جهة أخرى. وقيمتها تزداد عندما يكون ارتباط المتغير التابع بالمتغيرات المستقلة أكبر ما يمكن، وعندما ترتبط المتغيرات المستقلة ببعضها بعضاً أدنى ارتباطات، واختلقت هذه النتيجة مع دراسة محمود (2003)، واتفقت مع دراسة ستفنز (Stevens, 1980).

وفيما يتعلق باقتران الدلالة الإحصائية بالدلالة العملية عند استخدام اختبار Z للفرق بين معاملي ارتباط مستقلين، فقد بلغ عدد الفرضيات التي فحصت دلالة الفرق ثلاث فرضيات، منها فرضيتان كانتا ذات دلالة إحصائية، وفرضية واحدة ليست ذات دلالة إحصائية بدلالة عملية صغيرة، والأخرى اقترنت بدلالة عملية متوسطة، وبشكل عام اقترنت (66.7%) من الفرضيات بدلالة عملية صغيرة، وما نسبته (33.3%) بدلالة عملية متوسطة، ولم تقترن أية فرضية بدلالة عملية كبيرة، ولا يمكن تعميم هذه النتيجة لأن عدد الفرضيات المستخدم قليل ولا يكفي لأغراض التعميم.

أما اقتران الدلالة الإحصائية بالدلالة العملية عند استخدام أسلوب تحليل التباين المتعدد باستخدام اختبار هوتلنج الخاص بالمتغيرات المستقلة التي لها مستويان، أما اختبار ولكس لأمدا فلم يتمكن الباحث من إيجاد الجداول الخاصة بمؤشرات قوة الاختبار أو الدلالة العملية. فقد أشارت البيانات في جدول (2) أن عدد الفرضيات الإحصائية التي تم اختبارها باستخدام اختبار هوتلنج قد بلغ (58) فرضية، بلغ عدد الدال منها إحصائياً (43) أي ما نسبته (74%)، وعدد غير الدال منها إحصائياً (15) أي ما نسبته (26%)، وكانت نسبة الفرضيات الدالة التي اقترنت بدلالة صغيرة (74.42%) ، وما نسبته (16.28%) من الفرضيات اقترنت بدلالة عملية متوسطة، في حين اقترنت (9.3%) بدلالة عملية كبيرة. أما نسبة الفرضيات غير الدالة التي اقترنت بدلالة عملية صغيرة فكانت (100%)، ولم تقترن أي من الفرضيات غير الدالة بدلالة عملية متوسطة أو كبيرة. وبشكل عام فقد اقترنت ما نسبته (81.03%) بدلالة صغيرة، و(12.07%) بدلالة متوسطة، و(6.9%) بدلالة كبيرة، وتتفق هذه النتيجة مع ما أشارت إليه دراسة محمود (2003) والتي توصلت لنتيجة مشابهة.

مما سبق تبين من النتائج أن الاختبارات المستخدمة (t) للفرق بين وسطين، F لتحليل التباين، ودلالة معامل الارتباط البسيط، و χ^2 ، وتحليل التباين المتعدد)، كانت جميعها قد اقترنت بدلالة عملية صغيرة وبنسبة عالية؛ وهذا يعني أن الدلالة الإحصائية ليست كافية للحكم على قيمة النتائج، واتخاذ القرارات العملية في ضوءها. وبالتالي فإن نسبة ما تفسره المتغيرات المستقلة من تباينات المتغيرات التابعة قليلة؛ وهذا يعني أن هناك عوامل أخرى لم يتم دراستها، وتؤثر في تفسير المتغيرات التابعة ولكنها غير معروفة، لذا فمن الضروري أن يتم تقدير حجم الأثر المتوقع للدراسة قبل البدء بجمع المعلومات والوصول إلى النتائج، مما يساعد الباحث في تقدير حجم العينة وقوة الاختبار حتى تكون النتائج مفيدة من الناحية العملية، وهذا ما أشار إليه (Daniel, 1993). كما يمكن عزو عدم الحصول على دلالة عملية كبيرة بالمستوى المطلوب إلى عدم دقة أدوات القياس المستخدمة، مما يدل على ضعف التصميم الإحصائي المستخدم في تلك الدراسات، وعدم تحديد التعريف الإجرائي المناسب للمفهوم، حتى يكون هناك اتساق بين التعريف والأداة المستخدمة. وهذه النتيجة كانت قد اتفقت مع ما أشارت إليه دراسة الدرايع (2003)، ودراسة محمود (2003)، ودراسة إبراهيم (2000)، ودراسة حجمات وعليان (1997).

كما تم استخدام اختبار مربع كاي χ^2 لفحص مدى استقلالية الدلالة الإحصائية عن الدلالة العملية لكل اختبار من الاختبارات المستخدمة وهي: (t، F_1 ، r ، χ^2 ، R^2 ، Z، F_2)، وقد بلغت قيم مربع كاي للاختبارات على التوالي على النحو الآتي: (46.3)، (19.8)، (16.7)، (16.63)، (21.2)، (15.81)، (14.69) وجميعها كانت ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = .05$)؛ أي أن الدلالة الإحصائية غير مستقلة عن الدلالة العملية. ولدى حساب قيم معامل التوافق (Contingency Coefficient) للاختبارات الإحصائية المستخدمة جميعها، فقد بلغت على الترتيب على النحو الآتي: (.356)، (.20)، (.20)، (.37)، (.46)، (.92)، (.45). على الترتيب. وجميعها أيضاً كانت ذات دلالة إحصائية، وهذا يؤكد بأن الدلالة الإحصائية ليست مستقلة عن الدلالة العملية، أي توجد علاقة بين الدالتين.

6 - 2 للإجابة عن السؤال الثاني المتعلق بنسبة الفرضيات الإحصائية التي اقترنت فيها الدلالة

الإحصائية بقوة الاختبار الإحصائي (متدنية ومتوسطة وعالية) للاختبارات الإحصائية المستخدمة في فحص الفرضيات الإحصائية في المجلة الأردنية في العلوم التربوية التي شكلت عينة الدراسة.

فقد تم إيجاد قوة الاختبارات الإحصائية المستخدمة وفق جداول كوهن، وتم تصنيفها تبعاً لتصنيف كوهن (Cohen, 1988)، وفقاً لنوع الاختبار المستخدم والدلالة الإحصائية، والجدول (3) يبين نتائج التحليل.

جدول 3: التكرارات والنسب المئوية لكل فئة من فئات قوة الاختبار والدلالة الإحصائية للاختبارات المستخدمة في فحص الفرضيات

الكلي		الدلالة الإحصائية				فئة قوة الاختبار	الاختبار المستخدم
		غير دال		دال			
النسبة المئوية	التكرار	النسبة المئوية	التكرار	النسبة المئوية	التكرار		
%50.47	161	%73.05	122	%25.66	39	متدنية - 0.39	t للفرق بين وسطين
%10.66	34	%8.38	14	%13.16	20	متوسطة - 0.4 0.59	
%38.87	124	%18.56	31	%61.18	93	مرتفعة - 0.6 1	
%100	319	%52.35	167	%47.65	152	الكلي	
%50.42	238	%68.02	168	%31.11	70	متدنية - 0.39	F لتحليل التباين
%11.23	53	%13.36	33	%8.89	20	متوسطة - 0.4 0.59	
%38.35	181	%18.62	46	%60	135	مرتفعة 1 - 0.6	
%100	472	%52.33	247	47.67	225	الكلي	
%25.39	97	%63.12	77	%7.69	20	متدنية - 0.39	r معامل ارتباط البسيط
%14.4	55	%21.31	26	%11.15	29	متوسطة - 0.4 0.59	
60.21	230	%15.57	19	%81.16	211	مرتفعة 1 - 0.6	
%100	382	%31.94	122	%68.06	260	الكلي	
%59.6	62	%75.3	58	14.82	4	صغير 0.39 - 0	

متوسط -0.4 0.59	3	%11.11	3	%3.9	6	%5.77	
كبير 1 - 0.6	20	%74.07	16	%20.8	36	%34.61	
الكلي	27	%25.96	77	%12.42	104	%100	
متدنية - 0 0.39	1	%4.17	0	%0	1	%4	تحليل الانحدار المتعدد
متوسطة - 0.4 0.59	5	%20.83	1	%100	6	%24	
مرتفعة 1 - 0.6	18	%75	0	%0	18	%72	
الكلي	24	%96	1	%4	25	%100	
متدنية - 0 0.39	0	%0	0	%100	0	%0	Z الفشرية للفرق بين معاملتي ارتباط
متوسطة - 0.4 0.59	0	%0	0	%0	0	%0	
مرتفعة 1 - 0.6	2	%100	1	%100	3	%100	
الكلي	2	66.67	1	%33.33	3	%100	
متدنية - 0 0.39	8	%18.6	12	%80	20	%34.48	تحليل التباين المتعدد
متوسطة - 0.4 0.59	3	%6.98	2	%13.33	5	%8.62	
مرتفعة 1 - 0.6	32	%74.42	1	%6.67	33	%56.9	
الكلي	43	%74.14	15	%25.86	58	%100	
	733	%53.78	630	%46.22	1363	%100	الكلي

تبيّن من الجدول (3) أن الفرضيات الدالة إحصائياً التي تم فحصها باستخدام اختبار t، واقتزنت بمستوى قوة متدنية بما نسبته (25.66%)، وبمستوى قوة متوسطة بما نسبته (13.16%)، وبمستوى قوة عالية بما نسبته (61.18%). وأما الفرضيات غير الدالة إحصائياً فقد اقترن ما نسبته (73.05%) بمستوى قوة متدنية، وما نسبته (8.38%) بمستوى قوة متوسط، وما نسبته (18.58%) بمستوى قوة

عالية. وبشكل عام فإن نسبة الفرضيات التي استخدم اختبار t في فحصها، والتي اقترنت بمستوى قوة صغيرة بلغت (50.47%)، وبمستوى قوة متوسطة (10.66%)، وبمستوى قوة عالية (38.87%).

وعند استخدام اختبار F_1 المستخدم في أسلوب تحليل التباين، فقد بلغت نسبة الفرضيات الدالة إحصائياً التي اقترنت فيها الدلالة الإحصائية بقوة اختبار صغيرة (31.11%)، وبقوة اختبار متوسطة (8.89%)، وبقوة اختبار عالية (60%). أما نسب الفرضيات غير الدالة إحصائياً، والتي اقترنت فيها الدلالة الإحصائية بقوة اختبار متدنية، فقد بلغت (68.02%)، وبقوة اختبار متوسطة (13.36%)، وبقوة اختبار عالية (18.62%). وبشكل عام فقد بلغت نسبة الفرضيات التي استخدمت اختبار F ، و اقترنت الدلالة الإحصائية فيها بمستوى قوة اختبار صغيرة (50.42%)، وبمستوى قوة متوسطة (11.23%)، وبمستوى قوة عالية (38.35%).

وعند فحص دلالة معامل الارتباط البسيط (r)، وفيما يتعلق بالفرضيات الدالة إحصائياً التي وجد أنها دالة إحصائياً، فقد وجد أن نسبة الفرضيات التي اقترنت بمستوى قوة متدنية (7.69%)، وبمستوى قوة متوسطة (11.15%)، وبمستوى قوة عالية (81.16%). أما فيما يتعلق بالفرضيات غير الدالة إحصائياً، فقد وجد أن ما نسبته (63.12%) قد اقترنت بمستوى قوة صغيرة، و (21.31%) بمستوى قوة متوسطة، وما نسبته (15.57%) بمستوى قوة عالية. وبشكل عام فقد بلغت نسبة الفرضيات التي فحصت دلالة معامل الارتباط البسيط و اقترنت بمستوى قوة متدنية (25.39%)، وبمستوى قوة متوسطة (14.4%)، وبمستوى قوة عالية (60.21%).

وفيما يتعلق باقتران الدلالة الإحصائية بقوة الاختبار لاختبار χ^2 المستخدم، فقد وجد أن ما نسبته (14.82%) من الفرضيات الدالة إحصائياً قد اقترنت بمستوى قوة متدنية، في حين أن ما نسبته (11.11%) قد اقترنت بمستوى قوة متوسطة، و (74.07%) اقترنت بمستوى قوة عالية، أما الفرضيات غير الدالة إحصائياً فقد اقترنت (75.3%) منها بمستوى قوة متدنية، و (3.9%) بمستوى قوة متوسطة، و (20.8%) بمستوى قوة عالٍ. أما نسبة إجمالي الفرضيات التي استخدمت χ^2 التي اقترنت بمستوى قوة متدنية فكانت (59.6%) و (5.77%) اقترنت بمستوى قوة متوسطة، في حين أن (34.61%) اقترنت بمستوى قوة عالية.

وعند استخدام أسلوب تحليل الأعداد في فحص الفرضيات المستخدمة في المجلة الأردنية في العلوم التربوية، فقد بلغت عدد الفرضيات التي استخدمت هذا الأسلوب (25) أي ما نسبته (1.83%) من إجمالي الفرضيات منها (24) فرضية ذات دلالة إحصائية وفرضية واحدة ليست ذات دلالة إحصائية. بلغت نسبة الفرضيات الإجمالية التي اقترنت بمستوى قوة متدنية (4%)، أما نسبة الفرضيات التي اقترنت بمستوى قوة متوسطة (24%)، في حين أن ما نسبة (72%) منها اقترنت بمستوى قوة عالٍ.

وعند استخدام اختبار Z للعلامات الفشرية لفحص دلالة الفرق بين معاملي ارتباط مستقلين، فقد بلغ عدد الفرضيات التي استخدمت هذا الأسلوب ثلاث فرضيات منها فرضيتان كانتا ذوات دلالة إحصائية اقترنت بمستوى قوة عالية، ولم تقترن أية فرضية بمستوى قوة صغير أو متوسط، في حين أن الفرضية غير الدالة إحصائياً، قد اقترنت أيضاً بمستوى قوة اختبار عالٍ.

أما عند استخدام أسلوب تحليل التباين المتعدد وتحديداً اختبار هوتلنج، فقد بلغ عدد الفرضيات التي استخدمت هذا الأسلوب (58) فرضية، منها (43) ذات دلالة إحصائية، و(15) فرضية لم تكن ذات دلالة إحصائية، بلغت نسبة الفرضيات ذات الدلالة الإحصائية التي اقترنت بمستوى قوة متدنية (18.6%) وبمستوى قوة متوسطة (6.98%) وبمستوى قوة كبير (74.4%). في حين أن الفرضيات غير الدالة إحصائياً التي اقترنت بمستوى قوة متدنية بلغت نسبتها (80%) وبمستوى قوة متوسطة (13.33%) وبمستوى قوة عالية (2.42%).

وبشكل عام فإن نسبة الفرضيات الدالة وغير الدالة التي اقترنت بمستوى قوة اختبار متدنية بلغت (44.63%)، أما التي اقترنت بمستوى قوة اختبار متوسطة (10.23%)، كما تم استخدام اختبار χ^2 لفحص استقلالية الدلالة الإحصائية عن قوة الاختبار الإحصائي لكل اختبار من الاختبارات المستخدمة وهي ($t, F_1, r, \chi^2, Z, R^2, F_2$)، وقد بلغت قيم χ^2 للاختبارات على النحو الآتي: (26.7)، (16.3)، (34.6)، (21.2)، (15.19)، (12.54)، (14.73) على التوالي، وجميعها كانت ذات دلالة إحصائية عند مستوى الدلالة $\alpha = .05$ ؛ أي أن الدلالة الإحصائية غير مستقلة عن قوة الاختبار، كما تم حساب معامل التوافق (Contingency Coefficient) للاختبارات الإحصائية المستخدمة، وقد بلغت قيمها (0.28)، (0.18)، (0.29)، (0.41)، (0.61)، (0.90)، (0.45) على الترتيب، وجميعها كانت ذات دلالة إحصائية ($\alpha = .05$)، وهذا

يؤكد مرة أخرى بأن الدلالة الإحصائية ليست مستقلة عن قوة الاختبار الإحصائي أيضاً، أي توجد علاقة بينهما وذات دلالة إحصائية.

وهكذا فقد شككت الفرضيات التي اقترنت بمستوى قوة اختبار متدنية أو متوسطة ما نسبته (54.86%) من مجموع الفرضيات التي تم اختبارها. فهذا يعني أن نسبة كبيرة من الفرضيات قد اقترنت بمستوى قوة متدنية، وهذا يشكك في صحة القرارات المتخذة حول الفرضيات الدالة إحصائياً من حيث التفسير المقدم لرفض الفرضية الصفرية. أما الفرضيات غير الدالة إحصائياً، فإن ذلك يعني أن هناك فروقاً معدومة أو أن هناك فروقاً، ولكن الاختبار الإحصائي المستخدم عجز عن كشفها، وهذا يفسر ما تم الحصول عليه من دلالة عملية صغيرة لهذه الاختبارات. واتفقت هذه النتيجة جزئياً مع دراسة الدرايبع (2003)، ودراسة محمود (2000)، ودراسة إبراهيم (2000). أما فيما يتعلق بأسلوب تحليل الانحدار المتعدد، وكذلك اختبار فحص دلالة معامل ارتباط التي أشارت إلى ارتفاع في مستوى قوة هذه الاختبارات للفرضيات الدالة وغير الدالة إحصائياً، فرمما يُعزى ذلك إلى أن النتائج كانت أكثر دقة، والقرارات المتخذة والتفسيرات المقدمة تكون أكثر إقناعاً سواء في رفض الفرضية الصفرية أو عدم رفضها. كما يمكن عزو النتيجة إلى أن التصميم الإحصائي المستخدم كان قوياً، أو أن الأدوات المستخدمة في عملية القياس أكثر دقة ومصداقية، وهذا ما فسره ارتفاع قيم الدلالة العملية التي اقترنت مع الدلالة الإحصائية، وربما يعود السبب في ذلك انخفاض نسبة الفرضيات التي استخدمت هذا الاختبار الإحصائي، وبالتالي لا يمكن تعميم تلك النتائج. كما يمكن أن تُعزى النتيجة الخاصة بمعامل الارتباط إلى تأثر الدلالة وقوة الاختبار بحجم العينة المستخدم في الدراسات.

6 - 3 للإجابة على السؤال الثالث المتعلق باقتران الدلالة العملية بقوة الاختبار الإحصائي

المستخدم للاختبارات الإحصائية المستخدمة في فحص الفرضيات الصفرية المنشورة في المجلة الأردنية في العلوم التربوية، فالجدول (4) يربط بين قوة الاختبار الإحصائي والدلالة العملية ممثلة بحجم الأثر الذي تم حسابه لكل اختبار من الاختبارات المستخدمة في تلك الدراسات.

جدول 4: التكرارات والنسب المئوية للاختبارات الإحصائية المستخدمة مصنفة وفق فئة كل من الدلالة العملية وقوة الاختبار

الكلي		فئة الدلالة العملية						فئة قوة الاختبار	الاختبار المستخدم
		كبير		متوسط		صغير			
النسبة المئوية	التكرار	النسبة المئوية	التكرار	النسبة المئوية	التكرار	النسبة المئوية	التكرار		
%50.47	161	%11.59	8	%16.33	8	%72.14	145	متدنية	t للفرق بين وسطين
%10.66	34	%5.79	4	%20.4	10	%9.95	20	متوسطة	
%38.87	124	%82.62	57	%63.27	31	%17.91	36	مرتفعة	
%100	313	%21.63	69	%15.36	49	%63.01	201	الكلية	
%50.42	238	%16.06	13	%13.72	7	%64.12	218	متدنية	F لتحليل التباين
%11.23	53	%8.64	7	%25.49	13	%9.7	33	متوسطة	
%38.35	181	%75.3	61	%60.78	31	%26.18	89	مرتفعة	
%100	472	%17.16	81	%10.81	51	%72.03	340	الكلية	
%25.39	97	%0	0	%8.33	2	%27.46	95	متدنية	r معامل ارتباط بيرسون
%14.39	55	%0	0	%0	0	%15.9	55	متوسطة	
%60.20	230	%100	12	%91.67	22	%56.65	196	مرتفعة	
%100	382	%3.14	12	%6.28	24	%90.58	346	الكلية	
%59.62	62	%0	0	%0	0	%68.13	62	متدنية	χ^2
%5.77	6	%0	0	%18.18	2	%4.4	4	متوسطة	
%34.62	36	%100	2	%81.82	9	%27.47	25	مرتفعة	
%100	104	%1.92	2	%10.58	11	%87.50	91	الكلية	
%4	1	%0	0	%0	0	%16.67	1	متدنية	تحليل الانحدار المتعدد
%24	6	%14.29	2	%40	2	%33.33	2	متوسطة	
%72	18	%85.71	12	%60	3	%50	3	مرتفعة	
%100	25	%56	14	%20	5	%24	6	الكلية	
%0	0	%0	0	%0	0	%0	0	متدنية	Z الفشرية للفرق بين
%0	0	%0	0	%0	0	%0	0	متوسطة	
%100	3	%0	0	%100	1	%100	2	مرتفعة	

معامل ارتباط	الكلية	2	%66.67	1	%33.33	0	%0	3	%100
تحليل التباين المتعدد	متدنية	14	%29.79	3	%42.86	3	%75	20	%34.48
	متوسطة	3	%6.38	1	%14.28	1	%25	5	%8.62
	مرتفعة	30	%63.83	3	%42.86	0	%0	33	%56.9
	الكلية	47	%81.03	7	%12.07	4	%6.9	58	%100
الكلية	1033	%75.79	148	%10.86	182	%13.35	1363	%100	

يشير الجدول 4 الذي يربط بين قوة الاختبار الإحصائي والدلالة العملية (حجم الأثر)، أن نسبة الفرضيات التي اقترنت بدلالة عملية صغيرة (%75.79)، والتي اقترنت بدلالة عملية متوسطة (%10.86)، أما الفرضيات التي اقترنت بدلالة عملية قوية قد بلغت (%13.35). وإذا ما أضفنا الفرضيات التي اقترنت بحجم أثر صغير ومتوسط وفق معيار كوهن (Cohen, 1988) معاً، تصبح نسبة الفرضيات (%86.65) من مجموع الفرضيات. وقد بلغت نسبة الفرضيات التي اقترنت منها بقوة اختبار مرتفعة (%42.27)، في حين أن نسبة الفرضيات التي اقترنت بقوة اختبار مرتفعة من الفرضيات التي اقترنت بحجم أثر عالٍ بلغت (%79.12)، وهذا يدل على أهمية حجم الأثر في تحديد قوة الاختبار الإحصائي. كما بيّنت نتائج الدراسة أن نسبة الفرضيات الدالة إحصائياً من بين الفرضيات التي تم اختبارها، والمنشورة في المجلة الأردنية في العلوم التربوية (%52.58) شكلت نسبة الفرضيات التي اقترنت بها بحجم أثر صغير أو متوسط ما نسبته (%79.79)، لذا فإن الدلالة الإحصائية لتلك الفرضيات على ما يبدو ليست راجعة لكبير حجم الأثر. وكذلك فإن نسبة الفرضيات التي تم اختبارها والتي اقترنت بحجم أثر كبير، وكذلك بقوة عالية هي نسبة متدنية بلغت (%10.56)، وهذا يتفق مع دراسات الدرايب (2003)، وجرادات وحوودة (2005)، وإبراهيم (2000) ومحمود (2003)، التي أشارت بأن غالبية الدراسات ذات القوة العالية قد اقترنت بحجم أثر متوسط أو صغير.

7- الخلاصة

لقد تبين أنه حتى تكون للنتائج قابلية للتعميم وموثوق بمصادقتها، لا بد للباحثين من الاهتمام بالدلالة العملية وقوة الاختبار الإحصائي المستخدم إضافة إلى الدلالة الإحصائية، لأن الدراسات التي تتمتع بقوة اختبار إحصائي مرتفعة، وكذلك دلالة عملية يمكن أن يساعد في تفسير نتائج الدراسات

التجريبية التي تبحث في أثر المتغيرات التابعة، وبالتالي يمكن أن يتم تعميم نتائجها، بحيث تكون قابلة للتطبيق. لذا ينبغي على الباحثين حساب الدلالة العملية لنتائج جهوداتهم البحثية ذات الدلالة الإحصائية، لإبراز مدى الاتفاق بين الدالتين الإحصائية والعملية من عدمه، إلا أن ذلك بطبيعة الحال غير كافٍ، إذ لا بُد من إعطاء معلومات شاملة ودقيقة ووافية عن أهمية المشكلات المدروسة وعن تصميماتها البحثية، وتقديم معلومات ومؤشرات قوية عن الخصائص السيكومترية للمقاييس المستخدمة، ومدى مناسبة التحليلات الإحصائية المستخدمة لاختبار الفرضيات الإحصائية، مما يترتب عليه نتائج غير دقيقة، وبالتالي لا يمكن الاستفادة من نتائج تلك البحوث في الواقع لأن نتائجها مشكوك بمصداقيتها. ولكن ذلك يحتاج بطبيعة الحال إلى جهودات بحثية مستقبلية من قبل باحثين آخرين، وكذلك يمكن لرؤساء تحرير المجالات البحثية الاستفادة من نتائج هذه الدراسة في طلب معلومات أكثر من الباحثين تؤكد دقة وصدق نتائج جهوداتهم البحثية كمؤشرات الدلالة العملية، وقوة الاختبارات الإحصائية، والتحقق من مناسبة الاختبارات الإحصائية المستخدمة في فحص تلك الفرضيات.

المراجع

المراجع العربية:

- إبراهيم، رشاد إبراهيم. (2000). واقع الدلالة الإحصائية والدلالة العملية وقوة الاختبارات الإحصائية المستخدمة في بحوث مجلة دراسات العلوم التربوية في الأعوام (1987-1998) في الجامعة الأردنية، رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.
- جرادات، ضرار وجودة، ماجد. (2005). قوة الاختبار الإحصائي وحجم الأثر وحجم العينة للدراسات المنشورة في مجلة أبحاث اليرموك - سلسلة العلوم الإنسانية والاجتماعية، المجلة الأردنية في العلوم التربوية، 1(1)، 21-29.
- جعفر، الطاف رمضان. (1998). واقع الدلالة الإحصائية والدلالة العملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير علم النفس التربوي في الجامعة الأردنية، رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.
- حجمات، تحسين وعليان، خليل. (1997). واقع الدلالات الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير الإرشاد النفسي والتربوي بالجامعة الأردنية، مجلة دراسات العلوم التربوية، 24 (4)، 398-408.
- الدرايع، ماهر. (2003). واقع الدلالات الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة في بحوث مجلة مؤتة للبحوث والدراسات وسلسلة العلوم الإنسانية والاجتماعية. مجلة مؤتة للبحوث والدراسات، 18، 151-176.
- الصياد، عبد المعطي أحمد. (1988). الدلالة العملية وحجم العينة المصاحبتين للدلالة الإحصائية لاختبار (ت) في البحث التربوي والنفسي العربي - دراسة تقويمية، بحوث مؤتمر البحث التربوي الواقع والمستقبل، المجلد الثاني، رابطة التربية الحديثة، المركز العربي للبحوث التربوية، القاهرة، 2-4 يوليو 1988 ، 199-230.

- ما وراء التحليل للأبحاث المنشورة في المجلة الأردنية.....د.الشريفين
- الطاف، رمضان. (1998). واقع الدلالة الإحصائية، الدلالة العملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير علم النفس التربوي في الجامعة الأردنية، رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.
 - علام، صلاح الدين. (1990). الأساليب الإحصائية الاستدلالية البارامترية واللابارامترية في تحليل البحوث النفسية والتربوية (الطبعة الأولى). القاهرة: دار الفكر العربي.
 - عودة، أحمد والخليلي، خليل. (1998). الإحصاء للباحث في التربية والعلوم الإنسانية (الطبعة الأولى)، عمان: دار الفكر للنشر والتوزيع.
 - الكيلاني، عبد الله والشريفين، نضال. (2011). مدخل إلى البحث في العلوم التربوية والاجتماعية: أساسياته، مناهجه، تصاميمه، أساليبه الإحصائية (الطبعة الثالثة). عمان: دار المسيرة للنشر والتوزيع.
 - محمود، فريال محمود. (2003). واقع الدلالة الإحصائية والدلالة العملية وقوة الاختبار للاختبارات الإحصائية المستخدمة في رسائل الماجستير الصادرة عن كلية العلوم التربوية في جامعة مؤتة، رسالة ماجستير غير منشورة، مؤتة، الكرك، الأردن.
 - النجار، عبد الله عمر. (1991). دراسة تقويمية مقارنة للأساليب الإحصائية التي استخدمت في تحليل البيانات في رسائل الماجستير في كل من كلية التربية بجامعة أم القرى بمكة المكرمة وكلية التربية بجامعة الملك سعود بالرياض، رسالة ماجستير غير منشورة، جامعة أم القرى، مكة المكرمة.
 - النجار، عبد الله. (2005). الدلالة الإحصائية والدلالة العملية لاختبار كا2 في البحوث الإدارية المنشورة "دراسة تقويمية". المجلة العربية للعلوم الإدارية، 12(5)، 169-193.

المراجع الأجنبية

- Aron, A. , & Aron, E. (1997). **Statistics for the behavioral and social sciences: A brief course**. New Jersey: Prentice Hall.
- Brewer, J. k, (1972). On the power of statistical tests in the American education research journal. **American Education Research Journal**, 9(39), 391-401
- Carver, P. R. (1993). The case against statistical significance testing, revisited. **The Journal of Experimental Education**,61,287-292.
- Cohen, J. (1988). **Statistical power analysis for the behavioral sciences** (1st edition). New York: Academic Press.
- Cohen, J. (1990). Things I have Learned (So far). **American Psychologist**, 45, 1304-1312.
- Daniel, L. (1997). **Statistical significance testing in educational and psychological measurement and other journals**. Paper Presented at the Annual Meeting of the National Council on Measurement in Education. Chicago: IL. , 22 September, 1997
- Daniel, T. D. (1993). **Statistical power analysis of the quntitive techniques used in the Journal of research in music education**. 1987 through 1991, Doctoral Dissertation, Auburn University.
- Fagley, N. S., & I. J. McKinney. (1983). Reviewer bias for statistically significant results. A reexamination. **Journal of Counseling Psychology**, 30, 298-300.

- Fan, X. (1999). **Statistical significance and effect size: two sides of a coin**. Paper Presented at the Annual Meeting of the American Evaluation Association. Orlando: FL. ,12-may, 1999.
- Hasse, T. (2001). Statistical Significance should not be considered one of life's guarantees: effect size are needed. **Educational and Psychological Measurement**,61, 219-224.
- Huston, L. (1993). **Meaningfulness, statistical significance, effect size, and power analysis: a general discussion with implications for MANOVA**. Paper presented at the annual Meeting of the Mid-South Educational Research Association, New Orleans, LA (ERIC Document, No. 364608).
- Kirk, R.E. (1996). Practical significance: a concept whose time has com. **Educational and Psychological Measurement**,56, 746-759.
- Konstantopoulos, S. (2008). Computing power of tests on the variance treatment effects in designs with two levels of nesting. **Multivariate Behavioral Research**,43, 327-352.
- Levin, J. (1998). What if there were no more bickering about statistical significance tests?. **Research in the Schools**, 5(2), 43-53.
- Mahadevan, L. (2000). **The effect size statistic: overview of various choices**. Paper presented at the annual meating of southwest educational Research Association Dallas: TX. ,5 june,2000.

- Mclean, J. & Ernest, J. (1998). The role of statistical significance testing in educational research. **Research in Schools**, 5(2), 15-22.
- Nix, T. & Barnette, J. (1998). A review of hypothesis testing revisited: rejoinder to Thompson, Knapp, and Levin. **Research in the Schools**, 5(2), 122-126.
- Park, H. M. (2008). **Hypothesis testing and statistical power of a test**. Working Paper. The Diversity Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University. Retrieved, 2008, from <http://www.indiana.edu/statmath>.
- Patty, W., Qi, S., Frank, B. & Ronald M. (2010). Meta- analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease. **American Journal of Clinical Nutrition**, 91(3), 535-546.
- Plucker, J. (1997). Debunking the myth of the highly significant result: effect size in gifted. **Education Research Roper Review**, 20(2), 122-126.
- Rea, L. & Parker, R. (1997). **Designing and conducting survey research: a comprehensive guide** (1st edition). San Francisco, CA: Jossey – Bass publishers.
- Stevens, J. P. (1980). Power of multivariate analysis of tests. **Psychological Bulletin**, 88, 728-737.

- Thalheimer, K. & Cook, M. (2002). **How to calculate effect sizes from published research: A simplified methodology.** Retrieved, August 22, 2002 from: <http://www.work-learning.com/effect-sizes.htm>.
- Thompson, B. & Synder, P. (1997). Statistical significance testing practices in the journal of experimental education. **The Journal of Experimental Education**, 66(1), 75-79.
- Vaske, 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.
- Wilkinson, L. (1992). **Statistical significance testing: alternatives and considerations.** paper presented at the educational research association, Houston, TX, January 31, 1992.
- Wilkinson, L. (1999). Statistical methods in psychology journals: guidelines and explanations. **American Psychologist**, 54(8), 594-604.
- Woolly, T. & Dawson, G. (1983). A follow-up power analysis of the tests used in the journal of research in science teaching. **Journal of Research in Science Teaching**, 20, 673-681.

<< وصل هذا البحث إلى المجلة بتاريخ 2011/4/10، وصدرت الموافقة على نشره بتاريخ 2011/7/24 >>