تقدير القوة الإحصائية للدراسات النفسية في المجلات العربية

(2000 -2012)

د. عبد الناصر السيد عامر

أستاذ مساعد القياس والتقويم

كلية التربية–جامعة قناة السويس

الملخص:

هدفت الدراسة إلى تقدير القوة الإحصائية للدراسات النفسية المنشورة في المجلات العربية في المدة من 2000 إلى 2012، وبلغت عينة الدراسة 49دراسة وتضمنت 233 اختبار(ت)(53.3%)، و176 معامل ارتباط بيرسون(40.3%)، و28 انحدار متعدد(6.4%). وقدر حجم التأثير والقوة الإحصائية باستخدام برنامج G-Power وأشارت النتائج إلى أن وسيط القوة لحجم التأثير الصغير 0.28 ولحجم تأثير متوسط0.98، ولحجم تأثير كبير 0.99. وعموما يجب أن يولي الباحث مزيدا من الانتباه لتحليل القوة القبلي والبعدي لأهميتها في جودة تصميم البحث وتعميم النتائج.

الكلمات المفتاحية: قوة الاختبار الإحصائي، اختبارات الفروض.

Abstract

The study aimed to estimate the statistical power in published psychological studies in Arabian Journals form 2000 to 2012. The research sample consist of 49 studies included 233 t test (53.3%), 176 Pearson correlation (40.3%), and 28 multiple Regression (6.4%). Effect Size and statistical power were estimated by G-Power program. The overall median

power for all tests was 0.28 for small effect size, 0.98 for medium effect, and 0.99 for large effect size. In general more attention should be placed for both a prior and post hoc power because of its importance of Research design's quality and result generalization.

Keywords: Statistical power, and Hypotheses tests.

مقدمة

تعتبر الإحصاء العمود الفقري لصناعة القرارات والوصول إلي المعلومات والمعارف في العلوم النفسية والإجتماعية وغيرها. وتقوم نظرية الاستدلال الإحصائي علي إستراتيجية اختبارات الفروض التي تقوم علي صياغة الفروض الإحصائية وتحديد الاختبار الإحصائي المناسب ومستوي الدلالة الإحصائية ثم اتخاذ القرار في ضوء قيمة p حيث إذا كانت $p > 0.05$ نقبل عدم الدلالة الإحصائية (قبول الفرض الصفري Ho) وإذا كانت $p \leq 0.05$ نقبل الدلالة الإحصائية (رفض الفرض الصفري Ho).

وقد أوضح Gigerenzer(2004) مدخلين لاستخدام اختبارات الفروض هما مدخل (1955) Fisher الذي يتضمن صياغة الفرض الإحصائي الصفري واتخاذ القرارات في ضوء قيمة p وتجاهل الخطأ من النوع الثاني (β) ويتم تقدير حجم العينة من خلال القواعد المتعارف عليها وشبه (2004) Gigerenzer هذا المدخل بمرحلة الأنا Ego عند فرويد، أما المدخل الثاني فهو مدخل Pearson & Neyman الذي انتقد مدخل Fisher في صياغة الفرض الصفري فقط، واقتراحاً صياغة فرضين علي الأقل أحدهما صفري والآخر بديل، وأهتم بالخطأ من النوع الأول(α) الخطأ من النوع الثاني (β)، ويتم تحديدهما قبل إجراء الدراسة وفي ضوء ذلك يتم تقدير حجم العينة الذي يجعل a و B أقل ما يمكن، وعلي ذلك فقد اهتم مدخل Neyman & Pearson بالقوة الإحصائية للدراسة وشبه (2004) Gigerenzer هذا المدخل بالأنا العليا Super Ego عند فرويد، ومن ثم يلعب مفهوم القوة الإحصائية

كحلقة وصل بين مدخل Fisher إلي مدخل Pearson & Neyman ويعتبر Cohen (1988) القوة الإحصائية المؤشر الكمي للحساسية الاحصائية للبحث أو للتجربة، واعتبرها Sawyer & Balla (1981) مظهراً أساسياً لصدق الاستنتاج الإحصائي Conclusion Validity Statistical، ويعتبرها البعض مؤشر هام لجودة تصميم الدراسة. وعلي الرغم من أن القوة الإحصائية جزء مهم في البحوث إلا أنه مازال الاهتمام منصباً علي الدلالة الإحصائية Verma &Goodle, 1995

وبالرغم من أهمية القوة الإحصائية في إجراءات اختيار الفروض، إلا أن الباحثين أهملوا هذا الجانب في إجراء البحوث ويوجد عدم اهتمام بها في تصميم وتفسير نتائج البحوث ,Cohen ;2004 (Baguley, 1988; Hallahan & Rosenhal, 1996).

وأن مفهوم القوة الاحصائية جزء من مرحلة التخطيط للدراسة لتحديد حجم العينة المطلوب للوصول إلي مستوي قوة مرغوبة وتفسير نتائجها، وأن عدم تضمينه يؤدي إلي سوء تفسير نتائج الدراسات والوصول إلي استنتاجات خاطئة من البيانات ,Baroudi &Orlikowski) 1989; Hallahan & Rosenthal, 1996).

تعتبر دراسة Cohen (1962) من أشهر الكلاسيكيات في مجال دراسات القوة الإحصائية، ثم تبعها بمرجع احصائي وهو تحليل القوة الإحصائية للعلوم السلوكية عام 1969، ومنذ ظهور ذلك المرجع بدأ الاهتمام بدراسات القوة الإحصائية وحجم التأثير في مختلف التخصصات.

وقد بدأ الاهتمام حديثاً بتضمين القوة الإحصائية في الدراسات للمساعدة في تفسير النتائج لأنه لو فشل الاختبار في إظهار التأثير أو العلاقة الموجودة فقد يكون هذا ليس بسبب عدم وجود تأثير ولكن ربما بسبب ضعف الأختبار في إظهار التأثيرات الدالة احصائياً.

ومن مخاطر القوة المنخفضة للاختبارات الاحصائية حدوث تعارض بين نتائج الدراسات السابقة التي تتناول نفس الظاهرة موضع الدراسة.

المشكلة:

في ضوء اطلاع الباحث علي كثير من الدراسات النفسية والتربوية المنشورة في المجلات العربية منذ العقد الأخير من القرن الماضي فإنه لم تشر أي دراسة من قريب أو من بعيد إلى القوة الإحصائية سواء كان في مرحلة تخطيط الدراسة وذلك لتحديد حجم العينة أو في مرحلة النتائج وتفسيرها، وأيضاً توجد ندرة في الدراسات التي تناولت تقدير القوة الإحصائية للدراسات النفسية التربوية في المجلات العربية وذلك لمعرفة الجودة الإحصائية وصدق الاستنتاج الإحصائي لنتائج الدراسات.

وتساءل(1992) Cohen ليس واضح تماماً لماذا يستمر الباحثون في تجاهل تحليل القوة في دراساتهم، ويرى(2004) Cashen &Geiger أن هذا الاستنتاج تم على أساس عدم وجود دراسة واحدة قامت بتقدير القوة للاختبارات المتضمنة فيها كما ورد في معظم دراسات تحليل القوة وعلى ذلك يمكن طرح التساؤلات الأتية للدراسة:

1.ما وسيط القوة الإحصائية للاختبارات الإحصائية وهي اختبارت للعينات المستقلة (T) ومعامل ارتباط بيرسون (r) والانحدار المتعدد (R) معاً وذلك لحجم تاثير صغير، ومتوسط وكبير لعينة من الدراسات النفسية التربوية المنشورة في المجلات التربوية العربية في المدة من 2000 حتى 2012 ؟

2.هل توجد علاقة بين القوة الإحصائية ونوع الاختبار (R ,r ,t) ؟

3.هل تتحسن القوة الإحصائية عبر الزمن (ما قبل 2009 وما وبعدها)؟

الهدف:

هدفت الدراسة إلى:

1.تقدير القوة الإحصائية لعينة من الدراسات النفسية المنشورة في مجلات تربوية ونفسية في البيئة العربية.

2.التعرف على العلاقة بين نوع الاختبار الاحصائي والقوة الاحصائية في الدراسات النفسية.

3.تحديد مدى وجود تحسن في القوة الاحصائية عبر الزمن (ما قبل 2009 وما بعدها).

الأهمية:

1.توجيه انتباه الباحثين إلى أهمية تضمين تقدير القوة الاحصائية في دراساتهم سواء في مرحلة تصميم الدراسة لتحديد حجم العينة أو في مرحلة تقييم نتائجها لزيادة الطمأنينة والثقة في تعميم نتائج الدراسة.

2.تشخيص الوضع الحالي للقوة الاحصائية في البحث النفسي العربي.

الإطار النظري والدراسات السابقة:

أولا: **مفهوم القوة الإحصائية**: صناعة القرار في نظرية اختبارات الفروض الكلاسيكية تتم بمقارنة قيمة P بمستوي الدلالة الإحصائية (α) وعلي ذلك يتم قبول الفرض الصفري (Ho) أو رفضه (قبول الفرض الإحصائي البديل HA) . وقبول Ho لا يعني عدم وجود فروق بالفعل ولكننا لم نمتلك الدليل والوسائل للرفض، وعلي ذلك فإن استخدام الإحصاء الاستدلالي يؤدى إلي الوقوع في أخطاء قراريه كالأتي:

جدول (1)

الأخطاء القرارية المصاحبة للإحصاء الاستدلالي

الفرض الصفري(H_o)		**المجتمع**	
		أرفض	**أقبل**
العينة	أرفض	لا خطأ	الخطأ من النوع الأول(α)
	أقبل	الخطأ من النوع الثاني(β)	لا خطأ

علي ذلك يوجد خطأين هما: الخطأ من النوع الأول (α) وتعرف بأنه رفض الفرض الصفري (H_o) علي مستوي الدراسة (العينة) ولكنه حقيقة علي مستوي المجتمع، بمعني أن الدراسة توصلت إلي وجود دلالة إحصائية (فروق أو علاقة) ولكن علي مستوي المجتمع (الحقيقي) لا توجد فروق أو علاقة، ويتم التحكم في هذا الخطأ قبل إجراء الدراسة من قبل الباحث من خلال تبني مستوي دلالة إحصائية معينة مثل 0.05 أو 0.01 و 0.001 وغيرها. والخطأ من النوع الثاني (β) ، ويعرف بأنه قبول الفرض الصفري علي مستوي الدراسة (العينة) ولكنه مرفوض علي مستوي المجتمع، بمعنى أن الدراسة توصلت إلي عدم دلالة إحصائية (لا فروق أو علاقة) علي مستوي عينة الدراسة ولكن علي مستوي المجتمع توجد فروق أو علاقة، أي الفشل في الوصول إلي دلالة إحصائية في الدراسة وفي الواقع يوجد ويصعب التحكم في هذا الخطأ قبل إجراء الدراسة ولا يستطيع الباحث التحكم فيه أو السيطرة عليه.

وترتبط القوة الإحصائية بالخطأ من النوع الثاني (β) وتعرف بإحتمال تجنب حدوث (β) أو الوصول إلي التأثير الحقيقي باحتمال (1-β) علي ذلك فالقوة الإحصائية مقياس للدرجة التي تكون من خلالها الدراسة قادرة علي إظهار الفروق أو العلاقات الحقيقية الموجودة في مجتمع الدراسة (Cohen, 1988) ، أو تعريف Wilson- Vanvoorhin &Morgan (2007) بأنها احتمالية رفض الفرض الصفري المرفوض في المجتمع، بمعني تجنب حدوث الخطأ من النوع الثاني (β)، أو قدرة الاختبار علي الحصول علي دلالة إحصائية عندم يكون الفرض الصفري غير حقيقي (Hallahan & Rosenhal,1996) فالقوة الإحصائية P=0.80 تعني وجود احتمال 80% للحصول علي دلالة احصائية وهي موجودة بالفعل بالتالي فأن فرصة ارتكاب (β) تساوي 0.2.

وعلي ذلك فأن دراسات تحليل القوة الإحصائية تجيب علي تساؤل ما هي النسبة الاحتمالية لرفض Ho ؟ ولا يوجد اتفاق بين الباحثين عن المستوي المقبول من القوة الإحصائية للدراسة. وهنا يوصي Cohen (1988) بأن علي الباحثين الوصول إلى مستوي قوة لا يقل عن 0.80 حيث انه مستوي مقبول وواقعي. والقوة أقل من 0.8 تعني ثقة منخفضة في مصداقية القرار (Rossi, 1990). وعلي ذلك يكون أقصي خطأ من النوع الثاني β = 0.20 ، ويري Verma & Goodale (1995) أن الخطأ a أشد خطورة مقارنة بخطأ (B) ولتحقيق قوة معقولة لابد أن تكون نسبة ألفا إلى بيتا هي (4إلي 1) أي أقصي ألفا يساوي 0.05 وأقصي بيتا يساوي 0.02.

ثانياً: **العوامل المؤثرة في القوة الإحصائية**: القوة الإحصائية هي دالة وظيفية لثلاثة متغيرات أساسية وهي: مستوي الدلالة الإحصائية، وحجم العينة، وحجم التأثير (Cohen, 1988; Fox & Mathers, 1998; Halahan & Rosenlhal, 1996)

، وعبر عنها Cohen (1988) بالمعادلة الآتية $P = F(\alpha, N, ES)$. وفيما يلي عرض لأهم المتغيرات المؤثرة علي القوة كما تناولتها الأدبيات البحثية:

حجم العينة (N) : كلما زاد حجم العينة قلت أخطاء المعاينة بالتالي تزيد قوة الاختبار،ويري Graf, Mazen, Kellogg & Hemmasi(1987) أن زيادة حجم العينة يؤدي إلي أخطاء أقل وثبات أعلي واحتمالية عالية لرفض الفرض الصفري والحصول علي دلالة إحصائية ويشير Saywer & Ball (1981) أن العشوائية في اختيار العينة وتجانسها يقلل الأخطاء المعيارية بالتالي يزيد من دقة القياسات.

2. **الخطأ من النوع الأول** (α): بجعل ألفا اقل صرامة تزيد قيمة القوة، فبدلا من جعل آلفا 0.01 يمكن أن تكون 0.05 فأكثر، ذلك لان العلاقة بين α و(β) علاقة عكسية، فإذا قلت قيمة a تزيد (β) ويدوره تقل القوة الإحصائية.

ولكن إذا جعل الباحث α أكثر صرامة فانه يتطلب زيادة حجم العينة، ويري البعض بأنه لا مانع من زيادة قيمة a حتى لو زادت عن 0.1 في سبيل ان تكون القوة عالية (Sink & Mvududu, 2010) .

3.**حجم التاثير** (ES): ويعتبر حجم التأثير الاعتبار الأساسي لحساب قوة الاختبار الاحصائي (Houser, 2007)، زيادة حجم الفروق بين المتوسطات أو قوة العلاقة بين المتغيرات يؤدي إلي احتمال كبير للحصول علي دلالة إحصائية بالتالي إلى قوة إحصائية عالية، وذلك لان حجم التأثير المرتفع يعني القدرة العالية لإظهار أو الكشف عن الفروق الموجودة بالفعل في المجتمع، وهذا يعني رفض الفرض الصفري والحصول علي قوة عالية (Mazen et al., 1987; Sink &,Mvududu, 2010) أي أن القوة الإحصائية تزيد عندما يكون للاختبار قدرة عالية علي إظهار حجم تأثير كبير مقارنة بحجم تأثير صغير.

4.**تصميم البحث**: حدد (Baguley, 2004; Baroudi& Houser, 2007; Sink& Orlikowski, 1989; Mvududu, 2010) عوامل تصميم البحث المؤثرة في القوة وهي:

أ-**اختبار المتغيرات**: دقة الباحث وشموليته في اختبار أهم المتغيرات المستقلة المؤثرة في الظاهرة يزيد من القوة واختيار متغيرات لا ترتبط بالظاهرة يزيد من القوة يؤدي إلى عدم دلالة إحصائية.

ب-**الثبات**: وضح Zimmerman & Williams (1989) طبيعة العلاقة بين الثبات والقوة فالثبات المرتفع للقياسات يؤدي إلي قوة إحصائية عالية.

ج- **توزيع مفردات العينة**: فالتوزيع المتساوي علي المجموعات المتعددة أو مستويات العامل يزيد القوة، وأن التوزيع غير المتساوي على مجموعات التصميم العاملي يؤدي إلى تقليص الارتباطات أو الفروق، ومن ثم تقليص حجم التأثير وهذا بدوره يؤدي إلي قوة منخفضة.

د- **المعالجة التجريبية**: إجراء معالجات وإجراءات تجريبية قوية ومعيارية وكذلك تصميمات القياسات المتكررة يؤدي إلي قوة عالية (Hallahan & Rosenthal, 1996; Sink & Mvududu, .2010)

هـ- **الاختبار الإحصائي**: الاختبارات الإحصائية البارمترية أكثر قوة من الاختبارات اللابارامترية وذلك لان تقطيع المتغيرات المتصلة الاعتدالية ينشأ عنها فقد كثير من المعلومات، كما أن الاختبارات الإحصائية الموجهة(ذو اتجاه واحد إلي أعلي أو أدنى) أكثر قوة من الاختبارات غير الموجهة ذو اتجاهين أو ذيلين) Fox & Maehersm, 1997; Sink & Mvududu, 2010)كما ان استخدم الاختبار في المقارنة بين أكثر من مجموعتين في تحليل واحد أفضل من المقارنة بين كل مجموعتين علي حدة Omnibus test لان هذا يزيد من القوة(Hallahan & Rosenthal,.1996)

و- **دراسات ما وراء التحليل**: دراسات ما وراء التحليل تكون أكثر قوة من الدراسات المفردة (Hallahan & Rosenthal, 1996).

ز- **استخدام بدائل الاختبارات الإحصائية الكلاسيكية**: استخدام إجراءات استدلالية رياضية Bayesian أكثر مرونة لاختيار الفروض الكلاسيكية ثنائية القطبية في القرار، ومن ثم تصبح عملية صنع القرار أكثر مرونة ومنطقية في رفض الفرض الصفري.

ثالثا: **أنواع تحليل القوة الإحصائية**: توجد عدة مداخل لتناول تحليل القوة الاحصائية حددها.

(Baguley, 2004; Baroudi& Or-Likowski, 1989; Erdfelder& :Buchner, 2007; wong, 2010)

1.**قبل أجراء الدراسة** Prospective or prior power analysis: في هذا النوع من التحليل يتبنى الباحث مستوي قوة إحصائية اللازمة للوصول إليها ولتكن P= 0.08، ومستوي دلالة إحصائية معينه مثلآ α= 0.05، وحجم تأثير مفترض في ضوء تراث الظاهرة أو دراسه استطلاعية. ويكون هدف الباحث تحديد حجم العينة المطلوب لتحقيق قوة إحصائية محددة مسبقا قبل إجراء الدراسة وهذا التحليل خطوة أساسيه في مرحله تصميم الدراسة وهذا التحليل للقوة يفيد في تقليل التكلفة والموارد نتيجة زيادة حجم العينة بدون داعي ويوصي Cohen (1992) بأهمية إجراء هذا النوع من التحليل للدراسة قبل إجرائها.

2.**أثناء الدراسة** Statistical Power during study:الاستخدام الثالث لتحليل القوة هو استخدامه في مجال تقيم البرامج أو التجارب عندما يستخدم تصميم القياسات التتبعية، فالسؤال الذي يطرح نفسه في هذا المعترك هو: إلى متى تستمر إجراءات البرنامج أو الدراسة لتحقيق مستوي قوة مرضي؟.

3. **بعد أجراء الدراسة** Retrospective, post hoc power analysis: وهنا يجري الباحث تحليل القوة بعد الانتهاء من الدراسة للتقييم البعدي لنتائج الدراسة، والتعرف علي احتمالية الوصول إلى

تفسيرات واستنتاجات صحيحة، ويتم تقدير القوة خاصة للنتائج غير الدالة أحصائيآ وذلك لأن عدم الدلالة قد لا يعود لانعدام الدلالة الفعلية في المجتمع (تأثير حقيقي صغير)، بل إن الاختبار الإحصائي له قوة إحصائية ضعيفة غير قادرة علي كشف الفروق الموجودة في المجتمع ويتم تقدير القوة للاختبار من خلال معرفة مستوي الدلالة الإحصائية a وحجم التأثير وحجم العينة.

4.**تحليل الحساسية** Sensitivity analysis: ينصب الاهتمام في تحليل الحساسية على تقدير حجم التأثير في المجتمع بمعلوميه aو pو N، ويكون مفيد في تقييم الدراسات المنشورة وهذا يجيب علي تساؤل وهو: ما هو حجم التأثير الذي بإمكان الاختبار تحقيقه في ظل قوة إحصائيه تساوي 0.80؟ ويتم تقدير الحساسية قبل إجراء الدراسة.

5.**تحليل المحكية** Criterion analysis: في تحليل المحكية يتم تقدير aكوظيفة للقوة وحجم التأثير وحجم العينة.

رابعا: **تقدير القوة الإحصائيه**: **يتم تقدير القوة الإحصائية بالطرق الآتية**:

1.**جداول** Cohen (1988): قدم Cohen (1988) جداول يتم من خلالها تقدير P لمعظم الاختبارات الإحصائية مثل R, F, r, X^2, t وغيرها من الاختبارات اللابارامترية، وذلك بمعلومية مستوى الدلالة الإحصائية وحجم العينه وحجم التأثير.

2.**البرامج الاحصائية**: ومنها G-power, power& Precision Thomas& Kerbs وحدد، NCSS, PASS, Stat, NQuery (1997) المواصفات المثالية لأي برنامج وهي:

- يغطي اختبارات عديدة بسيطة او متدرجة، وبارامترية ولا بارامترية.

- يسمح بمرونة في التعامل مع المواقف الجديدة.

- يعطي نتائج دقيقة ويسهل تفسير مخرجاته.

- يقدر القوة وحجم التأثير وحجم العينة أي يغطي معظم تحليلات القوة.

- يقدر معالم المدخلات اللازمة لتقدير القوة مثل حجم التأثير.

- يأخذ في حسبانه مؤشرات متنوعة لحجم التأثير.

- يتميز بالجودة في العرض سواء أشكال أو رسومات وجداول.

- يسهل توثيقه.

وأشار Faul et al. (2007) إلى توافر هذه المواصفات في برنامج G-power وهو أكثر استخداما في دراسات تقدير القوة.

الدراسات السابقة:

تم إجراء الدراسات السابقة الخاصة بالقوة الإحصائيه منذ زمن بعيد ففي مجال علم النفس قام Cohen (1962) في دراسته المشهورة بتقدير القوة الإحصائية للدراسات المنشورة في مجلة علم النفس الاجتماعي وغير العاديين عام 1960. وتوصل إلى أن متوسط القوة الإحصائية للاختبارات التي لها حجم صغير 0.18، ولحجم تأثير متوسط 0.48، ولحجم تأثير كبير 83. 0،، ومتوسط القوة العام 0.48 ،وهذا يعني أن نصف النتائج تعطي عدم دلالة إحصائية بالرغم من وجود فروق حقيقية ولكن الاختبارات غير قادرة على إظهارها. وبعد24 عاماً أعاد Sedlmeier & Gigerenzer (1989) دراسة كوهين علي 54 دراسة منشورة في نفس المجلة في مجلد 1984، وتوصل إلى أن دراستين أشارت إلى القوة الاحصائيه ولم تقدر أي منهما القوة وبلغ وسيط القوة لأحجام التأثير الصغيرة والمتوسطة والكبيرة 0.14 ,0.44 ,0.90

على التوالي، أي بعد 24 عاماً فإن وسيط القوة أقل قليلاً عما كان عليه فى عام 1962 وهذا يدل على أنه بمرور الزمن فإن قوة الاختبار تقل أو ثابتة نسبياً إلى حد ما.

وتوصل Chase & Chase (1976) إلى أن متوسط القوة الإحصائية لدراسات علم النفس التطبيقي لحجم تأثير صغير 0.25 وحجم تأثير متوسط0.67 ، وحجم تأثير كبير 0.86.

وقدر Rossi (1990) القوة الإحصائية لـ25 دراسة منشورة في مجلة علم النفس غير العادين ومجلة علم النفس الاجتماعي والشخصية وتوصل إلى أن متوسط القوة لحجم تأثير صغير 0.26، ولحجم تأثير متوسط 0.71، ولحجم تأثير كبير 0.85.

وفي مجال التربية: **قام** Brewer (1972) بتقدير القوة الإحصائية في ثلاث مجلات تربوية وتوصل إلى أن وسيط القوة في المجلة الأمريكية للبحوث التربوية لاحجام التأثير الصغيرة 0.14، للمتوسطة 0.58، وللكبيرة 0.78. ولمجلة البحوث في التدريس (التعليم) هي 0.87 ,0.22 ,0.71 على التوالي ولمجلة البحث ربع سنوية هي:,0.25 ,0.80 ,0.14 علي التوالي.

وقدر Haase (1974) القوة الإحصائية في مجلة تربية المرشد والإشراف وتوصل إلى أن متوسط القوة الإحصائية لحجم تأثير صغير و متوسط وكبير هي 0.10 ,0.37 ,0.74 علي التوالي.

وفي مجال التربية العلاجية: قدر Woolley (1983) القوة الإحصائية للدراسات المنشورة في مجلة التربية العلاجية في المدة من 1980 حتى 1988، وتوصل إلي أن 91% من الدراسات التي لها حجم تأثير صغيرة تمتلك قوة اقل من 0.50، و25% من الدراسات التي لها حجم تأثير متوسط لها قوة اقل من 0.50، و 4% من الدراسات التي لها حجم تاثير كبير لها قوة اقل من 0.50، ووسيط القوة لحجم تأثير صغير هو 0.14، ولحجم تأثير متوسط 0.71،حجم تأثير كبير0.99 .

وفي مجال التربية الموسيقية توصل Daniel (1993) إلى أن وسيط القوة الإحصائية لحجم تأثير صغير 0.13، والقوة لحجم تأثير متوسط 0.64، وحجم تأثير كبير 0.97.

وفي مجال الإدارة حسب Mazen et al. (1987) متوسط القوة الإحصائية للدراسات المنشورة في عدة مجلات للإدارة وكانت لحجم تأثير صغير 0.31، ولحجم تأثير متوسط 0.77، ولحجم كبير 0.86.

وقدر Baroudi& Orlikowski (1989) متوسط القوة في دراسات الإدارة منشورة في أربع مجلات في المدة من 1980 إلى 1985 وبلغت 0.19 لحجم تأثير صغير، و0.06لحجم تأثير متوسط، و0.83 لحجم تأثير كبير.

وقدر Verma & Goodle (1995) متوسط القوة لدراسات في مجلة الإدارة التطبيقية ومجلة علوم القرار وتوصل إلي أن 60% من الاختبارات لها قوة أقل من 0.80. وكانت الاختيارات التي لها أعلى قوة إحصائيه هي الانحدار يليه الارتباط ثم اختبارات لعنينين مستقلين.

وتوصل Casher &Geiger (2004) في تقيميه لـ 53 دراسة في مجال الإدارة ألى أن 4دراسات (9.3%) ناقشت القوة ومن هذه الدراسات الأربعة ثلاثة دراسات قدرت القوة الإحصائية تراوحت من 0.48 إلى0.98، وأن متوسط القوة لـ 77 اختبار هى 0.29.

وفي البيئة العربية قام النجار (2006) بتقدير قوة الاختبار في مجلة الإدارة في المملكة العربية السعودية وبلغ متوسط القوة لحجم تأثير صغير 0.19، ولحجم تأثير متوسط 0.79، ولحجم تأثير كبير 0.93.

وفي مجال التسويق قدر Sawyer & Ball (1981) القوة الاحصائية حيث بلغ وسيط القوة الإحصائية لحجم تأثير صغير0.41، ولحجم تأثير متوسط 0.89، ولحجم تأثير كبير 0.98.

ويتضح أن الدراسات السابقة اتبعت منهجية تقدير القوة الاحصائيه لكل الاختبارات معآ في ضوء ثلاثة مستويات من حجم التأثير الضعيف والمتوسط والكبير وذلك في ضوء محكات Cohen (1988) لحجم التأثير، ويلاحظ أن الدراسات التي لها حجم التأثير الضعيف والمتوسط لم تصل إلى المستوى المرغوب من القوة وهو 0.80 وذلك لجميع التخصصات.

الطريقة والإجراءات

1. **العينة**: تضمنت عينة الدراسة 49 دراسة نفسية منشورة في المجلات التربوية والنفسية في البيئة العربية في المدة من 2000 حتى 2012؛ وتم اختيارها عشوائيآ من مكتبة الدراسات العليا بكلية التربية بالإسماعلية، جامعه قناة السويس، وكانت أكثر المجلات استخداماً هي المجلة المصرية للدراسات النفسية (28.6%)، يليها مجلة كلية التربية الزقازيق (16.3%)، ثم مجلة العلوم النفسية والتربوية في البحرين (10.2%)، ثم مجلة كلية التربية بالمنصورة ومجلة كلية التربية بسوهاج وكلاهما بنسبة (61%)، ثم تنوعت الدراسات في مجلات كليه تربية طنطا وأسيوط والمنوفية والإسكندرية ورسالة الخليج العربي. اعتمدت الدراسة على تقدير القوة لاختبارات "ت" لعينتين مستقتلين ومعامل ارتباط بيرسون وأسلوب الانحدار المتعدد، وتوزعت كالأتي 223 اختبار "ت"(53.5%)، و 176 معامل ارتباط بيرسون (40.3%)، و28 تحليل الانحدار (6.4%)، وبلغ عدد الدراسات التي استخدمت المنهج الوصفي 40 دراسة (81.6%)، والتي استخدمت المنهج التجريبي (18.4%).وبلغ عدد الدراسات في تخصص علم النفس التربوي 35 دراسة (71.4%)، وفي تخصص الصحة النفسية 14دراسة (28.6%)وتوزعت عينات هذه الدراسات كالإتي:

جدول (2)

توزيع عينات الـ 49 دراسة.

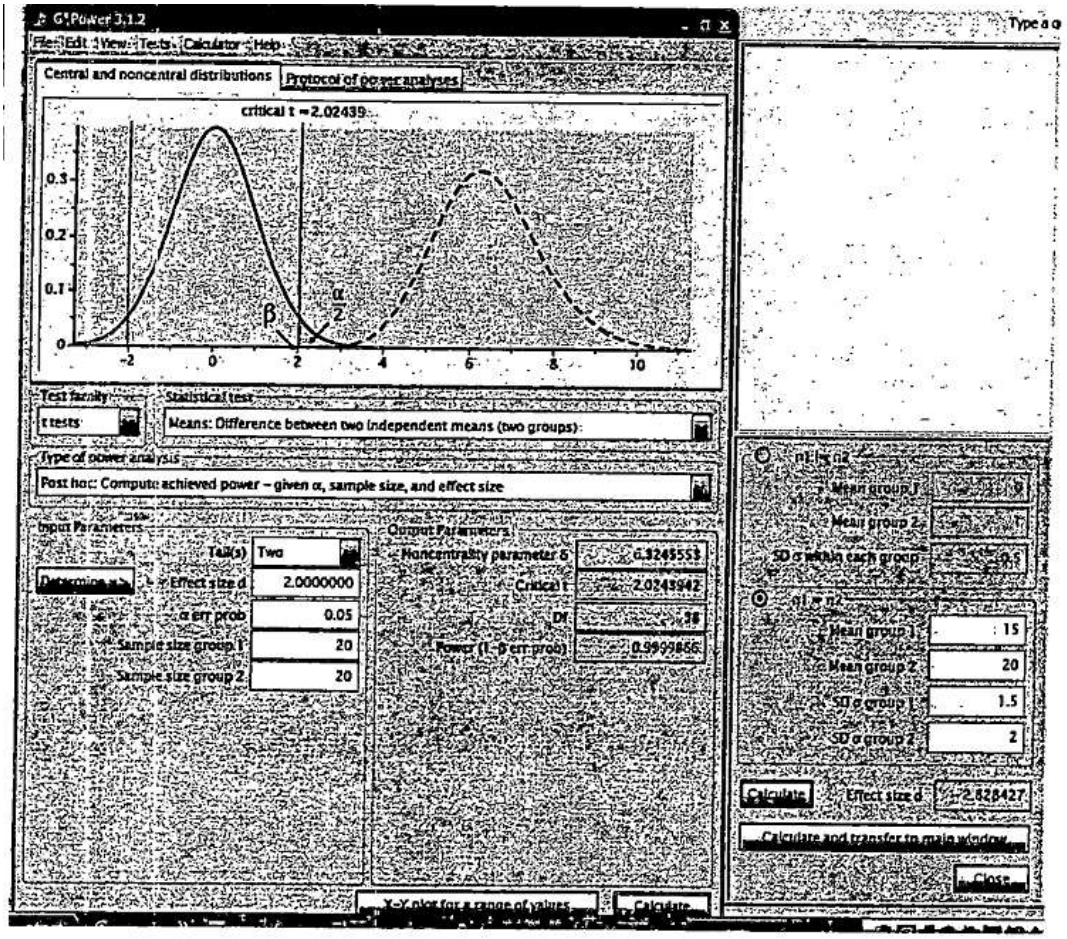
العينة	التكرار	النسبة المئوية
100≥	15	30%
200-101	15	32%
200 <	19	38%

وبلغ متوسط حجم العينة في الدراسات 210 فرداً، ووسيطها 160.00 فرداً وبانحراف معياري 178.78.

2.**البيانات**: تم جمع بيانات أساسية من كل دراسة وهى مجلة النشر، والتخصص(عام نفس تربوي- صحة نفسية)، والمنهج المستخدم (وصفي- تجريبي)، والعينة المستخدمة لتحليل كل اختبار في الدراسة (قد يختلف حجم العينة لكل اختبار)، والاختبار المستخدم (R, r, T)، واتجاهه (اتجاه واحد -اتجاهين)، وفي تحديد اتجاهه الاختبار وأحياناً لم يذكر صاحب الدراسة اتجاهه ولكن الباحث حدده من خلال صياغة فرض الدراسة، ومستوي الدلالة الإحصائية لكل اختبار (0.001,0.01)والدلالة الإحصائية (دال- غير دال)، وبعد ذلك تكويدها وإدخالها في ملف بيانات في برنامج SPSS 10.

3.**تقدير حجم التاثير والقوة الإحصائية**: لتقدير القوة الإحصائية للاختبار يستلزم تحديد ثلاثة معالم وهي مستوي الدلالة الإحصائية وحجم العينة وحجم التأثير ولكن حجم التأثير ولم يقدر من قبل صاحب الدراسة وعلي هذا قام الباحث الحالي بالاعتماد علي برنامج G- (Faul, Erdfelder, Long Power& Buchmer, 2007) لاستخدامه في غرضين هما تقدير حجم التأثير وتقدير قوة الاختبار،

وهذا البرنامج يتكون من شاشة بها قوائم ثم مربع أبيض لعرض الرسومات، وبعد ذلك تظهر أيقونة هى Test Family، وتتكون من Exact (تتضمن معامل ارتباط بيرسون والانحدار) وعائلة F-Test (وتتضمن تحليل التباين و التغاير و MANOVA)،وعائلة T-test تتضمن اختبار (ت) والاختبارات اللابارامترية) وعائلة Z. واعتمد الباحث الحالي علي عائلة T-test وعائلة Exact. وبعد ذلك تم تحديد نوع الاختبار ونوع تحليل القوة، كما في الشكل الآتي:



شكل (1): **مخرج برنامج G-Power في حالة اختبار"ت" للعينات المستقلة**.

واعتمدت الدراسة على نوع التحليل Post hoc: Compute achieved power، أي أن الهدف هو حساب القوة الإحصائية للاختبار بعد إجراء الدراسة. ولكن الباحث قام اولآ بتقدير حجم

التأثير وذلك بإدخال نوع اتجاهه الاختبار وα وحجم العينة لكل مجموعة (في حالة T-test) ثم ضغط علي Determine المقابلة لـ Effect size ظهرت شاشة فرعية بها حالتين من التحليل الحالة الأولي $n1 \neq n2$ والحالة الثانية $n_2=n_1$، وبعد ذلك تم إدخال متوسط كل مجموعة والانحراف المعياري في حالة $n1=n2$ وبالضغط علي Calculate يقوم البرنامج بتقدير حجم التأثير، ثم يقوم البرنامج بإدخالها في ملف تقدير القوة ثم تقدر القوة ودرجات الحرية وقيمة الاختبار.

برنامج G-power قائم علي معادلات(1988) Cohen لتقدير حجم التأثير يأخذ بحدود كوهين لتحديد أما إذا كان حجم تأثير صغير أو متوسط أو كبير وهي في حالة معامل ارتباط بيرسون والانحدار 0.01 صغير و0.09 متوسط و0.25 كبير، وفي حالة اختبار(ت) لعينتين مستقلتين 0.2 صغير و0.5 متوسط و0.8كبير ثم أدخلت قيمتي حجم التأثير والقوة إلي برنامج SPSS في ملف البيانات.

نتائج الدراسة:

1.عرض القوة الإحصائية للاختبارت الإحصائية معآ في ضوء حجم التأثير (صغير ،متوسط، كبير):

جدول (3)

التكرارت والتكرار المتجمع الصاعد لتوزيع القوة الإحصائية ل49 دراسة 437 اختبار)

مستوي القوة الإحصائية	حجم تأثير صغير		حجم تأثير متوسط		حجم تأثير كبير	
	التكرار	المتجمع الصاعد المئوي	التكرار	المتجمع الصاعد المئوي	التكرار	المتجمع الصاعد المئوي
>0.99	7	3.1	41	39.5	93	83.8
0.90-0.98	21	12.6	26	56	15	97.3
0.80-0.89	14	18.8	15	79.6	2	99.1
0.70-0.79	16	%26	9	88.3		
0.60-0.69	15	23.7	3	91.3		
0.05-0.59	18	40.8	5	96.1		
0.40-0.49	18	48.9	4	99	1	100
0.30-0.39	13	54.7	1	100		

				63.2	19	0.20-0.29
				79.4	36	0.10-0.19
				%100	46	<0.09
111		103		223		العدد الكلي
0.97		0.88		042		المتوسط
0.99		0.98		0.38		الوسيط
1.00		0.99		0.05		المنوال

ويتضح من الجدول بالنسبة لتقدير القوة في حالة حجم التأثير الصغير يتبين ان وسيط القوة الإحصائية للاختبارات 0.38وهي صغيرة وهذا متوقع لان العلاقة بين حجم التاثير والقوة عكسية ولكنها تزيد عن قيمة الوسيط أو المتوسط في مجال العلوم النفسية والتربوية (Brewer, 1972; Cohen,1962; Hase, 1974; Rossi, 1990; Sedlmeier & Gigetenzer, 1989; Chase & Chase,1978) وكذلك في مجالات الإدارة والتسويق مثل (Baroudi& Orlikowski, 1989; Mazen et al, 1987; Sawyer& Ball, 1981; Verma& Goodle, 1995). ويتضح من الجدول السابق أن 18.8% من الاختبارات لها مستوي قوة 0.8 فأكثروا وأن 59.1% من الاختبارات لديها فرصة اقل من 50% للحصول علي دلالة إحصائية. وبما أن β-1= P، فأن= β 0.62 وعلي

ذلك فإن نسبة ارتكاب الخطأ من النوع الثاني (β) إل الخطأ من النوع الأول(α) هي تقريبا 12.6(0.05/ 0.62) إلى الواحد الصحيح وهي نسبة عالية وتؤدي إلى مخاطر عالية في القرار.

وبالنسبة لتقدير القوة في حالة حجم تأثير متوسط يتبين من الجدول أن وسيط القوة للاختبارات الإحصائية هي 0.98، وهي تفوق نظيرتها في الدراسات النفسية والتربوية الأجنبية فوجدها (Cohen, 1990 0.48، و(1989) Sedlmeier &Gigetenzer 0.44 ، و(1990) Rossi 0.71و (1996) Chase &Chase 0.67، و(1972) Brewer 0.58. واتضح أن 79.6% من الاختبارات لها قوة إحصائية 0.80 فأكثروا و3.8% من الاختبارات لها قوة أقل من 0.50 وأن نسبة ارتكاب الخطأ من النوع الثاني إلى الخطأ من النوع الأول 0.40 إلي واحد، وهي نسبة مقبولة ومرضيه.

وبالنسبة لتقدير القوة في حالة حجم التأثير الكبير يتبين أن وسيط القوة الإحصائية 0.99 وهذا يدل علي أنه إذا كان حجم التأثير كبير فإن الاختبارات لها القدرة على كشف هذه الفروق بدرجة تصل تقريبا إلى 100% وأن نسبة الخطأ في النوع الثاني إلى الخطأ من النوع الأول هي 0.20 إلى واحد، وهي نسبة مثالية تؤدي إلى قرارات مصحوبة بأقل قدر من المخاطر القرارية.

2. **الاختبار الإحصائي**: **يعبر عن القوة الإحصائية في ضوء متغير بمستويين أقل من** 0.8 **و**0.8 **فأكثر**:

جدول (4)

التكرارات في مستوي القوة الاحصائية ونوعية الاختبار.

	الاختبار الاحصائي				
		المستقلة	معامل بيرسون	الانحدار	المجموع
قوة الاختبار	0.8	141 %60.5 %32.3	62 %35.2 %14.2	-	203 %100 %46.5
	≥0.8	92 %39.5 %26.1	114 %64.8 %26.1	28 %100 %6.4	234 %100 %53.5
	المجموع	233 %100 %53.3	176 %100 %40.3	28 %100 %6.4	437 %100 %100

حيث أن النسبة المئوية الأولي تعبر عن النسبة المئوية للعمود، والثانية للمجموع الكلي.

يتبن من جدول (4) أن 60.5% من اختبارات المستقلة و14.2% من معامل ارتباط بيرسون لا يتوفر لهما الحد المرغوب من القوة الاحصائية، وأن 39.5% من اختبارات المستقلة 64.8% من معامل ارتباط بيرسون و100% من الانحدار المتعدد يتوفر لهم قوة مرغوبة 0.80 فأكثر، ويظهر من الجدول أيضا أن 46.5% من اختبارات r, t لها قوة أقل من 0.80 و53.5% من الاختبارات لها قوة أكبر من 0.80 فأكثر كما يتضح أن 203 اختبار(46.4%) لا يتوفر لديهم الحد المقبول من القوة الإحصائية أقل من 0.8. كما توجد علاقة بين نوع الاختبار وقوة الاختبار وهي دالة إحصائية عند 0.01، (X^2= 51.73, df= 2, p= 0.00) وكذلك (Ø =0.34, P= 0.00).

وللتنبؤ من نوعية الاختبار الاحصائي بالقوة الاحصائية، تم استخدام اختبار معامل عدم التأكد Uncertainty Coefficient والمتبوع بمخرج X^2 وقد اتضح أن قيمة معامل التأكد UC= 0.104; P= 0.00 وهذا يعني أن نوعية الاختبار الاحصائي فسرت تقريبا 10% من تباين القوة الاحصائية.

3. تحديد مدى وجود تحسن في القوة الاحصائية عبر المدتين من 2009 وما قبلها، ومن 2010 وما بعدها:

وتم اختيار عام 2009 كحد فاصل بين فترتين وذلك مراعاة لوجود تناسب بين أعداد الدراسات قبلها وبعدها فقط.

جدول (5)

التكرارات لمستوي القوة الإحصائية عبر الفترتين الزمنيين

	الاختبار الاحصائي			
		2009≤	**2010≥**	**المجموع**
قوة الاختبار	<0.8	94 %51.4 %21.5	109 %42.9 %24.9	203 %46.5 %46.5
	≥ 0.8	89 %48.6 %20.5	145 %57.1 %33.2	234 %53.6 %53.6
		183 %100 %41.9	254 %100 %58.1	437 %100 %100

يظهر من الجدول ان 89 اختبار (48.6%) في المدة من 2000 حتى 2009 لها قوة 0.8 فأكثر، بينما 145 اختبار (57.1%) في المدة من 2010 حتى 2012 لها قوة تساوي أو أعلى من

0.08 ويتبين أن ($X^{2=}$ 3.05, df= 1, P= 0.08) أي أنها غير دالة إحصائيا بمعني أن القوة الإحصائية للاختبارات في المدة من 2009 فأقل لم تتحسن عن القوة الإحصائية في الفترة من 2010 فأكثر بمعنى عدم وجود تحسن للقوة الاحصائية بمرور الزمن. ولا توجد علاقة بين المتغيرين ($\emptyset = 0.084, \emptyset = 0.80$) وهذا يتفق مع Sedlmeier &Gigerenzer (1989) عندما توصل إلى عدم وجود تحسن في القوة الإحصائية للدراسات النفسية ومنذ دراسة Cohen (1962) إلى عام 1989.

المناقشة والتعليق

هدفت الدراسة إلي تقدير قوة الاختبار الإحصائي في الدراسات النفسية التربوية العربية، باعتبارها مكون هام لصناعة القرار الإحصائي، ويعتبر استخدام هذا المكون مهمل في الدراسات النفسية التربوية العربية. وتدخل القوة الإحصائية بداية من التخطيط للدراسة من خلال تحديد حجم العينة اللازم لتحقيق مستوي قوة مقبول. وبتقدير القوة الإحصائية لاختبارات R, r, t تبين أن الدراسات النفسية والتربوية العربية لها وسيط قوة إحصائية تفوق نظريتها الأجنبية سواء كان لحجم تأثير صغير، أو متوسط، أو كبير، بل المدهش أنه إذا كان حجم التأثير متوسط فإن الدراسات لها وسيط قوة بلغ 0.9 وهذا لم يتحقق في التراث الأجنبي سواء في العلوم الاجتماعية أو الإدارية أو غيرها ويتفق هذا تماما مع ما توقعه الصياد(1988) بأن القوة الإحصائية للبحث النفسي العربي أكبر من نظريتها الأجنبية لكن ما السبب لحدوث هذا التحسن الكبير؟، من وجهه نظر الباحث قد يعود إلى أسباب عديدة منها أن التصميم المنهجي. لعظم الدراسات السابقة التي قدرت القوة تضمنت اختبارات إحصائية متعددة وتنوعت بين البارامترية مثل(R, f, r, t) واللابارامترية مثل(مان وتيني، كروسكال واليس، ويلكوكسون، X^2) ولكن الدراسة الحالية اعتمدت علي (R, r, t) فقط، وأن استخدام الاختبارات اللابارامترية يؤدي إلى انخفاض القوة علي أساس أن القوة

الإحصائية للاختبارات اللابارامترية اقل من البارامترية. كذلك اعتمدت الدراسات السابقة على دراسات منشورة مجلة بعينها في مدة زمنية تتراوح من أربع إلى خمس سنات ما عدا (Mazen et al., 1987; Rossi, 1990) في حين اعتمدت الدراسة الحالية على مجلات تربوية متنوعة في مدة زمنية تزيد عن العشر سنوات وربما يعود هذا التفوق في الدراسة الحالية إلى أن الدراسات السابقة قامت بتثبت اتجاهه الاختبار إلى ذو اتجاهين لكل الدراسات التي قدر لها القوة وكذلك تثبت مستوي الدلالة الإحصائية عند 0.05 ولكن الدراسة الحالية اعتمدت على نوعية اتجاهه الاختبار كما حدده صاحب الدراسة التي تم تقيمها فقد يكون ذو اتجاهين(72.1%) أو ذو اتجاه واحد(27.9%)، وكما اعتمدت على مستوي الدلالة الإحصائية التي حددتها الدراسة 0.01(54.7%) أو(0.05 (45.3%).

ويري الباحث أن لحجم العينة دورا كبيرا في زيادة قوة الاختبار الإحصائي فبلغ متوسط حجم العينة 209.86 فردا في حين بلغت عند (1989) Baroudi &Orlikowski 84 فرد. وعند (1993) Daniel تراوح حجم العينة عبر الدراسات من75 إلأى 122 فردا بمتوسط 99 فردا علي ذلك فإن البحث العربي يتفوق علي نظيره الأجنبي من حيث حجم العينة وهذا ما أشار إليه الصياد(1988). وربما يرجع هذا التفوق إلى أن الدراسات النفسية التي تم تقيمها تتضمن (18.4%) دراسة تجريبية ودائما الدراسات التجريبية لها قوة إحصائية أعلى من الدراسات الوصفية فبلغ وسيط القوة للدراسات التجريبية 0.99 في حين بلغ وسيط الدراسات الوصفية(0.83).

أما بالنسبة لتقدير القوة لكل اختبار علي حدة، بلغ وسيط اختبار (ت) لعينتين مستقلتين(0.58)، ومعامل ارتباط بيرسون (0.91)، وتحليل الانحدار(0.99) وعلي ذلك فإن وسيط قوة لأسلوب الانحدار ومعامل ارتباط بيرسون أكبر من وسيط اختبار(ت) ولكن لماذا حدث هذا التفوق؟ ربما يرجع هذا إلى أن الاستخدام الشائع لاختبار (ت) يكون ذات طبيعة استكشافية(فروق بين ذكور

وإناث، فروق بين تخصصات وغيرها)، في حين أن الارتباط والانحدار يستخدم لأغراض توكيدية في ضوء نظرية قوية والبحث التوكيدي أكثر قوة من البحث الاستكشافي كما أن تفوق القوة لأسلوب تحليل الانحدار المتعدد عل معامل ارتباط بيرسون يرجع إلى وجود أكثر من متغير مستقل يسهم في تفسير تباين المتغير التابع بالتالي فإن حجم التأثير لتحليل الانحدار مرتفع نتيجة تعدد المتغيرات المستقلة عكس اختبار(ت) ومعامل ارتباط بيرسون وأيضا لطبيعة المتغيرات لهذه الأساليب فالارتباط والانحدار يتعاملا مع متغيرات متصلة في حين اختبار (ت) يتعامل مع متغيرات متصلة ومنفصلة معا.

في المجمل فإن وسيط القوة الإحصائية للاختبارات الثلاثة معا في الدراسات 0.84 وهذا يعني أن الاختبارات الإحصائية أو بمعنى اشمل الدراسات النفسية العربية لديها احتمال 84% للكشف عن الفروق الحقيقية عندما تكون موجودة وأن الخطأ من النوع الثاني هو 0.16 وهي نسبة مقبولة حيث لا يفضل أن تزيد عن 0.20 وعلى ذلك يمكن القول أن الدراسات النفسية تمتع بمستوى قوة احصائية مرغوب.

الفشل في رفض الفرض الصفري يترك للباحث في حيرة فهل يرجع هذا إلى قوة احصائية غير كافية للاختبار لرفض الفرض الصفي أم إلى عدم وجود فروق حقيقية للظاهرة؟.

ولكن ماذا لو كانت القوة الإحصائية للدراسة أقل من0.80؟ ربما يرجع هذا إلى أن الفروق الحقيقة موجودة وأن الاختبار ليس لديه القوة الكافية لإظهارها وهذا نتيجة إلى عيوب في انتقاء الاختبار، والتحقق من مسلماته أو القياسات أو في تصميم الدراسة أو غيرها من الأمور المنهجية والإحصائية، وقد يرجع انخفاض القوة إلى أن الفروق الحقيقة صغير جدا في الواقع الحقيقي وبالتالي فهذا ليس عيبا في الدراسة، إنما هو طبيعة الظاهرة تحت الدراسة.

ولكن هل توجد استراتيجيات للتعامل مع القوة المنخفضة؟ فيرى البعض أنه لزيادة القوة الاحصائية يجب زيادة مستوى الدلالة الإحصائية a لأنه من شأنه أن يقلل مستوي الخطأ من النوع الثاني (B) ولكن(Daniel(1993 يري أنه حلا غير مرغوبا كما يجب معرفة حجم التأثير للظاهرة تحت الدراسة من خلال تقدير حجم التأثير قبل إجراء الدراسة عن طريق مراجعة التراث وأيضا يمكن زيادة القوة من خلال زيادة أحجام العينات لأنها الضمانة الأكيدة لزيادة قوة الاختبار.

ومجال دراسات القوة الإحصائية في البيئة العربية يكاد يكون محدود جدا خاصة في الدراسات النفسية وعلي ذاك فإن الدراسة الحالية جاءت لتسد الفجوة في مجال المنهجية البحثية والإحصائية وهذا ما أسهمت به الدراسة في تشخيص الوضع الحالي للقوة الإحصائية في الدراسات النفسية العربية. وتوصلت الدراسة إلى عدة توصيات أهمها:

1.علي الباحثين حساب القوة الإحصائية لكل اختبار إحصائي في الدراسة وهذا يسمح للباحثين تفهم المخاطر الناتجة من عدم الدلالة الإحصائية للنتائج. فالدراسات مع قوة إحصائية عالية والتي ليس لها دلالة احصائية تعطي تدعيم قوى للقرار، بينما الدراسات مع قوة منخفضة تعطي تدعيم ضئيل سواء للفرض الصفري أو البديل.

2. اعتمدت الدراسة الحالية علي ثلاثة اختبارات هي R, r, t، ويمكن للدراسات المستقبلية أن نتناول اختبارات متنوعة من بارامترية ولا بارامترية ومن اختبارات النموذج البسيط، والمتعدد(تحليل التباين)، والمتعدد المتدرج(MANCOVA, MANOVA)، وغيرها ومن استقراء الدراسات السابقة.

3. على محرري المجلات التأكيد علي ضرورة تقدير القوة الإحصائية علي الأقل بعد إجراء الدراسة كشرط للنشر في المجلات باعتباره مؤشر للجودة الإحصائية والمنهجية للدراسة.

المراجع

1.الصياد، عبدالعاطي أحمد.(1988). الدلالة العلمية وحجم العينة المصاحبة للدلالة الإحصائية لاختبار(ت) في البحث التربوي والنفس العربي. بحوث مؤتمر البحث التربوي الواقع والمستقبل رابطة التربية الحديثة بالقاهرة، 2، 197- .233

2. النجار، عبدالله بن عمر.(2006). دراسة تحليلية لقوة الاختبار الإحصائي في البحوث الإدارية المنشورة. المجلة العلمية لجامعة الملك فيصل (العلوم الإنسانية والإدارية)، 7، 261- 293.

3. Baeuley, T (2004). Understanding statistical power in the context of applied research. Applied Ergonomics, 35, 73- 80.
4. Baroudi, J. J., & Orlikowski, W. J. (1989). The problem of statistical power in MIS research, 13, 87-106.
5. Brewer, J. (1972). On the power of statistical tests in the American Educational research journal. American & Educational Research Journal, 9, 391-401.
6. Casher, H. C., & Geiger, S. W. (2004). Statistical power and the testing of null hypotheses: a review of contemporary management Research and Recommendation for future studies.

   Organizational Research Methods, 7, 151 - 167.
7. Chase, J., & Chase, B. (1976). Statistical power analysis of applied. psychological research. Journal of Applied Psychology, 61,234 - 237.
8. Cohen, J. (1962). The statistical power of abnormal—social psychological research: A review.Journal of Abnormal and Social Psychology, 65,145- 153.

9. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale. NJ: Erlbaum.

10. Cohen, J. (1990). Things I have learned (So far). American Psychologist, 45, 1304 - 1312.

11. Cohen, J. (1992). A power premier. Psychological Bulliten, 112,155- 159.

12. DaNIel, T. D. (1993, November). A statistical power analysis of the Quantitative techniques used in the "journal research in music Education 1987 - 1991". paper presented at the Annual Meeting of the mid — south Educational Research Association. New Orleans, L.A.

13. Faul, F., ErdFelder, E., Lang, A. G., & Buchner, A. A. (2007). G - power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175- 219 .

14. Fisher, R. A. (1955). Statistical methods and scientific inference. Journal of Royal Statistical Society (B), 11,69-11.

15. Fox, N., & Mathers, N. (1997). Empowering research: statistical power in general practice research. Family Practice, 14, 324-329.

16. Gigerenzer , G. (2004). Mindless statistics. The Journal of Socio-economics, 33, 587 — 606.

17. Haase, R. F. (1974). Power analysis of research in counselor education. Counselor Education and Supervision, 14,124-132.

18. Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. Behavioral Research, 34, 489-499.

19.Houser, J. (2t)07). How many are enough? Statistical power analysis and sample size estimation in clinical research. Journal of Clinical Research Best Practice, 3, 1- 5.

20.Mazen, A. M., Graf, L. A., Kellogy, C. E., & Hemmasi, M. (1987). statistical power in contemporary management research. The Academy of Management Journal , 30 , 369 - 380.

21.Rossi, J. S. (1990). Statistical power of psychological research: what have we gained in 20 years?. Journal of Consulting and Clinical Psychology, 58, 646- 655.

22.Sedlmeier, p.,_ & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies?, psychological Bulletin, 105, 309- 316.

23.Sink, G. A., & Mvududu, N. H. (2010). Statistical power, sampling, and effect size: three keys to research relevancy. Counseling Outcome Research and Evaluation, 1, 1—18.

24.Sawyer, A. G., & Balia, A. (1981). Statistical power and effect size in marketing research. Journal of Marketing Review, XVIII, 275 — 290.

25.Thomas, L., & Krebs, C. J. (1997). A review of statistical power analysis software. Bulletin of the Ecological Society of American, 78, 128 — 139.

26.Wang, L. W. (2010). Retrospective statistical power: fallacies and Recommendation. New Born & Infant Nursing Reviews, 10, 55- 59.

27.Wilson-vanvoorhiis, C. R., & Morgan, B. L. (2007). Understanding Power and Rule of thumb for determining sample size. Tutorial in Quantitative Methods for psychology, 3, 43 — 50.

28. Woolley, T. W. (1983)._ A comprehensive power analytic investigation of research in medical education. Journal of Medical Research, 58, 710-715.

29. Verma, R., & Goodale, J. C. (1995). Statistical power in operations managementresearch. Journal of Operations Management, 13, 139 — 152.

30. Zimmerman, D. W., & Williams, R. H. (1989). Statistical power analysis and reliability of measurement. Journal of General Psychology, 116,359-369.