# دراسة تقويميه لاستخدامات تحليل التباين أحادي الاتجاه واختبار "ت" في الدراسات النفسية والتربوية في ضوء حجم العينة وقوة الاختبار وحجم التأثير

**د. ربيع عبده رشوان**
قسم علم النفس التربوي
كلية التربية-جامعة القصيم - المملكة العربية السعودية
rrshoan@qu.edu.sa

http: //dx.doi.org/10.12785/jeps/210404



# دراسة تقويميه لاستخدامات تحليل التباين أحادي الاتجاه واختبار "ت" في الدراسات النفسية والتربوية في ضوء حجم العينة وقوة الاختبار وحجم التأثير

**د. ربيع عبده رشوان**
قسم علم النفس التربوي
كلية التربية-جامعة القصيم - المملكة العربية السعودية

## الملخص

هدف البحث الحالي إلى لتعرف على واقع استخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه وذلك في ضوء مفاهيم قوة الاختبار وحجم التأثير وفقاً لحجم العينة، وتم تحليل عينة من البحوث التي تم نشرها في مجلة البحوث التربوية والنفسية التي تصدر عن جامعة البحرين في الفترة من ٢٠٠٨ إلى ٢٠١٨ والتي تم فيها استخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه بلغ قوامها ٩٤ بحثاً، وتوصل البحث في بعض نتائجه إلى أن قوة الاختبار الإحصائي تزداد بزيادة حجم العينة بينما لا يختلف حجم التأثير باختلاف حجم العينة، وأن نسبة كبيرة من البحوث المنشورة كانت حجوم التأثير فيها ضعيفة وضعيفة جداً، وأن نسبة الخطأ من النوع الأول في هذه البحوث تقريباً ٢١,٤٪ بينما نسبة الخطأ من النوع الثاني فيها تقريباً ٢١,٧٪، وتم تقديم مجموعة التوصيات والمقترحات في ضوء نتائج البحث.

**الكلمات المفتاحية:** الإحصاء الاستدلالي، اختبار "ت"، تحليل التباين، مستوى الدلالة، حجم التأثير، قوة الاختبار، حجم العينة.

* تاريخ تسلم البحث: ٢٠٢٠/٢/٢٠م
* تاريخ قبوله للنشر: ٢٠٢٠/٩/٢م

# An Evaluation Study of the Uses of One-Way ANOVA and "T-Test" in Psychological and Educational Studies in Light of Sample Size, Test Power and Effect Size

**Dr. Rabie A. Rashwan**
Faculty of Education
Qassim University (KSA)

## Abstract

The aim of the current research is to get acquainted with the reality of using the “T” test and the one-Way ANOVA test as inferential parametric statistical tests considering the concepts of test power and the effect size of according to the sample size.

The sample of research published in the Journal of Psychological and Educational Research issued by the Bahrain University in the period from 2008 to 2018, in which the “T” test and the one-way ANOVA were used, which reached 94 research.

The research concluded in some of its results that the power of the statistical test increases with the increase of the sample size, while the effect size does not differ according to the sample size, and that the effect sizes were weak and very weak, and the rate of the first type error in this research was approximately 21.4%, while the rate of the second type error contained approximately 21.7%.

**Keywords:** Inferential statistics, "T" test, ANOVA, Significance level, Effect size, Test power, Sample size.

# دراسة تقويميه لاستخدامات تحليل التباين أحادي الاتجاه واختبار "ت" في الدراسات النفسية والتربوية في ضوء حجم العينة وقوة الاختبار وحجم التأثير

**د. ربيع عبده رشوان**
قسم علم النفس التربوي
كلية التربية-جامعة القصيم - المملكة العربية السعودية

## المقدمة

تعد الاختبارات الإحصائية الاستدلالية البارامترية التي تستخدم في اختبار صحة الفروض الفارقة (تساوي متوسطات المجتمعات) في متغيرات تابعة معينة من أكثر الاختبارات الإحصائية شيوعاً في الدراسات التربوية والنفسية، ومن أهمها اختبار «ت» بحالاته المختلفة واختبار تحليل التباين بصوره المختلفة حسب التصميم البحثي؛ ولهذه الاختبارات البارامترية افتراضات قائمة عليها ويجب الالتزام بها وفي حالة الحيود عن تلك الافتراضات فإن النتائج التي يتم التوصل إليها نتائج غير صادقة وقيمتها من الناحية العملية ضعيفة، فالهدف من التحليلات الإحصائية الاستدلالية بصفة عامة هو الوصول لمدى صحة افتراضات حول الظاهرة في المجتمع من خلال عينات مشتقة من هذا المجتمع، وفي حالة عدم تمثيل العينة للمجتمع وعدم مراعاة افتراضات الاختبارات الإحصائية المستخدمة تكون النتائج التي نصل إليها مضللة في غالبية الأحيان.

ومن أهم الافتراضات التي يجب مراعاتها عند استخدام اختبار «ت» واختبار تحليل التباين تجانس تباين المجموعات المستقلة واعتدالية توزيع الدرجات في كل مجموعة من المجموعات موضوع المقارنة؛ وكذلك مدى مناسبة حجم العينة للاختبار الإحصائي المستخدم في ضوء حجم التأثير المتوقع، وهو ما يفرض على الباحثين في مجال الدراسات التربوية والنفسية الدقة في التحقق من مدى مناسبة الأسلوب الإحصائي الذي سيتم استخدامه، وفي حالة عدم مراعاة تلك الافتراضات وعدم مناسبة حجم العينة وبالتالي عدم دقة التصميم الإحصائي للبحث فإن الاختبار الإحصائي المستخدم لا يتمتع بالقوة الكافية وتكون النتائج غير دقيقة بدرجة كبيرة حيث تزداد احتمالية الخطأ الإحصائي من النوع الأول أو من النوع الثاني (Rasmussen, 1983; Yusof, Abdullah & Yahaya, 2012; Blanca, Alarcón, Arnau, Bono & Bendayan, 2017).

فكما هو معروف هناك نوعان من الأخطاء الإحصائية من الممكن الوقوع فيها عند استخدام الأساليب الإحصائية الاستدلالية، هما الخطأ من النوع الأول Type One Error ويعبر عن احتمالية رفض الفرض الصفري وهو صحيح والذي يعبر عنه بمستوى الدلالة ويرمز له بالرمز ($\alpha$)، بينما يعبر الخطأ من النوع الثاني Type Two Error عن احتمالية قبول الفرض الصفري وهو خاطئ ويرمز له بالرمز ($\beta$)، ويرتبط الخطأ من النوع الأول بمفهوم حجم التأثير بينما يرتبط الخطأ من النوع الثاني بقوة الاختبار الإحصائي.

ويمكن اللجوء للاختبارات الإحصائية اللابارامترية والي تتمتع بالمنعة الكافية مقارنة بالاختبارات الإحصائية البارامترية في حالة انتهاك الافتراضات القائمة عليها تلك الاختبارات (تعني منعة أو ضلاعة الاختبار الإحصائي عدم تأثر نتائجه «تقديراته» بانتهاك الافتراضات القائم عليها) والتي منها اختبار مان وتني Mann-Whitney واختبار ويلككسون Wilcoxon واختبار والش Welch واختبار كروسكال واليز Kruskal-Walles واختبار فريدمان Fredman وهي اختبارات لها قدرة كبيرة على التحكم في الخطأ من النوع الأول ولكنها غير شائعة الاستخدام وأيضا ترتفع فيها احتمالية الخطأ من النوع الثاني (Yusof et al., 2012).

والبحث الحالي يحاول إلقاء الضوء على هذه القضايا والتعرف على واقع استخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه فيما يتعلق بمفاهيم قوة الاختبار وحجم التأثير وفقاً لحجم العينة، ويعرض الباحث فيما يلي للإطار النظري والدراسات السابقة المرتبطة بموضوع البحث.

**أولاً: مستوى الدلالة وحجم التأثير:**

لتجنب الخطأ من النوع الأول يعتمد الباحث على الدلالة الإحصائية والتي عن طريقها يمكن التمييز بين الفروق الحقيقية بين المجتمعات والفروق التي ترجع للصدفة أو أخطاء المعاينة، ولكن الاعتماد على الدلالة الإحصائية قد يكون له نتائج مضللة حيث يُظهر الفرق الكبير الملحوظ على أنه فرق غير حقيقي عندما يكون حجم العينة صغيرًا، بينما يُظهر الفرق الصغير بين مجتمعين على أنه حقيقي ودال إحصائياً عندما يكون حجم العينة كبيرًا، ومن هنا تتضح حقيقة اعتماد مستوى الدلالة الإحصائية على حجم العينة (محمد، ٢٠١٣، ١٠٢)؛ والقرار برفض الفرض الصفري المتمثل في دلالة الفروق في المتغير التابع والناتجة عن اختلاف مستويات المتغير المستقل يتأثر عددًا من العوامل بجانب هذه الفروق ومن هذه العوامل حجم العينة وقوة الاختبار الاحصائي المستخدم ومستوى الدلالة الإحصائية والتشتت في المتغيرات

ذات العلاقة وقوة التأثير ومناسبة القياس الإحصائي والاختبارات الإحصائية المستخدمة وفروض البحث وتعقد التحليلات الإحصائية (باهي، ٢٠١٠، ٤١٨؛ حجمات، عليان، ١٩٩٧، ٢٩٩).

وفي الوقت الحالي هناك نقاشات ودعوات كثيرة إلى ضرورة عدم الاعتماد فقط على مستوى الدلالة في عرض نتائج معالجة الفرض الصفري والاعتماد بجانب ذلك على مقاييس لقوة العلاقة بين المتغيرات فيما يعرف بحجم التأثير وكذلك تدوين بيانات عن فترات الثقة (Pierce, Block & Aguinis, 2004, 917; Greenland et al., 2016, 337).

ويجب على الباحث أن يكون واعياً بأن معنى الدلالة الإحصائية ليس له علاقة بقوة النتائج، فكثيراً ما يعتقد الباحثون بأن مستوى الدلالة ٠,٠١ مثلاً يعبر عن نتائج أكثر أهمية عنها في حالة مستوى الدلالة ٠,٠٥، والحقيقة أن مستوى الدلالة لا يتعدى حد الثقة في النتائج، وأن المتغير المستقل له علاقة بالمتغير التابع، ولا يمتد ذلك لمعرفة قوة هذه العلاقة وهنا تظهر أهمية ما يعرف بحجم التأثير.

فحجم التأثير يعبر عن كمية الفروق بين مجموعتين أو أكثر، أو عن مقدار الفرق في النواتج التي تم تحقيقها بين المجموعتين التجريبية والضابطة، أو عن كم العلاقة بين متغيرين أو أكثر، وظهر مفهوم حجم التأثير من حقيقة أن مستوى الدلالة لا يعبر عن جوهرية الفروق أو العلاقة ومدى أهميتها من الناحية العملية؛ ويعتبر لاكينز Lakens (2013, 1) أن حجم التأثير من أهم النواتج في الدراسات والبحوث الوصفية والتجريبية، وذلك لثلاثة أسباب تتمثل في:

- يُمكن حجم التأثير الباحث من التعبير عن التأثير بطريقة معيارية والذي يمكن أن يفهم في ضوء طبيعة أداة القياس المستخدمة في التعبير عن المتغير التابع، وحجم التأثير المعياري Standardized Effect Size يوضح الأهمية العملية للنتائج أكثر من مجرد الاعتماد على الدلالة الإحصائية.
- يُمكن حساب حجم التأثير المعياري من مقارنة النتائج في دراسات التحليل البعدي.
- عندما تتاح معلومات عن حجم التأثير في الدراسات السابقة يمكن الاعتماد عليها في تخطيط الدراسات المستقبلية وخاصة فيما يتعلق بتحديد حجم العينة في ضوء قوة الاختبار الإحصائي وحجم التأثير ومستوى الدلالة الإحصائية.

ويشير فيدلر وسومبسون Fidler and Thompson (2001, 583) إلى أن حجم التأثير في حالة استخدام اختبار "ت" لفحص دلالة الفروق بين متوسطي مجتمعين يمكن حسابه من خلال حجم التأثير d لكوهين Cohan اعتماداً على متوسطي المجموعتين والانحرافات

المعيارية من العلاقة التالية:

$$d= \frac{M_1-M_2}{SD_p}$$

حيث أن $M_1$، $M_2$ هما متوسطًا المجموعتين بينما $SD_p$ هو الانحراف المعياري المجمع – المتوسط الوزني للانحرافات المعيارية للمجموعتين- أو مباشرة من خلال قيمة t المحسوبة وحجم العينتين $n_1$، $n_2$ في حالة عدم معرفة الانحرافات المعيارية من خلال المعادلة التالية:

أو من خلال معادلة إيتا تربيع التالية:

$$\eta^2= \frac{t^2}{t^2 + d.f}$$

وتعبر *d.f* عن درجات الحرية؛ ويعد مربع إيتا من حيث هو مؤشر لحجم التأثير من أكثر المؤشرات استخداماً خاصة في حالة استخدام تحليل التباين أو اختبار "ت" وذلك لسهولة استخدامه ولتضمينه في نتائج العديد من البرامج الإحصائية، ويؤخذ على هذا المؤشر ($\eta^2$) التحيز إلى الزيادة في تقدير قوة العلاقة بين المتغيرين التابع والمستقل عنها في المجتمع خاصة في حالة صغر حجم العينة، وفي المقابل يعتبر مربع أوميجا ($\omega^2$) ومربع إيبلسون ($\varepsilon^2$) من المقدرات الحصينة وغير المتحيزة لحجم التأثير ويجب أن تستخدم في حالة رغبة الباحث في تقرير حجم التأثير أكثر من مجرد كتابة ملخص وصفي لنتائج الدراسة اعتماداً على ($\eta^2$)؛ وفي حالة استخدام تحليل التباين أحادي الاتجاه يمكن حساب مربع إيتا من خلال المعادلة التالية وهي من المعادلات شائعة الاستخدام (Pierce et al., 2004, 917-918):

$$\eta^2= \frac{SS_{factor}}{SS_{total}}$$

حيث أن البسط $SS_{factor}$ يعبر عن مجموع المربعات بين المجموعات والذي يعبر عن الاختلافات التي ترجع إلى تأثير العامل أو المتغير المستقل بينما $SS_{total}$ تعبر عن مجموع المربعات الكلي، وهذه المعادلة لا تسعمل في حالة استخدام تحليل التباين العاملي حيث يكون هناك أكثر من متغير مستقل ونحتاج إلى استخدام مربع إيتا الجزئي Partial eta-squared

حيث يستبدل مقام المعادلة السابقة بمجموع المربعات الراجع لتأثير العامل مضافاً إليه مجموع مربعات الخطأ ($SS_{factor} + SS_{error}$) وحتى في حالة الاعتماد على قيمة Partial eta-squared التي يتم حسابها باستخدام برنامج SPSS فيجب الحذر لأن البرنامج يستخدم المعادلة الأولى والتي فيها المقام يمثل مجموع المربعات الكلي.

وفي حالة استخدام إيتا تربيع $f^2$ كمؤشر لحجم التأثير فإن حجم التأثير الضعيف هو الذي تقل قيمته عن (٠,٠٦) والمتوسط هو الذي تبدأ قيمته من (٠,٠٦) وتقل عن (٠,١٤) بينما حجم التأثير الكبير هو الذي تبدأ قيمته من (٠,١٤)؛ وقد افترض كوهين Cohen في حالة تحليل التباين مؤشر (f) لحجم التأثير، وعندما تساوي قيمة f (٠,١) يعتبر حجم التأثير ضعيف بينما حجم التأثير المتوسط هو الذي يبدأ من (٠,٢٥) وحجم التأثير الكبير هو الذي يبدأ من (٠,٤٠) (Karadag & Aktas, 2012, 129)؛ ويمكن تحويل إيتا تربيع لحجم التأثير f بواسطة المعادلة التالية (Cohen, 1988, 284):

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

وقد قام كوهين Cohen (1988. 40. 355) بتحديد مقاييس معيارية لحجم التأثير وهو ما يفيد في مقارنة نتائج الدراسات حتى مع اختلاف التدريج ووحدات القياس وتصنف حجوم التأثير في هذه الحالة إلى ضعيف ومتوسط وكبير، فعند مستوى دلالة ٠,٠٥ وقوة الاختبار ٠,٨ فإن حجم التأثير لاختبار "ت" واختبار تحليل التباين يمكن عرضها في الجدول التالي:

**جدول (١)**
**حجم التأثير المعياري لاختبار "ت" واختبار تحليل التباين**
**(مستوى الدلالة ٠,٠٥ وقوة الاختبار ٠,٨)**

| الاختبار المستخدم | حجم التأثير | ضعيفاً جداً | ضعيفاً | متوسطاً | كبيراً |
|---|---|---|---|---|---|
| اختبار «ت» | d | أقل من ٠,٢٠ | ٠,٢٠-٠,٥٠ | ٠,٥٠-٠,٨٠ | ٠,٨٠ فأكثر |
| تحليل التباين أحادي الاتجاه | f | أقل من ٠,١٠ | ٠,١٠-٠,٢٥ | ٠,٢٥-٠,٤٠ | ٠,٤٠ فأكثر |

وعلى الرغم من الاهتمام المتزايد بعرض معلومات عن حجم التأثير وقوة الاختبار الإحصائي كجزء مكمل لنتائج الدراسات والبحوث بصفة عامة وفي مجال العلوم التربوية والنفسية بصفة خاصة إلا أنه مازالت هذه المفاهيم غائبة عن الممارسات الشائعة في كتابة التقارير البحثية وما يتم التوصل إليه من نتائج.

**ثانياً: قوة الاختبار الإحصائي والخطأ من النوع الثاني:**

تعني قوة الاختبار الإحصائي «احتمالية رفض الفرض الصفري الخاطئ» (Lakens, 2013, 1)؛ بمعنى قدرة الاختبار الاحصائي على توفير معلومات يمكن في ضوئها رفض الفرض الصفري عندما يجب أن يرفض، ففي حالة وجود فروق حقيقية بين المجتمعات أو علاقات بين متغيرات معينة في المجتمع فإن قوة الاختبار الإحصائي تتمثل في احتمالية رفض الفرض الصفري المتعلق بدلالة هذه الفروق أو دلالة تلك العلاقة.

وظهر مفهوم قوة الاختبار الإحصائي للتغلب على مشكلة احتمالية الخطأ من النوع الثاني، وهو يعبر عن قدرة الاختبار الإحصائي على الكشف عن التأثيرات الضعيفة غير الصفرية؛ ومفهوم قوة الاختبار الاحصائي مفهوم قديم نسبياً ويرجع إلى ثلاثينيات القرن الماضي لجهود Jerzy Neyman و Egon S. Person ابن Karl Person الإحصائي البريطاني الشهير صاحب معادلة معامل الارتباط وعلى الرغم من أهمية المفهوم وتطبيقاته في التحليلات الإحصائية إلا أنه مازال غير متداولاً ولم يلقى الاهتمام الكافي من قبل الباحثين، وأعاد Jacob Cohen المفهوم إلى دائرة الاهتمام بداية من ستينيات القرن الماضي، عندما قام بمراجعة البحوث والدراسات المنشورة في مجلة علم النفس الاجتماعي وتوصل إلى أن الدراسات المنشورة لم تتسم بالقوة المقبولة لكي تصل إلى الدلالة الإحصائية في حالة حجوم التأثير الواقعية المتوسطة وبالتالي الضعيفة، وعلى الرغم من زيادة الاهتمام بالمفهوم في الكتابات النظرية المهتمة وزيادة وعي الباحثين بأهميته إلا أن ملاحظة Cohen عن ضعف قوة التصميمات البحثية لم تزل قائمة حتى الآن، وهو ما أكدته نتائج العديد من الدراسات التي تناولت تحليلات القوة في البحوث المنشورة في المجلات العلمية حتى الأن والتي تشير نتائجها إلى أن قوة الاختبارات الإحصائية في ظل تصميم تلك الدراسات لم يتعدَّ في أفضل الحالات ٥٧٪، وهو ما يؤكد عدم قدرة هذه الدراسات على رفض الفرض الصفري على الرغم من أن حجم التأثير واقعياً أعلى من المتوسط (زيادة نسبة الخطأ من النوع الثاني)، وبالتالي تكون نتائج دراسة من كل دراستين نتائج غير واقعية وليست ذات أهمية كبيرة من الناحية العملية (Descoteaux, 2007, 28).

وقد تتأثر كذلك قوة الاختبار الإحصائي بالتباين داخل المجموعة مما يعني انتهاك افتراض تجانس التباين، فكلما زاد التباين داخل المجموعات نحتاج إلى عينات أكبر حتى نحقق القوة المقبولة للاختبار الإحصائي (Prajapati, Dunne & Armstrong, 2010, 2).

وعموماً فهناك اتفاق بين المهتمين على أن قوة الاختبارات الإحصائية المقبولة يجب ألا تقل عن ٠,٨٠ بمعنى أن نسبة الخطأ من النوع الثاني يجب ألا تزيد عن ٠,٢٠ ففي هذه الحالة وإذا فرضنا أن مستوى الدلالة سيكون ٠,٠٥ فإن نسبة الخطأ من النوع الثاني إلى الخطأ من النوع الأول تعني أن احتمالية الخطأ من النوع الثاني ستكون أربعة أمثال نسبة الخطأ من النوع الأول وهي نسبة كبيرة، لكن الخطأ من النوع الثاني يعتبر أقل خطورة من الخطأ من النوع الأول، ويتشدد البعض في تقرير قوة الاختبارات الإحصائية ويعتمد على أن نسبة الخطأ من النوع الثاني يجب ألا تتعدي ٠,٠٥ بمعنى أن قوة الاختبار يجب ألا تقل عن ٠,٩٥ (Cunningham & McCrum-Gardner, 2007; Lakens, 2013; Fedler et al., 2012; Batterham & Atkinson, 2005).

**ثالثاً: حجم العينة وتأثيره على حجم التأثير وقوة الاختبار الإحصائي:**

الدقة في نتائج المعالجة الإحصائية للفرض الصفري تتوقف على أربعة عوامل مرتبطة ببعضها بعضًا تتمثل في قوة الاختبار الإحصائي Beta Level ومستوى الدلالة Alpha Level وحجم التأثير Effect Size وحجم العينة Sample Size؛ وأثناء التخطيط للدراسة في مرحلة جمع البيانات وقبل البدء في التحليل الإحصائي، فإننا إذا عرفنا قيم أي ثلاثة من هذه العوامل أو وضعنا تقديرات مبدئية لأي ثلاثة منها نستطيع التقدير المبدئي لقيمة العامل الرابع؛ وبالنسبة لحجم العينة بالتحديد يجب ألا يحدد بطريقة اعتباطية أو بناءً على أحكام غير علمية عشوائية وإنما يحدد في ضوء بعض التوقعات المبدئية عن مستوى الدلالة وقوة الاختبار الإحصائي وحجم التأثير (Algina & Olejnik, 2003, 289; Cunningham & McCrum-Gardner, 2007, 132; Descoteaux, 2007, 29; Karadag & Aktas, 2012, 127).

فهناك اعتقادات خاطئة لدى كثير من الباحثين بأن هناك حدًّا أدنى مناسب لحجم العينة يجب الالتزام به عند استخدام الاختبارات الإحصائية البارامترية، والقاعدة هنا تفيد بأن صغر حجم العينة يؤثر في قوة الاختبار الإحصائي بينما زيادة حجم العينة يزداد معه احتمالية الخطأ من النوع الأول، فقد نصل لمؤشرات مضللة لرفض الفرض الصفري مع أن حجم التأثير ضعيف جداً.

فالمشكلة هنا تكمن في العلاقة العكسية بين مستوى الدلالة وحجم التأثير من ناحية وقوة الاختبار من ناحية أخرى وفقاً لحجم العينة وبمعنى أخر تكمن في العلاقة العكسية بين الخطأ من النوع الأول والخطأ من النوع الثاني وفقاً لحجم العينة (Cunningham & McCrum-Gardner, 2007, 133).

وتتضح أهمية مناسبة حجم العينة وتحليلات القوة الإحصائية من التركيز المتزايد من قبل المجلات العلمية على ضرورة ذكر معلومات عن مدى مناسبة حجم العينة وقوة الاختبارات الإحصائية المستخدمة وذلك حتى تتحقق الشفافية العلمية، فمن ناحية أخلاقيات البحث العلمي هناك دائماً استفسارات عن مدى كفاية حجم عينة الدراسة وقوة الاختبارات الإحصائية المستخدمة، فعدم تصميم الدراسة بحيث تتصف الاختبارات الإحصائية الاستدلالية المستخدمة بالمنعة الكافية وقدرة إجراءاتها على كشف التأثيرات الفعلية نظراً إلى ضعف القوة الإحصائية وعدم مناسبة حجم العينة يعد من مظاهر الإخلال بأخلاقيات البحث العلمي (Pierce et al., 2004, 917; Prajapati et al., 2010, 1; Ahmed, Amin, Aleng & Mohamed, 2012, 51; Descoteaux, 2007, 29)

وهناك العديد من المجلات العلمية الأجنبية في الوقت الحالي تفرض عند النشر ضرورة تضمين معلومات عن حجم التأثير، وكذلك ينص الدليل الإجرائي للجمعية الأمريكية لعلم النفس بداية من الإصدار الرابع صراحة على ضرورة الاهتمام بهذا الجانب عند نشر البحوث والدراسات النفسية حتى يستطيع كل من يرجع للبحث الفهم الصحيح للنتائج التي يتم التوصل إليها (Fidler & Thompson, 2001, 576)؛ وهو ما يؤكد مرة أخرى على ضرورة تحري الشفافية في عرض النتائج؛ وفي هذا الصدد يجب أن تُأخذ هذه الجزئية بعين الاعتبار من قبل مسؤولي النشر بالمجلات العلمية العربية المختصة في التربية وعلم النفس.

وفي ضوء ما سبق لابد من توخي الحذر عند التحديد القبلي لحجم العينة المناسب لكي يتمتع الاختبار الإحصائي المستخدم بالقوة المناسبة لتقليل احتمالية الخطأ من النوع الثاني لأقل حد ممكن، وتزداد التعقيدات في تحديد حجم العينة المناسب وقوة الاختبار الإحصائي في حالة المقارنة بين أكثر من مجموعتين كما هو الحال في تحليل التباين أحادي الاتجاه نظراً لتعدد المقارنات بين المتوسطات بعكس المقارنة بين متوسطين فقط في حالة اختبار "ت" ويزداد الموقف تعقيداً في حالة عدم التجانس والتوزيعات الملتوية (Guo & Luh, 2008)؛ فمعظم التوزيعات الاحتمالية للاختبارات الإحصائية البارامترية (القيم الحرجة لهذه الاختبارات) تم اشتقاقها بافتراض اعتدالية التوزيع وانتهاك هذا الافتراض ينتج عنه التحفظ الشديد للخطأ من النوع الأول والخطأ من النوع الثاني وكنتيجة لذلك تفتقد هذه الاختبارات للكثير من قوتها الإحصائية (Rasmussen, 1983, 37).

وإذا كانت هناك قناعة بضرورة أن تكون العينات كبيرة بدعوى تمثيلها للمجتمع، ومع حقيقة أن زيادة حجم العينة يزيد من قوة الاختبارات الإحصائية المستخدمة؛ لكن الزيادة الكبيرة في حجم العينة غير مرغوب، ففي هذه الحالة سيتم التوصل إلى نتائج ذات دلالة

إحصائية ولكنها غير ذات أهمية علمية أو عملية، وبمعنى آخر تزداد احتمالية الخطأ من النوع الأول (Lenth, 2001, 2; Trajkovski, 2016, 24).

وقد يكون لزيادة حجم العينة قد يكون له تأثيرات عكسية على بعض العوامل الأخرى والتي منها حجم التأثير واحتمالية الخطأ من النوع الأول، وفي المقابل فصغر حجم العينة يكون له تأثيرات عكسية على احتمال الخطأ من النوع الثاني خاصة في حالة صغر حجم التأثير (Prajapati et al., 2010, 2)؛ ويؤكد لاكينز Lakens (2013،1) على أن تأثير حجم العينة على قوة الاختبار الإحصائي يمكن تفسيره في ضوء تأثير حجم العينة على ثبات القياس، وبصورة أكثر توضيحاً، فإن حجم العينة يؤثر على دقة تمثيل العينة للمجتمعات موضوع الدراسة؛ وكما يذكر كلٌ من باترام واتكنسون Batterham and Atkinson (2005, 158) أنه في حالة انخفاض الثبات وخاصة ثبات إعادة التطبيق فإننا نحتاج إلى عينات أكبر في الحجم لنزيد من قوة الاختبار الإحصائي.

والأفكار السابقة هي التي يتم الاستناد إليها في تحديد حجم العينة المناسب؛ فمثلاً في حالة التحليلات الإحصائية القبلية لقوة الاختبار وحجم العينة المناسب، يمكن تحديد حجم العينة المناسب وفقاً لحجم التأثير المتوقع (ضعيف، متوسط، كبير)، وقوة الاختبار الإحصائي والتي ينبغي أن تكون كبيرة ولا تقل عن ٠,٨٠ ومستوى الدلالة الذي سيتم الاعتماد عليه ٠,٠٥ أو ٠,٠١، وفي حالة التحليلات البعدية لقوة الاختبار يمكن بمعلومية القيم الفعلية لحجم التأثير ومستوى الدلالة وحجم العينة الحكم على قوة الاختبار الإحصائي، وبذلك يمكن أن نوضح التالي بشأن بعض الممارسات ذات الأهمية في الدراسات النفسية والتربوية:

- هناك جدل كبير حول تحديد مستوى الدلالة في صياغة الفروض قبل البدء في التحليلات الإحصائية، وما سبق يؤكد على ضرورة تحديد مستوى الدلالة الذي سيتم الاعتماد عليه في رفض الفرض الصفري، لما لذلك من أهمية في التحديد القبلي لحجم العينة في حالة التحليلات القبلية لقوة الاختبار الإحصائي الذي سيتم استخدامه، فلابد من تحديد أدنى قيمة مقبولة للخطأ من النوع الأول.
- في حالة ما إذا أشارت نتائج التحليلات الإحصائية في بحث ما إلى قبول الفرض الصفري، يجب أن يضع الباحث في ذهنه احتمالية أن يكون السبب في ذلك ضعف قوة الاختبار الإحصائي، وهو ما يستوجب تقرير قوة الاختبار الإحصائي في حالة قبول الفرض الصفري.

- يرتبط بالنقطة الأخيرة، أن يضع الباحث في ذهنه أثناء تفسير نتائج دراسته وقبوله للفرض الصفري أنه قد تكون تلك النتائج بسبب زيادة احتمالية الخطأ من النوع الثاني، وليس هناك ما يمنع من الإشارة إلى هذه المعلومة في تفسير النتائج، وهو ما يزيد من قيمة الشفافية العلمية لنتائج البحث العلمي.
- في حالة رفض الفرض الصفري يجب في جميع الأحول تقدير قيمة حجم التأثير حتى تتضح الأهمية العملية لرفض الفرض الصفري وتتضح حقيقة الدلالة الإحصائية في هذه الحالة.

ويذكر كلٌّ من ألجينا وأولجنيك Algina and Olejnik (2003, 289) أن اختيار حجم العينة الذي يُمكن من تحقيق القوة المطلوبة للاختبار الإحصائي يمر بثلاث خطوات تتمثل في:

- تحديد نسبة الخطأ من النوع الأول والحد الأدنى المرغوب تحقيقه من قوة الاختبار الإحصائي في التحليلات الإحصائية المختلفة، حيث نحتاج إلى عينات كبيرة الحجم في حالة تحديد نسبة صغيرة مسموح بها من الخطأ من النوع الأول والرغبة في زيادة قوة الاختبار الإحصائي.
- اعتماداً على التحليلات الإحصائية لقوة الاختبار، يجب تحديد حجم التأثير الجوهري أو حجم التأثير المتوقع في الدراسة كأن يقرر الباحث بأن حجم التأثير الجوهري في الدراسة ٠,٢٠ أو يقرر الباحث أن حجم التأثير المتوقع في الدراسة على الأقل متوسط، فكلما كان حجم التأثير المراد الكشف عنه صغيرًا كلما احتجنا إلى عينات كبيرة الحجم.
- يتم في هذه الخطوة تحديد حجم العينة المناسب بناءً على الشروط السابقة ويستخدم في ذلك العديد من البرامج الخاصة بالتحليلات الإحصائية لقوة الاختبار، أو بالاعتماد على الجداول والأشكال البيانية الموضوعة سابقاً لهذا الغرض.

فالاختبارات الإحصائية البارامترية والتي تتقيد بافتراضات معينة عن شكل التوزيع تتمتع بالقوة الإحصائية الكافية لكنها أقل منعة وذلك في حالة عدم مراعاة الافتراضات القائمة عليها، وفي المقابل الاختبارات الإحصائية اللابارامترية والتي لا تتقيد بافتراضات معينة عن شكل التوزيع وتتميز بالمنعة الكافية لكنها أقل قوة، وعلى الرغم من الاعتقاد الشائع بأن الاختبارات الإحصائية اللابارامترية تناسب العينات صغيرة الحجم إلا أن ضعف قوة بعض هذه الاختبارات يتطلب عينات كبيرة الحجم حتى نتغلب على هذه المشكلة، هذا من جهة ومن جهة أخرى فإنه في حالة البيانات الرتبية على سبيل المثال فإن الاختبارات اللابارامترية لها من القوة ما يمزها عن الاختبارات البارامترية وتزداد قوتها بزيادة حجم العينة (Prajapati et al., 2010, 3; Hecke, 2012, 241-242).

**رابعاً: التحليلات القبلية لقوة الاختبار الإحصائي:**

للتحليلات القبلية لقوة الاختبارات الإحصائية فائدتان على الأقل لها تتمثل الأولى منهما في الاختيار الدقيق لحجم العينة بناءً على أسس عقلانية بدلاً من اختياره بناءً على التوقعات التي قد تنطبق على ظروف الدراسة أو لا، أما الثانية فتتجلَّى في إمكانية تحديد حجم التأثير الجوهري أو المتوقع، وهو ما يتفق مع التأكيدات المتزايدة بضرورة تضمين تقارير البحوث المنشورة لمعلومات عن حجم التأثير وما إذا كان حجم التأثير الذي تم التوصل إليه يتفق مع التوقعات القبلية عن ذلك الحجم أم لا (Algina & Olejnik, 2003, 290).

وبالتالي يمكن تحديد أقل حجم مناسب للعينة الذي يمكننا من تحقيق الحد الأدنى لقوة الاختبار الإحصائي بناءً على تحديد حجم التأثير المتوقع وتحديد الدرجة المقبولة من احتمالية الخطأ من النوع الأول (مستوى الدلالة الذي سيتم الاعتماد عليه في رفض الفرض الصفري)؛ وليست هناك مشكلات في تقدير مستوى الدلالة أو قوة الاختبار الإحصائي المرغوبة لكن المشكلة في تحديد حجم التأثير.

فالتحليلات القبلية الخاصة بقوة الاختبار تتضمن تحديد حجم التأثير بناءً على حجم العينة ومستوى الدلالة الإحصائية الذي سيتم الاعتماد عليه في رفض الفرض الصفري من عدمه (Prajapati et al., 2010, 4)؛ وفي التحليلات القبلية لحجم العينة فإن حجم التأثير يكون غير معلوم ويمكن تحديده أو توقع قيمته بأكثر من طريقة وفي ضوء ما أشار إليه (Cohen, 1988, 276; Cunningham & McCrum-Gardner, 2007, 132, 133; Lenth, 2001, 189) يمكن الاستعانة بما يلي في تحديد حجم التأثير المتوقع في الدراسة:

مراجعة الدراسات السابقة في المجال ولو تضمنت نتائج هذه الدراسات معلومات عن حجم التأثير فإنّه يسهل على الباحث توقع حجم التأثير أو وضع تصورات تقريبية عنه؛ وهنا لا يجب وضع تقديرات مرتفعة وغير واقعية لحجم التأثير لأن ذلك سيقود إلى تخفيض حجم العينة عما هو مطلوب فعلا وبالتالي زيادة احتمالية الخطأ من النوع الثاني؛ وفي المقابل فإن الإقرار بأن حجم التأثير ضعيف سيؤدي إلى كبر حجم العينة المطلوبة وبالتالي زيادة احتمالية الخطأ من النوع الأول.

يمكن الاعتماد على الدراسات الاستطلاعية للتحديد المبدئي لحجم التأثير وهو ما يتطلب خلفية معرفية للباحث عن متغيراته بالإضافة إلى وعي إحصائي لمراعاة الجوانب المؤثرة في النتائج المحتملة للدراسة، وهذه الطريقة مفضلة حسب كثيراً من التوجهات العلمية في هذا الصدد، ويمكن هنا حساب حجم التأثير بناءً على الفرق بين متوسطي المجموعتين مقسوماً على التباين العام للمجموعتين وفي حالة تحليل التباين أحادي الاتجاه تكون المجموعتان هما

أعلى وأدني مجموعة في المتغير التابع.

وهناك العديد من البرامج الإحصائية كبرنامج GPower يمكن استخدامها في تحديد حجم العينة المناسب بمعلومية قوة الاختبار المرغوبة ومستوى الدلالة وحجم التأثير المتوقع، أو تحديد قوة الاختبار عند حجم عينة محدد ومستوى الدلالة وحجم التأثير؛ ويعد برنامج GPower من أفضل البرامج المستخدمة في التحديد المبدئي لحجم العينة وقوة الاختبار الإحصائي وأكثرها شيوعاً في هذا الصدد وذلك لسهولة استخدامه، حيث يمكن تقدير حجم العينة وقوة الاختبار الإحصائي أو الكشف عن مدى مناسبتهما بنوع من السهولة وعدم التعقيدات التقنية ويمكن من خلاله تحديد حجم العينة المناسب قبلياً (أي قبل البدء في التحليل) في ضوء تحديد قوة الاختبار وحجم التأثير ومستوى الدلالة أو تحديد قوة الاختبار الإحصائي بعدياً (أي بعد التحليل وتحديد القيم الفعلية لهذه الثلاث عوامل) بمعلومية حجم العينة وحجم التأثير ومستوى الدلالة (Cunningham & McCrum-Gardner, 2007, 132).

## مشكلة البحث

عند مراجعة نتائج الدراسات السابقة المتعلقة بتأثير متغير معين أو علاقة متغير بمتغيرات أخرى، نلاحظ كثيراً عدم استقرار النتائج وتناقضها في أحيان كثيرة؛ وهذا التناقض بين نتائج الدراسات السابقة قد يكون ناتجًا عن مشكلات في تطبيق أدوات القياس المختلفة، وجدية الاستجابات في بعض الأحيان، أو لعدم مناسبة الأسلوب الإحصائي المستخدم نظراً لعدم تحقق افتراضاته القائم عليها أو لعدم كفاية حجم العينة، والذي يترتب عليه عدم مصداقية النتائج، حيث تزداد احتمالية الوقوع في الخطأ الإحصائي من النوع الأول أو من النوع الثاني.

فمن الأمور التي تكاد تكون مهملة ولا يعي معناها الكثير من الباحثين تأثير حجم العينة على قوة الاختبار الإحصائي المستخدم؛ فقضية تحديد حجم العينة من القضايا الجوهرية والتي عادة ما يكون هناك أخطاء في فهم جوانبها من قبل الباحثين، ذلك من منطلق أن حجم العينة يجب أن يكون كافٍ للوصول إلى نتائج جديرة بالاهتمام، وألا تكون العينة كبيرة بدرجة تمثل عبئًا على الباحث والمشاركين في الدراسة إضافة إلى أن زيادة حجم العينة عن حد معين يكون له تأثير سلبيٌ على الخطأ من النوع الأول وحجم التأثير (Guo & Luh, 2008; Cunningham & McCrum-Gardner, 2007; Karadag & Aktas, 2012; Lenth, 2001)؛ وكذلك عدم الدقة في تقدير الحجم المناسب للعينة يكون له تأثير سلبيٌ على ثبات وصدق النتائج التي يتم التوصل إليها (Batterham & Atkinson, 2005).

ويذكر جرينلاند وآخرون Greenland et al. (2016, 341) أن عدم وجود دلالة إحصائية لا يعبر عن أن حجم التأثير ضعيف، فهذه المعلومة خاطئة خاصة عندما يكون حجم العينة صغيراً، فقد يكون حجم التأثير كبيراً ولا يتمكن الاختبار الإحصائي المستخدم من الكشف عن هذا التأثير، والذي يرجع للعديد من العوامل المؤثرة على دقة النموذج الإحصائي موضع الاختبار في هذه الحالة، وفي المقابل لا تعني الدلالة الإحصائية أهمية النتائج أو أهمية التأثير الدال إحصائياً خاصة عندما تكون حجوم العينات كبيرة، فقد يكون حجم التأثير محدوداً جداً ونصل إلى نتائج دالة إحصائياً ونفس الأمر في حالة انتهاك افتراضات النموذج الإحصائي حيث تزيد احتمالية الخطأ من النوع الأول.

فصغر حجم العينة وعدم دقة التصميم الإحصائي للبحث وكذلك ضعف حجم تأثير المتغيرات المستقلة موضوع الدراسة قد يؤدي إلى قبول الفرض الصفري وهو خاطئ (الخطأ من النوع الثاني)، ونادراً ما نجد أن الباحث في تفسيره لقبول الفروض الصفرية كنتيجة لدراسته يولى الانتباه لضعف قوة الاختبار الإحصائي، وعلى الرغم من أن بعض النتائج تؤكد على منعة اختبار تحليل التباين أحادي الاتجاه واختبار «ت» في حالة انتهاك بعض افتراضات تلك الاختبارات إلا أن هناك بعض النتائج التي تؤكد عكس ذلك.

فقد أكدت نتائج دراسة هيك Hecke (2012) أن قوة اختبار كروسكال واليز تفوق قوة اختبار تحليل التباين أحادي الاتجاه في حالة التوزيعات الملتوية وزيادة حجم العينة، بينما يكاد يكون هناك توافق في قوة الاختبارين في حالة التوزيعات الطبيعية؛ وبعد مراجعة نتائج الدراسات بداية من عام ١٩٣١ والتي تناولت منعة اختبار تحليل التباين أحادي الاتجاه في حالة انتهاك افتراض اعتدالية التوزيع توصلت دراسة بلانكا وآخرين Blanca et al. (2017) إلى أن اختبار تحليل التباين أحادي الاتجاه يتسم بالمنعة في حالة التواء التوزيع ولكن انتهاك افتراض تجانس التباين له تأثير كبير على منعة الاختبار؛ وأكد وانج وآخرون (Wang et al. 2016) على أن الانتهاك البسيط لافتراض تجانس التباين ربما لا يكون له تأثير كبيرٌعلى نتائج اختبار تحليل التباين ولكن الانتهاك الشديد لهذا الافتراض يجعل النتائج غير دقيقة.

## أسئلة البحث

البحث الحالي يهدف إلى التعرف على واقع استخدام اختبار «ت» واختبار تحليل التباين أحادي الاتجاه فيما يتعلق بمفاهيم قوة الاختبار وحجم التأثير وفقاً لحجم العينة، وتتحدد مشكلة البحث في الأسئلة التالية:

١- ما واقع قوة الاختبارات الإحصائية (اختبار «ت»، اختبار تحليل التباين أحادي الاتجاه) المستخدمة في تحليل بيانات البحوث التربوية والنفسية في ضوء حجم العينة؟.
٢- ما واقع حجوم التأثير الناتجة في تحليل بيانات البحوث التربوية والنفسية باستخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه في ضوء حجم العينة؟.
٣- ما واقع نتائج تحليل بيانات البحوث التربوية والنفسية (قبول أو رفض الفرض الصفري) في ضوء حجم التأثير وقوة الاختبار الإحصائي؟.
٤- ما حجم العينة المناسب لاختبار دلالة الفروق بين المتوسطات في الدراسات التربوية والنفسية في ضوء حجم التأثير وقوة الاختبار الإحصائي؟.

## أهداف البحث

يهدف البحث الحالي للتعرف على حجم العينة المناسب لاستخدام اختبار تحليل التباين أحادي الاتجاه واختبار «ت» في الدراسات التربوية والنفسية وفقاً لمفهوم قوة الاختبار وحجم التأثير وبصفة خاصة يهدف البحث الحالي إلى التعرف على:

- واقع قوة الاختبارات الإحصائية (اختبار «ت»، اختبار تحليل التباين أحادي الاتجاه) المستخدمة في تحليل بيانات البحوث التربوية والنفسية في ضوء حجم العينة.
- واقع حجوم التأثير الناتجة في تحليل بيانات البحوث التربوية والنفسية باستخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه في ضوء حجم العينة.
- واقع نتائج تحليل بيانات البحوث التربوية والنفسية (قبول أو رفض الفرض الصفري) في ضوء حجم التأثير وقوة الاختبار الإحصائي.
- حجم العينة المناسب لاختبار دلالة الفروق بين المتوسطات في الدراسات التربوية والنفسية في ضوء حجم التأثير وقوة الاختبار الإحصائي.

## أهمية البحث

تتمثل أهمية البحث الحالي في إلقاء مزيد من الضوء على قضايا ذات أهمية في مجال التحليلات الإحصائية المعتمدة على الأساليب الإحصائية الاستدلالية البارامترية وبصفة خاصة اختبار تحليل التباين أحادي الاتجاه واختبار «ت»، فنتائج البحث الحالي قد تسهم في زيادة وعي الباحثين بمفاهيم قوة الاختبار وحجم التأثير وفقاً لحجم العينة، وهو ما يسهم في زيادة دقة النتائج التي يمكن الوصول إليها عند استخدام هذه الاختبارات الإحصائية، خاصة

وأنَّ تلك الاختبارات من أكثر الاختبارات الإحصائية استخداماً في تحليل بيانات البحوث التربوية والنفسية.

## مصطلحات البحث

**اختبار «ت» T-Test:** يعرفه تراجكوفسكي Trajkovski (2016, 17) بأنه أحد الاختبارات الاستدلالية البارامترية وهو يُستخدمُ بهدف الكشف عن دلالة الفروق بين متوسطي مجموعتين مستقلتين أو مرتبطتين أو متوسط درجات مجموعة واحدة وقيمة افتراضية.

**تحليل التباين أحادي الاتجاه One Way ANOVA**: هو أحد الاختبارات الاستدلالية البارامترية ويهدف إلى الكشف عن دلالة الفروق بين متوسطات عدد K(K>2) من المجتمعات ويكون الفرض الصفري في هذه الحالة متمثلاً في: $H0: \mu1= \mu2= \mu3=\ldots \mu k$

فهو الاختبار الإحصائي المناسب لمقارنة أكثر من اثنين من المعالجات أو الكشف عن تأثير متغير مستقل له أكثر من مستويين على متغير تابع، وهنا يتم تقسيم التباين إلى مكوناته الأصلية والمتمثلة في التباين الراجع إلى تأثير المتغير المستقل ويعرف بالتباين بين المجموعات والتباين مجهول المصدر والذي يعرف بالتباين داخل المجموعات فإذا كانت المجموعات موضع المقارنة مشتقة من مجتمعات لها متوسطات مختلفة فإن التباين بين المجموعات سيكون أكبر من التباين داخل المجموعات، وفي حالة دلالة النسبة الفائية (قسمة التباين بين المجموعات على التباين داخل المجموعات) يمكن الإقرار بأن المجموعات لها متوسطات مختلفة، ويجب ملاحظة أنه في حالة استخدام تحليل التباين أحادي الاتجاه للمقارنة بين مجموعتين فقط فإن أداء اختبار تحليل التباين يكافئ أداء اختبار "ت" للمجموعات المستقلة وتكون ($F=t^2$) (Wang et al., 2016, 306; Trajkovski, 2016, 18).

**قوة الاختبار الإحصائي Power of Statistical Test :** وتعني احتمال الكشف عن الدلالة الإحصائية، بمعنى قدرة الاختبار الإحصائي على رفض الفرض الصفري الخاطئ، وبهذا المعني فقوة الاختبار الإحصائي تساوي (١-الخطأ من النوع الثاني) وفي مرحلة التخطيط لاختيار حجم العينة يتم تحديد قيمة معينة يتم قبولها للخطأ من النوع الثاني وبالتالي تتحدد قوة الاختبار الإحصائي حتى نتأكد من قدرة الاختبار على كشف تأثير المتغير المستقل على المتغير التابع في حالة وجود هذا التأثير أو الكشف عن العلاقات ذات الدلالة بين المتغيرات (Batterham & Atkinson, 2005, 153).

**حجم التأثير Effect Size:** يشير كلا من نيكس وبارنيتي Nix and Barnette في ١٩٩٨ إلى أن حجم التأثير يعبر عن درجة وجود الظاهرة في المجتمع، أو درجة الانحراف عن الفرض الصفري، فإذا كان الفرض الصفري خاطئًا لدرجة ما، فحجم التأثير هو قيمة غير صفرية في المجتمع تزداد مع زيادة الانحراف عن الفرض الصفري، وبالتالي إذا كان مستوى الدلالة يوضح ما إذا كانت الفروق ترجع للصدفة أما لا، فإن حجم التأثير يوضح حجم هذه الفروق في حالة الفروض الفارقة أو قوة العلاقة في حالة الفروض الارتباطية (في: محمد، ٢٠١٣، ١٠٣).

## منهجية البحث وإجراءاته

### منهج البحث:

اعتمد الباحث في الإجابة عن أسئلة البحث الحالي على المنهج الوصفي التحليلي والمقارن، حيث تم تحليل محتوى عينة من البحوث والدراسات التربوية والنفسية المنشورة في مجلة العلوم التربوية والنفسية التي تصدر عن جامعة البحرين في الفترة من ٢٠٠٨ إلى ٢٠١٨، وتمت المقارنة بين هذه النتائج في ضوء المتغيرات موضوع الاهتمام في البحث الحالي.

### عينة البحث:

وقع اختيار الباحث على البحوث المنشورة في مجلة العلوم التربوية والنفسية والتي تصدرها جامعة البحرين وذلك لعدة أسباب أهمها:

- مكانة المجلة كمجلة علمية متخصصة ومحكمة في مجال البحوث التربوية والنفسية.
- جميع أعداد المجلة متاحة على موقع الجامعة مما يُسهل على الباحث الوصول للدراسات والبحوث المستهدفة.
- معظم نتائج البحوث المنشورة في المجلة مذكور فيها مستوى الدلالة سواء في حالة رفض الفرض الصفري أو قبوله وهو ما يحتاجه البحث الحالي في التحليلات الخاصة بقوة الاختبار وحجم التأثير.
- تنوع البحوث والدراسات المنشورة في المجلة في مجال الدراسات التربوية والنفسية.

وقام الباحث بتحليل عينة عشوائية من البحوث التي تم فيها استخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه والتي تم نشرها في المجلة في الفترة من ٢٠٠٨ إلى ٢٠١٨ وذلك لتحديد الأسلوب الإحصائي وحجم العينة ونتائج الدراسة وجميع الإحصاءات الناتجة من التحليل الإحصائي والتي سيتم الاعتماد عليها في حساب حجم التأثير وقوة الاختبار

الإحصائي، وبلغ عدد البحوث التي تم تحليلها ٩٤ بحثاً تم فيها استخدام اختبار "ت" واختبار تحليل التباين أحادي الاتجاه، والجدول التالي يوضح أهم الإحصاءات المرتبطة بالعينة في ضوء الأسلوب الإحصائي المستخدم:

**جدول (٢)**
**وصف عينة البحث من البحوث والدراسات التي تم تحليل نتائجها والمنشورة بمجلة العلوم التربوية والنفسية بجامعة البحرين**

| الاختبار الإحصائي | عدد مرات الاستخدام | | رفض الفرض الصفري | | قبول الفرض الصفري | |
|---|---|---|---|---|---|---|
| | التكرار | النسبة | التكرار | النسبة | التكرار | النسبة |
| اختبار «ت» | ٣٦٢ | ٪٥٨,٨ | ١٤٣ | ٪٣٩,٥ | ٢١٩ | ٪٦٠,٥ |
| تحليل التباين الأحادي | ٢٥٤ | ٪٤١,٢ | ١١٧ | ٪٤٦,١ | ١٣٧ | ٪٥٣,٩ |
| العينة الكلية | ٦١٦ | ٪١٠٠ | ٢٦٠ | ٪٤٢,٢ | ٣٥٦ | ٪٥٧,٨ |

## نتائج البحث

### أولاً: نتائج السؤال الأول:

ينص السؤال الأول للبحث الحالي على «ما واقع قوة الاختبارات الإحصائية (اختبار «ت»، اختبار تحليل التباين أحادي الاتجاه) المستخدمة في تحليل بيانات البحوث التربوية والنفسية في ضوء حجم العينة؟.

تم استخدام برنامج GPower في حساب قوة اختبار «ت» وقوة اختبار تحليل التباين أحادي الاتجاه المستخدمة في تحليل بيانات البحوث التربوية والنفسية المنشورة في مجلة البحوث التربوية والنفسية في الفترة من ٢٠٠٨ إلى ٢٠١٨، وذلك للتعرف على واقع قوة اختبارات دلالة الفروق المستخدمة في تلك البحوث فكانت النتائج كما هو موضح في التالي:

**جدول (٣)**
**قوة اختبار «ت» واختبار تحليل التباين من واقع البحوث التربوية والنفسية في ضوء حجم العينة**

| قوة الاختبار | اختبار «ت» | | | | تحليل التباين أحادي الاتجاه | | | |
|---|---|---|---|---|---|---|---|---|
| | تكرار | نسبة | متوسط | انحراف معياري | تكرار | نسبة | متوسط | انحراف معياري |
| أقل من ٠,٥ | ١٢٣ | ٪٣٤,٠٠ | ٠,٣٧٢ | ٠,١٣٤ | ٢١ | ٪٨,٣٠ | ٠,٢٨ | ٠,١٤٦ |
| من ٠,٥ لأقل من ٠,٦ | ٩٧ | ٪٢٦,٨٠ | ٠,٥٢٥ | ٠,٠٢٥ | ٣٧ | ٪١٤,٥٠ | ٠,٥٧٣ | ٠,٠٢٦ |
| من ٠,٦ لأقل من ٠,٧ | ٣٧ | ٪١٠,٢٠ | ٠,٦٤ | ٠,٠٢٦ | ٨٦ | ٪٣٣,٩٠ | ٠,٦٥١ | ٠,٠٢٨ |
| من ٠,٧ لأقل من ٠,٨ | ٣٢ | ٪٨,٨٠ | ٠,٧٣٨ | ٠,٠٢٨ | ٥٩ | ٪٢٣,٢٠ | ٠,٧٥٣ | ٠,٠٢٦ |

**تابع جدول (٣)**

| قوة الاختبار | اختبار «ت» | | | | تحليل التباين أحادي الاتجاه | | | |
|---|---|---|---|---|---|---|---|---|
| | تكرار | نسبة | متوسط | انحراف معياري | تكرار | نسبة | متوسط | انحراف معياري |
| من ٠,٨ لأقل من ٠,٩ | ١٧ | ٤,٧٠٪ | ٠,٨٤٩ | ٠,٠٢٩ | ٣٤ | ١٣,٤٠٪ | ٠,٨٤ | ٠,٠٢٧ |
| من ٠,٩ فأكثر | ٥٦ | ١٥,٥٠٪ | ٠,٩٧٣ | ٠,٠٣٣ | ١٧ | ٦,٧٠٪ | ٠,٩٥٩ | ٠,٠٣٤ |
| العينة الكلية | ٣٦٢ | ١٠٠٪ | ٠,٥٨٨ | ٠,٢٢٨ | ٢٥٤ | ١٠٠٪ | ٠,٦٧٩ | ٠,١٦٦ |

وكما هو ثابت في غالبية الدراسات السابقة أن قوة الاختبار المقبولة يجب ألا تقل عن ٠,٨ نجد أنه:

في ٧٩,٨٪ من الدراسات التربوية والنفسية كان استخدام اختبار «ت» لا يتصف بالقوة المقبولة، وإذا ما اعتمدنا على أن قوة الاختبار يجب ألا تقل عن ٠,٧٠ كما في بعض الدراسات نلاحظ أن ٧١٪ من الدراسات التي استخدمت اختبار «ت» لم يتصف فيها الاختبار بالقوة المقبولة، ولم يختلف الحال كثيراً في حالة اختبار تحليل التباين أحادي الاتجاه حيث بلغت نسبة الدراسات التي لم تصل فيها قوة الاختبار إلى ٠,٨ حوالي ٧٩,٩٪، بينما في حالة الاعتماد على ٠,٧٠ كمحك للحكم على قوة الاختبار نجد أن ما نسبته ٥٦,٧٪ من الدراسات لم يتصف فيها الاختبار الإحصائي بالقوة المقبولة.

وهذه النتيجة تجعلنا نتشك كثيراً في نتائج الدراسات التي لم تستطع فيها الاختبارات الإحصائية المستخدمة في الكشف عن دلالة الفروق في المتغيرات موضوع الدراسة، ففي حالة اختبار «ت» نجد أن ثلاثًا من كل أربع دراسات تقريباً يحتمل وقوعها في الخطأ من النوع الثاني (قبول الفرض الصفري وهو خاطئ)، بينما في حالة تحليل التباين أحادي الاتجاه نجد أن هناك دراسة من كل دراستين يحتمل وقوعها في الخطأ من النوع الثاني.

وللتعرف على تأثير حجم العينة على قوة الاختبار الإحصائي تم استخدام تحليل التباين أحادي الاتجاه في التعرف على دلالة الفروق في قوة اختبار «ت» وفي قوة اختبار تحليل التباين أحادي الاتجاه والتي ترجع لاختلاف حجم العينة (أقل من ١٠٠، من ١٠٠ لأقل من ٢٠٠، من ٢٠٠ فأكثر) فكانت النتائج كما هي موضحة في التالي:

**جدول (٤)**
**المتوسطات والانحرافات المعيارية لقوة الاختبار باختلاف حجم العينة**

| الاختبار | اختبار «ت» | | | تحليل التباين أحادي الاتجاه | | |
|---|---|---|---|---|---|---|
| حجم العينة | العدد | المتوسط | الانحراف المعياري | العدد | المتوسط | الانحراف المعياري |
| أقل من ١٠٠ | ١٠٥ | ٠,٤٦٤ | ٠,٢٠٣ | ٥٨ | ٠,٦٣٤ | ٠,٢١٥ |
| من ١٠٠ لأقل من ٢٠٠ | ١٠٨ | ٠,٥٩ | ٠,٢١١ | ٩٠ | ٠,٦٦٢ | ٠,١٣٢ |
| من ٢٠٠ فأكثر | ١٤٩ | ٠,٦٧٤ | ٠,٢١٧ | ١٠٦ | ٠,٧١٨ | ٠,١٥٤ |

**جدول (٥)**
**دلالة الفروق في قوة الاختبار الإحصائي والتي ترجع لاختلاف حجم العينة**

| الاختبار الإحصائي | مصدر التباين | مجموع المربعات | درجة الحرية | متوسط المربعات | قيمة «ف» | مستوى الدلالة |
|---|---|---|---|---|---|---|
| اختبار «ت» | بين المجموعات | ٢,٧١٢ | ٢ | ١,٣٥٦ | ٣٠,٣٠٧ | ٠,٠١ |
| | داخل المجموعات | ١٦,٠٦١ | ٣٥٩ | ٠,٠٤٥ | | |
| | الكلي | ١٨,٧٧٣ | ٣٦١ | | | |
| اختبار تحليل التباين أحادي الاتجاه | بين المجموعات | ٠,٣٠٧ | ٢ | ٠,١٥٤ | ٥,٧٨٥ | ٠,٠١ |
| | داخل المجموعات | ٦,٦٦ | ٢٥١ | ٠,٠٢٧ | | |
| | الكلي | ٦,٩٦٧ | ٢٥٣ | | | |

ويتضح من جدول (٥) السابق أنه توجد فروق دالة إحصائياً عند مستوى ثقة ٠,٠١ في قوة الاختبار الإحصائي (اختبار «ت»، أو اختبار تحليل التباين أحادي الاتجاه) ترجع لاختلاف حجم العينة، وتم استخدام اختبار أقل فرق دال LSD كاختبار للمقارنات البعدية فكانت النتائج كما هي موضحة في الجدول التالي:

**جدول (٦)**
**المقارنات البعدية لقوة الاختبار الإحصائي للعينات مختلفة الحجم**

| | حجم العينة | أقل من ١٠٠ (م= ٠,٤٦٤) | من ١٠٠ لأقل من ٢٠٠ (م= ٠,٥٩٠) |
|---|---|---|---|
| اختبار «ت» | من ١٠٠ لأقل من ٢٠٠ (م= ٠,٥٩٠) | ٠,١٢٦** | |
| | من ٢٠٠ فأكثر (م= ٠,٦٧٤) | ٠,٢١٠** | ٠,٠٨٤** |
| تحليل التباين أحادي الاتجاه | حجم العينة | أقل من ١٠٠ (م= ٠,٦٣٤) | من ١٠٠ لأقل من ٢٠٠ (م= ٠,٦٦٢) |
| | من ١٠٠ لأقل من ٢٠٠ (م= ٠,٦٦٢) | ٠,٠٢٨ | |
| | من ٢٠٠ فأكثر (م= ٠,٧١٨) | ٠,٠٨٤** | ٠,٠٥٦** |

وجدول (٦) السابق يتأكد منه أن قوة الاختبار الإحصائي تزيد بزيادة حجم العينة، ففي حالة اختبار «ت» كانت هناك فروق دالة إحصائياً عند مستوى ٠,٠١ بين المجموعات المختلفة في حجم العينة ولصالح المجموعات كبيرة الحجم (من ١٠٠ لأقل من ٢٠٠، من ٢٠٠ فأكثر)، وفي حالة تحليل التباين أحادي الاتجاه كانت الفروق دالة إحصائياً عند مستوى ٠,٠١ ولصالح الدراسات التي كان فيها حجم العينة من ٢٠٠ فأكثر.

ويمكن تفسير زيادة قوة الاختبار الإحصائي بزيادة حجم العينة، في ضوء ما أشار إليه حجمات وعليان (١٩٩٧، ٤٠١) بأنه كلما زاد حجم العينة قل مقام الاختبار الإحصائي (جميع الاختبارات الإحصائية البارامترية تعتمد على الخطأ المعياري في المقام وبالتالي يكون مقام المعادلة مقسوماً على حجم العينة) م وبالتالي تزيد قوته، وتتفق النتائج الحالية كثيراً مع نتائج هذه الدراسة والتي أشارت إلى أن ٣١,٩٪ من الدراسات التي استخدمت اختبار تحليل التباين لم تصل فيها قوة الاختبار إلى ٠,٦.

وضعف قوة الاختبارات الإحصائية المستخدمة في الكشف عن دلالة الفروق بين المجتمعات في البحث الحالي يتفق مع نتائج دراسة محمد (٢٠١٣) والتي أكدت نتائجها على ضعف قوة الاختبارات الإحصائية المستخدمة في معالجة نتائج البحوث والدراسات التربوية المنشورة في مجلة كلية التربية جامعة المنيا.

**ثانياً: نتائج السؤال الثاني:**

ينص السؤال الثاني للبحث الحالي على «ما واقع حجوم التأثير الناتجة في تحليل بيانات البحوث التربوية والنفسية باستخدام اختبار «ت» واختبار تحليل التباين أحادي الاتجاه في ضوء حجم العينة؟.

تم حساب حجم التأثير d في حالة اختبار «ت» وحجم التأثير f في حالة اختبار تحليل التباين أحادي الاتجاه باستخدام المعادلات المعتمدة في هذا الصدد والسابق العرض لها في الإطار النظري للبحث الحالي، من خلال بيانات البحوث التربوية والنفسية المنشورة في مجلة البحوث التربوية والنفسية في الفترة من ٢٠٠٨ إلى ٢٠١٨، وذلك للتعرف على واقع حجم التأثير للمتغيرات التي تم دراستها في تلك البحوث فكانت النتائج كما هو موضح في التالي:

**جدول (٧)**

**حجوم التأثير (d, f) لاختبار «ت» واختبار تحليل التباين من واقع البحوث النفسية والتربوية**

| حجم التأثير | d في حالة اختبار "ت" | | | | f في حالة تحليل التباين أحادي الاتجاه | | | |
|---|---|---|---|---|---|---|---|---|
| | تكرار | نسبة | متوسط | انحراف معياري | تكرار | نسبة | متوسط | انحراف معياري |
| ضعيف جداً | ١٥٧ | ٪٤٣,٤٠ | ٠,٠٩١ | ٠,٠٥٩ | ٥٩ | ٪٢٣,٢٠ | ٠,٠٥٢ | ٠,٠٢٦ |
| ضعيف | ١٥٠ | ٪٤١,٤٠ | ٠,٣٢٤ | ٠,٠٨٥ | ١١٧ | ٪٤٦,١٠ | ٠,١٦٧ | ٠,٠٣٨ |
| متوسط | ٣٤ | ٪٩,٤٠ | ٠,٦٣٣ | ٠,٠٨٩ | ٥٩ | ٪٢٣,٢٠ | ٠,٣٢ | ٠,٠٤٦ |
| كبير | ٢١ | ٪٥,٨٠ | ١,٣٥٩ | ٠,٥٩٥ | ١٩ | ٪٧,٥٠ | ٠,٤٥٥ | ٠,٠٣٩ |
| العينة الكلية | ٣٦٢ | ٪١٠٠ | ٠,٣١٢ | ٠,٣٤٧ | ٢٥٤ | ٪١٠٠ | ٠,١٩٨ | ٠,١٢٤ |

ومن جدول (٧) السابق يتضح أنه ٨٤,٨٪ من الدراسات التي استخدمت اختبار «ت» كان حجم التأثير فيها ضعيفاً جداً وضعيفاً، و٩,٤٪ من الدراسات كان حجم التأثير فيها متوسطاً، وفي ٥,٨٪ من الدراسات كان حجم التأثير فيها كبيراً، بينما في حالة تحليل التباين أحادي الاتجاه كان ٦٩,٣٪ من الدراسات حجم التأثير فيها ضعيفاً جداً وضعيفاً، بينما كان حجم التأثير متوسطاً في ٢٣,٢٪ من الدراسات وكان حجم التأثير كبيراً في حوالي ٧,٥٪ من الدراسات.

وللتعرف على تأثير حجم العينة على حجم التأثير تم استخدام تحليل التباين أحادي الاتجاه في التعرف على دلالة الفروق في حجوم التأثير في حالة اختبار «ت» وفي حالة اختبار تحليل التباين أحادي الاتجاه والتي ترجع لاختلاف حجم العينة (أقل من ١٠٠، من ١٠٠ لأقل من ٢٠٠، من ٢٠٠ فأكثر) فكانت النتائج كما هي موضحة في التالي:

**جدول (٨)**

**المتوسطات والانحرافات المعيارية لحجم التأثير باختلاف حجم العينة**

| الاختبار الإحصائي | اختبار «ت» | | | تحليل التباين أحادي الاتجاه | | |
|---|---|---|---|---|---|---|
| | العدد | المتوسط | الانحراف المعياري | العدد | المتوسط | الانحراف المعياري |
| أقل من ١٠٠ | ١٠٥ | ٠,٣٤٢ | ٠,٢٠٢ | ٥٨ | ٠,١٨٣ | ٠,١٤٣ |
| من ١٠٠ لأقل من ٢٠٠ | ١٠٨ | ٠,٢٦٢ | ٠,٣١٨ | ٩٠ | ٠,٢١٤ | ٠,١٢١ |
| من ٢٠٠ فأكثر | ١٤٩ | ٠,٣٢٨ | ٠,٤٣٢ | ١٠٦ | ٠,١٩٢ | ٠,١١٣ |

**جدول (٩)**
**دلالة الفروق في حجم التأثير والتي ترجع لاختلاف حجم العينة**

| الاختبار الإحصائي | مصدر التباين | مجموع المربعات | درجة الحرية | متوسط المربعات | قيمة «ف» | مستوى الدلالة |
|---|---|---|---|---|---|---|
| اختبار «ت» | بين المجموعات | ٠,٤٠٨ | ٢ | ٠,٢٠٤ | ١,٧١٣ | ٠,١٨٢ غير دالة |
| | داخل المجموعات | ٤٢,٧٢٦ | ٣٥٩ | ٠,١١٩ | | |
| | الكلي | ٤٣,١٣٤ | | | | |
| اختبار تحليل التباين أحادي الاتجاه | بين المجموعات | ٠,٠٣٩ | ٢ | ٠,٠٢ | ١,٢٧٩ | ٠,٢٨٠ غير دالة |
| | داخل المجموعات | ٣,٨٢٦ | ٢٥١ | ٠,٠١٥ | | |
| | الكلي | ٣,٨٦٥ | ٢٥٣ | | | |

ويتضح من جدول (٩) السابق أنه لا توجد فروق دالة إحصائياً في حجوم التأثير في حالة استخدام اختبار «ت» أو اختبار تحليل التباين أحادي الاتجاه ترجع لاختلاف حجم العينة.

ومجمل ما تم التوصل إليه هنا من نتائج يؤكد أن حجوم التأثير في الدراسات النفسية والتربوية المنشورة في مجلة العلوم التربوية والنفسية في الفترة من ٢٠٠٨ إلى ٢٠١٨ غالبيتها ضعيفة وضعيفة جداً، وقد لاحظ الباحث أثناء فحص نتائج تلك الدراسات أنها خلت تماما من أي معلومات عن حجم التأثير وأن الباحثين يكتفون بالاعتماد على الدلالة الإحصائية في تقرير أهمية الدراسة ونتائجها من الناحية العملية وهي من الممارسات الخاطئة التي قد ترجع لضعف وعي الباحثين بمعنى وأهمية الدلالة العملية للنتائج.

وتتفق هذه النتيجة مع ما توصل إليه باهي (٢٠١٠، ٤٣٨) من أن حجم التأثير لا يرتبط بحجم العينة؛ كذلك تتفق مع نتائج دراسة نصار (٢٠٠٢) والتي أكدت أن حجم التأثير أقل تأثراً بحجم العينة من الأساليب الإحصائية التي تستخدم لفحص الفروض عن طريق الدلالة الإحصائية وكذلك أكدت نتائجها أن نسبة ٦٠٪ من حجوم التأثير للأساليب الإحصائية كانت ضعيفة ومتوسطة، كذلك أكدت نتائج دراسة الحاج (٢٠٠٣) أن حوالي ٧١٪ من النتائج التي تم فيها رفض الفرض الصفري كان فيها حجم التأثير ضعيفاً.

وضعف حجوم التأثير في الدراسات السابقة ويمكن تفسير في ضوء عدم دقة أدوات القياس المستخدمة في هذه الدراسات وعدم تعبير نتائج هذه الأدوات بدقة عن المتغيرات المراد قياسها، كذلك عدم الدقة في اختيار المتغيرات المستقلة التي تم دراسة تأثيرها، وهو ما ينتج عنه أن يكون التباين غير المفسر في المتغيرات التابعة أكبر بكثير من التباين الذي تفسره المتغيرات المستقلة موضوع التركيز.

وتتفق النتائج الحالية بدرجة كبيرة مع نتائج دراسة حجمات وعليان (١٩٩٧) التي أشارت إلى أن ٧٥٪ من الفرضيات البديلة التي تم فيها رفض الفرض الصفري في رسائل ماجستير الإرشاد النفسي والتربوي بالجامعة الأردنية كانت دلالتها العملية (حجم التأثير) ضعيفة؛ كذلك تتفق مع نتائج دراسة الصياد (١٩٨٨) التي أكدت أن حجم التأثير في حالة استخدام اختبار "ت" في البحوث التربوية والنفسية كان ضعيفاً، وأن الباحثين يميلون لزيادة حجم العينة للوصول لدلالة إحصائية.

**ثالثاً: نتائج السؤال الثالث:**

ينص السؤال الثالث للبحث الحالي على «ما واقع نتائج تحليل بيانات البحوث النفسية والتربوية (قبول أو رفض الفرض الصفري) في ضوء حجم التأثير وقوة الاختبار الإحصائي؟».

وللإجابة عن هذا السؤال تم مراجعة نتائج الدراسات والبحوث التربوية والنفسية موضوع البحث الحالي لتحديد ما إذا كان تم قبول أو رفض الفرض الصفري في حالة استخدام اختبار «ت» أو اختبار تحليل التباين أحادي الاتجاه، وكذلك تم تحديد حجم التأثير في كل حالة (ضعيف جداً، ضعيف، متوسط، كبير)، وكذلك تم تحديد قوة الاختبار الإحصائي في كل حالة (أقل من ٠,٨٠، من ٠,٨٠ فأكثر) حتى تتكون لدى الباحث صورة أكثر عمقاً وشمولاً للنتائج السابقة التي تم التوصل إليها في هذه الدراسات فكانت النتائج كما هي موضحة بجدول (١٠):

**جدول (١٠)**
**نتائج تحليل بيانات البحوث التربوية والنفسية في ضوء حجم التأثير وقوة الاختبار**

| الاختبار الإحصائي | النتيجة المتعلقة بالفرض الصفري | حجم التأثير ضعيف جداً | | | | حجم التأثير ضعيفاً | | | | حجم التأثير متوسطاً | | | | حجم التأثير كبيراً | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | قوة الاختبار | | | | قوة الاختبار | | | | قوة الاختبار | | | | قوة الاختبار | | | |
| | | أقل من ٠,٨ | | من ٠,٨ فأكثر | | أقل من ٠,٨ | | من ٠,٨ فأكثر | | أقل من ٠,٨ | | من ٠,٨ فأكثر | | أقل من ٠,٨ | | من ٠,٨ فأكثر | |
| | | ك | ٪ | ك | ٪ | ك | ٪ | ك | ٪ | ك | ٪ | ك | ٪ | ك | ٪ | ك | ٪ |
| اختبار «ت» | قبول | ١٢١ | ٣٣,٤ | ٣١ | ٨,٦ | ٦١ | ١٦,٩ | ٣ | ٠,٨ | ٣ | ٠,٨ | ٠,٠ | ٠,٠ | ٠,٠ | ٠,٠ | ٠,٠ | ٠,٠ |
| | رفض | ٥ | ١,٤ | ٠,٠ | ٠,٠ | ٧٧ | ٢١,٣ | ٩ | ٢,٥ | ٢٠ | ٥,٥ | ١١ | ٠,٠ | ٢ | ٠,٦ | ١٩ | ٥,٢ |
| تحليل التباين أحادي الاتجاه | قبول | ٣٩ | ١٥,٤ | ١٧ | ٦,٧ | ٥٦ | ٢٢ | ٦ | ٢,٤ | ١٢ | ٤,٧ | ٥ | ٢ | ٢ | ٠,٨ | ٠,٠ | ٠,٠ |
| | رفض | ٣ | ١,٢ | ٠,٠ | ٠,٠ | ٤٧ | ١٨,٥ | ٩ | ٣,٥ | ٣١ | ١٢,٢ | ١٠ | ٣,٩ | ١٣ | ٥,١ | ٤ | ١,٦ |
| العينة الكلية | قبول | ١٦٠ | ٢٦ | ٤٨ | ٧,٨ | ١١٧ | ١٩ | ٩ | ١,٥ | ١٥ | ٢,٤ | ٥ | ٠,٨ | ٢ | ٠,٣ | ٠,٠ | ٠,٠ |
| | رفض | ٨ | ١,٣ | ٠,٠ | ٠,٠ | ١٢٤ | ٢٠,١ | ١٨ | ٢,٩ | ٥١ | ٨,٣ | ٢١ | ٣,٤ | ١٥ | ٢,٤ | ٢٣ | ٣,٧ |

وحتى تتضح النتائج ومنعاً للتكرار سيكتفى الباحث بعرض النتائج للعينة الكلية من الدراسات والاختبارات الإحصائية المستخدمة فيها، حيث يتضح من الجدول السابق أنه:

**أ- بالنسبة لقبول الفرض الصفري في حالة العينة الكلية:**

- تم قبول الفرض الصفري في ٣٣,٨٪ من الدراسات وهي الدراسات التي كان فيها حجم التأثير ضعيفاً جداً، وهذه النتيجة تعد نتيجة منطقية ولن تختلف باختلاف قوة الاختبار الإحصائي.
- تم قبول الفرض الصفري في ٢٠,٥٪ من الدراسات وهي التي كان فيها حجم التأثير ضعيفاً، وكانت قوة الاختبار الاحصائي ضعيفة (أقل من ٠,٨) في ١٩٪ من هذه الدراسات.
- تم قبول الفرض الصفري في ٣,٥٪ من الدراسات التي كان فيها حجم التأثير متوسطاً أو كبيراً، وكانت قوة الاختبار الإحصائي ضعيفة (أقل من ٠,٨) في ٢,٧٪ من هذه الدراسات.

وبالتالي تكون نسبة الخطأ من النوع الثاني في هذه الدراسات تساوي تقريباً ٢١,٧٪ وهي النتائج التي تم فيها قبول الفرض الصفري في حالة انخفاض قوة الاختبار الإحصائي المستخدم.

**ب- بالنسبة لرفض الفرض الصفري في حالة العينة الكلية:**

- تم رفض الفرض الصفري في ٢٤,٣٪ من الدراسات التي كان فيها حجم التأثير ضعيفاً أو ضعيفاً جداً، ويلاحظ أن ما نسبته ٢١,٤٪ من هذه الدراسات كانت فيها قوة الاختبار منخفضة.
- تم رفض الفرض الصفري في ١٧,٨٪ من الدراسات وهي الدراسات التي كان فيها حجم التأثير متوسطاً أو كبيراً وهذه النتيجة منطقية خاصة في حالة زيادة قوة الاختبار الإحصائي.

وهنا يلاحظ أن نسبة الخطأ من النوع الأول تقريباً ٢١,٤٪ وهي النتائج التي تم فيها رفض الفرض الصفري على الرغم من أن حجم التأثير كان ضعيفاً أو ضعيفاً جداً خاصة مع ضعف قوة الاختبار الإحصائي مما يجعلنا لا نطمئن لهذه النتيجة، وقد تكون هذه النتائج مضللة ناتجه عن ظروف غير دقيقة في الإجراءات كالفرق الكبير بين حجوم المجموعات أو عدم تجانس التباين بين المجموعات مما أدى لظهور الفرق على الرغم من ضعف حجم التأثير وضعف قوة الاختبار الإحصائي المستخدم.

**رابعاً: نتائج السؤال الرابع:**

ينص السؤال الرابع للبحث الحالي على "ما حجم العينة المناسب لاختبار دلالة الفروق بين المتوسطات في الدراسات النفسية والتربوية في ضوء حجم التأثير وقوة الاختبار الإحصائي؟".

وللإجابة عن هذا السؤال تم استخدام برنامج GPower في اشتقاق حجوم للعينات في حالة استخدام اختبار "ت" باختلاف حجم التأثير المتوقع (صغيراً، متوسطاً، كبيراً) حيث كانت قيمة d على الترتيب (أقل من ٠,٥، أكبر من أو تساوي ٠,٥ وأقل من ٠,٨، أكبر من أو تساوي (٠,٨ وفي حالة استخدام تحليل التباين أحادي الاتجاه باختلاف عدد المجموعات موضع المقارنة (٣، ٤، ٥، ٦) وفي حالة اختلاف حجم التأثير المتوقع (صغيراً، متوسطاً، كبيراً) حيث كانت قيمة f على الترتيب (أقل من ٠,٢٥، أكبر من أو تساوي ٠,٢٥ وأقل من ٠,٤، أكبر من أو يساوي ٠,٤) وذلك بافتراض أن قوة الاختبار المرغوبة ٠,٨ عند مستوى ثقة ٠,٠٥، وعند مستوى ثقة ٠,٠١ في حالة (اختبار "ت"، اختبار تحليل التباين أحادي الاتجاه)، فكانت النتائج كما هو موضح في التالي:

**جدول (١١)**
**حجم العينة المطلوب في حالة استخدام تحليل التباين أحادي الاتجاه وفقاً لمستوى الثقة وحجم التأثير وعدد المجموعات في حالة قوة الاختبار المرغوبة تساوي ٠,٨**

| عدد المجموعات | مستوى الثقة α | حجم التأثير f | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ضعيفاً | | | متوسطاً | | | كبيراً | | | | |
| | | ٠,١ | ٠,٢ | ٠,٢٥ | ٠,٣ | ٠,٣٥ | ٠,٤ | ٠,٥ | ٠,٦ | ٠,٧ | ٠,٨ | ٠,٩ | ١ |
| ٣ | ٠,٠٥ | ٩٦٩ | ٢٤٦ | ١٥٩ | ١١١ | ٨٤ | ٦٦ | ٤٢ | ٣٠ | ٢٤ | ٢١ | ١٨ | ١٥ |
| | ٠,٠١ | ١٣٩٥ | ٣٥٤ | ٢٢٥ | ١٥٩ | ١٢٠ | ٩٣ | ٦٣ | ٤٥ | ٣٦ | ٢٧ | ٢٤ | ١٨ |
| ٤ | ٠,٠٥ | ١٠٩٦ | ٢٨٠ | ١٧٦ | ١٢٨ | ٩٦ | ٧٦ | ٤٨ | ٣٦ | ٢٨ | ٢٤ | ٢٠ | ١٦ |
| | ٠,٠١ | ١٥٥٢ | ٣٩٦ | ٢٥٦ | ١٨٠ | ١٣٢ | ١٠٤ | ٦٨ | ٥٢ | ٤٠ | ٣٢ | ٢٨ | ٢٤ |
| ٥ | ٠,٠٥ | ١٢٠٠ | ٣٠٥ | ٢٠٠ | ١٤٠ | ١٠٥ | ٨٠ | ٥٥ | ٤٠ | ٣٠ | ٢٥ | ٢٥ | ٢٠ |
| | ٠,٠١ | ١٦٨٥ | ٤٣٠ | ٢٧٥ | ١٩٥ | ١٤٥ | ١١٥ | ٧٥ | ٥٥ | ٤٥ | ٣٥ | ٣٠ | ٢٥ |
| ٦ | ٠,٠٥ | ١٢٩٠ | ٣٣٠ | ٢١٦ | ١٥٠ | ١١٤ | ٩٠ | ٦٠ | ٤٢ | ٣٦ | ٣٠ | ٢٤ | ٢٤ |
| | ٠,٠١ | ١٨٠٠ | ٤٥٦ | ٢٩٤ | ٢١٠ | ١٥٦ | ١٢٠ | ٨٤ | ٦٠ | ٤٨ | ٣٦ | ٣٦ | ٣٠ |

**جدول (١٢)**
**حجم العينة المطلوب في حالة استخدام اختبار "ت" وفقاً لمستوى الثقة وحجم التأثير في حالة قوة الاختبار المرغوبة تساوي ٠,٨**

| حالات اختبار «ت» | مستوى الثقة α | حجم التأثير d | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ضعيفاً | | | | متوسطاً | | | كبيراً | | |
| | | ٠,١ | ٠,٢ | ٠,٣ | ٠,٤ | ٠,٥ | ٠,٦ | ٠,٧ | ٠,٨ | ٠,٩ | ١ |
| المجموعات المستقلة | ٠,٠٥ | ٣١٤٢ | ٧٨٨ | ٣٥٢ | ٢٠٠ | ١٢٨ | ٩٠ | ٦٨ | ٥٢ | ٤٢ | ٣٤ |
| | ٠,٠١ | ٤٦٧٦ | ١١٧٢ | ٥٢٤ | ٢٩٦ | ١٩٢ | ١٣٤ | ١٠٠ | ٧٨ | ٦٢ | ٥٢ |
| المتوسطات المرتبطة | ٠,٠٥ | ٨٢٤ | ٢٠٨ | ٩٤ | ٥٤ | ٣٥ | ٢٥ | ١٩ | ١٥ | ١٣ | ١١ |
| | ٠,٠١ | ١٢٢٧ | ٣١٠ | ١٤٠ | ٨٠ | ٥٣ | ٣٨ | ٢٩ | ٢٣ | ١٩ | ١٦ |
| المجموعة الواحدة | ٠,٠٥ | ٧٨٧ | ١٩٩ | ٩٠ | ٥٢ | ٣٤ | ٢٤ | ١٩ | ١٥ | ١٢ | ١٠ |
| | ٠,٠١ | ١١٧٢ | ٢٩٦ | ١٣٤ | ٧٧ | ٥١ | ٣٦ | ٢٨ | ٢٢ | ١٨ | ١٦ |

ويلاحظ أنه في حالة التوقع بأن يكون حجم التأثير كبيراً فإن حجم العينة الكلي المطلوب يكون محدوداً ولكن في حالة كان حجم التأثير متوسطاً أو ضعيفاً فإن حجم العينة المناسب لتحقيق قوة الاختبار المطلوبة يكون كبيراً، ويمكن توضيح كيفية الاسترشاد بالجداول السابقة في تحديد حجم العينة في الدراسات والبحوث التربوية والنفسية بالأمثلة التالية:

- عندما نرغب في الكشف عن دلالة الفروق في التحصيل على سبيل المثال بين أربع مجموعات تم تصنيفهم وفقاً لمتغير توجهات أهداف الإنجاز في إطار النموذج الرباعي (إتقان/ إقدام، إتقان/ إحجام، أداء/ إقدام، أداء/ إحجام)، ونحتاج لتحديد حجم العينة المناسب لاستخدام تحليل التباين أحادي الاتجاه في المقارنة بين المجموعات الأربع لنحقق قوة اختبار قدرها ٠,٨ كحد أدنى (احتمالية الخطأ من النوع الثاني ٠,٢) وسوف نعتمد على مستوى ثقة ٠,٠٥ في اختبار دلالة الفرض الصفري، تبقى المشكلة الآن في تحديد حجم التأثير والذي يمكن الوصول لصورة تقريبية له عن طريق أحد أو كل الخطوات التالية:
- مراجعة الدراسات السابقة والإطار النظري المرتبط بموضوع الدراسة ومحاولة الوصول إلى قرار بشأن تأثير توجهات أهداف الإنجاز على الجوانب الأكاديمية المرتبط بالتعلم.
- مراجعة حجم التأثير في الدراسات السابقة ومحاولة الوصول إلى قيمة تقريبية لحجم التأثير.

- يمكن إجراء دراسة استطلاعية والتوصل لمؤشر تقريبي لحجم التأثير بناءً على الفرق بين متوسطي أعلى مجموعة وأقل مجموعة في المتغير التابع مقسوماً على التباين العام للمجموعتين، وهذه الطريقة مفضلة حسب كثيراً من التوجهات العلمية في هذا الصدد.

من خلال الخطوات السابق أمكن للباحث الإقرار بأن حجم التأثير متوسطٌ وليكن ٠,٣ (أعلى من ٠,٢٥ وأقل من ٠,٤) وبالتالي يكون حجم العينة المناسب لاستخدام تحليل التباين الأحادي ١٢٨ مقسماً على أربع مجموعات، أما إذا كان مستوى الثقة المطلوب ٠,٠١ فيكون حجم العينة المناسب ١٨٠ مقسماً على أربع مجموعات

ويكون الأمر نفسه في حالة استخدام اختبار "ت"، وهنا تظهر مشكلة أن هذه الجداول وما شابهها في الدراسات السابقة يفترض فيها تساوي حجوم المجموعات موضع المقارنة، وفي هذا الصدد تؤكد الدراسات والبحوث السابقة أن منعة الاختبارات الإحصائية كاختبار تحليل التباين واختبار "ت" لا تتأثر كثيراً بهذه القضية، ومعنى ذلك في مثالنا السابق بأن المجموعة الواحدة لا تقل عن (١٢٨ ÷ ٤ = ٣٢ فرداً) في حالة الاعتماد على مستوى ثقة ٠,٠٥ أما في حالة الاعتماد على مستوى الثقة ٠,٠١ في تحديد الحجم التقريبي للعينة فنجد أن المجموعة الواحدة يجب ألا تقل عن ٤٥ فرداً.

وفي الدراسات التجريبية التي تعتمد على التصميم التجريبي الذي يتضمن أكثر من مجموعتين، يكون هنا حجم التأثير المتوقع كبيراً وليكن ٠,٥ على سبيل المثال ولذلك نجد أن حجم العينة الكلي في حالة ثلاث مجموعات مثلاً هو ٤٢ في حالة الاعتماد على مستوى الثقة ٠,٠٥، و٦٣ في حالة الاعتماد على مستوى الثقة ٠,٠١، ويكون عدد أفراد المجموعة الواحدة يجب ألا يقل عن ١٤ و٢١ على الترتيب.

ولا نَنْسَى هنا ضرورة التأكيد على أهمية تحقق شرطِ اعتدالية توزيع الدرجات في المجموعات موضوع المقارنة وتجانس تباين المجموعات.

## التوصيات والمقترحات

- ضرورة أن يتم إلزام الباحثين في مرحلة الدراسات العليا من قبل المشرفين والمناقشين للرسائل العلمية بتضمين النتائج لمعلومات عن حجم التأثير وقوة الاختبار الإحصائي المستخدم حتى تتحقق المصداقية والشفافية في نتائج تلك الدراسات، وكذلك تتضح الأهمية العملية للنتائج التي يتم التوصل إليها.

- ضرورة أن يكون من ضمن شروط النشر في المجلات التربوية والنفسية أن يتضمن تقرير البحث معلومات وافيه عن حجوم التأثير وقوة الاختبار الإحصائي، وهو ما يفيد في الدراسات المستقبلية ويزيد من القيمة العملية للنتائج التي يتم التوصل إليها.
- أن يكون هناك دليل إجرائي لكيفية عرض نتائج البحوث والدراسات في البيئة العربية على غرار الدليل الذي يصدر عن الجمعية الأمريكية لعلم النفس، بحيث يتم الالتزام بهذا الدليل عند نشر البحوث والرسائل العلمية، ويمكن أن يتم إعداد هذا الدليل بالتعاون مع الجمعيات المختلفة لعلم النفس كالجمعية السعودية والجمعية المصرية للعلوم النفسية.
- العمل على زيادة وعي الباحثين وطلاب الدراسات العليا بأهمية ومعنى الدلالة العملية (حجم التأثير) بجانب الدلالة الإحصائية، وأن قبول الفروض الصفرية وعدم الكشف عن تأثيرات المتغيرات المستقلة على المتغيرات التابعة أو الكشف عن العلاقات بينها قد يكون بسبب ضعف قوة الاختبارات الإحصائية المستخدمة في معالجة تلك الفروض.

وفي ضوء نتائج البحث الحالي يمكن اقتراح الدراسات التالي:

- تنفيذ نفس فكرة البحث الحالي على مجلات وأوعية نشر مختلفة وفي تخصصات أخرى.
- دراسة وعي الباحثين بالمفاهيم الإحصائية المختلفة وخاصة طلاب مرحلة الدراسات العليا.
- تصميم البرامج التدريبية التي تهدف لرفع وعي وكفاءة الباحثين بالجوانب الإحصائية ودراسة فاعلية مثل هذه البرامج.

## المراجع


باهي، مصطفى حسين (٢٠١٠). العلاقة بين الدلالة الإحصائية وحجم التأثير في البحوث التربوية والنفسية. المؤتمر العلمي السادس عشر لكلية التربية بجامعة حلوان، مستقبل إعداد المعلم في كليات التربية وجهود الجمعيات العلمية في عمليات التطوير بالعامل العربي، ، ٢٨-٢٩ مارس المجلد الثاني، ٤١٥-٤٤٤.

الحاج، فريال محمود (٢٠٠٣). واقع الدلالة الإحصائية والدلالة العملية وقوة الاختبار للاختبارات الإحصائية المستخدمة في رسائل الماجستير الصادرة عن كلية العلوم التربوية. رسالة ماجستير غير منشورة، جامعة مؤتة، الأردن.

حجمات، تحسين وعليان، خليل (١٩٩٧). واقع الدلالات الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير الإرشاد النفسي والتربوي بالجامعة الأردنية. دراسات. العلوم التربوية. الأردن، ٢٤(٢)، ٣٩٨-٤٠٨.

الصياد، عبدالعاطي أحمد (١٩٨٨). الدلالة العملية وحجم العينة المصاحبان للدلالة الإحصائية لاختبار "تاء" في البحث التربوي والنفسي العربي. مؤتمر بحوث البحث التربوي، الواقع والمستقبل، المجلد الثاني، رابطة التربية الحديثة، المركز القومي للبحوث التربوية، القاهرة.

محمد، محمد إبراهيم (٢٠١٣). تحليل قوة الاختبار الإحصائي وعلاقتها بمستوى الدلالة وحجم التأثير في البحوث التربوية. دراسات عربية في التربية وعلم النفس. ٣٧(٣)، ١٠٠-١٢٥.

نصار، يحي (٢٠٠٢). حجم الأثر كأسلوب إحصائي مكمل لفحص الفرضيات الإحصائية. مركز بحوث كلية التربية، جامعة الملك سعود، ١٧٦، ١-٤١.

Ahmed, W.; Amin, W.; Aleng, N. & Mohamed, N. (2012). Some Practical Guidelines for Effective Sample-Size Determination in Observational Studies. *Aceh International Journal of Science and Technology, 1*(2), 51-53. DOI: https://doi.org/10.13170/aijst.1.2.127

Algina, J. & Olejnik, S. (2003). Conducting Power Analyses for ANOVA and ANCOVA in Between-Subjects Designs. *Evaluation & the Health Professions, 26*(3), 288-314. DOI: https://doi.org/10.1177/0163278703255248

Batterham, A. & Atkinson, G. (2005). How big does my sample need to be? A primer on the murky world of sample size estimation. *Physical Therapy in Sport, 6*, 153–163.

Blanca, M.; Alarcón, R.; Arnau, J.; Bono, R. & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. *Psicothema, 29*(4), 552-557. DOI: 10.7334/psicothema2016.383.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Mahwah, NJ: Lawrence Erlbaum

Cunningham, J. B. & McCrum-Gardner, E. (2007). Power, effect and sample size using GPower: practical issues for researchers and members of research ethics committees. *Evidence Based Midwifery, 5*(4), 132-6

Descoteaux, J. (2007). Statistical power: An historical introduction. *Tutorials in Quantitative Methods for Psychology, 3*(2), 28 34. DOI: 10.20982/tqmp.03.2.p028

Fidler, F. & Thompson, B. (2001). Computing Correct Confidence Intervals for ANOVA Fixed- And Random-Effects Effect Sizes. *Educational and Psychological Measurement, 61*(4), 575-604. DOI: https://doi.org/10.1177/00131640121971383

Greenland, S.; Senn, S.; Rothman, K.; Carlin, J.; Poole, C.; Goodman, S. & Altman, D. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350, DOI: 10.1007/s10654-016-0149-3.

Guo, J. & Luh, W. (2008). Approximate Sample Size Formulas for Testing Group Mean Differences When Variances Are Unequal in One-Way ANOVA. *Educational and Psychological Measurement, 68*(6), 959-971. DOI: https://doi.org/10.1177/0013164408318759

Hecke, T. (2012). Power study of ANOVA versus Kruskal-Wallis test. *Journal of Statistics and Management Systems, 15*(3), 241-247. DOI: https://doi.org/10.1080/09720510.2012.10701623

Karadag, O. & Aktas, S. (2012). Optimal Sample Size Determination for the ANOVA Designs. *International Journal of Applied Mathematics & Statistics, 25*(1), 127-134.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology,* 4:863, DOI:10.3389/fpsyg.2013.00863.

Lenth, R. (2001). Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician, 55*(3), 187-193. DOI: https://doi.org/10.1198/000313001317098149

Pierce,C.;Block,R.&Aguinis,H.(2004).CautionaryNoteonReportingEta-Squared Values from Multifactor ANOVA Designs. *Educational and Psychological Measurement, 64*(6), 916-924. DOI: 10.1177/0013164404264848.

Prajapati, B.; Dunne, M. & Armstrong, R. (2010). Sample size estimation and statistical power analyses. *Optometry Today*, July, 1-9

Rasmussen, J. (1983). *Parametric Vs Nonparametric Tests on Non-Normal and Transformed Data*. Ph. D. Dissertation, Graduate School of Tulane University, New Orleans

Trajkovski, V. (2016). How to Select Appropriate Statistical Test in Scientific Articles. *Journal of Special Education and Rehabilitation, 17*(3-4), 5-28. DOI: 10.19057/jser.2016.7

Wang, Y.; Gil, P.; Chen, Y.; Kromrey, J.; Kim, E.; Pham, T.; Nguyen D. & Romano, J. (2016). Comparing the Performance of Approaches for Testing the Homogeneity of Variance Assumption in One-Factor ANOVA Models. *Educational and Psychological Measurement, 77*(2), 305-329. DOI: https://doi.org/10.1177/0013164416645162

Yusof, Z.; Abdullah, S. & Yahaya, S. (2012). Type I Error Rates of Parametric, Robust and Nonparametric Methods for two Groups Cases. *World Applied Sciences Journal, 16*(12), 1815-1819.