

العنوان:	نحو معيار موحد لتقدير حجم التأثير لاختبار "ت" لعينتين مستقلتين في البحث التربوي النفسي العربي
المصدر:	المجلة المصرية للدراسات النفسية
الناشر:	الجمعية المصرية للدراسات النفسية
المؤلف الرئيسي:	الصيد، عبدالعاطي أحمد
مؤلفين آخرين:	عامر، عبدالناصر السيد(م. مشارك)
المجلد/العدد:	مج23, ع80
محكمة:	نعم
التاريخ الميلادي:	2013
الشهر:	يوليه
الصفحات:	5 - 21
رقم MD:	1011910
نوع المحتوى:	بحوث ومقالات
اللغة:	Arabic
قواعد المعلومات:	EduSearch
مواضيع:	البحوث التربوي، الإحصاء الوصفية، العلاقات الارتباطية
رابط:	http://search.mandumah.com/Record/1011910

نحو معيار موحد لتقدير حجم التأثير لاختبار "ت"
لعينتين مستقلتين في البحث التربوي النفسي العربي

أ.د. عبد العاطي أحمد الصياد د. عبد الناصر السيد عامر

الملخص:

هدفت الدراسة إلى عرض لبعض مؤشرات حجم التأثير المستخدمة في اختبار "ت" المستقلة ($d, r, W2, \gamma^2$) ومعرفة هل المؤشرات الأربعة تعطي نفس التفسير؟ وأيضاً محاولة اقتراح معايير لكل مؤشر من المؤشرات الأربعة حتى نصل إلى نفس التفسير.

ولتحقيق ذلك تم اختيار (٢٧) دراسة نفسية مصرية وعربية منشورة في المجالات النفسية والتربوية وتضمنت (١٤٩) اختباراً لـ "ت" وتم تقدير حجم التأثير للمؤشرات المختلفة باستخدام المعادلات المذكورة في متن الدراسة ٣، ٥، ٦، ٧.

وتم تحليل البيانات باستخدام الاحصاء الوصفي ومعامل ارتباط بيرسون والتحليل العاملي الاستكشافي.

وتوصلت النتائج إلى وجود اختلاف في تفسير المؤشرات الأربعة سواء في ضوء مقياس المتوسط والوسيط، وكذلك وجود علاقات ارتباطية موجبة عالية بينها، كذلك وجود عامل عام تشبعت عليه هذه المؤشرات، وتم اقتراح معايير جديدة.

نحو معيار موحد لتقدير حجم التأثير لاختبار "ت" لعينتين مستقلتين في البحث التربوي النفسي العربي

أ.د. عبد العاطي أحمد الصياد¹ د. عبد الناصر السيد عامر²

مقدمة:

تعرضت منهجية اختبارات الفروض الصفرية Null Hypothesis Testing (NHT) لانتقادات شديدة لاعتمادها على قيمة الخطأ من النوع الأول p-Value في مقارنتها بمستوي الدلالة الاحصائية الذي ارتضاه الباحث (α) لصناعة القرار. وأصبح استخدام قيمة p محل جدل ونقاش لدي الباحثين (الصياد، 1988، Cohen, 1988, 1998; Kirk, 1996; Thompson, 1998). وقد أشار Nickerson (2000) إلى وجود خطأ شائع في تفسير الدلالة الاحصائية وهو الاعتقاد السائد بأن القيمة الصغيرة لـ P تعني أن تأثير المعالجة كبير والقيمة الكبيرة تعني أن تأثير المعالجة صغير، وأوصى الباحثون باستخدام مؤشر حجم التأثير Effect size كجزء مكمل لاختبارات الفروض الصفرية، وحجم التأثير ليس مفهوماً جديداً ولكنه أصبح أكثر شيوعاً مع استخدام أسلوب ما وراء التحليل Meta-Analysis (أسلوب لمراجعة الدراسات السابقة كميًا) منذ بداية عام 1970. وأصبح استخدامه أكثر انتشاراً في الدراسات النفسية مع تأكيد الجمعية النفسية الأمريكية (American Psychological Association (APA) (1994) على استخدامه في الدراسات التي تنشر في المجلات التابعة لها، كذلك جمعية

¹ أستاذ الاحصاء التربوي ومناهج البحث - كلية التربية بالاسماعيلية - جامعة قناة السويس

² أستاذ القياس والتقويم التربوي المساعد - كلية التربية بالاسماعيلية - جامعة قناة السويس

الأبحاث التربوية الأمريكية American Educational Research Association (AERA) (2006).

وأشار Wetzal, Matzke, Lee, Rouder, Iverson, & Wagenmakers (2011) بضرورة أن تدعم اختبارات الفروض الصفرية بمؤشرات أخرى مثل حجم التأثير وفترات أو حدود الثقة ومؤشر بيز Bayes factor وهو مؤشر من الاحصاء الاستدلالي لبيز Bayesian لتكميم وتدعيم التأثير التجريبي وهو مؤشر مأخوذ عن مؤشر Odds Ratio.

ويعرف حجم التأثير بأنه تكميم لحجم الفروق أو لقوة العلاقة بين متغيرين فأكثر (Coe, 2004; Cohen, 1988)، ولذلك فهو مؤشر هام في تكميم فاعلية برامج التدخل وتفسر مقدار الفاعلية (Coe, 2004)، وعلى الرغم من أهمية هذا المؤشر كجزء مكمل لاختبارات الفروض الصفرية إلا أن استخدامه في الدراسات النفسية التربوية في البيئة العربية محدوداً في حدود علم الباحثين. وعلى ذلك فإنه ليس من الضروري أن يسأل الباحثون عن وجود تأثير أو علاقة بل تزداد الحاجة إلى السؤال الضروري وهو كم مقدار هذا التأثير أو العلاقة؟، بل وأحياناً إلى هل هذا التأثير كاف لأن يؤخذ في الاعتبار؟ (Kirk, 1996; Mc-Grath & Meyer, 2006).

والتصميم البحثي الأكثر انتشاراً في البحث النفسي التربوي والسلوكي هو الذي يتضمن مقارنة بين متوسط مجموعتين على متغير تابع كمي، ومن الاختبارات الشائعة الاستخدام في الدراسات النفسية في البيئة العربية هو اختبار "ت" المستقلة حيث نسبة استخدامه في البحث النفسي التربوي في البيئة العربية ٢٤,٧% وفي البيئة المصرية ٣٨,٢% وإجمالي بنسبة ٣١,٧% (عامر، ٢٠١٢). وأصبح استخدام اختبار "ت" مقروناً بقيمة P في اختبارات الفروض الصفرية محل انتقادات شديدة (Cohen, 1994; Cumming, 2008; Nickerson, 2000).

مشكلة الدراسة:

يري الصياد (١٩٨٨) أن الباحث التربوي والنفسي يصارع من أجل الحصول على دلالة احصائية ويختبر فروضاً أطلق عليها بالفروض ذات القيمة القرارية المحدودة.

واستقراءً للدراسات النفسية في المجالات التربوية العربية والمصرية لاحظ الباحثان استخدام محدود لمؤشرات حجم التأثير المصاحبة لاختبار 'ت' المستقلة، وكذلك للعينة الواحدة.

واستخدمت الدراسات مؤشرات مختلفة لتقدير حجم التأثير لاختبار 'ت' المستقلة، فبعضها استخدم مؤشر d لكوهين (Cohen, 1969)، والبعض استخدم مؤشر مربع إيتا η^2 ، والبعض استخدم مؤشر مربع اوميغا w^2 ليهيز (Hayes, 1963) والبعض استخدم معامل الارتباط الثنائي الأصيل (r_{pb})، وإن تم تقدير حجم التأثير فإن بعض الباحثين قام بتفسير قيمته والبعض الآخر لم يفعله.

وتتنوع مؤشرات تقدير حجم التأثير في اختبار 'ت' المستقلة وهذا التنوع يتضح في ضوء ثلاثة مداخل وهي:

أولاً: مؤشرات الفروق المعيارية بين المتوسطات: وهي تقدر حجم التأثير في ضوء تكميم حجم الفروق وتتضمن مؤشرات عديدة أهمها مؤشر d لكوهين (Cohen, 1969)، ومؤشر نلتا لجلال (Glass, 1976)، ومؤشر g لبيدجز (Hedges, 1981)، ويعتبر مؤشر كوهين أشهرهم وأكثرهم استخداماً في حساب حجم التأثير وأيضاً توظيفاً في دراسات ماوراء التحليل، ويتم تقدير مؤشر d بالمعادلة التالية:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{pooled}} \quad \longrightarrow \quad (1)$$

حيث \bar{x}_1, \bar{x}_2 هما متوسطي المجموعتين كلاً على حدة، σ_p هي الانحراف المعياري المشترك ويتحدد بالمعادلة الآتية:

$$\sigma_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}} \quad \longrightarrow \quad (2)$$

حيث $S_1, 2S_2$ هما تباين درجات المجموعتين المضابطة والتجريبية كلا علي حدة ،
وكلاً من n_1, n_2 حجم الأفراد في المجموعتين علي حدة ، ويتم تقدير قيمة d من قيمة t
المحسوبة كالتالي:

$$d = t \left[\frac{n_1 + n_2}{\sqrt{(n_1 \cdot n_2) df}} \right] \longrightarrow (3)$$

حيث df درجات الحرية وتساوي $(n_1 + n_2 - 2)$.

واعتبر (1988) Cohen أن قيمة ٠,٢٠ حجم تأثير صغير ولكنه يستحق الإهتمام
و ٠,٥ حجم تأثير متوسط، و ٠,٨ فأكثر حجم تأثير كبير.

ثانياً: مؤشر العلاقة: هو مؤشر يقيس قوة العلاقة بين متغيرين وهو معامل الارتباط
الثنائي الأصيل (Point-Biserial Correlation r_{pb}) ، وهو حالة خاصة من معامل
ارتباط بيرسون ، ويستخدم عندما يكون للمتغير المستقل تصنيفي بمستويين: (0-1)
والتابع متصل وتتحدد قيمته كالتالي (Mc-Grath & Meyer, 2006):

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_2}{S_y} \sqrt{p_1 p_2} \longrightarrow (4)$$

حيث \bar{y}_1, \bar{y}_2 الأمتوسطي درجات المتغير التابع للمجموعتين الأولى والثانية علي
التوالي و S_y هي الانحراف المعياري للمتغير التابع للمجموعتين معاً ، P_1, P_2 نسبة
الأفراد في المجموعتين علي التوالي.

وتقدر قيمة r_{pb} من قيمة t المحسوبة كالتالي (Rosenthal, Rosnow & Rubin, 2000):

$$r = \frac{t}{\sqrt{t^2 + df}} \longrightarrow (5)$$

وتعتبر القيمة ٠,١٠ حجم تأثير صغير ، ٠,٣٠ حجم تأثير متوسط ، ٠,٥٠

حجم تأثير كبير.

ثالثاً: مؤشرات نسبة التباين المفسر: وهي مؤشرات تقيس نسبة التباين المفسر في المتغير التابع نتيجة للمتغير المستقل ومن أهمها:

١. مؤشر مربع أوميغا ω^2 : ويمكن تقديره من قيمة t كالآتي (Hayes, 1988):

$$\omega^2 = \frac{t^2 - 1}{t^2 + n_1 + n_2 - 1} \longrightarrow (6)$$

٢. مؤشر مربع أيتا η^2 : وتتحدد قيمته من t كالآتي (Richardson, 2011):

$$\eta^2 = \frac{t^2 - 1}{t^2 + n_1 + n_2 - 2} \longrightarrow (7)$$

وبالنسبة لمؤشرات التباين المفسر تعتبر القيمة ٠,٠١ حجم تأثير صغير و ٠,٠٦ حجم تأثير متوسط ، ٠,١٤ حجم تأثير كبير.

وقدم الصياد (١٩٨٨) عرضاً مفصلاً للمؤشرات المستخدمة في حالة اختبارات مثل مؤشرات كوهين ، مربع أوميغا ، ومربع أيتا ، ومربع إيسيلون وأوضح العلاقة بينهم.

ويوجد قليل من الاتفاق حول أي مؤشر لحجم التأثير يستخدم في أي موقف ، ففي العلوم السلوكية يعتبر مؤشر كوهين أكثر انتشاراً في البحوث التجريبية ، بينما مؤشر r أكثر استخداماً في البحوث المسحية (Kraemer, Morgan, Leech, Gliner, Vaske, & Harmon, 2003).

ويؤيد بعض الباحثين استخدام مؤشر d في إجراء دراسات ما وراء التجليل (Cohen, 1988; Lipsey & Wilson, 2001) بينما يري البعض ضرورة استخدام مؤشر r لمعياريته ومرورته في الاستخدام عبر تصميمات بحثية وتحليلات احصائية مختلفة (Rosenthal, et al., 2000).

ومما سبق يتضح تعدد مؤشرات حجم التأثير المستخدمة في اختبارات المستقلة وإن كان مؤشر d أكثر استخداماً ، ولكن هذا التعدد لهذه المؤشرات لنفس الاختبار يسبب نوع من الحيرة والارتباك بين الباحثين وهذا ما عبر عنه (Coe, 2004).

وفي ضوء العرض السابق للمؤشرات الأربعة يمكن عرض التساؤلات الآتية:

1. هل تختلف القيمة التأثيرية لمؤشرات حجم التأثير الأربعة المستخدمة في اختبار 'ت' المستقلة (d, r, w^2, γ^2) ؟
2. ما هي المعايير المقترحة للمؤشرات الأربعة بحيث لا تختلف القيمة التفسيرية من مؤشر لآخر؟
3. ما حجم العلاقة الارتباطية الموجبة بين قيم المؤشرات الأربعة؟
4. ما هو البناء العاملي الذي ينظم العلاقة بين مؤشرات حجم التأثير الأربعة؟

الهدف:

تهدف الدراسة في اطار منهجها الوصفي إلى:

1. التعرف على ما إذا تم تقدير حجم التأثير باستخدام مؤشرات مختلفة لاختبارات المستقلة تعطي نفس التفسير أم تختلف.
2. اقتراح معايير وحدود لمؤشرات d, r, w^2, γ^2 في ضوء مؤشر مئيني.
3. تقدير حجم العلاقة الارتباطية بين المؤشرات d, r, w^2, γ^2 لاختبار 'ت' المستقلة.
4. التعرف على طبيعة البناء العاملي للمؤشرات الأربعة..

الأهمية:

1. التأكيد على أهمية تقدير حجم التأثير كإجراء مكمل لاختبارات الفروض الصفرية في حالة اختبار 'ت' المستقلة.
2. توجيه انتباه الباحثين إلى تعدد مؤشرات حجم التأثير المستخدمة في اختبار 'ت' المستقلة.
3. اقتراح معايير جديدة لمؤشرات d, r, w^2, γ^2 في ضوء بيانات الدراسة.

المنهجية:

١. العينة: لتحقيق أهداف الدراسة تم اختيار (٢٧) دراسة منشورة في المجالات التربوية بجمهورية مصر العربية بالزقازيق وطنطا والمنيا والمينصورة وقنا وسوهاج والمنوفية والجمعية النفسية للدراسات النفسية وبالمدول العربية بالسعودية وقطر وسوريا في الفترة من ١٩٩٤ حتى ٢٠١١ ، وباعتبار أن وحدة التحليل هي الاختبار وليس الدراسة ، علي ذلك تضمنت السبعة وعشرين دراسة (١٤٤) اختباراً لـ 'ت' المستقلة بمتوسط (٥,٣) اختبار لكل دراسة وأقتصر الباحثان على اختبار 'ت' المستقلة الدالة احصائياً عند مستوى دلالة ٠,٠٥ أو ٠,٠١ ، وتم استبعاد الاختبارات غير الدالة احصائياً.

٢. تقدير حجم التأثير: تم تقدير حجم التأثير لمؤشر d باستخدام الصيغة (٣) ومؤشر r باستخدام الصيغة (٥) ، ولمؤشر w^2 باستخدام الصيغة (٦) ومؤشر γ^2 باستخدام الصيغة (٧) ، وتم تقديرها باستخدام مقدرين.

٣. المعالجة الاحصائية: تمت المعالجة الاحصائية بادخال كود الدراسة والمجلة وتاريخ النشر وعينة الدراسة ومستوى الدلالة الاحصائية وقيم مؤشرات d, r, w^2, γ^2 المقدره يدوياً في برنامج SPSS وتم إجراء:

أ. احصاء وصفي لكل مؤشر والمئينيات.

١. معامل ارتباط بيرسون بين قيم المؤشرات الأربعة.

ب. التحليل العاملي الاستكشافي باستخدام طريقة المكونات الاساسية.

النتائج:

١. للإجابة على التساؤل الأول: هل تختلف القيمة التأثيرية لمؤشرات حجم التأثير الأربعة المستخدمة في اختبار 'ت' المستقلة: (d, r, w^2, γ^2) ؟

تم إجراء أمر التكرارات مقرونناً بالاحصاء الوصفي لمعرفة الوصف الاحصائي

لكل مؤشر علي حدة.

جدول (1)

الاحصاء الوصفي لمؤشرات d, r, w^2, γ^2 .

المتوسط	الوسيط	الانحراف المعياري	أدنى قيمة	أعلى قيمة	المتوال	الانقواء
0,759	0,505	0,600	0,19	2,60	0,49	2,10
0,318	0,255	0,187	0,07	0,87	0,26	1,19
0,133	0,057	0,166	0,00	0,75	0,012	1,99
0,140	0,067	0,167	0,005	0,757	0,062	1,95

والمستقرى للجدول (1) يتضح أنه في ضوء المتوسط فإن هذه الدراسات لها حجم تأثير متوسط في ضوء مؤشرات d, r, w^2 ، بينما لها حجم تأثير كبير في ضوء مؤشر γ^2 أي أن تفسير مؤشر مربع إيتا يختلف عن تفسير مؤشري d, r, w^2 (حجم تأثير متوسط)، ولكن بالنظر إلى توزيع المؤشرات الأربعة نلاحظ أن توزيعها ملتوي للتواء موجباً وعلي ذلك فإن مؤشر النزعة المركزية الأنسب هو الوسيط، ويتبين من الجدول أنه في ضوء معيار الوسيط فإن تفسير حجم التأثير يكون متوسط لمؤشري كوهين d ومربع إيتا γ^2 ، بينما يكون ضعيفاً لمؤشري r, w^2 ومن ثم يتبين أن اختلاف المؤشر يمكن أن يؤدي إلى استنتاجات مختلفة خاصة للمؤشرات الأكثر انتشاراً في دراسات ماوراء التحليل وهما d, r وهذا يتفق مع (Mc-Grath & Meyer (2006).

ومن حيث قيمة حجم التأثير فإن هذه النتيجة تتفق تماماً مع ما توصل إليه الصياد (1988) بأن وسيط مؤشري مربع إيتا ومربع أوميغا هي علي التوالي 0,063 و 0,057 لاختبار "ت" للدراسات في علم النفس للباحثين المحترفين (ما بعد الدكتوراه) ومتوسطهم هو 0,14 و 0,13 ولكن القيمة التفسيرية للباحثين الحاليين تختلف مع ما أشار إليه الصياد (1988) حيث فسر حجم التأثير بالمتوسط ولكن الدراسة الحالية تتفق معه بأن حجم التأثير من نوع المتوسط (في ضوء مقياس الوسيط) لمؤشر مربع إيتا γ^2 ، ومن النوع الضعيف لمؤشر مربع أوميغا w^2 (أيضاً في ضوء مقياس الوسيط).

وأيضاً نتيجة الدراسة تتفق مع جرادات وجودة (٢٠٠٥) حيث توصلنا من خلال مسح لاختبار 'ت' في الدراسات المنشورة في مجلة جامعة اليرموك وبلغ عينة اختبار 'ت' (١٦٦) وقدّر حجم التأثير باستخدام مؤشر كوهين d وتوصل إلى أن متوسطه ٠,٦٣ وهذا حجم تأثير من النوع المتوسط.

وعلى ذلك فإن متوسط مؤشر γ^2 يعطي تقديراً لحجم التأثير أعلي من متوسط مؤشرات d, r, w^2 ويبدو هذا مقبولاً لأن مؤشر γ^2 يعطي تقديراً متحيزاً موجباً (Olejnik & Algina, 2000). وعلى ذلك يختلف التفسير ولكن في ضوء مقياس الوسيط فإن تفسير مؤشري r, w^2 هو من النوع الضعيف وهذا يختلف عن تفسير مؤشري d, γ^2 وهو من النوع المتوسط.

٢. للإجابة على التساؤل الثاني: ما المعايير المقترحة للمؤشرات الأربعة بحيث لا تختلف القيمة التفسيرية من مؤشر لآخر ؟

في ضوء هذا الاختلاف في التفسير بين المؤشرات الأربعة فإن وضع حدود جامدة لمعايير Benchmarks لمؤشر ما غير مقبولة فمثلاً وضع (1988) Cohen معايير تختلف عن معايير (2001) Lipsey & Wilson ولتوحيد قيمة المؤشرات تم اقتراح مدخل آخر لهذا التوحيد ويتم الحصول عليه من خلال المئينيات وهي كالآتي:

جدول (٢) قيم المئينيات للمؤشرات الأربعة.

المئينيات	d	r	w^2	γ^2
١	٠,٢٠٣	٠,٠٨٣	٠,٠٠١	٠,٠٠٦
٥	٠,٢٤٠	٠,١١	٠,٠١٠١	٠,٠١٢
١٠	٠,٢٨٥	٠,١٣٠	٠,٠١٥	٠,٠١٩
٢٠	٠,٣٥٠	٠,١٧	٠,٠٢٤	٠,٠٢٨
٢٥	٠,٣٩٠	٠,١٨	٠,٢٠١	٠,٠٣٣
٣٠	٠,٤٢٠	٠,٢٠	٠,٠٣٧	٠,٠٤٠
٤٠	٠,٤٧٠	٠,٢٣	٠,٠٥٠	٠,٠٥٧
٥٠	٠,٥٠٥	٠,٢٥٥	٠,٠٥٧	٠,٠٦٧
٦٠	٠,٦٦٠	٠,٣٠٠	٠,٠٨٢	٠,٠٩٥
٧٠	٠,٨٢٠	٠,٣٧٠	٠,١٣٧	٠,١٤٦
٧٥	٠,٩٠٧	٠,٣٩٧	٠,١٦٠	٠,١٦٧
٨٠	١,٠٠	٠,٤٤	٠,١٩٣	٠,١٩٩
٩٠	١,٥٨٥	٠,٦٣٥	٠,٤٠٠	٠,٤٥٩
١٠٠	٢,٦٠٠	٠,٨٧٠	٠,٧٥٤	٠,٧٥٧

ويمكن تبسيط الجدول السابق في جدول (٣) ونلاحظ بالنسبة لمؤشر d فإنه لم تختلف المعايير بين Cohen (1988) والدراسة إلا في حجم التأثير الكبير ، أما بالنسبة لمؤشر r فنلاحظ اختلافاً واضحاً في المعايير بين Cohen (1988) والدراسة ، حيث أن حجم التأثير المتوسط والكبير عند كوهين هما $0.30 - 0.50$ ، أما في الدراسة فهما 0.25 و 0.37 ، والدراسة الحالية تتفق مع (Lipsey & Wilson, 2001; Mc-Grath & Meyer, 2006) وهي 0.01 (صغير)، 0.24 (متوسط)، 0.37 (كبير).

جدول (٣) معايير Cohen (1988) ومعايير الدراسة الحالية للمؤشرات الأربعة.

Cohen (1988)	d			r			w^2			γ^2		
	صغير	متوسط	كبير	صغير	متوسط	كبير	صغير	متوسط	كبير	صغير	متوسط	كبير
	0.20	0.50	0.80	0.10	0.30	0.50	0.01	0.06	0.14	0.01	0.06	0.14
الدراسة الحالية	0.20	0.50	0.82	0.082	0.25	0.37	0.01	0.057	0.137	0.006	0.067	0.146

وأفرزت الدراسة الحالية اختلافاً واضحاً بين معايير w^2 ومعايير γ^2 كما هو واضح في جدول (٣) حيث إن حجم تأثير متوسط لـ w^2 هو 0.057 ولمؤشر γ^2 هو 0.067 وحجم تأثير كبير لمؤشر w^2 0.137 ولمؤشر γ^2 0.146 وهذا يتفق جزئياً مع معايير Kirk (1996) لمؤشر w^2 وهي 0.01 (صغير)، 0.059 (متوسط)، 0.138 (كبير) وهذا لأن مؤشر γ^2 هو تقدير عالي لحجم التأثير Over estimated على أساس أن مؤشر w^2 هو مؤشر استدلالي وليس وصفي.

على تلك فالاعتماد على معايير الدراسة يؤدي للوصول إلى تفسيرات واحدة لكل المؤشرات سواء d أو r أو w^2 أو γ^2 في حين أن الاعتماد على معايير Cohen (1988) يؤدي إلى تفسيرات مختلفة.

٣. للإجابة عن التساؤل الثالث: "ما حجم العلاقة الارتباطية الموجبة بين قيم المؤشرات الأربعة؟"

ولمعرفة العلاقة الارتباطية بين المؤشرات الأربعة تم تقدير معامل ارتباط بيرسون بين المؤشرات الأربعة وهي كالآتي:

جدول (٤) مصفوفة الارتباط بين المؤشرات الأربعة.

γ^2	W^2	r	d	
			١,٠٠	D
		١,٠٠	٠,٩٥٥	R
	١,٠٠	٠,٩٦٦	٠,٩٧٥	W^2
١,٠٠	٠,٩٩٠	٠,٩٥٨	٠,٩٦٥	Γ^2

(*) دالة احصائياً عند ٠,٠١

ويظهر من الجدول السابق وجود علاقة ارتباطية موجبة عالية جداً بين المؤشرات الأربعة وكانت اعلام بين γ^2 و W^2 وهذا طبيعي لأنهم مؤشرين من نفس العائلة لتحديد نسبة التباين المفسر وكانت أقلهم بين r , d وهذا طبيعي لأنهم من عائلتين مختلفتين فأحدهما مقياس لحجم الفروق والآخر مقياس لقوة العلاقة.

٤. للإجابة عن التساؤل الرابع: ما هو البناء العاملي الذي ينظم العلاقة بين مؤشرات حجم التأثير الأربعة؟

تم اجراء التحليل العاملي الاستكشافي للمؤشرات الأربعة باستخدام طريقة المكونات الأساسية (PC) وفيما يلي نتائج التحليل :

جدول (٥) تشبعات المؤشرات الأربعة علي العامل العام:

م	المؤشر	التشبع
١	d	٠,٩٨٦
٢	r	٠,٩٧٦
٣	W^2	٠,٩٩٤
٤	γ^2	٠,٩٩١
	الجذر الكامن	٣,٩
	التباين المفسر	٩٧,٥٣

ومن الجدول السابق يظهر أن العامل العام فسر ٩٧,٥% من تباين الظاهرة، وأن تشبع كل مؤشر متقارب مع الآخر علي العامل العام مما يمكن القول أن أحد المؤشرات يمكن استخدامه محل الآخر.

المناقشة والتعليق:

أصبح الاعتماد علي الدلالة الإحصائية فقط لاتخاذ قرار غير كافي ولا بد أن يصاحبه دلائل أخرى أهمها حجم التأثير كمؤشر لتكميم حجم الفروق أو قوة العلاقة.

ومن أهم الاختبارات استخداماً في الدراسات النفسية هو اختبار "ت" المستقلة ، ويحتار الباحث في تقدير حجم التأثير لهذا الاختبار نظراً لتعدد مؤشراته فمنها مؤشر حجم الفروق مثل d وغيره ، ومؤشر قوة العلاقة r ، ومؤشر نسبة التباين المفسر مثل γ^2 ، w^2 ، والدراسة هدفت إلى تحديد ما إذا تم تقدير حجم التأثير باستخدام المؤشرات الأربعة لنفس الاختبار هل يصل الباحث إلى نفس التفسير؟ وكانت الإجابة لا فقد يختلف التفسير من مؤشر إلى آخر ويمكن تفسير ذلك لأن فلسفة كل مؤشر مختلفة عن الآخر ومعايير المؤشر مختلفة أيضاً، فمعايير مؤشر d مختلفة عن مؤشر r ، وكذلك عن مؤشري γ^2 ، w^2 . وتم التوصل إلى معايير غير تلك التي اقترحها (Cohen 1988) وهذه للمعايير لكل مؤشر تتفق جزئياً مع ما أشار إليه (Lipsey & Wilson, 2001; Mc-Grath & Meyer, 2006) فيما يخص مؤشر d ، وتتفق مع (Kirk 1996) فيما يخص مؤشر w^2 .

ولكن قد يسأل البعض ما المؤشر الأفضل للاستخدام في حالة اختبار "ت" المستقلة؟ الدراسة لم تستطع الوصول إلى اجابة قاطعة ولكن الباحثين حاولوا الإجابة عن هذا التساؤل حيث قام بعمل انحدار متعدد لكل مؤشر وذلك باعتبار أن مؤشر d متغير تابع والمؤشرات الثلاثة الأخرى مستقلة ثم جعل مؤشر r متغير تابع والمؤشرات الأخرى مستقلة وهكذا، واتضح أن مؤشر w^2 أفضل منبأ للمؤشرات الثلاثة (من حيث قيمة معامل الانحدار المعياري) يليه مؤشر γ^2 ، وهذا يتفق مع تشبعات المؤشرات الأربعة علي العامل العام في التحليل العاملي حيث كان أكبر تشبع لمؤشر w^2 يليه γ^2 و d و r علي التوالي.

وإذا اعتمد الباحثون علي مؤشرات مختلفة لحجم التأثير لنفس الاختبار فإن عليهم الحذر عند تفسير هذه المؤشرات لأنها قد تؤدي إلى اختلاف في التفسير، كما يجب أن تؤخذ المعايير المعدلة التي توصلت إليها الدراسة بعين الاعتبار لأنها تؤدي إلى الوصول إلى نفس التفسير وذلك في حالة اختبار "ت" المستقلة.

نحو معيار موحد لتقدير حجم الأثير لاختبار 'ت' لمينتين مستقلتين

ويمكن طرح دراسات مستقبلية مثيلة لهذه الدراسة عبر اختبارات أخرى مثل تحليل التباين الأحادي وت المرتبطة واختبار مربع كاي وغيرها. وكذلك معرفة ما هي التوزيعات العينية الاحصائية لمؤشرات حجم التأثير حتى يمكن الانتقال بها من وضعها الحالي المرتبط بالوصف فقط إلى الاستدلال الاحصائي حتى يمكن اجراء اختبارات فروض حولها كشرط ضروري وليس كافياً للتعميم.

المراجع:

١. الصياد، عبد العاطي أحمد (١٩٨٨). الدلالة العملية وحجم العينة المصاحبة للدلالة الاحصائية لاختبار (ت) في البحث التربوي والنفسي العربي. بحوث مؤتمر البحث التربوي الواقع والمستقبل ، رابطة التربية الحديثة بالقاهرة ، ٢، ١٩٧-٢٣٣.
٢. جرادات ، ضرار وجوده، ماجدة (٢٠٠٥). قوة الاختبار الاحصائي وحجم التأثير وحجم العينة للدراسات المنشورة في مجلة أبحاث اليرموك. سلسلة العلوم الانسانية والاجتماعية. المجلة الأردنية في العلوم التربوية، ١، ٢١-٢٩.
٣. عامر، عبد الناصر السيد. (٢٠١٢). النماذج والاختبارات الاحصائية المستخدمة في البحث النفسي المصري والعربي. المجلة المصرية للدراسات النفسية، ٢٢، ٣٥٦-٣٧٠.
4. American Educational research Association. (2006). Standards on reporting an empirical social science research in AERA publications. *Educational Researcher*, 35, 33- 40.
5. American psychological Association. (1994). Publication manual of the American Psychological Association (5th. Ed.). Washington, DC: Author.
6. Coe, R. (2004). Issues arising from the use of effect sizes in analyzing and reporting research. In I. Schagen, & K. Elliot. (Eds.,). But what does it means? The use of effect sizes in educational research. (p.p. 80- 100). Institute of

Education: University of London.

7. Cohen, J. (1969). Statistical power analysis for the behavioral sciences. London: Academic press.
8. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd. Ed.). Hillsdale, NJ: Erlbaum.
9. Cohen, J. (1994). The Earth is round ($p < 0.05$). *American Psychologist*, 49, 997- 1003.
10. Cumming, G. (2008). Replication and p intervals: p values predict the futures only vaguely, but confidence intervals do much better. *Perspectives on Psychological science*, 3, 286- 300.
11. Glass, G. V. (1976). Primary, Secondary, and meta analysis of research. *Educational Researcher*, 5, 3- 8.
12. Hayes, W. L. (1988). Statistics (4th. Ed.). New York: Holt, Rinehart, & Winston.
13. heges, L. V. (1981). Distribution theory of Glass 's estimator of effect size and related estimators. *Journal of Educational Behavioral Statistics*, 6, 107- 128.
14. Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746- 759.
15. kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *J. AM. ACAD Child Adolescent. Psychiatry*, 42, 1524- 1529.
16. Lipsey, M. W., & Wilson, D. B. (2001). Practical Meta-analysis. Thousand Oaks CA: Sage.

17. Mc-Grath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: the case of r and d . *Psychological Methods*, 11, 386-401.
18. Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, 5, 241-301.
19. Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: applicants, interpretations, and limitations. *Contemporary Educational Psychology*, 25, 241-286.
20. Richardson, J. T. E. (2011). Eta square and practical Eta squared as measures of effect size in educational research. *Educational Research Review*, 6, 135-147.
21. Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). Contrasts and effect sizes in behavioral research: a correlational approach. Cambridge university press.
22. Thompson, B. (1998). Statistical significance and effect size reporting: portrait of possible future. *Research in Schools*, 5, 33-38.
23. Wetzel, S. R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in environmental psychology: an empirical comparison using 8551 tests. *Perspectives on psychological science*, 6, 291-298.

Toward a New Benchmark to estimate effect size for Independent Sample "t" test in Arabic Educational Psychological Research

By:

Prof. Abdel- Atti Ahmed El-Sayyad

Dr. Abdel-Nasser El-Sayed Amer

Abstract:

Aim: This study aimed to Demonstrate some effect size indices used in independent sample "t" test as d , r , w^2 , and γ^2 and to investigate, Are the previous indices have the same Interpretation?, and Provide the new benchmark of each indices to have the same Interpretation for each index.

Method: the sample included (27) psychological study in Egyptian and Arabic Journals. The studies included (144) "t" test for included descriptive statistics, Pearson correlation, and factor analysis, have been used in this study.

Result: There are discrepancies among effect size indices regarding there interpretation the factorial structure for these indices produce just one general factor, finally the study suggested the new benchmark for the indices to give the same Interpretation.