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, γ^2) ومعرفة هل المؤشرات الأربعة تعطي نفس التفسير؟ وأيضاً محاولة اقتراح معايير لكل مؤشر من المؤشرات الأربعة حتي نصل إلى نفس التفسير.

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

وتم تحليل البيانات باستخدم 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 الصياد [1] د. عبد الناصر السيد عامر [2]

مقدمـــة:

تعرضـــت منهجيـــة اختبـــارات الفـــروض الـــصفرية Null Hypothesis Testing (NHT) لانتقادات شديدة لاعتمادهـا علـي قيمـة الخطـأ مـن النـوع الأول p-Value فـي مقارنتها بمستوي الدلالة الاحصائية الـذي ارتـضاه الباحـث (α) لـصناعة القـرار. وأصـبح استخدام قيمـة p محـل جـدل ونقـاش لـدي البـاحثين (الـصياد، ١٩٨٨؛ Cohen, 1988, 1994; Kirk, 1996; Thompson, 1998). وقـد أشـار Nickerson (2000) إلـى وجود خطأ شائع في تفسير الدلالة الاحصائية وهـو الاعتقـاد الـسائد بـأن القيمـة الـصغيرة لـــ P تعني أن تأثير المعالجة كبير والقيمـة الكبيـرة تعنـي أن تـأثير المعالجـة صـغير، وأوصى الباحثون بأستخدام مؤشـر حجـم التـأثير Effect size كجـزء مكمـل لاختبـارات الفروض الصفرية ، وحجم التأثير لـيس مفهومـاً جديـداً ولكنـه أصـبح أكثـر شـيوعاً مـع استخدام أسلوب ما وراء التحليـل Meta-Analysis (أسـلوب لمراجعـة الدراسـات الـسابقة كمياً) منذ بداية عام ١٩٧٠ وأصبح استخدامه أكثر انتشاراً في الدراسـات النفـسية مـع تأكيـد الجمعيـة النفـسية الأمريكيـة American Psychological Association (APA) (1994) علي استخدامه في الدراسات التي تنشر في المجـلات التابعـة لهـا ، كـذلك جمعيـة

---

[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 الامريكية American Educational Research Association (AERA) (2006).

وأشار Wetzel, Matzke, Lee, Rouder, Iverson, & Wagenmakers (2011) بضرورة أن تدعم اختبارات الفروض الصفرية بمؤشرات أخري 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) ، وعلي 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. وعلي 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 علي 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% واجمالي 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 علي دلالة احصائية ويختبر فروضاً أطلق عليها بالفروض ذات القيمة القرارية المحدودة.

واستقراءً للدراسات النفسية في المجلات 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 ايتا η^2 ، والبعض استخدم مؤشر مربع اوميجا w^2 لهيز (Hayes, 1963) والبعض استخدم معامل الارتباط الثنائي الأصيل (r_{pb}) ، وإن تم تقدير حجم التأثير فإن بعض الباحثين قام بتفسير قيمته والبعض الآخر لم يفسره.

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 دلتا لجلاس (Glass, 1976) ، ومؤشر g لهيدجز Hedges (1981) ، ويعتبر مؤشر كوهين أشهرهم وأكثرهم استخداماً في حساب حجم التأثير وأيضاً توظيفاً في دراسات ماوراء التحليل ، ويتم تقدير مؤشر d بالمعادلة التالية:

$$d = \frac{\overline{x_1} - \overline{x_2}}{\sigma_{pooled}} \longrightarrow (1)$$

حيث X_1, X_2 هما متوسطي المجموعتين كلاً على حدة، σ_p هي الانحراف المعياري المشترك ويتحدد بالمعادلة الأتية:

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

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

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

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

واعتبر Cohen (1988) 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 (r_{pb}) Point–Bisereal Correlation ؛ وهو حالة خاصة من معامل ارتباط بيرسون ، ويستخدم عندما يكون المتغير المستقل تصنيفي بمستويين (0-1) والتابع متصل وتتحدد قيمته كالتالي (Mc-Grath & Meyer, 2006):

$$r_{pb} = \frac{\overline{y_1} - \overline{y_2}}{S_y}\sqrt{p_1 p_2} \longrightarrow (4)$$

حيث $\overline{y_2}$,$\overline{y_1}$ متوسطي درجات المتغير التابع للمجموعتين الأولي والثانية على 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):

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

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 أوميجا 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 γ^2 : وتتحدد قيمته من t كالآتي (Richardson, 2011):

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

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 اختبار ت مثل مؤشرات كوهين ، مربع أوميجا ، ومربع أيتا ، ومربع ابسيلون 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 اجراء دراسات ما وراء التحليل (Cohen, 1988; Lipsey & Wilson, 2001) بينما يرى البعض ضرورة استخدام مؤشر r لمعياريته ومرونته في الاستخدام عبر تصميمات بحثية وتحليلات احصائية مختلفة (Rosenthal, et al., 2000).

ومما سبق يتضح تعدد مؤشرات حجم التأثير المستخدمة في اختبار ت 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 الأتية:

١. هل تختلف القيمة التأثيرية لمؤشرات حجم التأثير الأربعة المستخدمة في اختبار "ت" المستقلة (d, r, w^2, γ^2)؟
٢. ما هى المعايير المقترحة للمؤشرات الأربعة بحيث لا تختلف القيمة التفسيرية من مؤشر لأخر ؟
٣. ما حجم العلاقة الارتباطية الموجبة بين قيم المؤشرات الأربعة ؟
٤. ما هو البناء العاملي الذي ينظم العلاقة بين مؤشرات حجم التأثير الأربعة؟

الهدف:

تهدف الدراسة في اطار منهجها الوصفي إلى:

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

الأهمية:

١. التأكيد علي أهمية تقدير حجم التأثير كإجراء مكمل لاختبارات الفروض الصفرية في حالة اختبار "ت" المستقلة.
٢. توجيه انتباه الباحثين إلى تعدد مؤشرات حجم التأثير المستخدمة في اختبار "ت" المستقلة.
٣. 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 المجلات التربوية بجمهورية مصر العربية بالزقازيق وطنطا والمنيا والمنصورة وقنا وسوهاج والمنوفية والجمعية النفسية للدراسات النفسية وبالدول العربية بالسعودية وقطر وسوريا في الفترة من ١٩٩٤ حتي 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 واقتصر الباحثان على اختبار "ت" المستقلة الدالة احصائياً عند مستوي دلالة ٠,٠٥ أو ٠,٠١ ، وتم استبعاد الاختبارات غير الدالة إحصائياً.

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

٣. **المعالجة الاحصائية:** تمت المعالجة الإحصائية بادخال كود الدراسة والمجلة وتاريخ النشر وعينة الدراسة ومستوى الدلالة الاحصائية وقيم مؤشرات d, r, w^2, γ^2 المقدرة 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 إختبار "ت" المستقلة (d, r, w^2, γ^2)؟**

تم إجراء أمر التكرارات مقروناً بالاحصاء الوصفي لمعرفة الوضع الاحصائي لكل مؤشر علي حدة.

جدول (١)

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

	المتوسط	الوسيط	الانحراف المعياري	أدنى قيمة	أعلى قيمة	المنوال	الالتواء
d	٠,٧٥٩	٠,٥٠٥	٠,٦٠٠	٠,١٩	٣,٦٠	٠,٤٩	٢,١٠
r	٠,٣١٨	٠,٢٥٥	٠,١٨٧	٠,٠٧	٠,٨٧	٠,٢٦	١,١٩
w^2	٠,١٣٣	٠,٠٥٧	٠,١٦٦	٠,٠٠	٠,٧٥	٠,٠١٢	١,٩٩
γ^2	٠,١٤٠	٠,٠٦٧	٠,١٦٧	٠,٠٠٥	٠,٧٥٧	٠,٠٦٢	١,٩٥

والمستقرئ للجدول (١) 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^2أي أن تفسير مؤشر مربع ايتا يختلف عن تفسير مؤشري d, r, w^2 (حجم تأثير متوسط) ، ولكن بالنظر إلى توزيع المؤشرات الأربعة نلاحظ أن توزيعها ملتوي التواءً 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 ايتا γ^2 ، بينما يكون ضعيفاً لمؤشري r, w^2 ومن ثم يتبين أن اختلاف المؤشر يمكن أن يؤدي إلي 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 (١٩٨٨) بأن وسيط مؤشري مربع ايتا ومربع أوميجا هي علي التوالي ٠,٠٦٣ و ٠,٠٥٧ لاختبار "ت" للدراسات في علم النفس للباحثين المحترفين (ما بعد الدكتوراه) ومتوسطهم هو ٠,١٤ و ٠,١٣ ولكن القيمة التفسيرية للباحثين الحاليين تختلف مع ما أشار إليه الصياد (١٩٨٨) حيث فسر حجم التأثير بالمتوسط ولكن الدراسة الحالية تتفق معه بأن حجم التأثير من نوع المتوسط (في ضوء مقياس الوسيط) لمؤشر مربع ايتا γ^2 ، ومن النوع الضعيف لمؤشر مربع أوميجا w^2 (أيضاً في ضوء مقياس الوسيط).

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

وعلي 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 علي التساؤل الثاني: ما المعايير المقترحة للمؤشرات الأربعة بحيث لا تختلف القيمة التفسيرية من مؤشر لأخر ؟**

في ضوء هذا الاختلاف في التفسير بين المؤشرات الأربعة فإن وضع حدود جامدة لمعايير Benchmarks لمؤشر ما غير مقبولة فمثلاً وضع Cohen (1988) معاييراً تختلف عن معايير Lipsey & Wilson (2001) 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	٠,٣٠١	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	٠,٦٧
٦٠	٠,٦٦٠	٠,٣٠٠	٠,٠٨٢	٠,٩٥٠
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	٣,٦٠٠	٠,٨٧٠	٠,٧٥٤	٠,٧٥٧

ويمكن تبسيط الجدول السابق في جدول (٣) وملاحظ 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 ٠,٣٠ - ٠,٥٠ ، أما في الدراسة فهما ٠,٢٥٥ و ٠,٣٧ والدراسة الحالية تتفق مع (Lipsey & Wilson, 2001; Mc-Grath & Meyer, 2006) وهي ٠,٠١ (صغير)، ٠,٢٤٠ (متوسط)، ٠,٣٧ (كبير).

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

	d			r			w^2			γ^2		
	صغير	متوسط	كبير	صغير	متوسط	كبير	صغير	متوسط	كبير	صغير	متوسط	كبير
Cohen (1988)	٠,٢٠	٠,٥٠	٠,٨٠	٠,١٠	٠,٣٠	٠,٥٠	٠,٠١	٠,٠٦	٠,١٤	٠,٠١	٠,٠٦	٠,١٤
الدراسة الحالية	٠,٢٠	٠,٥٠٥	٠,٨٢	٠,٠٨٣	٠,٢٥٥	٠,٣٧٠	٠,٠١	٠,٠٥٧	٠,١٣٧	٠,٠٠٦	٠,٠٦٧	٠,١٤٦

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 ٠,٠٥٧ ولمؤشر γ^2 هو ٠,٠٦٧ وحجم تأثير كبير لمؤشر w^2 ٠,١٣٧ ولمؤشر γ^2 ٠,١٤٦ وهذا يتفق جزئياً مع معايير Kirk (1996) لمؤشر w^2 وهي ٠,٠١ (صغير)، ٠,٠٥٩ (متوسط)، ٠,١٣٨ (كبير) وهذا لأن مؤشر γ^2 هو تقدير عالي لحجم التأثير Over estimated على أساس أن مؤشر w^2 هو مؤشر استدلالي وليس وصفي.

علي ذلك فالاعتماد علي معايير الدراسة يؤدي للوصول إلي تفسيرات واحدة لكل المؤشرات سواء d أو r أو w^2 أو γ^2 في حين أن الاعتماد علي 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.

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

(*) دالة احصائياً عند ٠,٠١

ويظهر من الجدول السابق وجود علاقة ارتباطية موجبة عالية جداً بين المؤشرات الأربعة وكانت اعلاهم بين γ^2 ,w^2 وهذا طبيعي لأنهم مؤشرين من نفس العائلة لتحديد نسبة التباين المفسر وكانت أقلهم بين d ,r وهذا طبيعي لأنهم من عائلتين مختلفتين فاحدهما 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 للقول أن أحد المؤشرات يمكن استخدامه محل الآخر.

المناقشة والتعليق:

أصبح الاعتماد علي الدلالة الاحصائية فقط لاتخاذ قرار غير كافي ولابد أن يصاحبة دلائل آخري 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 ω^2, γ^2، والدراسة هدفت إلى تحديد ما إذا تم تقدير حجم التأثير باستخدام المؤشرات الأربعة لنفس الاختبار هل يصل الباحث إلي نفس التفسير؟ وكانت الاجابة لا فقد يختلف التفسير من مؤشر إلي 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 ω^2, γ^2. وتم التوصل إلى معايير غير تلك التي اقترحها Cohen (1988) وهذه المعايير 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 ω^2.

ولكن قد يسال البعض ما المؤشر الأفضل للاستخدام في حالة اختبار "ت" المستقلة؟ الدراسة لم تستطع الوصول إلى اجابة قاطعة ولكن الباحثين حاولا الإجابة عن هذا التساؤل حيث قام بعمل انحدار متعدد لكل مؤشر وذلك باعتبار أن مؤشر d متغير تابع والمؤشرات الثلاثة الأخرى مستقلة ثم يجعل مؤشر r متغير تابع والمؤشرات الأخري مستقلة وهكذا، واتضح أن مؤشر ω^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 ω^2 يليه γ^2 و d و r علي التوالي.

وإذا اعتمد الباحثون علي مؤشرات مختلفة لحجم التأثير لنفس الإختبار 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.

ويمكن طرح دراسات مستقبلية مثيلة لهذه الدراسة عبر اختبارات أخري مثل تحليل التباين الأحادي و"ت" 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 حتي 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 وجودة، 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.