النمذجة البنائية للعلاقات بين المتغيرات المؤثرة في حجم التأثير لاختبار "ت" في البحوث النفسية العربية

إعداد

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

د.محمود علي موسى *** السيد أحمد موسى ****

مقدمة:

تعد الإحصاء علم صنع القرار التربوي في ضوء حالات عدم اليقين. ويتعامل معها الباحث على أنها أداة لتطبيق الأساليب الإحصائية في معالجة قضايا السلوك الانساني. ويشرع الباحث النفسي في مستهل دراسته لتعريف مصطلحات دراسته نفسياً، ثم إجرائياً في ضوء أدوات القياس، ثم يأتي دور تحديد متغيرات الدراسة وتحديد كيفية قياسها وصياغة الفروض البحثية التي تعتبر تفسير واستنتاج متوقع لسؤال مشكلة الدراسة وعادة ما يكتب بصورة تقريرية أو كيفية، ثم ترجمة فرض البحث احصائياً بحيث يختبر صحة أو عدم صحة هذا الفرض إحصائياً باستخدام تحليل إحصائي أمثل (عبد العاطي الصياد وعبد الناصر عامر، ٢٠٠٨).

* أستاذ القياس والتقويم النفسي والتربوي.

** أستاذ القياس والاحصاء التربوي – قسم علم النفس التربوي، كلية التربية بالاسماعيلية، جامعة قناة السويس.

*** مدرس بقسم علم النفس التربوي – – قسم علم النفس التربوي، كلية التربية بالاسماعيلية، جامعة قناة السويس.

**** باحث دكتوراة في الصحة النفسية – قسم الصحة النفسية، كلية التربية، جامعة الزقازيق.

وقد يصاغ الفرض الإحصائي موجهاً أو صفرياً ويمكن اختبار الفرض الموجه بما يسمى بالاختبار ذي الذيل الواحد أما اختبار دلالة الفرض الصفري قد تم في ضوء اختبار ذي ذيلين ويكون الاختبار ذا الذيل الواحد أقوى من اختبار ذي الذيلين، وبالتالي فالقوة الإحصائية في الاختبار بذيل واحد تختلف بحسب قيمة المتوسط بالفرض من حيث كونها أكبر أو أصغر من قيمة المتوسط (عبد الله النجار، ٢٠٠٦).

والمتتبع لتوجه الباحث النفسي والتربوي فإنه من السهولة بمكان أن نكتشف أنه يصارع من أجل الوصول للدلالة الاحصائية لذا فإنه يختبر فروضاً بسيطة؛ فالدلالة الإحصائية ليست كافية لصنع القرار التربوي أو النفسي فهي شرط ضروري وليس كافياً؛ وتتحقق هذه الكفاية من خلال حساب قوة العلاقة بين المتغير المستقل والتابع (عبد العاطي الصياد، ١٩٨٨).

والقرار الذي يتخذه الباحث فيما يتعلق بالفرض الصفري المراد اختباره أو التحقق منه يتطلب وجود قاعدة يستند إليها في هذا الشأن، فيحاول الباحث غالباً الوصول إلى أدلة من البيانات التي قام بجمعها تمكنه من رفض الفرض الصفري وتأييد الفرض البحثي الذي يشتق من إطار نظري يتبناه ويرى أنه يفسر الظاهرة تفسيراً منطقياً، لذلك ينبغي أن يحدد الباحث قبل شروعه في جمع البيانات قيمة احتمالية معينة تبين مقدار الخطأ الذي يقبل أن يقع فيه نتيجة رفضه للفرض الصفري، وبعبارة أخرى إذا قرر الباحث على أساس البيانات التجريبية التي حصل عليها رفض الفرض الصفري، فإن احتمال خطأ هذا القرار يكون أقل من أو مساوياً هذه القيمة التي يطلق عليها مستوى الدلالة الإحصائية أو ألفا. ويمكن استخلاص التفسيرات التالية في دراسة وتفسير الدلالة الاحصائية:

١. قيمة الدلالة الاحصائية هي احتمال أن تكون الفروق بين المجموعتين ناتجة عن الصدفة (Brockhoff, 2003).
٢. قيمة الدلالة الاحصائية هي احتمال أن يكون الفرض الصفري صحيحاً (وبالتالي فقيمة $1-p$ هي احتمال أن يكون الفرض البديل صحيح). وهذا هو التفسير الصحيح المستمد من تحليل النظرية البيزانية Bayesian analysis، والتي سعى معظم الباحثين لمعرفتها (Burk, 2005).

٣. قيمة الدلالة الاحصائية المحسوبة باستخدام المداخل الكلاسيكية Classical approaches، يمكن أن تنتج قياسات عالية مضللة للبيانات المستخدمة لرفض الفرض الصفري (Burk, 2005).
٤. حجم التأثير هو الوجه المكمل للدلالة الاحصائية، بمعني أن الدلالة الصغيرة تعطي تأثيرات عالية كبيرة وهذا ما أوضحه Burk (2005) في دراسته.

ويرى الصياد (١٩٨٨) أن العلاقة الارتباطية بين مستوى الدلالة الاحصائية التي يقررها الباحث وقيمة حجم الأثر علاقة عكسية وهذا يعني تقرير الباحث حجم العينة مسبقاً ثم جمع البيانات وتحليلها إحصائيا، ثم يقرر قيمة ألفا التي تعطيه الدلالة الاحصائية؛ أي تحديد قيمة مستوى الدلالة قبل جمع البيانات وبعد ضبط وتحديد حجم العينة.

حجم التأثير حول نقطة وفترة:

ويعرف Cohen (1988) حجم التأثير بأنه الدرجة التي تكون عليها الظاهرة في المجتمع أو الدرجة التي تكون فيها فرضية العدم غير صحيحة، وبالرغم من إتاحة بيانات المجتمع ككل فعادة ما يستخدم حجم التأثير للإشارة إلى نموذج يستند إلى تقدير حجم العلاقة بين المتغيرات المستقلة والتابعة في ضوء عينة غير ممثلة للمجتمع الاصلي الذي سحبت منه (Rosenthal, 1994). وبالرغم من وجود عدة مؤشرات لحساب حجم التأثير، إلا أنه يمكن تصنيفها إلى نوعين: قياسات الفروق المعيارية Standardized difference بين متوسطات المجموعة وقياس التباين (Kirk, 1996).

وحدد Cohen (1988) متطلبات حجم التأثير المتوقع وهي تحديد نوع الاختبار المستخدم، مستوى الدلالة الاحصائية، حجم العينة المستخدم، ولكنه يعاني بعض المحددات منها أنه عالج حجم الأثر وقوة الاختبار حتى النموذج المتعدد فحسب، ولم يظهر إلا بعض المداخل الاجتهادية فيما يتصل للنموذج المتدرج والنموذج الشامل.

وبدلاً من الاعتماد على اختبارات الدلالة فيجب توجه الباحثين إلي حجوم التأثير، وفترات الثقة. وفترات الثقة هي الأكثر معلوماتية عن كون دلالة الاختبار من عدمه، وقد حددها (Breaugh, 2003; Hubbard & Armstrong, 2005) فيما يلي:

أولاً: اعتماد التوجهات المعاصرة على التوصل إلى تقديرات ذات مصداقية لحجم التأثير وفترات الثقة لتوسيع مدى هذه التقديرات المعبرة عن المجتمع.

ثانياً: فترة الثقة تسهل قياس الصدق أو دقة التقدير، كما توفر الموضوعية بدلاً من الدلالة الإحصائية؛ ويرجع ذلك إلى صياغة حدود الثقة في ضوء فترة متصلة، تؤدي إلى معقولية القيم في هذه الفترة عند تفسيرها في سياق المشكلة، على العكس من اختبارات الدلالة الإحصائية التي تكون أكثر عرضة للخطأ من النوع الأول (α).

ثالثاً: تستخدم فترات الثقة كاختبار دلالة فعلى سبيل المثال ٩٥ % حد ثقة التي لا تتضمن قيمة العدم Null Value، تعادل رفض الفرضية عند مستوي ٠.٠٥.

وقد قارن Parker & Vannest (2009) بين حجم التأثير عند قيمة معينة باستخدام مؤشر كوهين d وحجم التأثير حول فترة Effect size over Interval، وقد توصل إلى فعالية الفترة في حساب حجم التأثير وذلك بسبب وقوع متوسط الفروق بين حدي الفترة. كما أن حجوم العينات الصغيرة تؤدي إلى فترات ثقة واسعة تتسع إلى (٣٤٪ حتى ٧٠٪)، وحجوم العينات الكبيرة تؤدي إلى فترات ضيقة (٤٧.٦٪ حتى ٥٦.٤٪) (Florey, 1993).

وتتأثر قيمة حجم التأثير في الدراسات الوصفية باختلاف أحجام العينة واعتدالية البيانات بغض النظر عن مدى تجانس البيانات ويرى Olejnik & Huberty (2001) أن القوة التنبؤية باستخدم النماذج الانحدارية تزداد تحت شروط العينات الكبيرة واعتدالية توزيع بيانات العينة وقد توصل لتلك النتيجة اعتماداً على بيانات محاكاة.

وتعتبر القوة الإحصائية أحد الاختبارات المكملة للدلالة الإحصائية وهي احتمالية أن يرفض الفرضية الصفرية بطريقة صحيحة (Huston, 1993)، وقد لاحظ Thompson (1999) أن ارتفاع القوة الإحصائية للنتائج غير الدالة إحصائيا يسهم في تفسير المعرفة؛ كما أن قبول الفرض الصفري دون معرفة القوة الإحصائية يحدث خطأ النوع الثاني.

ويزداد حجم التأثير بتدني مستوى الدلالة الإحصائية وزيادة حجم العينة واعتدالية بيانات العينة ولكن القوة الاحصائية وحجم التأثير تتمتعان بالضلاعة بشرط تجانس العينات (Brockhoff, 2003; Brewer, 1972; Keselman & Penfield, 2005; Nandy, 2012)، فإذا كان مستوى الدلالة وتباين درجات العينة ثابتاً فإن زيادة حجم العينة يزيد القوة الإحصائية وحجم التأثير، والإسهام الأكبر لمستوى الدلالة وتباين المجتمع بجانب حجم العينة، فإذا كانت قيمة ألفا ثابتة وكذلك حجم العينة، فإن قيمة بيتا تقل بزيادة الفرق بين المتوسطين، ومعنى هذا أنه كلما ازداد الفرق بين المتوسطين، فإن إحتمال قبول فرض العدم يقل. أما إذا كان الفرق بين المتوسطين حجم العينة ثابتاً، ازدادت قيمة بيتا بنقص قيمة ألفا، فانخفاض ألفا يؤدي إلى الفشل في رفض فرض العدم بالرغم من وجود فروق بين المتوسطين (عادل بابطين، ٢٠٠٢).

وهنا يجب ملاحظة أن العينات الكبيرة تجعل أي فرق بسيط بين الإحصاءة والمعلم المناظر لها فرقاً ذي دلالة عملية وعندئذ يختفي خطأ المعاينة، بالرغم من إشارة الإختبار الإحصائي إلى وجود فروق جوهرية، يصبح الاختبار أشد كفاءة وحساسية، لأنه يصل إلي مستويات الدلالة الإحصائية والعملية بعينة أصغر من غيرها (زكريا الشربيني، ٢٠٠١). ونظراً لتجانس المجموعتين الضابطة والتجريبية وصغر حجم العينة وارتباط متغيرات الدراسة مما يعني وجود قدر مشترك من التباين، وهذا يحدث تفاعلاً بين عملية الاختيار وعملية المعالجة، وهذا التفاعل نتيجة لعملية التحيز في الإختيار لا يهدد الصدق الداخلى فحسب بل الصدق الخارجى للتجربة؛ كما يتعرض الصدق الداخلى للخطر لأن المجموعتين التجريبية والضابطة لم يتم اختيارهما عشوائياً من مجتمع الدراسة، ولهذا فإن نتائج التجربة لا يمكن تعميمها بموثوقية لأي مجتمع دراسي وتسمى عملية المعرفة المسبقة بطبيعة أفراد العينة وتوقعاتهم وتحيزهم تسمى عملية التلويث الإحصائي Statistical Contamination، فمن أجل تعميم النتائج على مجتمع افراد الدراسة فمن الضرورى أن تمثل المجموعات مجتمع الدراسة الذي اختيرت منه تمثيلاً صادقاً.

وبالتالي فتضخم مؤشرات حجم التأثير الناتجة عن تلك الدراسات تعتبر مهدد قوي للصدق الخارجي للتجربة البحثية أو القدرة على تعميم ما يتم التوصل إليه من نتائج أثناء التجربة على أفراد آخرين وعلى المجتمع الذي يمثله الأفراد الداخلين في التجربة فقد تكون البيانات البحثية خاضعة للتزييف أو التحايل على الدلالة الإحصائية باستخدام أسلوب أو نموذج إحصائي غير الواجب استخدامه أو التحيز في تطبيق الإختبار أو الفترة الزمنية الطويلة بين التطبيق القبلي والبعدي (عامل التاريخ) والإختيار العشوائي لعينة مجتمع الدراسة. وهنا يخرج الباحثون بالعديد من الإستقراءات لمؤشرات حجم التأثير في البحث النفسي وهي كالأتي:

١. يتعامل الباحث في الحقل التربوي مع مؤشرات الدلالة العملية علي أنها ظواهر ذرية وتفتيت البناء الكلي للظاهرة وتجزيئها والإعتماد علي مؤشرات كلاسيكية كمؤشر كوهين (Cohen, 1988) ومربع ايتا باعتبارها مؤشرات فريدة في علوم النفس والتربية وعدم استخدام الأساليب الحديثة كفترات الثقة Confidence interval وتحليل الإعادة Replication analysis وغيرها من الأساليب الحديثة.
٢. عدم وجود مؤشرات للدقة الإحصائية في النماذج ابتداء من النموذج المتدرج وحتى النموذج الشامل، وبالرغم تعددها في النماذج الاحصائية الأخرى إلا أنها مداخل تنافسية من وجهة نظر الباحثون حيث لم يثبت فعالية أي من هذه المداخل، وبالرغم من التقدم الهائل في أساليب حسابها إلا أن التوجه نحو الحسابات اليدوية مازال قائماً.
٣. اختلاف قيمة الدلالة العملية باختلاف معادلة الحساب داخل المؤشر الواحد وبالتالي وجود فجوه في الحكم علي قيمة المؤشر.
٤. القصور في حساب قوة الإختبار الإحصائي الناتج عن استخدام بعض مؤشرات الدلالة العملية كمربع ايتا ومربع ابسلون ومؤشر دلتا وكراميز وغيرها من المؤشرات.

حساسية التصميم Design sensitivity:

يعتمد حجم التأثير على ظاهرة حساسية التصميم Design sensitivity ويقصد بها العلاقة بين حجم العينة (N) واعتدالية البيانات ومستويات الدلالة

الاحصائية (α) والتصميم التجريبي Research design المستخدم في الدراسة وثبات أدوات القياس Instruments reliability بغض النظر عن مدى تجانس بيانات العينة Heterogeneity.

وأكد Nandy (2012) وجود علاقة طردية بين حجم الأثر وثبات أدوات القياس. وأكد الصياد (١٩٨٨) أن تدني قيمة حجم الأثر ترجع إما لسوء اختيار الباحث لمتغيراته المستقلة عن طريق المصادفة، أو أنه حديث العهد بتلك المتغيرات في بحثه العلمي وبالتالي يقوم بضبط مستوى الدلالة الاحصائية والتخفي وراء حجم العينة.

ويرى Cohen (1962, 1988) أن العلاقة بين حجم الأثر وحجم العينة علاقة طردية في حالة ضبط التصميم التجريبي والمبالغة في ضبطه كاختيار مجموعات متطرفة في الصفات موضع القياس تؤدي إلى تضخم الفروق أو حدوث التلويث الإحصائي والذي يبدو في الأبحاث العلمية لبعض التخصصات فقد تجتاز قيمة الدلالة العملية ٣ وهذا غير مقبول نظرياً بالنسبة لبعض المراحل العمرية التي يطبق عليها البحث العلمي مما يعيق من تطبيق الصدق التعميمي لهذه النتائج.

مشكلة الدراسة والدراسات السابقة:

حسب Kroll & Chase (1975) حجم التأثير لعددين من مجلة بحوث اضطرابات التواصل وقد توصل إلى أن متوسطات حجم التأثير لاختبار ت (٠.١٦ ضعيف، ٠.٤٤ متوسط، والمرتفع ٠.٧٣) على الترتيب، وقدر Pillai, Katsikeas & Presi (2012) حجم تأثير المادة المنشورة للقراء في الصحف الرسمية على معدل تسويق تلك المنتجات باستخدام ماوراء التحليل وقد بلغ حجم الأثر الكلي ٠.٤٧.

حسب Brewer (1972) متوسط حجم تأثير كوهين لـ ٣٧٣ دراسة من مجلة AERJ التي استخدمت اختبار "ت" لحجوم عينات مختلفة ومستويات دلالة مختلفة (٠.٠١، ٠.٠٥)، وقد كانت متوسطات حجم التأثير (٠.١٧ ضعيف، ٠.٧٠ متوسط، ٠.٩٨ مرتفع) عند مستوى دلالة ٠.٠١، (٠.٠٦ ضعيف، ٠.٤٥ متوسط،

٠.٩١مرتفع) لمستوى دلالة ٠.٠٠٥. وحسب Slavin & Smith (2009) العلاقة بين أحجام العينة وحجم التأثير لـ ١٨٥ دراسة منشورة في بحوث التعلم المنظومي لتخصص المناهج وطرق التدريس وقد أسفرت دراسته عن علاقة سالبة دالة احصائياً بين حجم العينة والقوة الاحصائية، وقد كانت الفروق في حجم الأثر للتجارب الصغيرة أكبر من التجارب العشوائية والمرتبطة. وفسر الصياد (١٩٨٨) العلاقة العكسية بين حجم العينة وحجم التأثير بأن الباحث أصبح لديه حس عن قيمة الأثر فإذا ما توقع حجم تأثير كبير فإنه يختار حجم عينة صغير والعكس صحيح؛ وذلك بعد تحديده لمستوى الدلالة مسبقاً وجعلها ثابته على طول الدراسة. واتفق (Brewer, 1972; Nandy, 2012) على وجود علاقة طردية بين حجم الأثر والقوة الاحصائية وحجم العينة وثبات أدوات القياس.

وقدر Sawyer & Ball (1981) القوة الاحصائية وحجم تأثير كوهين لـ ٤٧٥ دراسة بمجلة بحوث التسويق للاختبارات الاحصائية المستخدمة وقد بلغت متوسطات حجم التأثير ٠.١٠ ضعيف، ٠.٣٠ متوسط، ٠.٥٠ مرتفع وذلك للدراسات التي استخدمت اختبار ت. وقيم Mazen, Hemmasi & Lewis (1987) حجم التأثير وقوة الاختبار للبحوث الاستراتيجية لـ ٧٠ مقال و٢٠٨٨ دراسة وقد بلغ وسيط حجم التأثير لاختبار ت ٠.١٣ ضعيف، ٠.٥٣ متوسط، ٠.٨٨ مرتفع. بينما استخدم Woodside, Beretich & Lauricella (1993) ما وراء التحليل لحجوم التأثير للصفقات التجارية المباشرة لـ ٢١١ دراسة منشورة وقد استخدمت الدراسة معدلات الاستجابة التجارية كمتغير تابع واختبار ت وقد تراوحت حجوم التأثير الكلية بين (٠.٠٣- ٠.٣٥) وهو حجم تأثير ضعيف.

وقد قيم Stout & Ruble (1995) استخدام الدلالة العملية للنتائج الأمبيريقية لبحوث التعلم المحاسبي، باستخدام دراسة وصفية وتوصل إلى أن ١٪ من الدراسات حجم التأثير بها ضعيف، ٩٪ استخدم حجم تأثير متوسط، ٢٥٪ حجم تأثير مرتفع في الدراسات التي استخدمت معامل الارتباط، بينما بلغت نسب حجوم التأثير للدراسات التي استخدمت أسلوب ت النسب ١٪ ضعيف، ٦٪ متوسط، ١٤٪ مرتفع. قيم Ellis (2010) ٢٠٤ دراسة منشورة للمجلة الدولية للبحوث التجارية في الفترة من ٢٠٠٣ وحتى ٢٠٠٥، وقد أسفرت الدراسة عن نقص قيم القوة الاحصائية وذلك نظراً لنقص معاملات ثبات الأدوات، وقد كانت تكرارات الدراسات حجم الأثر بالنسبة للدراسات ذات القوة الاحصائية (٠.٨ فأكثر) للأثر المنخفض

٢١ (١٥٪)، وللأثر المتوسط ١٨٢ (٨٩٪)، وللأثر المرتفع ٢٠٢ (٩٩٪)، أما مستويات القوة (٠.٥٠ حتى ٠.٧٩) بلغت ٢٥ (٦٧٪) للأثر المنخفض، ١٣ (٩٦٪) للأثر المتوسط، ١ (١٠٠٪) للأثر المرتفع.

وقدر عبد الناصر عامر (٢٠١٣) حجم التأثير في الدراسات النفسية التربوية العربية، ولتحقيق ذلك تم اختيار ٦٠ دراسة منشورة في مجلات تربوية مصرية وعربية في المدة من ١٩٩٤ وحتي ٢٠١١ وتضمنت هذه الدراسات ٤٨٩ اختباراً واتضح أن ١٣.٣٪ من الدراسات قدرت حجم التأثير، و ٧٥٪ فسرت هذه المؤشرات، وكان استخدام مؤشرات حجم التأثير لاختبار معامل ارتباط بيرسون (مؤشر r^2) ١٦.٣٪ ولاختبار T ٤.٤٪ ولاسلوب الانحدار المتعدد ٤٥.٨٪ و لكاى تربيع ولاختبار F صفر%، وتراوح وسيط مؤشرات حجم التأثير من المستوى المتوسط لاختبارات T, R, r، إلى المستوى الضعيف لـ X^2, F. وأخيراً أقترحت الدراسة مبادئ ارشادية للباحثين لاختيار وتقدير وتفسير مؤشرات حجم التأثير.

أما بالنسبة لشروط استخدام حجم التأثير فقد حسب Brockhoff (2003) القوة الاحصائية باستخدام بيانات المحاكاة لاختبار ذي الحدين وقد أسفرت دراسته عن القوة الاحصائية كدالة في عدد مرات تحليل الاعادة وأن تطرف البيانات ومتسع الانحراف المعياري للبيانات يؤثر سلباً على القوة الاحصائية وحجم التأثير. بينما قارن Campbell (2004) بين أربعة مؤشرات لحجم التأثير وذلك لـ ١١٧ دراسة و١٨١ مقالاً لبحوث تشخيص التوحد والتي اعتمدت على تصميم المجموعة الواحدة وتم حساب مؤشرات الاحصاء الوصفي لاستخدام تلك المؤشرات في الدراسات ونسبة تداخل البيانات، وقد أشارت جميع حجوم التأثير إلى فعالية العلاج السلوكي للتوحد تحت شرط الاعتدالية خصوصاً في مؤشر Cohen's d، وقد أخفق مؤشرات حجم التأثير القائمة على الانحدار في تحسين فهم نتائج العلاج أحادية الموضوع.

بينما قارن Keselman & Penfield (2005) بين ثمان فترات لثماني حدود للثقة لتصميم القياسات المتكررة (بمستويين)، وقد اختلفت فترات الثقة وذلك بسبب ضلاعة التقدير بالاضافة إلى اختلاف التوزيع النظري والامبيريقي للعينة وقد أسفرت الدراسة عن اختلاف تقدير حجم التأثير باختلاف اعتدالية البيانات، ومدى

تجانس التباين. وقارن Iversson, Andersen, Johnson & Lindwall (2013) الفروق بين حجم التأثير المصحح والغير مصحح خصوصاً في العينات الصغيرة، بالإضافة إلى وجود فروق حجم التأثير البارامتري واللابارامتري وقد أعزي هذه الفروق للتحيز الناتج عن الاحصاء اللابارامتري.

ودرس Florey (1993) أثر حجم العينة على حجم التأثير حول فترة وقد أسفرت الدراسة عن حجوم العينات الصغيرة تؤدي إلى فترات ثقة واسعة تتسع إلى (٣٤٪ حتى ٧٠٪)، وقد أسفرت حجوم العينات الكبيرة تؤدي إلى فترات ضيقة (٤٧.٦٪ حتى ٥٦.٤٪)، وقدر Fan (2001) نسبة الاختبارات الدالة في ضوء حجوم عينات مختلفة (٢٠، ٤٠، ٨٠، ١٦٠، ٢٤٠) وحجوم تأثير مختلفة باستخدام مؤشري d, r لحجم التأثير.

حسب Weisburd, Petrosino & Mason (1993) متوسطات حجوم التأثير للبحوث الوصفية عند نفس أحجام عينات السابقة وبلغت (٠.٥٢، ٠.٢٢، ٠.٢١، ٠.٢٨، ٠.٠٩)، كما حسب متوسطات حجم الأثر للبحوث التجريبية للبحوث التجريبية عند نفس أحجام العينات السابقة وقد بلغت على الترتيب (٠.٥٩، ٠.١٩، ٠.١٨). كما درس Carlson & Schmidt (1999) أثر التصميم التجريبي على حجم التأثير باستخدام ماوراء التحليل وقد استخدم الباحثان تصميم المجموعة الواحدة بقياسين (قبلي وبعدي) وتصميم المجموعتين بقياس بعدي واحد وقد بلغت حجوم الأثر الكلية لتصميم المجموعة الواحدة بلغ "٠.٦٦٨"، تصميم المجموعتين بقياس بعدي واحد "٠.٧٧٢".

وقارن Olejnik & Algina (2000) بين مؤشرات ($w2$, γ^2, ε^2) للدراسات التي استخدمت عائلة اختبارات تحليل التباين والتصميم البحثي ذو المجموعة الواحدة وقد تفوق مؤشر ايتا في حساب حجم التأثير. وقارن Algina & Keselman (2003) بين دقة التقدير لفترات الثقة التقليدية باستخدام Spss وفترات الثقة لـ Bird لتصميم المجموعات المرتبطة (القياسات المتكررة) وقد أسفرت دراسته عن تقارب طول فترات بيرد عند حد ثقة ٠.٩٥ وقد كانت فترات الثقة التقليدية أكثر دقة احصائية في حساب حجم التأثير.

وقدر Charles, Giraudeau, Dechartres, Baron & Ravaud (2014) حجم العينة المطلوب في التصميمات التجريبية وذلك من خلال تقييم ٢١٥

دراسة عددين يناير ٢٠٠٥ وديسمبر ٢٠٠٦ من المجلة الطبية وقد توصلت الدراسة إلى وجود ١٨٣ (٩٦٪) استخدمت مستوى دلالة احصائية ٠.٠٥، وعدد الاختبارات التي استخدمت التصميمات التجريبية ذوات المجموعتين ١٦٥ (٨١٪)، عدد الدراسات التي حسبت تأثير المعالجة ١٨٦ (١٩٪)، بلغت عدد الدراسات التي راعت المعايير اللازمة لتحديد حجم العينة ١١٣ (٥٣٪).

قدر عبد العاطي الصياد وعبد الناصر عامر (٢٠١٣) بعض مؤشرات حجم التأثير المستخدمة (d, r, w2, γ^2) لاختبار "ت" المستقلة، تم اختيار (٢٧) دراسة نفسية مصرية وعربية منشورة في المجلات النفسية والتربوية وتضمنت (١٤٩) اختباراً لـ "ت" وتوصلا إلى وجود تناقض في تفسير المؤشرات الأربعة سواء في ضوء معيار المتوسط والوسيط، وتم اقتراح معايير جديدة للمؤشرات الأربعة.ومن العرض السابق للتراث السيكولوجي، فيمكن صياغة مشكلة الدراسة في التساؤلات التالية: ما مدى مطابقة النموذج البنائي المنظم للعلاقات بين العوامل المؤثرة في حجم التأثير؟

هدف الدراسة:

تحديد الاسهام النسبي لحجم العينة ومستوى الدلالة الاحصائية والمنهجية وجنسية الباحث في تفسير حجم التأثير لاختبار ت.

أهمية الدراسة:

تحديد مدى تأثير كل من مستويات الدلالة وحجم العينة ونوع التصميم التجريبي على حجم الأثر. وتسليط الضوء على ظاهرة منهجية هي حساسية التصميم.

الطريقة والاجراءات

أولاً: المنهج: استخدام المنهج الوصفي الارتباطي لدراسة مطابقة النموذج البنائي للعلاقات بين العوامل المؤثرة في حجم التأثير لاختبار "ت" في البحث النفسي العربي.

ثانياً: عينة الدراسة: تضمنت عينة الدراسة ٤٩ دراسة نفسية منشورة في المجلات التربوية والنفسية في البيئة العربية في المدة من ٢٠٠٠ حتى ٢٠١٢؛ وتم اختيارها عشوائياً من مكتبة الدراسات العليا بكلية التربية، جامعة قناة السويس، وكانت أكثر المجلات استخداماً هي المجلة المصرية للدراسات النفسية (٢٨.٦٪) يليها مجلة كلية التربية الزقازيق (١٦.٣٪)، ثم مجلة العلوم النفسية والتربوية في البحرين (١٠.٢٪)، ثم مجلة كلية التربية بالمنصورة ومجلة كلية التربية بسوهاج وكلاهما بنسبة (٦١٪)، ثم تنوعت الدراسات في مجلات كلية تربية طنطا وأسيوط والمنوفية والإسكندرية ورسالة الخليج العربي. والدراسات التي قام بها باحثون ماجستير ودكتوراه ١٨١ دراسة (٨٨.٧٪).

ثالثاً: إجراءات الدراسة:

- **اختيرت الدراسات النفسية في المجلات النفسية والتربوية للباحثين العرب والمصريين التي أجريت على عينات عربية.**
- **اختيرت الدراسات في المدى الزمني من ٢٠٠٠ حتى ٢٠١٢ وقد اختيرت بصورة عشوائية من مكتبة كلية التربية، جامعة قناة السويس.**
- **قسمت الدراسات في ضوء العينة إلى ثلاث فئات (أقل من ١٠٠ و ١٠١ – ٢٠٠ و أكبر من ٢٠٠).**
- **اختيرت تخصصات البحوث النفسية في مجالات علم النفس والصحة النفسية.**
- **انتقيت الدراسات في ضوء منهجية الدراسة التي اعتمدت على المنهج الوصفي والمنهج التجريبي.**
- **صنفت الدراسات في ضوء مستوى الدلالة الاحصائية إلى ثلاثة فئات (٠.٠١ و ٠.٠٥ والدراسات التي لم تحدد مستوى الدلالة الاحصائية).**
- **صنفت الدراسات في ضوء المستوى الزمني إلى قسمين الأول من ٢٠٠٠ إلى ٢٠٠٦ والثاني ٢٠٠٧ إلى ٢٠١٢.**
- **صنفت الدراسات في ضوء حجم التأثير إلى ثلاثة فئات هي (صغير ومتوسط وقوي).**

- اعتمدت الدراسة على الأساليب الاحصائية الوصفية مثل التكرارات والنسب المئوية، ومتوسطات حجم التأثير ووسيط حجم التأثير لبعض الدراسات التي تطرفت بها قيم حجم التأثير. ونمذجة المعادلة البنائية.

نتائج الدراسة

مؤشرات الاحصاء الوصفي لمتغيرات الدراسة:

حسبت مؤشرات الاحصاء الوصفي لمؤشرات حجم التأثير جدول ولوحظ لمؤشر كوهين (Cohen, 1988) والجدول (١) يوضح قيم التكرارات والنسب المئوية لكل من متغيرات الدراسة على النحو التالي:

جدول (١): توصيف عينة الدراسة.

تصنيف العينة	ن < ١٠٠	١٠١- ٢٠٠	ن >٢٠٠	تخصص الباحث	علم النفس	الصحة النفسية
	٣٩	٦٦	٩٩		٣٥	١٤
	٪١٩.١	٪٣٢.٤	٪٤٨.٥		٪٧١.٤	٪٢٨.٦
مستوى الدلالة الاحصائية	٠.٠١	٠.٠٥	لم تحدد	منهج الدراسة	وصفي	تجريبي
	١٣٩	٦٤	١		٤٠	٩
	٪٦٨.١	٪٣١.٤	٪٠.٥		٪٨١.٦	٪١٨.٤
مستوى حجم التأثير	صغير	متوسط	كبير	الفاصل الزمني	٢٠٠٠- ٢٠٠٦	٢٠٠٧ - ٢٠١٢
	صفر	١٤٨	٥٦		١١٠	٩٤
	صفر%	٪٧٢.٥	٪٢٧.٥		٪٥٣.٩	٤٦.١
جنسية الباحث	مصري	٢٤	٪٤٨.٩	لغة الدراسة	عربية	٤٩ (١٠٠٪)
	عربي	٢٥	٪٥١.١		أجنبية	صفر (صفر%)

فقد أسفرت النتائج عن عدم اعتدالية قيم حجم الأثر المحسوب بالدراسات النفسية والتربوية؛ وهذا يبرره معامل الالتواء ٢.٢٧٣ ومعامل التفرطح ٦.٣٣٤ ويدل ارتفاع معامل التفرطح على أن عدد قليل من الدراسات النفسية كان حجم التأثير لها مرتفع، في حين أن الالتواء المرتفع يدل على تدني قيم حجم التأثير بهذه الدراسات.

كما حسب الباحث مؤشرات الاحصاء الوصفي لمؤشرات حجم التاثير في الجدول (٢) على النحو التالي:

جدول (٢): مؤشرات الاحصاء الوصفي لمؤشرات حجم التأثير لاختبار "ت".

المؤشر	اختبار كولمجروف سيمرنوف		اختبار التجامس		مقاييس الشكل		الانحراف المعياري	مقاييس النزعة المركزية		
	القيمة	د.ح	الدلالة	القيمة	الدلالة	الالتواء	تفرطح		المتوسط	الوسيط	المنوال
d	٠,٢٠	٢٠٤	٠,٠٠	٠,٤٩	٠,٠٠	٢,٢٧	٦,٣٣	٠,٥٤	٠,٩٦	٠,٤٩	٠,٤٩
r	٠,٢٠	٢٠٤	٠,٠٠	٠,٢٤	٠,٠٠	١,٣٠	١,١٧	٠,١٨	٠,٣٠	٠,٢٤	٠,٢٦
w^2	٠,٢٤	٢٠٤	٠,٠٠	٠,٠٥٣	٠,٠٠	٢,٣١	٥,٣١	٠,١٥	٠,١١	٠,٠٥٧	٠,٠١٨
γ^2	٠,٢٢	٢٠٤	٠,٠٠	٠,٠٦	٠,٠٠	٢,٢٢	٤,٩٤	٠,١٥	٠,١٢	٠,٠٦٣	٠,٠٢٤

بلغ متوسط مؤشر كوهين ٠.٩٦١ وهي قيمة متطرفة تتعارض مع معاملات الالتواء والتفرطح المحسوبة، بينما بلغ وسيط ومنوال مؤشر كوهين ٠.٤٩ وهي قيمة منخفضة كما أوضح Cohen (1988) ويبدو تطرف قيم المؤشر في ارتفاع الحد الأعظم لحجم تأثير كوهين والذي بلغ "٣.٦". وقد تم حساب الارباعيات لمضاهاة قيم التأثير وهي ٠.٣٦ ضعيف، ٠.٤٩ متوسط، ٠.٨٥٧ قوي. وهي تختلف مع تقييم Mazen et al. (1987) لوسيط حجم التأثير للبحوث الاستراتيجية لاختبار ت ٠.١٣ للتاثير ضعيف، ٠.٥٣ للتأثير متوسط، ٠.٨٨ للتأثير مرتفع؛ متوسطات تأثير دراسة Kroll & Chase (1975) (٠.١٦ ضعيف، ٠.٤٤ متوسط، والمرتفع ٠.٧٣) وهذا يعني تدني قيمة الدلالة العملية لاختبار "ت" في البحث النفسي العربي وهذا ما أكدته دراسة الصياد (١٩٨٨).

وبالتأمل في قيم الارباعيات لمؤشر r فإنه يمكن تفسير حجم التأثير ٠.٢٤ متوسط، ٠.٣٩ مرتفع وهي معايير متدنية عما حدده Rosenthal, Rosnow & Rubin (2000). وهي ٠.١٠، ٠.٢٤، ٠.٥٠، مما يعني تدني حجم التأثير في البحث النفسي العربي عن نظيره في البحث النفسي الاجنبي وهذا قد يرجع إلى عزوف الباحث الاجنبي عن اختبار فروض قرارية تافهة كتلك التي نسعى لاختبارها في المجتمع العربي، بالاضافة غلى اعتماده في حساب حجم الأثر على حدود وفترات الثقة بدلاً من الاعتماد على مستويات دلالة وأحجام عينة تقود الباحث إلى دلالة مصطنعة لا معنى لها في الواقع التطبيقي وبالتالي عدم تحقق الصدق الخارجي.

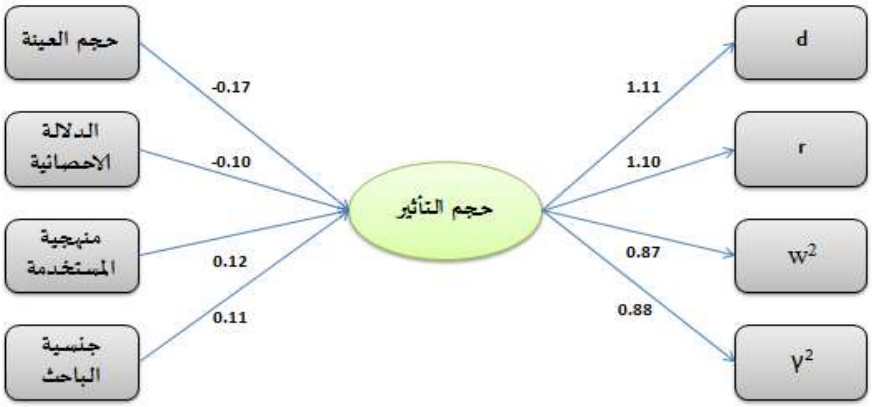
ونظراً لتطرف قيم متوسطات مؤشرات حجم التأثير فإنه يمكن تفسيره في ضوء الوسيط ومن حيث قيمة حجم التأثير فإن هذه النتيجة تتفق تماماً مع ما توصل إليه الصياد (١٩٨٨) بأن وسيط مؤشري مربع ايتا ومربع أوميجا هي على التوالي ٠.٠٦٣ و ٠.٠٥٧ لاختبار "ت" للدراسات في علم النفس للباحثين المحترفين ومتوسطهم هو ٠.١٤ و ٠.١٣ ولكن القيمة التفسيرية للدراسة الحالية تختلف مع

ما أشار إليه الصياد (١٩٨٨) في أنه قرر تفسير حجم التأثير في ضوء المتوسط الحسابي ولكن يلاحظ تطرف متوسطات حجوم الأثر في الدراسة الحالية إلا أن حجم التأثير من نوع المتوسط (في ضوء مؤشر الوسيط) في ضوء مؤشر مربع ايتا γ^2 ، ومن النوع الضعيف في ضوء مؤشر r.

للاجابة عن السؤال الأول: ما مدى مطابقة النموذج البنائي المنظم للعلاقات بين العوامل المؤثرة في حجم التأثير؟

أجري التحليل العاملي الاستكشافي Exploratory factor analysis (EFA) للمؤشرات الأربعة باستخدام طريقة المكونات الأساسية Principle component analysis (PCA) وتمحورت المؤشرات الأربعة علي عامل عام فسر ٩٧.٥ % من تباين الظاهرة كما أن كل مؤشر متقارب مع الآخر في المساهمة، وتشبع المؤشرات الأربعة علي العامل العام بلغ ٠.٩٨٦ لمؤشر كوهين d كما بلغ ٠.٩٨٠ لمؤشر r بينما بلغ ٠.٩٩٤ لمؤشر مربع w^2 وأخيراً بلغت ٠.٩٩١ لمؤشر γ^2.

واستخدام إجراء النمذجة المختلطة Mixture modeling وهو أحد تطبيقات المعادلة البنائية التي تهتم بدراسة أثر المتغيرات المتصلة والمنفصلة الكامنة أو المشاهدة على البناء العاملي لمؤشرات حجم التأثير، وقد تم معالجة بيانات الدراسة باستخدام برنامج الليزرل 8.51 Lisrel باستخدام طريقة الارجحية العظمي Maximum Likelihood (ML) والشكل (١) يوضح النموذج البنائي بمساراته الدالة:



شكل (١): النموذج البنائي المختلط المنظم للعلاقات المؤثرة على حجم التأثير لاختبار "ت".

وقد أسفرت نتائج ليزرل عن عدم مطابقة النموذج في ضوء مؤشري X^2, RMSEA وفيما يلي مؤشرات حسن المطابقة:

جدول (٣): مؤشرات المطابقة لنموذج العوامل المؤثرة في حجم التأثير لاختبار "ت".

المؤشر	RMSEA	90% CI		X^2	df (p)	NFI	NNFI
القيمة	٠.٢٠٠	(٠.٢٦٨-٠.٥٧٠)		٩١.٣٤	١٠ (٠.٠٠٠)	٠.٩٦	٠.٩٠
المؤشر	CFI	IFI	RFI	GFI	AGFI	PGFI	PNFI
القيمة	٠.٩٧	٠.٩٧	٠.٩٠	٠.٩٠	٠.٦٣	٠.٢٥	٠.٣٤

وقعت معظم مؤشرات المطابقة في المدى المثالي، إلا أن النموذج يعني عدم مطابقة البحوث النفسية العربية للمعايير المنهجية والاحصائية المتعارف عليها. إلا مؤشر RMSEA فقد يرجع إلى انتهاك بيانات الدراسة لشرط الاعتدالية للمتغير التابع.

واتفقت النتائج جزئياً مع Campbell (2004) في تفوق مؤشر كوهين عن المؤشرات الأخرى في تفسير حجم التأثير وهذا يتضح في ارتفاع قيمة التشبع من العامل الكامن (حجم التأثير) إلى مؤشر كوهين، واختلفا في الاختبار المستخدم حيث استخدمت الدراسة الحالية اختبار "ت" والآخرى استخدمت اختبار تحليل الانحدار، وتختلف أيضاً مع Olejnik & Algina (2000) اللذان قارنا بين مؤشرات (w^2, ε^2, γ^2) وقد تفوق مؤشر مربع ايتا إلا أن الدراسة الحالية اعتمدت على الدراسات المنهج الوصفي أو التجريبي أو كليهما إن وجد.

وبلغت قيمة التأثير (β = −0.17) من حجم العينة إلى حجم التأثير وهو دال احصائياً، وهذا عكس القواعد المتعارف عليها وهذا يعني شروع الباحث إلى اختيار

اختبار خاطئ سعياً وراء الدلالة الاحصائية، أو قد ترجع هذه الدلالة إلى عدم أمانة الباحث العلمية كتلفيق الباحث لبيانات بحثه ومن ثم تدني معاملات ثبات أدوات القياس وبالتالي انخفاض حجم التأثير، أو قد ترجع هذه النتيجة إلى أن الباحث العربي قد لا يراعي المعايير اللازمة لاختيار الأسلوب الاحصائي المناسب وهذا يعتبر قصوراً في الدراسات العربية وهذا يتفق مع Charles et al. (2014) الذي حدد ٥٣٪ من إجمالي الدراسات في البيئة الاجنبية راعت شروط انتقاء العينة. أو شروط اختيار الأسلوب الاحصائي الأمثل وهذا ما حدده (Brockhoff, 2003; Brewer, 1972; Keselman & Penfield, 2005; Nandy, 2012).

وترجع عدم الدلالة من حجم العينة إلى حجم التأثير بسبب حدوث التلويث الاحصائي الذي ينتج عن اختيار مقصود لبعض الدراسات التي تؤيد وجهة نظر الباحث بصرف النظر عن محددات هذه الدراسات والظروف التي تم فيها اختبار القرار وطبيعة المجتمع، و تتفق الدراسة الحالية منطقياً مع Iversson et al. (2013) تحيز الباحث في اختيار أفراد العينة من أجل وصوله إلى النتيجة القرارية التي يتبناها.

كما بلغت قيمة التأثير (β = −0.10) من مستوى الدلالة الاحصائية إلى حجم التأثير وهو دال احصائياً، وهذا يعني أن انخفاض مستوى الدلالة يؤدي إلى ارتفاع حجم التأثير وهذا يتفق كلياً مع Brewer (1972) ويختلف مع دراسة Charles et al. (2014) التي قيمت دراسات جميعها استخدمت مستوى دلالة ٠.٠٥.

وتم حساب النسب المئوية لاستخدام مستوى الدلالة الاحصائية في البحث النفسي العربي فبلغ ٦٨.١٪ التي اختارت مستوى دلالة ٠.٠١، بينما بلغت ٣١.٤٪ اختارت مستوى دلالة قيمته ٠.٠٥، وهذه النسب تفوق النسب التي حسبها الصياد (١٩٨٨) في الدراسات النفسية العربية والتي بلغت ٥٤.٣٪ لمستوى دلالة ٠.٠١، و٣٥.٨٪ اختارت مستوى دلالة ٠.٠٥، وتتفق الدراسة مع تفسير الصياد (١٩٨٨) لارتفاع نسب اختيار مستوى ألفا مساوي ٠.٠١ وهو ضبط مستوى الدلالة وبالتالي العلاقة السلبية بين حجم العينة وحجم الأثر وهذا ما اتفق عليه Slavin & Smith (2009).

وبلغت قيمة التأثير (β = 0.12) من التصميم التجريبي إلى حجم التأثير وهو دال احصائياً، وهذا يعني تفوق البحوث التجريبية عن نظيرتها الوصفية. وقد بلغ متوسط حجم العينات المستخدمة في الدراسة الحالية ٢٤٩.٧ حالة بمتوسط حجم أثر ٠.٦٩١ وهذا المعيار متفوق عن نظيره في الدراسات الاجنبية التي قيمها Weisburd et al. (1993) لنفس حجم العينة والذي بلغ القيمة ٠.١٨ للدراسات الوصفية، ٠.٢٨ للدراسات التجريبية. وبلغ ٠.٦٦٨ للدراسات الوصفية عند Carlson & Schmidt (1999).

وقد اتفقت نتائج الدراسة جزئياً مع دراسة Carlson & Schmidt (1999) والتي تفوقت على الدراسة الحالية في حجم الأثر الكلي ٠.٧٧٢ للدراسات التجريبية وهذا قد يرجع إلى تفوق دراسة في المنهجية المستخدمة حيث أعتمد على ما وراء التحليل.

وبلغت قيمة التأثير (β = 0.11) من جنسية الباحث إلى حجم التأثير وهو دال احصائياً، وهذا يعني أن كلما اختلفت قومية الباحث اختلفت نتائج دراسته وهذا يبدو في رفض بعض المجلات العلمية المحكمة لبعض البحوث النفسية والتربية وهذا الرفض قد لا يعني قصور الدراسة من الناحية الاحصائية أو المنهجية فحسب إلا أن تخصصية المجلة بات التحدي الأكبر للنشر؛ وهذا يبدو في وضع بعض المجلات العربية في وضع نموذج انسيابي يقيم في ضوئه البحث النفسي المنشور ومن أمثلة تلك المجلات المجلة المصرية للدراسات النفسية، مجلة رسالة الخليج العربي، مجلة جامعة نايف للعلوم الأمنية.

المناقشة والتعليق

توصلت الدراسة إلى أن حجم التأثير في الدراسات العربية بلغ ٧٢.٥٪ تأثير متوسط، ٢٧.٥٪ وهذا يختلف مع ما قدره Stout & Ruble (1995) لاستخدام الدلالة العملية للنتائج الأمبيريقية لاختبار "ت" لبحوث التعلم المحاسبي و ١٪ ضعيف و ٦٪ متوسط، ١٤٪ مرتفع، وما قيمه Ellis (2010) لحجوم الأثر في دراسات المجلة الدولية للعلوم التجارية والتي بلغت (٦٧٪) للأثر المنخفض، (٩٦٪) للأثر المتوسط، (١٠٠٪) للأثر المرتفع وهذه النسب للبحث العربي تفوق نظيرتها في البحث الاجنبي؛ وهذا يعني سعي الباحث وراء الدلالة العملية إلا أن منوال حجم

تأثير كوهين بلغ ٠.٤٩ وهذا يؤيد وجهة نظر الصياد (١٩٨٨) في ضبط مستوى الدلالة وزيادة حجم العينة بالاضافة إلى ذلك يرى الباحثون قصور الدلالة العملية في البحوث التربوية بمقارنتها بالمجال الطبي أو التجاري.

وبلغت عدد الدراسات التي استخدمت مستوى دلالة احصائية ٠.٠١ القيمة ٦٨.١٪، بينما مستوى الدلالة ٠.٠٥ بلغ استخدامه القيمة ٣١.٤٪ وهذا يختلف مع ما قدره Charles et al. (2014) (٩٦٪) استخدمت مستوى دلالة احصائية ٠.٠٥، وذلك يرجع لأنه قدر هذه النسبة لدراسات المجلات الطبية.

وهذا يتعارض مع متوسطات حجم تأثير كوهين الذي قدرته دراسات Sawyer & Ball (1981) والتي بلغت ٠.١٠ ضعيف، ٠.٣٠ متوسط، ٠.٥٠ مرتفع، وهذا يعني تضخم مؤشرات كوهين للدراسات العربية وهذا يعني سعي الباحث العربي وراء دلالة عملية مصطنعة قد تنتج من صرامة التصميم أو تحيز في اختيار العينة أو ضبط مستوى الدلالة أو تلفيق بيانات الدراسة.

واتضح من نتائج الدراسة قصور الدراسات النفسية في البيئة العربية في مجال حجم التاثير بالرغم من تضخمها وذلك يرجع لصرامة ضبط التصميم التجريبي الذي يتبناه الباحث مع العلم بأنه يمكن استخدام أساليب احصائية مثلى لتحقيق مثل هذا الضبط مثل اختبار تحليل التغاير الاحادي وتحليل التغاير المتعدد، بالاضافة إلى عشوائية الباحث العربي في انتقاء متغيراته المستقلة، بالاضافة إلى ضبط الباحث لمستوى الدلالة من عدمه والسعي وراء مستويات دلالة تجعله يقبل بالفرض البديل، بالاضافة إلى أنه قد لا يعبأ بتحكيم أدوات بحثه مما يؤدي إلى الاضرار بثباتها، وأخيراً التحيز الذي يقع في الباحث في انتقاء أفراد المجموعة التجريبية.

وقدمت الدراسة إسهاماً هاماً في البيئة العربية في مجال النمذجة المختلطة التي تدرس أثر بعض المتغيرات على البنية العاملية لمجموعة من العوامل الكامنة، إلا أن الدراسة تعاني من بعض المحددات منها: الاقتصار على الدراسات التي استخدمت اختبار ت، لم يتم تحديد المجتمع لكل الدراسات النفسية والتربوية في البحث النفسي العربي، اقتصرت على تخصص علم النفس التربوي والصحة النفسية وإهمال بقية التخصصات التربوية.

المراجع

زكريا أحمد الشربيني. (٢٠٠١). الاحصاء اللابارامتري مع استخدام Spss في العلوم النفسية والتربوية والاجتماعية. القاهرة: مكتبة الانجلو المصرية.

عادل أحمد بابطين. (٢٠٠٢). مشكلات الدلالة الاحصائية في البحث التربوي وحلول بديلة.جامعة أم القري. كلية التربية. قسم علم النفس.

عبد العاطي أحمد الصياد. (١٩٨٥). النماذج الاحصائية في البحث التربوي والنفسي والعربي بين ما هو قائم وما يجب ان يكون – دراسة تقويمية للواقع الاحصائي للنماذج الاحصائية. رسالة الخليج العربي ، مكتب التربية العربي لدول الخليج بالرياض، ١٦، ٥.

عبد العاطي الصياد وعبد الناصر عامر. (٢٠٠٨). الاستشاري في الاحصاء التربوي ومناهج البحث. مذكرة غير منشورة. كلية التربية بالاسماعيلية. جامعة قناة السويس.

عبد العاطي الصياد ومني الألفي. (٢٠٠٩). العوامل والمتغيرات النفسية والاجتماعية في دراسات وبحوث التورط في المخدرات: دراسة نقدية للمنهج. ورقة منشورة عرضت بندوة المخدرات والامن الاجتماعي المنعقدة في كلية التربية بالسويس في الفترة من ٣٠ مارس وحتي ١ أبريل.

عبد العاطي الصياد. (١٩٨٨).الدلالة العملية وحجم العينة المصاحبتين للدلالة الاحصائية لاختبار "ت" في البحث التربوي و النفسي والعربي: دراسة تقويمية. بحوث مؤتمر البحث التربوي الواقع والمستقبل، ٢.

عبد الله النجار. (٢٠٠٦). دراسة تحليلية لقوة الاختبار الاحصائي في البحوث الادارية المنشورة .المجلة العلمية لجامعة الملك فيصل. المجلد السابع، العدد ٢.

Breaugh, J. A. (2003). Effect size estimation: factors to consider and mistakes to avoid. *Journal of management*, 29, 1, 79-97.

Brewer, J. K. (1972). On the power of statistical tests in the American Educational Research Journal. *American Educational Research Journal*, 9, 3, 391- 401.

Brockhoff, P. B. (2003). The statistical power of replications in difference tests. *Food Quality and Preference*, 14, 405- 417.

Burk, A.R. (2005). New trends in ecology research. New York: Nova science publishers.

Cohen, J. (1988). Statistical power analysis for behavioral sciences.

Ellis, P. D. (2010, December). Effect sizes and interpretation of research results in international business. *Journal of International Business*, 41, 9, 1581- 1588.

Florey, C. D. V. (1993, May). Sample size for beginners. *British Medical Journal*, 306, 6886, 1181- 1184.

Hubbard, R., & Armstrong, J. S. (2005). Why we don't really what "statistical significance" means: a major educational failure. Available at: marketing.wharton.upenn.edu/ideas/pdf/Armstrong /statisticalsignificance.pdf

Huston, H. L. (1993). Meaningfulness, statistical significance, effect size, and power analysis: a general discussion with implications for MANOVA. Paper presented at annual meeting of the Mid-South Educational Research Association (2nd, New Orleans, LA, November 9-12.

Keselman, H. J., & Penfield, R. D. (2005). Effect Sizes and their Intervals: The Two-Level Repeated Measures Case. *Educational and Psychological Measurement*, 65, 2, 241- 258.

Kirk, R. (1996). Practical significance: a concept whose time has come. In Breaugh, J.A. (2003). Effect size estimation: factors to consider and mistakes to avoid. *Journal of management*, 29, 1, 79-97.

Kroll, R. M., & Chase, L. J. (1975). Communication disorders: a power analytic assessment of recent research. *Journal of Communication Disorders*, 8, 237- 247.

Mazen, A. M., Hemmasi, M., & Lewis, M. F. (1987). Assessment of statistical power in contemporary strategy research. *Strategic Management Journal*, 8, 4, 403- 410.

Nandy, K. (2012). Understanding and quantifying effect sizes. Unpublished lectures. School of Nursing, Department of Biostatistics, University of California, Los Angeles (UCLA)

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25, 241- 286.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single case research: nonoverlap of all pairs. *Behavior Therapy*, 40, 357- 367.

Pillai, K. G., Katsikeas, C. S., & Presi, C. (2012). Print advertising: Type size effects. *Journal of Business Research*, 65, 865- 868.

Rosenthal, R. (1994). Parametric measures of effect size. In Cooper, H., & Hedges, L.V. (Eds). The handbook of research synthesis. New York: Russel Sage.

Sawyer, A. G., & Ball, A. D. (1981, August). Statistical power in marketing research. *Journal of Marketing Research*, 18, 3, 275- 290.

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31, 4, 500- 506.

Thompson, B. (1999). Improving Research Clarity and usefulness with effect size Indices as supplement to statistical significance tests. *Exceptional children*, 65, 3, 329-337.

Woodside, A. G., Beretich, T. M., & Lauricella, M. A. (1993). A Meta analysis of effect sizes based on direct marketing campaigns. *Journal of Direct Marketing*, 7, 2, 19- 34.

النمذجة البنائية للعلاقات بين المتغيرات المؤثرة في حجم التأثير لاختبار "ت" في البحوث النفسية العربية

مستخلص:

هدفت الدراسة إلى التحقق من التأثيرات السببية لكل من المتغيرات المؤثرة في حجم التأثير لاختبار ت المستقلة والمرتبطة في البحوث العربية. وحددت الدراسة المتغيرات حجم العينة والدلالة الاحصائية ومنهجية الدراسة وجنسية الباحث كمتغيرات مستقلة مشاعدة وحجم التأثير كمتغيرات كامن تابع بينما اعتبرت المؤشرات d و r و γ^2 و w^2 مؤشرات لمتغير حجم التأثير. بلغ حجم العينة ٤٩ دراسة من المجلات العربية والمصرية في المدى الزمني من ٢٠٠٠ حتى ٢٠١٢. واعتمدت الدراسة على المنهج الوصفي الارتباطي لاختبار النموذج البنائي للعلاقات بين متغيرات الدراسة. وتوصلت الدراسة إلى وجود تأثيرات سببية سالبة من حجم العينة والدلالة الاحصائية إلى متغير حجم التأثير. بينما وجدت تأثيرات سببية موجبة من منهجية الدراسة وجنسية الباحث إلى حجم التأثير.

الكمات المفتاحية: النمذجة البنائية، حجم التاثير، اختبار ت.

Structural equation modeling of variables influencing on t-test effect sizes' in Arabian Psychological researches

Abstract:

This study aimed as testing the structural effects of variables influencing the effect size of t-test techniques. Sample size, statistical significance, research method, researcher nationality specified as the observed independent variables and effect size specified as a dependent variable. 49 Arabian and Egyptian studies had been selected in educational psychology and mental health. The time scale of those study in which years 2000 and 2012. Descriptive Correlational approach was drawn. The study revealed that there were negative statistical causal effects from sample size and statistical significance on effect size. Finally there were positive statistical causal effects from research methods and researcher nationality on effect size.

Keywords: structural equation modeling, effect size, t-tests.