الجمهورية العربية السورية

جامعة دمشق

كلية التربية

قسم القياس والتقويم التربوي والنفسي

واقع الدلالات العلمية في الأبحاث التجريبية وأثر التدريب فيها وفي الخصائص السيكومترية للاختبارات

أُطروحة مقدمة للحصول على درجة الدكتوراه في القياس والتقويم التربوي والنفسي

إعداد الطالبة

سُميه حسين سويد

إشراف

د. إيمان إبراهيم عز

الأستاذة في قسم القياس والتقويم التربوي والنفسي

1443-1444هـ

2022/2021 م

Damascus University
Faculty of Education
Educational and Psychological
Measurement and Evaluation Department

The Reality of Scientific Significances in Experimental Research and the Effect of Training in them and in the Psychometric Properties of Tests

A Dissertation submitted to obtain the Doctor's Degree in Educational and Psychological Measurement and Evaluation

Prepared By

Sumaiah Houssin Swaid

Supervised By

Dr. Iman Ibrahim Ezz

A Professor in Educational and Psychological
Measurement and Evaluation Department

1443 -1444 H.A
2021- 2022 D.A

الفصل الأول

الإطار المنهجي للبحث

الفصل الأول

الإطار المنهجي للبحث

مقدمة:

تُشكل الدلالات العلمية بأنواعها الثلاث: الإحصائي، العملي، والعيادي مفاهيماً رئيسة في أي بحثٍ علمي، بل هناك تأكيدٌ بضرورة اعتمادها جميعاً في الأبحاث العلمية للوصول إلى قراراتٍ علميةٍ صحيحةٍ ودقيقةٍ قدر المستطاع.

من المعروف تماماً، أن أي بحثٍ علميٍّ كمي «أكاديميٍّ كان أم غير أكاديمي» ينطوي على جزأين اثنين: أحدهما نظري، وآخرٌ ميداني، ورغم أهمية الجانب النظري، تبقى للميداني أهمية نوعية وكمية أخرى، فأي باحث يحتاج لتنفيذه إلى أدوات قياس وتقويم، وإلى عينات تجمع منها البيانات، ومعالجات إحصائية لها ليصل في النهاية إلى نتيجةٍ محددةٍ، وإلى قرارٍ علميٍّ بشأنها، وهو في هذا كله يعتمد أطراً مرجعية إحصائية للحكم على جودة قراره الإحصائي والبحثي في آن واحد، وغالباً ما تكون الدلالات العلمية أحد هذه الأطر، وإن كانت الدلالة الإحصائية هي الأكثر استخداماً وانتشاراً بالمقارنة مع الدلالتين الأخريين، بل قد تكون في كثير من الأحيان هي الوحيدة المستخدمة في غالبية الأبحاث.

ظهر مقابل التوجه لاعتماد مؤشرات الدلالة الإحصائية فقط في القرار الإحصائي، تأكيدٌ علمي من عدد من الجهات العلمية المختصة وأدلة النشر، أنها غير كافية وحدها لاتخاذ القرار الإحصائي والبحثي، فقد شدد دليل نشر الأبحاث السابع الصادر عن الجمعية النفسية الأمريكية "APA" على أن الدلالة الإحصائية ليست سوى نقطة الانطلاق في حساب نتائج الأبحاث، ولابد أن تتبع بالدلالة العملية للأبحاث الكمية التجريبية وغير التجريبية، وبالعيادية في الأبحاث التجريبية. ولعل هذا التشديد لم يختص فقط بمعالجة النتائج، وإنما بكل إجراء فيه معالجات إحصائية، سواء ارتبط بأداة البحث، أو حجوم العينات، أو معالجة البيانات وتحليلها.

من هنا كان البحث الحالي محاولةً لبيان واقع الدلالات العلمية بأنواعها الثلاث، وأهمية اعتمادها، بغية الوصول لقراراتٍ بحثيةٍ علميةٍ أكثر دقة وصحة، من خلال عينة من الأبحاث التجريبية أو ذات التصميم التجريبي المنجزة، من جهة، وعبر برنامج تدريبي يُظهر مدى تأثير التدريب فيها وفي الخصائص القياسية[1] للاختبار، التي تعتمد أساساً على القرار الإحصائي لتحقق مؤشراتها، من جهةٍ أخرى.

[1] سيعتمد مصطلح الخصائص القياسية في المتن تماشياً مع تعريبه، على الرغم من استخدام مصطلح الخصائص السيكومترية في نص العنوان

1- مشكلة البحث:

يعتمد الجانب الميداني في أي بحث كمي مجموعة عناصر، يأتي في مقدمتها الأفراد الذين تدرس عندهم الظواهر أو المتغيرات، وأدوات القياس التي تكشف مقدار ما يمتلكون من هذه المتغيرات، والبيانات التي تمثل القياسات التي تم الحصول عليها نتيجة ذلك، إضافةً إلى المعالجات الإحصائية التي تعالج بها هذه البيانات للوصول إلى نتيجةٍ تبيّن تحقق الفرضية من عدم تحققها، وتجيب عن الأسئلة البحثية المختلفة التي يصمم البحث أساساً من خلالها.

ويُعد القرار الإحصائي جذعٌ مشترك الاستخدام والأهمية بين هذه العناصر جميعها، فمن خلاله يتم بيان جودة العينة المستخدمة من حيث حجمها وتمثيلها للمجتمع الأصلي، وخصائصها الديموغرافية المختلفة، وباعتماده يُكشف مدى صلاحية أدوات القياس، وتمتعها بالخصائص القياسية اللازمة والضرورية للاطمئنان لاستخدامها، وبالاستناد إليه تتم معالجة البيانات واتخاذ القرار البحثي المرتبط بالفرضيات والأسئلة البحثية التي تنطلق منها الأبحاث عموماً.

حتى وقتٍ ليس بالبعيد كان مفهوم الدلالة الإحصائية Statistical Significance (Ss) المفهوم الوحيد المعتَمد لاتخاذ القرار الإحصائي عند التحقق من الفرضيات الإحصائية، وبالأخص الفرضية الصفرية التي تعني بصورتها العامة مدى إمكانية تعميم نتيجة الفرضية الصفرية للعينة على المجتمع الأصلي (بحاش، 2020، ص. 377)، إذ تقارن مقادير زيادة نتائج المعالجات الإحصائية أو نقصانها مع مستويات دلالة إحصائية محددة، ليتم اتخاذ القرار البحثي فيما بعد بشأن تحقق الفرضيات، والإجابة عن الأسئلة.

لقد بين العديد من اختصاصي الإحصاء والقياس، أنه رغم أهمية الدلالة الإحصائية في أي قرار إحصائي، إلا أنها غير كافية لتحقيق دقته المطلوبة، ودقة النتيجة البحثية المعتمدة عليه، بسبب تأثُّرها بعوامل عديدة، من مثل، حساسيتها لحجوم العينات وتجانسها، وتشتت البيانات، واعتمادها قيماً محددة لمستوى الدلالة (Mclean & Ernest, 1997, pp.7-10). لذا لم يكن من المستغرب دعوة الجمعية الأمريكية للإحصاء ASA عام (2016) الباحثين إلى التوقف الفوري عن استخدامها كأداةٍ وحيدةٍ عند اتخاذ القرارات الإحصائية، مؤكدةً عبر الكثير من الدراسات، عدم كفايتها كمؤشر ودليل وحيد على دقة النتائج، ومشددةً على ضرورة تبنّي مستويات أكثر دقة من الدلالات العلمية Scientific Significances (SS) للوثوق بالنتائج المستخلصة (Hansen et al., 2018, P.802). كمفهومي الدلالة العملية Practical Significance (PS)، والدلالة العيادية Clinical Significance(CS).

ظهر مفهوم الدلالة العملية الممثل بحجم الأثر(Effect Size (ES، كوجهٍ مكملٍ للدلالة الإحصائية (Huston, 1993,p.11)، يتم من خلاله معرفة حجوم الفروق، أو قوة العلاقات بين المتغيرات، وبه يُقرر إلى أي مدى يمكن تفسير التغيُّر الذي يحدث في المتغيّر التابع بواسطة المتغيّر أو المتغيّرات المستقلة التي ترتبط به أو تؤثر عليه (محمد حسن، 2011، ص.265).

وبذلك، بدأ توجهٌ علميٌّ جديد اعتمدته العديد من الجامعات ومراكز الأبحاث وأدلة النشر العالمية، تؤكد من خلاله ضرورة وجود مؤشرات الدلالة العملية في أي بحث كمي، أكان تجريبياً أم غير تجريبي[2]. مما يعني أن حجوم الأثر لابد أن تُرصد في الأبحاث الكمية التي تعتمد على أشخاص مشاركين أو على أفراد عينات –أكانت احتمالية أم غير احتمالية– عند استخراج بياناتها ومعالجتها إحصائياً، بما يتناسب مع طبيعة هذه العينات وحجومها من ناحية، وطبيعة المتغيرات وتوزع بياناتها من ناحيةٍ أخرى، لما لها من دورٍ مهم في حساب القوة الإحصائية Statistical Power، وفي إجراء المقارنات الكمية بين نتائج دراستين أو أكثر، وكونها تُعد أداةً مهمة في منهج التحليل البعدي Meta - Analysis (Ferguson, 2009, P.P 532-533, Thompson, 2004, P.480) الذي يسعى للوصول إلى استنتاجات عامة حول موضوع ما، استناداً إلى نتائج عدد كبير من الأبحاث المنجزة سابقاً.

يعتمد حساب الدلالتين الإحصائية والعملية على إحصاءات العينة كاملةً، باعتبارها كتلةً واحدة، ولا يقف عند إحصاءات كل فرد من أفرادها. وهذا قد يكون كافياً في الأبحاث الكمية غير التجريبية، ولكنه قد لا يكون كذلك في الأبحاث التجريبية، فغالباً ما يكون الهدف الرئيس لهذه الأبحاث أياً كانت: علاجية، أم إرشادية، أم تعليمية، أم تأهيلية، مساعدة الأفراد لامتلاك معارف ومهارات وقدرات نوعية، أو تنميتها بما يضمن لهم حياة أفضل. وهي، والحال هذه، إنما تسعى لبيان مدى فاعلية المعالجات والبرامج التي تبنيها في إحداث التغيُّر الإيجابي والمطلوب لدى كل فرد من أفراد مجموعتها التجريبية، وليس على مستوى المجموعة ككل. ولكن واقع الحال غير ذلك، فهي بالنتيجة تقدم نتائجها على كامل أفراد العينة أو المجموعة، دون بيان بالتغير الحاصل على مستوى كل فرد. مما يعني، حاجة هذه الأبحاث إلى نوع آخر من الدلالات غير الإحصائية والعملية تأخذ بالحساب فعالية أو تأثير البرنامج التدريبي على أداء كل فرد على حده. فظهور التغير – أو عدم ظهوره– للمجموعة كاملة لا يعني أنه قد حدث فعلاً لدى كل فرد، وبالمقدار ذاته. فقد تصل هذه الأبحاث إلى دلالات إحصائية وعملية مهمة، ولكنها لا تملك أدلة كافية لوجود مؤشرات تغيير حقيقي، أو لدلالة فعلية على مستوى الفرد.

[2] وذلك على اعتبار أن تصاميم البحث الكمي حسب دليل النشر APA السابع تقسم إلى ستة أنواع رئيسية، يمكن وضعها في مجموعتين: الأولى هي التصاميم التجريبية وتتضمن ثلاثة أنواع هي: التصاميم التجريبية التامة، التصاميم قبل التجريبية والتصاميم شبه التجريبية. والثانية هي التصاميم غير التجريبية وتتضمن الأبحاث الوصفية والمسحية والارتباطية

لذا بدأ الاهتمام بضرورة رصد هذا التغيُّر على مستوى كل فرد، مهما كانت مؤشرات الدلالتين، وظهر – ضمن التشديد ذاته على ضرورة الحصول على دقة أكبر للقرار الإحصائي والبحثي– مفهوم الدلالة العيادية أو السريرية في الدراسات الطبية، أو مؤشر التغيُّر الموثوق Reliable Change Index في الدراسات النفسية والتربوية[3]. فقد تكون نتائج اعتماد الدلالتين الإحصائية والعملية مهمة ودالة، ولكن نتائج مؤشر التغيُّر الموثوق غير مهم، أو قليل الأهمية، وقد يكون العكس تماماً، كأن تكون الدلالتين الإحصائية والعملية غير مهمة، ولكن مؤشر التغيُّر الموثوق مهم جداً. وهذا كله، يعني، أنه في مرحلةٍ ما من مراحل البحث التجريبي قد يحدث الخطأ من النوع الأول، فيُرفض الفرض الصفري بينما هو صحيح، وفي مرحلة أخرى، قد يحدث الخطأ من النوع الثاني، فيُقبل الفرض الصفري بينما هو خاطئ، سواءٌ على مستوى المجموعة كاملة، أو على مستوى كل فرد من أفرادها.

من زاويةٍ أخرى، يبدو أن الأبحاث الكمية كلها تعتمد على أدوات قياس لجمع البيانات عن المتغيرات المدروسة، ولابد أن تتوافر فيها الخصائص القياسية اللازمة. ومما لا شك فيه هنا، أن القرار الإحصائي بمدى تمتع أداة القياس بالخصائص المطلوبة، لا يختلف عن أي قرار إحصائي يتعلق بنتائج المعالجات الإحصائية من حيث الأهمية والاعتماد. وبالتالي، لابد أن يستند إلى مستويات عدة للدلالة العلمية، شأنه في ذلك، شأن القرارات الإحصائية الأخرى.

لقد بين (Dodeen,2009) أن الاختبارات بأنواعها المختلفة توفر نتائج لنوعين مستقلين من المعرفة عن المفحوص، يرتبط الأول بمقدار ما يمتلكه من السمة المقاسة، بينما يتعلق الثاني بسلوكه وأدائه في الموقف الاختباري ومدى توظيفه للمهارات التي تمّكّنه من التصرف بفاعلية أثناء إجرائه (Dodeen,2009, p.36). وهذا الموقف الاختباري قد يكون عند تجريب الاختبار في أحد مراحل بنائه المتعلقة بالتحقق من خصائصه القياسية، وقد يكون عند استخدامه لتحقيق الأغراض المرجوة منه فيما بعد. ورغم أهمية كلا الموقفين، يبدو أن الموقف الأول هو الأهم لأنه الأساس للوثوق بالاختبار واستخدامه لاحقاً.

تتأثر الخصائص القياسية لأي اختبار بعدد من العوامل المعرفية والنفسية، كمستوى دافعية الأفراد المشاركين في عينات صدق الاختبار وثباته، وجديتهم، وضبطهم الذاتي، وتصرفاتهم، وتوظيفهم لاستراتيجياتهم الخاصة في الموقف الاختباري، من مثل قدرتهم على إدارة الوقت، وفهمهم لتعليمات الاختبار وأسئلته بالشكل الصحيح.

فنقص الدافعية وعدم معرفة المفحوص باستراتيجيات الأداء الملائمة في الموقف الاختباري يضعف من قدرته على تأكيد امتلاكه للمعرفة المطلوبة في الاختبار، ويُنقص من درجته المستحقة، مما قد ينتج

[3] تستخدم الباحثة في البحث الحالي مصطلحي مؤشر التغيُّر الموثوق، والدلالة العيادية، للتعبير عن المفهوم ذاته

عنه عدم تعبيرها بصدق عن المستوى الحقيقي للسمة المدروسة لديه (Pohl et al.,2016, p.3, Barry and Finney, 2009, pp. 4-5).

لقد عدّ ثورندايك (Thorndike, 1951) التدريب على حكمة التصرف في الموقف الاختباري مصدراً مهماً وعاملاً مفسراً يعزى له جزء كبير من تباين درجات الأفراد في الاختبارات المختلفة، وبيّن أن تدريب المفحوص على أداء الاختبار يمكن أن يؤدي إلى زيادة صدقه وثباته؛ من خلال التقليل من التباين غير المفسر في الدرجات (أبو حطب وآخرون، 2003، ص ص. 103- 104، Banks and Eaton, 2014, p.207). مما يعني من زاويةٍ أخرى، أن الخصائص القياسية قد تتغير إذا تم تدريب الأفراد على كيفية أداء الاختبار، والتصرف الصحيح في الموقف الاختباري من دون أن تتغير بنوده، أو يتعدل محتواها أو صياغتها، كما أن القيم الإحصائية التي ستظهر قد تتغير، وقد تختلف دلالاتها إذا تم اعتماد مستوى دلالة علمي آخر إلى جانب الدلالة الإحصائية.

لقد كشفت معظم الأبحاث التجريبية التربوية والنفسية المحلية والعربية والعالمية، على مدى السنوات العشرين الأخيرة– وقد يكون قبل ذلك– عن قراراتٍ ونتائج بحثية هامة وواضحة عن فعالية برامجها استناداً للدلالتين الإحصائية والعملية على مستوى المجموعات، دون أن ترصد أية مؤشرات عن مقدار التغيُّر الحقيقي على مستوى الأفراد داخل المجموعات التجريبية، مما قد يثير تساؤلاً بحثياً مهماً حول ما إذا كانت هذه القرارات دقيقة فعلاً في إثبات فاعلية هذه البرامج في إحداثها التغير المطلوب لدى جميع أفراد المجموعة التجريبية، أم أن الأمر غير ذلك.

إلى جانب ذلك كله، يبدو أن جميع ما يتضمنه الأدب النظري والميداني في خطوات بناء المقاييس وإعدادها–الذي راجعته الباحثة و بلا استثناء– يُظهر أن الدلالة الإحصائية، هي الوحيدة فقط، المعتَمدة في تحديد مؤشرات الخصائص القياسية دون غيرها من الدلالات. وبالتالي، تبدو مسألة دقة القرار الإحصائي في كشف الخصائص القياسية للمقاييس وعلاقته بالدلالات العلمية الأخرى بحاجة إلى مزيد من الدراسة والبحث شأنها في ذلك شأن أي قرار إحصائي آخر.

ولعل هذا ما كان فعلاً، إذ دفعت أهمية أنواع الدلالات العلمية الثلاث في الأبحاث الكمية عدداً من الباحثين إلى ضرورة إخضاعها لمزيد من الأبحاث والدراسات التي انتهت في معظمها إلى تأكيد أهميتها بشكلٍ أو بآخر، وضرورة اعتمادها. غير أن مراجعة الأدب النظري حولها تبين أن بعض هذه الأبحاث استهدف التمييز بينها، وبيان أهمية كل منها في الأبحاث مثل بحث بالكين ولينز (Balkin & Lenz, 2021)، (سكوبر وآخرين Schober & et al., 2018) ، (نصار، 2016)، وبعضها الآخر اختص بالتركيز على موضوع الدلالة العيادية، والمقارنة الإحصائية بين الطرائق المختلفة لحساب نتائجها، مثل

رونك وآخرون (Ronk & et al., 2016)، شدا (Shada, 2013)، أتكينز (Atkins & et al, 2005)، باور وآخرون (Bauer & et al, 2004)، بينما تناول البعض الثالث واقع استخدام الدلالتين الإحصائية والعملية في الأبحاث مثل (بحاش، 2020)، (وهبي وآخرون، 2020)، هانسن وآخرون (Hansen & et al., 2018)، (عصفور، 2018)، ولكن لم يكن هناك – من جميع ما اطلعت عليه الباحثة – أي بحثٍ استهدف تحليل واقع استخدام الدلالات الثلاث معاً، أو دراسة العلاقات بينها جميعاً والقدرة التنبؤية لكل نوع منها بالنوعين الآخرين، وبيان مدى كفاية كل منها سواءٌ في الإبلاغ عن نتائج الأبحاث أو الخصائص القياسية للاختبارات، ومدى تأثرها بعوامل تتعلق بالموقف الاختباري أو التدريبي الذي يخضع له المفحوصون. مما يعني ضرورة إخضاعها لمزيد من البحث علّه يتم الوصول إلى مزيد من المعرفة عنها، وبما قد يسهم في إغناء أدبيات البحث النظرية والعملية حولها، أكانت بتأكيد اعتمادها أو التخلي عنها. ولعل البحث الحالي يستطيع أن يقدم مساهمة علمية تضاف إلى ذلك المخزون العلمي في ميدان الدلالات العلمية والقرار الإحصائي.

يؤسس البحث الحالي معالجته لموضوع الدلالات العلمية على منحيين اثنين: يرتبط الأول بالقرار الإحصائي الخاص بنتائج الأبحاث التجريبية، بينما يتعلق الثاني بالقرار الإحصائي حول الخصائص القياسية للاختبار، في ضوء افتراضين بحثيين تنطلق منهما الباحثة: يتلخص الأول "بأنه يمكن الوصول إلى دقة أكبر للقرار الإحصائي المتعلق بنتائج الأبحاث التجريبية وبالخصائص القياسية لأدوات القياس إذا تم الاعتماد على الدلالات العلمية الثلاث"، ويتلخص الثاني "بأن متغير التدريب قد يؤثر في نتائج هذه الدلالات، وفي رفع أو خفض مؤشرات الخصائص القياسية للاختبار في ضوء نتائج الدلالات العلمية".

إنه يسعى لبيان مدى كفاية كل نوع من أنواع الدلالات العلمية في الإبلاغ عن نتائج الأبحاث التجريبية وأهمية التكامل فيما بينها، من ناحية، ولمعرفة كيف يمكن لمتغير التدريب أن يؤثر في نتائجها، وفي رفع أو خفض مؤشرات الخصائص القياسية للاختبار، من دون تغيير محتواه، من ناحيةٍ ثانية. مع تأكيد أن الهدف الأساسي هنا، ليس التحقق من الخصائص القياسية للاختبار، وإنما بيان مقدار تغيُّر هذه الخصائص بتأثير متغير التدريب.

استناداً لكل ما سبق، تحدّدت مشكلة البحث الحالي بكشف واقع الدلالات العلمية في الأبحاث التجريبية، وبيان أثر متغير التدريب في نتائجها وفي الخصائص القياسية للاختبار.

2- أهمية البحث:

تتجلى أهمية البحث الحالي في النقاط الآتية:

- الحاجة للوصول لدقةٍ أكبر في القرار الإحصائي والبحثي عند قبول نتائج الأبحاث، والتحقق من مؤشرات الخصائص القياسية للاختبارات والمقاييس، باعتماد مستويات ودلالات إحصائية علمية متعددة، وعدم الاكتفاء بالدلالة الإحصائية كنقطة مرجعية وحيدة في الإبلاغ عن نتائج الأبحاث.
- أهمية الكشف عن التغيرات التي يمكن أن تحصل لدى كل فرد من أفراد العينة، إلى جانب الكشف عن التغيرات التي تحصل على كامل العينة، في الأبحاث التجريبية مهما كان نوعها: تدريبية، تعليمية، علاجية، بما يلائم طبيعة المتغير المدروس، وطبيعة البرنامج المستخدم ومحتواه.
- إمكانية الوصول لتنظيم إرشادات استخدام قوانين الدلالات العلمية في الأبحاث، بما يمكّن من اعتمادها كإجراء في الأبحاث المستقبلية، مع إعادة النظر في إمكانية اعتمادها في الأبحاث المنجزة إن كان ذلك ممكناً بما يكفل الوصول إلى تأكيد دقة النتائج.
- يُعد البحث الحالي الأول من نوعه – في حدود علم الباحثة – الذي يبحث في الدلالات العلمية الثلاث معاً في نوعين من الأبحاث الكمية: الوصفية التحليلية والتجريبية، من ناحية، ويعتمد منهج التحليل البعدي للوصول إلى استنتاجات علمية حول الدلالات العلمية، وإمكانية التنبؤ بها من خلال التكامل بينها، من ناحية ثانيةٍ، ودراسة أثر التدريب على دقة القرار الإحصائي المعتمد في الدلالات العلمية الثلاث، وفي استخراج مؤشرات الخصائص القياسية للاختبار، من ناحيةٍ ثالثة.

3- أهداف البحث:

يسعى البحث الحالي إلى قياس:

- واقع استخدام الدلالات العلمية في الأبحاث الأكاديمية التجريبية: الماجستير والدكتوراه المنجزة في كلية التربية.
- مدى كفاية الدلالات العلمية في دقة القرار الإحصائي.
- القدرة التنبؤية لنتائج كل نوع من أنواع الدلالات العلمية بنتائج النوعين الآخرين.
- أثر متغير التدريب في نتائج الدلالات العلمية.
- طرائق حساب مؤشرات التغيُّر الموثوق الملائمة لنوع البرنامج وطبيعته، والتفاضل فيما بينها.
- أثر متغير التدريب في مؤشرات الخصائص القياسية للاختبار.

4- أسئلة البحث:

تتعدد الأسئلة التي يحاول البحث الإجابة عنها، وهي:

السؤال الأول: ما واقع استخدام الدلالات العلمية في الأبحاث الأكاديمية التجريبية المنجزة في كلية التربية – جامعة دمشق؟

السؤال الثاني: ما مدى كفاية الدلالة الإحصائية في الإبلاغ عن نتائج تأثير البرامج التجريبية؟

السؤال الثالث: ما مدى كفاية الدلالة العملية في الإبلاغ عن نتائج تأثير البرامج التجريبية؟

السؤال الرابع: ما إمكانية التنبؤ بنتائج أي نوع من الدلالات العلمية من خلال نتائج النوعين الآخرين؟

السؤال الخامس: إلى أي مدى يُغيّر التدريب في مؤشرات الدلالات العلمية لنتائج أفراد العينة على اختبار الذكاء[4]؟

السؤال السادس: إلى أي مدى تختلف قيم مؤشر التغيُّر الموثوق وتصنيفاته العيادية لدرجات أفراد العينة التجريبية على اختبار الذكاء باختلاف طريقة حسابه؟

السؤال السابع: إلى أي مدى يُغيّر التدريب في مؤشرات الخصائص القياسية لاختبار الذكاء؟

5- منهج البحث:

لتحقيق أهداف البحث والإجابة عن أسئلته اعتُمدت ثلاثة مناهج بحثية هي: المنهج الوصفي التحليلي، ومنهج التحليل البعدي، والمنهج التجريبي.

6- مجتمع البحث وعينته:

مجتمع البحث:

يقتصر مجتمع البحث على مجتمعٍ إحصائيٍّ واحد يرتبط بمنهجي البحث الوصفي التحليلي، والتحليل البعدي، ويتمثل بجميع رسائل الماجستير والدكتوراه ذات التصميم التجريبي المنجزة في كلية التربية جامعة دمشق بالأقسام الآتية: المناهج وطرائق التدريس، علم النفس، الإرشاد النفسي، والتربية الخاصة، بين عامي 2000- 2020 الميلاديين، والفصل الأول من العام الدراسي 2020 – 2021م.

عينة البحث: تتكون من عينتين أساسيتين هما:

- عينة الرسائل الأكاديمية: تتألف من جميع الرسائل الأكاديمية التجريبية في أقسام المجتمع الأصلي المحددة ذاتها، خلال الفترة الزمنية الممتدة بين عامي 2000 – 2020 الميلاديين، والفصل الأول من العام الدراسي 2020 – 2021م.
- عينة البحث التجريبية: وهي عينة عرضية متيسرة تتكون من طلبة السنة الثانية من اختصاص علم النفس، عدد أفرادها (125) طالباً وطالبةً.

7- متغيرات البحث:

تتوزع متغيرات البحث الحالي إلى نوعين أساسيين هما:

- المتغيرات المستقلة: تتمثل بمتغير التدريب على الاختبار.
- المتغيرات التابعة: وهي قيم كلٍ من:
  - الدلالات العلمية للاختبار.
  - الخصائص القياسية للاختبار.

[4] يُقصد باختبار الذكاء: اختبار المصفوفات المتتابعة رافن- النسخة الموازية: SPM-PV

8- أدوات البحث:

لتحقيق أهداف البحث والإجابة عن تساؤلاته جرى استخدام الأدوات الآتية:

- اختبار المصفوفات المتتابعة رافن- النسخة الموازية: SPM-PV.
- اختبار الذكاء غير اللفظي Toni -4 الصورة A.
- البرنامج التدريبي الذي أعدته الباحثة حول استراتيجيات أداء الاختبار.

9- مصطلحات البحث العلمية:

- **الواقع Reality :**

ورد في معجم المعاني الجامع أن واقع الأمر يعني وضعه الحقيقي، كما ورد في معجم اللغة العربية المعاصرة أن الأمر الواقع هو: الحقيقي والفعلي والحاصل.

- **إجرائياً:** يمكن تعريف واقع الدلالات العلمية بأنه: الحجم الحقيقي لاستخدام كل نوع منها في حساب وتفسير نتائج الأبحاث التجريبية، وأكثر الأخطاء الشائعة المرتبطة بهذا الاستخدام.
- **الدلالات العلمية:**

على الرغم من أنه لم يتم التوصل إلى تعريف علمي متّفق عليه لمصطلح الدلالات العلمية، فقد أجمعت المراجع على أن هناك ثلاثة أنواع من الدلالات، تُمثل مستويات متعددة من الدقة الإحصائية لتفسير النتائج في الأنواع المختلفة من الأبحاث العلمية، وأن العلاقة بين هذه الأنواع علاقة تراتبية هرمية.

- **الدلالة الإحصائية:**

مؤشرٌ إحصائي يساعد على تحديد ما إذا كانت النتيجة ترجع إلى الصدفة أو لبعض المتغيرات المحددة موضع البحث، كما يُبين للباحث إلى أي حد يستطيع أن يتأكد من دقة وموثوقية نتائجه، وإذا كان يستطيع تعميم النتائج التي حصل على المجتمع الأصلي لعينة بحثه (Gallo, 2016, p.1).

- **الدلالة العملية:**

مؤشرٌ إحصائي، يدل على مقدار التأثير المدروس بين المتغيرات موقع البحث، ويقيس مقدار التباين في المتغير التابع الذي يمكن تفسيره من خلال المتغير المستقل الذي يرتبط به أو يؤثر فيه، ويحدد إذا ما كانت الفروق الإحصائية أو العلاقات بين المتغيرات كبيرة إلى درجة تبرر عملية الأخذ بنتائجها (Akbaryan, 2013, p.5).

- **الدلالة العيادية:**

مؤشرٌ إحصائي يحدد القيمة العملية لتأثير التدخل أو المعالجة التجريبية لدى كل فرد من أفراد المجموعة، كما يحدد قدرة التدخل على إحداث تغيُّر دال على المستوى الفردي من خلال حساب مؤشر التغيُّر الموثوق (RCI) لكل فرد في المجموعة (Kazdin,1999, P.1).

- **مؤشر التغير الموثوق:**

مؤشرٌ إحصائي يدل على مقدار التغيُّر الحقيقي الناتج عن التدخل، والذي لا يعود لأخطاء القياس(Peterson, 2008, P. 5).

- **الأبحاث التجريبية:**

هي الأبحاث التي تعتمد منهج البحث التجريبي، والتي تقوم على إجراء تغيير متعمد ومضبوط للشروط المحددة للواقع أو الظاهرة، التي تكون موضوعاً للدراسة، وملاحظة ما ينتج عن هذا التغيُّر من آثار في هذا الواقع أو الظاهرة (المحمودي، 2019، ص.65).

- **إجرائياً:** يُقصد بها جميع الأبحاث التي يجرى فيها تدريب مجموعة من المشاركين لتلبية حاجات أو تحقيق أهداف محددة، من خلال برنامج تدريبي معد أو مبني لذلك.

- **البرنامج التدريبي:**

خطة متكاملة من الأنشطة والعمليات المنظمة، التي تستهدف تزويد المتدربين بمعارف ومهارات لتحقيق أغراض محددة، تُبنى في ضوء نظريات علمية وفق مراحل بناء متسلسلة، وتُنفّذ على شكل جلسات لها قواعد ومحتوى واضح من العناصر، تسير حسب خطة زمنية محددة (محمود، 2018، ص. 706).

- **إجرائياً:** أي برنامج تجريبي يرتبط بتحقيق أهداف محددة لدى المتدربين، سواءٌ أكان نوعه تعليمياً أم تأهيلياً أم إرشادياً أم علاجياً.

- **التدريب على الاختبار Training on test:**

إكساب المفحوصين خبرة التعرض للموقف الاختباري، من خلال إعطائهم تعليمات تتصل بالاختبار موضع التنفيذ، وتتضمن ظروف إجرائه، وشكله العام وتعليماته الأساسية العامة وكذلك الخاصة بكل جزء من أجزائه1 -13 Ferrier & et al, 2011, P.P).

- **الخصائص القياسية (السيكومترية):**

هي مؤشرات إحصائية، تعبّر عن مدى جودة وقوة الأدوات البحثية في تحقيق الأهداف المرجوة منها، ويُقصد بالخصائص القياسية للاختبار مدى توافر معاملات الصدق والثبات له في بيئةٍ بحثيةٍ محددة (بودلال ومور اس، 2021، ص.115).

- **إجرائياً:** يُقصد بها قيم معاملات الصدق المحكّي والتمييزي، وكذلك قيم معاملات ثبات ألفا كرونباخ والتجزئة النصفية لاختبار الذكاء في كل من التطبيق القبلي والبعدي المباشر.

- **منهج التحليل البعدي Meta - Analysis :**

يشير التحليل البعدي أو التِلوي إلى مجموعةٍ من التقنيات التي يستخدم فيها الباحث النتائج من مجموعة من الدراسات ذات الصلة لاستخلاص استنتاج عام (توليف)، بناءً على الأبحاث الموجودة حول موضوعٍ ما (عز وكشيك، 2022، ص.14).

- **الكفاية:**

عرفها معجم المعاني الجامع بأنها: مصدر كفى، وتعني المقدرة والكفاءة، وما يلزم بالضبط على قدر الحاجة، وأنها الحد الذي يفي بالغرض ويغني عن غيره.

- **إجرائياً: يُقصد بكفاية الدلالة الإحصائية في البحث الحالي:** مدى قدرتها على الكشف بدقة عن حجم تأثير البرامج التدريبية من ناحية، ومدى قدرتها على الكشف عن التغيُّرات الفعلية على مستوى كل فرد في العينة التجريبية نتيجة البرنامج التدريبي من ناحيةٍ أخرى.
- **إجرائياً: يُقصد بكفاية الدلالة العملية في البحث الحالي:** مدى قدرتها على الكشف عن التغيّرات الفعلية على مستوى كل فرد في العينة التجريبية نتيجة البرنامج التدريبي.

خلاصة:

عرض هذا الفصل مقدمة عامة عن البحث الحالي، متضمناً الإطار المفاهيمي العام لموضوع الدلالات العلمية وأنواعها، وأثر متغير التدريب فيها وفي الخصائص القياسية للاختبار، ووضح أهمية البحث وأهدافه والأسئلة المنبثقة منها، وقدم تعريفاً بأدوات البحث الأساسية المستخدمة فيه، والمنهجية البحثية المعتمدة، مع حديثٍ عن مجتمعه وعينته، إضافةً إلى تعريف بالمصطلحات الأساسية المتصلة بكل جانب من جوانبه أكانت تعريفات وصفية أم إجرائية.

الفصل الثاني

أبحاثٌ ودراساتٌ سابقة

الفصل الثاني

أبحاثٌ ودراساتٌ[5] سابقة

تمهيد:

يتضمن الفصل الحالي مجموعة أبحاث ودراسات سابقة استطاعت الباحثة الوصول إليها، ذات صلة بموضوع البحث، سيتم عرضها في محاورٍ ثلاث وفق متغيرات عنوان البحث، وفي تسلسلٍ زمني يبدأ من الأحدث وينتهي بالأقدم، ومن دون الفصل بين ما هو محلي أو عربي أو أجنبي. ابتداءً بأبحاث ودراسات الدلالات العلمية التي تعرضت لأهميتها وللتمييز بين أنواعها المختلفة، وتحليل واقع استخدامها، ثم بأبحاث الدلالة العيادية وطرائق حساب مؤشر التغير الموثوق، والمفاضلة بين هذه الطرائق الإحصائية ودقة نتائجها في ضوء طبيعة البحث ومتغيراته وخصائص عيناته، وانتهاءً بأبحاثٍ تناولت متغير التدريب وأثره في الخصائص القياسية للاختبار.

كما ضمَ الفصل تعقيباً ومناقشةً لمعلومات الأبحاث السابقة، أُتبع بجدولٍ يلخص أهم ما ورد في هذه الأبحاث، ثمّ بيان موقع البحث الحالي في خارطة الأبحاث السابقة.

1- أبحاث الدلالات العلمية:

- بحث بالكين ولينز (Balkin & Lenz, 2021) في الولايات المتحدة الأمريكية، بعنوان:

" قضايا معاصرة في الإبلاغ عن الدلالة الإحصائية، والعملية، والعيادية في الأبحاث الإرشادية "

Contemporary Issues in Reporting Statistical, Practical, and Clinical Significance in counseling Research.

هدف البحث إلى تسليط الضوء على المخاوف المتصلة بصحة نتائج الفرضية الصفرية، وتقديم مقترحات لكيفية كتابة تقارير نتائج الأبحاث الإرشادية المعاصرة، بحيث تتضمن تطبيق وتفسير قوانين حجم الأثر، وحساب فترات الثقة، ومؤشرات التغيُّر الموثوق (RCI) التي تعبر عن قيم الدلالة العيادية، والرسومات التوضيحية المرئية للنتائج إلى جانب اختبارات الدلالة الإحصائية.

وقد كانت العينة عبارة عن مثالين لتقارير إرشادية، هذين المثالين من بيانات أبحاث سابقة، تمثلت منهجية البحث بإعادة تحليل نتائجهما، بحيث يتضمن التقرير البحثي لكلٍ منهما مؤشرات الدلالات الثلاث، ليستعين بهما الباحثون عند الإبلاغ عن نتائج أبحاثهم. وقد انتهى البحث إلى وجود قصور في نتائج الأبحاث المستمدة من الدلالة الإحصائية، لذا فإن التوسع في حساب حجوم الأثر وفترات الثقة

[5] تأخذ الباحثة بالفرق بين مفهومي البحث والدراسة في عرض هذا الفصل من أن البحث عمل ميداني يبحث في الظواهر من خلال جمع بيانات عن مشاركين باستخدام أدوات استقصاء ثم تعالج إحصائياً، بينما الدراسة عمل علمي نظري مكتبي يجمع معلوماته البحثية من الكتب والمراجع والدوريات والمؤلفات المختلفة. لذا عمدت لعرض الدراسات السابقة باستخدام كلمة بحث لارتباطها بالعمل الميداني وليس بالعمل النظري المكتبي.

ومؤشرات الدلالة العيادية لنتائج الأبحاث الإرشادية أمرٌ ضروري لتعزيز فهم نتائج هذه الأبحاث، وزيادة المنفعة العلمية لأصحاب القرار.

- **بحث بحاش (2020) في الجزائر، بعنوان:**

"الدلالة الإحصائية والعملية لفرضيات الأبحاث النفسية والتربوية"

سعى البحث نحو تقديم تغذية راجعة عن استخدام اختبارات الدلالتين الإحصائية والعملية في نتائج الأبحاث النفسية والتربوية لدرجة الماجستير والدكتوراه، المنجزة في مجال علم النفس والتربية في بعض الجامعات الجزائرية.

تكونت عينة البحث من (347) بحثاً، تضمنت (5123) فرضية صفرية، وتمثلت أداته بقائمة لجمع البيانات قام الباحث بتصميمها استناداً إلى الأدبيات والأبحاث المتصلة ببحثه. وقد توصل من خلالها إلى أن أكثر اختبارات الدلالة الإحصائية استخداماً هو اختبار (T-test)، وأن نسبة استخدام حجم الأثر في فرضيات هذه الأبحاث ضعيفة جداً، إذ بلغت (2.71%) من إجمالي عدد الفرضيات المحسوب، كما أن تصنيفات قيم حجوم الأثر المستخرجة من هذه الاختبارات كانت منخفضة بالمجمل.

- **بحث وهبي وآخرين (2020) في سلطنة عمان، بعنوان:**

"درجة التوافق بين الدلالة الإحصائية ومؤشرات الدلالة العملية في الاختبارات الإحصائية برسائل الماجستير بالجامعات العمانية"

سعى الباحثون إلى تقييم درجة التوافق بين الدلالتين الإحصائية والعملية في (628) رسالةً من رسائل الماجستير التربوية، منذ بداية تطبيق برامج الماجستير فيها وحتى نهاية عام 2018 ميلادية.

قام الباحثون بإعداد استمارة تحليل محتوى، واستخدامها أداةً لهذا البحث، تألفت من ثلاثة أقسام يختص الأول باختبارات (ت)، والثاني بتحليل التباين، في حين يختص القسم الثالث بمعامل الارتباط، وقد بيّنت نتائجهم وجود علاقة إيجابية بين الدلالة الإحصائية وكل مستوى من مستويات الدلالة العملية، وأن أغلب الرسائل التي بحثت العلاقة بين المتغيرات في الجامعات العمانية وقعت في الخطأ من النوع الأول، كما أظهرت النتائج وجود ضعف في التصميمات الإحصائية المستخدمة لحساب معاملات الانحدار في رسائل بعض الجامعات العمانية.

- **بحث هانسن وآخرين (Hansen & et al., 2018) في الولايات المتحدة الامريكية، بعنوان:**

"تعزيز فهم النتائج ذات الأهمية العيادية: منظور البحث العيادي"

Enhancing the understanding of clinically meaningful results: A clinical research perspective.

هدف البحث إلى التحقق من فوائد حساب مقاييس حجم الأثر لاسيما مقياسي (d) لكوهين ومعامل ارتباط بيرسون، وإبراز أهمية حساب فترات الثقة إلى جانب اختبارات الدلالة الإحصائية، كما سلط الضوء على مشكلة استخدام الإحصائيات الاستدلالية الخاطئة في الأبحاث.

تمثلت منهجية هذا البحث بإجراء مقارنة بين قيم معاملات حجم الأثر والدلالة الإحصائية لأبحاثٍ سابقة. وبينت نتائجه أن القيمة الاحتمالية لا توفر مقياساً جيداً ودليلاً كافياً على صحة الفرضية أو عدم صحتها، مع أنها القيمة الأكثر شيوعاً واستخداماً من قبل الباحثين، وأكدت على ضرورة استكمالها بطرائق إحصائية أخرى مثل حجوم الأثر، وأشارت إلى ازدياد التوجه نحو حساب إحصائيات بايز في الأبحاث.

- بحث سكوبر وآخرين (Scober & et al., 2018) في هولندا، بعنوان:

"الدلالة الإحصائية مقابل الأهمية العيادية لحجوم الأثر الملحوظة: ما الذي تمثله حقاً القيم الاحتمالية *P* وفترات الثقة؟"

Statistical Significance Versus Clinical Importance of Observed Effect Sizes: What Do P- Values and Confidence Intervals Really Represent?

لقد هدف سكوبر وزملاؤه من بحثهم إلى شرح مفاهيم حجم الأثر، وفترات الثقة، إلى جانب مفهوم القيمة الاحتمالية ودراسة العلاقات بينها، وكيف يمكن استخدام فترات الثقة وحجم الأثر لتوضيح وزيادة دقة نتائج الدلالة الإحصائية المحسوبة للمعالجات العيادية، من خلال محاكاة بيانات عينة مؤلفة من 30 مريضاً، وتكرار هذه المحاكاة 25 مرة.

لقد بينت النتائج أن 23 من أصل 25 فترة ثقة محسوبة، أي ما يعادل 92% من فترات الثقة، تغطي متوسط المجتمع الأصلي الحقيقي البالغ 30، وأن هناك علاقة وثيقة بين فترات الثقة وقيم حجم الأثر، حيث توفر فترات الثقة نطاقاً من القيم لتقدير حجم الأثر، ويمكن عدّها مقياساً لقوة الحكم على صحة القرار الإحصائي.

- بحث عصفور (2018) في الجمهورية العربية السورية، بعنوان:

"الدلالة الإحصائية والعملية وقوة الاختبار الإحصائي في البحوث التربوية المنشورة في مجلة جامعة دمشق للعلوم التربوية والنفسية (2004 – 2014م)"

هدفت الباحثة إلى تقييم حجم الأثر وقوة الاختبار الإحصائي، وتحديد نسبة الفرضيات التي اتفقت فيها نتائج الدلالة الإحصائية مع الدلالة العملية وقوة الاختبار، وإلى تحديد نسبة استخدام مفهوم الدلالة العملية إلى جانب الدلالة الإحصائية، في الأبحاث التربوية المنشورة في مجلة جامعة دمشق للعلوم التربوية والنفسية (2004 – 2014 م).

لقد بينت نتائجها وجود ارتباط بين معاملات الدلالتين الإحصائية والعملية، إذ بلغت نسبة الاتفاق بينهما(64.71%)، ووضحت أن نسبة حساب الدلالة العملية في إجمالي الأبحاث–الوصفية التحليلية والتجريبية– بلغت (8.31%)، في حين أن نسبة حسابها في الأبحاث التجريبية كانت (94.28%)، وقد أرجعت الباحثة ذلك للاعتقاد الخاطئ لدى الباحثين أن حجم الأثر يُحسب فقط للأبحاث التجريبية، وأن الأبحاث الوصفية التحليلية لا تحتاج إلى حساب حجم الأثر.

– بحث نصار (2016) في الإمارات العربية المتحدة، بعنوان:

"الدلالة الإكلينيكية للبحوث النفسية والتربوية المستخدمة للتصاميم التجريبية"

هدف البحث إلى تمييز أنواع الدلالات الثلاث: الإحصائية والعملية والإكلينيكية لنتائج الأبحاث التجريبية، واستعراض طرائق حساب الدلالة العيادية مع التركيز على طريقة (JT) في حساب هذا النوع من الدلالة، لأنها برأي الباحث تعطي تقديراً متوسطاً لأثر المعالجة، فلا تقلل ولا تضخم من أثر المعالجة التجريبية.

وقدم الباحث ضمن نتائجه نموذجاً مقترحاً لفحص الدلالة العيادية لأثر المعالجة التجريبية على مستوى الفروق بين المجموعتين التجريبية والضابطة، من خلال الربط بين التصنيفات الناتجة عن طريقة (JT) ونوع المجموعات "تجريبية وضابطة" باستخدام الاختبار الإحصائي كاي مربع، مؤكداً قدرة هذا النموذج على فحص الدلالة العيادية، والتمييز بين تصنيفات المجموعتين الضابطة والتجريبية. وبين أيضاً أنه عندما تكون النتائج دالة عيادياً على مستوى المجموعة، فهي بالضرورة ستكون دالة من الناحيتين الإحصائية والعملية، ولكن العكس غير صحيح.

– بحث البارقي (2012) في المملكة العربية السعودية، بعنوان:

"واقع الدلالة الإحصائية والعملية للبحوث المنشورة بمجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية في المدة من 1425 – 1430 هـ"

سعى البحث إلى الكشف عن نسبة استخدام مفهوم الدلالة العملية إلى جانب الدلالة الإحصائية في الاختبارات المستخدمة في أبحاث مجلة جامعة أم القرى، والبحث عن الفروق في تصنيفات حجوم الأثر المحسوبة.

بلغ العدد الكلي لاختبارات الفروض المدروسة (2223) اختباراً، وبيّنت النتائج أن نسبة الاختبارات الإحصائية التي تمّ حساب حجم الأثر لها بلغت (0.76%) وهي نسبة ضعيفة جداً، وأن (52.27%) من حجوم الأثر المحسوبة كان تصنيفها منخفض، في حين أن (25.24%) منها كان تصنيفها مرتفع، و(22.49%) منها كان تصنيفها متوسط.

– بحث ماكميلان وفولي (McMillan & Foley, 2011) في الولايات المتحدة الأمريكية، بعنوان:

"الإبلاغ عن حجم الأثر ومناقشته: يبقى الطريق الأقل ارتياداً؟"

Reporting and Discussing Effect Size: Still the Road Less Traveled?

صُمم البحث لتحديد النسب المئوية لحساب قيم حجم الأثر، ومناقشتها وتفسيرها، في أربع مجلات علمية هي: مجلة علم النفس التربوي، علم النفس التربوي المعاصر، الأبحاث التربوية، ومجلة التعليم التجريبي، تكونت عينة البحث من (417) بحثاً جرى اختيارها بشكل عشوائي من هذه المجلات.

أظهرت النتائج أن النسب المئوية لحساب حجم الأثر في هذه الأبحاث انحصرت بين 64% و82%، وأن قيم حجوم الأثر المحسوبة ذات التصنيف المنخفض هي الأكثر انتشاراً، كما بينت أن ما يقارب نصف هذه الأبحاث لا يتضمن أية مناقشات أو تفسيرات عن نتائج حجوم الأثر المحسوبة، لذا من الصعب معرفة ما إذا كان الباحثون على دراية بمعناه وأهميته الفعلية، أم تم حسابه فقط كإجراء إحصائي روتيني.

- **بحث إيسين وآخرين(Eisen & et al,2007) في الولايات المتحدة الامريكية، بعنوان:**

"قياس التغيُّر المهم عيادياً في متابعة معالجة الصحة النفسية"

Measuring Clinically Meaningful Change Following Mental Health Treatment.

كان الهدف الرئيسي من البحث هو تقييم الاستجابة على مقياس تحديد السلوك والأعراض BASIS-24)، باستخدام ثلاثة أساليب لتقييم التغير الدال عيادياً هي: مؤشر التغير الموثوق (RCI)، وحجم الأثر(ES)، والخطأ المعياري للقياس (SEM).

تكونت العينة من بيانات (2248) مريضاً نفسياً، وبيّنت النتائج أن حساب التغيُّرات على مستوى المجموعة باستخدام قيم كل من قيم حجم الأثر والخطأ المعياري للقياس، قد أغفل حقيقة عدم ظهور أي تغيّرٍ حقيقي بالاتجاه المطلوب لدى نسبة كبيرة من الأفراد، إضافةً إلى أن بعضهم قد أظهر تراجعاً دالاً عيادياً بعد تطبيق البرنامج.

2- أبحاث الدلالة العيادية:

- **بحث رونك وآخرين (Ronk & et al., 2016) في أستراليا، بعنوان:**

"صدق تصنيفات التغيُّر الدال عيادياً الناتجة عن طريقتي (JT) و(HA)"

Validity of clinically significant change classifications yielded by Jacobson Truax and Hageman – Arrindell methods.

هدف البحث إلى مقارنة نسب التوافق بين التصنيفات الناتجة عن طريقتي (JT) و (HA) في حساب الدلالة العيادية، بتطبيق مقياس الاكتئاب والقلق والشدة (DASS-21)، ومقياس تقييم التعافي (RAS)، ومقياس الرضا عن نوعية الحياة – النسخة المختصرة (LES-Q) على (718) مريضاً نفسياً داخلياً.

بينت نتائج البحث عدم وجود فروق ذات دلالة إحصائية بين التصنيفات الناتجة عن كلتا الطريقتين، خاصةً تلك المتصلة بتصنيف التعافي الدال عيادياً.

ولأنه لم تكن هناك ميزة ملحوظة لاستخدام طريقة (HA) على طريقة (JT)، فقد أوصى الباحثون باستخدام طريقة (JT) لأنها الأبسط والأسرع حساباً.

- **بحث رونك وآخرين (Ronk & et al., 2014) في أستراليا، بعنوان:**

"هل يمكننا الوثوق بالدلالة العيادية عند تقييم النتائج السريرية؟ تحقيق في ثبات منهجية الدلالة العيادية وصدقها ومرونتها واستقرارها"

Can we trust clinical significance when evaluating clinical outcomes? An investigation of the reliability, validity, flexibility, and stability of clinical significance methodology.

من خلال هذا البحث تمّ تقييم مدى اتساق تصنيفات الدلالة العيادية الناتجة عن حساب الطرائق الآتية: (JT, GLN, NK, EN, HA) لدرجات (2676) مريضاً نفسياً، عولجوا في مصحة نفسية على مدى أربع سنوات (2004–2008)، تمثلت أدوات البحث باستخدام ثلاثة مقاييس للتقرير الذاتي هي: مقياس الاكتئاب والقلق والإجهاد (DASS-21)، مقياس نوعية الحياة والشعور بالرضا (Q-LES-Q)، والمقياس الفرعي من مقياس الصحة العقلية العامة –النسخة القصيرة– (SF-14).

وبينت النتائج أن التصنيفات العيادية للطرائق الأربعة (JT, GLN, NK, EN) كانت متسقة، إذ بلغت نسب الاتفاق بينها (93.7%)، بينما كانت طريقة (HA) أقل توافقًا مع الطرائق الأربع الأخرى، حيث بلغت نسبة الاتفاق (81.3٪)، وصنفت عدداً أقل من تصنيفات التعافي مقارنةً بالطرائق الأربع الأخرى، لذا ينصح باستخدام الطريقة الأبسط لحساب الدلالة العيادية وهي طريقة (JT).

وقد ناقش الباحثون الافتراض القائل "أن منهجية الدلالة العيادية تعدّ التصنيفات نتائج وليست إجراءات" مبيّنين أن تصنيفات الدلالة العيادية التي يتم الحصول عليها ليست بالضرورة مستقرة بمرور الوقت، ويجب عدّ مفاهيم التعافي والتحسن إجراءات وليس نتائج نهائية للمعالجات التجريبية.

- **بحث شدا (Shada, 2013) في الولايات المتحدة الامريكية، بعنوان:**

"تكرار مقارنة التقنيات الإحصائية المستخدمة في طرائق حساب الدلالة العيادية"

Replication of a comparison of statistical techniques using clinical significance methods.

كان الهدف من هذا البحث مقارنة نتائج طرائق حساب الدلالة العيادية الأربع (JT، GLN، EN ، HA)، لدرجات عينة مؤلفة من (395) مريضاً نفسياً على مقياس الخبرة الشخصية (OQ-45). وقد بينت النتائج وجود اختلافات في تصنيفات الدلالة العيادية الناتجة عن الطرائق الأربع باستثناء طريقتي (JT) و(GLN)، وكانت نسبة الاتفاق بين هاتين الطريقتين (95%).

ومن الجدير بالذكر أن نتائج هذا البحث بينت أن طريقة (HA) في حساب الدلالة العيادية قد أسفرت عن معدلات عالية جداً من تصنيفات التحسن الدال عيادياً، خلافاً لنتائج الأبحاث السابقة التي أظهرت أن هذه الطريقة هي الأكثر تحفظاً في إظهار معدلات التحسن الدال عيادياً، وقد ردّ الباحث هذا

التناقض إلى التعقيدات الإحصائية في مقام المعادلة، وبيّن أنها قد تكون هي سبب الاختلاف، كما بيّن بالإجماع مع الأبحاث السابقة أن هذه الطريقة هي الأكثر تعقيداً في حساب الدلالة العيادية.

وأظهرت النتائج أيضاً أن طريقة (JT) قدمت تقديراً معتدلاً للتغيير الموثوق وتصنيفات الدلالة العيادية المتصلة به، لذا أوصت باستخدامها، كما أوصت بضرورة عمل المزيد من المقارنات بين هذه الطرائق في أبحاث مستقبلية، بغية التوصل إلى اتفاق على مقياس موحد لحساب هذا النوع من الدلالة لتوحيد نتائج عمليات التدخل والمعالجة.

- **بحث أتكينز (Atkins & et al, 2005) في الولايات المتحدة الأمريكية، بعنوان:**

"تقييم الدلالة العيادية: هل يهم أية طريقة نستخدم؟"

Assessing Clinical Significance: Does It Matter Which Method We Use?

هدف هذا البحث إلى مقارنة نتائج استخدام طرائق (JT)، (GLN)، (EN)،(HA).لحساب الدلالة العيادية، وبيان إلى أي مدى تختلف في تصنيفاتها العيادية لعينات سريرية تم استخدام درجاتها قبل المعالجة وبعدها، وإجراء طريقة المحاكاة وتوليد البيانات لهذه الدرجات واستخراج تصنيفاتها العيادية وفق الطرائق السابقة، ثم حساب نسب الاتفاق بين التصنيفات الناتجة عنها باستخدام الاختبار Kappa.

أظهرت النتائج وجود اتفاق دال إحصائياً بين التصنيفات الناتجة عن الطرائق الأربع، وأن نسب الاتفاق تزداد في الحالات التي تكون فيها معاملات الثبات مرتفعة، وأن طريقة (EN) هي الأكثر قدرة على الكشف عن حالات التحسن الدالة، على عكس طريقة (HA) الأكثر تحفظاً في تصنيف الحالات المرضية التي تعافت. وأفاد الباحثون أن طريقتي (JT) و (GLN) أدتا إلى تصنيفات عيادية متطابقة تقريباً، على الرغم من أن طريقة (GLN) أُسست بالأصل بناءً على الانتقادات الموجهة لطريقة (JT).

- **بحث باور وآخرين (Bauer & et al, 2004) في ألمانيا، بعنوان:**

"طرائق حساب الدلالة العيادية: مقارنة بين التقنيات الإحصائية"

Clinical Significance Methods: A Comparison of Statistical Techniques.

سعى هذا البحث نحو فحص الاختلاف في التصنيفات الناتجة عن حساب مؤشرات التغيُّر الموثوق بتطبيق طرائق (JT, GLN, EN, HA, HLM) في حساب الدلالة العيادية على بيانات (386) مريضاً خارجياً.

وباعتماد منهج التحليل البعدي أظهرت النتائج وجود فروق ذات دلالة إحصائية في التصنيفات العيادية الناتجة عن الطرائق السابقة جميعها، باستثناء طريقتي (JT) و (GLN)؛ إذ أنتجت هاتان الطريقتان أعلى نسبة اتفاق بلغت (0.93)، في حين كانت أدنى نسب الاتفاق بين طريقتي (EN) و(HA)؛ إذ بلغت (0.41)، كما بينت النتائج أن طريقة (JT) قدمت تقديرات معتدلة لتأثير العلاج، لذا أوصت باستخدامها في حساب نتائج الدراسات والأبحاث.

3- أبحاث التدريب والخصائص القياسية:

- **بحث رشوان (2017) في السعودية، بعنوان:**

"إستراتيجيات أداء الاختبار وتأثيرها في مستوى الأداء والخصائص القياسية للاختبار وبنوده"

هدف الباحث من هذا البحث إلى تعرُّف إستراتيجيات أداء الاختبار الأكثر استخداماً لدى عينة مؤلفة من (181) طالباً وطالبةً من طلبة الدبلوم في جامعة القصيم، ومدى اختلاف درجاتهم في الاختبار التحصيلي، تبعاً لاختلاف درجة استخدامهم لإستراتيجيات أداء الاختبار. كما هدف إلى بيان مدى تأثُّر الخصائص القياسية للاختبار التحصيلي باختلاف استخدام إستراتيجيات أداء الاختبار لدى أفراد عينة البحث.

تمثلت أداة البحث بمقياس من إعداد الباحث مؤلف من 49 بنداً حول استراتيجيات أداء الاختبار، بينت نتائجه أن معدلات درجات الاختبار التحصيلي للطلبة تزداد بازدياد درجة استخدامهم لإستراتيجيات أداء الاختبار، وأنه يمكن التنبؤ بدرجات الطلبة في الاختبار التحصيلي من خلال درجاتهم على مقياس إستراتيجيات أداء الاختبار، كما أظهرت تحسُّن الصدق التنبؤي والتلازمي للاختبار التحصيلي بازدياد استخدام هذه الإستراتيجيات، وكذلك تحسُّن الثبات ولكن بدرجة أقل من تحسُّن الصدق.

- **بحث بوهل وآخرين (Pohl & et al., 2016) في ألمانيا، بعنوان:**

"اختبار الطلبة ذوي الاحتياجات التعليمية الخاصة في – التقييمات واسعة النطاق – الخصائص القياسية لدرجات الاختبارات وعلاقتها بسلوك إجراء الاختبار"

Testing Students with Special Educational Needs in –Large Scale Assessment – Psychometric Properties of Test Scores and Associations with Test Taking Behavior .

سعى هذا البحث إلى فحص الكفاءات القرائية عند الطلبة الذين يعانون من صعوبات التعلم في التقييمات واسعة النطاق؛ والتي يقصد بها التقييمات الوطنية العامة، ودراسة أثر الاختلافات في سلوك أداء الاختبار، وبالتحديد، سلوكي التخمين وعدم الإجابة، في تفسير انخفاض معاملات الخصائص القياسية لدرجات هؤلاء الطلبة.

تكونت عينة البحث من مجموعتين من طلبة الصف التاسع، سُحبت الأولى من مدارس التعليم العام، وبلغ عدد أفرادها (13897) طالباً، وسُحبت الثانية من المدارس الخاصة بالطلبة ذوي صعوبات التعلم، وبلغ عدد أفرادها (401) طالباً.

بينت النتائج أن سلوك التخمين العشوائي يرتبط بانخفاض الخصائص القياسية لنتائج الاختبار في مجموعة طلبة صعوبات التعلم، كما أظهرت أيضاً أن سلوك الإجابات المفقودة؛ أي عدم الإجابة عن السؤال، تفسر الجودة القياسية المنخفضة في العينة بأكملها، وأن مؤشرات ملاءمة البند الناتجة عن

تطبيق نموذج Rasch في التحليل الإحصائي، ومعاملات ثبات القياس لمقاييس الكفاءة القرائية كانت منخفضة لدى المجموعة الفرعية من الطلبة الذين جرى تصنيفهم على أنهم يظهرون استجابات مفقودة عالية.

- **بحث ماسبونز ولابر (Maspons & Llabre, 1985) في الولايات المتحدة الأمريكية، بعنوان:**

"أثر تدريب ذوي الأصول الإسبانية على أداء الاختبار في الخصائص القياسية"

The Influence of Training Hispanics in Test Taking on Psychometric Properties of a test.

هدف البحث إلى معرفة أثر تدريب عينة من الطلبة الإسبانيين على مهارات أداء الاختبار، في خصائصه القياسية الآتية: الاتساق الداخلي لاختبار تحديد المستوى الحسابي، والصدق التنبؤي لهذا الاختبار بدرجات التحصيل النهائي.

تألفت العينة من (535) طالباً وطالبةً من مرحلة التعليم الجامعي، جرى تقسيمها بشكل عشوائي إلى مجموعتين: تجريبية؛ تم تدريب أفرادها على مهارات أداء الاختبار، وضابطة؛ تلقى أفرادها معلومات عن البرامج التعليمية للكلية.

بينت النتائج انخفاضاً طفيفاً في معامل الاتساق الداخلي المحسوب لدرجات المجموعة التجريبية في الاختبار البعدي حيث بلغ (0.855) مقارنةً بهذا المعامل لدرجات المجموعة الضابطة، والبالغ (0.889). في حين أظهرت ارتفاع معامل الصدق التنبؤي لاختبار تحديد المستوى الحسابي بنتائج التحصيل الدراسي، لأفراد المجموعة التجريبية مقارنةً بالمجموعة الضابطة.

4- تعقيبٌ ومناقشةٌ:

- تنوعت الأبحاث السابقة في عرضها موضوع الدلالات العلمية، فمنها ما حاول الجمع بين الأنواع الثلاث من الدلالات من حيث التمييز بينها، وتوضيح أهميتها، ونقاط القوة والضعف في كل نوع منها في جانبها النظري، ولكن دون تحليل واقع استخدامها جميعاً، أو دراسة العلاقات بين مؤشراتها ومدى كفاية كل منها في الإبلاغ عن النتائج في جانبها العملي، مثل بحث (Balkin & Lenz, 2021) الذي جمع بين مؤشرات الدلالات الثلاث من خلال تقديم مثالين عن كيفية الإبلاغ عن التقارير البحثية، دون دراسة العلاقات بين هذه المؤشرات أو توضيح مدى كفاية كل منها في تقديم رؤيا أوضح للنتائج، الأمر الذي جعل أهميته تقتصر على تقديم هذين المثالين، وبقيت الحاجة موجودة لأبحاثٍ علمية توضح العلاقات بين مؤشرات الدلالات، كذلك بحث (نصار، 2016) الذي قدم شرحاً مستفيضاً عن أنواع الدلالات في جانبه النظري، لكنه اقتصر في منهجيته وإجراءاته العملية على عرض مثالين الأول نفسي والثاني تربوي مُستخدماً فيهما بيانات افتراضية بهدف توضيح كيفية حساب مؤشر التغير الموثوق باستخدام طريقة (JT) من ناحية، وتقديم نموذج مقترح لفحص الدلالة

العيادية لأثر المعالجة التجريبية من ناحيةٍ أخرى، دون أن يدرس علاقتها ببقية الدلالات، وبحث إيسين وآخرين(Eisen& et al,2007) الذي بالرغم من اعتماده مؤشرات الدلالات العلمية الثلاث في تقييم استجابة المرضى النفسيين، غير أنه لم يدرس العلاقات بين هذه المؤشرات أو مدى كفاية كل منها في حساب نتائج الأبحاث العيادية.

ومن الأبحاث السابقة ما استهدف تحليل واقع استخدام الدلالتين الإحصائية والعملية فقط دوناً عن الدلالة العيادية مثل (بحاش، 2020) الذي اعتمد في عينته على أبحاث أكاديمية، وبينت نتائجه تدني نسبة حساب حجم الأثر في فرضيات هذه الأبحاث، دون أن يحدد مستويات حجم الأثر المحسوبة أوعلاقتها بالقرار الإحصائي، وبحثي (البارقي، 2012) و(McMillan & Foley, 2011) اللذين هدفا إلى تحليل واقع استخدام هاتين الدلالتين في الأبحاث العلمية، وتحديد نسب مستويات حجم الأثر المحسوبة، دون دراسة علاقتها مع مؤشرات الدلالة الإحصائية. على خلاف أبحاث (وهبي وآخرين، 2020) و(عصفور، 2018) و(Scober & et al., 2018) إذ لم يكتفِ الباحثون فيها بتحليل واقع استخدام هاتين الدلالتين بل درسوا أيضاً العلاقات بين مؤشراتهما، وقد أسفرت نتائجهم عن وجود علاقة إيجابية بين معاملات الدلالة الإحصائية وكل مستوى من مستويات الدلالة العملية، وبينت نتائج (عصفور، 2018) أن نسبة الاتفاق بين معاملات هاتين الدلالتين بلغت (64.71%). كما أظهرت نتائج (Scober & et al., 2018) وجود علاقة وثيقة بين فترات الثقة وحجم الأثر، وأكدت إمكانية عدّها مقياساً إضافياً مهماً لقوة الحكم على صحة القرار الإحصائي.

- في حين اقتصر بعض الأبحاث على تسليط الضوء على الدلالة العيادية وحدها باعتبارها النوع الأحدث من الدلالات، والذي يُمثل أعلى مستويات الدقة الإحصائية، دون أن يسعى إلى دراسة العلاقة بين تصنيفاتها العيادية ومؤشرات الدلالتين الإحصائية والعملية، مثل أبحاث: (Ronk & et al, 2016)، و(Ronk & et al, 2014)، و(Shada, 2013)، و(Atkins & et al, 2005)، و(Bauer & et al, 2004). والتي تنوعت هي أيضاً في طرائق حسابها مؤشرات الدلالة العيادية، فمنها ما اقتصر على طريقة واحدة وهي طريقة (JT)، باعتبارها الأسهل والأسرع في حساب هذا النوع من الدلالة مثل (نصار، 2016)، ومنها ما استهدف مقارنة طريقتين فقط من طرائق حسابها مثل (Ronk & et al, 2016) الذي سعى إلى مقارنة تصنيفات طريقتي (JT) و (HA) على وجه التحديد باعتبار الأولى هي الأسهل حساباً، والثانية هي الأكثر تعقيداً، وأظهرت نتائجه عدم وجود فرق دال بين تصنيفات الطريقتين، لذا أوصى باستخدام طريقة (JT) باعتبارها الأسهل. ومنها ما صُمّم لمقارنة نتائج أربع طرائق في حساب الدلالة العيادية مثل بحثي (Shada, 2013)، و(Atkins & et al, 2005) اللذين أسفرا عن نتائج متعارضة؛ ففي حين بينت نتائج الأخير وجود اتفاق دال بين التصنيفات الناتجة عن الطرائق الأربعة، وضحت نتائج (Shada, 2013) وجود اختلاف في التصنيفات الناتجة عن تلك الطرائق باستثناء طريقتي (JT) و(GLN). ومنها ما هدف

إلى مقارنة طرائق حساب الدلالة العيادية الخمس جميعها مثل بحثَي (Ronk & et al, 2014) و(Bauer & et al, 2004) اللذين أسفرا أيضاً عن نتائج متعارضة؛ ففي حين بينت نتائج الأخير وجود فروق دالة إحصائياً في تصنيفات الطرائق الخمس، باستثناء طريقتي (JT) و (GLN)؛ أظهرت نتائج (Ronk & et al, 2014) أن أربعاً من الطرائق الخمس كانت تصنيفاتها العيادية متسقة، إذ بلغ متوسط نسب الاتفاق بينها (93.7%)، باستثناء طريقة (HA) التي كانت أقل توافقًا مع الطرائق الأخرى، وفي النهاية أوصى كلا البحثان باستخدام طريقة (JT) في حساب هذا النوع من الدلالة.

- وتنوعت الأبحاث التي تناولت أثر متغير التدريب في الخصائص القياسية من حيث منهجيتها ففي الوقت الذي اعتمد فيه (رشوان، 2017) و(Pohl & et al., 2016) على رصد استراتيجيات أداء الاختبار المُستخدمة من قبل أفراد عينتيهما من خلال أدوات، مثل مقياس مُعد لهذا الغرض، وتحليل الاستجابات على الاختبار لمعرفة الاستراتيجيات المستخدمة من قبل أفراد العينة، اعتمد بحث (Maspons and Llabre, 1985) على برنامج تدريبي على هذه الاستراتيجيات متبعاً خطوات منهج البحث التجريبي للوصول إلى نتائج بحثه، ولعل ما تجدر الإشارة إليه اعتماد هذه الأبحاث جميعها في تحليل نتائجها على مؤشرات الدلالة الإحصائية وحدها دون الأخذ بالحسبان مؤشرات الدلالتين العملية والعيادية، رغم أهميتهما في دقة القرار الإحصائي المتعلق بالخصائص القياسية، والذي لابد أن يستند إلى مستويات عدة من الدلالة العلمية ليكون على قدرٍ كافٍ من الدقة والموثوقية.

كما تباينت هذه الأبحاث أيضاً من حيث استراتيجيات أداء الاختبار المستهدفة فيها، إذ تناول رشوان خمس استراتيجيات أساسية هي: تخطيط وتنظيم الوقت، تخطيط وتنظيم الإجابة، تحليل الأسئلة، الضبط والسيطرة، وإستراتيجية ما بعد الاختبار، واستهدف (Maspons and Llabre, 1985) أربع استراتيجيات هي: إدارة الوقت، تجنب الخطأ، التخمين، والاستدلال الاستنباطي، في حين اقتصر (Pohl & et al., 2016) على إستراتيجيتي التخمين وعدم الإجابة، ودراسة أثرهما في الخصائص القياسية للاختبار.

- وقد اختلفت نتائج هذه الأبحاث حول أثر التدريب في الخصائص القياسية، فقد أظهرت نتائج (رشوان، 2017) تحسُّن الصدق التنبؤي والتلازمي للاختبار بازدياد استخدام هذه الإستراتيجيات، وكذلك تحسُّن الثبات ولكن بدرجة أقل من تحسُّن الصدق، وبينت نتائج (Pohl & et al., 2016) أن سلوكي التخمين العشوائي وترك الإجابات يفسران الجودة المنخفضة للخصائص القياسية، في حين بينت نتائج (Maspons and Llabre, 1985) انخفاضاً طفيفاً في معامل الاتساق الداخلي المحسوب، وارتفاعاً في معامل الصدق التنبؤي للاختبار بتأثير متغير التدريب.

- وتعددت أنواع العينات المُستخدمة في الأبحاث السابقة، فقد اعتمد بعضها على عينات بحثية تكونت إما من رسائل أكاديمية مثل (بحاش، 2020) (وهبي وآخرين، 2020) أو من أبحاث مجلات علمية مثل (عصفور، 2018)، و(البارقي، 2012)، و(McMillan & Foley, 2011) بهدف تحليل واقع استخدام الدلالتين الإحصائية والعملية فيها.

واستخدم بعضها بيانات مُجمَّعة لمرضى نفسيين مثل(Ronk & et al, 2016)، و(Ronk & et al, 2014) و(Shada, 2013) و(Bauer & et al, 2004)، في حين اعتمد بعضها على أسلوب المحاكاة وتوليد البيانات مثل بحثي(Scober & et al., 2018) و(Atkins & et al., 2005)، واقتصر (نصار، 2016) في بحثه على أمثلة تضمنت بيانات افتراضية، وأخيراً تناولت بعض الأبحاث عينات تجريبية مكونة من متعلمين، سواءٌ جامعيين مثل بحث(رشوان، 2017)، و (Maspons & Llabre, 1985) أم تلاميذ مثل بحث (Pohl & et al., 2016).

- كما تنوعت أنواع البرامج التجريبية المستهدفة في الأبحاث السابقة، فقد استهدف بحث بالكين ولينز (Balkin & Lenz, 2021) البرامج التجريبية في الأبحاث الإرشادية، وركز بحث (نصار، 2016) على البرامج التدريبية في الأبحاث النفسية والتربوية، بينما اعتمد كل من(Ronk & et al., 2016)، و(Ronk & et al, 2014) و(Shada, 2013) و (Eisen& et al.,2007) و(Bauer & et al, 2004) على برامج علاجية لعينات سريرية، في حين استهدف (Maspons & Llabre, 1985) تدريب أفراد العينة التجريبية على برنامج تدريبي حول استراتيجيات أداء الاختبار.

والجدول الآتي يبين تلخيصاً لأهم ما ورد في الأبحاث السابقة من معلوماتٍ ونتائج.

جدول (1)

ملخص الأبحاث السابقة

البحث	الهدف	أهم النتائج	نقاط القوة	نقاط الضعف
Balkin & Lenz, (2021	الوصول إلى مقترحات لكتابة تقارير الأبحاث الإرشادية	– وجود قصور في نتائج الدلالة الإحصائية. – ضرورة حساب حجوم الأثر وفترات الثقة والدلالة العيادية	– الاعتماد على مؤشرات الدلالات العلمية الثلاث.	– الاقتصار على طرح مثالين فقط عن كيفية الإبلاغ عن التقارير
(بحاش، 2020)	تحليل واقع استخدام الدلالتين الإحصائية والعملية في نتائج الرسائل الأكاديمية	– ضعف حساب حجم الأثر في فرضيات هذه الأبحاث	– الاعتماد على تحليل عينات من رسائل أكاديمية	– الاقتصار على تحليل واقع استخدام الدلالتين الإحصائية والعملية فقط
(وهبي وآخرين، 2020)	تقييم درجة التوافق بين الدلالتين الإحصائية والعملية في رسائل الماجستير	– وجود علاقة إيجابية بين الدلالة الإحصائية وكل مستوى من مستويات الدلالة العملية	– حجم العينة كبير يغطي فترة زمنية طويلة، لحدود 20 عاماً	– الاقصار على رسائل الماجستير دوناً عن الدكتوراه. – الاقتصار على تحليل الدلالتين الإحصائية والعملية فقط

(Hansen & et al., 2018)	دراسة فوائد حساب مقاييس حجم الأثر وفترات الثقة، إلى جانب اختبارات الدلالة الإحصائية	– لا توفر القيمة الاحتمالية دليلاً كافياً على صحة القرار البحثي. – ضرورة حساب حجم الأثر، وفترات الثقة إلى جانب الدلالة الإحصائية	– دراسة العلاقة بين مؤشرات الدلالتين الإحصائية والعملية	– إهماله مفهوم الدلالة العيادية مع أنه ينطلق من منظور البحث العيادي.
(Schober & et al., 2018)	دراسة العلاقات بين حجم الأثر، وفترات الثقة والقيمة الاحتمالية	– هناك علاقة وثيقة بين فترات الثقة وحجم الأثر – يمكن عدّ فترات الثقة مقياساً لقوة الحكم على صحة القرار الإحصائي	– اعتماد أسلوب محاكاة وتوليد البيانات	– الخلط بين مفهومي الدلالة العيادية وحجم الأثر.
عصفور ، (2018	تقييم حجم الأثر وقوة الاختبار الإحصائي، وتحديد نسبة استخدام الدلالة العملية إلى جانب الإحصائية في الأبحاث	– هناك نسبة اتفاق جيدة بين معاملات الدلالتين الإحصائية والعملية – تدني نسبة حساب الدلالة العملية في الأبحاث الوصفية والتجريبية	– حجم العينة الكبير	– اعتماد أبحاث مجلات دوناً عن الرسائل الأكاديمية. – غياب الدلالة العيادية في عملية التحليل.
(نصار، 2016)	تمييز أنواع الدلالات العلمية لنتائج الأبحاث التجريبية، وتقديم أمثلة لحساب الدلالة العيادية.	– طريقة (JT) هي الأنسب لحساب الدلالة العيادية – ارتفاع مؤشرات الدلالة العيادية يرافقه ارتفاع مؤشرات الدلالتين الإحصائية والعملية ولكن العكس غير صحيح	– التمييز بين أنواع الدلالات الثلاث	– اعتماد بيانات افتراضية فقط دون الواقعية – الاقتصار على طريقة (JT) في حساب الدلالة العيادية.
(البارقي ، 2012	تحديد نسبة استخدام الدلالة الإحصائية والعملية في الأبحاث	– تدني نسبة حساب الدلالة العملية – ارتفاع نسبة حجم الأثر المنخفض في العينة المُحللة	– عرض معظم قوانين حجم الأثر في الجانب النظري	– غياب الدلالة العيادية في عملية التحليل – اعتماد أبحاث مجلات دوناً عن الرسائل الأكاديمية
(McMillan & Foley, (2011	تحديد نسب حساب حجم الأثر ومناقشتها في الأبحاث	– انحصرت نسب حساب حجم الأثر بين 64% و 82%. – حجوم الأثر المنخفضة هي الأكثر انتشاراً – نصف الأبحاث تقريباً لا تتضمن أية مناقشات أو تفسيرات لقيم حجم الأثر	– حجم العينة الكبير	– غياب الدلالة العيادية في عملية التحليل. – اعتماد أبحاث مجلات دوناً عن الرسائل الأكاديمية
(Eisen et al ,2007)	مقارنة نتائج الاستجابة على مقياس BASIS-(24) باستخدام: مؤشر التغيُّر الموثوق، وحجم الأثر، والخطأ المعياري	– حساب التغيُّرات على مستوى المجموعة أغفل حقيقة عدم ظهور أي تغيّرٍ بالاتجاه المطلوب لدى نسبة كبيرة من الأفراد، وأن بعضهم أظهروا تراجعاً عيادياً	– الجمع بين الدلالات الثلاث في حساب ومقارنة النتائج – حجم العينة الكبير	– إغفال دراسة العلاقات بين مؤشرات الدلالات الثلاث
Ronk) & et	مقارنة نسب الاتفاق بين تصنيفات طريقتي	– عدم وجود فرق ذو دلالة إحصائية بين التصنيفات	– حجم العينة الكبير – التنوع في أدوات البحث	– غياب حساب الدلالتين الإحصائية والعملية

al, (2016	(JT) و (HA)	العيادية الناتجة عن طريقتي (JT) و(HA) – التوصية باستخدام طريقة (JT) لأنها الأسهل		– الاقتصار على مقارنة تصنيفات طريقتي (JT) و(HA) فقط
Ronk) & et al, (2014	مقارنة نسب الاتفاق بين تصنيفات الطرائق الخمس لحساب الدلالة العيادية	– يوجد اتفاق بين تصنيفات أربع طرائق باستثناء طريقة (HA) – صنفت طريقة (HA) عدداً أقل من حالات التحسن والتعافي مقارنةً ببقية الطرائق	– مقارنة تصنيفات الطرائق العيادية جميعها – حجم العينة الكبير – التنوع في أدوات البحث	– غياب حساب الدلالتين الإحصائية والعملية
(Shada , 2013)	مقارنة نسب الاتفاق بين تصنيفات أربع طرائق لحساب الدلالة العيادية	– وجود اختلافات في التصنيفات بين الطرائق الأربع باستثناء طريقتي (JT) و(GLN) – أسفرت طريقة (HA) عن تصنيفات متناقضة مع بقية الطرائق – قدمت طريقة (JT) تقديراً معتدلاً للتغيير الموثوق	– مقارنة تصنيفات أربع طرائق عيادية – أهمية نتائجها المتعلقة بطريقة (HA) واختلافها عن بقية الطرائق	– غياب حساب الدلالتين الإحصائية والعملية
Atkins & et al, (2005	مقارنة نسب الاتفاق بين تصنيفات أربع طرائق لحساب الدلالة العيادية	– وجود اتفاق دال بين تصنيفات الطرائق الأربع – طريقة (EN) أكثر قدرةً على كشف حالات التحسن خلافاً لطريقةHA – طريقتا (JT) و (GLN) أدتا إلى تصنيفات عيادية متطابقة تقريباً	– مقارنة تصنيفات أربع طرائق لحساب الدلالة العيادية – اعتماد أسلوب محاكاة وتوليد البيانات	– غياب حساب الدلالتين الإحصائية والعملية
Bauer) & et al, (2004	مقارنة نسب الاتفاق بين تصنيفات الطرائق الخمس لحساب الدلالة العيادية	– وجود فروق ذات دلالة بين التصنيفات الناتجة عن الطرائق الخمس باستثناء (JT) و (GLN) – التوصية باستخدام طريقة (JT)	– مقارنة تصنيفات الطرائق العيادية جميعها	– غياب حساب الدلالتين الإحصائية والعملية
(رشوان، 2017)	دراسة مدى تأثّر الخصائص القياسية للاختبار التحصيلي باختلاف استخدام أفراد العينة لإستراتيجيات أداء الاختبار	– تحسُّن الصدق التنبؤي والتلازمي للاختبار التحصيلي بازدياد استخدام الإستراتيجيات. – تحسُّن الثبات بتأثير استخدام الإستراتيجيات ولكن بدرجة أقل من الصدق	– تنوع الإستراتيجيات المدروسة كماً ونوعاً – دراسة استراتيجية ما بعد الاختبار	– عدم استخدام برنامج تدريبي – اقتصار الأدوات على مقياس للإستراتيجيات – دراسة أثر التدريب في ضوء الدلالة الإحصائية وإغفال الدلالتين العملية والعيادية
Pohl) & et al., (2016	دراسة أثر الاختلافات في سلوك أداء الاختبار، - وبالتحديد التخمين وعدم الإجابة-	– سلوك التخمين يرتبط بانخفاض الخصائص القياسية لنتائج مجموعة صعوبات التعلم.	– التحليل باستخدام نموذج Rash	– عدم استخدام برنامج تدريبي – الاكتفاء بسلوكي التخمين وعدم الإجابة

	في تفسير انخفاض معاملات الخصائص القياسية.	– سلوك الإجابات المفقودة يفسر انخفاض الخصائص القياسية في درجات العينة الكلية		– دراسة النتائج استناداً إلى الدلالة الإحصائية وإغفال الدلالتين العملية والعيادية.
Maspons & Llabre (, 1985	دراسة أثر تدريب الطلبة الإسبانيين على مهارات الأداء في الخصائص القياسية للاختبار	– انخفاض معامل الاتساق الداخلي في المجموعة التجريبية بشكل طفيف عن مثيله في الضابطة. – ارتفاع معامل الصدق التنبؤي لأفراد المجموعة التجريبية بتأثير التدريب.	– الاعتماد على برنامج تدريبي حول سلوكات أداء الاختبار.	– دراسة أثر التدريب استناداً إلى الدلالة الإحصائية، وإغفال الدلالتين العملية والعيادية

5- موقع البحث الحالي من الأبحاث السابقة:

تشابه البحث الحالي في عرض موضوع الدلالات العلمية مع الأبحاث التي ركزت على تمييز الأنواع الثلاثة من الدلالة، ولم تكتف بنوعٍ واحدٍ أو نوعين منها، وجمع في عيناته الأنواع المختلفة من العينات المستخدمة في الأبحاث السابقة والمتمثلة في: عينة الرسائل الأكاديمية، وعينة الأبحاث المجمعة إضافةً إلى عينة البحث التجريبية، واتفق مع الأبحاث السابقة التي استهدفت مقارنة أربع طرائق في حساب الدلالة العيادية، كما في بحثي (Shada, 2013) و (Atkins & et al, 2005).

في حين تميّز هذا البحث عن الأبحاث السابقة جميعها في تناوله تحليل واقع استخدام الدلالات العلمية الثلاث معاً، ودراسة العلاقات بينها ومدى كفاية كل منها في الإبلاغ عن نتائج البحوث، كما جمع البحث الحالي المناهج البحثية الثلاث: الوصفي التحليلي، والتحليل البعدي، والمنهج التجريبي، خلافاً للأبحاث السابقة جميعها التي اكتفى كلٌ منها بمنهجٍ بحثيٍ واحد. وتميز بدراسته أثر متغير التدريب في النتائج والخصائص القياسية للاختبار في ضوء الدلالات العلمية الثلاث، خلافاً لجميع الأبحاث السابقة التي اعتمدت على الدلالة الإحصائية وحدها في دراسة هذا المتغير.

خلاصة:

قدّم هذا الفصل عرضاً مفصلاً لأبحاثٍ سابقة تناولت موضوع الدلالات العلمية، مبتدئاً بالأحدث وحتى الأقدم، عُرِضت ضمن المحاور الثلاثة الأساسية: الدلالات العلمية، الدلالة العيادية، وأثر متغير التدريب في النتائج البحثية والخصائص القياسية للاختبار.

وقد بيّن هذا الفصل وجود عدد لابأس به من الدراسات والأبحاث التي تناولت كل من المحورين الأول والثاني، ووجود ثلاث دراسات فقط تناولت المحور الثالث، من دون أن يكون الهدف الأساسي في أي منها تحليل واقع استخدام الدلالات الثلاث معاً، أو دراسة العلاقات بينها جميعاً، وكفاية كل منها في الإبلاغ عن نتائج الأبحاث، ومدى تأثرها بعوامل تتعلق بالموقف الاختباري أو التدريبي الذي يخضع له المفحوصون.

الفصل الثالث

الإطار النظري

الفصل الثالث

الإطار النظري

تمهيد:

يتناول هذا الفصل عرضاً لمفهوم الدلالة الإحصائية، ومستوى الدلالة، والقيمة الاحتمالية، ويوضح أنواع الفروض، والأخطاء التحليلية والتفسيرية المرتبطة بها، ثم يعرض مفهوم الدلالة العملية وأهميته، ويبين أهم أنواع مقاييس حجم الأثر وفوائد استخدامها، والانتقادات الموجهة لها، ويقدم شرحاً لمفهومي الدلالة العيادية ومؤشر التغيُّر الموثوق، ويعرض الطرائق الرياضية المُستخدمة في حساب هذا النوع من الدلالة، ومجالات استخدامها وتوظيفها، وأهم الانتقادات الموجهة لها. كما يتناول مفهوم التدريب على الاختبار، وأثره في الخصائص القياسية، ويوضح معنى إستراتيجيات أداء الاختبار، والتصنيف المُعتمد لها في البحث الحالي، ويبين أهمية توظيف هذه الإستراتيجيات في المجالات المختلفة.

أولاً: أنواع الدلالات العلمية:

على الرغم من أن الدلالات العلمية الثلاث تُعد مفاهيم أساسية وتمثّل مستوياتٍ من الدقة الإحصائية لابد من توافرها في أي بحثٍ علمي كمي، إلا أن الأدبيات والأبحاث السابقة وثقت الخلط الكبير في استخدام هذه المفاهيم في أبحاث العلوم السلوكية، إذ أشار طومسون (Thompson, 2006) إلى أنه جرى الخلط ولعقودٍ طويلةٍ من الزمن بين مفهومي الدلالة الإحصائية والعملية، وأكدت بترسون (Peterson, 2008) أنه غالباً ما تمّ الخلط بين مفهومي الدلالة العيادية والدلالة العملية، وأن ذلك يعود لافتراض الباحثين أن الدلالة العيادية مجرد طريقة إضافية لوصف الدلالة العملية، وحجوم الأثر المحسوبة للاختبارات الإحصائية (Peterson, 2008, P.P 3 - 6). لذا كان لا بد من توضيح نشأة ومفهوم كل نوع من الدلالات آنفة الذكر، وبيان وأهميته وطرائق حسابه، والانتقادات الموجهة له، والتي أدت لظهور النوع الذي يليه، وصولاً إلى الدلالة العيادية.

1- الدلالة الإحصائية:

هي مصطلحٌ إحصائي يُستخدم في اختبار الفرضيات، وقد استخدم العلماء اختبارات الدلالة الإحصائية لأغراض البحث العلمي منذ أوائل القرن الثامن عشر، وعلى الرغم من تقدم تطبيقات الاختبار الإحصائي بشكلٍ كبير، لاسيما مع ظهور الكمبيوتر ووسائل التكنولوجيا الحديثة، فقد ظلّت الكثير من الاختبارات الإحصائية إلى هذا اليوم تستند إلى نفس المنطق المُستخدم في الاختبارات الإحصائية الأولى، المُقدَّمة في أوائل القرن العشرين من خلال عمل فيشر وبيرسون ونيمان (Fisher, Pearson &

Neyman, 1925)، ومازال إلى الآن اختبار الدلالة، واختبار الفرضية الصفرية يشكلان حجر الزاوية في الأوراق والأبحاث العلمية (Mclean & Ernest, 1997, p.p4-5).

تهدف اختبارات الدلالة الإحصائية إلى اشتقاق استنتاجات عن خصائص المجتمع من خلال دراسة خصائص عينة مأخوذة منه، وتسعى إلى الكشف عن مدى اقتراب المقاييس الإحصائية للعينات من مقاييس المجتمع، وتزداد الثقة في مقاييس العينة كلما اقتربت من مقاييس المجتمع الأصلي، أو كلما كان تذبذبها حوله ضيقاً، أي بمعنى آخر كلما كان انحرافها عن مقاييس الأصل صغيراً، وتحاول هذه الاختبارات الإجابة عن أسئلة مثل: أي من النتائج التي تم الحصول عليها ترجع إلى الصدفة؟ (عايش، 2019، ص90، البهي السيد، 2005، ص309).

كما تبين مقاييس الدلالة الإحصائية للباحث إلى أي حد يستطيع أن يتأكد من ثبات نتائجه ويثق بها، وإلى أي حد يستطيع إرجاعها إلى عامل المصادفة (أبو النيل، 1987، ص.222).

والدلالة الإحصائية تعني ببساطة الندرة الإحصائية، فالنتيجة من وجهة النظر الإحصائية تكون دالة إحصائياً لأنها تحدث بشكلٍ نادرٍ جداً في العينة العشوائية تحت ظروف الفرضية الصفرية، وهي بذلك تسمح للباحث باتخاذ قرار (نعم / لا) فيما يتعلق برفض أو عدم رفض الفرض الصفري(Huston, 1993, P.3).

- **مستوى الدلالة والقيمة الاحتمالية:**

يشير مستوى الدلالة إلى الحد الاقصى لاحتمال الوقوع في الخطأ من النوع الأول أو خطأ ألفا (α) عند اختبار الفرض الصفري، وتعني كلمة دلالة أن الفرق بين القيمة النظرية للبارامتر في المجتمع والقيمة الناتجة من العينة فرق حقيقي ولا يرجع إلى المصادفة.

توجد في اختبارات الدلالة الإحصائية قيمتان احتماليتان: الأولى هي (α) النظرية وتسمى أيضاً $p_{critical}$ ، والثانية هي القيمة الاحتمالية المحسوبة $p_{calculated}$، فعند حساب القيمة الاحتمالية p المحسوبة تُقارن بمستوى الدلالة **(α)**؛ أي بقيمة p النظرية، فإذا كانت أقل منها، أو تساويها يُرفض الفرض الصفري، وتحدد الفرضية على أنها دالة إحصائياً، أما إذا كانت أكبر من مستوى الدلالة **(α)** يتم قبول الفرض الصفري، وتتحدد الفرضية على أنها غير دالة إحصائياً (محمد حسن، 2011، ص.201، بابطين،2001 ، ص.16).

أما حجم أو مقدار هذا المستوى الذي يجب أن يحدده الباحث لكي يرفض الفرض الصفري، فهناك شبه اتفاق على أن مستويات الدلالة (0.05)، (0.01)، (0.001) هي التي يُفضّل اتخاذها معياراً للحكم، وعلى الرغم من أن الاتفاق على استخدام هذه القيم من مستويات الدلالة تساعد الباحثين على مقارنة نتائج أبحاثهم مع نتائج الأبحاث والتجارب الأخرى، لأن اختيار كل باحث لمستوى دلالة حسب رغبته سيؤدي إلى صعوبة في إجراء هذه المقارنات، إلا أن هذا التحديد لمستويات الدلالة الإحصائية انتُقد من قبل العديد من العلماء والباحثين مثل هوبكنز(Hopkins, 2014) وماكلين وإرنيست (Mclean & Ernest, 1997) وكافر (Carver,1993) وهيوستن (Huston, 1993) وغيرهم، ووصفوها بالمستويات التعسفية، وأشار بعض النقاد إلى أن فيشر نفسه (Fisher, 1925) الذي اقترح المستوى (0.05) معياراً للدلالة الإحصائية، ذكر أنه لا يوجد عمل علمي لديه مستوى ثابت من مستويات الدلالة في جميع الظروف، بل دعا الباحثين إلى تحديد وتفسير قيم مستوى الدلالة في ضوء طبيعة أبحاثهم، والسياق العام الذي أجريت فيه (عايش، 2019، ص. 95، حبيب، 2000، ص ص. 331-332).

- **أنواع الفروض:**

الفرض هو جملة علمية تعبر عن توقع أو احتمال أو تخمين، أو إجابة مؤقتة لتساؤلٍ علمي، يحاول الباحث من خلال سلسلة من الإجراءات أن يتحقق منها، وتحتاج صحتها إلى تحقق وإثبات، لذلك يستخدم الباحث الوسائل المناسبة لجمع الأدلة والبيانات التي تثبت صحة الفرض أو تدحضه (محمد حسن،2011 ، ص.190).

تنقسم الفروض بوجهٍ عام إلى نوعين أساسيين هما: العلمي والإحصائي:

- **الفرض العلمي أو البحثي:** هو حل مقترح للمشكلة ويصاغ بشكل استنتاجي، ويتم التحقق منه في ضوء التحليل النظري للخبرات أو المعلومات السابقة، ويجري اشتقاقه بطريقة مباشرة من إطار نظري معين، مثال: يزداد رضا المعلم عن العمل بانخفاض شعوره بضغط العمل، وهذا النوع من الفروض لا يمكن اختباره بصورة مباشرة، لأنه لا يحدد بارامترات معينة لإجراء المقارنات أو دراسة العلاقات بينها؛ لذا يتجه الباحث للفرض الإحصائي لاختبار فروض دراسته. إذ يرى (فهمي،2005) أن أية دراسة يستخدم فيها الباحث الفروض البحثية يحتاج إلى تحويلها إلى فروض إحصائية، وبالتالي ترجمة الفرض البحثي رياضياً، واختباره إحصائياً، بحيث يحدد نوع العلاقات بين المتغيرات، ومقدارها واتجاهها (فهمي، 2005، ص.331).
- **الفرض الإحصائي:** عندما يتم التعبير عن الفرض العلمي أو البحثي بصورة رمزية عددية يسمى فرضاً إحصائياً، وهو الفرض الذي يضعه الباحث في صورة جملة رقمية عن أحد معالم –بارامترات–

مجتمع البحث، ويتم التحقق منه بأسلوب إحصائي، لذا فإن الاختبار الإحصائي هو الذي يحدد قبوله أو رفضه. والفروض الإحصائية نوعان هما:

- **الفرض الصفري أو فرض العدم:** يرمز له بالرمز (H_0)، وهو الذي ينفي وجود الظاهرة، وبالتالي ينفي وجود فرق أو علاقة، أو تأثير للمتغير المستقل في المتغير التابع.
- **الفرض البديل:** يرمز له بالرمز (H_1)، وهو الذي يمكن أن يقبل بديلاً للفرض الصفري عندما يتم رفضه، ويؤكد بوجهٍ عام وجود الظاهرة، أي وجود الفرق أو العلاقة أو التأثير للمتغير المستقل في المتغير التابع (محمد حسن، 2011، ص ص.191- 194، علام،2005، ص ص.95–99).

- الأخطاء المرتبطة باختبارات الفروض:

عند اتخاذ قرار برفض أو قبول الفرض الصفري هناك أربعة احتمالات لهذا القرار هي:

الاحتمال الأول: رفض الفرض الصفري وهو في حقيقة الأمر صحيح، وفي هذه الحالة يستنتج الباحث وجود فروق أو علاقات غير موجودة أصلاً، ويسمى بالخطأ من النوع الأول (Type I Error)، ويرمز له ألفا Alpha (α).

وقد يكون السبب في حصول هذا النوع من الخطأ عدم تمثيل العينة للمجتمع، فيكون متوسطها في الذكاء مثلاً أعلى من متوسط المجتمع، أو أن البرنامج التجريبي مثلاً فعال بالنسبة للعينة لسبب ما، أكثر مما هو فعال بالنسبة للمجتمع.

الاحتمال الثاني: قبول الفرض الصفري وهو في حقيقة الأمر خاطئ، أي فشل الباحث في اكتشاف الفروق أو العلاقات بين المجموعات عندما تكون حقيقية وموجودة في الواقع، ويسمى بالخطأ من النوع الثاني (Type II Error)، ويرمز له بيتا Beta (β). مثل أن يفشل في اكتشاف الفروق بين الذكور والإناث في متغير ما، مع أن هناك فروقاً دالة وواقعية، لكن المشكلة أن العينة كانت غير ممثلة، أو طريقة الاختبار التي تعرضوا لها غير مناسبة.

الاحتمال الثالث: قبول الفرض الصفري وهو في حقيقة الأمر صحيح، وفي هذه الحالة لا يكتشف الباحث وجود فروق أو علاقات بين المتغيرات، وهي في الواقع غير موجودة، وهذا ما يدعى بمستوى الثقة ويساوي $(1-\alpha)$.

الاحتمال الرابع: رفض الفرض الصفري وهو في حقيقة الأمر خاطئ، أي إن الباحث يتوصل إلى وجود فروق أو علاقات ذات دلالة إحصائية، وهي حقيقية وموجودة في الواقع، ويسمى هذا قوة الاختبار الإحصائي Power of the test، ويساوي $(1-\beta)$ (المنيزل وغرايبة، 2006، ص.211).

وتشير أدبيات البحث إلى أن الأخطاء التحليلية والتفسيرية المرتبطة باختبارات الدلالة الإحصائية لا تقتصر فقط على نوعي الخطأ الأول والثاني، المعروفين لدى معظم الباحثين والإحصائيين، بل هناك أخطاءٌ أخرى تتعلق باختبار الفرضيات مثل:

- **الخطأ من النوع الثالث Type III Error:** يتم الوقوع فيه عندما يقوم الباحث بعمل استدلالات خاطئة حول اتجاهية النتيجة، كأن يقوم برفض الفرض الصفري بشكل صحيح، ولكن الاستدلال بشكل خاطئ على اتجاه الفرق أو التأثير.
- **الخطأ من النوع الرابع Type IV Error:** يتم الوقوع فيه عندما يقوم الباحث بتقديم تفسير خاطئ لفرض مرفوض بشكل صحيح.
- **الخطأ من النوع الخامس Type V Error:** يتصل بخطأ الإعادة أو التكرار الداخلي، ويحدث عند الوقوع في الخطأ من النوع الأول أو الثاني في أثناء سلسلة الإعادة أو التكرار عند استخدام إجراءات منهجية متسلسلة.
- **الخطأ من النوع السادس Type VI Error:** يتصل بخطأ تعميم الثبات، ويحدث عندما يفشل الباحثون في الأخذ بالاعتبار تقديرات الثبات المتفاوتة للعينات الفرعية ضمن مجموعة البيانات.
- **الخطأ من النوع السابع Type VII Error:** يتصل بخطأ عدم تجانس التباين أو الانحدار، ويتم الوقوع فيه عندما تكون البيانات المفحوصة في تحليل التباين غير مستوفية افتراضات التجانس.
- **الخطأ من النوع الثامن Type VIII Error:** يتم الوقوع فيه عندما يضع الباحثون فروضاً بديلة موجهة، ويتم اختبار النتائج باختبارات دلالة الطرفين وليس الطرف الواحد.
- **الخطأ من النوع التاسع Type IX Error:** يتصل بخطأ تحيز المعاينة.
- **الخطأ من النوع العاشر Type X Error:** يتصل بخطأ درجات الحرية (محمد حسن، 2011، ص ص. 199- 201).

- العوامل المؤثرة في اختبارات الدلالة الإحصائية والانتقادات الموجهة إليها:

ساد الكثير من الجدل في الأدبيات والأبحاث المتصلة بالدلالة الإحصائية حول سوء استخدام اختباراتها وتفسير نتائجها، وتطرقت العديد من المقالات والأبحاث إلى المشكلات المتصلة باختبارات هذا النوع من الدلالة، والتي أثارت موجة من الانتقادات الموجهة إليها، وقد تركزت بمعظمها في النقاط الآتية:

- **تأثير حجم العينة في نتيجة اختبارات الدلالة الإحصائية:** لعل الانتقاد الأكثر تكراراً في أدبيات البحث، والموجه لاختبارات الدلالة الإحصائية، هو حساسيتها لحجم العينة، فقد بيّن هيوستن (Huston, 1993) أنه عند زيادة حجم العينة إلى حدٍ معين يمكن أن تصبح أية نتيجة غير دالة إحصائياً ذات دلالة، والعكس صحيح، وأكد أنه يجب دائماً الإبلاغ عن نتائج الاختبارات الإحصائية في سياق عدد أفراد العينة (Huston, 1993, P.3).

كما أكد طومسون (Thompson, 1989)، وشنايدر ودارسي (Schneider& Darcy, 1984) وميهل(Meehl, 1967) وشافير (Shaver, 1993) أهمية حجم العينة في عملية صنع القرار المرتبط بالفرض الصفري، حيث بيّن طومسون أن تغييراً طفيفاً في حجم العينة قد يؤثر في هذا القرار، إذ يمكن الحصول على فروق دالة إحصائياً لمتوسطات متقاربة القيم فقط لأن الباحث قد اختار عينة كبيرة في دراسته، ووضح ماكلين وإرنست (Mclean & Ernest, 1997) أن الدلالة الإحصائية يمكن تحقيقها عند حجم معين من العينة، أي إن المسألة ببساطة تصبح في الحصول على عددٍ كافٍ من أفراد العينة لتحقيقها (Mclean & Ernest, Shaver, 1993, P.302-303 1997, pp.7-10,)، ذلك لأن قيمة الخطأ المعياري للقياس تقل كلما كبر حجم العينة، وبذلك يكون تحقُّق الدلالة الإحصائية نتيجة لحجم العينة، بدلاً من أن يكون نتيجة للفروق العملية والواقعية بين المتغيرات، وبالتالي يمكن لأي باحث يعتقد بوجود علاقة ضعيفة بين متغيرات دراسته، أن يزيد حجم عينته للحصول على نتائج ذات دلالة إحصائية (عايش، 2019، ص.94، البارقي، 2012، ص.30).

- **مستوى الدلالة الإحصائية:** من الانتقادات الموجهة لاختبارات الدلالة الإحصائية اعتمادها قيماً محددة لمستوى الدلالة، وقد وصِفت هذه القيم بالمستويات التعسفية من قبل العديد من العلماء والباحثين، مثل هوبكنز(Hopkins, 2014)، وماكلين وإرنيست (Mclean & Ernest, 1997)، وكافر(Carver,1993)، وهيوستن (Huston, 1993). حيث ذكر شنايدر (Schneider, 2011) أن القيم (0.05) أو (0.01) أو (0.001) لمستوى الدلالة تُستخدم بشكل روتيني من قبل الباحثين منذ أن اقترحها فيشر (Fisher, 1925)، وبمجرد أن تكون قيمة P المحسوبة أصغر من هذه القيم تُرفض الفرضية الصفرية، ويتم الإبلاغ عن نتيجة ذات دلالة إحصائية (Schneider, 2011, p.5)، ووضح بينتيا (Pintea, 2010) أنه قد يتم في بعض الأحيان إساءة تفسير قيم P المحسوبة بناءً على مستوى الدلالة، كأن يتم تفسير قيم P الأصغر من مستوى الدلالة (0.01) على أنها أقوى من القيم الأصغر من مستوى الدلالة (0.05)، وهذا الاعتقاد في حقيقة الأمر خاطئ، فربما نتجت تلك القيم عن حجم العينة المُستخدمة، وربما كانت قيم P الدالة عند مستوى الدلالة (0.05) مرتبطة بحجم أثر أكبر؛ لذا فإنه لا يكفي معيار الدلالة الإحصائية وحده لفهم وتفسير معنى النتائج (Pintea, 2010, P.102).
- **التشتت في المتغيرات ذات العلاقة:** يُعتبر من العوامل الأساسية التي تؤثر في اختبارات الدلالة الإحصائية، إذ يُلاحَظ أن تشتت الدرجات يؤثر في القيمة الاحتمالية المحسوبة لـ (*P*)، إذ أن زيادة هذا التشتت والتغاير ينتج قيمة أعلى لـ (*P*) المحسوبة، وعلى العكس يؤدي انخفاض التشتت في

الدرجات إلى انخفاض قيمة (*P*) المحسوبة، وبالتالي الحصول على نتائج دالة إحصائياً (بابطين، 2001، ص.24).

ومن ناحيةٍ أخرى يرى طومسون (Thompson, 1993) أن حساب القيم الاحتمالية (P) لا تستطيع تكوين منظومة تساعد على استخلاص نتائج حقيقية عن قيم أفراد المجموعة، وأنه لا تُعزى أية قيم إنسانية على مستوى الأفراد إلى نتائج الفرض الصفري، ذلك لأنها عبارة فقط عن قيم احتمالية، مما يجعل من الصعب اتخاذ قرار فيما إذا كانت نتائج الأبحاث ذات أهمية للأطباء، أو المعالجين، أو المرشدين النفسيين، أو المعلمين ،أو أي شخص يأمل في استخدام وتوظيف المعرفة المكتسبة من البحث في العالم الحقيقي والواقعي (Peterson, 2008, p.4).

ومع كل الانتقادات الموجهة إلى اختبارات الدلالة الإحصائية فقد ظلت المعيار الأساسي في تحليل البيانات للعديد من المجالات البحثية، إذ أكد كارفر(Carver, 1993) استمرار هيمنة هذا النوع من الدلالة في تفسير البيانات الكمية في أبحاث العلوم الاجتماعية، لذا قدم في مقالٍ له عام (1993) أربع اقتراحات على الباحثين الالتزام بها، للتقليل من الأهمية المتصورة للدلالة الإحصائية في أثناء الإبلاغ عن نتائج أبحاثهم، وهذه الاقتراحات على النحو الآتي:

- ❖ الإصرار على إدراج كلمة إحصائياً أمام كلمة هام أو ذو دلالة، عند كتابة تقارير الأبحاث.
- ❖ الإصرار على تفسير النتائج في ضوء طبيعة الأبحاث وسياقها العام، بما في ذلك نوعية البيانات، وحجوم العينات المستخدمة فيها.
- ❖ الإصرار على حساب حجوم الأثر وتفسيرها إلى جانب اختبارات الدلالة الإحصائية، بغض النظر عما إذا كانت النتيجة ذات دلالة أم لا، والحكمة في ذلك هي أن حجوم الأثر حتى التافهة أو المنخفضة منها قادرة على إظهار النتائج بوضوح أكبر، وأن النتائج ذات حجوم الأثر المنخفضة بغض النظر عن وجود الدلالة الإحصائية، هي مؤشرات على عدم وجود أدلة كافية لرفض الفرضية الصفرية.
- ❖ الإصرار على أن يقدم محررو المجلات العلمية وجهات نظرهم وآرائهم في نتائج اختبارات الدلالة الإحصائية قبل اختيار الأبحاث، وذلك لضمان اعتماد الأبحاث الجيدة والجديرة بالنشر فقط (Carver, 1993, PP. 288-289).

2- الدلالة العملية وحجم الأثر:

- مفهوم الدلالة العملية وحجم الأثر:

يُقصد بالدلالة العملية مقدار التباين في المتغير التابع الذي يمكن تفسيره من خلال المتغير المستقل الذي يرتبط به أو يؤثر فيه (النجار، 2006، ص.267)، ويحدد هذا النوع من الدلالة المقدار الفعلي للفروق بين المتوسطات، وقوة العلاقات بين المتغيرات؛ ويبيّن ما إذا كانت هذه الفروق أو العلاقات كبيرة إلى الدرجة التي تبرر الأخذ بنتائجها (أبو جراد، 2013، ص.352).

فالدلالة الإحصائية تقيّم فقط علاقات ترتيبية، مثلاً: إذا كان هناك فرق دال بين متوسطي مجموعتين، وما إذا كان أحدهما أكبر من الآخر، في الوقت الذي يريد الباحث فيه معرفة حجم الفرق بين المتوسطين، وهل هو جدير بالاهتمام أم لا (Thompson, 2002, p.p 65).

وفي عام 1962 أطلق كوهين أول استفسار عن قوة الاختبار الإحصائي في الأبحاث المنشورة، وبرز ما يعرف باسم الدلالة العملية، ومنذ ذلك الحين أصبح الجدل بخصوص الدلالة الإحصائية والدلالة العملية يمثل قضية بحثية في العلوم الاجتماعية (Huston, 1993, p.11).

وقد أوصى العديد من الباحثين والجمعيات الأكاديمية، وسياسات التحرير، وأدلة النشر بوجوب تضمين أحد مقاييس حجم الأثر في نتائج الأبحاث جنباً إلى جنب مع اختبارات الدلالة الإحصائية، وأن يتم تفسير كلاً منهما، وبينت الجمعية الأمريكية للإحصاء "ASA" عام 2016 أنه لا يكفي الاعتماد على الدلالة الإحصائية وحدها مؤشراً ودليلاً على دقة وموثوقية النتائج، وصحة القرارات العلمية المرتبطة بها، بل لا بد من اعتماد مستوى أكثر دقة من الدلالة يتمثل بحساب حجوم الأثر (Hansen et al., 2018, P.802).

وأكد دليل نشر الأبحاث السابع الصادر عن جمعية علم النفس الأمريكية "APA" أن اختبارات الفرضيات الإحصائية ليست سوى نقطة الانطلاق في حساب النتائج، وأنه لابد من حساب قيم حجم الأثر في بيان نتائج الأبحاث الوصفية والتجريبية على حدٍ سواء، واعتبر أن الإبلاغ عن قيم حجم الأثر وتفسيرها في فصل النتائج علامة على البحث الموثوق، لأنها تزوّد القارئ بمعلوماتٍ كافية لتقييم مقدار التأثير أو العلاقة المشاهدة، مما يسهم في فهم النتائج بشكلٍ أوضح، ويتيح فرصة الاستفادة منها. (APA, 2020, P.P 154 -155)

تقوم الدلالة العملية في جوهرها على حساب حجم الأثر، وهو كما يعرفه كوهين (Cohen, 1988) يدل على الدرجة التي توجد فيها الظاهرة في المجتمع، أو الدرجة التي يكون فيها الفرض الصفري خاطئاً (Akbaryan, 2013, p.5). ويوضح علام أنه عندما يكون الفرض الصفري خاطئاً، فإنه يكون خاطئاً بدرجة معينة، وأن حجم الأثر يُعد مقياساً لدرجة خطأ الفرض الصفري (علام، 2005، ص. 123).

وفي الأبحاث التجريبية يُعرَّف حجم الأثر بأنه: أسلوبٌ إحصائيٌ يحدد مقدار تأثير العلاج أو التدخل في أية دراسة بحثية، أي أنه يحدد الحجم النسبي للمعالجة التجريبية (Peterson, 2008, p.4).

يتبين مما سبق أن الحاجة ملحة لضرورة تضمين مقاييس حجم الأثر في نتائج الأبحاث العلمية على اختلاف أنواعها، سواءٌ أكانت تجريبية أم وصفية تحليلية، إذ أصبح من الضروري أن يضع الباحثون في اعتبارهم أسئلة مهمة مثل: هل الأثر المُلاحظ حقيقي أم يرجع إلى المصادفة؟ وإذا كان الأثر حقيقياً فما حجمه؟ وهل هو كبير بدرجة كافية لأن يكون مفيداً؟، وهذه المعلومات لا يستطيع الباحثون الحصول عليها إلا من خلال حساب قيم حجوم الأثر.

- **فوائد حساب حجم الأثر:**

لحساب حجم الأثر فوائد عديدة، يمكن إجمالها في النقاط الآتية:

- تسمح مؤشراته بتجاوز التبسيط الموجود في اختبارات الدلالة الإحصائية، فبدلاً من مجرد إصدار حكم بوجود فروق ذات دلالة أم لا، تساعد حجوم الأثر في إجراء العمليات الأكثر تعقيداً والتي تسمح بالحكم على مدى الأهمية والقيمة الفعلية لهذه الفروقات، وهي بذلك تزود الباحثين بمعلومات تتجاوز حدود اختبارات الدلالة الإحصائية، وتمنع سوء تفسيرها كمؤشرات للأهمية (Coe, 2002, p.1).
- تخفف من تأثير حجم العينة، وهيمنة مستوى الدلالة، في تفسير النتائج (أبو هاشم،2005، ص.20).
- تشير مؤشراته إلى درجة وجود الظاهرة في المجتمع بمقياس متصل، بحيث يعني الصفر عدم وجود الظاهرة (بابطين، 2001، ص.94).
- يمكن استخدام مؤشراته في المقارنات الكمية بين نتائج الدراسات والأبحاث المختلفة، إذ يُعد أداةً مهمة في منهج التحليل البعدي (Thompson, 2004, p. 480).
- يُعد مقياساً دقيقاً حتى في حالة النتائج الصغيرة وغير الواضحة؛ أي ذات حجم الأثر الضئيل.
- تُعد مؤشراته عناصر أساسية في كتابة التقارير البحثية، كما تلعب دوراً مهماً في الجوانب التفسيرية لنتائج الأبحاث وتوضيح أهميتها (بدوي، 2018، ص.445).
- يمكن استخدامه في حساب وتحليل قوة الاختبار الإحصائي، إذ تُعد المقاييس المعيارية لحجم الأثر قاعدة أساسية للجداول والعمليات الحسابية، التي تُستخدم في حساب القوة الإحصائية (Akbaryan, 2013, p.5).

- **أنواع مقاييس حجم الأثر:**

يُطلق مصطلح حجم الأثر "ES" على عائلة من القوانين الرياضية التي تقيس مقدار الفروق أو العلاقات بين المتغيرات، تختلف وتتنوع طرائق حسابه تبعاً للأسلوب الإحصائي المُستخدم لاختبار الفرض الصفري، وقد ذكر هيوستن ببحثه أن ماكسويل وديلاني صنفا(Maxwell & Delaney, 1990) الصيغ الرياضية لقوانين حجم الأثر في طائفتين أساسيتين هما:

- مقاييس مبنية على أساس الفروق المعيارية بين المتوسطات، تُسمى مقاييس الفروق المعيارية، مثل: d ل Cohen و g ل Hedges، وبالمجمل لها الشكل الرياضي الآتي:

$$d = \frac{x\ group1 - x\ group2}{SD_{pooled}}$$

- مقاييس مبنية على أساس الارتباط أو قوة العلاقة بين المتغيرات، وتُسمى مقاييس التباين المفسر؛ لأنها توضح مقدار التباين في المتغير التابع الذي يمكن تفسيره من خلال التباين في المتغير المستقل الذي يرتبط به أو يؤثر فيه، وهذه المقاييس مثل: $\boldsymbol{\eta^2}$ و R^2، وتعتمد بالمجمل الصيغة الرياضية الآتية:

$$R^2 = \frac{Sum\ of\ Squares_{group}}{Sum\ of\ Squares_{Total}}$$

(Huston, 1993, p.13)

وفيما بعد تمكّن كيرك من وصف 40 مقياساً لحساب حجم الأثر، جرى تصنيفها في بُعدين أساسيين هما:

- مقاييس الفروق المعيارية في مقابل مقاييس التباين المفسر.
- مقاييس حجم الأثر غير المُصححة في مقابل مقاييس حجم الأثر المُصححة (Kirk, 1996, p. 748).

ويُقصد بمقاييس حجم الأثر المُصححة: تلك التي تعزل إحصائياً التحيز الصاعد في تقدير القيمة الحقيقية للمجتمع من بيانات العينة، لأن جميع إحصائيات حساب التباين غير المصححة هي طرائق ترابطية، تستفيد من جميع التباينات في العينة بما في ذلك تباين أخطاء العينة، فهي منحازة إيجابياً وتبالغ في التقديرات، لذا تأخذ مقاييس حجم الأثر المُصححة في الحسبان العوامل التي تسهم في أخطاء العينة مثل:

- **حجم العينة:** فكلما صغر حجم العينة ارتفع احتمال ازدياد أخطاء المعاينة.
- **عدد المتغيرات في البحث:** كلما زاد عدد المتغيرات في البحث الواحد ارتفع احتمال ازدياد أخطاء المعاينة (Thompson, 1999, p.171).

وفيما يلي عرضاً لأكثر هذه المقاييس انتشاراً، كما أوردتها العديد من أدبيات البحث ذات الصلة مثل:

(Kirk,1996, Becker, 2000, Hedges, 1981, Cohen, 1992, Thalheimer &Cook, 2002, Nakagawa & Cuthill, 2007, Ferguson, 2009, Christopher, 2006, Tomczak , 2014, Richardson, 1996, , Fritz et al., 2012, Xitao, 1999, Tellez et al., 2015, أبو هاشم،2005 ، بابطين،2001، مراد، 2011، بدوي، 2018).

جدول (2)

مقاييس حجم الأثر لفروق المتوسطات

اسم المقياس	رمزه	اختبار الدلالة الإحصائية المرافق له	قانونه الحسابي	تفسير حجم الأثر	
				المجال	التفسير
معامل d لكوهين	D	الفرق بين متوسطي مجتمعين	$\boldsymbol{d = \frac{\mu_1 - \mu_2}{s}}$	[0.49 – 0.20]	منخفض
				[0.79 – 0.50]	متوسط
				0.80 فما فوق	مرتفع
معامل G لـ هيدجز	G	الفرق بين متوسطي عينتين	$\boldsymbol{g = \frac{\bar{x}_1 - \bar{x}_2}{S_{pooled}}}$ And $\boldsymbol{g = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{MSwithin}}}$	[0.49 – 0.20]	منخفض
				[0.79 – 0.50]	متوسط
				0.80 فما فوق	مرتفع
معامل جلاس دلتا Δ	Δ	الفرق بين متوسطين + اختبارات التحليل البعدي	$\boldsymbol{g = \frac{\bar{x}\,Experment - \bar{x}\,Control}{SD_{control}}}$	[0.49 – 0.20]	منخفض
				[0.79 – 0.50]	متوسط
				0.80 فما فوق	مرتفع
معامل إيتا تربيع **η2**	**η2**	قانون T- Student للفرق بين متوسطي عينتين مترابطتين	$\boldsymbol{\eta 2 = \frac{t^2}{t^2 + N - 1}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
معامل إيتا تربيع **η2**	**η2**	قانون T- Student للفرق بين متوسطي عينتين مستقلتين	$\boldsymbol{\eta 2 = \frac{t^2}{t^2 + df}}$ And $\boldsymbol{\eta 2 = \frac{t^2}{t^2 + (N_1 + N_2 - 2)}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع

(Christopher, 2006, p.404, Kirk,1996, p.6, Becker, 2000, p.p 2 -4, Hedges, 1981, p.p 109-110, Cohen, 1992, p.4, Thalheimer &Cook, 2002, p.p 4-7، مراد، 2011، ص. 247).

حيث:

$\boldsymbol{\mu_1}$: متوسط المجتمع الأول

$\boldsymbol{\mu_2}$: متوسط المجتمع الثاني

$\boldsymbol{s}$: الانحراف المعياري الكلي

$\boldsymbol{\bar{x}_1}$: متوسط المجموعة الأولى "التجريبية"

$\boldsymbol{\bar{x}_2}$:متوسط المجموعة الثانية "الضابطة"

$\boldsymbol{SD_{control}}$: الانحراف المعياري للمجموعة الضابطة

$\boldsymbol{SD_{pooled}}$: الانحراف المعياري المُشترك أو التجميعي

$\boldsymbol{MSwithin}$: متوسط المربعات داخل المعالجات

$\boldsymbol{df}$: درجات الحرية

الإحصاءة (t): قيمة الاختبار الإحصائي T-student لاختبار الفرق بين متوسطين.

ملاحظة: المعادلة الأولى لكوهين هي معادلة حجم الأثر في شكلها الأساسي، كما أشار إليها كوهين (Cohen, 1988)؛ أي إن قيمة (d) تمثل الفرق بين متوسطي مجتمعين مقسوماً على الانحراف المعياري الكلي، إلا أنه من العقبات الأساسية التي تواجه الباحثين والتي أشار إليها كوهين، أنه نادراً ما يكون لديهم القيم الحقيقية لبارامترات المجتمع، والتي يمكن أن يعوضوا بها في تلك المعادلة، لذا يمكن لهم أن يحصلوا على تقديرات المجتمع من خلال إحصائيات عيناتهم، وبذلك تصبح المعادلة السابقة بعد استبدال متوسطي المجتمعين بمتوسطي العينتين، وافتراض تجانس تباين المجموعتين على الصورة الآتية:

$$\boldsymbol{d = \frac{\bar{x}_1 - \bar{x}_2}{s}}$$

حيث $(\boldsymbol{s})$ = الانحراف المعياري لإحدى العينتين في حالة تساوي حجم العينتين وتجانس تباينهما.

وفي حالة عدم تساوي حجم العينتين، وعدم تساوي قيم الانحراف المعياري في فيهما، مع وجود شرط التجانس، فإن قيمة الانحراف المعياري المستخدم في المعادلة السابقة يسمى الانحراف المعياري المُشترك أو المُجمَّع أو المُدمَج s_{pooled} ، ويحسب على النحو الآتي:

$$\boldsymbol{s_{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

حيث:

(s_1) و(s_2) هما الانحراف المعياري للعينتين على الترتيب

(n_1) و(n_2) هما عدد أفراد العينتين على الترتيب

وفي حالة انتهاك افتراض تجانس تباين العينتين، فإنه لا يصلح استخدام الانحراف المعياري المشترك، بل يُستخدم الانحراف المعياري للمجموعة الضابطة $(\boldsymbol{S_{control}})$ في مقام المعادلة، والمنطق النظري وراء ذلك يكمن في أن درجات المجموعة الضابطة لم تتأثر بالمعالجات المتصلة بالمتغير المستقل، فهي تعكس تقديرات المجتمع بصورة نقية، على خلاف درجات المجموعة التجريبية التي تأثرت بهذا المتغير، وفي هذه الحالة يُسمى حجم الأثر بمعامل دلتا لـ Glass. (محمد حسن، 2011، ص ص. 268- 269)

جدول (3)

مقاييس حجم الأثر للتباينات غير المصححة

اسم المقياس	رمزه	اختبار الدلالة الإحصائية المرافق له	قانونه الحسابي	تفسير حجم الأثر	
				القيمة	التفسير
معامل الارتباط المتعدد ويُسمى معامل التحديد	R^2	تحليل الانحدار البسيط والمتعدد	$\frac{SS_{regression}}{SS_{total}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
معامل إيتا تربيع	$\eta 2$	تحليل التباين الأحادي ANOVA تحليل التغاير الأحادي ANCOVA تحليل التباين الأحادي ذو القياسات المتكررة	$\frac{SS_{between}}{SS_{total}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
معامل إيتا تربيع	$\eta 2$	تحليل التباين والتغاير و الانحدار الخطي البسيط	$\frac{F.df}{F.df + df}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
معامل إيتا تربيع متعدد المتغيرات	$\eta 2$	تحليل التباين المتعدد المتغيرات MANOVA وتحليل التغاير المتعدد MANCOVA تحليل الانحدار متعدد المتغيرات وتحليل التباين المتعدد المتغيرات ذو القياسات المتكررة	$1 - \Lambda^{\frac{1}{2}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع

(Nakagawa, & Cuthill, 2007, p. 598, Ferguson, 2009, p.533-534, Christopher, 2006, p.404, Tomczak , 2014, p.p 21-22,
بدوي، 2018، ص ص. 446 – 447).

حيث:

$SS_{regression}$: sum squares regression: مجموع مربعات الانحدار

$SS_{between}$: sum squares between: مجموع المربعات بين المعالجات

SS_{total} : sum squares total: المجموع الكلي للمربعات.

F **:** قيمة النسبة الفائية

Λ : قيمة معامل توزيع ويلكس لامدا Wilkss lambda distribution، ويحسب في أساليب التحليل المتعدد المتغيرات.

ملاحظة:

- يُفسَّر معامل R^2 على أنه النسبة المئوية من التباين المفسر في درجات المتغير التابع بواسطة المتغيرات المستقلة أو المتنبئات، كما هو الحال عند تفسير مربع إيتا η^2.
- ثمة أسلوبان لتفسير معامل ارتباط بيرسون:

الأول: تفسير قيمة معامل الارتباط على أنها حجم تأثير، أي إن قيمة r هي نفسها قيمة حجم التأثير.

والثاني: يتم فيه تربيع قيمة معامل الارتباط للحصول على (r^2) وهو ما يسمى معامل التحديد، والذي يدل على النسبة المئوية من التباين المفسر في درجات المتغير التابع بواسطة المتغير المستقل إذا كان البحث تجريبياً، أو يدل على أنه نسبة التباين المشترك الذي يمكن أن يرجع إلى أحد المتغيرين إذا كان البحث ارتباطياً.

ويُفضَّل الأسلوب الأول على الثاني للطبيعة الاتجاهية الأوضح لمعامل الارتباط إذا لم يتم تربيعه، وهو الأنسب في التحليلات البعدية، لأن معامل الارتباط بعد التربيع لا يُعرف اتجاهه أو إشارته الأصلية، وبالتالي فهو لا يحمل قطبي العلاقة لأنه موجب دائماً (محمد حسن، 2011، ص.278).

وقد بيّن منصور(1997) أن معامل الارتباط يشكل حلقة الوصل بين أسلوبي الدلالة الإحصائية والدلالة العملية، فحجم المعامل بصرف النظر عن حجم العينة المبني عليها هو حجم تأثير وقوة ترابط، أما دلالته إحصائياً فتتوقف على علاقته بمقدار حجم العينة (أبو هاشم،2005، ص.20).

جدول (4)

مقاييس حجم الأثر للتباينات المصححة

اسم المقياس	رمزه	اختبار الدلالة الإحصائية المرافق له	قانونه الحسابي	تفسير حجم الأثر	
				القيمة	التفسير
آر تربيع المضبوطة أو مربع معامل الارتباط المتعدد المصحح	R^2_{adj}	نماذج تحليل الانحدار المتعدد	$1-(1-R^2)\frac{N-1}{N-K-1}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
إيتا تربيع الجزئية Partial Eta Squared	η_p^2	تحليل الانحدار الخطي البسيط، تحليل التباين الأحادي ANOVA تحليل التغاير الأحادي ANCOVA تحليل التباين الأحادي ذو القياسات المتكررة	$\frac{SS_{between}}{SS_{between}+SS_{error}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
معامل هايز أوميجا تربيع Hays omega squared	ω^2	تحليل الانحدار الخطي البسيط، تحليل التباين الأحادي ANOVA تحليل التغاير الأحادي ANCOVA تحليل التباين الأحادي ذو القياسات المتكررة	$\frac{SS_{between}-(v-1)MS_{within}}{SS_{total}+MS_{within}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
مربع إبسيلون Epsilon Squared	ϵ^2	تحليل الانحدار الخطي البسيط، تحليل التباين الأحادي ANOVA تحليل التغاير الأحادي ANCOVA تحليل التباين الأحادي ذو القياسات المتكررة	$\frac{SS_{between}-(v-1)MS_{within}}{SS_{total}}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
معامل هيرزبرج Herzberg	R^2	تحليل الانحدار الخطي البسيط، تحليل التباين الأحادي ANOVA تحليل التغاير الأحادي ANCOVA تحليل التباين الأحادي ذو القياسات المتكررة	$1-(1-R2)\left[\frac{n-1}{n-k-1}\right]\left[\frac{n+k+1}{n}\right]$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع

(Richardson, 1996, p. 18, Ferguson, 2009, p.535, Christopher, 2006, p.405, Fritz et al., 2012, p.p 9- 12, Xitao, 1999, p 16)

حيث:

R^2 : مربع معامل الارتباط المتعدد المُصحح

N : عدد أفراد العينة

K : عدد المتغيرات المنبئة

$SS_{between}$: sum squares between: مجموع المربعات بين المجموعات

SS_{error} : sum squares error: مجموع مربعات الخطأ المعياري للقياس

MS_{within} : متوسط المربعات داخل المجموعات، ويساوي مجموع المربعات داخل المجموعات مقسوماً على درجات الحرية داخل المجموعات

SS_{total} : المجموع الكلي للمربعات

v **:** عدد المجموعات، أو عدد مستويات المعالجات موضع المقارنة

جدول (5)

مقاييس حجم الأثر للاختبارات اللامعلمية

اسم المقياس	رمزه	اختبار الدلالة الإحصائية المرافق له	قانونه الحسابي	تفسير حجم الأثر	
				القيمة	التفسير
معامل فاي	Φ	كاي مربع في حالة الجداول ثنائية التصنيف	$\sqrt{\frac{x^2}{n}}$	[0.29 – 0.10]	منخفض
				[0.49 – 0.30]	متوسط
				0.50 فما فوق	مرتفع
معامل فاي لكرامر Cramer's V كما يُرمز له ϕ_c	ϕ_c	كاي مربع في حالة الجداول أكبر من (2×2)	$\sqrt{\frac{x^2}{n\,(df)}}$	[0.29 – 0.10]	منخفض
				[0.49 – 0.30]	متوسط
				0.50 فما فوق	مرتفع*
Z	Z	Wilcoxon البديل اللامعلمي لاختبار T للعينات المترابطة	$r = \frac{Z}{\sqrt{N}}$	[0.29 – 0.10]	منخفض
				[0.49 – 0.30]	متوسط
				0.50 فما فوق	مرتفع
Z	Z	Mann-whitney البديل اللامعلمي لاختبار T للعينات المستقلة	$\mathbf{r} = \frac{z}{\sqrt{n_1 + n_2}}$	[0.29 – 0.10]	منخفض
				[0.49 – 0.30]	متوسط
				0.50 فما فوق	مرتفع
إيتا تربيع لاختبار كروسكال - والس η_H^2	η_H^2	كروسكال - والس البديل اللامعلمي لاختبار تحليل التباين، واختبار T للعينات المستقلة	$\frac{H - K + 1}{n - k}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع
W	W	معامل فريدمان البديل اللامعلمي لاختبار تحليل التباين، واختبار T للعينات المترابطة	$\frac{x_W^2}{n\,(k-1)}$	[0.05 – 0.01]	منخفض
				[0.13 – 0.06]	متوسط
				0.14 فما فوق	مرتفع

(Tellez et al., 2015, p.p37-38, Fritz et al., 2012, p. 12, Tomczak , 2014, p.p23-24, محمد حسن،2011، ص ص.279-280، أبو هاشم،2005، ص ص.27- 29، بابطين،2001، ص.98)

حيث:

x^2 : قيمة إحصاء كاي مربع

N : عدد أفراد العينة

Z: قيمة الإحصاء U للاختبار الإحصائي مان ويتني البديل اللامعلمي للاختبار T لعينتين مستقلتين، وقيمة الاختبار الإحصائي ويلكوكسون البديل اللامعلمي لاختبار T لعينتين مترابطتين

T1 : مجموع الرتب ذات الإشارة الموجبة

N في اختبار ويلككوسون: عدد أزواج الدرجات

MR_1 : متوسط رتب المجموعة الأولى، وهي التجريبية

MR_2 : متوسط رتب المجموعة الثانية، وهي الضابطة

n_1: عدد أفراد المجموعة الأولى، وهي التجريبية

n_2: عدد أفراد المجموعة الثانية، وهي الضابطة

H : قيمة اختبار كروسكال- والس

K : عدد المجموعات المستقلة موضع المقارنة (عدد مستويات المعالجة)

x_w^2 : قيمة اختبار فريدمان

K: عدد مرات القياس (عدد مستويات المعالجة)

*جدول (6)

محكات حجوم الأثر حسب درجات الحرية

عدد المجموعات (ودرجات الحرية للبعد الأقل في جدول التوافق)	تفسير قيم حجم الأثر		
	منخفض	متوسط	مرتفع
مجموعتان (درجة الحرية df = 1)	0.10	0.30	0.50
ثلاث مجموعات (درجات الحرية df = 2)	0.07	0.21	0.35
أربع مجموعات (درجات الحرية df = 3)	0.06	0.17	0.29

(محمد حسن، 2011، ص.277)

- مفهوم قوة الاختبار الإحصائي:

بيّن كوهين (Cohen, 1977) أن القوة الإحصائية أو كما تُسمى الحساسية الإحصائية، تعني احتمال رفض الفرض الصفري بشكل صحيح – أي عندما يكون خاطئاً–، وبشكلٍ مكافئ قبول فرض بديل صحيح، فهي احتمال الإصابة في رفض الفرض الصفري عندما يكون في حقيقة الأمر خاطئاً (Akbaryan, 2013, p.9). ويتم حسابها من المعادلة الآتية: Statistical Power = 1 – β (علام، 2005، ص ص. 110 –111).

ولحساب قوة الاختبار الإحصائي فوائد مهمة، يمكن تلخيصها على النحو الآتي:

- تزوّد الباحثين بمعلوماتِ عن نتائج البحث بشكل كلي، فعندما تكون قوة الاختبارات الإحصائية منخفضة لدراسة ما فإن النتائج ستكون محل تساؤل وشك في موثوقيتها.
- تفيد في تفسير النتائج الصغيرة وشبه المعدومة.
- ويبين كوهين (Cohen, 1990) أنها تفيد في زيادة الوعي بأهمية تقديرات حجم الأثر، كما أنها تجعل الباحثين أكثر وعياً بالعلاقة بين العوامل الأربعة المُتضمنة والمطلوبة في عملية الاستدلال الإحصائي، وهي: مستوى الدلالة، وحجم العينة، وحجم الأثر، والخطأ من النوع الثاني (Huston, 1993, p.p 11- 13).

- استخدامات حجم الأثر:

تشير أدبيات البحث إلى وجود ثلاثة استخدامات أساسية لحجوم الأثر في الدراسات والأبحاث النفسية والتربوية، وهي:

- استخدام حجم الأثر في تحليل القوة الإحصائية: إذ بيّن كوهين (Cohen, 1988) أن السبب وراء تلك الإحصاءات المعيارية – حجوم الأثر– عدم إمكانية تجهيز وبناء جداول القوة الإحصائية، بالنسبة لكل وحدة قياس جديدة. وبالتالي تُمثل المقاييس المعيارية لحجوم الأثر قاعدة أساسية للجداول والعمليات الحسابية التي استخدمها كوهين في حساب القوة، والتي قدّم بناءً عليها قيم تصنيف مؤشرات حجم الأثر: منخفض، متوسط، ومرتفع.
- استخدام حجم الأثر في الأبحاث التي تعتمد منهج التحليل البعدي.
- استخدام حجم الأثر في إعداد تقارير الأبحاث والإبلاغ عن النتائج (بدوي، 2018، ص ص.444– 445، الشاردي،2011، ص ص.16 – 17).

- الانتقادات الموجهة إلى الدلالة العملية:

على الرغم من أهمية مقاييس حجم الأثر في توضيح نتائج الأبحاث، إلا أنها كأية تقنية إحصائية لا يخلو استخدامها من بعض المشكلات، ويرى البعض أن الاعتماد المفرط عليها دون التنبُّه إلى

حدودها قد يوقع الباحثين في مشكلاتٍ جديدة، تماماً مثلما حدث عند خضوعهم لاختبارات الدلالة الإحصائية، ومن أبرز هذه المشكلات:

- مؤشرات الدلالة العملية لا تغني عن الدلالة الإحصائية، لأنها لا تزود الباحثين بأي إثبات أو برهان عن احتمالية حدوث النتائج بعيداً عن المصادفة، لذا يجب تضمين قيم الدلالتين الإحصائية والدلالة العملية مترافقين معاً في نتائج الأبحاث (Xitao, 1999, p.11).
- وجود جدل بين الباحثين حول مقدار حجم الأثر الذي يجب أن يتحقق حتى تُعد النتائج ذات قيمة وفائدة عملية، ومع أن كوهين (Cohen, 1988) وضع محكات لتقييم معاملات حجم الأثر: منخفض، متوسط، مرتفع، إلا أن الحكم على مقدار الأثر يعتمد بالأساس على تقييم الباحث بناءً على تساؤلات بحثه، ونوعية تصميمه، وسياقه العام (محمد حسن، 2011، ص289). وقد أعرب كوهين عن حذره الشديد تجاه هذه المحكّات، وبيّن أنها تقديرات نسبية، ليس بالنسبة لبعضها البعض فحسب، وإنما بالنسبة لمجال علم السلوك بشكلٍ عام، وللمحتوى المحدد قيد البحث على وجه الخصوص (Akbaryan, 2013, p.7).
- أغلب مؤشرات حجم الأثر شديدة التأثر بتباين واختلاف تجانس العينات، لذا فهي قد لا تقيس المقدار الحقيقي لتأثير المعالجة بنفس وحدة قياس المتغير التابع (بابطين، 2001، ص.99).
- تواجه مؤشرات حجم الأثر بعض الصعوبات في تفسير نتائجها في حالة الأبحاث متعددة المتغيرات (بدوي، 2018، ص.446).
- ليست كل معادلات نسب التباين المفسر تعطي نتائج متماثلة، وفي الحقيقة هناك قضايا عدة يجب أخذها بالاعتبار في تحديد المعادلات التي يجب استخدامها مثل: التقديرات المتحيزة مقابل التقديرات غير المتحيزة، تقديرات نسبة التباين المفسر للدراسات أحادية المتغيرات مقابل متعددة المتغيرات (بابطين، 2001، ص.102).
- تتركز جميع حجوم الأثر على حساب وتفسير التغيُّرات التي تحصل على مستوى المجموعات، مع عدم وجود أية مؤشرات على التغيُّرات التي تحدث على مستوى الأفراد، في الوقت الذي أكد فيه العديد من الباحثين مثل (Jacobson, Follette, Revenstorf, 1984) على أن حساب نتائج المعالجات التجريبية على مستوى المجموعة لا يعطي أية معلومات حقيقية عن كيفية استفادة الأفراد منها، وأن التطورات والتحسينات الفردية لا تقل أهمية عن التطورات على مستوى المجموعة (Peterson, 2008, p.p 5-6).

يتبين مما سبق أن مؤشرات الدلالتين الإحصائية والعملية تخدمان أغراضاً مختلفة، ويدعم كلٌ منها الآخر، وأن الاعتماد الوحيد على أي منهما قد يكون مضللاً، فمؤشرات الدلالة العملية تقيس حجم أثر التغيُّر، بصرف النظر عن الثقة التي نوليها للنتائج، في حين تهتم اختبارات الدلالة الإحصائية بالثقة في

نتائج البحث بصرف النظر عن حجمها، وأن كلا هذين النوعين من الدلالة يكشفان الفروق على مستوى جميع الأفراد في المجموعة بوصفها وحدة متكاملة، ولا يظهران مقدار التغيُّرات التي تحدث على مستوى الأفراد، على الرغم من أهمية حساب هذه التغيُّرات خاصةً في الأبحاث التي تعتمد على تطبيق البرامج التجريبية، وبالتالي لا يكفي حساب هذين النوعين من الدلالة عند الإبلاغ عن نتائج الأبحاث، لأنهما يرصدان التغيُّرات على مستوى المجموعة، ولا يأخذان برصدها على المستوى الفردي.

3- الدلالة العيادية ومؤشر التغيُّر الموثوق:

- **مفهوم الدلالة العيادية:**

تُعتبر حركة جديدة بالنسبة للدلالتين الإحصائية والعملية في علم النفس والعلوم السلوكية، أضافت نوعاً ثالثاً من الدلالة إلى اللغة البحثية يختص بالأبحاث التجريبية. تتركز هذه الدلالة بجوهرها على حساب مقدار التغيُّر الناتج عن المعالجة التجريبية، على مستوى كل فرد من أفراد المجموعة بالطرائق الإحصائية. ذلك لأن الهدف الأساسي من المعالجات التجريبية مهما كان نوعها، هو التحقق من كونها تساعد الأفراد على التغيُّر نحو الأفضل، ولا يُعد أثر هذه المعالجات مهماً إلا إذا أحدث التغيُّر المطلوب، والذي لابد من قياس مدى تحققه لدى كل فرد في المجموعة.

أسس هذا النوع من الدلالة كل من جاكبسون، وفوليت وريفنستورف (Jacobson, Follette & Revenstorf, 1984) الذين أشاروا إلى أن حساب الفروق بين المتوسطات القبلية والبعدية للمجموعة استناداً إلى مقاييس الدلالتين الإحصائية والعملية، لا يعطي معلومات حقيقية عن كيفية استفادة الأفراد من المعالجات التجريبية، فعلى الرغم من أن حجوم الأثر المرتفعة من المرجح أن تكون ذات دلالة وأهمية عيادية أكثر من المنخفضة، غير أنها لا تحدد نسب الأفراد الذين تحسنوا أو تراجعوا نتيجة المعالجة، لذا أكدوا على ضرورة رصد التغيُّر الحاصل لدى كل فرد، بدلاً من التعامل مع المتوسطات الكلية للمجموعات (Jacobson & et al., 1999, p.1).

طوّر بعد ذلك جاكبسون وترواكس اختبارات الدلالة العيادية من خلال إدخالهم مفهوم مؤشر التغيُّر الموثوق (RCI)، والذي يتم حسابه على المستوى الفردي كخطوةٍ أولى لحساب مؤشرات هذا النوع من الدلالة، واقترحا تصنيفات للتغيير الموثوق تحدد معناه واتجاهه، بدلاً من مجرد الإقرار بوجود التغير الدال مقابل غير الدال، كما هو الحال في اختبارات الدلالة الإحصائية (Ronk, 2014, p.20).

ومنذ ذلك الحين أصبح أسلوبهم الإحصائي في حساب التغيُّر الموثوق هو الأساس في مجال حساب هذا النوع من الدلالة، وظهر لها العديد من التعريفات. إذ عرفها بدايةً جاكبسون وترواكس بأنها مؤشر للتغيير المهم الناتج عن التدخل أو المعالجة، والذي يحدد مدى تحرك الفرد الذي يعاني اضطراباً وظيفياً معيناً خارج نطاق المجتمع المضطرب، باتجاه نطاق المجتمع الوظيفي أو الطبيعي (Jacobson& Truax, 1991, p.12).

ثم عرّف هسو التغيُّر الموثوق: بأنه التغيُّر الحقيقي الذي يعكس مقدار التقدم أو التراجع الموثوق لدى الفرد، والذي لا يمكن إرجاعه أو نسبه إلى أخطاء القياس (Hsu, 1998, p.1195).

وعرف كازدين الدلالة العيادية: بأنها القيمة العملية أو التطبيقية، التي تُقرر ما إذا كان التدخل التجريبي يحدث فرقاً حقيقياً وملموساً، وعملياً، في الحياة اليومية للأفراد المستفيدين منه، وحياة الآخرين الذين يتعامل معهم (Kazdin, 1999, p.1).

يتضح من التعريفات السابقة أن الدلالة العيادية تدل على القيمة الفعلية لحجم التدخل التجريبي على المستوى الفردي، من خلال حساب التغيُّر الدال عيادياً، وهو التغيُّر الحقيقي والفعلي، الذي لا يمكن إرجاعه أو نسبه إلى أخطاء القياس.

أتبعها ظهور العديد من التأكيدات في أدبيات البحث على ضرورة تضمين مؤشرات هذا النوع من الدلالة عند الإبلاغ عن النتائج، إذ أكد كامبلز وهولون أنه لا يكفي أن تكون تأثيرات العلاج دالة إحصائياً وعملياً؛ بل يجب أن تكون كبيرة بما يكفي لتكون دالة عيادياً للإبلاغ عنها (Hollon, 1998, p.11 & Chambless).

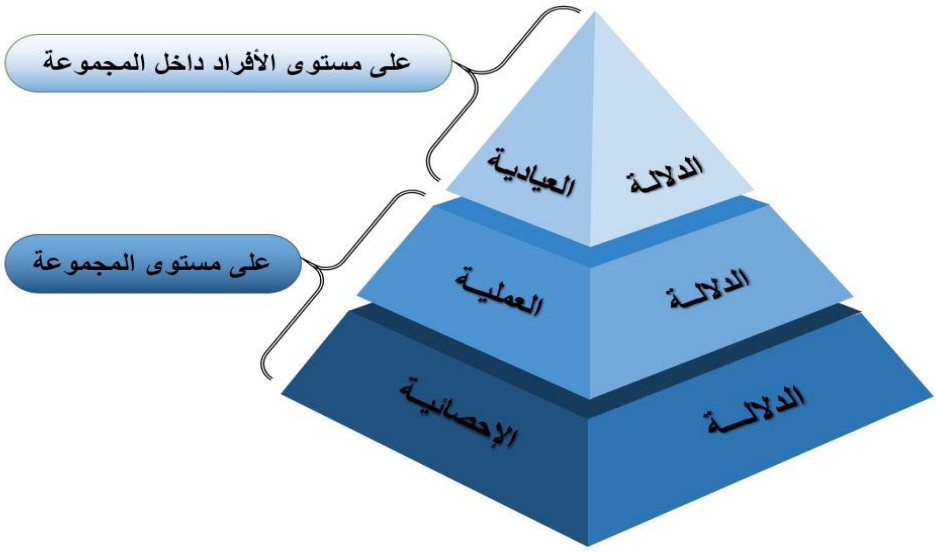
وأشار كيندال أنه لا بد من تضمين مؤشرات الدلالة العيادية في نتائج الأبحاث المنشورة، إلى جانب مؤشرات الدلالتين الإحصائية والعملية (Kendall, 1999, p.283).

وشدد كازدين (Kazdin, 1999) على أهمية حساب هذا النوع من الدلالة في نتائج الأبحاث التجريبية مهما كان نوعها، سواءٌ أكانت علاجية، أم تعليمية، أم تأهيلية (Kazdin, 1999, p.1).

بناءً على ما سبق يمكن القول أن الدلالة العيادية باتت تشكل ركناً أساسياً في حساب نتائج الأبحاث التجريبية، وأنها خطوة إلى الأمام في مجال الدلالات العلمية، تتقدم على الدلالتين الإحصائية والعملية وتكمل مؤشراتهما الإحصائية.

إذ تبنى الأنواع السابقة من الدلالات بعضها على بعض في تسلسلٍ هرمي يوضحه الشكل رقم (1)، بدءاً من الدلالة الإحصائية، ثم العملية المتمثلة بحجوم الأثر، وصولاً إلى العيادية التي تُعد الأساس في صحة القرار بتحقق أو عدم تحقق الفرض الصفري، وبالتالي فإن مؤشرات الدلالة العيادية في رأس هرم الدلالات العلمية، وتتضمن الأنواع السابقة من الدلالات جميعها، ولكن العكس غير صحيح.

شكل (1)
هرم الدلالات العلمية ***" نقلاً عن عز، 2021"***



- **مفهوم مؤشر التغيُّر الموثوق:**

تقوم منهجية الدلالة العيادية على حساب التغيُّر ذي المغزى أو التغيُّر الهادف، والذي يرمز له (RCI) اختصاراً للمصطلح Reliable Change Index ، وهو مؤشر التغيُّر الموثوق.

يشير هذا المصطلح إلى حدوث تغيُّر حقيقي وهادف في عمل كل فرد، ويسمح للباحثين بالتركيز على الأداء الفردي، والتعامل معه على أنه حالة فريدة لا علاقة له ببقية المشاركين، بدلاً من التركيز على الأداء الكلي للمجموعات (Bauer & et al., 2004, p.60).

جرى تطوير هذا المؤشر من قبل جاكبسون وترواكس منذ ما يقارب 30 عاماً (Jacobson & Truax, 1991) بهدف قياس مقدار التغيُّر في درجة كل فرد عقب المعالجة التجريبية مقارنةً بدرجته القبلية، والذي يعتبر تحديده أمراً بالغ الأهمية في جميع عمليات التدخل على اختلاف أنواعها، لأنه يسمح للقائمين على هذه العمليات بتتبع مقدار تقدم أو تراجع الفرد، ومعرفة كيفية استجابته مع عملية التدخل، وبالتالي يمكّنهم من اتخاذ القرار الأنسب بإنهاء التدخل أو الاستمرار به، أو حتى تغييره بما يتناسب مع استجابة الفرد، ولهذا السبب يُعد حسابه الخطوة الأولى والأساسية في الدلالة العيادية، والتي يتم تجاهلها في الدراسات والأبحاث النفسية والتربوية لصالح اختبارات الدلالة الإحصائية (Zahra & Hedge, 2010, p.14).

بيّن كل من جاكبسون وترواكس (Jacobson & Trauax, 1991) أن التغيُّر الناتج عن عملية التدخل يمكن أن يحدث نتيجة عوامل عديدة، بما في ذلك التغيُّر الحقيقي، وأخطاء القياس، وتأثيرات التطبيق، والتحيز، وغير ذلك. ووصفا معنى التغيُّر الموثوق بسؤالٍ بسيط هو: هل تغير الفرد بالقدر الكافي لتكون واثقاً من أن هذا التغير يتجاوز المقدار الذي يمكن نسبه أو إرجاعه إلى أخطاء القياس؟

بناءً على ذلك كانت الصيغة الحسابية لحساب هذا المؤشر تقوم على إيجاد الفرق بين درجتي كل فرد في الاختبارين القبلي والبعدي، ومن ثم قسمته على قيمة الـ Sdiff، وهو الخطأ المعياري للفرق بين درجات هذين الاختبارين، والذي يعتمد في حسابه بشكلٍ مباشر على قيمة الخطأ المعياري للقياس (SE)، وذلك للتأكُّد من أن التغيُّر الحاصل في درجات الفرد حقيقي وموثوق، ولا يعود إلى أخطاء القياس، وذلك على النحو الآتي:

$$\text{Sdiff} = \sqrt{2(S_E)^2}$$

حيث: $S_E = S_1\sqrt{1 - r_{xx}}$

S_1: هو الانحراف المعياري لدرجات الاختبار القبلي

r_{xx}: هو معامل ثبات الاختبار (Jacobson& Truax, 1991, p.14).

ومما تجدر الإشارة إليه وجود العديد من المناقشات حول معامل الثبات الأنسب والأكثر دقة في حساب هذا المؤشر، ففي البداية أوصت معظم أدبيات الدلالة العيادية باستخدام معامل ثبات إعادة الاختبار؛ أي معامل الاستقرار، ولكن فيما بعد توجهت التوصيات إلى استخدام معامل ثبات ألفا كرونباخ، خاصةً بالنسبة للعينات السريرية، والسبب في ذلك هو محاولة عزل التأثيرات التي تحدث في درجات الأفراد البعدية نتيجة عملية التدخل أو المعالجة، إذ يمكن لبعض الأفراد أن يستجيب للمعالجة والبعض الآخر قد لا يستجيب، مما قد يؤثر في قيمة وموثوقية معامل الثبات نتيجةً للاختلافات الفردية في الاستجابة لعملية التدخل (Ronk, 2014, p.p23-24).

وقد قدم مارتينوفيتش ، ساندرز وهاورد (Martinovich, Saunders &Howard, 1996) وصفاً مفصلاً لمزايا ومساوئ كلا البديلين في حساب معامل الثبات، وفي النهاية أوصت دراستهم باستخدام معامل ثبات ألفا بدلاً من معامل الثبات بالإعادة للأسباب السابقة ذاتها، كما توصل بحث تينجي وآخرين (Tingey et al., 1996) إلى أن أفضل طريقة لحساب مؤشرات التغيُّر الموثوق هي معامل ثبات ألفا كرونباخ، ذلك لأن المقاييس التي تُستخدم لقياس التدخل أو المعالجة تحاول قياس الخصائص التي يجب أن تتعدل ويطرأ عليها تغيير بالاتجاه المطلوب، أكثر من قياسها لسمات الشخصية التي تتصف بحكم طبيعتها بأنها مستقرة مع مرور الزمن؛ لذا فإن معامل الثبات بالإعادة أو كما يسمى معامل الاستقرار يُعد الأضعف والأقل موثوقية لحساب قيمة هذا المؤشر (Bauer & et al., 2004, p.61).

وبعد حساب قيمة مؤشر التغيُّر الموثوق، يتم مقارنتها بالدرجة الحدية (1.96) المقابلة لمستوى الدلالة (0.05) في اختبارات الدلالة الإحصائية، فإذا كانت أكبر من الدرجة (1.96) وفي الاتجاه

المطلوب، يُفسّر التغيُّر في درجة الفرد على أنه تحسن ذو دلالة عيادية، وإذا كانت ضمن المجال [+1.96 و– 1.96] يُفسّر ذلك بأن الفرد لم يحقق تحسناً أو تراجعاً ذا دلالة بل بقي أداؤه ثابتاً، أما إذا تجاوزت قيمة مؤشر التغيُّر الموثوق الدرجة (1.96) في الاتجاه السلبي؛ أي المعاكس للمطلوب، عندها يُفسّر التغيُّر على أنه تراجعٌ ذو دلالة عيادية (Jacobson & et al., 1999, P.P 4-5).

وفي هذا السياق لا بد من التأكيد على أن التغيُّر يجب أن يكون باتجاه الهدف والغاية المطلوب تحقيقها من البرنامج التجريبي، فبعض البرامج التجريبية التي يكون نوعها علاجياً أو إرشادياً، وتكون الغاية منها خفض شدة أعراض واضطرابات معينة لدى الفرد كالاكتئاب مثلاً، تُصنّف قيمة مؤشر التغيُّر الموثوق التي تتجاوز الدرجة الحدية (–1.96) على أنها تحسناً دالاً عيادياً، لأنها تدل على أن البرنامج قد نجح في خفض شدة الاضطراب لدى الفرد، أما البرامج التعليمية والتدريبية كالبرنامج الذي جرى إعداده في البحث الحالي، والتي تتمثل الغاية منها في رفع كفاءة الفرد في مجالٍ معين، فإنها تُصنّف قيمة مؤشر التغيُّر الموثوق الأكبر من الدرجة الحدية (+1.96) على أنها تحسُّناً دالاً عيادياً، لأنها تدل على أن البرنامج قد نجح في رفع كفاءة الفرد وتنمية قدراته في المجال المطلوب.

- **طرائق حساب الدلالة العيادية:**

ابتكر جاكبسون وآخرون (Jacobson & et al., 1984) أول طريقة لحساب الدلالة العيادية، ثم جرى تعديلها من قبل جاكبسون وترواكس (Jacobson & Trauax, 1991)، وبعد ذلك تمّ اقتراح العديد من الطرائق الإحصائية لحساب هذا النوع من الدلالة، طورها باحثون مختلفون لاعتقادهم بأنها أكثر دقة وحساسية في اكتشاف مؤشرات التغيُّر الموثوق، بعضها يعتمد على حساب مؤشرات التغير الموثوق مثل طريقة (JT) و(GLN) و(HA)، وبعضها الآخر يعتمد على تحديد فترات ثقة للتحقُّق من التغيُّر الدال عيادياً، مثل طريقة (EN)، وفيما يلي شرحاً مفصلاً لهذه الطرائق:

- **طريقة جاكبسون وترواكس:**

تُعد طريقة Jacobson – Truax Method (JT) هي الأولى في حساب الدلالة العيادية، وتُعرَف بالنهج الكلاسيكي في الأبحاث التجريبية. تنطوي في جوهرها على خطوتين أساسيتين، تتمثل الأولى بإنشاء درجة تسمى بنقطة القطع، وهي درجة على المقياس يُفترض أن تكون قادرة على الفصل والتمييز بين المجتمع المضطرب وظيفياً؛ أي الذي يحتاج أفراده إلى تدخل أو معالجة، والمجتمع الوظيفي أو الطبيعي، الذي يتكون من أفرادٍ أسوياء. وقد اقترح جاكبسون وترواكس ثلاث طرائق ممكنة لحساب درجة القطع، يقوم الباحث باختيار واحدة منها تبعاً لإمكانية توفر البيانات المعيارية المطلوبة لحسابها، وهذه الطرائق هي:

(1) Cutoff A = $M_1 + 2S_1$

(2) Cutoff B = $M_0 - 2S_0$

(3) Cutoff C = $\frac{(M_1+M_0)}{2} = \frac{S_0M_1+S_1M_0}{S_0+S_1}$

حيث:

M_1: متوسط درجات الأفراد في الاختبار القبلي، والتي تمثل درجات المجتمع المضطرب وظيفياً، قبل المعالجة.

M_0: متوسط درجات أفراد المجتمع الطبيعي.

S_1: الانحراف المعياري القبلي.

S_0: الانحراف المعياري لدرجات أفراد المجتمع الطبيعي (Jacobson & Truax, 1991, p.p 13-14)

يتبيّن من الصيغ الرياضية السابقة أن:

- نقطة القطع (A): هي متوسط الدرجات القبلية + 2 مضروباً بالانحراف المعياري للدرجات القبلية؛ وبالتالي فهي تقع خارج نطاق المجتمع المضطرب وظيفياً، وتمتد بنطاق انحرافين معياريين باتجاه المجتمع الطبيعي، وهي الطريقة التي جرى اعتمادها في هذا البحث، وفي بحث (Atkins & et al., 2005).
- نقطة القطع (B): هي متوسط درجات المجتمع الطبيعي– 2 مضروباً بالانحراف المعياري لدرجات نفس المجتمع؛ أي إنها تمتد بنطاق انحرافين معياريين ضمن متوسط درجات المجتمع الطبيعي، ومن الواضح أن هذه الطريقة تفترض توفر معلومات عن توزيع درجات المجتمع الطبيعي في السمة موضع القياس، وفي حال عدم توفرها فإنه لا يمكن استخدام هذه الطريقة لحساب درجة القطع.
- نقطة القطع (C): هي نقطة موزونة تقع بين متوسط المجتمع الطبيعي والمجتمع المضطرب وظيفياً، وهي نقطة التداخل بين توزيعي هذين المجتمعين، وتوصف بالنقطة الأقل تعسفية لحساب درجة القطع. ويمكن حسابها بشرطين: الأول أن تتوفر بيانات عن توزيع المجتمعين الطبيعي والمضطرب وظيفياً، والثاني أن يكون هناك تداخلاً بين هذين التوزيعين.

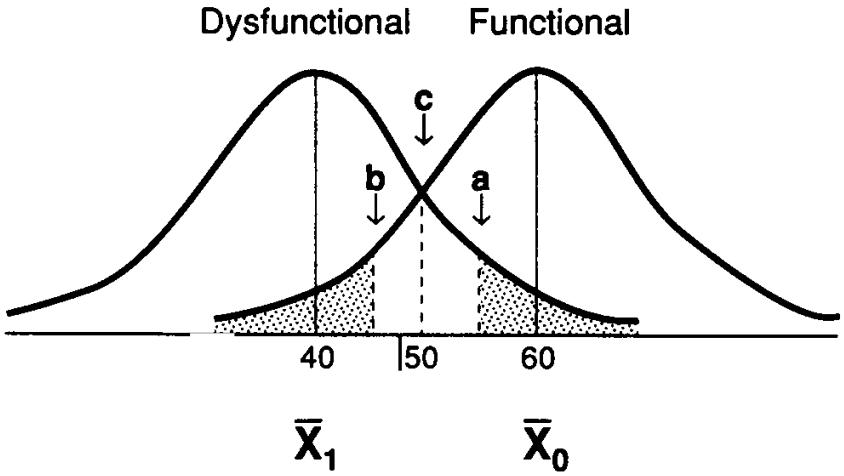
وكما هو موضح بالصيغة الحسابية رقم (3) فإنه يمكن حساب هذه النقطة عن طريق حساب مجموع متوسطي المجتمعين الطبيعي والمضطرب وظيفياً وقسمته على 2، أو على النحو الآتي:

[(الانحراف المعياري للمجتمع الطبيعي مضروباً بمتوسط المجتمع المضطرب وظيفياً) + (الانحراف المعياري للمجتمع المضطرب وظيفياً مضروباً بـمتوسط المجتمع الطبيعي)] / مجموع الانحرافين المعياريين للمجتمعين الطبيعي والمضطرب وظيفياً.

والشكل التوضيحي الآتي يوضح الطرائق الثلاث لتحديد نقطة القطع:

شكل (2)

منحنى بياني لطرائق تحديد نقطة القطع ***(Jacobson & Truax, 1991, P.14)***



يتبيّن من الشكل التوضيحي: على افتراض أن متوسط المجتمع الطبيعي= 60 ، وانحرافه المعياري = 7.5، ومتوسط المجتمع المضطرب وظيفياً = 40، وانحرافه المعياري = 7.5 ، فإن:

نقطة القطع A، وهي عبارة عن انحرافين معياريين خارج متوسط المجتمع المضطرب وباتجاه المجتمع الوظيفي أي: $\boldsymbol{M_1 + 2S_1}$ = 40 + 15 = 55

ونقطة القطع B، وهي عبارة عن انحرافين معياريين ضمن متوسط المجتمع الطبيعي أي:

$\boldsymbol{M_0 - 2S_0}$ = 60 − 15 = 45.

ونقطة القطع C، وهي ببساطة نقطة التداخل بين التوزيعين عند الخط الرأسي الموضح بالشكل السابق، عند النقطة C=50

كما يمكن حسابها على النحو الآتي:

$$C = \frac{(\boldsymbol{M_1 + M_0})}{\boldsymbol{2}} = \frac{\boldsymbol{S_0 M_1 + S_1 M_0}}{\boldsymbol{S_0 + S_1}} = \frac{\boldsymbol{7.5 * 40 + 7.5 * 60}}{\boldsymbol{15}} = 50$$

وبالتالي فإن نقطة القطع C تُعد الوسط، وهي الأفضل في حال توفر معلومات عن توزيع درجات المجتمعين الطبيعي والمضطرب وظيفياً؛ لكن نادراً ما تكون هذه البيانات متوفرة ومتاحة للباحثين، وحتى في حال توفرها فقد بيّن (Jacobson & Truax, 1991) أنه إذا كان توزيع درجات المجتمعين لا يتداخلان فلا معنى لاستخدام درجة القطع (C). وفي المقابل يُعد استخدام الطريقتين (A) و(B) أكثر منطقية في حال عدم توفر البيانات الكافية لحساب نقطة القطع (C). وقد أكد جاكبسون أنه في حال عدم توفر بيانات معيارية فإن نقطة القطع (A) تُعد الخيار الأنسب، خاصةً وأن النقطة (B) تفترض أيضاً توفر معلومات عن توزيع درجات المجتمع الطبيعي (Jacobson & Truax, 1991, P.14).

والخطوة الثانية في حساب قيم الدلالة العيادية وفق طريقة (JT) هي حساب قيمة مؤشر التغيُّر الموثوق، والذي جرى شرحه في الفقرة السابقة، وفي ضوء هذين المعيارين؛ نقطة القطع المحسوبة ومؤشر التغيُّر الموثوق، يجري تصنيف أفراد عينة البحث إلى أحد التصنيفات الأربعة الآتية:

- التصنيف الأول: فئة الأفراد المُعافين إذا كان نوع التدخل علاجياً، والمُتقنين إذا كان نوع التدخل تدريبياً Recovered Group ، وهم الأفراد الذين استفادوا تماماً من عملية التدخل، ولم يعودوا بحاجة لأية معالجة أو تدريب، ويُعرّفون وفق طريقة (JT): بأنهم أولئك الأفراد الذين تجاوزت درجاتهم في الاختبار البعدي نقطة القطع المحسوبة، وحصلوا على قيمة مؤشر تغيير موثوق أكبر من الدرجة الحدية (1.96) وبالاتجاه المطلوب. وهي النتيجة المُثلى لأية عملية تدخل، وتشير إلى القضاء على الأعراض السلبية إذا كان البرنامج علاجياً، والوصول إلى درجة الإتقان إذا كان البرنامج تدريبياً أو تعليمياً.
- التصنيف الثاني: فئة الأفراد الذين تحسنوا Improved Group، وهم الأفراد الذين تحسنت درجاتهم البعدية مقارنةً بدرجاتهم القبلية، ولكنهم مازالوا بحاجة للمعالجة أو التدريب، ويُعرّفون وفق طريقة (JT): بأنهم أولئك الأفراد الذين لم تتجاوز درجاتهم في الاختبار البعدي نقطة القطع المحسوبة، لكنهم حصلوا على قيمة مؤشر تغيير موثوق أكبر من الدرجة الحدية (1.96) وبالاتجاه المطلوب.
- التصنيف الثالث: فئة الأفراد الذين لم يتغيروا Unchanged Group، وهم أولئك الأفراد الذين لم يحققوا أي تغيير دال عيادياً نتيجة تعرضهم للمعالجة التجريبية؛ ومازالوا بحاجة للمعالجة أو التدريب، ويُعرّفون وفق طريقة (JT): بأنهم أولئك الأفراد الذين لم تتجاوز درجاتهم في الاختبار البعدي نقطة القطع المحسوبة، وحصلوا على قيمة مؤشر تغيُّر موثوق ضمن المجال [+1.96 و – 1.96].
- التصنيف الرابع: فئة الأفراد الذين تراجعوا Deteriorated Group، وهم الأفراد الذين تراجعت درجاتهم تراجعاً دالاً عيادياً في الاختبار البعدي، ويُعرّفون وفق طريقة (JT): بأنهم أولئك الأفراد

الذين لم تتجاوز درجاتهم في الاختبار البعدي نقطة القطع المحسوبة، وحصلوا على قيمة مؤشر تغيير موثوق أكبر من الدرجة الحدية (1.96) ولكن في الاتجاه المعاكس للاتجاه المطلوب من البرنامج التجريبي، أي في الاتجاه السلبي (Hsu, 1998, p.1195).

وقد تمّ توضيح الصيغة الرياضية لحساب معاملات الدلالة العيادية وفق هذه الطريقة في الجدول رقم (7)، كما عُرض مثالاً محسوباً يبيّن كيفية استخراج معامل الدلالة العيادية، والتصنيف الناتج عنه وفقاً لها في الملحق رقم (5).

الانتقادات الموجهة إلى هذه الطريقة:

- ❖ الانتقاد الأول والأساسي الذي وجهه هسو (Hsu, 1989) إلى طريقة (JT) أنها تهمل قيمة المتوسط في حساب مؤشر التغيُّر الموثوق، حيث أكد هسو ضرورة تعديل الصيغة الحسابية لهذا المؤشر بحيث تشمل تقديرات متوسط المجتمع إضافةً إلى الانحراف المعياري.
- ❖ كما جرى انتقادها لأنها تعتمد في حساب قيمة مؤشر التغيُّر الموثوق على درجات الأفراد الخام وتعتبرها درجات حقيقية، إلا أن هذه الدرجات وكما هو معروف تشوبها أخطاء القياس، إلا إذا كان متسقاً اتساقاً تاماً، أي خالياً تماماً من الأخطاء، وهذا ما يصعب تحقيقه، وبالتالي إذا ما تمّ إعادة التطبيق وحساب مؤشر التغيُّر الموثوق فلن يتم الحصول على نفس النتيجة، إضافةً إلى أن الدرجات الخام لا تسمح بإجراء المقارنات المعيارية عبر الدراسات المختلفة (Atkins & et al, 2005, P.P 982-983).

- **طريقة جليكسين لورد نوفيك:**

طوّر طريقة Gulliksen- Lord- Novick Method (GLN): هسو (Hsu, 1989) لتلافي القصور القائم في طريقة (JT) من وجهة نظره، معتمداً في عمله الإحصائي على الأساليب الرياضية لجليكسين (Gulliksen, 1950) واللورد ونوفيك (Lord & Novick, 1969) لذلك سميت هذه الطريقة (GLN).

تأخذ هذه الطريقة بالاعتبار الانحدار نحو المتوسط في بسط الصيغة الحسابية لمؤشر التغيُّر الموثوق، ويتركز حسابها على طرح درجات الفرد في الاختبارين القبلي والبعدي من المتوسط الافتراضي للمجتمع، والتقسيم على قيمة الخطأ المعياري للقياس والذي يتضمن قيمة الانحراف المعياري، ولأن المتوسطات والانحرافات المعيارية للمجتمع الافتراضي نادراً ما تكون معروفة من قبل الباحثين، لذا دعا هسو(Hsu, 1998) إلى استخدام الدرجات القبلية لأفراد عينة البحث في حساب هذه التقديرات، لأنها تعكس تقديرات المجتمع بصورة نقية قبل أن تتأثر بالمعالجات التجريبية (Hsu, 1998, p.1196, Bauer & et al., 2004, p.64). وبذلك كان هسو أول من تعامل رياضياً مع القيود المفروضة على حساب مؤشر التغيُّر الموثوق وفق الطريقة (JT).

وقد تمّ توضيح الصيغة الرياضية لحساب معاملات الدلالة العيادية وفق هذه الطريقة في الجدول رقم (7)، كما عُرض مثالاً محسوباً يبيّن كيفية استخراج معامل الدلالة العيادية، والتصنيف الناتج عنه وفقاً لها في الملحق رقم (5).

- **طريقة إدواردز ونانللي:**

أسس طريقة Edwards- Nunnally Method (EN): سبير (Speer, 1992) لحساب مؤشر التغيُّر الموثوق، الذي كان ينتقد طريقة (JT) للأسباب نفسها التي طرحها هسو، مُعتمداً في عمله على الأفكار التي وضعها إدواردز (Edwards, 1978) ونانللي (Nunnally, 1965)، ولهذا سُميت هذه الطريقة (EN).

اقترح سبير طريقة (EN) وسيلة لمنع التضخم المصطنع في معدلات التحسن، وتعتمد في حسابها على فترة ثقة تُحسب وفق الصيغة الرياضية الآتية: فترة الثقة $= \pm 2SD_{pre}\sqrt{1 - r_{xx}}$

تُضاف القيمة الناتجة عن فترة الثقة إلى قيمة الدلالة العيادية المحسوبة وفق قانون (EN)، والموضح بالجدول رقم (7) فتكوّن الحد الأعلى من فترة الثقة، وتُطرح منها للحصول على الحد الأدنى.

ثم تُقارن درجة الفرد البعدية بحدي فترة الثقة المحسوبة، بدلاً من مقارنتها بدرجة واحدة كالدرجة الحدية (1.96)، فإذا كانت درجته البعدية ضمن حدي فترة الثقة المحسوبة يُصنّف الفرد على أنه غير مُتغيّر؛ أي لم يحقق تحسناً أو تراجعاً ذا دلالة عيادية، وإذا تجاوزت درجته البعدية الحد الأعلى من فترة الثقة المحسوبة بالاتجاه المطلوب، يُصنّف الفرد على تحسن تحسناً ثابتاً ودالاً عيادياً، أما إذا كانت درجته البعدية أدنى من الحد الأدنى لفترة الثقة؛ وبالاتجاه المعاكس للتغيير المطلوب، فيُصنّف الفرد على أنه تراجع تراجعاً دالاً عيادياً (Peterson, 2008, P.P 10-11, Atkins & et al, 2005, P.983).

وقد تمّ توضيح الصيغة الرياضية لحساب معاملات الدلالة العيادية وفق هذه الطريقة في الجدول رقم (7)، كما عُرض مثالاً محسوباً يبيّن كيفية استخراج معامل الدلالة العيادية، والتصنيف الناتج عنه وفقاً لها في الملحق رقم (5).

- **طريقة هاجمان وآرينديل**:

طوّر طريقة Hageman-Arrindell Method (HA): كلٌ من هاجمان وآريندال (Hageman-Arrindell, 1999) تعديلاً لطريقة حساب مؤشر التغيُّر الموثوق التقليدي، تستند أساساً إلى قانون كرونباخ وجلاسر عام 1959 في حساب الثبات، وعلى غرار طريقتي هسو(GLN) وسبير (EN) اعتمدت طريقة (HA) على الدرجات الحقيقية التي تأخذ بالحسبان الانحدار إلى المتوسط، بدلاً من اعتمادها الدرجات الخام في حساب مؤشر التغيُّر الموثوق.

وقد أدخلت طريقة (HA) مؤشرات جديدة لتقييم الدلالة العيادية تمثلت في حساب معامل موثوقية فرق الدرجات، ويرمز له ا- r_{dd} ، والذي يتطلب حسابه قدراً كبيراً من العمليات الحسابية الموضحة في الجدول رقم (7)، وتقارن الدرجات الناتجة عن مؤشر التغيُّر الموثوق وفقاً لهذه الطريقة بالدرجة الحدية (1.65) بدلاً من الدرجة (1.96) المُستخدمة في الطرائق السابقة، وبذلك يمكن القول إن هذه الطريقة أدخلت أهم التنقيحات على طريقة (JT)، كما أنها تُعد الطريقة الأكثر صرامةً في حساب هذا النوع من الدلالة، وتتطلب قدراً أكبر من التغيُّر في درجات الفرد البعدية حتى يتم تصنيفه على أنه حقق تحسناً ذا دلالة. (Shada, 2013, P.12, Ronk, 2014, p.p57-58).

وقد تمّ توضيح الصيغة الرياضية لحساب معاملات الدلالة العيادية وفق هذه الطريقة في الجدول رقم (7)، كما عُرض مثالاً محسوباً يبيّن كيفية استخراج معامل الدلالة العيادية، والتصنيف الناتج عنه وفقاً لها في الملحق رقم (5).

- **الطريقة الخطية الهرمية:**

طوّر طريقة Hierarchical Liner Method (HLM): كل من سبير وجرينباوم (Speer & Greenbaum, 1995)، تنطوي في جوهرها على تحليل نماذج منحنى تزايد الدرجات (Growth curve models)، أحد أهم مزاياها أنها تتطلب أقل ما يمكن ثلاث نقاط من درجات الفرد لحساب مؤشر التغيُّر الموثوق، بدلاً من الاكتفاء بدرجتيه في الاختبارين القبلي والبعدي كما هو الحال في طرائق حساب الدلالة العيادية السابقة.

تستخدم طريقة (HLM) تقديرات بايز(Bayes) في تحديد مؤشر التغيُّر الموثوق، الذي يُحسب بواسطة برامج حاسوبية خاصة مثل البرامج الهرمية الخطية، وبرامج النمذجة اللاخطية، ويعتقد أنصار هذه الطريقة أنها أكثر دقةً ومرونة في حساب مؤشرات التغيُّر الموثوق من النماذج الكلاسيكية السابقة (Peterson, 2008, P 12, Nezu & Maguth, 2008, P.P 362-375).

والجدول الآتي يوضح الصيغ الرياضية المُستخدمة في حساب قيم الدلالة العيادية وفق كل طريقة من الطرائق السابقة.

جدول (7)

قوانين حساب الدلالة العيادية

اسم الطريقة	مؤسسها	قانونها الحسابي	وصفٌ عام لها
Jacobson – Truax Method (JT)	**Jacobson, Follett,& Revenstorf, 1984 وتم مراجعتها من قبل Jacobson و Truax عام 1991**	$\text{RCI}=\frac{X_{post}-X_{pre}}{S_{diff}}$	**تقوم على حساب نقطة القطع ومؤشر التغيُّر الموثوق، تعتمد بشكل أساسي على الخطأ المعياري للقياس، ولا تأخذ بالاعتبار الانحدار إلى المتوسط**
Gulliksen-Lord-Novick Method (GLN)	**Hsu (1989)**	$\frac{[X_{post}-M_{pop}]-r_{xx}[X_{pre}-M_{pop}]}{SD_{pop}\sqrt{1-r_{xx}^2}}$	**عدلت طريقة (JT) من خلال الأخذ بالاعتبار الانحدار نحو المتوسط في حساب قيم مؤشر التغيُّر الموثوق**
Edwards-Nunnally Method (EN)	**Speer (1992)**	$[r_{xx}(X_{pre}-M_{pre})+M_{pre}]\pm 2SD_{pre}\sqrt{1-r_{xx}}$	**تعتمد في حسابها على فترتي ثقة، بهدف وضع معايير ثابتة لتحديد التحسن أو التراجع الدال عيادياً**
Hageman-Arrindell Method (HA)	**Hageman&Arrindell (1999)**	$\frac{(X_{post}-X_{pre})r_{dd}+(M_{post}-M_{pre})(1-r_{dd})}{(\sqrt{r_{dd}})\left(\sqrt{2S_E^2}\right)}$	**أدخلت أهم التنقيحات على طريقة (JT)، وتُعد الطريقة الأكثر صرامةً في حساب هذا النوع من الدلالة**
Hierarchical Liner Method (HLM)	**Speer & Greenbaum (1995)**	$\frac{B^*}{\sqrt{V^*}}$	**تعتمد في حسابها على أكثر من نقطتين من درجات كل فرد لحساب مؤشر التغيُّر الموثوق، وتستخدم تقديرات بايز(Bayes) في حسابه**

حيث:

$\boldsymbol{x_{post}}$ = درجة الفرد الخام في الاختبار البعدي.

$\boldsymbol{x_{pre}}$ = درجة الفرد الخام في الاختبار القبلي.

$\boldsymbol{S_{diff}}$ هو الخطأ المعياري للفرق بين درجات الاختبارين القبلي والبعدي، ويحسب من الصيغة الرياضية الآتية: $\sqrt{2S_E^2}$

حيث: $S_E = S\sqrt{1 - r_{xx}}$

$\boldsymbol{M_{pop}}$ = المتوسط الافتراضي للمجتمع، الذي إذا لم يتوفر يمكن استبداله بالمتوسط القبلي.

$\boldsymbol{S_{pop}}$ =الانحراف المعياري للمجتمع، وإذا لم يتوفر يمكن استبداله بالانحراف المعياري القبلي.

$\boldsymbol{r_{xx}}$ = قيمة معامل ثبات المقياس.

M_{pre} = متوسط العينة في الاختبار القبلي.

SD_{pre} = الانحراف المعياري للعينة في الاختبار القبلي.

r_{dd} يحسب وفق الصيغة الآتية:

$$r_{dd} = \frac{SD_{pre}^2 r_{pre} + SD_{post}^2 r_{post} - 2SD_{pre}SD_{post}r_{pre*post}}{SD_{pre}^2 + SD_{post}^2 - 2SD_{pre}SD_{post}r_{pre*post}}$$

$$r_{pre} = \frac{SD_{pre}^2 - S_E^2}{SD_{pre}^2}$$

$$r_{post} = \frac{SD_{post}^2 - S_E^2}{SD_{post}^2}$$

B^* = درجات بايز التجريبية المحسوبة وفق الانحدار الخطي.

V^* = الانحراف المعياري لدرجات بايز التجريبية المحسوبة.

(Jacobson & Truax,1991, P.P 13-14, Peterson, 2008, P.P 10-12, Kazdin,1999, P.P3-7, , Atkins & et al, 2005, P.P 982-983, Nezu & Maguth, 2008, P.P 362-375)

أخيراً لا بد من الإشارة إلى أن الطرائق السابقة في حساب الدلالة العيادية يصعب تقييمها أو تفضيل بعضها على الآخر، لأنها جميعاً مبنية وفق أسس وبراهين رياضية صحيحة، وكلٌّ منها قائم على افتراضات ومعطيات عمل مختلفة، وما يصلح حسابه لحالاتٍ معينة قد لا يصلح لحالاتٍ أخرى، وعلى الباحث أن يختار الطريقة التي تتناسب مع طبيعة بحثه والهدف منه، ونوع عيناته وتوصيف خصائصها بشكلٍ دقيق، وهذا ما أظهرته أدبيات البحث ونتائج الدراسات السابقة التي جرى الرجوع إليها، فقد بينت بالمجمل أنه لا يزال هناك اختلافٌ عام حول الطريقة الأمثل لحساب الدلالة العيادية، ومع أن الطرائق الأحدث لحساب هذا النوع من الدلالة تحاول أن تكون أكثر دقة وصرامة من خلال التحكم بالخطأ المعياري للقياس، ومعالجة مشكلة الانحدار إلى المتوسط، إلا أن نتائج الأبحاث بينت بالمجمل أن القيم والتصنيفات الناتجة عن الطرائق الأحدث تقارب كثيراً تلك الناتجة عن طريقة جاكبسون وترواكس، وهي الأبسط والأكثر أصالة في حساب هذا النوع من الدلالة.

ومن الجدير بالذكر أن نتائج البحث الحالي أظهرت عدم ملائمة طرائق الدلالة العيادية جميعها للأنواع المختلفة من البرامج التجريبية، حيث جرى التمييز بين نوعين أساسيين من التدخل التجريبي، الأول يشمل البرامج التي تهدف إلى رفع درجات الأفراد وتنمية كفاءاتهم في مجالٍ معين، كالبرامج التعليمية والتدريبية والتأهيلية، في حين يشمل النوع الثاني البرامج التي تهدف إلى خفض درجات الأفراد عقب المعالجة التجريبية، من خلال خفض شدة اضطرابات وأعراض معينة لديهم كالبرامج العلاجية وبعض البرامج الإرشادية. حيث أظهرت النتائج أن النوع الأول من البرامج التجريبية يصلح معها استخدام جميع طرائق الدلالة العيادية باستثناء طريقة هاجمان آريندال (HA)، التي تصلح للنوع الثاني من برامج التدخل على وجه الخصوص، إذ من المحتمل أن يعود السبب في ذلك إلى أن التعقيدات الإحصائية التي جرى إدخالها على هذه الطريقة قد تفترض بأن حالات التدخل التي يجري حساب مؤشرات الدلالة العيادية لها جميعها مرضية سريرية، وحسب هذا الافتراض فإن درجة الشدة المرضية للفرد تكون مرتفعة في الاختبار القبلي ومن المفترض أن تنخفض في الاختبار البعدي؛ أي بعد التعرض للمعالجة التجريبية، مما جعل هذه الطريقة تتناسب مع البرامج العلاجية ولا تصلح للبرامج التي تهدف إلى رفع درجات الفرد في الاختبار البعدي. ولدى قيام الباحثة بحساب قيم الدلالة العيادية وفق طريقة (HA) لدرجات أفراد العينة التجريبية في البحث الحالي، والقائم على برنامج تدريبي، تَبيّن أنها أنتجت معدلات تحسن دال عيادياً لجميع الأفراد دون استثناء خلافاً لبقية الطرائق، وهذا التحسن في الواقع غير حقيقي. هذا ما أكدته نتائج (Shada, 2013) التي أظهرت وجود اختلافات وتناقضات في تصنيفات الدلالة العيادية الناتجة عن طريقة (HA) مقارنةً بالطرائق الأخرى من جهة، ومقارنةً بنتائج الأبحاث السابقة من جهةٍ أخرى، من حيث إظهارها نسباً عالية جداً من تصنيفات التحسن والتعافي الدال عيادياً، وقد فسر الباحث ذلك بأن التعقيدات الإحصائية في مقام المعادلة جعلت التصنيفات الناتجة عن هذه الطريقة غير منطقية (Shada, 2013, P.P 31 -32)

- **مجالات توظيف الدلالة العيادية:**

ظهرت أولى محاولات حساب هذا النوع من الدلالة في ميدان علم النفس والعلاج النفسي، حيث جرى حساب معاملاتها في البداية على حالات مرضية سريرية، لتقييم فاعلية العلاج المُقدّم لكل فرد، والتأكد من انتقاله من نطاق الاضطراب الوظيفي إلى النطاق والتصنيف الطبيعي، كما هو الحال في الأبحاث المبكرة لجاكبسون وترواكس (Jacobson & Truax, 1991)، غير أن الدراسات والأبحاث التي أجريت فيما بعد على هذا النوع من الدلالة أكّدت أن أهميتها لا تقتصر فقط على البرامج العلاجية، فقد وضح كازدين (Kazdin, 1999) أن الدلالة العيادية لا تنحصر في تقييم الحالات السريرية فحسب، بل تمتد أهميتها لتشمل النواحي التعليمية والتأهيلية، وأكّد بحث بترسون (Peterson, 2008) على أن الدلالة العيادية تضيف معلومات مهمة لا يمكن الاستغناء عنها في مجال العلوم التربوية والاجتماعية، فهي تساعد على تحديد نوع واتجاه التغيُّر الذي يطرأ على أداء كل فرد عقب التدخل التجريبي، مهما كان نوعه سواء أكان تعليمياً أم تدريبياً أم تأهيلياً.

كما تشمل أهمية حساب هذا النوع من الدلالة ميدان التربية الخاصة، إذ تحدد تصنيفات الدلالة العيادية ما إذا كانت الإستراتيجيات أو البرامج العلاجية تفيد في تحسن الأفراد ذوي الاحتياجات الخاصة بعد تدخل معين، وبالتالي تساعد القائمين على هذه البرامج باتخاذ القرار الأنسب بالاستمرار بالمعالجة أو إيقافها أو تغييرها بما يتناسب مع تقييم حالة كل فرد. وتفيد أيضاً في تقييم مشكلات النطق واللغة مثل حالات التأتأة، والحالات التي تكون فيها المشكلات الفردية متباينة، أي تختلف في شدتها ونوعيتها من فرد لآخر.

وقد سلط بالكين ولينز (Balkin & Lenz, 2021) الضوء على أهمية حسابها في مجال الإرشاد النفسي، وأكدا أهميتها ودورها في تفسير نتائج البرامج الإرشادية، وأشار نصار (نصار، 2016) إلى أهمية حساب مؤشراتها في مجال صعوبات التعلم، وبالتالي يمكن القول إن حساب الدلالة العيادية أمرٌ بالغ الأهمية في جميع الأبحاث التجريبية، التي تعتمد الدرجات القبلية والبعدية معياراً للحكم على فاعلية البرامج المُعدّة فيها على اختلاف أهدافها وأنواعها.

- **الانتقادات الموجهة إلى الدلالة العيادية:**

على الرغم من إجماع أدبيات البحث على أهمية النقلة النوعية التي أضافتها الدلالة العيادية إلى مؤشرات الدلالتين السابقتين في الأبحاث التجريبية، فقد وجهت إليها مجموعة من الانتقادات، يمكن إجمالها بالنقاط الآتية:

- تعدُّد طرائق حساب الدلالة العيادية، وعدم وجود إجماع على طريقة محددة لها إلى الآن، أدّى إلى اختلاف نتائج الأبحاث باختلاف هذه الطرائق، كما أثّر أيضاً في ثبات النتائج والتصنيفات العيادية الناتجة عنها (Ronk, 2014, p.13).
- الترجمة الحرفية للتصنيفات الناتجة عن الدلالة العيادية، والقرارات التي تُبنى على أساسها، دون الأخذ بالاعتبار السياق العام للحالات موضع البحث. فقد لا يكون تصنيف التعافي أو العودة للأداء الطبيعي دقيقاً ومناسباً لجميع الاضطرابات مثل الفصام والتوحد، وفي هذا السياق يجب تأكيد ضرورة عدم الاكتفاء بمقياس وحيد للسمة موضع القياس، والاستعانة بأكثر من مقياس للتأكد من موثوقية التغيُّر الحاصل في أداء الفرد، كما لابد من تعاضد وتلازم الاعتماد على أكثر من وسيلة لوصف التغيُّر كالمقابلات التشخيصية، ودراسة الحالة في البرامج العلاجية والإرشادية، والمعاينة السريرية في الحالات الطبية لدعم نتائج الطرائق الإحصائية (Wise, 2004, p.51).
- الافتراض الضمني بأن تصنيفات الدلالة العيادية مستقرة وليست ديناميكية؛ فعلى سبيل المثال يتم تصور التعافي نتيجة يجب الوصول إليها وأنها حالة مستقرة، ولكن في حقيقة الأمر هي عملية يتعين تجربتها واختبارها، وهي ليست ثابتة بل يتوجب تتبع استقرارها وثباتها مع مرور الزمن.
- لم يتم بعد إثبات أن منهجية الدلالة العيادية مرنة بما فيه الكفاية لمراعاة التغيُّر الموثوق الذي تعانيه مختلف الحالات المرضية، خاصةً الحالات شديدة الخطورة (Ronk, 2014, p.14).

ولكن على الرغم من الانتقادات السابقة التي تعرضت لها الدلالة العيادية، إلا أنه ثمة إجماع واضح في أدبيات البحث على أهمية حساب والإبلاغ عن هذا النوع من الدلالة في التقارير البحثية، وهي كما وصفها كيندال (Kendall, 1999) خطوة بالغة الأهمية في تطوير نتائج البحث العلمي، ومبنية على أساس منهجي قوي ساعد على انتشارها في الآونة الأخيرة أكثر مما سبق، وجعلها قادرة على الصمود في وجه النقاش الموجه ضدها (Wise, 2004, p.57).

وفي نهاية المطاف لا بد من تأكيد أهمية اعتماد الأنواع الثلاثة من الدلالات لأنها ليست بديلة عن بعضها، ولا يلغي أيٌ منها الآخر، فلكلٌ منها دوره وأهميته في الإبلاغ عن نتائج الأبحاث، ولابد من استخدامها على نحو يكمل بعضها بعضاً للوصول إلى نتائج وقرارات بحثية على قدرٍ كافٍ من الدقة والموثوقية.

ثانياً: التدريب وأثره في الخصائص القياسية:

1- مفهوم الخصائص القياسية:

تعتمد جودة أدوات القياس المستخدمة في الأبحاث الكمية على مدى تمتعها بمؤشرات مناسبة من الخصائص القياسية، وهي تلك الخصائص المتعلقة بصدق الأدوات وثباتها، والتي يتم حسابها بعد تجريب الاختبار على عينة ممثلة للمجتمع الإحصائي (مراد وسليمان، 2002، ص.350).

فالصدق يشير إلى كون الاختبار يقيس فعلاً ما أُعدَّ لقياسه، ويُقصد به أن تكون هناك دلائل مقبولة حول ما يقيسه المقياس وكيفية قياسه، وعلى الرغم من أن كل اختبار يحمل مُسمى يشير إلى ما يقيسه إلا أن هذا المُسمى غالباً ما يكون عاماً وغير محدد، كما أنه لا يُعد – في حد ذاته- دليلاً على ما يقيسه الاختبار، حيث يفيد كأداة تعريفية فقط، لذلك من الضروري التحقُّق من صدق الاختبار بأساليب منهجية محددة، قبل الشروع في الاعتماد على النتائج المستخلصة منه (فرج، 2007، ص ص. 238 – 239)

ولصدق الاختبار أنواعٌ عديدة منها: صدق المحتوى أو الصدق التمثيلي الذي يرتكز على الفحص المنظم لمحتوى الاختبار، لتحديد ما إذا كان عينة ممثلة لمحتوى الموضوع ومجال السلوك الذي يُراد قياسه، والصدق الظاهري والذي يُستخدم للتأكُّد من أن المقياس يبدو بنظر المفحوصين قادراً على قياس السمة أو الخاصية موضع القياس، وعلى الرغم أن هذا النوع من الصدق ليس علمياً بالمعنى الدقيق للكلمة، إلا أنه عنصرٌ أساسي لإثارة الدافعية لدى المفحوصين، فإذا كان المفحوصون لا يعتقدون أن هذا المقياس يقيس بدقة القدرة، أو أنه غير مرتبط بالسمة موضع القياس، فمن الممكن أن لا يعيروه الاهتمام اللازم، وبالتالي يُفشلوا عملية القياس. ومن أنواعه أيضاً الصدق البنيوي أو الافتراضي: وهو مفهومٌ شاملٌ يتضمن معظم أنواع الصدق، ويُعد شرطاً مهماً للاختبارات النفسية والتربوية بأنواعها المختلفة، إذ يتطلب هذا النوع من الصدق اللجوء إلى الاستنتاجات المنطقية بالإضافة إلى الوسائل التجريبية والإحصائية، ومن الأساليب المستخدمة في تأسيسه الصدق العاملي، والصدق التجريبي، والذي يتم التحقق منه بطرائق متعددة منها الصدق المحكي، وصدق المقارنات الطرفية أو الفرق المتقابلة (مخائيل،2008، ص ص162- 163، بودلال ومورواس، 2021، ص ص.116-118).

أما مفهوم ثبات درجات الاختبار فيقصد به مدى خلوّها من الأخطاء غير المنتظمة التي تشوب القياس، أي مدى قياس الاختبار للمقدار الحقيقي للسمة التي يهدف لقياسها، فدرجات الاختبار تكون ثابتة إذا كان الاختبار يقيس سمة معينة قياساً متسقاً في الظروف المتباينة التي قد تؤدي إلى أخطاء القياس، فالثبات بهذا المعنى يعني الاتساق أو الدقة في القياس (علام، 2002، ص131).

ومفهوم الثبات بالمعنى العام يُستخدم للدلالة على مدى اعتماد الفروق في درجات الاختبار على الفروق الحقيقية في السمة أو الخاصية المقيسة من جهة، ومدى خلوها من أخطاء القياس العشوائية أو الناتجة عن الصدفة من جهة أخرى (مخائيل، 2008، ص ص181 – 182).

وقد تمّ إدراج مفهومين أساسيين لثبات درجات الاختبار هما:

- الحصول على النتائج نفسها تقريباً عند إعادة التطبيق.
- أن يكون التباين الحقيقي أكبر ما يمكن بالنسبة للتباين العام، وتباين الخطأ أقل ما يمكن بالنسبة للتباين العام (عبد الرحمن، 1998، ص ص 165- 166).

توجد أساليب عديدة لحساب الثبات أهمها: الثبات بالإعادة أو ثبات الاستقرار؛ يقوم هذا المعامل على تطبيق المقياس نفسه على الأفراد ذاتهم، ولكن في نقطتين زمنيتين مختلفتين وبظروف متشابهة تماماً، ثم حساب الارتباط بين درجات التطبيقين، وثبات التجزئة النصفية؛ يقوم هذا المعامل على إمكانية تقسيم المقياس إلى نصفين بحيث تكون بنود كل نصف مجمّعة لتشكل مقياساً جديداً، ثم تطبيق الاختبارين بالوقت نفسه، وحساب الارتباط بين درجات النصفين، وتتم التجزئة باستخدام أساليب عديدة منها:

- الأسلوب الفردي والزوجي.
- أسلوب جزأي الاختبار.
- تحديد معامل السهولة لكل سؤال ثم توزيع الأسئلة على نصفي الاختبار بحسب هذه المعاملات، وهو ما تمّ اعتماده في البحث الحالي.

ومن أساليب حساب الثبات أيضاً معامل ألفا كرونباخ، الذي يجري حسابه وفق المعادلة الآتية:

$$a = \frac{K}{K-1}\left[1 - \frac{\sum S_i^2}{S_t^2}\right]$$

حيث K: عدد أسئلة الاختبار.

و $\sum S_i^2$: مجموع تباين كل سؤال من أسئلة الاختبار، أي

$$S_1^2 + S_2^2 + \cdots + S_i^2$$

و s_t^2: تباين درجة الاختبار الكلية.

وأيضاً معامل ثبات الصور المتكافئة؛ و يستلزم هذا المعامل إعداد صورتين متكافئتين على الأقل من الاختبار وتطبيقهما على الأفراد ذاتهم، بحيث يتساوى عدد البنود وطريقة الإجابة والتصحيح، والزمن، إضافةً إلى تساوي معاملات السهولة والصعوبة والتمييز والتباين والمتوسط الحسابي والانحراف المعياري، ثم يقوم الباحث بتصحيح الصورتين وحساب معامل الارتباط بين نتائج المجموعة ذاتها على كلٍ منهما (جميلة ومحمد، 2017، ص ص.214– 215، النبهان، 2004، ص ص. 248- 249).

2- أثر التدريب في الخصائص القياسية:

يشير مصطلح التدريب على الاختبار إلى إكساب المفحوصين خبرة التعرض للموقف الاختباري، من خلال تعريفهم بنوع الاختبار موضع التطبيق، والهدف منه، وشكله العام، وتعليماته الأساسية سواءٌ العامة أم الخاصة بكل نوع من بنوده الاختبارية إذا كان يتناول أنواعاً مختلفةً منها، كما يتضمن أيضاً تدريبهم على حل بنود اختبارية تمهيدية لها نفس شكله العام، ولكنها لا تعرض معلومات عن محتواه الحقيقي وبنوده الفعلية، الغاية منها فقط عملية تهيئة المفحوصين وتدريبهم على طريقة إجراء الاختبار، بحيث لا يُعطى لها درجات، كما لا يُحسب زمن الإجابة عنها من الزمن المخصص للاختبار إذا كان موقوتاً.

تطوّر هذا المفهوم في الأساس على يد ثورنديك (Thorndike, 1951)، وهو أول من أعدّ تصنيفاً كاملاً لمصادر تباين الدرجات الاختبارية في كتابه عن الثبات، إذ اعتبر فيه التدريب على التصرف الصحيح في الموقف الاختباري أحد مصادر التباين الأساسية التي يمكن أن تؤثر في ثبات الدرجات (Sarnacki, 1979, P 252).

وأضاف ستانلي فيما بعد (Stanley, 1951) بأن هناك عاملين أساسيين مسببين للتباين ينطويان تحت هذا المفهوم هما: القدرة العامة على فهم التعليمات، والخبرة بأداء الاختبار، وهما عاملان يكادان يؤثران في كل درجة اختبارية، إذ يلعب العامل الأول دوراً مهماً على وجه الخصوص عندما يكون الاختبار جديداً وتعليماته معقدة، أما العامل الثاني فيتمثّل فيما يظهره المفحوص الخبير بالاختبار من خصائص تختلف عن المفحوص الأقل خبرة، فالمفحوص الذي تعرّض مسبقاً للاختبار يكون عادةً أقل انفعاليةً وقلقاً، وأكثر وعياً بالمنبهات الثانوية والخارجية التي تفيد في الإجابة. وبشكلٍ عام يمكن التقليل من أثر هذين العاملين عن طريق تحسين تعليمات الاختبار من ناحية، وتقديم قدر من التدريب على بنوده للمفحوصين من ناحيةٍ أخرى (أبو حطب وآخرون، 2003، ص ص. 103 – 104).

وقد أشارت العديد من أدبيات البحث إلى أهمية هذا المتغير، حيث وضح (أبو حطب وآخرون، 2003) أن أحد أهم مصادر تباين الدرجات التي تؤثر في الثبات هي الخصاص المستمرة والعامة في الفرد، والتي تتضمن بشكل أساسي طريقة تفاعله مع الموقف الاختباري، فالتباين في درجات الأفراد ينشأ في العادة نتيجةً لما يظهره الأفراد من مقادير مختلفة في سمات أو خصائص معينة في الموقف الاختباري (المرجع السابق، ص.103).

وأكد عبد الخالق أن درجات الأفراد على الاختبارات بأنواعها المختلفة تتأثر بعدد كبير من المتغيرات والعوامل، لذا فإن الباحث عندما يحصل على الدرجات المستخرجة من المقاييس يميل إلى التفكير في أنها ممثلة حقاً للقدرة الحقيقية أو السمة التي هدف إلى قياسها أم لا، ومن أهم هذه المتغيرات والعوامل: طبيعة موقف القياس، وسمات المفحوص ومدى خبرته بالموقف الاختباري (عبد الخالق، 2005، ص ص. 106- 107).

وعلى الرغم من أهمية متغير التدريب على أداء الاختبار، وأثره المباشر في الدرجات الاختبارية، والذي أظهرته العديد من الأبحاث مثل بحث هونج (Hong et al.,2006) الذي بينت نتائجه أن الطلبة منخفضي التحصيل يعانون من ضعف في مهارات أداء الاختبار، في حين يهتمّ الطلبة ذوي التحصيل المرتفع باستخدام وتفعيل هذه المهارات(Hong et al.,2006, p.151)، وبحث دودين (Dodeen, 2014) الذي أظهر وجود علاقة ارتباطية إيجابية بين التدريب على استراتيجيات أداء الاختبار والاتجاه نحو تعلُّم الرياضيات (Dodeen, 2014, p.11)، وبحث (رشوان، 2017) الذي بيّنت نتائجه ازدياد معدلات درجات الاختبار التحصيلي للطلبة بازدياد درجة استخدامهم إستراتيجيات أداء الاختبار(رشوان، 2017، ص.154)، غير أن عدداً قليلاً جداً من هذه الأبحاث استهدف دراسة أثر هذا المتغير في الخصائص القياسية للاختبار، حيث لم تتمكن الباحثة من الوصول إلا لثلاثة أبحاث فقط تناولت ذلك هي: بحث (Maspons & Llabre, 1985) الذي بينت نتائجه انخفاضاً طفيفاً في معامل الاتساق الداخلي المحسوب بتأثر متغير التدريب، وارتفاع معامل الصدق التنبؤي للاختبار بتأثير هذا المتغير، وبحث بوهل وآخرين (Pohl & et al., 2016) الذي أكدت نتائجه أن انخفاض الخصائص القياسية لمقاييس الكفاءة القرائية مرتبطة بسلوكات التخمين العشوائي والاستجابات المفقودة من قبل طلبة، والناتجة عن نقص التدريب على أداء الاختبار، وبحث (رشوان، 2017) الذي أظهرت نتائجه تحسُّن الصدق التنبؤي والتلازمي للاختبار التحصيلي بازدياد استخدام إستراتيجيات الأداء، كذلك تحسُّن الثبات ولكن بدرجة أقل من الصدق.

ونظراً لعدم وجود اتفاق تام في نتائج الأبحاث حول أثر هذا المتغير في الخصائص القياسية، كان من المناسب إخضاعه لمزيد من الدراسة في البحث الحالي، في ضوء نتائج الدلالات العلمية الثلاث. ولأن البرنامج التدريبي المُعد في هذا البحث قائمٌ على التدريب على إستراتيجيات الأداء في الاختبار كان من الضروري شرحها وتوضيحها بالتفصيل في هذا الفصل.

3- مفهوم إستراتيجيات أداء الاختبار:

تُعد إستراتيجيات أداء الاختبار[6] من العوامل الأساسية التي تؤثر في درجات الفرد الاختبارية، ارتبط هذا المفهوم في أدبيات البحث بمفهوم التدريب على الاختبار وتطور معه. تعبر هذه الإستراتيجيات عن جملة من المعارف والمهارات الأدائية التي تتيح للفرد التصرف بطريقة مناسبة: قبل، وفي أثناء، وبعد إجراء الاختبار، ولكن سيتم التركيز في هذا البحث على إستراتيجيات الأداء في أثناء إجراء الاختبار على وجه التحديد، وهي تلك التي تركز على سلوك الفرد وقدرته على ترجمة ما تعلمه من معارف ومهارات بشكلٍ صحيح في أثناء العملية الاختبارية؛ كالقدرة على إدارة الوقت، والفهم الدقيق للبنود والتعليمات

[6] فضلت الباحثة ترجمة مصطلح " Test Taking Strategies" بإستراتيجيات أداء الاختبار بدلاً من إستراتيجيات أخذ الاختبار أو إستراتيجيات تناول الاختبار، لأن هذه الترجمة توضح المقصود بالمفهوم بصورة أدق؛ من حيث أنها تصف ما يقوم به الفرد في أثناء إجرائه العملية الاختبارية.

الاختبارية، ويعود السبب في ذلك لأن أداة القياس المُستخدمة في البرنامج التدريبي الحالي هي اختبار الذكاء، وهذا النوع من الاختبارات لا يحتاج لإعدادٍ أو تحضيرٍ مسبق لإجرائه، كما هو الحال في الاختبارات التحصيلية.

وقد عرّف كوهين إستراتيجيات أداء الاختبار بأنها: عملياتٌ مختارة بوعي، يستخدمها المتقدمون للاختبار، لمواجهة مشكلات ومتطلبات الإجابة عن العناصر والمهام المطلوبة لإجرائه (Takallou et al., 2016, p.178).

وعرفها دودين: بأنها القدرات المعرفية التي تمّكن الفرد من التعامل مع أي موقف اختباري بطريقة مناسبة، ومعرفة ما الذي يجب فعله في أثناء الاختبار(Dodeen, 2009,P.36).

كما عرفها ستينلاند وآخرون: بأنها مجموعة من الإجراءات التي تُستخدم لفهم أسئلة الاختبار والإجابة عنها بطريقة صحيحة (Stenlund et al, 2017, P 5).

وعرفها رشوان (رشوان، 2017) بأنها: المهارات والفنيات المتصلة بكيفية الأداء في الاختبار، والتي تمكن الفرد من إدارة وقت الاختبار، والسيطرة على مشاعره السلبية، وعرض وتنظيم إجاباته في الاختبار بالطريقة التي تتيح له الحصول على أعلى درجة ممكنة، بما يتوافق مع مستواه في السمة موضع القياس (رشوان، 2017، ص.122).

وترى الباحثة أن هذه الإستراتيجيات تُعبّر عن خطط متعددة، توضع لتحقيق أهداف معينة عند إجراء الاختبار، تتضمن كل منها مهارات فرعية تساعد الفرد على تنفيذها، وهذه الخطط مثل: إدارة وقت الاختبار، وتحليل التعليمات والأسئلة الاختبارية، والتعامل بشكل صحيح مع ورقة الإجابة، وتوظيف الاستدلال الاستنباطي والتخمين الذكي لتحديد الجواب الصحيح، والعمل على ضبط القلق والتوتر في الاختبار، وإجراء المراجعة والتقويم الذاتي للإجابات قبل تسليم الورقة، وأن هذه الإستراتيجيات والخطط عامة؛ أي يمكن تطبيقها في جميع أنواع الاختبارات، ولا تقتصر على نوع محدد، وتتيح للمفحوص الحصول على أعلى قدر ممكن من الدرجات بما يتناسب مع مستواه في السمة موضع القياس.

4- التداخل بين الإستراتيجيات والحكمة الاختبارية:

لاحظت الباحثة لدى مراجعة أدبيات البحث وجود تداخل بين مصطلحي إستراتيجيات أداء الاختبار، والحكمة الاختبارية Test Wiseness؛ إذ لم يتمّ الفصل بشكل واضح بين هذين المصطلحين، وفي كثير من الأحيان استُخدما كمترادفين.

حيث عرف ميلمان وبيشوب وإيبل الحكمة الاختبارية بأنها: مقدرة المفحوص على الاستفادة من خصائص الاختبار أو موقف تناوله، للحصول على درجة أعلى (Millman et al, 1965, P 707)، وهو التعريف الأساسي الذي التي اعتمدت عليه معظم الدراسات السابقة المتصلة بالحكمة الاختبارية.

وعرفها النبهان بأنها: قدرة الفرد على الاستفادة من إستراتيجيات التقييم، والضعف في صياغة الفقرات والمنبهات الموجودة في الاختبار (النبهان، 2004، ص.125).

وعرفها سارنكي بأنها: قدرة معرفية أو مجموعة من المهارات، يستخدمها الفرد لتحسين درجته، وليست مرتبطة بمحتوى المجال المقاس، وهي مهددة محتملة لصدق الاختبار(Sarnacki, 1979, P 252).

ويرى هوبكينز وانتيس أن حكمة الاختبار ما هي إلا موهبة قد يمتلكها مؤدي الاختبار، تسمح له بأن يتفوق على معد الاختبار، الذي من الممكن أن يقع في أخطاء القياس (Hopkins&Antes,1985,p. 480).

وبيّن بينج أنه تم استخدام مصطلح حكمة الاختبار على نطاق واسع، لوصف تكتيكات الاختبار التي يقوم بها الطلاب، للحصول على درجات عالية (Peng,2005, p.17).

يتضح من التعريفات السابقة أن الحكمة بهذا المعنى تعني قدرة المفحوص على الاستفادة من خصائص ومواصفات الاختبار، للحصول على درجات مرتفعة، نتيجة الإفادة من خبرة التعامل مع الاختبار، وتخمين ما يهدف إليه واضع الاختبار، وفحص ورقة الامتحان للبحث عن مفاتيح إجابة إضافية تسهم في الحصول على درجة أعلى دون إتقان فعلي للمحتوى، ومحاولة المفحوص الاستفادة من أخطاء القياس للإجابة عن أسئلة لا يعرفها، والحصول على درجات لا يستحقها.

وقد أكّد ستينلاند وآخرون (Stenlund et al, 2017) أن مهارات الحكمة الاختبارية بمعناها الشائع تناسب بشكل خاص اختبارات الاختيار من متعدد، والتي تعتمّد أساساً على التخمين، كما أشار (رشوان،2017) إلى أنه يمكن اعتبار إستراتيجيات أداء الاختبار جزءاً من الحكمة الاختبارية بمعناها الواسع الشامل وليس بمعناها التقليدي الشائع، وبيّن إكسو وزميله وو (Xu & Wu, 2012) أن معظم المهارات المتصلة بالحكمة الاختبارية بمعناها التقليدي تناسب أنواعاً معينة من الأسئلة الاختبارية، كاختبارات الكفاءة اللغوية وليس لها صفة العمومية، في حين أن إستراتيجيات أخذ الاختبار أو كما سماها بوهل وآخرون (Pohl et al, 2016) سلوكات أخذ الاختبار Test-Taking Behavior لها معنى أوسع وأشمل؛ إذ لا تتضمن فقط السلوكات الصريحة للطالب أثناء الموقف الاختباري المندرجة تحت مهارات الحكمة الاختبارية، ولكنها تتضمن أيضاً القرارات المعرفية والأفكار الإيجابية التي تمّكن الطالب من ترجمة معارفه ومهاراته، والاستفادة منها في أثناء الإجابة عن أسئلة الاختبار، مثل: الفهم الدقيق لتعليمات الاختبار، والتسلسل في عرض الإجابة، والاستدلال الاستنباطي للوصول إلى الجواب الصحيح،

وضبط التأثير السلبي للقلق، ويمكن تطبيقها في أنواع الاختبارات كافة (رشوان، 2017، ص.125 .(Stenlund et al, 2017, p. 5 , Pohl et al,2016, p.7,

وقد ميّز كوهين في تصنيفه لإستراتيجيات أخذ الاختبار بين ثلاثة أنواع أساسية منها هي: إستراتيجيات لغة المتعلم؛ وتعني الطريقة التي يبرز بها المتعلم معارفه ومهاراته، إستراتيجيات إدارة الاختبار، وإستراتيجيات حكمة الاختبار؛ وهو بذلك ضمّن حكمة الاختبار وعدّها واحدة من إستراتيجيات أخذ الاختبار (Cohen, 2006, p.311).

كذلك الأمر في كلٍ من بحثي بينج (Peng, 2005) وستينلاند وآخرون (Stenlund et al, 2017) حيث ضمّن الباحثون فيها مبادئ الحكمة الاختبارية ضمن إستراتيجيات أداء الاختبار، واعتبروها إحدى الإستراتيجيات التي يتبعها الفرد في تناوله الموقف الاختباري.

يتبيّن مما سبق أن مصطلح الحكمة الاختبارية بمعناه التقليدي الشائع يمثل جانباً جزئياً من مصطلح إستراتيجيات أداء الاختبار، وعلاقة الحكمة الاختبارية بإستراتيجيات أداء الاختبار تمثّل علاقة الجزء بالكل، فمصطلح إستراتيجيات أخذ الاختبار ينطوي على إستراتيجيات الحكمة الاختبارية من حيث اشتماله إستراتيجية التخمين المبني على المعلومات، ويتعداها إلى إستراتيجيات أخرى أعم وأشمل غير متضمنة في إطار مصطلح الحكمة الاختبارية، مثل: تحليل التعليمات الاختبارية، وتنظيم عرض الإجابة، والأفكار الإيجابية، والقرارات النفسية المتصلة بضبط القلق والتوتر، والسيطرة على المشاعر السلبية الناتجة عن طبيعة الموقف الاختباري، والتقويم الذاتي من خلال إجراء مراجعة شاملة للأسئلة والإجابات المطلوبة في ورقة الاختبار.

5- تصنيفات الإستراتيجيات كما وردت في الأبحاث:

تعددت تصنيفات الباحثين لإستراتيجيات أداء الاختبار تبعاً للتعريفات التي تبناها كلٌ منهم، والهدف من وراء هذه التصنيفات تسهيل تعرُّف هذه الإستراتيجيات، ودراستها بشكل أدق وأشمل، وقد بدأت هذه التصنيفات على يد ميلمان وبيشوب وإيبل (Millman et al., 1965) في تصنيف مبادئ الحكمة الاختبارية إلى فئتين رئيستين هما: إستراتيجيات مستقلة عن مصمم وهدف الاختبار، وتتضمن: استخدام وقت الاختبار، وتجنب الخطأ، والتخمين، والاستدلال الاستنباطي. وإستراتيجيات معتمدة على تصميم وهدف الاختبار وتتضمن: إستراتيجيتي مراعاة القصد أو الهدف، واستخدام الدليل أو الرمز. ثم صنفها كوهين (Cohen, 1998) إلى: إستراتيجيات لغة المتعلم، وإستراتيجيات إدارة الاختبار، وإستراتيجيات حكمة الاختبار (Cohen, 2006, p.311)، وصنّفها بينج (Peng, 2005) إلى: الإستراتيجيات المعرفية، والإستراتيجيات ما وراء المعرفية، والإستراتيجيات التحفيزية والعاطفة متضمنةً

خفض قلق الاختبار، وأخيراً تكتيكات أو حكمة الاختبار، والتي تتضمن خمس مهارات أساسية هي: الإزالة، والتوقع، والتخمين، وتجنب الخطأ، واستخدام التلميحات.

وصنف هونج وزملاؤه **(Hong et al,2006)** هذه الإستراتيجيات إلى ثلاثة أنواع هي: إستراتيجيات التحضير للاختبار، وإستراتيجيات الوعي بالاستعداد للاختبار، وإستراتيجيات أداء الاختبار. وأسفرت نتائج التحليل العاملي الاستكشافي في بحث بيكاك (Bicak,2013) عن وجود جزأين أساسيين للمقياس المعد لاستراتيجيات اداء الاختبار، هما: إستراتيجيات التحضير للاختبار، وإستراتيجيات أداء الاختبار.

وفي بحث دودين (Dodeen, 2014) صُنّفت إستراتيجيات أداء الاختبار في المقياس الذي بناه الباحث إلى: إستراتيجيات قبل الاختبار، وإستراتيجية إدارة الوقت، وإستراتيجيات أثناء الاختبار، وإستراتيجيات بعد الاختبار. أما في بحث ستينلاند وآخرين (Stenlund et al, 2017) فقد صُنّفت هذه الإستراتيجيات إلى نوعين أساسيين هما: إستراتيجيات الإعداد للاختبار، وإستراتيجيات إجراء الاختبار، والجزء المتعلق بإستراتيجيات إجراء الاختبار جرى تضمينه ثلاث مجموعات من الإستراتيجيات هي: إستراتيجيات التنظيم الهيكلي، وإدارة الوقت، والحكمة الاختبارية.

وصنف(رشوان، 2017) إستراتيجيات أداء الاختبار في بحثه إلى: إستراتيجية تخطيط وتنظيم الوقت، إستراتيجية تخطيط وتنظيم الإجابة، إستراتيجية تحليل الأسئلة، إستراتيجية الضبط والسيطرة، وإستراتيجية ما بعد الاختبار. وفي بحث (الطوطو، 2018) تألف مقياس حكمة الاختبار من أربعة أبعاد هي: إستراتيجيات قبل البدء بالاختبار، إستراتيجيات في أثناء أداء الاختبار، إستراتيجيات بعد الانتهاء من الإجابة عن الاختبار، وإستراتيجيات الحكمة الاختبارية المتصلة باختبار الاختيار من متعدد. وأخيراً صنف (Maspons and Llabre, 1985) هذه الاستراتيجيات إلى: إدارة الوقت، تجنب الخطأ، التخمين، والاستدلال الاستنباطي.

6- الإستراتيجيات المُستخدمة في البحث الحالي:

باستقراء تصنيفات الأبحاث السابقة لإستراتيجيات أداء الاختبار، وفي ضوء أهداف البحث الحالي تبنّت الباحثة مجموعة من الإستراتيجيات، التي جرى التأكيد على أهميتها في معظم الأبحاث من ناحية، والتي ركزت على الإستراتيجيات التي يوظّفها المفحوص في أثناء إجراء الموقف الاختباري على وجه الخصوص من ناحيةٍ أخرى، وفيما يلي عرضاً مفصلاً لهذه الإستراتيجيات:

- **إدارة وقت الاختبار:**

تتمثل هذه الإستراتيجية في القدرة على تنظيم وقت الاختبار، والالتزام به واستخدامه بطريقةٍ فعالة، وتتضمن العديد من المهارات الفرعية، منها:

- البدء بالإجابة عن أسئلة الاختبار بأسرع وقت ممكن، مع المحافظة على أكبر قدر ممكن من الدقة والإتقان.
- توزيع زمن الاختبار في ضوء صعوبة الأسئلة، ونوعها: مقالية أو موضوعية، والدرجات المخصصة لها. من خلال البدء بالإجابة عن الأسئلة الأسهل ثم الأصعب، والإجابة عن الأسئلة التي تنتج درجات أكثر في وقت أقل، وتجنب إضاعة الوقت في الإجابة عن سؤال يستغرق الكثير من الوقت ودرجته أقل.
- وضع إشارة أو تسجيل أرقام الأسئلة التي تم تأجيل الإجابة عنها، أو التي يعتقد المفحوص أنها تحتاج إلى مراجعة، وذلك لسرعة الوصول إليها لاحقاً.
- وضع تصوّر لما يُفترض إنهاؤه من الإجابات بعد مرور زمن معين.
- كتابة الأفكار الرئيسة دون الاهتمام الكبير بالتفاصيل، في حال ضيق الوقت في الأسئلة المقالية.
- عدم ترك الاختبار قبل انتهاء الوقت المخصص له، والاستفادة من الوقت المتبقي قدر الإمكان في المراجعة.
- الحرص على ترك بعض الوقت في آخر الاختبار للمراجعة (حماد، 2010 ، ص ص. 15 – 16، المالكي، 2010، ص.77، مطلك، 2009، ص. 574، Dodeen,2015, p112).

ومن الجدير بالذكر أن هناك العديد من العوامل التي تتدخل في إستراتيجية إدارة وقت الاختبار لدى الفرد، وتؤثر في خطة تنظيمه لها، منها ما يتعلق بالاختبار ذاته مثل: عدد الأسئلة ونوعها مقالية أو موضوعية، والوقت المحدد للاختبار، ونوعية تعليمات الإجابة المطلوبة؛ كالإجابة على نفس الورقة أو على ورقة منفصلة، ومن هذه العوامل ما يتعلق بطبيعة المفحوص نفسه وخصائص شخصيته، مثل: إمكاناته واستعداده للاختبار، وقدرته على فهم المطلوب من أسئلته بدقة، وخبرته بالعملية الاختبارية. وهي عوامل نسبية تختلف من فرد لآخر، ومن موقف اختباري لآخر، ولكن ما يجب تأكيده في هذا السياق، والذي يشير إليه النبهان هو أنه مهما كانت الخطة المتبعة في إدارة وقت الاختبار، فإنها يجب أن تهدف إلى استثمار الوقت بحكمة، بحيث تضمن الإجابة عن كافة المهام المطلوبة خلال الوقت المحدد (النبهان، 2004 ص.123).

- **تحليل التعليمات والأسئلة الاختبارية:**

تتضمن هذه الإستراتيجية جانبين أساسيين، هما:

- القراءة والفهم الدقيق والمتمعّن لتعليمات الاختبار؛ أي تحليل التعليمات الاختبارية.
- القراءة والفهم الدقيق والمتمعّن للأسئلة الاختبارية؛ أي تحليل الأسئلة الاختبارية.

ويتضمن كل جانب منهما العديد من المهارات الفرعية التي يوضحها الجدول الآتي.

جدول (8)

مهارات إستراتيجية تحليل التعليمات والأسئلة

	تحليل التعليمات الاختبارية	تحليل الأسئلة الاختبارية
المهارات الفرعية	– قراءة التعليمات المتصلة بزمن الاختبار. – الانتباه للتعليمات المتصلة بالبيانات الشخصية المطلوبة مثل تسجيل: الاسم الثلاثي، الرقم الامتحاني، السنة الدراسية، الفرع أو الاختصاص، تاريخ إجراء الامتحان ... – الانتباه للتعليمات الخاصة بالإجابة على نفس الورقة أو الإجابة على ورقة منفصلة كما هو الحال في الأسئلة المؤتمتة. – الانتباه للتعليمات الخاصة بالمساحة المخصصة لتسجيل الإجابة، ففي بعض الأوراق الاختبارية مساحة من الورقة تُعد مسودة لا تُصحح. – معرفة نوع ولون الأقلام المسموح باستخدامها. – الالتزام بالأدوات المسموح باستخدامها كالآلة الحاسبة، والأدوات الهندسية.	– فهم وتحليل المطلوب من كل سؤال من الأسئلة الاختبارية. – الانتباه للأسئلة الإجبارية والأسئلة الاختيارية، وعدم الإجابة عن أسئلة غير مطلوبة، أو ترك أسئلة مطلوبة. – الانتباه إلى الأسئلة المركبة أو المتضمنة أكثر من مهمة، وتجزئتها لتجنب نسيان أي جزء من السؤال. – وضع خط تحت عبارات النفي إن وجدت في نص السؤال، من مثل: ماعدا، لا، لم، غير صحيح. – الانتباه لوجود كلمات مثل: فقط، أبداً، دائماً، في نص السؤال لأنها تساعد على تحديد المطلوب بدقة. – وضع خط تحت الكلمات المهمة في نص السؤال.

(أبو هاشم، 2008، ص.226، الزعبي، 2018، ص.562)

- **التعامل مع ورقة الإجابة:**

تُعد من الإستراتيجيات المهمة في الاختبارات المقالية على وجه الخصوص، والتي تحتاج لتنظيم أكثر للإجابة، والوضوح والمنطقية في عرضها، ويسمي رشوان هذه الإستراتيجية بإستراتيجية تخطيط وتنظيم الإجابة (رشوان، 2017، ص.135).

وتركز بشكلٍ أساسي على طريقة تسجيل الإجابة بالشكل الصحيح على الورقة، كما تتضمن عدّة مهارات فرعية منها:

- كتابة الإجابة في المكان المخصص لها، والتأكُّد من أن رقم المفردة التي يجيب عنها المفحوص على ورقة الإجابة مطابقاً لرقم السؤال على ورقة الأسئلة.
- عند ترك أية مفردة دون إجابة يجب ترك المكان المخصص لها فارغاً على ورقة الإجابة.
- كتابة رقم السؤال على ورقة الإجابة في حال لم يكن موجوداً.
- يُفضَّل ترك مساحة خالية على الورقة عقب الانتهاء من الإجابة عن كل سؤال، إن كان ذلك ممكناً، حتى تتيح الفرصة للمفحوص فيما بعد لإضافة معلومات أخرى يتذكرها لتكملة الإجابة أو تعديلها إن لزم الأمر.
- التخطيط جيداً للإجابة قبل البدء بكتابتها خاصةً في الأسئلة المقالية، من خلال التفكير بالإجابة وتجهيزها في الذهن، والاهتمام بعرض عناصرها في تسلسلٍ منطقي، وتحديد الأفكار الرئيسية والأفكار الفرعية التي تتضمنها، كما يُفضل إنشاء مسودة يتم فيها التخطيط للإجابة قبل البدء بكتابتها.
- التركيز في الإجابة عن سؤالٍ واحدٍ فقط، وعدم تشتيت الذهن بالتفكير في أسئلة عدة بوقتٍ واحد.
- الكتابة بخطٍ واضحٍ ومقروء.
- تجنب الشطب في ورقة أو دفتر الإجابة قدر الإمكان (القرشي،2017، ص.338، عبد الوهاب،2007، ص ص. 29-31).

- **الاستدلال الاستنباطي**:

تُعبر هذه الإستراتيجية عن مهارة المفحوص في الاستدلال على الإجابة الصحيحة، من خلال التحليل المنطقي للمعلومات المستمدة من البند الاختباري ذاته، أو من بنودٍ أخرى في الاختبار، وذلك إذا لم يعرف الإجابة الصحيحة على وجه التحديد، من خلال استخدام المعلومات المُستمدة من البنود الاختبارية، والإفادة من معلومات المحتوى ذات الصلة مع غيرها من عناصر الاختبار، في محاولة الاستدلال على الإجابة الصحيحة (سليمان، 2014، ص.249، حماد، 2010، ص.17).

وقد عرّفها محمد علي بأنها: الاستفادة من المعلومات المتاحة في الاختبار لأداء المهمة المطلوبة (محمد علي، 2016، ص. 264)، وتتضمن هذه الإستراتيجية العديد من المهارات الفرعية مثل:

- حذف الأبدال التي يكون المفحوص متأكداً من خطئها، ثم الاختيار من بين الأبدال المتبقية.
- استبعاد البديلين اللذين يدلان ضمنياً أو يؤديان إلى نفس المعنى، إذا كان المطلوب اختيار إجابة واحدة صحيحة.
- اختيار أو حذف أحد البديلين المتعاكسين، والسبب في ذلك أنه إذا كان أحدهما يمثل البديل الصحيح فالآخر حتماً خاطئ.
- حصر الإجابة في البديل الذي يتضمن أكبر قدر ممكن من المعلومات الصحيحة.

- الاستفادة من معلومات المحتوى ذات الصلة، المكتسبة من السؤال نفسه أومن أسئلة أخرى في الاستدلال على الإجابة الصحيحة (Millman et al, 1965, P.p 711- 712, Sarnacki, 1979, p.255).

- **التخمين:**

يشير التخمين إلى السلوك الذي يلجأ إليه الفرد للإجابة عن فقرات الاختبار، عندما لا تكون لديه المعرفة الكافية والأكيدة، التي تمكنه من تحديد الإجابة الصحيحة من ضمن مجموعة الأبدال المطروحة للسؤال (الزبون، 2011، ص.8)، وقد جرى التمييز بين نوعين من التخمين هما:

- **التخمين الأعمى أو العشوائي:** وهو التخمين الذي لا يتروّى فيه الفرد في أثناء إجابته عن الأسئلة التي لا يكون متأكداً من صحتها، ويتم إجراؤه دون ترجيح لأحد الأبدال، كأن يقوم بإجراء القرعة بين الأبدال عند الإجابة عن الأسئلة الموضوعية، وهنا يكون احتمال حصوله على الإجابة الصحيحة هو 1/عدد الأبدال. ومن النادر جداً أن يلجأ المفحوص إلى الإجابة بطريقة التخمين العشوائي إلا إذا انتفت لديه الدافعية، أو كان الزمن المتاح للاختبار غير كافٍ (الخرشة، 2016، ص.3).
- **التخمين الذكي أو المبني على معلومات:** هو التخمين الذي يعتمد فيه المفحوص على محكات ودلائل تقود إلى الإجابة الصحيحة، ويستفيد فيه من المنبهات أو المفاتيح التي قد توجد في الاختبار، من أجل الإجابة عن المفردات التي يصعب عليه الإجابة عنها في ضوء معرفته بالمحتوى الذي تتناوله هذه المفردات، وقد تكون هذه الدلائل من صيغة السؤال نفسه أو بعض الكلمات في الأبدال المختلفة للسؤال، والتي قد تدل على الإجابة الصحيحة (الزبون، 2011، ص.8).

ويتطلب هذا النوع من التخمين القدرة على التفكير الاستدلالي في الإجابة، ولعل المفاتيح أو المنبهات التي قد تكون في أسئلة الاختبار تعد بمنزلة مقدمات للمفحوص، تمكنه من أن يخمن تخميناً ذكياً، يستدل من خلاله على الإجابة الصحيحة بدلاً من أن يترك السؤال دون إجابة، أما إذا لم يقم المفحوص بالانتباه إلى هذه المنبهات والاستفادة منها فإن تخمينه سيكون عشوائياً، ومن المفاتيح أو المنبهات التي قد يستخدمها المفحوص في أثناء الإجابة ما يلي:

- وجود أبدال غير معقولة أي مؤكد عدم صحتها، يقوم المفحوص باستبعادها، مما يزيد احتمالية التوصل إلى الإجابة الصحيحة.

- عدم وجود اتفاق نحوي أو لغوي بين المتن والأبدال غير الصحيحة، مثل عدم وجود اتفاق في الإفراد والجمع، أو التذكير والتأنيث، أو علامة الإعراب، أو صياغة وبناء الجملة.
- احتواء الأبدال على كلمات قد تشير إلى صحة البديل أو العبارة مثل: غالباً، بشكلٍ عام، أو كلمات قد تشير إلى خطأ البديل أو العبارة مثل: لا يوجد، يستحيل، أبداً.
- وجود بديلين لهما المعنى نفسه، مما يعني أن كليهما ليس البديل الصحيح.
- وجود بديلين يتضادان في المعنى، مما يعني أن أحدهما في الأغلب صحيح والآخر خاطئ.
- وجود البديل كل ما سبق صحيح، يستدل على صحته إذا وجد بديلان صحيحان، ويستدل على خطئه إذا وجد بديل واحد من المؤكد عدم صحته.
- وجود البديل كل ما سبق خاطئ، يستدل على خطئه إذا وجد بديل واحد صحيح.
- قد تكون مفاتيح الإجابة عن سؤالٍ ما في فقرات أخرى من الاختبار، وهذا يرتبط ارتباطاً وثيقاً بإستراتيجية الاستدلال الاستنباطي (المالكي، 2010، ص ص. 64 – 69).

<u>وتتمثل تعليمات إستراتيجية التخمين حسب معادلة التصحيح من أثر التخمين على النحو الآتي:</u>

- خمن دائماً إذا لم يتم تطبيق معادلة تصحيح للدرجات، أي إذا كانت تعليمات السؤال لا تتضمن أن الإجابة الخاطئة تذهب درجة إجابة أو عدة إجابات صحيحة.
- خمن دائماً إذا كانت عقوبة التخمين أقل صرامة من العقوبة الناتجة عن تطبيق معادلة التصحيح.
- خمن دائماً إذا استطعت استبعاد بعض الأبدال حتى مع تطبيق معادلة التصحيح.
- خمن حتى لو كان التصحيح من أثر التخمين أكثر خطورة، بشرط استبعاد الأبدال التي تعطي فرصة كافية للاستفادة (Millman et al., 1965, P. 711).

- **الضبط الذاتي:**

وتُسمى أيضاً بإستراتيجية خفض قلق الاختبار، وتعني قدرة الفرد على ضبط أفكاره ومعتقداته السلبية عن كفاءته وقدراته، من خلال محاولة الحد من أية أفكار أو مؤثرات سلبية قد تؤثر في جودة أدائه في الاختبار (رشوان، 2017، ص.136).

ويمثل قلق الاختبار أحد الأنماط الخاصة من القلق، ويمكن تحديده بأنه حالة نفسية ترتبط بالخوف من الاختبارات، تظهر على شكل مجموعة من الإثارات الفيزيولوجية والنفسية، مثل الشعور بالقلق والتوتر وتشتت الفكر، وبعض الأعراض البدنية المؤلمة التي قد تظهر قبيل أو في أثناء

الاختبار، وتؤثر سلباً في قدرة الفرد على التفكير السليم واسترجاع المعلومات وتنظيمها خلال العملية الاختبارية، لدرجة قد تؤدي إلى إضعاف مستوى الأداء في الاختبار، وهذه الحالة يمكن أن تؤثر سلباً في إنجاز الفرد بصفة عامة، كما تؤثر في سلوكه وتوازنه الانفعالي ومفهومه عن ذاته (كانوري،2013، ص.11).

وتعد إستراتيجية الضبط الذاتي من إستراتيجيات أداء الاختبار المهمة، التي تساعد الفرد على التحكم في مشاعر القلق والتخفيف منها قدر الإمكان. ويُعرَّف ضبط الذات على أنه: قدرة الفرد على التحكم شعورياً وعقلانياً بسلوكه ودوافعه، وإبقائها تحت السيطرة وتوجيهها، وإيقاف ردّات الفعل المباشرة، مع التعزيز الإيجابي للذات عند تحقيق الأهداف المرجوة، أو القيام بالسلوك المطلوب، وهي مهارة يمكن للفرد أن يتعلمها على نحو منتظم، ويزداد إتقانها كلما تدرب عليها ومارسها.

وتتضمن إستراتيجية الضبط الذاتي العديد من المهارات والآليات التي تساعد على التخفيف من قلق الاختبار، منها ما يفيد تنفيذه قبل الموقف الاختباري مثل: الاستعداد الجيد للاختبار، وتحديد أهداف واقعية، وعمل تدريبات كما لو أن الشخص في قاعة الاختبار، ومنها ما يفيد تنفيذه في أثناء الموقف الاختباري مثل: استخدام أساليب الاسترخاء عن طريق التنفس العميق، والتفكير الإيجابي، وتنظيم الأفكار قدر الإمكان، وحصرها بفقرات الاختبار وأسئلته للحد من الأفكار السلبية الأخرى، ومنها ما يفيد تنفيذه بعد إجراء الاختبار: مثل التعبير عن المشاعر والانفعالات، من خلال الحديث مع أفراد إيجابيين عن مشاعر القلق تجاه الاختبار، ليخفف الفرد من ضغطه النفسي، ويتخلص تدريجياً من القلق (أبو هاشم، 2008، ص.225).

- **المراجعة:**

تتمثل هذه الإستراتيجية في تقويم الفرد لتجربته في الاختبار، والتأمل والتمعن الموجه لكافة العمليات التي يجريها في الموقف الاختباري، من خلال إجراء تقويم ذاتي لاستجاباته في كل المهام المطلوبة منه في الاختبار، وتتضمن مهارات فرعية عديدة مثل:

- إعادة قراءة ورقتي الأسئلة والإجابة بتمعُّن.
- التأكد من عدم ترك أية أسئلة دون إجابات، من خلال مطابقة عدد الأسئلة المطلوب الإجابة عنها مع عدد الإجابات المدونة.
- التأكد من تسجيل كل إجابة في مكانها الصحيح على ورقة الإجابة، وإزالة الإجابات التي قام المفحوص بتغييرها إزالة تامة.
- التحقق من استيفاء الإجابات لكل ما هو مطلوب في نص السؤال.
- مراجعة البيانات الشخصية قبل تسليم الورقة.

- تحديد نوع الخطأ إن وجد، ليتسنى مراجعته عقب الانتهاء من العملية الاختبارية (عبد الوهاب، 2007، ص.31، Dodeen,2015, p112).

7- أهمية إستراتيجيات أداء الاختبار:

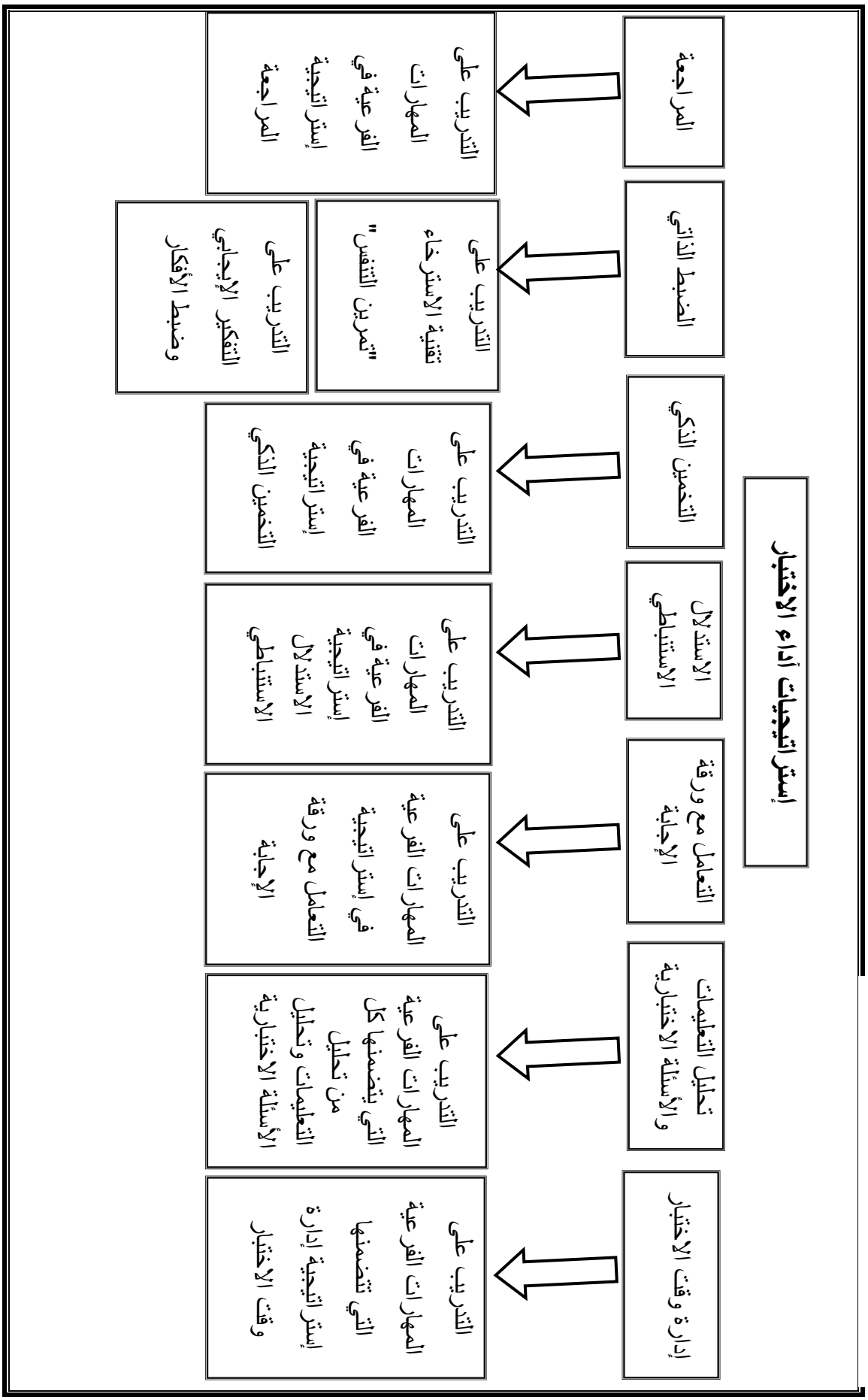
تتسم إستراتيجيات أداء الاختبار بعددٍ من المزايا التي تجعل لها أهمية كبيرة ليس فقط في العملية الاختبارية، وإنما أيضاً في مجال القياس والتقويم، فقد زاد الاهتمام بهذه الإستراتيجيات باعتبارها مصدراً مهماً وعاملاً مفسراً يعزى إليه جزء مهم من تباين الدرجات، يمكن أن يؤثر في صدق تعبير الدرجات عن السمة المُقاسة، حيث أكد جوين (Guion, 1998) ضرورة وضع هذه الإستراتيجيات على أجندة البحث العلمي في مجال القياس والتقويم، بوصفها متغيراً أساسياً غير خاضع للقياس، يؤثر في دقة التنبؤات حول السمة المقاسة، وأشار كوهين وأبتون (Cohen & Upton ،2007) إلى أن تدريب المفحوصين على هذه الإستراتيجيات يمكن أن يؤدي إلى زيادة صدق الاختبار ووثباته؛ من خلال التقليل من التباين غير المفسر في الدرجات (Ghafournia & Afghari,2013, p.141, Sarnacki,1979, p.253)

إضافةً إلى ذلك فقد أكدت معظم الأبحاث التي تناولت دراسة العلاقة بين التدريب على إستراتيجيات أداء الاختبار ونتائج الاختبارات التحصيلية وجود علاقة طردية دالة إحصائياً بين هذين المتغيرين، كما دعت العديد من أدبيات البحث إلى جعل هذه الإستراتيجيات جزءاً مكملاً للمنهاج الدراسي، إذ بيّن بحث بينج وآخرين (Peng et al., 2014) الحاجة الضرورية لتدريب المتعلمين على استخدام هذه الإستراتيجيات، لأنها تساعدهم على نقل معارفهم التعليمية إلى أجواء الاختبار.

ويرى آخرون أن هذه الإستراتيجيات تتسم بخاصية انتقال أثر التعلُّم، إذ يستطيع الفرد الاستفادة منها وتوظيفها في حياته العملية؛ كالاستخدام الفعال للوقت، والقدرة على وضع الأهداف، والعمل بسرعة ودقة، وإيجاد الأدلة للبرهنة على صحة الأفكار، مما يجعل لها أهمية كبيرة ليس فقط في العملية الاختبارية، بل في خصائص شخصية الفرد وقدرته على التعامل مع ظروف الحياة المختلفة (Dodeen, 2014, p.1، Khosravani &Zanjani, 2017, p.89). والشكل الآتي يوضح هذه الإستراتيجيات.

شكل (3)

إستراتيجيات أداء الاختبار



خلاصة:

قدّم هذا الفصل عرضاً مفصلاً لأدبيات البحث المتصلة بموضوع الدلالات العلمية، ووضح مفهوم كل نوع منها وأهميته وطرائق حسابه، وجوانب القصور فيه التي استدعت ظهور النوع الذي يليه، وصولاً إلى الدلالة العيادية، وذلك بهدف التمييز بينها، وتوضيح أهمية التكامل في حساب مؤشراتها للوصول إلى نتائج وقرارات بحثية على قدرٍ كافٍ من الدقة والموثوقية. كما عرض حديثاً عن الخصائص القياسية للاختبار، أتبعه بآخر عن أثر التدريب فيها، وفق إستراتيجيات أداء الاختبار المُعتَمدة في البحث الحالي.

الفصل الرابع

منهج البحث وإجراءاته

الفصل الرابع

منهج البحث وإجراءاته

تمهيد:

يتناول هذا الفصل إجراءات تنفيذ البحث ميدانياً، متضمناً منهج البحث، ومجتمعه وعيناته، ومتغيراته وأدواته، ويوضح وصفاً لإجراءات تنفيذه المتصلة بكل منهج من المناهج البحثية المُستخدمة فيه، وأخيراً يعرض موجزاً للأساليب الإحصائية المُستخدمة في معالجة البيانات وتحليلها للوصول إلى النتائج.

1- منهج البحث:

لتحقيق أهداف البحث والإجابة عن أسئلته اعتُمِدت ثلاثة أنواع من المناهج البحثية على النحو الآتي:

- منهج البحث الوصفي التحليلي في الإجابة عن السؤال الأول، المتصل بتحليل واقع استخدام الدلالات العلمية الثلاث في الرسائل الأكاديمية التجريبية المنجزة في كلية التربية.
- منهج التحليل البعدي في الإجابة عن أسئلة البحث: الثاني والثالث والرابع، المتصلة بدراسة مدى كفاية كل من الدلالتين الإحصائية والعملية في الإبلاغ عن نتائج الأبحاث التجريبية، والقدرة التنبؤية لكل نوع من الدلالات بالنوعين الآخرين.
- واتُبعت خطوات منهج البحث التجريبي في الإجابة عن الأسئلة المتصلة بالبرنامج التدريبي، وحساب نتائج تطبيقه وتفسيرها استناداً لأنواع الدلالات العلمية.

2- مجتمع البحث وعينته:

<u>مجتمع البحث:</u>

لم يقتضِ اعتماد البحث على ثلاثة مناهج بحثية وجود ثلاثة مجتمعات إحصائية أصلية، وإنما اقتصر على مجتمع إحصائي واحد فقط ارتبط بمنهجي البحث الوصفي التحليلي، والتحليل البعدي، وتمثل بجميع رسائل الماجستير والدكتوراه ذات التصميم التجريبي المنجزة في كلية التربية – جامعة دمشق بالأقسام الآتية: المناهج وطرائق التدريس، علم النفس، الإرشاد النفسي، والتربية الخاصة، بين عامي 2000-2020 الميلاديين، والفصل الأول من العام الدراسي 2020 – 2021م. مع الأخذ بالحسبان عدم تضمين قسم تربية الطفل ضمن المجتمع الأصلي لعدم وجود أرشفة خاصة بأبحاثه قبل العام الدراسي 2007- 2008م، وهو عام تأسيسه.

بالمقابل، لم يتم تحديد أي مجتمع إحصائي خاص باعتماد المنهج التجريبي، كون عيناته غير احتمالية، ولا تتطلب تعميم نتائجها على مجتمع إحصائي أساساً.

عينة البحث: تكونت من عينتين أساسيتين هما:

أ- **عينة الرسائل الأكاديمية:** تألفت من جميع رسائل الماجستير والدكتوراه التجريبية أو ذات التصميم التجريبي، الصادرة عن كلية التربية خلال الفترة الزمنية الممتدة بين عامي 2000 – 2020 الميلاديين، والفصل الأول من العام الدراسي 2020 – 2021م في أقسام المجتمع الأصلي المحددة.

أي كان حجم العينة مساوياً لحجم المجتمع الأصلي للفترة الزمنية المستهدفة وللأقسام الأربعة، وبنسبة تمثيل 100%.

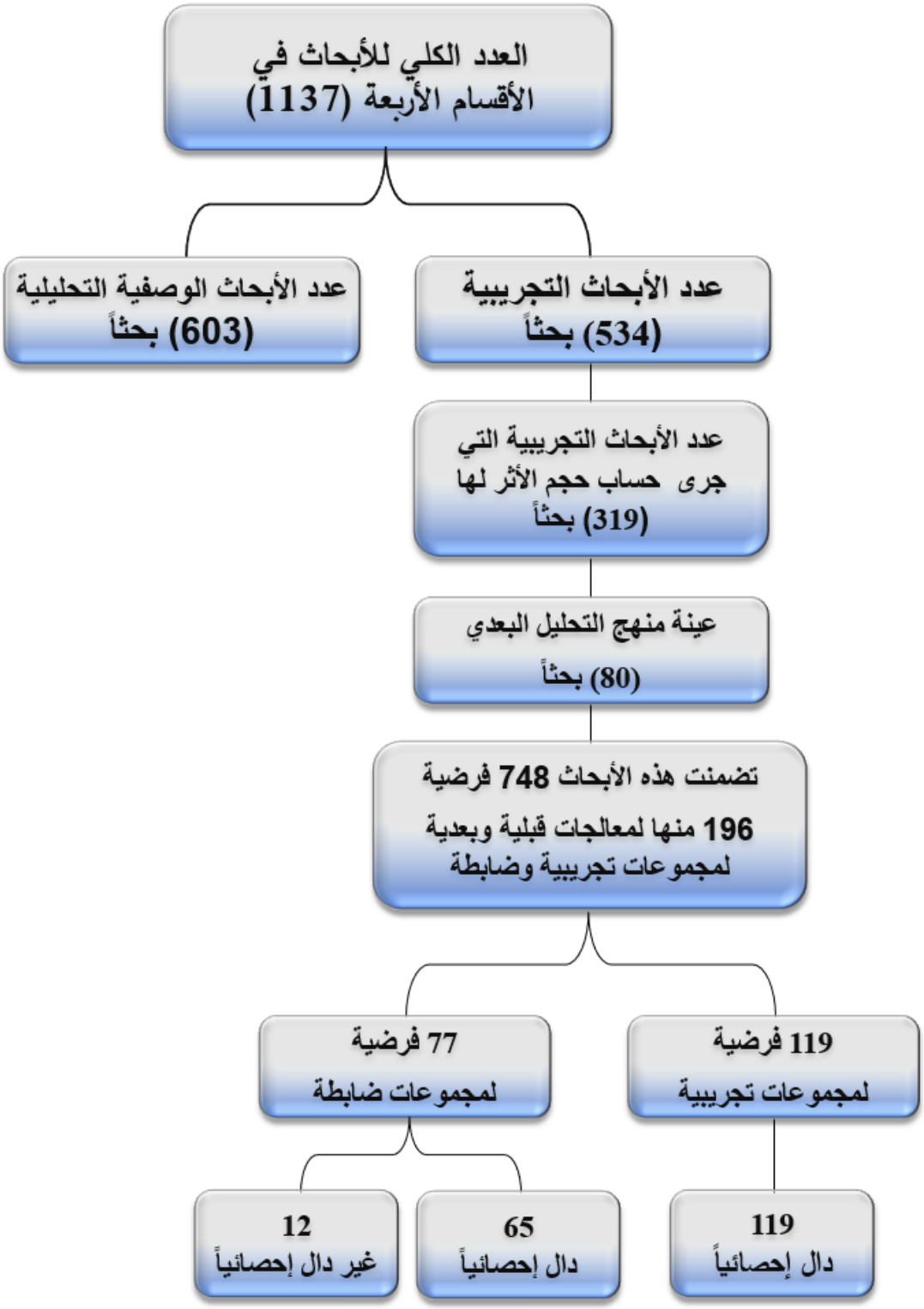
وقد استهدفت الباحثة هذه الأقسام لأن رسائلها تتضمن عينة جيدة من الأبحاث التجريبية، ولا تقتصر على الأبحاث الوصفية التحليلية؛ وبالتالي تسمح بدراسة أنواع الدلالات العلمية الثلاث المُستخدمة فيها، كما تسمح بحساب معاملات الدلالة العيادية لنتائجها ومقارنة قيمها مع الدلالتين الإحصائية والعملية. ومن الجدير بالذكر أن قسم تربية الطفل يحتوي عينة جيدة من الأبحاث التجريبية، ولكن نظراً لعدم وجود أبحاث مؤرشفة لهذا القسم من عام 2000 وحتى عام 2008م فقد تم استبعاده من عملية التحليل، والاكتفاء بالأقسام الأربعة السابقة الموجودة منذ عام 2000م في أرشيف مكتبة كلية التربية.

بلغ العدد الكلي للأبحاث في الأقسام الأربعة عينة البحث (1137) بحثاً، (603) منها وصفي تحليلي، و(534) منها تجريبي. وبلغ عدد الأبحاث التجريبية التي جرى حساب حجم الأثر في نتائجها (319) بحثاً، نسبتها المئوية (59.74%) من مجمل الأبحاث التجريبية. سحبت منها الباحثة عينة الأبحاث المجمعة لمنهج التحليل البعدي بلغ عددها (80) بحثاً، نسبتها المئوية (25.08%) من مجمل عينة الأبحاث التجريبية التي جرى حساب حجم الأثر لها.

بلغ العدد الكلي للفرضيات في الأبحاث الثمانين (748) فرضيةً، قامت الباحثة بحساب قيم وتصنيفات الدلالة العيادية ل(196) فرضية منها، وهي التي اختصت بدراسة الفروق بين المتوسطات القبلية والبعدية وتعادل نسبتها (26.2%) من مجمل الفرضيات، (119) فرضية منها لمجموعاتٍ تجريبية، و(77) فرضية منها لمجموعاتٍ ضابطة، والشكل البياني الآتي يوضح تدفق هذه العينة.

شكل (4)

تدفق عينة الرسائل الأكاديمية



ب- عينة البحث التجريبية: وهي عينة عرضية متيسرة تكونت من طلبة السنة الثانية من اختصاص علم النفس، بزمرها الأربع، بلغ عدد أفرادها 125 طالباً وطالبة، جرى توزيع أفرادها إلى مجموعتين، تتضمّن

كل مجموعة زمرتين وذلك فقط لسهولة عملية التدريب، وضمان توفير بيئة تدريبية مريحة تسمح بإجراء أنشطة البرنامج التدريبي، مع العلم أن هاتين المجموعتين تعرضتا لجلسات البرنامج ذاتها، وبالتسلسل ذاته دون أي تغيير، والجدول الآتي يوضح الخصائص الديموغرافية لعينة البحث.

جدول (9)

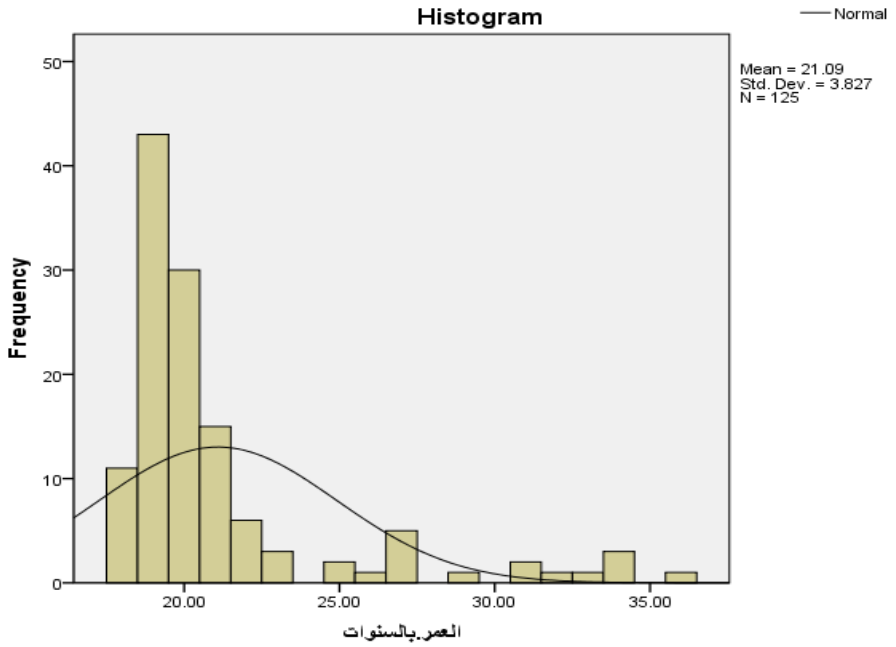
الخصائص الديموغرافية للعينة

نوع العينة		الاختصاص	السنة الدراسية	العدد	النسبة المئوية	متوسط العمر	الانحراف المعياري
التجريبية	الإناث	علم نفس	الثانية	108	%86.4	21.27	4.07
	الذكور			17	%13.6	19.94	0.97
	المجموع			125	%100	21.09	3.83

والشكل البياني الآتي يوضح توزّع أفراد عينة البحث حسب العمر بالسنوات.

الشكل (5)

أعمدة بيانية لتوزع أعمار أفراد العينة



3- متغيرات البحث: توزّعت المتغيرات في البحث إلى نوعين أساسيين هما:

- **المتغيرات المستقلة:** وهي: متغير التدريب على الاختبار.

- **المتغيرات التابعة: وهي:**

- قيم الدلالات العلمية.
- قيم الخصائص القياسية للاختبار.

4- أدوات البحث: لتحقيق أهداف البحث والإجابة عن تساؤلاته استُخدمت الأدوات الآتية:

أ- اختبار المصفوفات المتتابعة المعياري رافن - النسخة الموازية: SPM-PV:

- **وصف الاختبار:**

يتألف هذا الاختبار من كراسة أسئلة، وورقة إجابة، تتكون كراسة الأسئلة من مجموعات من الأشكال المتعاقبة يرتبط كل منها بعلاقة ما، وتُقدّم مع خلية واحدة فارغة، ويتعين على المفحوص أن يختار من بين الأبدال المُعطاة له، ذلك البديل أو الشكل الذي يتسق مع الأشكال الأخرى، بعد اكتشاف العلاقة القائمة بينها لملء الخلية الفارغة (مخائيل،2006، ص.491). يتكون الاختبار بصورته الكلية من (60) بنداً اختبارياً أو مصفوفة، موزعة على خمس مجموعات فرعية يتضمن كل منها (12) بنداً، مرتبةً وفق مبدأ التدرج المتصاعد في الصعوبة ضمن كل مجموعة منها، وتتضمن كل مصفوفة إحدى المهمات العقلية الآتية:

- إكمال تصميم هندسي واحد أو مساحة ناقصة.
- إكمال عدة تصميمات هندسية متشابهة أو متماثلة.
- التغيُّر المنتظم في التصميمات الهندسية.
- إعادة ترتيب التصميم الهندسي أو تغييره بطريقة منتظمة.
- تحليل التصميمات الهندسية إلى أجزاء على نحوٍ منتظم، وإدراك العلاقات بينها(مراد وسليمان،2002، ص ص. 269 - 270).

- **دراسة الخصائص القياسية للاختبار:**

دُرست الخصائص القياسية لهذا الاختبار من قبل الباحثة يوسف(2015) في بحثٍ منشور، في العدد الثالث من مجلة جامعة البعث، على عينة عشوائية مؤلفة من (365) طالباً وطالبةً من طلبة المرحلة الجامعية في جامعة دمشق، واستخرجت دلالات الصدق المحكي للمقياس بدلالة اختبار الذكاء Toni-4، وقد أسفرت النتائج عن معامل ارتباط بلغت قيمته (0.48)، وهو دال إحصائياً عند مستوى الدلالة (0.01)، ودُرِس الصدق البنيوي للاختبار باستخدام التحليل العاملي بطريقة المكونات الأساسية، وقد أسفرت نتائجه عن وجود عامل واحد تشبعت عليه جميع المجموعات، وهو العامل غير اللفظي الذي استوعب 71.14% من التباين الكلي للمصفوفات، كما دُرِس ثبات الاختبار، وبلغت قيمة معامل ثبات ألفا كرونباخ (0.97)، وبلغت قيمة معامل التجزئة النصفية للاختبار(0.798) ، وبلغت قيمة معامل الثبات بطريقة الإعادة (0.922)، وهي بالمجمل معاملات جيدة، ومتقاربة مع النتائج الأساسية في دليل الاختبار، مما يعطي الثقة بصلاحية استخدامه في البيئة السورية (يوسف، 2015، ص ص.11- 31).

ب- اختبار الذكاء غير اللفظي Toni -4 الصورة (A):

- **وصف الاختبار:**

يستخدم اختبار (Toni -4) الصيغة غير اللفظية ليقيس الذكاء العام، ويتكون هذا الاختبار من (60) بنداً مرتباً تدريجياً من حيث الصعوبة، تتضمن كل صفحة من صفحاته سؤالاً واحداً، على شكل مجموعة من الأشكال أو الرسوم بينها مربع أو مجموعة من المربعات الفارغة، وعلى المفحوص أن يقوم باختيار الشكل المناسب من مجموعة أبدال الإجابة الموجودة أسفل كل سؤال.

وتتضمّن العلاقات المجردة في فقرات هذا الاختبار واحداً أو أكثر من الملامح الآتية: الشكل، الموقع، الاتجاه، الدوران، الاستمرارية، التظليل، والحجم والحركة، وتحتوي الفقرات الأكثر صعوبة العديد من هذه المواصفات، بينما تحتوى الفقرات الأسهل واحدة أو اثنين منها فقط، ومن الجدير بالذكر أن كلا اختباري SPM-PV و4- Toni يستندان إلى نظرية العاملين لسبيرمان Spearman .

- **دراسة الخصائص القياسية للاختبار:**

قامت الباحثة موسى (2017) بتقنين هذا الاختبار على عينة من العاديين والصم في محافظة دمشق، للفئة العمرية من 13-18 سنة، في بحثها المُعد لنيل لدرجة الماجستير في القياس والتقويم.

وفي البحث الحالي دُرِست خصائصه القياسية على عينة من طلبة مرحلة التعليم الجامعي، بهدف استخدامه محكاً لاختبار الذكاء SPM-PV.

إذ قامت الباحثة بدراسة الخصائص القياسية لهذا الاختبار على عينة مؤلفة من (320) طالباً وطالبةً من طلبة المرحلة الجامعية في كلية التربية – جامعة دمشق، والجدول الآتي يوضح الخصائص الديموغرافية لهذه العينة، واختصاصاتها الدراسية.

جدول (10)

الخصائص الديموغرافية لعينة Toni-4

القسم	السنة الدراسية	العدد	النسبة المئوية	متوسط العمر	الانحراف المعياري
الإرشاد النفسي	الرابعة	52	%16.3	22.54	2.30
التربية الخاصة	الأولى	84	%26.3	20.1	2.49
رياض الأطفال	الثالثة	39	%12.2	23.23	3.29
علم النفس	الثانية	95	%29.7	21.25	4.10
معلم الصف	الثالثة	50	%15.6	20.76	1.87
المجموع		320	%100.1	21.32	3.22

جرى التحقق من ثبات الاختبار من خلال حساب معامل ثبات ألفا كرونباخ، وبينت النتائج أن قيمة هذا المعامل بلغت (0.82)، كذلك جرى التحقق من ثبات التجزئة النصفية للاختبار، وقد بينت النتائج أن قيمة معامل سبيرمان براون بين جزأي الاختبار بلغت (0.86) وهي معاملات جيدة.

كما جرى التحقق من الصدق المحكي لاختبار Toni -4 بدلالة محك اختبار الذكاء SPM-PV، وبينت النتائج أن قيمة معامل ارتباط بيرسون بين درجات أفراد العينة على الاختبارين بلغت (0.613)** وهو معامل مقبول.

ودُرِس الصدق التمييزي للاختبار، من خلال مقارنة الفرق بين متوسطي درجات الفئتين العليا والدنيا (الثلث الأعلى والأدنى)، باستخدام الاختبار الإحصائي T-Student لعينتين مستقلتين، والجدول الآتي يوضح نتائج هذا الاختبار.

جدول (11)

فرق متوسطي الفئتين على اختبار TONI-4

نوع الفئة	العدد	المتوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة T-Student	القيمة الاحتمالية P-Value	القرار	إيتا تربيع
الفئة العليا	107	45.25	3.14	212	26.41	0.000	دال	0.77
الفئة الدنيا	107	33.67	3.26					

يتبين من الجدول السابق أن قيمة الاختبار الإحصائي T- Student بلغت (26.41)، وأن القيمة الاحتمالية لها بلغت (0.000) وهي أصغر من مستوى الدلالة (0.05)؛ ما يدل على وجود فرق ذي دلالة إحصائية بين متوسطي الفئتين العليا والدنيا من أفراد العينة لصالح الفئة العليا، كما بلغت قيمة حجم الأثر لهذا الفرق (0.77)، وهي تدل على حجم أثر مرتفع.

كما جرى التحقق من دلالة الفرق بين متوسطي درجات هاتين الفئتين على الاختبار المحكي SPM-PV، باستخدام الاختبار الإحصائي T-Student لعينتين مستقلتين، والجدول الآتي يوضح نتائج هذا الاختبار.

جدول (12)

فرق متوسطي الفئتين على الاختبار المحكّي

نوع الفئة	العدد	المتوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة T-Student	القيمة الاحتمالية P-Value	القرار	إيتا تربيع
الفئة العليا	107	45.91	4.41	212	13.85	0.000	دال	0.47
الفئة الدنيا	107	37.67	4.29					

يتبين من الجدول السابق أن قيمة الاختبار الإحصائي T- Student بلغت (13.85)، وأن القيمة الاحتمالية لها بلغت (0.000) وهي أصغر من مستوى الدلالة (0.05)؛ ما يدل على وجود فرق ذي دلالة إحصائية بين متوسطي الفئتين العليا والدنيا لصالح الفئة العليا، كما بلغت قيمة حجم الأثر لهذا الفرق (0.47)، وهي تدل على حجم أثر مرتفع، ما يشير إلى أن اختبار الذكاء Toni-4 يتمتع بقدرة تمييزية بين الفئتين العليا والدنيا.

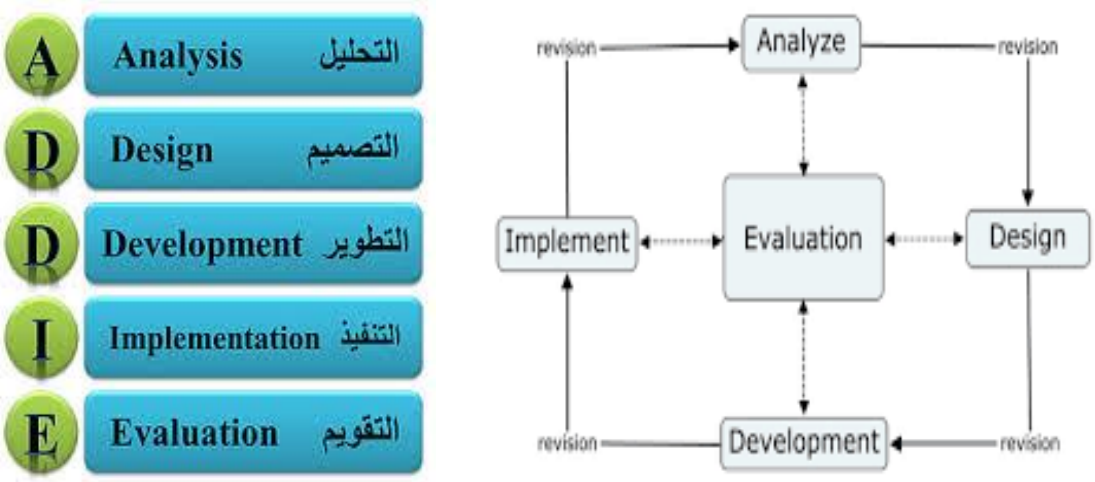
تُبين نتائج دراسة الخصائص القياسية لاختبار الذكاء Toni-4 أنه يتمتع بمعاملات صدق وثبات جيدة بالمجمل؛ ما يعطي الثقة بصلاحية تطبيقه على طلبة المرحلة الجامعية في البيئة السورية.

ج- البرنامج التدريبي:

بهدف دراسة أثر متغير التدريب في نتائج الدلالات العلمية والخصائص القياسية، جرى بناء برنامج تدريبي حول إستراتيجيات الأداء في الاختبار، وفق مراحل النموذج (ADDIE) في تصميم البرامج، وهو نموذجٌ عام لبناء وتصميم كافة أنواع البرامج التعليمية، وأثبت فاعليته في تصميم البرامج التعليمية والتدريبية، ويتضمن هذا النموذج خمس مراحل أساسية يوضحها الشكل الآتي.

شكل (6)

النموذج العام لتصميم التعليم



http://www.nwlink.com/~donclark/history_isd/addie.html

وفيما يلي شرحاً مفصلاً لخطوات بناء البرنامج التدريبي على إستراتيجيات الأداء في الاختبار في ضوء هذا النموذج:

أولاً: مرحلة التحليل: تضمنت هذه المرحلة عملية تحديد النقاط الآتية:

- **تحديد الاحتياج من البرنامج التدريبي:** من خلال تحديد المهارات والمعارف التي يهدف البرنامج لتنميتها وتطويرها لدى المتدربين، حيث يهدف هذا البرنامج إلى تنمية مهارات المتدربين وخبراتهم في التعامل مع الموقف الاختباري بشكل صحيح، في ظل نقص التدريب على الاختبارات والتعامل مع المواقف الاختبارية محلياً، وزيادة الاهتمام بهذا الجانب على المستوى العالمي.
- **تحديد الفئة المستهدفة بالتدريب:** حيث جرى استهداف فئة طلبة مرحلة التعليم الجامعي في البحث الحالي، على الرغم من أن الاحتياج للتعامل الصحيح مع الموقف الاختباري احتياجٌ عام لدى جميع المتعلمين وفي جميع المراحل التعليمية، لكنّ برامجاً تدريبياً واحداً غير قادر على تغطية كافة شرائح المتعلمين، لذا جرى تضمين مقترحات البحث إضافة التدريب على استراتيجيات الاختبار كجزء مكمل للمنهج الدراسي في المراحل التعليمية المختلفة.
- **تحديد نوع البرنامج المتناسب مع الاحتياج وأهدافه العامة:** وهو برنامج تدريبي، وتحليل الأساس النظري والمنطقي لهذا البرنامج: وذلك بالرجوع إلى الأساس النظري والأدبيات والأبحاث ذات الصلة، للاستفادة من محتوى هذه الأدبيات والأبحاث في تعريف المفهوم العام لإستراتيجيات أداء الاختبار، والتوصيف الدقيق لمحاوره الفرعية المتمّثلة بالإستراتيجيات الفرعية لأداء الاختبار، وقد جرى التركيز في هذا البحث على الإستراتيجيات التي يُطبقها المتعلمون في أثناء الموقف الاختباري، علماً أن مصطلح إستراتيجيات أداء الاختبار يتضمن العديد من الإستراتيجيات التي يتم توظيفها قبل وأثناء، وبعد إجراء الاختبار، ولكن تمّ الاقتصار على الإستراتيجيات في أثناء إجرائه؛ لأن الهدف من البرنامج الحالي هو التدريب على اختبار الذكاء، وهذا النوع من الاختبارات لا يحتاج لإعدادٍ أو تحضيرٍ مسبق لإجرائه كما هو الأمر في الاختبارات التحصيلية.
- **أما فيما يخص الأساس الفلسفي للبرنامج، والنظرية التربوية الداعمة له:** فقد استندت الباحثة في بناء جلسات البرنامج التدريبي، وأنشطته، وإجراءات تنفيذه إلى مبادئ النظرية المعرفية في التعلُّم Cognitive Learning Theory التي تهتم بالعمليات الذهنية لدى المتعلم، والمعالجات والتدخلات المستمرة في موضوع التعلم، بهدف تنظيمه وإدماجه في بيئة التعلم المعرفية.

ومن المسلمات الأساسية للنموذج المعرفي وفقاً لهذه النظرية أن هدف عملية التعلُّم هو انتقال أثر الخبرات والتجارب والدراسات إلى مجالات الحياة الواقعية، من خلال نقلها من بيئة التعلم إلى مجالات حياتية مختلفة، وقد جرى التأكيد في جميع جلسات البرنامج التدريبي أن الغاية من عملية التدريب لا تقتصر على زيادة معدلات درجات المتدربين الاختبارية في هذا البرنامج، وإنما الاحتفاظ بأبدال وطرائق حل المشكلات المُتضمنة في المصفوفات المختلفة، والاستفادة منها في حل مشكلات أخرى مماثلة فيما بعد.

ومن الجدير بالذكر أن إستراتيجيات أداء الاختبار التي جرى التدريب عليها خلال جلسات البرنامج التدريبي تتسم بخاصية انتقال أثر التعلُّم، ويمكن للفرد الاستفادة منها وتوظيفها في المواقف المختلفة من حياته: كالاستخدام الفعال للوقت، والقدرة على التخطيط، والعمل بسرعة ودقة، ما يجعل لها أهمية كبيرة في العملية الاختبارية من ناحية، وفي خصائص شخصية الفرد وقدرته على التعامل مع ظروف الحياة المختلفة من ناحيةٍ أخرى (Dodeen, 2014, p.1).

ومن مهارات التعلم المعرفي وفق هذه النظرية: حل المشكلات، وبناء المعنى، وفهم التمثيلات وفق صورها المختلفة، والتعاون (قطامي، 2013، ص. 44).

كما جرى تحديد مجالات أساسية لإستراتيجيات التدريس وفق النظرية المعرفية على النحو الآتي:

- الاكتشاف والاكتشاف الموجه.
- الملاحظة.
- المشاركة الموجهة.
- الشرح المباشر، وتقديم أمثلة ونماذج عملية.
- التدريس فرداً لفرد. والتي جرى تضمينها في جلسات البرنامج التدريبي.

- تتناسب هذه النظرية مع طبيعة التدريب الذي يستلزمه البرنامج التدريبي الحالي، فقد حدد جانييه (Gagne, 1985) تسعة أحداث أساسية للتعليم وفق إستراتيجية التعلم المعرفي من أجل دعم عمليات الانتباه الاختياري، والترميز والاسترجاع التي ينبغي توفرها من أجل حدوث التعلم لدى المتعلم، وهذه الأحداث هي:
  - جذب انتباه المتعلم.
  - إعلام المتعلم بالهدف.
  - إثارة الخبرات السابقة لدى المتعلم.
  - تقديم المعلومات الجديدة.
  - توجيه التعليم وأحداثه.
  - استدعاء أداءات المتعلم المعرفية.
  - تزويد المتعلم بالتغذية الراجعة المناسبة.
  - تقويم الأداء.
  - تعزيز انتقال أثر المعارف إلى مواقف وخبرات جديدة (المرجع السابق، ص ص. 51 –53).

وقد بُنيت جلسات البرنامج التدريبي في ضوء هذه الأحداث والخطوات الأساسية، مع التركيز على التطبيقات العملية عقب كل فقرة في أثناء سير الجلسات التدريبية.

ثانياً: مرحلة التصميم:

تعتمد هذه المرحلة على مخرجات عملية التحليل، وفيها قامت الباحثة بتحديد هيكلية المحتوى التدريبي وتقسيمه إلى جلسات، وقد تضمنت هذه المرحلة الأمور الآتية:

- تحديد الأهداف الفرعية المتصلة بالجلسات التدريبية وخطوات سيرها.
- تحديد عناصر المحتوى التدريبي.
- تحديد الأساليب التدريبية وطرائق تنفيذها.
- تحديد الوسائل وأدوات التدريب، والتصوُّر الأنسب لتوظيفها.
- تحديد الخطوط العريضة لعملية التقويم.
- تحديد تصور أولي للجدول الزمني.

وعلى الرغم من أن عمليات التطبيق المتعلقة بالبرنامج تأتي في مرحلة التنفيذ غير أن ضرورات البرنامج الحالي استدعت إجراء نوعين من التطبيق الاستطلاعي منفصلين عن التطبيق التجريبي والنهائي، حيث استندت الباحثة إلى نتائجهما في المرحلة التالية؛ وهي مرحلة التطوير وإنشاء محتوى الجلسات التدريبية، وفيما يلي عرضاً مفصلاً لهذين التطبيقين:

أ- تطبيقٌ استطلاعي لحساب معاملات سهولة المصفوفات:

وذلك بهدف تدريج هذه المصفوفات وفق معاملات سهولتها، لتكون عملية التدريب منظمة وفق منهجية متدرجة من السهل إلى الصعب، ومن البسيط إلى المعقد، والسبب في ذلك أن المصفوفات في الاختبار الأساسي متدرجة في سهولتها وفق كل مجموعة من مجموعاته الخمس، وبالتالي فإن المصفوفات الأخيرة من كل مجموعة أصعب من المصفوفات الأولى في بداية المجموعة التي تليها، لهذا الأمر كان لابد إعادة فرز وتصنيف هذه المصفوفات وفق معاملات سهولتها على مستوى الاختبار الكلي.

قامت الباحثة بإجراء هذا التطبيق خلال الفصل الأول من العام الدراسي 2020/2019م، على عينة من طلبة التعليم الجامعي بلغ عدد أفرادها (125) طالباً وطالبةً من الاختصاصات الأتية: (48) طالباً وطالبة سنة رابعة اختصاص معلم صف، (37) طالباً وطالبة سنة ثالثة اختصاص إرشاد نفسي، و(40) طالباً وطالبة سنة ثالثة اختصاص تربية خاصة، متوسط العمر لأفراد هذه العينة (21.37) سنة، والانحراف المعياري (1.20)، ووفقاً لهذا التطبيق جرى تدريج مصفوفات اختبار الذكاء SPM-PV تصاعدياً من الأسهل إلى الأصعب، والجدول الآتي يوضح هذه المعاملات.

جدول (13)

معاملات سهولة مصفوفات اختبار الذكاء

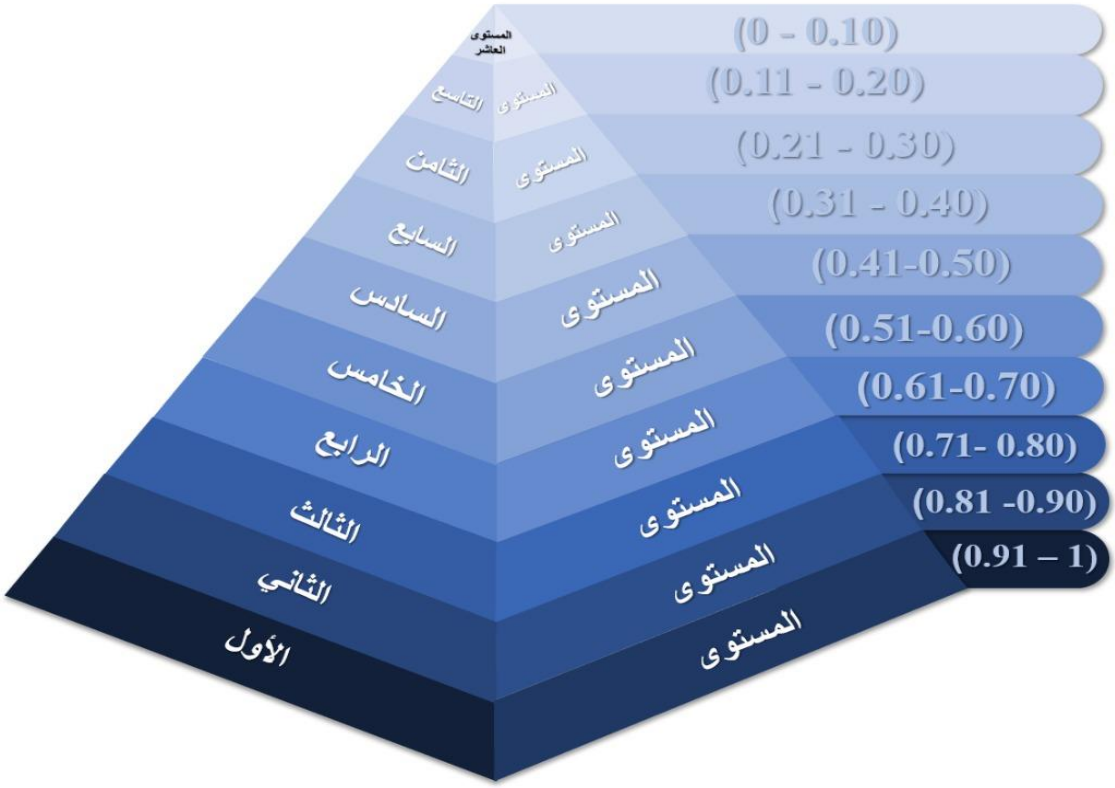
المجموعة (A)		المجموعة (B)		المجموعة (C)		المجموعة (D)		المجموعة (E)	
رقم المصفوفة	معامل السهولة	رقم المصفوفة	معامل السهولة	رقم المصفوفة	معامل السهولة	رقم المصفوفة	معامل السهولة	رقم المصفوفة	معامل السهولة
1	1	1	0.992	1	0.976	1	0.944	1	0.824
2	1	2	0.968	2	0.944	2	0.904	2	0.712
3	1	3	0.952	3	0.928	3	0.832	3	0.616
4	0.992	4	0.896	4	0.872	4	0.824	4	0.528
5	0.992	5	0.824	5	0.80	5	0.752	5	0.512
6	0.984	6	0.824	6	0.80	6	0.72	6	0.432
7	0.96	7	0.816	7	0.784	7	0.712	7	0.224
8	0.896	8	0.848	8	0.672	8	0.64	8	0.20
9	0.896	9	0.80	9	0.688	9	0.576	9	0.20
10	0.88	10	0.744	10	0.568	10	0.336	10	0.096
11	0.792	11	0.696	11	0.496	11	0.312	11	0.088
12	0.688	12	0.60	12	0.216	12	0.088	12	0.048

بعد ذلك قامت الباحثة بفرز وتصنيف هذه المصفوفات وفق هرم جرى تسميته إجرائياً بهرم مستويات معاملات السهولة، وهو الهرم الذي يُمثل تدرّجات مصفوفات هذا الاختبار وفق معاملات سهولتها ويوضحه الشكل رقم (7). إذ تتضمن المستويات الدنيا من هذا الهرم المصفوفات ذات معاملات السهولة الأكثر ارتفاعاً، ويبدأ المستوى الأول بالمصفوفات التي تنحصر معاملات سهولتها بين 0.90 والواحد الصحيح أي المصفوفات السهلة جداً، وتتدرج المصفوفات في تناقص معاملات سهولتها كلما تمّ التوجّه بالمستويات نحو الأعلى، وصولاً إلى المصفوفات شديدة الصعوبة، حيث يتضمن المستوى العاشر والأخير من هذا الهرم المصفوفات التي تنحصر معاملات سهولتها من الصفر وحتى (0.10).

وتجدر الإشارة إلى أن نوع العلاقات بين القواعد الناظمة للمصفوفات في المستويات المختلفة من هذا الهرم تراكمية؛ بمعنى أن القواعد الناظمة لمصفوفات كل مستوى من مستويات هذا الهرم تشمل جميع القواعد في المستويات الأدنى منها، ولا يمكن الانتقال لحل المصفوفات في المستويات العليا دون التمكُّن من حل جميع قواعد العلاقات في مصفوفات المستويات الأدنى، والشكل الآتي يوضح تدرجات هذا الهرم، والمدى الذي تنحصر فيه معاملات سهولة المصفوفات في كل مستوى من مستوياته:

شكل (7)

هرم مستويات معاملات السهولة



ويبين الجدول الآتي أرقام المصفوفات وعددها في كل مستوى من مستويات هذا الهرم.

جدول (14)

معاملات سهولة المصفوفات وفق المستويات

المستويات	معاملات السهولة	أرقام المصفوفات	عدد المصفوفات
العاشر	0 – 0.10	D12– E10 – E11 – E12	4
التاسع	0.11 – 0.20	E8 – E9	2
الثامن	0.21 – 0.30	C12 – E7	2
السابع	0.31 – 0.40	D10 – D11	2
السادس	0.41 – 0.50	C11– E6	2
الخامس	0.51 – 0.60	B12 – C10 – D9 – E4 – E5	5
الرابع	0.61 – 0.70	A12– B11– C8 – C9 – D8 – E3	6
الثالث	0.71 – 0.80	A11– B9 – B10 – C5 – C6 – C7 – D5 – D6 – D7– E2	10
الثاني	0.81 – 0.90	A8 – A9 – A10– B4 – B5 – B6 – B7 – B8 – C4 – D3 –D4 – E1	12
الأول	0.91 – 1	A1 – A2 – A3 – A4 – A5 – A6 – A7 – B1 – B2 – B3 – C1 – C2 – C3 – D1 - D2	15

كما يوضح الملحق (8) جداول معاملات السهولة لمصفوفات كل مجموعة من مجموعات الاختبار وفق التطبيق الاستطلاعي.

ب- **تطبيقٌ استطلاعي لتوزيع زمن الإجابة عن الاختبار:**

قامت الباحثة بإجراء تطبيق استطلاعي فردي، على عينة مؤلفة من (60) فرداً ممن تجاوزوا مرحلة التعليم الثانوي، واستخدام المؤقت لتعيير الزمن الذي يستغرقه كل فرد منهم في حل كل مصفوفة من مصفوفات الاختبار، متوسط أعمار هذه العينة (24.8) سنة وانحرافها المعياري (5.38)، وذلك بهدف وضع تصوّر للحد الأقصى من الزمن اللازم للانتهاء من الإجابة عن مصفوفات كل مستوى من مستويات هرم معاملات السهولة، علماً أن دليل الاختبار يشير إلى أن الزمن الكلي المخصص للإجابة عن هذا الاختبار ينحصر بين 40 و60 دقيقة، وهو بالمتوسط 50 دقيقة، والجدول الآتي يوضح ذلك.

جدول (15)

متوسط زمن الإجابة عن المصفوفات

المستويات	الزمن الكلي للإجابة بالثانية	متوسط زمن الإجابة بالثانية	الانحراف المعياري	عدد المصفوفات	متوسط زمن الإجابة عن كل مصفوفة مقدراً بالثانية
الأول	7211	120.18	29.02	15	ثواني≈ 8 8.01
الثاني	11537	192.28	52.16	12	ثانية≈ 16 16.02
الثالث	11992	199.87	37.36	10	ثانية≈ 20 19.99
الرابع	12469	207.82	17.23	6	ثانية≈ 35 34.64
الخامس	12633	210.55	57.88	5	ثانية≈ 42 42.11
السادس	10166	169.43	30.86	2	ثانية≈ 85 84.72
السابع	14480	241.33	47.70	2	ثانية≈ 121 120.67
الثامن	18026	300.43	54.74	2	ثانية≈ 150 150.22
التاسع	24040	400.66	72.09	2	ثانية≈ 200 200.33
العاشر	57643	960.72	146.10	4	ثانية≈ 240 240.18

مع التنويه في أثناء عملية التدريب وخلال سير الجلسات بأن هذا التوقيت هو توقيتٌ نسبي، يختلف من شخصٍ لآخر حسب قدراته العقلية، ومستوى ذكائه، وتعرضه المُسبق لبنود اختبارية مشابهة، ولكن جرى الاستعانة به في البرنامج التدريبي الحالي ليكون بمثابة ضابط ومؤشر يمكن الاعتماد عليه في تنظيم وإدارة وقت الاختبار، وذلك من خلال تدريب أفراد المجموعة التجريبية على الانتباه أكثر لزمن الإجابة عن مصفوفات الاختبار في كل مستوى من المستويات، ومحاولة عدم تجاوز هذا الزمن في أثناء الحل قدر الإمكان.

ثالثاً: مرحلة التطوير:

وتتضمن هذه المرحلة الخطوات الآتية:

- كتابة محتوى جلسات البرنامج التدريبي.
- تصميم الأنشطة التدريبية الخاصة بكل جلسة.
- تصميم أدوات التدريب وشرائح العرض الخاصة بكل جلسة تدريبية.
- تصميم مصفوفات التقويم المرحلي والنهائي الخاصة بكل جلسة.
- تجميع العناصر السابقة في إطارٍ منهجي ووضع الصورة الأولية للبرنامج.

رابعاً: مرحلة التنفيذ: سيجري شرحها بالتفصيل في فقرة إجراءات المنهج التجريبي.

خامساً: مرحلة التقويم: وهو بمثابة تقويم مرحلي يتضمن عمليتي التحكيم العلمي، وإجراء التجربة الاستطلاعية على النحو الآتي:

أ- التحقق من صدق محتوى البرنامج التدريبي ومصفوفات التقويم المرحلي والنهائي، من خلال التحكيم العلمي:

عُرِض البرنامج المُعد بصورته الأولية على مجموعة من السادة المحكمين الواردة أسماؤهم واختصاصاتهم العلمية في الملحق رقم (1)، بهدف تقويم جلساته التدريبية، وأدواته المتمثلة في مصفوفات التقويم المرحلي والتقويم النهائي الخاصة بكل جلسة، وإبداء آرائهم في سلامة فقرات هذا البرنامج من حيث الدقة العلمية، والصياغة اللغوية، وغيرها، وعُدِّل البرنامج وفق الآراء التي اتفقت بنسبة 30% على الأقل، وكانت أهم نقاط التعديل على النحو الآتي:

- توحيد المصطلحات العلمية وتوضيح الفرق بين مصطلحي الإستراتيجية والمهارة.
- توحيد المصطلحات المتصلة بأسئلة وبنود وفقرات الاختبار، والمصطلحات المتصلة بزمن / وقت الجلسات.
- إغناء التطبيقات العملية للجلسات التدريبية.
- تعديل مُسمّى إستراتيجية الضبط والسيطرة إلى إستراتيجية الضبط الذاتي.
- إضافة تدريبات عملية على مهارات الاسترخاء التي تتضمنها إستراتيجية الضبط الذاتي، مثل التدريب على مهارة تنظيم وضبط عملية التنفس.
- التركيز على استنتاج المتدربين القواعد قبل شرح المدربة لها في الجلسة التدريبية.
- ضبط وقت الجلسات بدقة أكبر من خلال التجريب الاستطلاعي، واختصار زمن الجلسات.

ب- **إجراء التجربة الاستطلاعية للبرنامج وأدواته:**

بعد تعديل البرنامج التدريبي في ضوء آراء السادة المحكمين، قامت الباحثة بإجراء تطبيق تجريبي للبرنامج التدريبي بكامل جلساته، على عينة استطلاعية بلغ عدد أفرادها ثمانية أفراد، ممن هم في مرحلة التعليم الجامعي أو المعهد المتوسط الذي يلي شهادة الثانوية العامة، أي الذين تجاوزوا الثامنة عشر من العمر، ولديهم الرغبة بحضور هذا النوع من البرامج تدريبية. وتتحدد الأهداف الأساسية لإجراء التجربة الاستطلاعية بالنقاط الآتية:

- تعرُّف الصعوبات التي قد تعترض سير التجربة النهائية لتلافيها.
- تعرُّف الأخطاء المُحتملة، والنقاط غير الواضحة في الأدوات لتعديلها.
- التأكد من وضوح المصفوفات في الشرائح ومناسبة أوراق العمل وأدوات التقويم المرحلي.
- ضبط الزمن اللازم لسير الجلسات وتطبيق الأدوات.
- التأكد من مناسبة الأنشطة والأساليب التدريبية وقابليتها للتطبيق.

والجدول الآتي يوضح تواريخ سير هذه الجلسات ورقم وعنوان كل جلسة منها.

جدول (16)

جلسات البرنامج في التجربة الاستطلاعية

رقم الجلسة	عنوانها	تاريخ التطبيق
الأولى	الجلسة التمهيدية: الترحيب والتعارف	الأحد 28/ 6/ 2020م
الثانية	القواعد الناظمة للعلاقات في مصفوفات المستوى الأول من هرم معاملات السهولة، وإستراتيجية أداء الاختبار المتصلة بإدارة وقت الاختبار.	الاثنين 29/ 6/ 2020م
الثالثة	القواعد الناظمة للعلاقات في مصفوفات المستوى الثاني من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بتحليل التعليمات والأسئلة الاختبارية، وإدارة وقت الاختبار.	الثلاثاء 6/30 / 2020م
الرابعة	القواعد الناظمة للعلاقات في مصفوفات المستوى الثالث من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالاستدلال الاستنباطي، وإدارة وقت الاختبار.	الأربعاء 1/ 7/ 2020م

الخامسة	القواعد الناظمة للعلاقات في مصفوفات المستوى الرابع من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالتعامل مع ورقة الإجابة، وإدارة وقت الاختبار.	الأحد 5 / 7/ 2020م
السادسة	القواعد الناظمة للعلاقات في مصفوفات المستوى الخامس من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالضبط الذاتي، وإدارة وقت الاختبار.	الاثنين 6 / 7/ 2020م
السابعة	القواعد الناظمة للعلاقات في مصفوفات المستويات السادس والسابع والثامن من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالضبط الذاتي وإدارة وقت الاختبار.	الثلاثاء 7 / 7/ 2020م
الثامنة	القواعد الناظمة للعلاقات في مصفوفات المستويين التاسع والعاشر من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالمراجعة وإدارة وقت الاختبار.	الأربعاء 8/ 7/2020م

وقد تمثّلت أهم النقاط التي جرى تعديلها في ضوء التجربة الاستطلاعية بالأمور الآتية:

- ضبط زمن الجلسات.
- توضيح طباعة المصفوفات في عرض الشرائح.
- تعديل صياغة بعض أسئلة التقويم المرحلي.
- إضافة جلسة تدريبية حول التخمين الذكي "المبني على المعلومات" واختصار جلسة من جلسات الضبط الذاتي لتصبح الجلسة السادسة في التطبيق النهائي عن "التخمين الذكي" والجلسة السابعة عن "الضبط الذاتي"، بالإضافة للجزء المتصل بالتدريبي على مصفوفات اختبار الذكاء.

5- إجراءات البحث: توزعت إجراءات البحث وفق ثلاثة محاور هي:

أ- إجراءات المنهج الوصفي التحليلي:

للإجابة عن السؤال الأول المتصل بتحليل واقع استخدام الدلالات العلمية في الأبحاث التجريبية، جرى حساب النسب المئوية لهذه الأبحاث من مجمل الأبحاث التي نوقشت خلال عامي 2000-2020 الميلاديين في أقسام عينة البحث الأربعة ، وذلك لتتمكن الباحثة من حصرها، ودراسة واقع الدلالات العلمية المستخدمة في حساب وتفسير نتائجها، والجدول والشكل البياني الآتيان يوضحان هذه النسب.

جدول (17)

عدد الأبحاث ونسبها المئوية

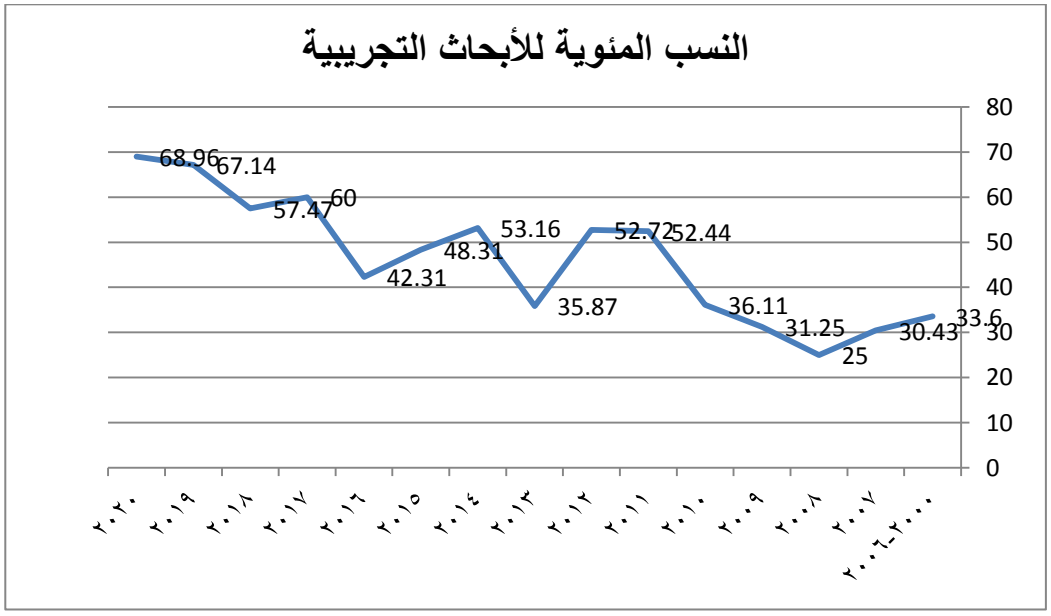
القسم	الأبحاث التجريبية		الأبحاث الوصفية التحليلية		المجموع
	العدد	النسبة المئوية	العدد	النسبة المئوية	
المناهج وطرائق التدريس	319	63.55	183	36.45	502
علم النفس	45	18.44	199	81.56	244
الإرشاد النفسي	41	25.15	122	74.85	163
التربية الخاصة	129	56.58	99	43.42	228
المجموع	534	46.97	603	53.03	1137

يتبين من الجدول السابق: أن النسب المئوية للأبحاث التجريبية في الأقسام الأربعة عينة البحث انحصرت بين (18.44%) في أبحاث قسم علم النفس، و(63.55%) في أبحاث قسم المناهج وطرائق التدريس، وقد بلغت النسبة المئوية الكلية لمجموع الأبحاث التجريبية في الأقسام الأربعة (46.96%).

ومن الجدير بالذكر أن النسب المئوية لهذا النوع من الأبحاث في تنامٍ مستمر، فحسب الملحق رقم (3) الذي يوضح بالتفصيل النسب المئوية للأبحاث التجريبية في كل عام، ولكل قسم من الأقسام الأربعة، يتبين أن نسبة هذه الأبحاث في الفترة الممتدة من عام 2000 وحتى 2006م بلغت (33.6%)، وتزايدت هذه النسبة مع التقدم بالأعوام حتى بلغت (68.96%) في عام 2020م.

شكل (8)

خط بياني للنسب المئوية للأبحاث التجريبية



بعد ذلك جرى الرجوع إلى هذه الأبحاث التجريبية في الأقسام السابقة، وتفريغ البيانات الضرورية منها في جداول تحليل تم إعدادها لهذا الغرض، تتضمن هذه الجداول: عام الرسالة، نوعها (ماجستير / دكتوراه)، منهجها (تجريبي/ وصفي تحليلي)، القسم والاختصاص، الفرضيات المُتضمنة فيها، نوع الاختبارات الإحصائية المُستخدمة فيها، نتيجة هذه الاختبارات والقرارات الإحصائية الناتجة عنها، كذلك استخدام حجم الأثر، ونوع القانون المستخدم لحسابه، وقيمة حجم الأثر الناتجة وتفسيرها، كما تتضمن حقلاً خاصاً لرصد ما إذا جرى استخدام أحد قوانين الدلالة العيادية في حساب نتائجها أم لا، والملحق رقم (2) يتضمن نموذجاً لجدول التحليل المُعد لتنظيم عملية استخراج هذه المعلومات.

ب- إجراءات منهج التحليل البعدي:

للإجابة عن كلٍ من أسئلة البحث الثاني والثالث والرابع جرى اتباع خطوات منهج التحليل البعدي؛ وهي على النحو الآتي:

- تحديد الموضوع بشكل واضح دقيق: وهو دراسة مدى كفاية الدلالتين الإحصائية والعملية في الإبلاغ عن نتائج الأبحاث التجريبية.
- اختيار الأبحاث المرتبطة بشكلٍ مباشر بالموضوع: وقد تم اختيار هذه الأبحاث في ضوء المحددين الآتيين:

- أن تكون الأبحاث التي يجري اختيارها تجريبية؛ أي قائمة على برنامج معين أياً كان نوعه سواءٌ أكان: تدريبياً أم إرشادياً أم تأهيلياً أم علاجياً أم تعليمياً.
- أن يجري حساب حجم الأثر للفرق بين درجات الاختبارين القبلي والبعدي في نتائج هذه الأبحاث، وذلك حتى تسمح بإمكانية إجراء التحليل البعدي فيما يتعلق بمقارنة نتائج الدلالتين الإحصائية والعملية لهذه الأبحاث من جهة، ومقارنة نتائج هاتين الدلالتين مع نتائج تصنيفات الدلالة العيادية التي ستقوم الباحثة بحسابها من جهةٍ أخرى.

وقد جرى جمع (80) بحثاً تجريبياً إما من الباحثين أنفسهم أو من خلال ملاحق الأبحاث التجريبية، تضمنت هذه الأبحاث (748) فرضية، قامت الباحثة بحساب قيم وتصنيفات الدلالة العيادية ل(196) فرضية منها، وهي تلك الفرضيات التي اختصت بدراسة الفروق بين المتوسطات القبلية والبعدية للمجموعات الضابطة والتجريبية.

ثم جرى حساب التكرارات والنسب المئوية للأفراد في كل تصنيف عيادي، وتفريغ هذه النسب وكذلك تفريغ قيم الدلالتين الإحصائية والعملية في جداول يوضحها الملحقان رقم (6) و(7)، وذلك لتنظيم المعلومات وتسهيل عملية مقارنة نتائج الدلالات العلمية للأبحاث المجمعة باستخدام منهج التحليل البعدي.

ج- إجراءات المنهج التجريبي:

تمثلت هذه الإجراءات بمرحلة التطبيق النهائي – أي مرحلة التنفيذ- للبرنامج التدريبي، التي جرى سيرها وفق الخطوات الآتية:

- الاستعداد لتطبيق التجربة النهائية من خلال التأكُّد من البيئة المادية والتقنية في مخبر الدكتور نعيم الرفاعي – وهو المخبر الذي أخذت الباحثة الموافقة لتجري فيه الجلسات التدريبية النهائية- إذ جرى التأكُّد من جهاز العرض والتوصيلات الكهربائية، والتنسيق مع الأساتذة والدكاترة المدرسين لطلبة الزمر الأربع للمجموعة التجريبية، وتحديد موعد إجراء الاختبارات: القبلية، البعدية المباشرة، والبعدية المؤجلة.
- البدء بتنفيذ تجربة البحث النهائية من خلال تطبيق جلسات البرنامج التدريبي على أفراد المجموعة التجريبية، ولأن عدد أفراد هذه المجموعة كبير ويتضمن 4 زمر، قامت الباحثة بتوزيع أفرادها إلى مجموعتين تجريبيتين تتضمن المجموعة التجريبية الأولى طلبة الزمرتين الأولى والثانية، وتتضمن المجموعة التجريبية الثانية طلبة الزمرتين الثالثة والرابعة، مع العلم أن المجموعتين التجريبيتين تعرضتا لجلسات البرنامج التدريبي ذاتها، وبنفس التسلسل دون أي تغيير، والغاية من هذا التوزيع هو فقط سهولة عملية التدريب، وضمان توفير بيئة تدريبية مريحة تسمح بإجراء أنشطة البرنامج التدريبي.
- بدأ تنفيذ أولى الجلسات التدريبية، وهي جلسة التمهيد والتعارف يوم الثلاثاء بتاريخ 2020/11/17م لأفراد المجموعة التجريبية الأولى، ويوم الخميس بتاريخ 19 /11 /2020م لأفراد المجموعة التجريبية الثانية.
- وبعد الانتهاء من تطبيق جلسات البرنامج التدريبي فوراً قامت الباحثة بإجراء التطبيق البعدي المباشر لاختباري الذكاء SPM-PV وToni-4 على أفراد هذه العينة بتاريخ 2020/12/17م.
- ومع بداية الفصل الدراسي الثاني قامت الباحثة بإجراء التطبيق البعدي المؤجل لاختباري الذكاء SPM-PV و Toni-4، بتاريخ الخميس 2021/6/10 على أفراد المجموعة التجريبية الأولى، والخميس 2021/6/17 على أفراد المجموعة التجريبية الثانية، وبلغت الفترة الزمنية الفاصلة بين الاختبارين البعدي المباشر والبعدي المؤجل ما يقارب ستة أشهر، والجدول الآتي يوضح الخطة الزمنية لسير الجلسات التدريبية بالتفصيل.

جدول (18)

جلسات البرنامج في التجربة النهائية

الأسبوع	المجموعة	الجلسة	التاريخ	مدة الجلسة	عنوان الجلسة
الأسبوع الأول	التجريبية الأولى	-	الاثنين 2020/11/16	كل جلسة 50 دقيقة تقريباً	التطبيق القبلي لاختباريSPM-PV و Toni-4
	التجريبية الثانية	-	الاثنين 2020/11/16	كل جلسة 50 دقيقة تقريباً	التطبيق القبلي لاختباريSPM-PV و Toni-4
	التجريبية الأولى	الأولى والثانية	الثلاثاء 2020/11/17	الأولى 45/ د والثانية 60/د	جلسة التعارف + القواعد الناظمة للعلاقات في مصفوفات المستوى الأول من هرم معاملات السهولة، وإستراتيجية أداء الاختبار المتصلة بإدارة وقت الاختبار.
	التجريبية الثانية	الأولى "عطل فني"	الخميس 2020/11/19	45/ د	جلسة التعارف
الأسبوع الثاني	التجريبية الأولى	الثالثة	الثلاثاء 2020/11/24	75/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الثاني من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بتحليل التعليمات والأسئلة الاختبارية، وإدارة وقت الاختبار.
	التجريبية الثانية	الثانية تعويضية	الأربعاء 2020/11/25	60/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الأول من هرم معاملات السهولة، وإستراتيجية أداء الاختبار المتصلة بإدارة وقت الاختبار.
	التجريبية الثانية	الثالثة	الخميس 2020/11/26	75/د	القواعد الناظمة للعلاقات في مصفوفات المستوى الثاني من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بتحليل التعليمات والأسئلة الاختبارية، وإدارة وقت الاختبار.
	التجريبية الثانية	الرابعة	الخميس 2020/11/26	75/د	القواعد الناظمة للعلاقات في مصفوفات المستوى الثالث من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالتعامل مع ورقة الإجابة، وإدارة وقت الاختبار.
	التجريبية الأولى	الرابعة	الخميس 2020/11/26	75/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الثالث من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالتعامل مع ورقة الإجابة، وإدارة وقت الاختبار.

الأسبوع الثالث	التجريبية الأولى	الخامسة	الثلاثاء 2020/12/1	75/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الرابع من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالاستدلال الاستنباطي، وإدارة وقت الاختبار.
	التجريبية الثانية	الخامسة	الخميس 2020/12/3	75/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الرابع من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالاستدلال الاستنباطي، وإدارة وقت الاختبار.
	التجريبية الثانية	السادسة	الخميس 2020/12/3	75/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الخامس من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالتخمين الذكي "المبني على المعلومات"، وإدارة وقت الاختبار.
	التجريبية الأولى	السادسة	الخميس 2020/12/3	75/ د	القواعد الناظمة للعلاقات في مصفوفات المستوى الخامس من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالتخمين الذكي "المبني على المعلومات"، وإدارة وقت الاختبار.
الأسبوع الرابع	التجريبية الأولى	السابعة	الثلاثاء 2020/12/8	80/ د	القواعد الناظمة للعلاقات في مصفوفات المستويات السادس والسابع والثامن من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالضبط الذاتي، وإدارة وقت الاختبار.
	التجريبية الثانية	السابعة	الاربعاء 2020/12/9	80/ د	القواعد الناظمة للعلاقات في مصفوفات المستويات السادس والسابع والثامن من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالضبط الذاتي، وإدارة وقت الاختبار.
	التجريبية الأولى	الثامنة	الخميس 2020/12/10	80/ د	القواعد الناظمة للعلاقات في مصفوفات المستويين التاسع والعاشر من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالمراجعة، وإدارة وقت الاختبار.
	التجريبية الثانية	الثامنة	الخميس 2020/12/10	80/ د	القواعد الناظمة للعلاقات في مصفوفات المستويين التاسع والعاشر من هرم معاملات السهولة، وإستراتيجيتي أداء الاختبار المتّصلتين بالمراجعة، وإدارة وقت الاختبار.
الأسبوع الخامس	التجريبية الأولى	-	الخميس 2020/12/17	كل جلسة 50 دقيقة تقريباً	التطبيق البعدي المباشر لاختباريSPM-PV و Toni-4
	التجريبية الثانية	-	الخميس 2020/12/17		التطبيق البعدي المباشر لاختباريSPM-PV و Toni-4
مجموع الجلسات	12 جلسة مع الاختبارات القبلية والبعدية				

وبعد الانتهاء من التطبيق أجرت الباحثة عمليات التصحيح، واستخراج الدرجات وفق سلمي تصحيح كل من اختباري SPM-PV و Toni-4، تمهيداً للمعالجات الإحصائية للنتائج، واستخراج قيم معاملات دلالاتها العلمية الثلاث الإحصائية والعملية والعيادية، والشكل الآتي يوضح التصميم التجريبي للبحث.

شكل (9)

التصميم التجريبي للبحث



6- المعالجات الإحصائية:

- اختبار T-Student للعينات المستقلة والمترابطة.
- الاختبار الإحصائي تحليل الانحدار البسيط والمتعدد.
- قوانين الدلالة العيادية (JT) ، (GLN)، (EN)، (HA)
- إحصاء كابا كوهين وهو مقياس للتوافق بين المتغيرات ذات الفئات المستقلة.
- الاختبار الإحصائي كاي تربيع x^2 لاختبار حسن المطابقة بين التصنيفات العيادية.
- الاختبار الإحصائيV الخاص بمعرفة دلالة الفروق بين عدة معاملات ارتباط بيرسون، يتبع توزيع كاي تربيع بدرجات حرية تساوي(عدد معاملات الارتباط−1)، ويحسب من المعادلة الآتية:

$$V = \sum_{1}^{K} (ni - 3)(Z_i - U)^2$$

وتحسب قيمة U من المعادلة الآتية:

$$U = \frac{[\sum_1^k (n_{i-3}) Z_i]}{\sum_1^k (n_{i-3})}$$

حيث n_i : عدد أفراد العينة في النموذج (i)

و z_i : قيمة فيشر المقابلة لمعامل الارتباط (Hays, 1977, P. 664)

- الاختبار الإحصائي M الخاص بمعرفة دلالة الفروق بين عدة معاملات ثبات ألفا كرونباخ، ويتبع توزيع كاي تربيع بدرجات حرية تساوي (عدد معاملات الثبات−1)، ويحسب بصورته العامة من المعادلة الآتية:

$$M = \sum_{K=1}^{K} A_K - \frac{\left[\sum_{K=1}^{K} A_K (1 - r_{ak})^{-1/3}\right]^2}{\sum_{k=1}^{k} A_K (1 - r_{ak})^{-2/3}}$$

حيث A_K:

$$A_K = \frac{(J_K - 1)(9n_k - 11)^2}{18J_K(n_k - 1)}$$

J : عدد فقرات الاختبار

K : رقم النموذج

r_{ak} : معامل الارتباط للنموذج k

n_k : عدد أفراد النموذج k (Hakstian & Whalen, 1976, PP. 222- 223)

خلاصة:

تضمن هذا الفصل توضيحاً لمجتمع البحث وعيناته، وأدواته، والمناهج البحثية المُستخدمة فيه، والتصميم التجريبي له، والإجراءات التي اتبعتها الباحثة للإجابة عن أسئلته، وصولاً إلى القوانين الإحصائية المستخدمة لتحليل النتائج ومعالجتها إحصائياً، ليتم فيما بعد عرضها بالتفصيل في الفصل الخامس.

الفصل الخامس

نتائج البحث وتحليلها

الفصل الخامس

نتائج البحث وتحليلها

تمهيد:

يتضمن الفصل الحالي عرضاً لنتائج البحث والإجابة عن أسئلته المتصلة بكل منهج من المناهج البحثية الثلاثة المُستخدمة فيه، مبتدئاً بعرض نتائج السؤال الأول المتصل بالمنهج الوصفي التحليلي، يليها نتائج الأسئلة المتصلة بمنهج التحليل البعدي، وأخيراً تلك المتصلة بمنهج البحث التجريبي، وفيما يلي عرضاً لهذه النتائج:

1- الإجابة عن السؤال الأول: ما واقع استخدام الدلالات العلمية في الأبحاث الأكاديمية التجريبية المنجزة في كلية التربية-جامعة دمشق؟

للإجابة عن هذا السؤال جرى حساب التكرارات والنسب المئوية لاستخدام اختبارات كل من الدلالتين العملية والعيادية، إلى جانب اختبارات الدلالة الإحصائية في نتائج الأبحاث التجريبية، في رسائل أقسام عينة البحث الأربعة ، والجدول الآتي يوضح هذه النسب.

جدول (19)

نسب الدلالة العملية إلى جانب الإحصائية

القسم	عدد الأبحاث التجريبية	العدد الكلي للاختبارات الإحصائية	عدد اختبارات الدلالة العملية المحسوبة لها	نسبتها المئوية
المناهج وطرائق التدريس	319	3683	1107	%30.06
علم النفس	45	542	158	%29.15
الإرشاد النفسي	41	392	108	%27.55
التربية الخاصة	129	642	177	%27.57
المجموع	534	5259	1550	%29.47

يتبين من الجدول السابق أن عدد الاختبارات الإحصائية المُستخدمة في الأبحاث التجريبية انحصر بين (392) اختباراً في رسائل قسم الإرشاد النفسي، و(3683) اختباراً في رسائل قسم المناهج وطرائق التدريس، وبلغ العدد الكلي لهذه الاختبارات في أبحاث الأقسام الأربعة (5259) اختباراً.

وأن النسب المئوية للاختبارات الإحصائية التي جرى التحقق من قيم دلالتها العملية انحصرت بين (27.55%) في قسم الإرشاد النفسي، و(30.06%) في قسم المناهج وطرائق التدريس، وبلغت النسبة المئوية الكلية لحساب قوانين حجم الأثر في نتائج هذه الأبحاث (29.47%)؛ مما يدل على أن أقل من ثلث الاختبارات الإحصائية المحسوبة في نتائج رسائل الأقسام الأربعة قد جرى التحقق من قيم دلالتها العملية.

كما بينت النتائج أن النسب المئوية لحساب قوانين الدلالة العملية في تزايد مستمر عبر الأعوام، حيث بلغت هذه النسبة من عام 2000إلى 2006 (4.98%)، في حين بلغت عام 2020 (57.79%)، والجدول والشكل البياني الآتيان يوضحان هذه النسب عبر الأعوام.

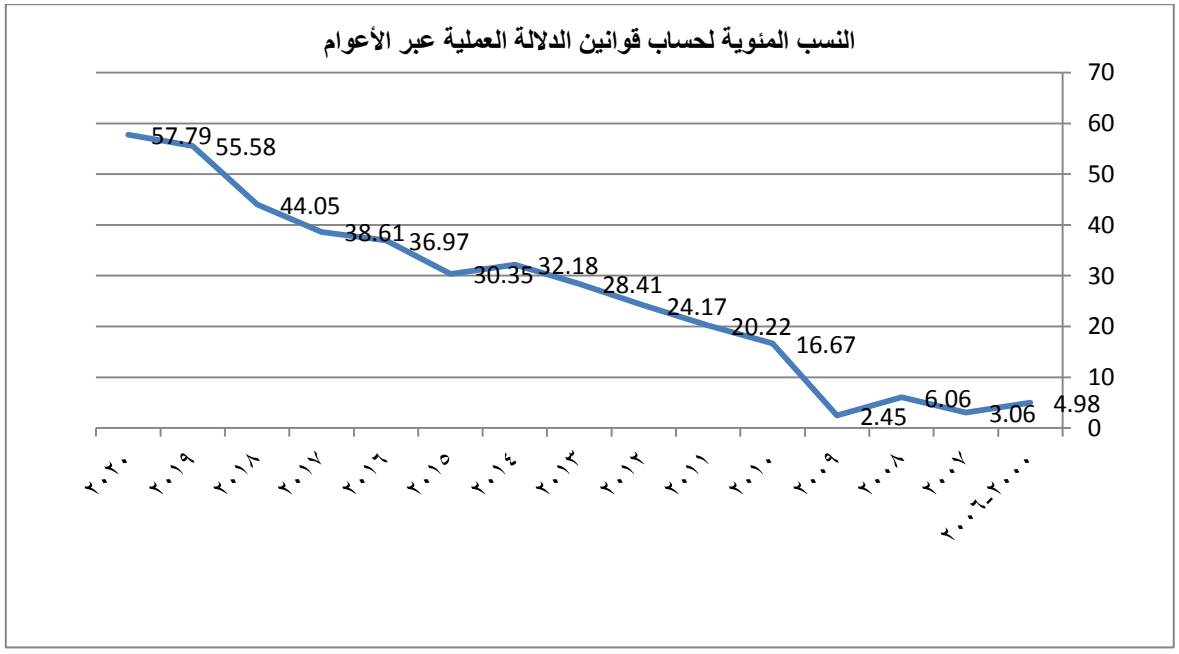
جدول (20)

نسب حجوم الأثر

العام	عدد الاختبارات الإحصائية في الأقسام الأربعة	عدد اختبارات الدلالة العملية	النسبة
2006-2000	402	20	%4.98
2007	98	3	%3.06
2008	99	6	%6.06
2009	204	5	%2.45
2010	306	51	%16.67
2011	539	109	%20.22
2012	695	168	%24.17
2013	433	123	%28.41
2014	348	112	%32.18
2015	346	105	%30.35
2016	284	105	%36.97
2017	303	117	%38.61
2018	420	185	%44.05
2019	493	274	%55.58
2020	289	167	%57.79
المجموع	5259	1550	%29.47

شكل (10)

خط بياني لنسب تزايد حجوم الأثر



وقد أظهرت نتائج حساب كاي تربيع x^2 لاختبار حسن المطابقة بين توزيعي العدد الكلي للاختبارات الإحصائية المحسوبة، وعدد الاختبارات التي استخدمت أحد قوانين حجم الأثر، وجود فرق ذي دلالة إحصائية بين هذين التوزيعين؛ مما يدل على عدم تناسب عدد الاختبارات التي جرى حساب حجم الأثر لها مع العدد الكلي للاختبارات المحسوبة. والجدول الآتي يوضح ذلك.

جدول (21)

كاي تربيع للدلالتين الإحصائية والعملية

قيمة معامل Chi- Square	درجات الحرية	Sig	معامل Cramer's v	تفسير المعامل
1593.00	90	0.00	1.00	حجم أثر مرتفع

وللتحقق من الدلالة العملية لهذه النتيجة حُسب معامل Cramer's v، وقد بينت النتائج أن قيمة هذا المعامل بلغت (1)، وهي تدل على حجم أثر مرتفع.

- أما فيما يتعلق بنسب استخدام الدلالة العيادية في حساب نتائج الأبحاث التجريبية، فقد أظهرت النتائج أنه لم يتم استخدام أي من قوانين الدلالة العيادية عند حساب نتائج هذه الأبحاث على الإطلاق.

2- الإجابة عن السؤال الثاني: ما مدى كفاية الدلالة الإحصائية في الإبلاغ عن نتائج تأثير البرامج التجريبية؟

للإجابة عن هذا السؤال جرى حساب النسب المئوية لكل نوع من تصنيفات حجوم الأثر: المرتفع، والمتوسط ، والمنخفض، التي تزامنت مع نوعي القرار الإحصائي: الدال وغير الدال، في نتائج الأبحاث التجريبية المُجمّعة التي تضمنها عينة التحليل البعدي، والجدول الآتي يوضح ذلك.

جدول (22)

تصنيفات حجوم الأثر والقرار الإحصائي

القرار الإحصائي	تصنيفات حجم الأثر					
	المرتفع		المتوسط		المنخفض	
	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
الدال	182	%98.9	1	%0.5	1	%0.5
غير الدال	1	%8.3	9	%75	2	%16.7

يتبين من الجدول السابق النقاط الآتية:

- تزامَن القرار الإحصائي الدال مع تصنيفات حجوم الأثر جميعها، ومع أن أعلى نسب تصنيفات حجم الأثر المترافقة مع هذا القرار كانت من نوع التصنيف المرتفع، إذ بلغت (98.9%)، إلا أن هذا القرار قد ترافق أيضاً مع حجم أثر متوسط، وكذلك منخفض بلغت نسبة كل منهما (0.5%).
- تزامَن القرار الإحصائي غير الدال – والذي ظهر فقط في المجموعات الضابطة– مع تصنيفات حجوم الأثر جميعها، وكانت أعلى نسب تصنيفات حجم الأثر المترافقة مع هذا القرار من نوع التصنيف المتوسط، إذ بلغت (75%)، كما ترافق أيضاً مع نسبة من حجوم الأثر المرتفعة بلغت قيمتها (8.3%).

كما جرى حساب متوسط النسب المئوية لكل نوع من تصنيفات الدلالة العيادية: الإتقان، والتحسُّن، والثبات، والتراجع، التي تزامنت مع نوعي القرار الإحصائي في نتائج الأبحاث التجريبية المُجمعة، والجدول الآتي يوضح ذلك.

جدول (23)

تصنيفات الدلالة العيادية والقرار الإحصائي

القرار الإحصائي	تصنيفات الدلالة العيادية			
	الإتقان "التعافي"	التحسُّن	الثبات	التراجع
الدال	70.67	5.93	23.35	0.05
غير الدال	2.34	1.28	94.04	2.34

يتبين من الجدول السابق النقاط الآتية:

- تزامن القرار الإحصائي الدال مع جميع تصنيفات الدلالة العيادية، ومع أن أعلى متوسطات نسب التصنيفات العيادية المتزامنة مع هذا القرار كانت من نوع تصنيف الإتقان العيادي؛ إذ بلغ متوسط نسبته (70.67)، إلا أن القرار الدال قد ترافق أيضاً مع تصنيفات الثبات والتراجع العيادي.
- تزامن القرار الإحصائي غير الدال مع جميع تصنيفات الدلالة العيادية، ومع أن أعلى متوسطات نسب التصنيفات العيادية المتزامنة مع هذا القرار كانت من نوع تصنيف الثبات؛ إذ بلغ متوسط نسبته (94.04)، غير أن هذا القرار قد ترافق أيضاً مع تصنيفات الإتقان والتحسن العيادية.

3- الإجابة عن السؤال الثالث: ما مدى كفاية الدلالة العملية في الإبلاغ عن نتائج تأثير البرامج التجريبية؟

للإجابة عن هذا السؤال جرى حساب متوسط النسب المئوية لكل نوع من التصنيفات العيادية المتزامنة مع كل تصنيف من تصنيفات حجوم الأثر، في نتائج الأبحاث المجمعة، والجدول الآتي يوضح هذه القيم.

جدول (24)

تصنيفات الدلالة العيادية وحجوم الأثر

تصنيفات حجم الأثر	تصنيفات الدلالة العيادية			
	الإتقان "التعافي"	**التحسُّن**	**الثبات**	**التراجع**
المرتفع	71	5.95	23	0.05
المتوسط	2.19	1.80	94.87	1.14
المنخفض	5.56	0	88.88	5.56

يتبين من الجدول السابق النقاط الآتية:

- تزامنت حجوم الأثر المرتفعة مع جميع تصنيفات الدلالة العيادية، ومع أن أعلى متوسطات نسب التصنيفات العيادية المتزامنة معها كانت من نوع تصنيف الإتقان العيادي؛ إذ بلغ متوسط نسبته (71)، إلا أنها قد ترافقت أيضاً مع تصنيف الثبات، الذي ظهر في كلا نوعي المجموعات التجريبية والضابطة، وأيضاً مع تصنيف التراجع الذي ظهر فقط في المجموعات الضابطة.
- تزامنت حجوم الأثر المتوسطة مع جميع تصنيفات الدلالة العيادية، وكانت أعلى متوسطات نسب التصنيفات العيادية المتزامنة معها من نوع تصنيف الثبات العيادي، إذ بلغ متوسط نسبته (94.87)، وهذا يدل على أن الأبحاث المجمعة في عينة التحليل البعدي التي تُقِر نتائجها بوجود حجم أثر متوسط للبرنامج على مستوى المجموعة، تزامنت بنسب مرتفعة من الحالات الفردية التي لم تحقق أي تغيُّر بالاتجاه المطلوب، يعادل متوسط هذه النسب (94.87%).
- تزامنت حجوم الأثر المنخفضة مع جميع تصنيفات الدلالة العيادية، باستثناء تصنيف التحسن العيادي، وكانت أعلى متوسطات نسب التصنيفات العيادية المتزامنة معها من نوع الثبات العيادي، إذ بلغ متوسط نسبة هذا التصنيف (88.88)، كما ترافقت أيضاً مع تصنيفي الإتقان والتراجع العيادي.

4- الإجابة عن السؤال الرابع: ما إمكانية التنبؤ بنتائج أي نوع من الدلالات العلمية من خلال نتائج النوعين الآخرين؟

للإجابة عن هذا السؤال جرى حساب الاختبار الإحصائي تحليل الانحدار البسيط لدراسة القدرة التنبؤية لقيم كل نوع من أنواع الدلالات العلمية بالنوعين الآخرين، والجداول الآتية توضح ذلك.

أ- القدرة التنبؤية لقيم الدلالة الإحصائية بقيم الدلالة العملية:

جدول (25)

تحليل الانحدار للدلالتين الإحصائية والعملية

	قيم معاملات الانحدار			T	Sig.	النسبة الفائيةF	Sig.	Rمعامل الارتباط	*معامل التحديد* R^2
	B	Std. Error	Beta						
(Constant)	0.715	0.02	.432	36.14	.000	44.55	.000	0.432	0.187
معاملات الدلالة الإحصائية	0.004	0.001		6.67	.000				

حجم الأثر :Dependent Variable

يتبين من الجدول السابق أن النسبة الفائية لتحليل الانحدار البسيط بلغت (44.55)، وهي دالة إحصائياً عند مستوى الدلالة (0.05)، وهذا يعني أن المتغير المستقل: الدلالة الإحصائية، دال إحصائياً في توقعه للمتغير التابع: الدلالة العملية، وقد بلغت قيمة معامل الارتباط بين مؤشراتهما (0.43)، وبلغت قيمة معامل التحديد (0.187)، وهي تدل على حجم أثر مرتفع للقدرة التنبؤية، ويمكن صياغة معادلة التنبؤ على النحو الآتي: $\gamma = 0.715 + 0.004\, x$

وهي تدل على أنه كلما زادت قيم الدلالة الإحصائية درجة واحدة، تحدث زيادة في الدلالة العملية بمقدار 0.004 درجة تقريباً.

كما جرى التحقق من القدرة التنبؤية لقيم الدلالة العملية بالدلالة الإحصائية، وقد بينت النتائج وجود قدرة تنبؤية ذات دلالة إحصائية وعملية أيضاً لقيمها بالدلالة الإحصائية.

ب- القدرة التنبؤية لقيم الدلالة الإحصائية بتصنيفات الدلالة العيادية:

جدول (26)

تحليل الانحدار للدلالتين الإحصائية والعيادية

Dependent Variable		قيم معاملات الانحدار			T	Sig.	النسبة الفائية F	Sig.	Rمعامل الارتباط	*معامل التحديد* R^2
		B	Std. Error	Beta						
تصنيف الإتقان الدال عيادياً	(Constant)	53.69	3.394	.386	15.82	.000	33.95	.000	0.386	0.149
	قيم الدلالة الإحصائية	0.66	0.113		5.83	.000				
تصنيف التحسن الدال عيادياً	(Constant)	6.62	1.148	0.094	5.77	.000	1.734	.189	0.094	0.009
	قيم الدلالة الإحصائية	0.05	0.038		1.32	.189				
تصنيف الثبات الدال عيادياً	(Constant)	39.37	3.148	-.381	12.51	.000	32.97	.000	-0.381	0.145
	قيم الدلالة الإحصائية	-0.60	0.105		-5.74	.000				
تصنيف التراجع الدال عياديا	(Constant)	0.315	0.122	-.110	2.58	.011	2.374	.125	-0.11	0.012
	قيم الدلالة الإحصائية	-0.006	0.004		-1.54	.125				

يتبين من الجدول السابق: وجود قدرة تنبؤية دالة إحصائياً لقيم الدلالة الإحصائية بقيم تصنيف الإتقان العيادي؛ إذ بلغت النسبة الفائية لتحليل تباين الانحدار البسيط (33.95)، وهي دالة إحصائياً عند مستوى الدلالة (0.05)، كما بلغت قيمة معامل الارتباط بينهما (0.386)، وبلغت قيمة معامل التحديد (0.149) وهي تدل على حجم أثر مرتفع للقدرة التنبؤية، ويمكن صياغة معادلة التنبؤ على النحو الآتي:

$$\gamma = 53.692 + 0.66\,x$$

كما يتبين وجود قدرة تنبؤية دالة إحصائياً لقيم الدلالة الإحصائية بقيم تصنيف الثبات العيادي؛ إذ بلغت النسبة الفائية لتحليل تباين الانحدار البسيط (32.97)، وهي دالة إحصائياً عند مستوى الدلالة (0.05)، كما بلغت قيمة معامل الارتباط بينهما (0.38−)، وبلغت قيمة معامل التحديد (0.145) وهي تدل على حجم أثر مرتفع للقدرة التنبؤية، ويمكن صياغة معادلة التنبؤ على النحو الآتي:

$$\gamma = 39.37 - 0.60x$$

في حين يتبين عدم وجود قدرة تنبؤية دالة إحصائياً لقيم الدلالة الإحصائية، بكلٍ من قيم تصنيفي التحسن والتراجع العياديين.

كما جرى دراسة القدرة التنبؤية لكل تصنيف من تصنيفات الدلالة العيادية بقيم الدلالة الإحصائية، وبينت النتائج وجود قدرة تنبؤية ذات دلالة إحصائية وعملية لكلٍ من تصنيفي الإتقان والثبات العيادي بقيم الدلالة الإحصائية، وعدم وجود قدرة تنبؤية ذات دلالة إحصائية وعملية لكلٍ من تصنيفي التحسن والتراجع العيادي بقيم الدلالة الإحصائية.

ج- القدرة التنبؤية لقيم الدلالة العملية بتصنيفات الدلالة العيادية:

جدول (27)

تحليل الانحدار للدلالتين العملية والعيادية

Dependent Variable	Predictor	قيم معاملات الانحدار			T	Sig.	النسبة الفائيةF	Sig.	Rمعامل الارتباط	معامل التحديد R^2
		B	Std. Error	Beta						
تصنيف الإتقان الدال عيادياً	(Constant)	75.74	23.19	.405	3.266	.001	38.09	.000	0.405	0.164
	قيم الدلالة العملية	48.73	7.89		6.172	.000				
تصنيف التحسن الدال عيادياً	(Constant)	4.55	7.92	.093	0.57	.566	1.68	.197	0.093	0.009
	قيم الدلالة العملية	3.49	2.69		1.295	.197				
تصنيف الثبات الدال عياديا	(Constant)	174.21	20.95	-.451	8.316	.000	49.53	.000	-0.451	0.203
	قيم الدلالة العملية	-50.21	7.13		-7.038	.000				
تصنيف التراجع الدال عياديا	(Constant)	6.07	0.73	-.501	8.273	.000	64.94	.000	-0.501	0.251
	قيم الدلالة العملية	-2.01	0.25		-8.058	.000				

يتبين من الجدول السابق: وجود قدرة تنبؤية دالة إحصائياً لقيم الدلالة العملية بقيم تصنيف الإتقان العيادي؛ إذ بلغت النسبة الفائية لتحليل تباين الانحدار البسيط (38.09)، وهي دالة إحصائياً عند مستوى الدلالة (0.05)، كما بلغت قيمة معامل الارتباط بينهما (0.405)، وبلغت قيمة معامل التحديد (0.164) وهي تدل على حجم أثر مرتفع للقدرة التنبؤية، ويمكن صياغة معادلة التنبؤ على النحو الآتي:

$$\boldsymbol{\gamma = 75.73 + 48.73\, x}$$

كما يتبين وجود قدرة تنبؤية دالة إحصائياً لقيم الدلالة العملية بقيم تصنيف الثبات العيادي؛ إذ بلغت النسبة الفائية لتحليل تباين الانحدار البسيط (49.53)، وهي دالة إحصائياً عند مستوى الدلالة (0.05)، كما بلغت قيمة معامل الارتباط بينهما (0.451−)، وبلغت قيمة معامل التحديد (0.203) وهي تدل على حجم أثر مرتفع للقدرة التنبؤية، ويمكن صياغة معادلة التنبؤ على النحو الآتي:

$$\boldsymbol{\gamma = 174.21 - 50.21\, x}$$

كذلك يتبين وجود قدرة تنبؤية دالة إحصائياً لقيم الدلالة العملية بقيم تصنيف التراجع العيادي؛ إذ بلغت النسبة الفائية لتحليل تباين الانحدار البسيط (64.94)، وهي دالة إحصائياً عند مستوى الدلالة (0.05)، كما بلغت قيمة معامل الارتباط بينهما (0.501−) وبلغت قيمة معامل التحديد (0.251) وهي تدل على حجم أثر مرتفع للقدرة التنبؤية، ويمكن صياغة معادلة التنبؤ على النحو الآتي:

$$\boldsymbol{\gamma = 6.06 - 2.01\, x}$$

أخيراً يتبين عدم وجود قدرة تنبؤية دالة إحصائياً وعملياً لقيم الدلالة العملية بقيم تصنيف التحسن العيادي.

كما جرى دراسة القدرة التنبؤية لكل تصنيف من تصنيفات الدلالة العيادية بقيم الدلالة العملية، وبينت النتائج وجود قدرة تنبؤية ذات دلالة إحصائية وعملية لكلٍ من تصنيفات: الإتقان والثبات والتراجع العيادي بقيم الدلالة العملية، وعدم وجود قدرة تنبؤية ذات دلالة إحصائية وعملية لتصنيف التحسن العيادي بقيم الدلالة العملية.

5- الإجابة عن السؤال الخامس: إلى أي مدى يُغيّر التدريب في مؤشرات الدلالات العلمية لنتائج أفراد العينة على اختبار الذكاء؟

للإجابة عن هذا السؤال تم حساب الفرق بين متوسطي درجات العينة التجريبية على اختبار الذكاء، في كلٍ من الاختبارين: القبلي والبعدي المباشر، القبلي والبعدي المؤجل، البعدي المباشر والبعدي المؤجل، باستخدام الدلالتين الإحصائية والعملية، علماً بأن قيم درجات الحرية في الحالات السابقة جميعها تساوي (124)، والجدول الآتي يوضح ذلك.

جدول (28)

الدلالتان الإحصائية والعملية لأنواع تطبيق الاختبار

نوع الاختبار	الإحصاء الوصفي		اختبار T-testللعينات المترابطة		قيم الدلالة العملية	
	المتوسط الحسابي	الانحراف المعياري	قيمة T المحسوبة	القيمة الاحتمالية P-Value	قيمة معامل d- Cohen	قيمة معامل إيتا تربيع
قبلي	41.85	5.32	33.41	0.00	1.98	0.90
بعدي مباشر	52.42	3.11				
قبلي	41.85	5.32	21.76	0.00	1.41	0.79
بعدي مؤجل	49.37	3.89				
بعدي مباشر	52.42	3.11	10.43	0.000	0.87	0.47
بعدي مؤجل	49.37	3.89				

يتبين من الجدول السابق وجود فرق ذي دلالة إحصائية وعملية بين متوسطي درجات المجموعة التجريبية في أنواع المقارنات الثلاثة، وهذه الفروق لصالح الاختبار البعدي المباشر في المقارنتين الأولى والثالثة أي بعد تعرُّض أفراد هذه المجموعة للتدريب على هذا المتغير مباشرةً، ولصالح الاختبار البعدي المؤجل مقارنةً بالقبلي في المقارنة الثانية، وأن جميع حجوم الأثر المحسوبة للفروق في المقارنات الثلاث كانت مرتفعة.

ولمعرفة مقدار واتجاه هذه التغيُّرات على مستوى كل فرد في العينة التجريبية، وفي كل نوع من المقارنات السابقة جرى حساب نقطة القطع وفق الطريقة (A) لكل نوع من المقارنات السابقة، وهي عبارة عن المتوسط القبلي مضافاً له انحرافين معياريين للدرجات القبلية، وبذلك تساوي هذه النقطة 52.49 في المقارنتين الأولى والثانية، وتساوي 58.64 في المقارنة الثالثة، كما جرى حساب مؤشر التغيُّر الموثوق لدرجات كل فرد وفق الصيغة الرياضية لطريقة (JT) لاستخراج التصنيف العيادي المقابل له، ثم استخراج النسب المئوية لعدد الأفراد في كل تصنيف من التصنيفات العيادية الأربعة وهي: الإتقان، والتحسن، والثبات، والتراجع، في كل نوع من المقارنات، والجدول الآتي يوضح ذلك.

جدول (29)

مؤشرات التغيُّر الموثوق لأنواع تطبيق الاختبار

رقم الفرد	الاختبار القبلي	الاختبار البعدي مباشر	الاختبار البعدي المؤجل	فبلي/بعدي مباشر		فبلي/بعدي مؤجل		بعدي مباشر/ مؤجل	
				مؤشر التغير	التصنيف العيادي	مؤشر التغير	التصنيف العيادي	مؤشر التغير	التصنيف العيادي
1	49	54	50	1.48	ثابت	0.29	ثابت	-1.75	ثابت
2	39	49	48	2.97	تحسن	2.61	تحسن	-0.44	ثابت
3	36	51	43	4.45	تحسن	2.03	تحسن	-3.49	تراجع

4	39	48	45	2.67	تحسن	1.74	ثابت	-1.31	ثابت
5	42	53	54	3.26	أتقن	3.48	أتقن	0.44	ثابت
6	41	51	50	2.97	تحسن	2.61	تحسن	-0.44	ثابت
7	36	47	46	3.26	تحسن	2.9	تحسن	-0.44	ثابت
8	47	57	50	2.97	أتقن	0.87	ثابت	-3.06	تراجع
9	37	51	50	4.15	تحسن	3.77	تحسن	-0.44	ثابت
10	44	55	51	3.26	أتقن	2.03	تحسن	-1.75	ثابت
11	34	51	49	5.04	تحسن	4.35	تحسن	-0.87	ثابت
12	35	51	50	4.75	تحسن	4.35	تحسن	-0.44	ثابت
13	45	54	51	2.67	أتقن	1.74	ثابت	-1.31	ثابت
14	47	56	52	2.67	أتقن	1.45	ثابت	-1.75	ثابت
15	48	57	56	2.67	أتقن	2.32	أتقن	-0.44	ثابت
16	39	52	51	3.86	تحسن	3.48	تحسن	-0.44	ثابت
17	54	58	57	1.19	ثابت	0.87	ثابت	-0.44	ثابت
18	39	52	48	3.86	تحسن	2.61	تحسن	-1.75	ثابت
19	28	45	40	5.04	تحسن	3.48	تحسن	-2.18	تراجع
20	51	54	55	0.89	ثابت	1.16	ثابت	0.44	ثابت
21	41	53	52	3.56	أتقن	3.19	تحسن	-0.44	ثابت
22	38	49	42	3.26	تحسن	1.16	ثابت	-3.06	تراجع
23	41	53	50	3.56	أتقن	2.61	تحسن	-1.31	ثابت
24	42	55	49	3.86	أتقن	2.03	تحسن	-2.62	تراجع
25	39	52	50	3.86	تحسن	3.19	تحسن	-0.87	ثابت
26	38	51	44	3.86	تحسن	1.74	ثابت	-3.06	تراجع
27	44	54	51	2.97	أتقن	2.03	تحسن	-1.31	ثابت
28	37	48	50	3.26	تحسن	3.77	تحسن	0.87	ثابت
29	44	56	51	3.56	أتقن	2.03	تحسن	-2.18	تراجع
30	38	51	41	3.86	تحسن	0.87	ثابت	-4.37	تراجع
31	42	52	46	2.97	تحسن	1.16	ثابت	-2.62	تراجع
32	51	56	55	1.48	ثابت	1.16	ثابت	-0.44	ثابت
33	44	53	50	2.67	أتقن	1.74	ثابت	-1.31	ثابت
34	40	53	51	3.86	أتقن	3.19	تحسن	-0.87	ثابت
35	46	47	44	0.3	ثابت	-0.58	ثابت	-1.31	ثابت
36	31	48	38	5.04	تحسن	2.03	تحسن	-4.37	تراجع

37	41	52	50	3.26	تحسن	2.61	تحسن	-0.87	ثابت
38	48	60	53	3.56	أتقن	1.45	ثابت	-3.06	تراجع
39	49	55	45	1.78	ثابت	-1.16	ثابت	-4.37	تراجع
40	45	57	49	3.56	أتقن	1.16	ثابت	-3.49	تراجع
41	39	52	44	3.86	تحسن	1.45	ثابت	-3.49	تراجع
42	33	51	40	5.34	تحسن	2.03	تحسن	-4.8	تراجع
43	44	56	55	3.56	أتقن	3.19	أتقن	-0.44	ثابت
44	39	53	45	4.15	أتقن	1.74	ثابت	-3.49	تراجع
45	43	52	50	2.67	تحسن	2.03	تحسن	-0.87	ثابت
46	38	50	46	3.56	تحسن	2.32	تحسن	-1.75	ثابت
47	39	51	49	3.56	تحسن	2.9	تحسن	-0.87	ثابت
48	42	46	48	1.19	ثابت	1.74	ثابت	0.87	ثابت
49	38	47	44	2.67	تحسن	1.74	ثابت	-1.31	ثابت
50	41	50	53	2.67	تحسن	3.48	أتقن	1.31	ثابت
51	47	57	49	2.97	أتقن	0.58	ثابت	-3.49	تراجع
52	44	55	52	3.26	أتقن	2.32	تحسن	-1.31	ثابت
53	37	50	52	3.86	تحسن	4.35	تحسن	0.87	ثابت
54	44	53	50	2.67	أتقن	1.74	ثابت	-1.31	ثابت
55	32	50	49	5.34	تحسن	4.93	تحسن	-0.44	ثابت
56	45	57	48	3.56	أتقن	0.87	ثابت	-3.93	تراجع
57	39	51	50	3.56	تحسن	3.19	تحسن	-0.44	ثابت
58	47	54	51	2.08	أتقن	1.16	ثابت	-1.31	ثابت
59	41	53	46	3.56	أتقن	1.45	ثابت	-3.06	تراجع
60	40	49	50	2.67	تحسن	2.9	تحسن	0.44	ثابت
61	50	57	56	2.08	أتقن	1.74	ثابت	-0.44	ثابت
62	42	53	54	3.26	أتقن	3.48	أتقن	0.44	ثابت
63	43	50	51	2.08	تحسن	2.32	تحسن	0.44	ثابت
64	45	54	52	2.67	أتقن	2.03	تحسن	-0.87	ثابت
65	48	54	54	1.78	ثابت	1.74	ثابت	0	ثابت
66	37	48	46	3.26	تحسن	2.61	تحسن	-0.87	ثابت
67	35	50	51	4.45	تحسن	4.64	تحسن	0.44	ثابت
68	34	46	49	3.56	تحسن	4.35	تحسن	1.31	ثابت
69	41	54	52	3.86	أتقن	3.19	تحسن	-0.87	ثابت

70	50	56	55	1.78	ثابت	1.45	ثابت	-0.44	ثابت
71	40	52	48	3.56	تحسن	2.32	تحسن	-1.75	ثابت
72	45	56	51	3.26	أتقن	1.74	ثابت	-2.18	تراجع
73	46	51	50	1.48	ثابت	1.16	ثابت	-0.44	ثابت
74	39	52	44	3.86	تحسن	1.45	ثابت	-3.49	تراجع
75	47	54	51	2.08	أتقن	1.16	ثابت	-1.31	ثابت
76	47	58	50	3.26	أتقن	0.87	ثابت	-3.49	تراجع
77	41	53	49	3.56	أتقن	2.32	تحسن	-1.75	ثابت
78	42	54	49	3.56	أتقن	2.03	تحسن	-2.18	تراجع
79	52	56	48	1.19	ثابت	-1.16	ثابت	-3.49	تراجع
80	40	53	51	3.86	أتقن	3.19	تحسن	-0.87	ثابت
81	44	55	48	3.26	أتقن	1.16	ثابت	-3.06	تراجع
82	46	54	52	2.37	أتقن	1.74	ثابت	-0.87	ثابت
83	53	58	56	1.48	ثابت	0.87	ثابت	-0.87	ثابت
84	40	50	53	2.97	تحسن	3.77	أتقن	1.31	ثابت
85	40	53	46	3.86	أتقن	1.74	ثابت	-3.06	تراجع
86	33	48	47	4.45	تحسن	4.06	تحسن	-0.44	ثابت
87	50	57	55	2.08	أتقن	1.45	ثابت	-0.87	ثابت
88	37	49	51	3.56	تحسن	4.06	تحسن	0.87	ثابت
89	39	50	53	3.26	تحسن	4.06	أتقن	1.31	ثابت
90	42	53	50	3.26	أتقن	2.32	تحسن	-1.31	ثابت
91	46	55	52	2.67	أتقن	1.74	ثابت	-1.31	ثابت
92	40	49	49	2.67	تحسن	2.61	تحسن	0	ثابت
93	32	46	41	4.15	تحسن	2.61	تحسن	-2.18	تراجع
94	46	53	52	2.08	أتقن	1.74	ثابت	-0.44	ثابت
95	42	55	51	3.86	أتقن	2.61	تحسن	-1.75	ثابت
96	40	52	48	3.56	تحسن	2.32	تحسن	-1.75	ثابت
97	35	50	41	4.45	تحسن	1.74	ثابت	-3.93	تراجع
98	37	50	44	3.86	تحسن	2.03	تحسن	-2.62	تراجع
99	44	53	51	2.67	أتقن	2.03	تحسن	-0.87	ثابت
100	46	54	52	2.37	أتقن	1.74	ثابت	-0.87	ثابت
101	55	59	58	1.19	ثابت	0.87	ثابت	-0.44	ثابت
102	42	52	49	2.97	تحسن	2.03	تحسن	-1.31	ثابت
103	45	55	54	2.97	أتقن	2.61	أتقن	-0.44	ثابت

104	52	57	56	1.48	ثابت	1.16	ثابت	-0.44	ثابت
105	31	50	44	5.64	تحسن	3.77	تحسن	-2.62	تراجع
106	44	55	48	3.26	أتقن	1.16	ثابت	-3.06	تراجع
107	41	48	49	2.08	تحسن	2.32	تحسن	0.44	ثابت
108	47	58	52	3.26	أتقن	1.45	ثابت	-2.62	تراجع
109	46	49	52	0.89	ثابت	1.74	ثابت	1.31	ثابت
110	36	49	47	3.86	تحسن	3.19	تحسن	-0.87	ثابت
111	43	51	48	2.37	تحسن	1.45	ثابت	-1.31	ثابت
112	43	52	50	2.67	تحسن	2.03	تحسن	-0.87	ثابت
113	53	55	54	0.59	ثابت	0.29	ثابت	-0.44	ثابت
114	40	52	52	3.56	تحسن	3.48	تحسن	0	ثابت
115	47	56	51	2.67	أتقن	1.16	ثابت	-2.18	تراجع
116	39	51	50	3.56	تحسن	3.19	تحسن	-0.44	ثابت
117	38	52	47	4.15	تحسن	2.61	تحسن	-2.18	تراجع
118	43	53	49	2.97	أتقن	1.74	ثابت	-1.75	ثابت
119	35	49	40	4.15	تحسن	1.45	ثابت	-3.93	تراجع
120	45	49	51	1.19	ثابت	1.74	ثابت	0.87	ثابت
121	35	52	43	5.04	تحسن	2.32	تحسن	-3.93	تراجع
122	33	49	48	4.75	تحسن	4.35	تحسن	-0.44	ثابت
123	43	54	51	3.26	أتقن	2.32	تحسن	-1.31	ثابت
124	39	50	47	3.26	تحسن	2.32	تحسن	-1.31	ثابت
125	39	52	50	3.86	تحسن	3.19	تحسن	-0.87	ثابت

يتبيّن من الجدول السابق أن: (49) فرداً من العينة التجريبية تعادل نسبتهم (39.2%) حققوا تصنيفات إتقان دال عيادياً في الاختبار البعدي المباشر مقارنةً بالاختبار القبلي، وأن (59) فرداً من أفراد هذه العينة تُعادل نسبتهم (47.2%) حققوا معدلات تحسن دال عيادياً نتيجة تعرضهم للبرنامج التدريبي، في حين أن (17) فرداً بقي أداؤهم ثابتاً في الاختبار البعدي المباشر مقارنةً بأدائهم في الاختبار القبلي، ونسبة هؤلاء الأفراد (13.6%).

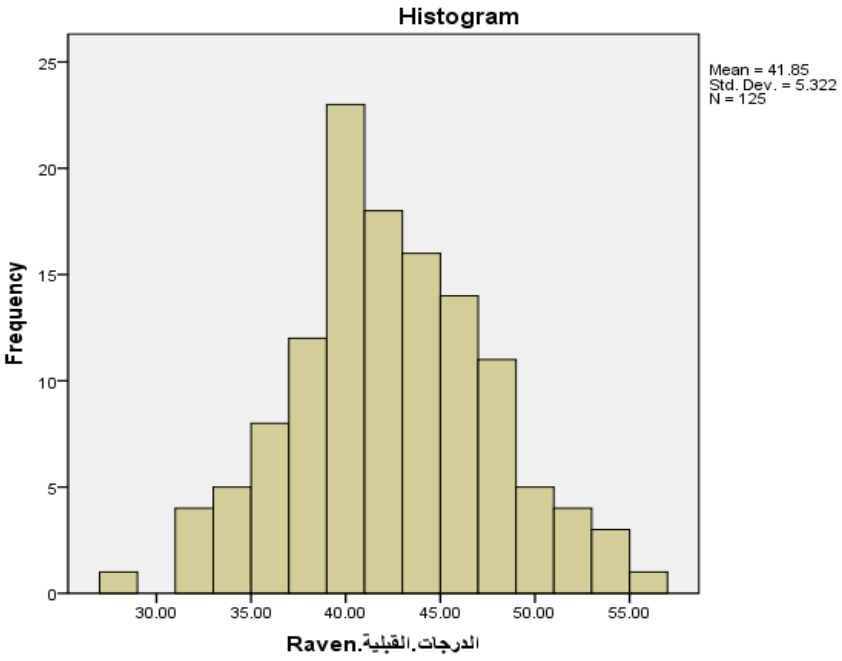
ويُلاحظ أن: (8) أفراد من العينة التجريبية تُعادل نسبتهم (6.4%) احتفظوا بتصنيفات إتقان دال عيادياً في الاختبار البعدي المؤجل مقارنةً بالاختبار القبلي، وأن (62) فرداً من أفراد هذه العينة تعادل نسبتهم (49.6%) احتفظوا بتصنيفات تحسن دالة عيادياً في الاختبار البعدي المؤجل، في حين أن (55) فرداً بقي أداؤهم ثابتاً في الاختبار البعدي المؤجل مقارنةً بأدائهم في الاختبار القبلي، ونسبة هؤلاء الأفراد (44%).

كما يتبيّن أن: (89) فرداً من العينة التجريبية تُعادل نسبتهم (71.2%) احتفظوا بتصنيفات ثبات دال عيادياً في الاختبار البعدي المؤجل مقارنةً بالبعدي المباشر، في حين أن (36) فرداً من هذه العينة تُعادل نسبتهم (28.8%) تراجع أداؤهم في الاختبار البعدي المؤجل تراجعاً دالاً عيادياً عنه في الاختبار

البعدي المباشر، والأشكال البيانية الآتية توضح توزع درجات أفراد المجموعة التجريبية في كل من الاختبار القبلي، والبعدي المباشر، والبعدي المؤجل.

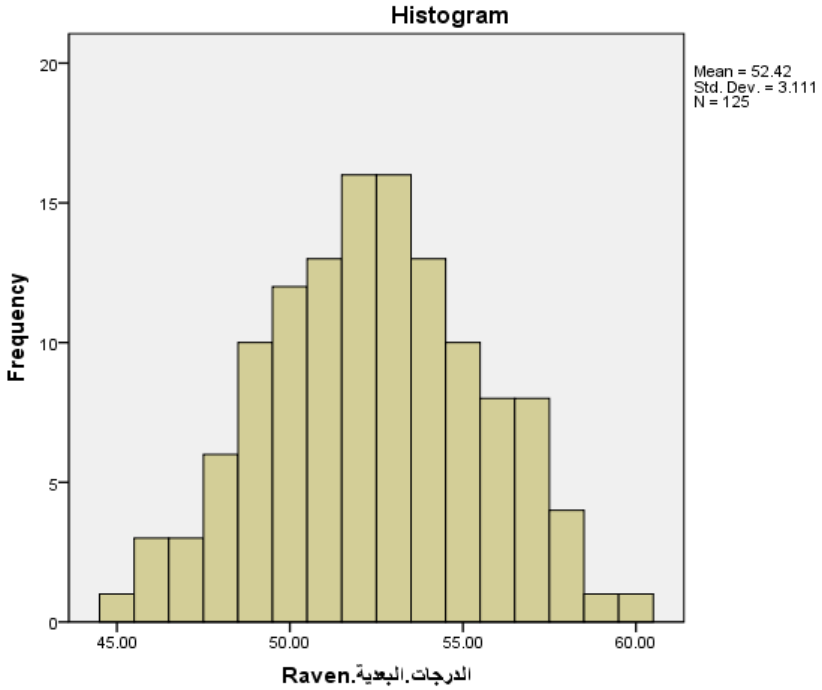
شكل (11)

تمثيل بياني لتوزع *الدرجات في الاختبار القبلي*



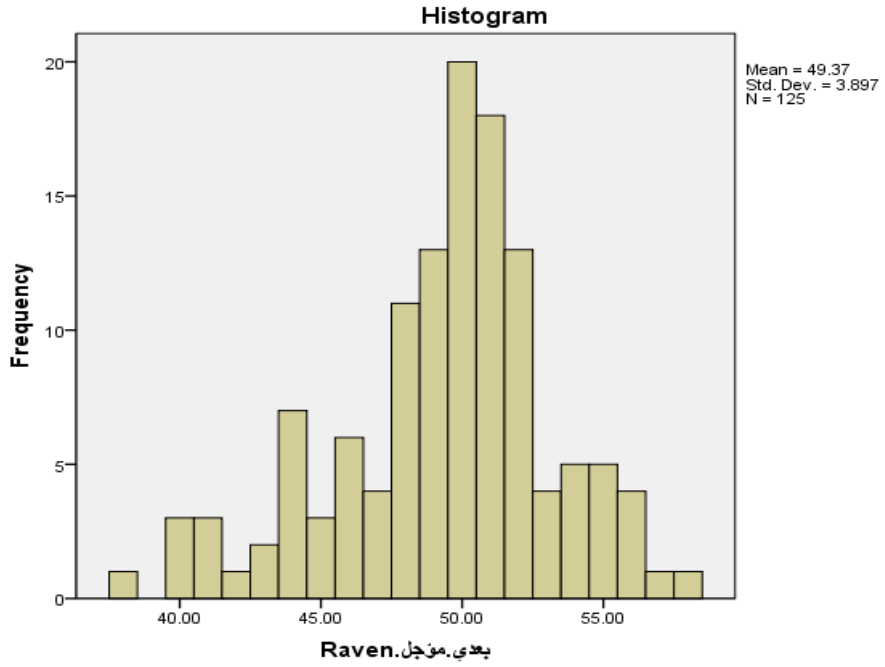
شكل (12)

تمثيل بياني لتوزع الدرجات في الاختبار البعدي المباشر



شكل (13)

تمثيل بياني لتوزع الدرجات في الاختبار البعدي المؤجل



ولتوضيح النتائج وتسهيل المقارنات بين نسب التصنيفات العيادية وفق المقارنات الثلاث، جرى تجميع هذه النسب وفق الجدول الآتي.

جدول (30)

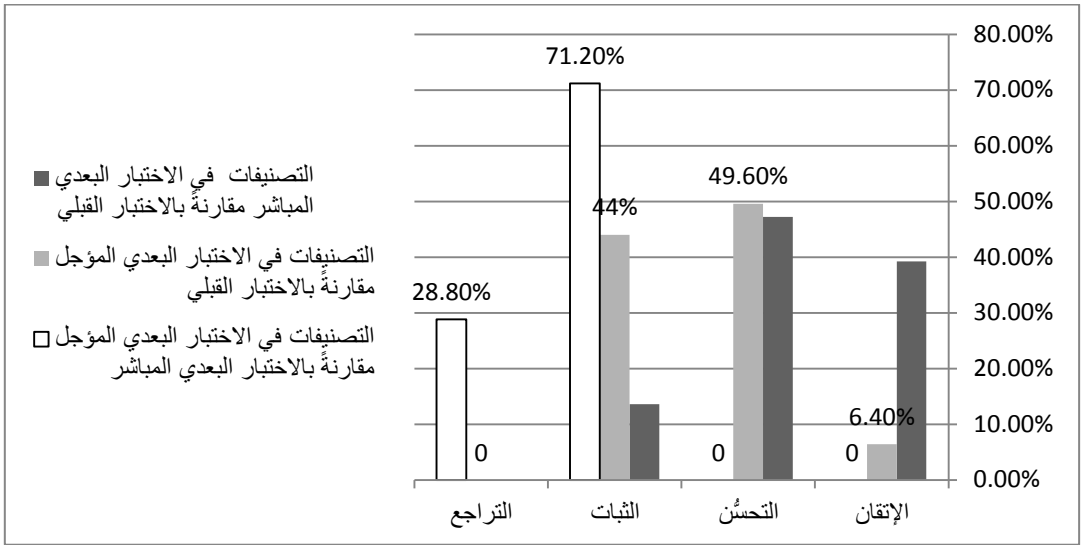
التصنيفات العيادية لمراحل التطبيق

نوع المقارنة	التصنيف العيادي			
	الإتقان	التحسُّن	الثبات	التراجع
في الاختبار البعدي المباشر مقارنةً بالاختبار القبلي	%39.2	%47.2	%13.6	-
في الاختبار البعدي المؤجل مقارنةً بالاختبار القبلي	%6.4	%49.6	%44	-
في الاختبار البعدي المؤجل مقارنةً بالاختبار البعدي المباشر	-	-	%71.2	%28.8

يتبين من الجدول السابق أن أعلى نسب تصنيفات الإتقان الدالة عيادياً كانت في التطبيق البعدي المباشر مقارنةً بالقبلي، وذلك بعد تعرُّض أفراد المجموعة التجريبية للبرنامج التدريبي على إستراتيجيات أداء الاختبار مباشرةً دون وجود فاصل زمني طويل، تراجع أداء المجموعة التجريبية في الاختبار البعدي المؤجل مقارنة بالاختبار القبلي لتصل نسبة الاتقان إلى 6.4% فقط، وتحوّلت معظم هذه التصنيفات إلى تحسُّن وثبات دال عيادياً. ويُلاحظ عدم وجود تصنيفات تحسن أو إتقان دال عيادياً في الاختبار البعدي المؤجل مقارنةً بالبعدي المباشر، والشكل البياني الآتي يوضح هذه النسب.

شكل (14)

تمثيل بياني للتصنيفات العيادية وفق مراحل تطبيق الاختبار



6- الإجابة عن السؤال السادس: إلى أي مدى تختلف قيم مؤشر التغيُّر الموثوق وتصنيفاته العيادية لدرجات أفراد العينة التجريبية على اختبار الذكاء باختلاف طريقة حسابه؟

للإجابة عن هذا السؤال تمّ حساب قيم مؤشرات التغيُّر الموثوق (RCI) والتصنيفات الناتجة عنها، لدرجات كل فرد من أفراد العينة التجريبية، على اختبار الذكاء في التطبيق البعدي المباشر مقارنةً بدرجته في التطبيق القبلي، وفق طرائق حساب الدلالة العيادية الآتية: (JT) و(GLN) و(EN) و(HA)، ولم يتم حساب مؤشرات التغيُّر الموثوق وفق الطريقة الخطية الهرمية (HLM) لعدم توافر البرامج اللازمة لحسابها، والجدول الآتي يوضح هذه القيم والتصنيفات.

جدول (31)

قيم الدلالة العيادية وفق طرائق حسابها

رقم الفرد	درجته في الاختبار القبلي	درجته في الاختبار البعدي	طريقة (JT)		طريقة (GLN)		طريقة (EN)		
			قيم RCI	التصنيف العيادي	قيم RCI	التصنيف العيادي	الحد الأعلى	الحد الأدنى	التصنيف العيادي
1	49	54	1.48	ثابت	2.02	أتقن	52.33	42.81	أتقن
2	39	49	2.97	تحسن	2.96	تحسن	44.33	34.81	تحسن
3	36	51	4.45	تحسن	4.34	تحسن	41.93	32.41	تحسن
4	39	48	2.67	تحسن	2.64	تحسن	44.33	34.81	تحسن
5	42	53	3.26	أتقن	3.46	اتقن	46.73	37.21	أتقن
6	41	51	2.97	تحسن	3.08	تحسن	45.93	36.41	تحسن
7	36	47	3.26	تحسن	3.08	تحسن	41.93	32.41	تحسن

8	47	57	2.97	أتقن	3.46	اتقن	50.73	41.21	أتقن
9	37	51	4.15	تحسن	4.09	تحسن	42.73	33.21	تحسن
10	44	55	3.26	أتقن	3.58	اتقن	48.33	38.81	أتقن
11	34	51	5.04	تحسن	4.84	تحسن	40.33	30.81	تحسن
12	35	51	4.75	تحسن	4.59	تحسن	41.13	31.61	تحسن
13	45	54	2.67	أتقن	3.02	اتقن	49.13	39.61	أتقن
14	47	56	2.67	أتقن	3.14	اتقن	50.73	41.21	أتقن
15	48	57	2.67	أتقن	3.21	اتقن	51.53	42.01	أتقن
16	39	52	3.86	تحسن	3.89	تحسن	44.33	34.81	تحسن
17	54	58	1.19	ثابت	2.02	اتقن	56.33	46.81	أتقن
18	39	52	3.86	تحسن	3.89	تحسن	44.33	34.81	تحسن
19	28	45	5.04	تحسن	4.46	تحسن	35.53	26.01	تحسن
20	51	54	0.89	ثابت	1.51	ثابت	53.93	44.41	أتقن
21	41	53	3.56	أتقن	3.71	اتقن	45.93	36.41	أتقن
22	38	49	3.26	تحسن	3.21	تحسن	43.53	34.01	تحسن
23	41	53	3.56	أتقن	3.71	اتقن	45.93	36.41	أتقن
24	42	55	3.86	أتقن	4.08	اتقن	46.73	37.21	أتقن
25	39	52	3.86	تحسن	3.89	تحسن	44.33	34.81	تحسن
26	38	51	3.86	تحسن	3.83	تحسن	43.53	34.01	تحسن
27	44	54	2.97	أتقن	3.23	اتقن	48.33	38.81	أتقن
28	37	48	3.26	تحسن	3.14	تحسن	42.37	33.21	تحسن
29	44	56	3.56	أتقن	3.89	اتقن	48.33	38.81	أتقن
30	38	51	3.86	تحسن	3.83	تحسن	43.53	34.01	تحسن
31	42	52	2.97	تحسن	3.14	تحسن	46.73	37.21	تحسن
32	51	56	1.48	ثابت	2.14	اتقن	53.93	44.41	أتقن
33	44	53	2.67	أتقن	2.96	اتقن	48.33	38.81	أتقن
34	40	53	3.86	أتقن	3.96	اتقن	45.13	35.61	أتقن
35	46	47	0.3	ثابت	0.57	ثابت	49.93	40.41	ثابت
36	31	48	5.04	تحسن	4.65	تحسن	37.93	28.41	تحسن
37	41	52	3.26	تحسن	3.39	تحسن	45.93	36.41	تحسن
38	48	60	3.56	أتقن	4.15	اتقن	51.53	42.01	أتقن
39	49	55	1.78	ثابت	2.33	اتقن	52.33	42.81	أتقن
40	45	57	3.56	أتقن	3.96	اتقن	49.13	39.61	أتقن

41	39	52	3.86	تحسن	3.89	تحسن	44.33	34.81	تحسن
42	33	51	5.34	تحسن	5.09	تحسن	39.53	30.01	تحسن
43	44	56	3.56	أتقن	3.89	اتقن	48.33	38.81	أتقن
44	39	53	4.15	أتقن	4.21	اتقن	44.33	34.81	أتقن
45	43	52	2.67	تحسن	2.89	تحسن	47.53	38.01	تحسن
46	38	50	3.56	تحسن	3.52	تحسن	43.53	34.01	تحسن
47	39	51	3.56	تحسن	3.58	تحسن	44.33	34.81	تحسن
48	42	46	1.19	ثابت	1.26	ثابت	46.73	37.21	ثابت
49	38	47	2.67	تحسن	2.58	تحسن	43.53	34.01	تحسن
50	41	50	2.67	تحسن	2.77	تحسن	45.93	36.41	تحسن
51	47	57	2.97	أتقن	3.46	اتقن	50.73	41.21	أتقن
52	44	55	3.26	أتقن	3.58	اتقن	48.33	38.81	أتقن
53	37	50	3.86	تحسن	3.77	تحسن	42.73	33.21	تحسن
54	44	53	2.67	أتقن	2.96	اتقن	48.33	38.81	أتقن
55	32	50	5.34	تحسن	5.02	تحسن	38.73	29.21	تحسن
56	45	57	3.56	أتقن	3.96	اتقن	49.13	39.61	أتقن
57	39	51	3.56	تحسن	3.58	تحسن	44.33	34.81	تحسن
58	47	54	2.08	أتقن	2.52	اتقن	50.73	41.21	أتقن
59	41	53	3.56	أتقن	3.71	اتقن	45.93	36.41	أتقن
60	40	49	2.67	تحسن	2.71	تحسن	45.13	35.61	تحسن
61	50	57	2.08	أتقن	2.71	اتقن	53.13	43.61	أتقن
62	42	53	3.26	أتقن	3.46	اتقن	46.73	37.21	أتقن
63	43	50	2.08	تحسن	2.27	تحسن	47.53	38.01	تحسن
64	45	54	2.67	أتقن	3.02	اتقن	49.13	39.61	أتقن
65	48	54	1.78	ثابت	2.27	اتقن	51.53	42.01	أتقن
66	37	48	3.26	تحسن	3.14	تحسن	42.73	33.21	تحسن
67	35	50	4.45	تحسن	4.27	تحسن	41.13	31.61	تحسن
68	34	46	3.56	تحسن	3.27	تحسن	40.33	30.81	تحسن
69	41	54	3.86	أتقن	4.02	اتقن	45.93	36.41	أتقن
70	50	56	1.78	ثابت	2.39	اتقن	53.13	43.61	أتقن
71	40	52	3.56	تحسن	3.65	تحسن	45.13	35.61	تحسن
72	45	56	3.26	أتقن	3.65	اتقن	49.13	39.61	أتقن
73	46	51	1.48	ثابت	1.83	ثابت	49.93	40.41	تحسن

74	39	52	3.86	تحسن	3.89	تحسن	44.33	34.81	تحسن
75	47	54	2.08	أتقن	2.52	اتقن	50.73	41.21	أتقن
76	47	58	3.26	أتقن	3.77	اتقن	50.73	41.21	أتقن
77	41	53	3.56	أتقن	3.71	اتقن	45.93	36.41	أتقن
78	42	54	3.56	أتقن	3.77	اتقن	46.73	37.21	أتقن
79	52	56	1.19	ثابت	1.89	ثابت	54.73	45.21	أتقن
80	40	53	3.86	أتقن	3.96	اتقن	45.13	35.61	أتقن
81	44	55	3.26	أتقن	3.58	اتقن	48.33	38.81	أتقن
82	46	54	2.37	أتقن	2.77	اتقن	49.93	40.41	أتقن
83	53	58	1.48	ثابت	2.27	اتقن	55.53	46.01	أتقن
84	40	50	2.97	تحسن	3.02	تحسن	45.13	35.61	تحسن
85	40	53	3.86	أتقن	3.96	اتقن	45.13	35.61	أتقن
86	33	48	4.45	تحسن	4.15	تحسن	39.53	30.01	تحسن
87	50	57	2.08	أتقن	2.71	اتقن	53.13	43.61	أتقن
88	37	49	3.56	تحسن	3.46	تحسن	42.73	33.21	تحسن
89	39	50	3.26	تحسن	3.27	تحسن	44.33	34.81	تحسن
90	42	53	3.26	أتقن	3.46	اتقن	46.73	37.21	أتقن
91	46	55	2.67	أتقن	3.08	اتقن	49.93	40.41	أتقن
92	40	49	2.67	تحسن	2.71	تحسن	45.13	35.61	تحسن
93	32	46	4.15	تحسن	3.77	تحسن	38.73	29.21	تحسن
94	46	53	2.08	أتقن	2.45	اتقن	49.93	40.41	اتقن
95	42	55	3.86	أتقن	4.08	اتقن	46.73	37.21	اتقن
96	40	52	3.56	تحسن	3.65	تحسن	45.13	35.61	تحسن
97	35	50	4.45	تحسن	4.27	تحسن	41.13	31.61	تحسن
98	37	50	3.86	تحسن	3.77	تحسن	42.73	33.21	تحسن
99	44	53	2.67	أتقن	2.96	اتقن	48.33	38.81	اتقن
100	46	54	2.37	أتقن	2.77	اتقن	49.93	40.41	اتقن
101	55	59	1.19	ثابت	2.08	اتقن	57.13	47.61	اتقن
102	42	52	2.97	تحسن	3.14	تحسن	46.73	37.21	تحسن
103	45	55	2.97	أتقن	3.33	اتقن	49.13	39.61	اتقن
104	52	57	1.48	ثابت	2.20	اتقن	54.73	45.21	اتقن
105	31	50	5.64	تحسن	5.28	تحسن	37.93	28.41	تحسن
106	44	55	3.26	أتقن	3.58	اتقن	48.33	38.81	اتقن
107	41	48	2.08	تحسن	2.14	تحسن	45.93	36.41	تحسن
108	47	58	3.26	أتقن	3.77	اتقن	50.73	41.21	اتقن
109	46	49	0.89	ثابت	1.2	ثابت	49.93	40.41	ثابت

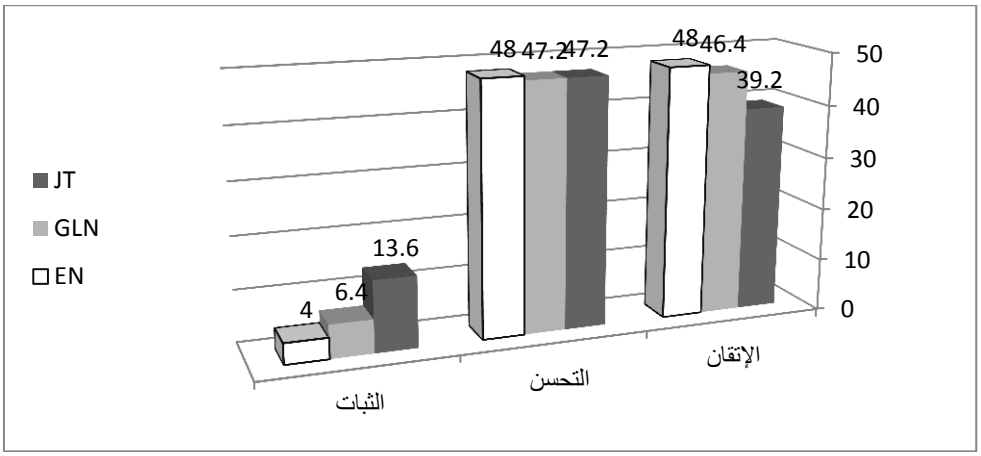
110	36	49	3.86	تحسن	3.71	تحسن	41.93	32.41	تحسن
111	43	51	2.37	تحسن	2.58	تحسن	47.53	38.01	تحسن
112	43	52	2.67	تحسن	2.89	تحسن	47.53	38.01	تحسن
113	53	55	0.59	ثابت	1.33	ثابت	55.53	46.01	ثابت
114	40	52	3.56	تحسن	3.65	تحسن	45.13	35.61	تحسن
115	47	56	2.67	أتقن	3.14	اتقن	50.73	41.21	اتقن
116	39	51	3.56	تحسن	3.58	تحسن	44.33	34.81	تحسن
117	38	52	4.15	تحسن	4.15	تحسن	43.53	34.01	تحسن
118	43	53	2.97	أتقن	3.21	اتقن	47.53	38.01	اتقن
119	35	49	4.15	تحسن	3.96	تحسن	41.13	31.61	تحسن
120	45	49	1.19	ثابت	1.45	ثابت	49.13	39.61	ثابت
121	35	52	5.04	تحسن	4.89	تحسن	41.13	31.61	تحسن
122	33	49	4.75	تحسن	4.46	تحسن	39.53	30.01	تحسن
123	43	54	3.26	أتقن	3.52	اتقن	47.53	38.01	اتقن
124	39	50	3.26	تحسن	3.27	تحسن	44.33	34.81	تحسن
125	39	52	3.86	تحسن	3.89	تحسن	44.33	34.81	تحسن

يتبيّن من الجدول السابق أن: نسبة تصنيفات الإتقان وفق طريقة (JT) بلغت (39.2%)، ونسبة تصنيفات التحسن بلغت (47.2%)، في حين أن نسبة تصنيفات الثبات الدال عيادياً بلغت (13.6%).

كما يتبين أن نسبة تصنيفات الإتقان وفق طريقة (GLN) بلغت (46.4%)، ونسبة تصنيفات التحسن بلغت (47.2%)، في حين أن نسبة تصنيفات الثبات بلغت (6.4%). ويُلاحظ أن نسبة تصنيفات الإتقان وفق طريقة (EN) بلغت (48%)، ونسبة تصنيفات التحسن بلغت (48%)، في حين أن نسب تصنيفات الثبات بلغت (4%)، كما يُلاحظ عدم وجود أية تصنيفات تراجع ذي دلالةٍ عياديةٍ وفقاً الطرائق السابقة جميعها. والشكل البياني الآتي يوضح هذه النسب.

شكل (15)

تمثيل بياني للتصنيفات العيادية وفق طرائق حسابها



- كما جرى حساب قيم مؤشرات التغيُّر الموثوق (RCI) والتصنيفات الناتجة عنها، وفق طريقة (HA) في حساب الدلالة العيادية، وقد أسفرت النتائج وفق هذه الطريقة عن معدلات مرتفعة جداً من تصنيفات التحسن والإتقان الدال عيادياً، التي تتناقض مع التصنيفات الناتجة عن بقية الطرائق من ناحية، كما تتناقض مع قيم معدلات الارتفاع في الدرجات التي لم تتجاوز درجة واحدة أو درجتين لدى بعض الأفراد من ناحية أخرى.

ولحساب نسب الاتفاق بين القيم والتصنيفات الناتجة عن الطرائق الثلاث المستخدمة لحساب الدلالة العيادية – مع استبعاد طريقة (HA)– ، تم حساب معاملات إحصاء كابا كوهين، لمقارنة حالات التوافق المشاهدة مع حالات التوافق المتوقعة بسبب المصادفة، علماً بأن هذه الطريقة تعتمد على المقارنات الثنائية، وهي ذاتها المستخدمة في جميع الأبحاث السابقة التي درست التوافق بين التصنيفات الناتجة عن الطرائق المختلفة لحساب الدلالة العيادية، مما يُمكّن من إجراء المقارنة بين نتائج هذا البحث ونتائج الأبحاث السابقة، وقد **وضع كوهين محكات ومعايير لتفسير نسب التوافق الناتجة عن إحصاء كابا (k) على النحو الآتي:**

- 0.1 إلى 0.20 توافق ضئيل جداً
- 0.21 إلى 0.40 توافق ضعيف
- 0.41 إلى 0.60 توافق متوسط
- 0.61 إلى 0.80 توافق عالٍ (جوهري)
- 0.81 إلى 0.99 توافق عالي جداً
- الواحد الصحيح توافق تام (Anthony & et al,2005, p.p 360 – 363)، والجدول الآتي يوضح ذلك.

جدول (32)

معامل كابا كوهين لتصنيفات الدلالة العيادية

الطرائق العيادية	قيم إحصاء كابا كوهين	تفسير نسب الاتفاق	Sig	القرار
(JT)*(GLN)	0.88	اتفاق عالٍ جداً	0.000	الاتفاق دال إحصائياً
(JT) * (EN)	0.83	اتفاق عالٍ جداً	0.000	الاتفاق دال إحصائياً
(GLN)*(EN)	0.96	اتفاق عالٍ جداً	0.000	الاتفاق دال إحصائياً

يُلاحظ من الجدول السابق: أن نسب الاتفاق بين طرائق حساب الدلالة العيادية الثلاث جميعها كانت دالة إحصائياً، وتدل على اتفاق عالٍ جداً وفق إحصاء كابا كوهين ومحكّات تفسيره، وأن أعلى نسبة اتفاق كانت بين طريقتي (EN) و(GLN) حيث بلغت (0.96)، يليها نسبة الاتفاق بين طريقتي (GLN) و (JT)، وأخيراً كانت أدنى نسبة اتفاق بين طريقتي (EN) و(JT) وقد بلغت (0.83).

7- الإجابة السؤال السابع: إلى أي مدى يُغيّر التدريب في مؤشرات الخصائص القياسية لاختبار الذكاء؟

للإجابة عن هذا السؤال جرى التحقق من أثر متغير التدريب في معاملات صدق اختبار الذكاء SPM-PV وثباته، وذلك على النحو الآتي:

أ- دراسة أثر متغير التدريب في معاملات صدق الاختبار:

جرى التحقق من أثر هذا المتغير في الصدق المحكي للاختبار بدلالة اختبار الذكاء المحكي Toni-4، من خلال حساب معامل ارتباط بيرسون لدرجات العينة التجريبية في التطبيق القبلي لهذين الاختبارين، وحسابه أيضاً في التطبيق البعدي المباشر، ثم مقارنة الفرق بين قيمتي معاملي الصدق المحكي باستخدام الاختبار الإحصائي V، والجدول الآتي يوضح نتائج هذا الاختبار.

جدول (33)

نتائج (V) لدلالة الفرق بين معاملي الصدق المحكي

	قيمة معامل الارتباط	كاي مربع	درجات الحرية	قيمة V	القيمة الاحتمالية	القرار
معامل الصدق المحكي في الاختبار القبلي	0.49**	3.84	1	5.24	0.02	دال إحصائياً
معامل الصدق المحكي في الاختبار البعدي المباشر	0.68**					

يتبين من الجدول السابق وجود فرق ذي دلالة إحصائية بين معاملي الصدق المحكي القبلي والبعدي المباشر، حيث بلغت القيمة الاحتمالية للاختبار الإحصائي V (0.02)، وهذا الفرق لصالح معامل الصدق المحكّي البعدي، والذي بلغت قيمته (0.68).

كما جرى التحقق من أثر متغير التدريب في الصدق التمييزي للاختبار، من خلال مقارنة الفرق بين متوسطي درجات الفئتين العليا والدنيا من العينة التجريبية في كلٍ من التطبيقين القبلي والبعدي المباشر، باستخدام الاختبار الإحصائيT-Student للمجموعتين المستقلتين، والجدول الآتي يوضح نتائج هذا الاختبار.

جدول (34)

فروق متوسطات الفئتين العليا والدنيا في الاختبارين

نوع الاختبار	نوع الفئة	العدد	المتوسط الحسابي	الانحراف المعياري	درجات الحرية	قيمة ت – ستودنت	القيمة الاحتمالية P-Value	حجم الأثر
الاختبار القبلي	الفئة العليا	42	47.67	2.95	82	18.20	0.00	0.80
	الفئة الدنيا	42	36.24	2.80				
الاختبار البعدي	الفئة العليا	42	55.83	1.54	82	20.04	0.00	0.83
	الفئة الدنيا	42	49.02	1.57				

يتبين من الجدول السابق وجود فرق ذو دلالة إحصائية وعملية بين متوسطي الفئتين العليا والدنيا في كل من الاختبارين القبلي والبعدي المباشر، وعلى الرغم من أن قيمة حجم الأثر المحسوبة للفرق في الاختبار البعدي المباشر كانت أعلى، إلا إن كلا حجمي الأثر المحسوبَين في التطبيقين القبلي والبعدي المباشر كان تصنيفهما مرتفع.

كما جرى التحقق من توزع أفراد الفئتين العليا والدنيا في الاختبارين القبلي والبعدي، علماً بأن مدى درجات الفئة الدنيا في الاختبار القبلي ينحصر بين (28 و 39)، ومدى درجات الفئة العليا في الاختبار القبلي ينحصر بين (44 و 55)، في حين أن مدى درجات الفئة الدنيا في الاختبار البعدي ينحصر بين (45 و 51)، ومدى درجات الفئة العليا في الاختبار البعدي ينحصر بين (54 و 60)، والجدول الآتي يبين توزع أفراد هاتين الفئتين في التطبيقين القبلي والبعدي.

جدول (35)

توزع أفراد الفئتين العليا والدنيا في الاختبارين

الاختبار القبلي				الاختبار البعدي			
رقم الفرد في الفئة الدنيا	درجته في الاختبار القبلي	رقم الفرد في الفئة العليا	درجته في الاختبار القبلي	رقم الفرد في الفئة الدنيا	درجته في الاختبار البعدي	رقم الفرد في الفئة العليا	درجته في الاختبار البعدي
19	28	101	55	19	45	38	60
36	31	17	54	48	46	101	59
105	31	83	53	68	46	17	58
55	32	113	53	93	46	76	58
93	32	79	52	7	47	83	58
42	33	104	52	35	47	108	58
86	33	20	51	49	47	8	57
122	33	32	51	4	48	15	57
11	34	61	50	28	48	40	57
68	34	70	50	36	48	51	57
12	35	87	50	66	48	56	57
67	35	1	49	86	48	61	57
97	35	39	49	107	48	87	57
119	35	15	48	2	49	104	57
121	35	38	48	22	49	14	56
3	36	65	48	60	49	29	56

56	32	49	88	47	8	36	7
56	43	49	92	47	14	36	110
56	70	49	109	47	51	37	9
56	72	49	110	47	58	37	28
56	79	49	119	47	75	37	53
56	115	49	120	47	76	37	66
55	10	49	122	47	108	37	88
55	24	50	46	47	115	37	98
55	39	50	50	46	35	38	22
55	52	50	53	46	73	38	26
55	81	50	55	46	82	38	30
55	91	50	63	46	91	38	46
55	95	50	67	46	94	38	49
55	103	50	84	46	100	38	117
55	106	50	89	46	109	39	2
55	113	50	97	45	13	39	4
54	1	50	98	45	40	39	16
54	13	50	105	45	56	39	18
54	20	50	124	45	64	39	25
54	27	51	3	45	72	39	41
54	58	51	6	45	103	39	44
54	64	51	9	45	120	39	47
54	65	51	11	44	10	39	57
54	69	51	12	44	27	39	74
54	75	51	26	44	29	39	89
54	78	51	73	44	33	39	116

يتبين من الجدول السابق أن (8) أفراد من الفئة الدنيا في الاختبار القبلي تحسن أداؤهم بحيث لم يبقَ تصنيفهم ضمن الفئة الدنيا في الاختبار البعدي، وتعادل نسبتهم (19.05%)، وهؤلاء الأفراد هم ذوو الأرقام: (121، 117، 16، 18، 25، 41، 44، 74).

وأن (6) أفراد من الفئة العليا في الاختبار القبلي لم يبق تصنيفهم ضمن هذه الفئة في الاختبار البعدي وتعادل نسبتهم (14.28%) وهؤلاء الأفراد هم ذوو الأرقام: (35 ، 73، 94، 109، 120، 33).

مع الإشارة إلى أن 4 منهم تغيّر تصنيفهم ليصبحوا من أفراد الفئة الدنيا وهم الأفراد ذوو الأرقام: (35، 73، 109، 120)، وللتأكد من نوع التغيُّر الحاصل في درجات هؤلاء الأفراد الذين تغير تصنيفهم في الفئتين الدنيا والعليا جرى حساب مؤشر التغيُّر الموثوق لكلٍ منهم، والجدول الآتي يوضح ذلك.

جدول (36)

مؤشر التغيُّر الموثوق للأفراد الذين تغيرت فئاتهم

نوع الفئة	رقم الفرد	درجته في الاختبار القبلي	درجته في الاختبار البعدي	قيمة مؤشر التغيَّر الموثوق (RCI)	التصنيف العيادي
الفئة الدنيا	121	35	52	5.04	تحسن
	117	38	52	4.15	تحسن
	16	39	52	3.86	تحسن
	18	39	52	3.86	تحسن
	25	39	52	3.86	تحسن
	41	39	52	3.86	تحسن
	44	39	53	4.15	أتقن
	74	39	52	3.86	تحسن
الفئة العليا	35	46	47	0.3	ثابت
	73	46	51	1.48	ثابت
	94	46	53	2.08	أتقن
	109	46	49	0.89	ثابت
	120	45	49	1.19	ثابت
	33	44	53	3.86	أتقن

يتبين من الجدول السابق أن: (7) من أفراد الفئة الدنيا الذين تغيرت فئتهم في الاختبار البعدي حققوا تصنيفات تحسن دال عيادياً، وواحداً منهم أتقن عملية أداء الاختبار، ولكن دون أن ينتقلوا إلى الفئة العليا التي ينحصر مدى درجاتها في الاختبار البعدي بين (54 و 60).

ويتبين أن (4) من أفراد الفئة العليا الذين تغيرت فئتهم في الاختبار البعدي، أصبح تصنيفهم ضمن الفئة الدنيا التي ينحصر مدى درجاتها في الاختبار البعدي بين (45 و 51)، ولكن دون أن يتراجع أداؤهم تراجعاً دالاً عيادياً بل بقي ثابتاً.

ب- دراسة أثر متغير التدريب في معاملات ثبات الاختبار:

جرى حساب معاملي ثبات ألفا كرونباخ لدرجات أفراد المجموعة التجريبية في كل من الاختبارين القبلي والبعدي المباشر، واختبار الفرق بين هذين المعاملين باستخدام الاختبار الإحصائي M. كما جرى

حساب معاملي ثبات سبيرمان براون للتجزئة النصفية لدرجات أفراد المجموعة التجريبية في كل من الاختبارين القبلي والبعدي المباشر، واختبار الفرق بين هذين المعاملين باستخدام الاختبار الإحصائي V، والجدول الآتي يوضح نتائج هذين الاختبارين علماً بأن درجات الحرية لكلٍ منهما تساوي (1)، وقيمة كاي مربع تساوي (3.84).

جدول (37)

فروق معاملي ثبات التطبيقين القبلي والبعدي المباشر

نوع مُعامل الثبات	قيمة المُعامل	قيمة M	قيمة V	القيمة الاحتمالية P-Value	القرار
ألفا كرونباخ للاختبار القبلي	0.756	5.538	-	0.02	دال إحصائياً
ألفا كرونباخ للاختبار البعدي	0.835				
التجزئة النصفية للاختبار القبلي	0.77	-	4.087	0.04	دال إحصائياً
التجزئة النصفية للاختبار البعدي	0.85				

يتبين من الجدول السابق وجود فرق ذي دلالة إحصائية بين قيمتي معاملي ثبات ألفا كرونباخ في الاختبارين القبلي والبعدي المباشر، حيث بلغت القيمة الاحتمالية للاختبار الإحصائي M (0.02)، وهذا الفرق لصالح معامل الثبات في الاختبار البعدي المباشر، والذي بلغت قيمته (0.835). كما يتبين وجود فرق ذي دلالة إحصائية بين قيمتي معاملي ثبات التجزئة النصفية في الاختبارين القبلي والبعدي المباشر، حيث بلغت القيمة الاحتمالية للاختبار الإحصائي V (0.04)، وهذا الفرق لصالح معامل التجزئة النصفية في الاختبار البعدي المباشر، والذي بلغت قيمته (0.85).

خلاصة:

قدم هذا الفصل عرضاً لنتائج البحث المتصلة بكلٍ من: منهج البحث الوصفي التحليلي لدراسة واقع الدلالات العلمية في الأبحاث التجريبية والنسب المئوية لاستخدامها، ومنهج التحليل البعدي للتحقق من مدى كفاية كل من الدلالتين الإحصائية والعملية في بيان نتائج البرامج التجريبية والقدرة التنبؤية لمعاملات كل من هاتين الدلالتين بقيم وتصنيفات الدلالة العيادية، وأخيراً منهج البحث التجريبي لدراسة أثر متغير التدريب في نتائج الدلالات العلمية الثلاث وفي الخصائص القياسية للاختبار.

الفصل السادس

مناقشة النتائج والمقترحات

الفصل السادس

مناقشة النتائج والمقترحات

تمهيد:

طالما أن البحث الحالي ركز على تحليل واقع استخدام الدلالات العلمية، وسعى إلى التحقُّق من العلاقات بينها، ودراسة أثر متغير التدريب فيها وفي الخصائص القياسية للاختبار، من خلال الأسئلة التي عُرضت في الفصل الأول، وجرى الإبلاغ عن نتائجها في الفصل الخامس، فستتم مناقشتها في هذا الفصل استناداً إلى المناهج البحثية المُستخدمة في الإجابة عنها من ناحية، وفي ضوء إجراءات البحث وأهدافه والدراسات السابقة والأدبيات المتصلة بموضوعه من ناحيةٍ أخرى.

1- مناقشة نتائج البحث:

- لقد انتهت نتائج الإجابة عن السؤال الأول حول واقع استخدام الدلالات العلمية في الأبحاث التجريبية إلى أن نسبة الاختبارات الإحصائية التي جرى التحقق من دلالتها العملية في نتائج هذه الأبحاث بلغت (29.47%)، ووضحت أن النسب المئوية لحساب قوانين الدلالة العملية في تزايدٍ مستمر عبر الأعوام، وهو مؤشرٌ جيدٌ يدل على تزايد إدراك الباحثين لأهمية حساب هذا النوع من الدلالة إلى جانب اختبارات الدلالة الإحصائية، مع الأخذ بالحساب ذلك التزايد المستمر أيضاً في عدد الأبحاث التجريبية عبر الأعوام، وزيادة ميل الباحثين لاعتماد المنهج التجريبي، وهذا بحد ذاته يشكل مدعاة أكبر لضرورة حساب مؤشرات الدلالات العلمية الثلاث فيها، للوصول إلى دقةٍ أكبر في الإبلاغ عن النتائج والقرارات البحثية المستمدة منها. من ناحيةٍ أخرى لم تنته نتيجة السؤال الأول بعدم مناسبة عدد اختبارات الدلالة العملية المحسوبة مع العدد الكلي لاختبارات الدلالة الإحصائية فقط، بل كان هناك عدد من الأخطاء الإحصائية شائعة الاستخدام في حساب الدلالتين الإحصائية والعملية بشكل ملفتٍ للانتباه، جاء في مقدمتها:
  - ❖ الحكم على تصنيفات حجم الأثر المحسوبة، وتفسير قيمها وفق التصنيفات التفسيرية الخاصة بكل نوع من أنواع قوانينه، مثل حساب الصيغة الرياضية لقانون إيتا تربيع، والحكم على قيمتها استناداً إلى تصنيفات حجم الأثر الخاصة بمعامل d لكوهين.
  - ❖ الحكم على صحة القرار الإحصائي برفض أو قبول الفرض الصفري، استناداً للقيمة الاحتمالية (P-Value)، عند مقارنتها بمستوى دلالة معين، مثل الحكم على القيمة الاحتمالية (0.266) على أنها دالة إحصائياً على الرغم من أنها أكبر من مستويي الدلالة (0.05) و(0.01).
  - ❖ الخلط في استخدام القوانين المختلفة لحجم الأثر، من مثل استخدام قوانين إيتا تربيع للعينات المترابطة والعينات المستقلة، واستخدام معادلة نسبة الكسب سواءٌ الأساسية لـ McGuigan أم

المعدلة لـ Blake كمعيارٍ وحيد للحكم على تأثير البرامج التجريبية، وهي ليست بالقليلة إذ تعادل نسبة تكرارها (70%) في مجمل الأبحاث التجريبية.

ولعل من المهم هنا الانتباه لمسألة تتعلق بمعادلة نسبة الكسب سواء الأساسية أم المعدّلة، إذ في الوقت الذي لم يهدف فيه أي بحث من الأبحاث السابقة إلى تحليل واقع استخدام هذه المعادلة كمؤشر للدلالة العملية، أو أن يتضمن إطاره النظري أي شرح عنها، فإن أدبيات البحث الحالي المتصلة بهذا الجانب أيضاً لم تتطرق لها ولم تسلط الضوء عليها، باستثناء (محمد حسن، 2011) الذي أكد من خلال أمثلة محسوبة أن هذه المعادلة غير حساسة للفروق بين درجات الاختبارين القبلي والبعدي، ومن الأمثلة التي أوردها أن "الدرجة (60) في الاختبار البعدي مقابل (0) في الاختبار القبلي حظيت بنسبة كسب معدلة بلغت (1.20)" وهي القيمة التي تمثل الحد الأدنى للحكم على البرنامج بأنه فعال أو مقبول، في حين أن "الدرجة (59) في الاختبار البعدي مقابل (0) في الاختبار القبلي حظيت بنسبة كسب معدلة بلغت (1.18)" وهي بذلك لا تصل إلى نسبة الكسب المقبولة، لذا أوصى بضرورة حساب قيم حجم الأثر جنباً إلى جنب فيما لو تم حساب نسبة الكسب المعدلة للحكم على فعالية البرامج التجريبية (محمد حسن، 2011، ص298).

من زاويةٍ أخرى هناك ما يلفت النظر عند مراجعة معادلات حساب نسبة الكسب، إذ إنها تعتمد فقط على المتوسط الحسابي دون تضمين قيم الانحراف المعياري أو الخطأ المعياري للقياس فيها، كما شأن معادلات حجم الأثر المتفق عليها في أدبيات الدلالة العملية.

فعلى الرغم من أهمية المتوسط الحسابي في وصف تجمُّع درجات المجموعة بقيمة واحدة يُستعاض بها عن قيم أفراد المجموعة، وتُمكّن معرفتها من المقارنة بين متوسطات مجموعات مختلفة، غير أن حساب المتوسط وحده قد لا يكفي لوصف قيم درجات المجموعة وصفاً دقيقاً، لأن قيمته غير حساسة للفروق بين درجات الأفراد داخل المجموعة، ومقدار انتشارها وابتعادها عنه؛ فمثلاً إذا جرى اختبار مجموعتين من الأفراد في اختبار قدرة خاصة، وكانت درجات المجموعة الأولى هي: (صفر، 25، 50)، ودرجات أفراد المجموعة الثانية (24.5، 25، 25.5)، فمن الواضح أن قيمة المتوسط الحسابي لكل منهما هو (25)، ولكن لا يمكن القول أن المجموعتين متعادلتان في القدرة، لأن قيم درجات المجموعة الأولى متباعدة، في حين أن قيم درجات المجموعة الثانية متقاربة، وبالتالي لا بد من حساب قيمة الانحراف المعياري لكل من المجموعتين إضافةً لقيمة المتوسط الحسابي، للتعبير الإحصائي بصورة أدق عن الدرجات وتوضيح مقدار انتشارها. لعل هذا ما يؤكد عدم كفاية الاعتماد على المتوسط وحده في معادلات حساب نسب الكسب بين المتوسطات القبلية والبعدية لتطبيق البرامج التجريبية (محمد خيري، 1997، ص69، علام، 2002، ص ص.93-94، فهمي، 2005، ص.186).

قد تشير النتيجة السابقة إلى أن دقة القرارات الإحصائية المحسوبة استناداً إلى اختبارات الدلالة الإحصائية، أو حتى تلك التي جرى التحقق من دلالتها العملية في نتائج الأبحاث المُحلّلة، من المحتمل أن تكون موضع شك، وأن حساب الدلالات العلمية فيها لم يكن في المستوى الدقيق تماماً، سواءٌ أكان ذلك فيما يتصل بنسب استخدامها، أم فيما يتصل بدقة حسابها وتفسير نتائجها. ولعل السبب في ذلك يعود إلى عدم توجيه الباحثين إلى ضرورة الاهتمام بحساب الأنواع المختلفة من الدلالات، أو لأن التوجه العام في إنجاز البحوث الأكاديمية مازال غير معتمد على نتائج الأبحاث والدراسات التي أكدت أهمية حساب الأنواع المختلفة من الدلالات وضرورة حسابها في أي بحثٍ كمي كان. مما قد يؤكد وجوب توجيه الباحثين لتضمين الدلالات جميعها وفقاً لطبيعة أبحاثهم وخصائصها، وتدريبهم – قبل ذلك– على امتلاك مهارات حسابها وتفسير نتائجها.

لا تختلف هذه النتيجة عن نتائج عمليات التحليل التي أظهرتها الأبحاث السابقة لواقع استخدام الدلالتين الإحصائية والعملية؛ إذ بينت نتائج (بحاش، 2020) أن نسبة حساب حجم الأثر في فرضيات أبحاث الجامعات الجزائرية المدروسة ضعيفة جداً بلغت (2.71%) من إجمالي عدد الفرضيات المحسوبة، وبينت وأظهرت نتائج (البارقي،2012) أن نسبة الاختبارات الإحصائية التي جرى حساب حجم الأثر لها بلغت (0.76%) وهي نسبة ضعيفة جداً، كما أكدت نتائج (Balkin & Lenz, 2021) أن أوجه القصور في نتائج الدلالة الإحصائية موثقة جيداً، وأن التوسع في حساب حجوم الأثر وفترات الثقة ومؤشرات الدلالة العيادية لنتائج الأبحاث الإرشادية أمرٌ ضروري لتعزيز فهم النتائج. في حين بينت نتائج (McMillan & Foley, 2011) أنه على الرغم من أن النسب المئوية لحساب حجوم الأثر في أبحاث المجلات العلمية المستهدفة تراوحت من 64% إلى 82% إلا أن حوالي نصف هذه الأبحاث لا تتضمن أي مناقشات أو تفسيرات لنتائج حجم الأثر، لذا من الصعب معرفة ما إذا كان الباحثون على دراية بمعنى حجم الأثر ومفهومه، أم جرى حسابه فقط كإجراء إحصائي روتيني (McMillan & Foley, 2011, p.8).

من ناحيةٍ أخرى، فقد أظهرت نتائج البحث الحالي أنه لم يتم حساب أية دلالة عيادية في أي بحثٍ من الأبحاث المستهدفة على الإطلاق، شأنها في ذلك شأن الدراسات السابقة –سواءٌ العربية منها أم الأجنبية–؛ إذ لم يستهدف أيٌ منها تحليل واقع استخدام هذه الدلالة، بل اقتصرت جميعها على تحليل الدلالتين الإحصائية والعملية فقط.

- وأسفرت نتائج التحليل البعدي في السؤالين الثاني والثالث حول مدى كفاية الدلالات العلمية في دقة القرار الإحصائي عن مجموعة من النتائج يلخصها الجدول الآتي.

جدول (38)

تزامُن نتائج الدلالات العلمية

أنواع الدلالات		نسب تصنيفات حجم الأثر			متوسط نسب تصنيفات مؤشر التغيُّر الموثوق (الدلالة العيادية)			
		المرتفع	المتوسط	المنخفض	الإتقان	التحسُّن	الثبات	التراجع
القرار الإحصائي	الدال	98.9	0.5	0.5	70.67	5.93	23.35	0.05
	غير الدال	8.3	75	16.7	2.34	1.28	94.04	2.34
تصنيفات حجم الأثر	المرتفع				71	5.95	23	0.05
	المتوسط				2.19	1.80	94.87	1.14
	المنخفض				5.56	0	88.88	5.56

يتبين من الجدول السابق أن كلا نوعي القرار المستندين إلى الدلالة الإحصائية قد ارتباطا بتصنيفات الدلالتين العملية والعيادية جميعها بنسبٍ متفاوتة. فمن الواضح، أنه بالرغم من أن القرار الإحصائي الدال قد تزامن بنسبٍ ضئيلة مع حجوم أثر متوسطة ومنخفضة، وحالات تراجع دال عيادياً، فإن ذلك لا يمنع من تأكيد أهمية هذه النتيجة، إذ تُبيّن أن هناك تأثير منخفض لبعض البرامج التدريبية التي أقرّت نتائجها بفروقٍ دالة إحصائياً بين المتوسطات القبلية والبعدية، وتشير إلى وجود حالات فردية قد تراجعت تراجعاً دالاً عيادياً نتيجة تعرضها لهذه البرامج.

وفي هذا السياق لا بد من تأكيد ظهور أثر لتضخم الدرجات في نتائج أبحاث التحليل البعدي، فمن أصل (196) فرضية اختبرت دلالة الفروق بين متوسطات مجموعات تجريبية وضابطة في الاختبارات القبلية والبعدية، كانت النتائج تقر بوجود فروق دالة إحصائياً لكامل فرضيات المجموعات التجريبية والبالغة (119) فرضية، ولـ (65) من أصل (77) فرضية من فرضيات المجموعات الضابطة، وبالتالي فإن نسبة الفرضيات الدالة إحصائياً تعادل (93.88%) من مجمل الفرضيات. في حين اقتصر عدد الفرضيات غير الدالة إحصائياً على (12) فرضية تعادل نسبتها (6.12%) فقط من مجمل الفرضيات.

ولعل ما تجدر الإشارة إليه أن نتائج التحليل البعدي بينت ظهور حجوم الأثر المنخفضة والتراجع العيادي المتزامنة مع القرار الدال في كلا المجموعات الضابطة والتجريبية على حدٍ سواء.

إن عدم قدرة الدلالة الإحصائية على إظهار التأثير المنخفض لبعض البرامج وحالات التراجع العيادي، قد يعود لآلية حساب هذا النوع من الدلالة الذي يقتصر على استخراج قيمة الفرق الناتج عن

المعالجة التجريبية على مستوى المجموعة، ولا يأخذ باستخراجها على مستوى الأفراد. الأمر الذي قد يحجب حقيقة عدم استفادة بعض الأفراد أو تراجع أدائهم نتيجة تعرضهم للبرنامج التجريبي. هذا ما أكدته نتائج (Eisen& et al,2007) التي أظهرت أن حساب التغيُّرات على مستوى المجموعة قد أغفل حقيقة عدم ظهور أي تغيّرٍ حقيقي بالاتجاه المطلوب لدى نسبة كبيرة من الأفراد، بل ظهر لدى بعض الأفراد تراجعاً دالاً عيادياً بعد تطبيق البرنامج (Eisen& et al,2007, p. 286).

وبالمقابل أظهرت النتائج أن (75%) من القرارات غير الدالة إحصائياً قد تزامنت مع حجوم أثر متوسطة، و(8.3%) مع حجوم أثر مرتفعة، وتصنيفات إتقان وتحسن دالة عيادياً بلغت متوسط نسبتيهما (2.34) و(1.28) على التوالي؛ وهذا يعني أنه حتى في بعض الحالات التي يكون فيها القرار الإحصائي غير دال، فقد تُظهر قيم الدلالة العملية حجوم أثر متوسطة، وقد تكون مرتفعة في بعض الأحيان. كما من المحتمل أن يترافق القرار غير الدال ببعض الحالات الفردية التي حققت تصنيفات إتقان وتحسن دال عيادياً، مما يؤكد أن القرار الإحصائي وحده غير قادر على الإبلاغ بدقة عن الأثر الواقعي والحقيقي للبرامج التجريبية من ناحية، وعن التغيُّرات الفعلية التي تحدث على مستوى الأفراد واتجاهها من ناحية أخرى.

وقد تزامنت التصنيفات المختلفة لحجوم الأثر وفق نتائج التحليل البعدي بجميع تصنيفات الدلالة العيادية، باستثناء حجم الأثر المنخفض الذي لم يترافق مع تصنيف التحسن الدال عيادياً؛ إذ ترافقت بعض حجوم الأثر المرتفعة مع تصنيفات الثبات والتراجع العيادي، وتزامنت بعض حجوم الأثر المنخفضة مع تصنيفات الإتقان العيادي؛ مما قد يشير إلى عدم كفاية التصنيفات المختلفة لحجوم الأثر في الإبلاغ عن نتائج الأبحاث التجريبية. ولعل السبب في ذلك كما في الدلالة الإحصائية يعود لأن آلية حساب الدلالة العملية يتم على مستوى المجموعة فقط، وليس على مستوى الأفراد.

استناداً لكل ما سبق يمكن القول أن كلا نتائج الدلالتين الإحصائية والعملية قاصر عن الإبلاغ الدقيق عن نتائج الأبحاث التجريبية، ذلك لأن القرارين الإحصائيين الدال وغير الدال، والتصنيفات المختلفة لحجوم الأثر قد ترافقت مع تصنيفات الدلالة العيادية جميعها، الأمر الذي قد يثير الشك في صحة ودقة النتائج المستمدة منها، ويُضعِف الثقة بالقرارات البحثية التي تبنى على أساسها. ويتوقع فيما لو جرى توسيع عينة البحث لتشمل جميع الأبحاث التجريبية المنجزة أن تتأكد هذه النتيجة وتتوضح بصورة أكبر.

لقد استدعت النتيجة آنفة الذكر ضرورة حساب القدرة التنبؤية للدلالتين الإحصائية والعملية بقيم وتصنيفات الدلالة العيادية، والتي أسفرت نتائجها في السؤال الرابع عن وجود قدرة تنبؤية لقيم الدلالة الإحصائية بقيم تصنيفي الإتقان والثبات الدال عيادياً، ووجود قدرة تنبؤية لقيم الدلالة العملية بقيم تصنيفات: الإتقان والثبات والتراجع الدال عيادياً، إلا أن كلا هذين النوعين من الدلالة لم يستطع التنبؤ

بقيم وتصنيفات الدلالة العيادية جميعها. ولعل كل ما سبق يبيّن أن حساب أي من نوعي الدلالة الإحصائية أو العملية، أو حتى حساب كليهما معاً، غير كافٍ لاستخراج نتائج الأبحاث التجريبية، ذلك لأن قيمهما غير قادرة على التنبؤ بجميع تصنيفات مؤشر التغيُّر الموثوق (الدلالة العيادية) التي تحدد نوع التغيُّر الفعلي الذي يحدث على مستوى كل فرد في المجموعة؛ وبالتالي لا غنى عن حساب معاملات الدلالة العيادية في الإبلاغ عن نتائج هذه الأبحاث وتفسيرها. وبات من الواضح تماماً ضرورة أن يضع الباحث في الحسبان تصوراَ محدداً ودقيقاً لأنواع الدلالات العلمية التي عليه حسابها، وتفسير نتائج بحثه في ضوئها منذ بداية تصميم بحثه، بما يتفق مع طبيعته ونوع منهجه وعيناته، ومتغيراته. وفي هذا السياق لا بد أيضاً، من تأكيد ضرورة حساب الدلالتين الإحصائية والعملية معاً في تصاميم الأبحاث غير التجريبية مهما كان نوعها، وضرورة حساب الأنواع الثلاثة من الدلالات العلمية معاً في تصاميم الأبحاث التجريبية بجميع أنواعها.

إن عدم وجود دراسات سابقة استهدفت تحليل العلاقات بين الدلالات الثلاث لم يمكّن الباحثة من مناقشة نتائج هذا السؤال إلا في ضوء أدبيات البحث التي تناولت هذا الجانب، وكذلك منهجه ومتغيراته وخصائص عيناته.

- وبينت نتائج السؤال الخامس المتصلة بمنهج البحث التجريبي، والتي اختصت بدراسة أثر متغير التدريب في نتائج الدلالات العلمية والخصائص القياسية للاختبار، أن هذا المتغير قد أثر في نتائج الدلالات العلمية لدرجات المجموعة التجريبية في التطبيق البعدي المباشر تأثيراً ذا دلالة إحصائية وعملية وعيادية، حيث بلغت القيمة الاحتمالية للفرق بين المتوسطين القبلي والبعدي المباشر (0.00) وبلغت قيمة الدلالة العملية لهذا الفرق (0.90) وهي تدل على حجم أثر مرتفع، كما أسفرت نتائج الدلالة العيادية عن معدلات إتقان وتحسن بلغ مجموع نسبتيهما (86.4%) من المجموع الكلي لأفراد المجموعة التجريبية.

بالمقابل، انخفضت قيم الدلالتين العملية والعيادية لهذا الفرق في التطبيق البعدي المؤجل مقارنةً بالقبلي، لتصبح قيمة حجم أثر الفرق (0.79)، ومجموع نسب الإتقان والتحسن الدال عيادياً (56%)، في حين ارتفعت نسبة تصنيف الثبات الدال عيادياً في التطبيق البعدي المؤجل إلى (44%) لتدل على تراجع درجات بعض الأفراد في الاختبار البعدي المؤجل وعودتها إلى معدلاتها الأساسية كما في الاختبار القبلي.

كذلك، فقد انخفضت أيضاً قيم الدلالات العلمية في التطبيق البعدي المؤجل مقارنة بالتطبيق البعدي المباشر، لتكون نتيجة الدلالة الإحصائية لصالح البعدي المباشر، ولتدل قيم الدلالة العيادية على أن (71.2%) من أفراد المجموعة التجريبية قد احتفظوا بأثر التدريب وبقيت معدلاتهم ثابتة، في حين تراجع أداء (28.8%) من أفراد هذه المجموعة وعادت معدلات ذكائهم إلى قيمها الأصلية قبل عملية التدريب.

تفسر الباحثة هذه النتيجة بأن متغير التدريب قد أثّر بارتفاع درجات بعضاً من أفراد المجموعة التجريبية في الاختبار البعدي المباشر، نتيجة تعرضهم لجلسات البرنامج التدريبي، مع عدم وجود فاصل

زمني بين التدريب والتطبيق المباشر للاختبار، دون أن يؤثر هذا المتغير في حقيقة الأمر بقدراتهم العقلية ومستويات ذكائهم الأساسية، بناءً على انخفاض وتراجع درجات نسبة لا بأس بها من أفراد هذه المجموعة والتي تقارب الثلث في الاختبار البعدي المؤجل مقارنة بالاختبار البعدي المباشر، لتعود إلى معدلاتها الأصلية كما في الاختبار القبلي، ولعل ذلك قد يعود لمرور فترة زمنية لم يتلقوا فيها أي تدريب كما لم يتعرضوا فيها لأداء هذا النوع من الاختبار، مما قد أدى إلى فقدانهم جزءاً لا بأس به من أثر هذا المتغير، وبعبارةٍ أخرى، يمكن القول بأن ارتفاع الدرجات في الاختبار البعدي المباشر يعود للأثر المباشر التدريب، والذي تنخفض قيمته تدريجياً مع مرور الوقت.

وفي هذا السياق لا بد من التأكيد أنه عند الإبلاغ عن تأثير البرامج التجريبية في نتائج الأبحاث، لا يكفي حساب حجم أثر هذه البرامج من خلال مقارنة الفرق بين قيم الاختبارين القبلي والبعدي المباشر فحسب، بل لابد من الإبلاغ عن حجم أثر الاحتفاظ، من خلال حساب قيمة ودلالة الفرق بين الاختبارين البعدي المؤجل والبعدي المباشر من جهة، والبعدي المؤجل والقبلي من جهةٍ أخرى، ليس فقط على مستوى المجموعة من خلال مؤشرات الدلالتين الإحصائية والعملية، بل أيضاً على مستوى كل فرد في المجموعة من خلال مؤشرات التغيُّر الموثوق.

- لقد تمت المفاضلة في الإجابة عن السؤال البحثي السادس بين الطرائق الإحصائية المختلفة لحساب مؤشرات الدلالة العيادية، ومقارنة التصنيفات الناتجة عنها، لعدم وجود اتفاق في أدبيات البحث والدراسات السابقة حول طريقةٍ علميةٍ جامعةٍ مانعةٍ ومثلى لحساب هذا النوع من الدلالة، ولوجود توصية في أدبيات البحث بضرورة إجراء المزيد من المقارنات بين هذه التقنيات الإحصائية، وذلك بغية الوصول إلى قرارٍ حول طريقةٍ محددة لحساب هذا النوع من الدلالة، وإجراء المقارنات بين نتائج الأبحاث المنجزة.

بناءً على ذلك بينت النتائج وجود نسب اتفاق دالة إحصائياً وعالية جداً بين القيم والتصنيفات المُستخرجة من الطرائق الثلاث (JT) و(GLN) و(EN)، وقد كانت أعلى نسبة اتفاق بين طريقتي (EN) و(GLN) إذ بلغت (0.96)، يليها نسبة الاتفاق بين طريقتي (GLN) و (JT) وقد بلغت (0.88)، وأخيراً كانت أدنى نسبة اتفاق بين طريقتي (EN) و(JT) والتي بلغت (0.83).

فعلى الرغم من أن كل من طريقتي (GLN) و(EN) جرى تأسيسهما بناءً على الانتقادات الموجهة لطريقة (JT) وقد كان أبرزها إهمالها قيمة المتوسط الحسابي، حيث أكد (Hsu, 1989) ضرورة تعديل الصيغة الرياضية لهذه الطريقة لتشمل تقديرات المتوسط إضافةً إلى الانحراف المعياري، كما اقترح (Speer, 1992) مؤسس طريقة (EN) وسيلة لمنع التضخم في معدلات التحسن تعتمد في حسابها على فترة ثقة مكونة من انحرافين معياريين مضروباً بالخطأ المعياري للقياس، إلا أن الأساس النظري لهذه الطرائق الثلاث متقارب جداً، حيث تعتمد جميعها على قيم الخطأ المعياري للقياس، والتغيُّر الأساسي في طريقتي (GLN) و(EN)هو إضافتهما قيمتي المتوسط الحسابي والانحراف المعياري للصيغة الحسابية

لطريقة (JT)، ومن الجدير بالذكر أن قيمة الانحراف المعياري هي من القيم الأساسية لحساب الخطأ المعياري الذي يمثل مقام الصيغة الرياضية لطريقة (JT)، كما أن فترتي الثقة المُقترحتين من قبل سبير هما عبارة عن انحرافين معياريين مضروبان بالخطأ المعياري للقياس؛ وبالتالي، يبدو أن هاتان الصيغتان لم تختلفا كثيراً عن الصيغة الرياضية لطريقة (JT).

على الرغم من تباين نتائج الدراسات السابقة حول نسب الاتفاق المحسوبة لتصنيفات هذه الطرائق إلا أنها بينت بالمجمل أن نسب الاتفاق بين هذه الطرائق الثلاث كانت عالية، حيث وضحت نتائج دراسة رونك وآخرين (Ronk & et al., 2014) اتساق التصنيفات الناتجة عن الطرائق (JT, GLN, EN)

وأن نسب الاتفاق بينها بلغت (93.7%)، وأظهرت نتائج دراسة أتكينز (Atkins & et al, 2005) وجود اتفاق دال إحصائياً بين التصنيفات الناتجة عن الطرائق الأربع الآتية: (JT, GLN, EN, HA)، مع الإشارة إلى أن طريقة (HA) هي الاكثر تحفظاً في إظهار حالات التعافي الدال عيادياً، وقد أكدت نتائجها أيضاً أن طريقتي (JT) و (GLN) أدتا إلى تصنيفات عيادية متطابقة تقريباً، مع أن طريقة (GLN) أُسست بالأصل بناءً على الانتقادات الموجهة لطريقة (JT)، في حين بينت دراسة باور وآخرين (Bauer & et al, 2004) وجود فروق ذات دلالة إحصائية بين تصنيفات الطرائق الآتية (JT, GLN, EN, HA, HLM)، باستثناء طريقتي (JT) و (GLN)؛ إذ أنتجت هاتان الطريقتان أعلى نسبة اتفاق بلغت (0.93).

وفي هذا السياق تجدر الإشارة إلى أن معظم الدراسات السابقة أكدت أن طريقة (JT) تعطي تقديرات معتدلة لتأثير المعالجة التجريبية، كما أنها الأبسط والأسهل استخداماً، لذا أوصت باعتمادها في حساب نتائج الدراسات والأبحاث المستقبلية، مثل دراسة (Ronk & et al., 2016)، و(نصار، 2016)، و(Ronk & et al., 2014)، و(Shada, 2013)، و(Bauer & et al, 2004).

أما بالنسبة لطريقة (HA) فقد أظهرت نتائجها معدلات مرتفعة جداً من تصنيفات التحسن الدال عيادياً، التي تتناقض مع التصنيفات الناتجة عن بقية الطرائق، ولعل ما يلفت النظر أن هذه الطريقة أنتجت تصنيفات تحسن دالة عيادياً حتى في الحالات التي كانت فيها معدلات الارتفاع في درجة الفرد البعدية لا تتجاوز درجة أو درجتين. ولعل الباحثة ترجح أن السبب في ذلك يعود إلى التعقيدات الإحصائية في بسط الصيغة الحسابية لهذه الطريقة، فمن المحتمل أن التطورات الناتجة عن التعديلات التي جرى إدخالها على الصيغة الرياضية لهذه الطريقة أدت لجعلها تتناسب أكثر مع الحالات المرضية السريرية، التي تكون فيها درجة الشدة المرضية للفرد قبل إجراء التدخل مرتفعة، ويُفتَرض أن تنخفض نتيجة التعرض للمعالجة، مما يجعلها ملائمةً للبرامج العلاجية التي تهدف إلى تخفيض المؤشرات

المرضية، وخفض درجة الفرد في الاختبار البعدي عنها في الاختبار القبلي، ولا تصلح للبرامج التدريبية التي تهدف إلى تنمية قدرات الفرد ومهاراته، ورفع درجته في الاختبار البعدي عنها في الاختبار القبلي.

ولدى قيام الباحثة بحساب هذه الطريقة لدرجات عينات سريرية كما في المثال الموضّح في الملحق رقم (5)، والمحسوب لدرجة مريض بعد تعرضه لبرنامج علاجي لخفض شدة المخاوف النوعية، أظهرت النتائج قيمة منطقية لمؤشر (RCI) تتناسب مع القيم المحسوبة باستخدام الطرائق الأخرى.

تتفق هذه النتيجة مع دراسة (Shada, 2013) التي بينت نتائجها أن طريقة (HA) في حساب الدلالة العيادية أنتجت نسب تحسن وتعافٍ دال عيادياً مرتفعة جداً، وتتناقض مع التصنيفات الناتجة عن الطرائق الأخرى في حساب الدلالة العيادية، وأرجع الباحث هذه النتيجة إلى التعقيدات الإحصائية في مقام المعادلة، مفسراً إياها بأنها السبب الذي جعل التصنيفات الناتجة عن هذه الطريقة غير منطقية في إظهار معدلات عالية جداً من تصنيفات التحسن الدال عيادياً، والتي تتناقض مع تصنيفات الطرائق الأخرى من جهة، ومع نتائج الدراسات السابقة التي أكدت بأن هذه الطريقة هي الأكثر تحفظاً في إظهار تصنيفات التحسن والتعافي الدال عيادياً (Shada, 2013, P.P 31 -32).

كما بينت نتائج دراسة (Ronk & et al., 2016) فيما يخص هذه الطريقة عدم وجود فروق ذات دلالة إحصائية بين تصنيفاتها وتصنيفات طريقة(JT) خاصةً فيما يتعلق بتصنيف التعافي الدال عيادياً، ولأنه لم تكن هناك ميزة ملحوظة لاستخدام طريقة (HA)على طريقة (JT)، فقد أوصت نتائجها باستخدام طريقة (JT).

يتبين مما سبق أن طرائق حساب الدلالة العيادية لا تصلح جميعها للحساب والاستخدام في جميع أنواع البرامج التجريبية، وأن اختيار الباحث لإحدى هذه الطرائق واعتمادها لحساب نتائج برنامجه التجريبي يجب أن يستند إلى قواعد استخدام يتم اعتمادها وفقاً لنوع برنامجه والهدف منه، وطبيعة عيناته وتوصيف خصائصها بشكل دقيق. وباستقراء نتائج البحث الحالي والدراسات السابقة وأدبيات البحث المتصلة بهذا المجال، يمكن استنتاج المؤشرات الآتية التي تساعد الباحثين في اختيار الطريقة الأنسب لحساب الدلالة العيادية:

- تصلح الطرائق الاتية: (JT) (GLN) (EN) للاستخدام في أبحاث البرامج التجريبية التي يكون الهدف الأساسي منها تنمية مهارات وكفاءات الأفراد.
- تصلح طريقة (HA) للاستخدام في أبحاث البرامج العلاجية والإرشادية التي تتضمن عينات سريرية، تهدف إلى خفض حدة أعراض واضطرابات معينة لدى أفرادها.
- تصلح طريقة (JT) للاستخدام في جميع أنواع البرامج نظراً لسهولة حسابها، ودقتها في إعطاء تقديرات معتدلة لأثر المعالجات التجريبية، سواءٌ أكانت برامج علاجية أم تدريبية.

- وانتهت نتائج السؤال السابع والأخير الذي يتناول دراسة أثر متغير التدريب في الخصائص القياسية لاختبار الذكاء: إلى أن هذا المتغير قد أثّر بارتفاع معامل الصدق المحكي في التطبيق البعدي المباشر لعملية التدريب ارتفاعاً ذا دلالة إحصائية، في حين لم يؤثر في معاملات الصدق التمييزي للاختبار تأثيراً ذا دلالة إحصائية وعملية، فعلى الرغم من أن قيمة الاختبار الإحصائي **T-Student** للفرق بين متوسطي الفئتين العليا والدنيا وقيمة حجم الأثر المحسوبة لها كانت أعلى بعد التدريب، إلا أن كلا القيمتين القبلية والبعدية لهذا الاختبار وكلا حجمي الأثر المحسوبَين له كانا دالين إحصائياً ومرتفعين عملياً.

وعند التحقُّق من توزع أفراد الفئتين الدنيا والعليا في الاختبارين القبلي والبعدي، تبيّن أن (8) أفراد من الفئة الدنيا و(6) أفراد من الفئة العليا تغير توزعهم في الاختبار البعدي، ولم تعد تصنيفاتهم في الفئتين الدنيا والعليا ذاتها كما كانت في الاختبار القبلي. وعند القيام بحساب مؤشر التغيُّر الموثوق لدرجات هؤلاء الأفراد بينت النتائج أن (7) من أفراد الفئة الدنيا الذين تغيرت فئاتهم قد تحسن أداؤهم تحسناً ذا دلالة عيادية، وفرداً واحداً منهم قد أتقن الأداء في الاختبار دون أن ينتقل تصنيفهم إلى الفئة العليا، في حين أن (4) أفراد من الفئة العليا الذين تغيرت فئتهم في التطبيق البعدي المباشر، أصبح تصنيفهم ضمن الفئة الدنيا دون أن يتراجع أداؤهم تراجعاً دالاً عيادياً.

فبالرغم من أن متغير التدريب لم يؤثر تأثيراً ذا دلالة إحصائية وعملية في معاملات الصدق التمييزي لكنه أثر تأثيراً ذا دلالة عيادية على مستوى أفراد الفئتين الدنيا والعليا، ولأن كل فرد ضمن هاتين الفئتين قد تأثر بهذا المتغير بطريقةٍ مختلفة، فقد يكون من المهم تأكيد ضرورة حساب مؤشرات التغيُّر الموثوق عند التحقق من هذا النوع من صدق الاختبار.

ووضحت النتائج أن متغير التدريب قد أثر في ارتفاع قيمتي معاملي ثبات ألفا كرونباخ والتجزئة النصفية ارتفاعاً ذا دلالة إحصائية في الاختبار البعدي المباشر، مما يمكّن من القول أن متغير التدريب قد أثر بارتفاع معظم قيم مؤشرات الخصائص القياسية، ولعل هذا يدل على أهميته ودوره في تحسين هذه الخصائص، ورفع مؤشرات صدق وثبات الاختبار.

ولعل السبب في ظهور هذه النتيجة يعود لتعرّض أفراد المجموعة التجريبية لجلسات البرنامج التدريبي الذي – من المحتمل– قد قلل من التباين في درجاتهم والناتج عن عدم معرفتهم باختبار الذكاء من ناحية، وكذلك قلة خبرتهم في التعامل مع الموقف الاختباري بشكلٍ صحيح من ناحيةٍ أخرى، إذ أن تعرضهم للتدريب على الجوانب السابقة قد أسهم في تحقيق التكافؤ بينهم من حيث قدرتهم على التصرّف الصحيح في الموقف الاختباري، وأدى إلى دقة أكبر في تقدير معدلات ذكائهم بسبب استبعاد العوامل العائدة إلى أخطاء القياس، وبالتالي ارتفاع معاملات ثبات الاختبار.

يؤيد هذا التفسير العديد من أدبيات البحث التي تؤكد أن سلوكيات الفرد في الموقف الاختباري، والتي يأتي في مقدمتها مدى توظيفه لاستراتيجيات أداء الاختبار تُعد من مصادر تباين الخطأ في

الدرجات إلى جانب التباين الحقيقي في السمات موضع القياس (Sarnacki, 1979, p. 252, Maspons & Llabre, 1985, p.177, Pohl et al., 2016, p.3, Stenlund et al, 2017, p.4).

ومن جانبٍ آخر من المحتمل أن عملية التدريب قد ساهمت في رفع مؤشرات الصدق، إذ مكّنت المفحوصين من إظهار مستوياتهم الحقيقية في السمة المقاسة بدقة أكبر؛ ذلك لأن عدم استخدامهم لاستراتيجيات أداء الاختبار قد يؤدي في بعض الأحيان، إلى عدم قدرتهم على الإجابة عن بنود الاختبار بما يتفق مع امتلاكهم للمقدار الحقيقي للسمة المقاسة، وبالتالي نقص درجاتهم المستحقة في الاختبار. إذ من المحتمل أن يزيد التدريب على هذه الاستراتيجيات من صدق تعبير درجات الأفراد عن مستواهم الحقيقي في الذكاء، ومن ثم رفع معاملات صدق الاختبار، سواء أكان هذا الارتفاع ذو دلالة إحصائية وعملية أم لم يكن.

تتفق هذه النتيجة مع بحث (رشوان، 2017) الذي أظهرت نتائجه ارتفاع معاملات الصدق التنبؤي والتلازمي للاختبار التحصيلي بتأثير ازدياد إستراتيجيات أداء الاختبار لدى أفراد العينة، كذلك تحسُّن معاملات الثبات ولكن بدرجة أقل من تحسُّن الصدق، وأيضاً مع بحث (Pohl & et al., 2016) الذي أكدت نتائجه أن انخفاض الخصائص القياسية لمقاييس الكفاءة القرائية المرتبطة بسلوكات التخمين العشوائي والاستجابات المفقودة من قبل طلبة صعوبات التعلم، والناتجة عن عدم وجود القدرة الكافية لديهم في التعامل مع بنود الاختبار، ونقص التدريب على المهارات التي تمكنهم من التصرف بشكلٍ صحيح في الموقف الاختباري، في حين تتفق مع نتائج بحث (Maspons & Llabre, 1985) من حيث ارتفاع معامل صدق الاختبار بتأثير هذا المتغير، وتتعارض معها من حيث انخفاض معامل الاتساق الداخلي المحسوب بتأثير متغير التدريب، ولكن على أية حال أظهرت نتائجها أيضاً بأن الانخفاض في قيمة هذا المعامل كان ضئيلاً بحيث لم يكن له أهمية إحصائية.

2- قوة البحث وحدوه:

تمثلت نقاط القوة الأساسية في البحث الحالي بما يأتي:

- اعتماده ثلاثة مناهج بحثية في الإجابة عن أسئلته وتفسير نتائجه.
- اعتماده حجوم عينات كبيرة في مناهجه الثلاثة، إذ بلغ عدد الأبحاث (534) بحثاً أكاديمياً في المنهج الوصفي التحليلي، و(80) بحثاً تجريبياً في منهج التحليل البعدي، وبلغت عينة المنهج التجريبي(125) طالباً وطالبةً، والذين هم جميع طلبة السنة الثانية من اختصاص الإرشاد النفسي.

- إعادة تحليل وحساب الدلالتين العملية والعيادية لفرضيات المجموعات الضابطة والتجريبية التي تضمنتها عينة التحليل البعدي، ومن ثم التكامل في نتائج الدلالات العلمية الثلاث المحسوبة ودراسة العلاقات بينها.
- اعتماد أكثر من طريقة في حساب مؤشرات التغيُّر الموثوق (RCI)، والمفاضلة بين هذه الطرائق تبعاً لنوع البرنامج التدريبي المستخدم والهدف منه.
- استنتاج مؤشرات استخدام أنواع طرائق حساب الدلالة العيادية الملائمة لنوع البرنامج المستخدم وأهدافه.

وعلى الرغم من نقاط القوة السابقة، هناك بعض القيود أو الحدود التي يمكن ملاحظتها في البحث الحالي، منها:

- اقتصار عمليتي التحليل، سواءٌ أكانت المرتبطة بواقع الدلالات العلمية، أو بالتحليل البعدي على الأبحاث التجريبية دون الأبحاث الوصفية التحليلية التي لا تقل أهميةً في ضرورة اعتمادها.
- عدم تمكّن الباحثة من الحصول على أبحاثٍ سابقة حول المقارنة بين الأنواع الثلاث من الدلالات العلمية، ودراسة العلاقات بينها، للوصول إلى نتائج دراسات وأبحاث توضح وتفسر نتائج البحث الحالي.

- مشكلة التضخم في درجات نتائج الأبحاث التجريبية المُستخدمة في عينة التحليل البعدي، حيث أن نسبة قليلة جداً من فرضيات هذه الأبحاث أسفرت عن نتائج غير دالة إحصائياً، الأمر الذي أثر في وضوح النتائج والعلاقات بين الدلالات.
- عدم حساب قيم الدلالة العيادية وفق طريقة (HLM)، لعدم تمكن الباحثة من الحصول على برامج إحصائية تعتمد لغة البايز الأساسية فيها.

3- الصعوبات المتعلقة بإجراء البحث:

واجهت الباحثة مجموعة من الصعوبات لم تمكنها من تنفيذ بحثها وفق مخططٍ منهجي وزمني محددين ولازمين، منها:

- عدم تمكن الباحثة من الحصول على الدرجات الخام القبلية والبعدية للأبحاث التجريبية الأكاديمية جميعها، إما لقدمها، أو لعدم تعاون الكثير من الباحثين الذين أجروا هذه الأبحاث معها.

- اعتماد الباحثين في تصميماتهم البحثية على المجموعات التجريبية فقط، دون المجموعات الضابطة، الأمر الذي قلل من ظهور النتائج غير الدالة في عملية التحليل، والتي كان من الممكن أن تكشف العلاقات بين الدلالات بوضوحٍ أكبر.

4- مقترحات البحث:

في ضوء النتائج السابقة توصل البحث إلى المقترحات الآتية:

- العمل على اعتماد الدلالات العلمية الثلاث في الأبحاث التجريبية بأنواعها المختلفة، واعتماد الدلالتين الإحصائية والعملية في الأبحاث الوصفية التحليلية.
- إجراء دراسات وأبحاث تهدف إلى تحليل واقع استخدام الدلالة العملية إلى جانب الإحصائية في الأبحاث الوصفية التحليلية.
- العمل على تمكين طلبة الدراسات العليا في المجالات والأقسام المختلفة من المعارف والمهارات اللازمة لحساب الأنواع الثلاث من الدلالات العلمية، وتفسير نتائجها.
- إضافة إستراتيجيات أداء الاختبار كجزءٍ مكمل للمنهج الدراسي في المراحل التعليمية المختلفة، لأهميتها في رفع قدرات المتعلمين على التعامل مع الاختبارات في ظل نقص الاهتمام بهذا المتغير على المستوى المحلي، وزيادة الاهتمام به والتركيز على أهميته على المستوى العالمي.
- إجراء برامج تدريبية على إستراتيجيات الأداء في الاختبارات النفسية والتربوية الأخرى، كالاختبارات التحصيلية، واختبارات الشخصية.

خلاصة واستنتاج:

انتهى البحث في ضوء إجراءاته وعيناته ومناهجه إلى جملةٍ من النتائج أكدت بمجملها مدى أهمية الدلالات العلمية في البحوث الأكاديمية – وحتى بحوث المجلات- في الإبلاغ عن النتائج ودورها في دقتها العلمية، مما يؤكد ما انتهت إليه نتائج الأبحاث السابقة، وما اعتمدته بعض الجامعات، والجمعيات العلمية، وأدلة النشر بعدم الاكتفاء بالدلالة الإحصائية كمؤشرٍ وحيد لدقة القرار الإحصائي والبحثي في كافة الأبحاث الكمية واعتماد الدلالة العملية فيها، وعدم كفاية الدلالتين الإحصائية والعملية كمؤشرين وحيدين لدقة القرار الإحصائي والبحثي في الأبحاث الأكاديمية التجريبية.

وضرورة توجيه الباحثين إلى تضمين مؤشرات الدلالات العلمية في إجراءات بحوثهم، ومعالجة بياناتهم بما يكفل الوصول إلى دقةٍ أكبر لنتائجهم البحثية، باعتماد الأساليب والطرائق التي تتناسب مع تصميماتهم البحثية وعيناتهم وبياناتهم وأهدافهم الأساسية، والمفاضلة بينها، إذ لا يمكن اعتماد طريقة إحصائية واحدة في حساب الدلالات العلمية تلائم جميع الأبحاث، وإنما هناك طريقة حسابٍ مثلى لنوعٍ مخصصٍ أو محددٍ من الأبحاث قد لا تكون هي الفضلى لبحثٍ آخر.

ملخص البحث باللغة العربية

واقع الدلالات العلمية في الأبحاث التجريبية وأثر التدريب فيها وفي الخصائص السيكومترية للاختبارات.

دفعت أهمية الدلالات العلمية التي تزايدت التأكيدات على أهمية حسابها في أدبيات البحث وأدلة والنشر والجمعيات العلمية، عدداً كبيراً من الباحثين إلى إخضاعها للدراسة والبحث، سواءٌ أكان ذلك من خلال التمييز بين أنواعها المختلفة أم من حيث رصد واقع استخدام بعضها، إلا أن أياً منها لم يسع إلى تحليل واقع استخدامها جميعاً، أو دراسة العلاقات بين مؤشراتها معاً، ومدى كفاية كل منها في الإبلاغ عن نتائج الأبحاث التجريبية، كما لم يهدف أي منها لدراسة مدى تأثر نتائجها بعوامل تتعلق بالموقف الاختباري أو التدريبي الذي يخضع له المفحوصون. من هنا فإن البحث الحالي يهدف إلى تحليل واقع استخدام الدلالات الثلاث في الأبحاث التجريبية، ودراسة العلاقات بين مؤشراتها، ومدى كفاية كل منها في دقة القرار الإحصائي، كما يسعى لدراسة كيف يمكن أن يؤثر متغير التدريب في نتائجها، وفي رفع أو خفض مؤشرات الخصائص القياسية للاختبار.

تمحورت أسئلة البحث حول تحليل واقع استخدام هذه الدلالات في الأبحاث التجريبية، ودراسة مدى كفايتها في الإبلاغ عن نتائجها، والقدرة التنبؤية بين معاملاتها من جهة، ودراسة أثر متغير التدريب في نتائجها وفي الخصائص القياسية للاختبار من جهةٍ أخرى. بالإضافة إلى المفاضلة بين الطرائق المستخدمة لحساب الدلالة العيادية لتحديد الطرائق الأنسب للبرامج التجريبية المُعدة من قبل الباحثين.

اعتُمد في الإجابة عن أسئلة البحث ثلاثة مناهج أساسية هي: الوصفي التحليلي، والتجريبي، والتحليل البعدي، واقتصر مجتمع البحث على مجتمع إحصائي واحد ارتبط بمنهجي البحث الوصفي التحليلي، والتحليل البعدي، وتمثل بجميع الأبحاث التجريبية في الأقسام الآتية: المناهج وطرائق التدريس، الإرشاد النفسي، علم النفس، والتربية الخاصة، بين عامي 2000–2020 الميلاديين، والفصل الأول من العام الدراسي 2020 – 2021م.

أما عينة البحث فقد تكونت من عينتين أساسيتين، الأولى عينة الرسائل الأكاديمية: تألفت من جميع رسائل الماجستير والدكتوراه التجريبية في أقسام المجتمع الأصلي ذاتها، أي كان حجم العينة مساوياً لحجم المجتمع الأصلي للفترة الزمنية المستهدفة وللأقسام الأربعة، وبنسبة تمثيل 100%.

والثانية هي عينة البحث التجريبية: وهي عينة عرضية متيسرة تكونت من طلبة السنة الثانية من اختصاص علم النفس، بلغ عدد أفرادها (125) طالباً وطالبةً.

توصل البحث إلى النتائج الآتية:

- بلغت نسبة الاختبارات الإحصائية التي جرى التحقق من دلالتها العملية في نتائج الأبحاث التجريبية (29.47%) فقط، وهذا لا يتناسب مع عدد اختبارات الدلالة الإحصائية الكلية المحسوبة فيها استناداً إلى نتائج اختبار كاي مربع، ولم يتم حساب أي من اختبارات الدلالة العيادية في أبحاث عينة التحليل على الإطلاق.
- كلا نوعي القرار المستندين إلى الدلالة الإحصائية قد ارتباطا بتصنيفات الدلالتين العملية والعيادية جميعها بنسبٍ متفاوتة، مما يدل على أن القرار الإحصائي وحده غير قادر على الإبلاغ بدقة عن الأثر الواقعي والحقيقي للبرامج التجريبية من ناحية، وعن التغيُّرات الفعلية التي تحدث على مستوى الأفراد واتجاهها من ناحية أخرى.
- تزامنت التصنيفات المختلفة لحجوم الأثر وفق نتائج التحليل البعدي بجميع تصنيفات الدلالة العيادية، –باستثناء حجم الأثر المنخفض الذي لم يترافق مع تصنيف التحسن– مما يشير إلى عدم كفاية التصنيفات المختلفة لحجوم الأثر في الإبلاغ عن نتائج الأبحاث التجريبية.
- وجود قدرة تنبؤية لمعاملات الدلالة الإحصائية بتصنيفي الإتقان والثبات الدال عيادياً، ووجود قدرة تنبؤية لمعاملات الدلالة العملية بتصنيفات: الإتقان والثبات والتراجع الدال عيادياً، إلا أن كلا هذين النوعين من الدلالة لم يستطع التنبؤ بقيم وتصنيفات الدلالة العيادية جميعها. وبالتالي فإن حساب أي من نوعي الدلالة الإحصائية أو العملية، أو حتى حساب كليهما معاً، غير كافٍ لاستخراج نتائج الأبحاث التجريبية.
- أثر متغير التدريب في قيم الدلالات العلمية لدرجات المجموعة التجريبية في التطبيق البعدي المباشر تأثيراً ذا دلالة إحصائية وعملية وعيادية، دون أن يؤثر في حقيقة الأمر بالقدرات العقلية لأفراد عينة البحث وبمستويات ذكائهم الأساسية، إذ انخفضت وتراجعت درجات نسبة لا بأس بها من أفراد هذه العينة في الاختبار البعدي المؤجل لتعود إلى معدلاتها الأصلية كما في الاختبار القبلي.
- وجود نسب اتفاق دالة إحصائياً وعالية جداً بين القيم والتصنيفات المُستخرجة من الطرائق الثلاث (JT) و(GLN) و(EN)، أما بالنسبة لطريقة (HA) فقد أظهرت نتائجها معدلات مرتفعة جداً من تصنيفات التحسن الدال عيادياً، ترجح الباحثة أن السبب في ذلك يعود إلى التعقيدات الإحصائية في بسط الصيغة الحسابية لهذه الطريقة مما جعلها تتناسب مع طبيعة برامج التدخل العلاجية السريرية.

- أثّر متغير التدريب بارتفاع معامل الصدق المحكي لاختبار الذكاء ارتفاعاً ذا دلالة إحصائية، في حين لم يؤثر في معاملات الصدق التمييزي للاختبار تأثيراً ذا دلالة إحصائية وعملية، ولكنه أثر في بعض قيم الدلالة العيادية لأفراد الفئتين العليا والدنيا.
- أثّر متغير التدريب بارتفاع معاملي ثبات ألفا كرونباخ والتجزئة النصفية ارتفاعاً ذا دلالة إحصائية في الاختبار البعدي المباشر، ولعل هذا يدل على أهميته ودوره في تحسين الخصائص القياسية للاختبار.

وفي ضوء النتائج السابقة تم تقديم مجموعة من المقترحات أهمها:

- العمل على اعتماد الدلالات العلمية الثلاث في الأبحاث التجريبية بأنواعها المختلفة، واعتماد الدلالتين الإحصائية والعملية في الأبحاث الوصفية التحليلية.
- إجراء دراسات وأبحاث تهدف إلى تحليل واقع استخدام الدلالة العملية إلى جانب الإحصائية في الأبحاث الوصفية التحليلية.
- العمل على تمكين طلبة الدراسات العليا في المجالات والأقسام المختلفة من المعارف والمهارات اللازمة لحساب الأنواع الثلاث من الدلالات العلمية، وتفسير نتائجها.
- إضافة إستراتيجيات أداء الاختبار كجزءٍ مكمل للمنهج الدراسي في المراحل التعليمية المختلفة، لأهميتها في رفع قدرات المتعلمين على التعامل مع الاختبارات في ظل نقص الاهتمام بهذا المتغير على المستوى المحلي، وزيادة الاهتمام به والتركيز على أهميته على المستوى العالمي.
- إجراء برامج تدريبية على إستراتيجيات الأداء في الاختبارات النفسية والتربوية الأخرى، كالاختبارات التحصيلية، واختبارات الشخصية.

المراجع

أبو جراد، حمدي. (2013). قوة الاختبارات الإحصائية وحجم الأثر في الأبحاث التربوية المنشورة في مجلة جامعة القدس المفتوحة للأبحاث والدراسات. *مجلة العلوم التربوية والنفسية*، *14*(2)، 350-368.

أبو حطب، فؤاد وعثمان، أحمد. (2003). *التقويم النفسي* (ط.3). القاهرة. مكتبة الأنجلو.

أبو النيل، محمود السيد. (1987). *الإحصاء النفسي والاجتماعي والتربوي*. جامعة عين شمس. دار النهضة العربية للطباعة والنشر.

أبو هاشم، السيد محمد. (2005). مؤشرات التحليل البعدي Meta-Analysis لبحوث فعالية الذات في ضوء نظرية باندورا. *مركز بحوث كلية التربية بجامعة الملك سعود*، (238)،1 – 87.

أبو هاشم، محمد السيد.(2008). النموذج البنائي التنبؤي لمهارات البحث والحكمة الاختبارية لدى طلاب المرحلة الثانوية. *مجلة كلية التربية بالمنصورة، 68* (1)، 210- 270.

بابطين، عادل بن أحمد. (2001). *مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة*. [رسالة ماجستير في علم النفس تخصص إحصاء وبحوث]. كلية التربية. جامعة أم القرى.

البارقي، طلال هيازع. (2012). *واقع الدلالة الإحصائية والدلالة العملية للبحوث المنشورة بمجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية*. [رسالة ماجستير في علم النفس تخصص إحصاء وبحوث]. كلية التربية. جامعة أم القرى.

بحاش، عبد الحق. (2020). الدلالة الإحصائية والعملية لفرضيات الأبحاث النفسية والتربوية. *مجلة الجامع في الدراسات النفسية والعلوم التربوية*، (1)،371-404.

بدوي، عبير علي أحمد. (2018). مقارنة طرق قياس حجم الأثر لبعض الأساليب الإحصائية مع أحجام عينات مختلفة. *مجلة البحث العلمي في التربية*، (19)، 433- 491.

البهي السيد، فؤاد. (2005). *علم النفس الإحصائي وقياس العقل البشري* (ط.3). القاهرة: دار الفكر.

بودلال، فتحية ومووراس، محبوبة. (2021). دراسة تقييمية لقياس الخصائص السيكومترية (الصدق-والثبات) في البحوث العلمية. *مجلة ضياء الجزائرية للبحوث النفسية والتربوية*، *1* (2)، 111-131.

جميلة، زيدان، وبوجرادة محمد. (2017). الخصائص السيكومترية لأدوات القياس النفسي والتربوي والاجتماعي. *مجلة السراج الجزائرية في التربية وقضايا المجتمع، 1* (1)، 209 – 215.

حبيب، مجدي عبد الكريم. (2000). *التقويم والقياس في التربية وعلم النفس* (ط.1). القاهرة: مكتبة النهضة المصرية.

حماد، ديانا فهمي. (2010). *علاقة الحكمة الاختبارية بالأداء الناتج عن اختبار تحصيلي ذي الاختيار من متعدد مبني وفق نموذج راش لدى طالبات كلية التربية الأقسام الأدبية*. [متطلب تكميلي للحصول على درجة الدكتوراه في قسم القياس والتقويم]. جامعة أم القرى.

الخرشة، طه. (2016). أثر طرائق معالجة أثر التخمين على تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدد وفق النظرية الحديثة في القياس. *مجلة جامعة النجاح للأبحاث في جامعة تبوك للعلوم الإنسانية، 30* (12)، 2347-2366.

رشوان، ربيع عبده. (2017). إستراتيجيات أداء الاختبار وتأثيرها على مستوى الاداء والخصائص القياسية للاختبار وبنوده. *المجلة العلمية في جامعة أسيوط، 32* (6)، 112-191.

الزبون، حابس سعد. (2011). *أثر طرق التصحيح لضبط أثر التخمين في الخصائص القياسية لاختبارات الاختيار من متعدد في ضوء نظرية استجابة الفقرة* (ط.1). عمان.

الزعبي، آمال. (2018). بناء مقياس لإستراتيجيات أخذ الاختبار لطلبة جامعة اليرموك. *مجلة الدراسات التربوية والنفسية، 12*(3)، 553- 573.

سليمان، شاهر. (2014). حكمة الاختبار لدى طلاب جامعة تبوك في ضوء متغيرات التخصص والمعدل التراكمي والمستوى الدراسي. *مجلة العلوم التربوية والنفسية، 15*(2)، 246 – 273.

الشاردي، محمد. (2011). *تأثير حجم العينة على قوة الاختبار الإحصائي*. [رسالة ماجستير في كلية التربية قسم علم النفس]. جامعة أم القرى.

الطوطو، رانيه. (2018). *اليقظة العقلية والحكمة الاختبارية كمنبئين بضبط الذات دراسة ميدانية على عينة من طلبة جامعة دمشق*. [رسالة دكتوراه غير منشورة]. جامعة دمشق.

عايش، صباح. (2019). الدلالة الإحصائية في الأبحاث النفسية والتربوية: قضايا منهجية. *المجلة الدولية للأبحاث التربوية، 43* (2)، 86- 101.

عبد الخالق، أحمد محمد. (2005). *استخبارات الشخصية* (ط.3). جامعتا الإسكندرية والكويت: دار المعرفة الجامعية.

عبد الرحمن، سعد. (1998). *القياس النفسي، النظرية والتطبيق*. (ط.3). القاهرة: دار الفكر العربي.

عبد الوهاب، محمد محمود. (2007). *أثر برنامج تدريبي لمهارة حكمة الاختبار على مستوى الأداء التحصيلي وقلق الاختبار لدى عينة من طلبة كلية التربية بالمينا*. [متطلب تكميلي للحصول على درجة الماجستير في التربية]. جامعة المينا.

عز، إيمان وكشيك، منى. (2022). دليل إعداد الرسائل الأكاديمية رسالة الماجستير/ أطروحة الدكتوراه وفقاً لمنظومة APA السابعة. قيد النشر.

عصفور، إيفا. (2018). *الدلالة الإحصائية والعملية وقوة الاختبار الإحصائي في الأبحاث التربوية المنشورة في مجلة جامعة دمشق للعلوم التربوية والنفسية (2004 -2014م)*. [رسالة ماجستير غير منشورة]. جامعة دمشق.

علام، صلاح الدين محمود. (2005). *الأساليب الإحصائية الاستدلالية في تحليل بيانات الأبحاث النفسية والتربوية والاجتماعية: البارامترية واللابارامترية*. القاهرة: دار الفكر العربي.

علام، صلاح الدين محمود. (2002). *القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة: دار الفكر العربي.

فرج، صفوت. (2007). *القياس النفسي*. (ط.6). القاهرة: مكتبة الأنجلو المصرية.

فهمي، محمد شامل. (2005). *الإحصاء بلا معاناة، المفاهيم مع التطبيقات باستخدام برنامج SPSS. الجزء الأول*. الرياض: مكتبة الملك فهد الوطنية.

القرشي، خديجة. (2017). الحكمة الاختبارية وعلاقتها بأنماط التفكير وقلق الاختبار لدى طالبات كلية التربية بجامعة الطائف. *المجلة التربوية*، (47)، 326-387.

قطامي، يوسف. (2013). *النظرية المعرفية في التعلم*. (ط.1). عمان: دار المسيرة للنشر والتوزيع والطباعة.

كانوري، عبد القادر. (2013). *دليل إستراتيجيات خفض قلق الاختبار*. السعودية: المركز الوطني للقياس والتقويم في التعليم العالي.

المالكي، ذياب بن عايض. (2010). *علاقة قلق الاختبار بالحكمة الاختبارية لدى عينة من طلاب المرحلة الثانوية بمحافظة الليث التعليمية*. [متطلب تكميلي للحصول على درجة الماجستير في علم النفس]. جامعة أم القرى.

محمد حسن، عزت عبد الحميد. (2011). *الإحصاء النفسي والتربوي: تطبيقات باستخدام برنامجSPSS*. القاهرة: دار الفكر العربي.

محمد خيري، السيد. (1997). *الإحصاء النفسي*. القاهرة: دار الفكر العربي.

محمد علي، أحمد. (2016). إستراتيجيات الحكمة المعرفية والتقدير الثقافي الذاتي في ضوء بعض المتغيرات لدى معلمي مدينة الخارجة. *مجلة كلية التربية في جامعة الأزهر*، *171* (2)، 255-307.

محمود، خالد محجوب. (2018). بناء برنامج تدريبي مقترح لتنمية الكفايات التعليمية الأدائية لمعلمات التعليم قبل المدرسة أثناء الخدمة بولاية الجزيرة. *المجلة الدولية للدراسات النفسية والتربوية*، *3* (3)، 704 – 719.

المحمودي، محمد. (2019). *مناهج البحث العلمي* (ط.3). صنعاء: دار الكتب.

مخائيل، امطانيوس. (2006). *القياس النفسي* (ج. 1). دمشق: جامعة دمشق.

مخائيل، امطانيوس. (2008). *القياس النفسي* (ج. 2). دمشق: جامعة دمشق.

مراد، صلاح أحمد. (2011). *الأساليب الإحصائية في العلوم النفسية والتربوية والاجتماعية*. القاهرة: مكتبة الأنجلو المصرية.

مراد، صلاح وسليمان، أمين. (2002). *الاختبارات والمقاييس في العلوم النفسية والتربوية، خطوات إعدادها وخصائصها*. القاهرة: دار الكتاب الحديث.

مطلك، فاطمة عباس.(2009). قياس مستوى الحكمة الاختبارية لدى الطلبة المتميّزين والمتميّزات وأقرانهم الاعتياديين في المرحلة الإعدادية. دراسة مقارنة في مركز محافظة نينوى، *مجلة كلية الآداب في الموصل*، (91)، 566- 598.

المنيزل، عبدالله وعايش، غرايبة. (2006). *الإحصاء التربوي*(ط.1). الأردن: دار المسيرة.

موسى، دعاء. (2017). *تقنين اختبار توني – 4 للذكاء غير اللفظي على عينة من العاديين والصم في محافظة دمشق للفئة العمرية "13 – 18"*. [رسالة ماجستير غير منشورة]. جامعة دمشق.

النبهان، موسى. (2004). *أساسيات القياس في العلوم السلوكية*(ط.1).الأردن: دار الشروق للنشر.

النجار، عبد الله. (2006). دراسة تحليلية لقوة الاختبار الإحصائي في الأبحاث الإدارية المنشورة في مجلة الإدارة العامة. *المجلة العلمية لجامعة الملك فيصل للعلوم الإنسانية والإدارية*، *7* (2)، 261-293.

نصار، يحيى. (2016). الدلالة الإكلينيكية للبحوث النفسية والتربوية المستخدمة للتصاميم التجريبية "دراسة تحليلية". *مجلة الدراسات التربوية والنفسية في جامعة السلطان قابوس*، *11* (2)، 352-369.

وهبي، إبراهيم، الحدابي، وداد، والخروصي، حسين.(2020). درجة التوافق بين الدلالة الإحصائية ومؤشرات الدلالة العملية في الاختبارات الإحصائية برسائل الماجستير بالجامعات العمانية. *مجلة التنمية البشرية والتعليم للأبحاث التخصصية*، *6* (1)، 101- 130.

يوسف، علا. (2015). دراسة الخصائص القياسية لاختبار المصفوفات المتتابعة المعيارية لجون رافن (النسخة الموازية) على عينة من طلاب المرحلة الجامعية في جامعة دمشق. *مجلة جامعة البعث*، *37* (3)،11 – 34.

Akbaryan, F. (2013). *Effect Size*. Department of Rehabilitation Medicine, university of Alberta.

American Psychological Association (2010). *Publication manual of the American Psychological Association*. (6th Ed.).

American Psychological Association (2020). *Publication manual of the American Psychological Association*. (7th Ed.).

Anthony, J, Viera, MD, Joanne, M. (2005). Understanding Inter observer Agreement: The Kappa Statistic. *Fam Med Journal*, *37*(5), 360-3

Atkins, D., Bedics, J., & McGlincheym J. (2005). Assessing Clinical Significance: Dose It Matter Which Method We Use? *Journal of Counseling and Clinical Psychology, 73*(5), 982-989.

Balkin, R., & Lenz, A. (2021). Contemporary Issues in Reporting Statistical, Practical, and Clinical Significance in counseling Research. *Journal of Counseling & Development*, Volume 99, 227-237.

Banks, T., & Eaton, I. (2014). Improving Test-Taking Performance of Secondary At-Risk Youth and Student with Disabilities. *Preventing School Failure*, *58*(4), 207-213.

Barry, C., & Finney, S. (2009). Exploring Change in Test – Taking Motivation. *Paper Presented at Annual Meeting of the Northeastern Educational Research Association*.

Bauer, S., Lambert, M., & Nielsen, S. (2004) Clinical Significance Methods: A Comparison of Statistical Techniques. *Journal of personality assessment*, *82*(1), 60-70.

Becker, L. (2000). *Effect Size (ES)*. From http://web.uccs.edu/lbecker/Psy590/es.htm

Bicak, B. (2013). Scale for Test Preparation and Test Taking Strategies. *Educational Consultancy and Research Center*. *13*(1), 279-289.

Carver, P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, *61*(4), 287-292.

Chambless, L., & Hollon, D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7-19.

Christopher, A. (2006). Practical Significance: The Use of Effect Sizes in School Counseling Research. *Professional School Counseling*, *9* (4), 401-411.

Coe, R. (2002). It ' s the Effect Size , What effect size is and why it is important. *Paper presented at the Annual Conference of the British Educational Research Association*, University of Exeter, England, 12-14.

Cohen, A. (2006). The Coming of Age of Research on Test- Taking Strategies. *Language Assessment Quarterly* , *3*(4), 307-331.

Cohen, J. (1992). A Power Primer. *American Psychological Association. 112* (1), 155-159.

Cohen, J. (1988). *statistical power analysis for the behavioral Sciences*, (2nd ed) . Hillsdale: New Jersey.

Dodeen, H. (2009). Test-Related Characteristics of OUAEU student: Test-Anxiety, Guessing, Attitudes toward tests, and Cheating. *Journal of Faculty of Education, 26*(1), 31–66.

Dodeen, H. (2014). Test-taking skills of secondary students: the relationship With motivation, attitudes, anxiety and attitudes towards tests. *South African Journal of Education*, *34*(2), 1-18.

Dodeen, H. (2015). Teaching Test- Taking Strategies: Importance and Techniques. *Psychology Research*, *5*(2), 108-113.

Eisen, s., Ranganathan, G., & Seal, P. (2007).Measuring Clinically Meaningful Change Following Mental Health Treatment. *Journal of Behavioral Health Services & Research, 34* (3), 272- 289.

Ferguson, J. (2009). An Effect Size Primer: A Guide for Clinicians and Researchers, *Texas A&M International University*, *40* (5), 532-538.

Ferrier, E., Lovett, J.,& Jordan, A. (2011). Construct - Irrelevant Variance in Achievement Test Scores*, department of psychology*, *Elmira college*.

Fritz, O., Morris, E., & Richler, J. (2012). Effect Size Estimates: Current Use, Calculations, and Interpretation, *Journal of Experimental Psychology*, *141* (1), 2–18.

Gallo, A. (2016). A Refresher on Statistical Significance. Analytics And Data Science, from: https://hbr.org/2016/02/a-refresher-on-statistical-significance

Ghafournia, N., & Afghari, A. (2013). The Interaction between Cognitive Test-Taking Strategies, Reading Ability, and Reading Comprehension Test Performance of Iranian EFL Learners. *English Language Teaching*, *6*(8), 139-150.

Hansen, A., Oien, R., Volkmar, F., Shic, F., & Cicchetti, D. (2018). Enhancing the understanding of clinically meaningful results: A clinical research perspective, *Psychiatry Research*, *270* (2018), 801–806.

Hedges, L.(1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, *6* (2), 107-128.

Hong, E, Sas, M and Sas, J. (2006). Test – taking Strategies of High and Low Mathematics Achievers, University of Nevada, Las Vegas, *Journal of Education Research*, *99*(3), 144- 155.

Hopkins, w. (2014). *A New View of Statistical*. On line Book.

Hsu, L. (1998). A comparison of three methods of identifying reliable and clinically significant client changes: commentary on Hageman and Arrindell. *Behavior Research and Therapy*, *37* (1999), 1195-1202.

Huston, L.(1993). Meaningfulness, Statistical Significance, Effect Size, and Power Analysis: A General Discussion with Implications for MANOVA. *APaper Presented at Annual Meeting of the Mid-South Educational Research Association, P.P* 9-12.

Jacobson, N., & Truax, P. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research, *Journal of Consulting and Clinical Psychology*, *59*(1),12-19.

Jacobson, N., Roberts, L., Berns, S., & McGlinchey , J. (1999). Methods for Defining and Determining the Clinical Significance of Treatment Effects, *Journal of Consulting and Clinical Psychology*, *67*(3), 300-307.

Kazdin, A. (1999). The Meanings and Measurement of Clinical Significance, *Journal of Consulting and Clinical Psychology*, *67*(3), 332-339.

Kendall, P. (1999). Clinical Significance. *Journal of Consulting and Clinical Psychology*, *67*(3), 283-284.

Khosravani, E., Zanjani, E., & Najafabadi, Z. (2017). The Relationship between Self-Esteem and Test-Taking Strategies and Their Relation with Test Performance of Iranian EFL Learners. *Journal of Applied Linguistics and Language Research*. *4*(4), 87-97.

Kirk, R. (1996). Practical Significance: A concept Whose Time Has Come. *Educational and Psychological Measurement*, *56*(5), 746-759.

Maspons, M., & Llabre, M. (1985). The Influence of Training Hispanics in Test Taking on Psychometric Properties of a test. *Journal for Research in Mathematics Education*, *16* (3), 177-183

McLean, E., & Ernest, M. (1997). Has Testing for Statistical Significance Outlived Its Usefulness? *Paper Presented at Annual Meeting of the Mid-South Educational Research Association, November 12*(1), 1-21.

McMillan, J., & Foley, J. (2011). Reporting and Discussing Effect Size: Still the Road Less Traveled? *Practical Assessment, Research & Evaluation*, *16*(14), 1 – 12.

Millman, J., Bishop,C., & Eble, R. (1965). An analysis of test wiseness. *Educational and Psychological Measurement*, *25*(3), 707-726.

Nakagawa, S., & Cuthill, C. (2007). Effect size, Confidence Interval and Statistical Significance: A Practical Guide For Biologists. *Biological Reviews, 82*(2). 591–605.

Nezu, A., & Maguth, C. (2008). *Evidence-Based Outcome Research A Practical Guide to Conducting Randomized Controlled Trials for Psychosocial Intervention*, On line Book.

Peng, Y. (2005). *Test preparation strategies and test taking strategies use in Chinese high school students*. A thesis submitted in partial fulfillment of the requirements for the Master of Science Degree in Educational Psychology. Department of Educational Psychology. University of Nevada, Las Vegas.

Peterson, S. (2008). Clinical Significance: "Clinical" Significance and "Practical" Significance Are Not The Same Things, *Paper presented at the annual meeting of the Southwest Educational Research Association, Texas University.*

Pintea, S. (2010). The Relevance Of Results In Clinical Research: Statistical, Practical, And Clinical Significance, *Journal of Cognitive and Behavioral Psychotherapies*, *10*(1), 101-114.

Pohl, S., Südkamp, A., Hardt, K., Carstensen, CH., & Weinert, S. (2016). Testing Students with Special Educational Needs in Large-Scale Assessments – Psychometric Properties of Test Scores and Associations with Test Taking Behavior. *Front Psychol Journal*, *23*(7) 154-160.

Richardson, J. (1996). Measures of effect size. *Behavior Research Methods, Instruments, & Computers*, *28* (1),12-22.

Ronk, F. (2014). *Can we trust clinical significance when evaluating clinical outcomes? An investigation of the reliability, validity, flexibility, and*

stability of clinical significance methodology, thesis presented for the degree of Doctor of Philosophy at The University of Western Australia.

Ronk, F., Hook, G., & Page, A. (2016).Validity of clinically significant change classifications yielded by Jacobson-Truax and Hageman-Arrindell methods, *BMC Psychiatry*, *16*(187), 1-9.

Sarnacki, E. (1979). An examination of Test – Wiseness in the cognitive test domain. *Review of Educational Research*. *49* (2), 252- 279.

Schober, P., Bossers, S., Schwarte, L. (2018). Statistical Significance Versus Clinical Importance of Observed Effect Sizes: What Do P Values and Confidence Intervals Really Represent? *Anesthesia & Analgesia*, *126*(3), 1068-1072.

Schneider, w. (2011). Caveats for using statistical significance tests in research assessments. *Danish Centre for Studies in Research and Research Policy*, Aarhus N, Denmark, *1*(1),1-26

Shada, K. (2013). *Replication of a comparison of statistical techniques using clinical significance methods*, Master's thesis, Pacific University.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education*, *61*(4), 293-316.

Stenlund,T., Eklöf, H., & Lyrén, E. (2017). Group differences in test-taking behavior: an example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice*, *24*(1), 4–20.

Takallou, F., Vahdany, F., Araghi, S. & Tabrizi, A. (2016). Test-Taking Strategy Instruction In The National University Entrance Examination (NUEE) English Section. *International Journal of Asian Social Science*, *6*(3), 177-191.

Téllez, A., Garcíab, C., & Verdugoc, V. (2015). Effect size, confidence intervals and statistical power in psychological research, *Psychology in Russia: State of the Art*, *8*(3), 27-46.

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research articles :A simplified methodology.
from http://work-learning.com/effect_sizes.htm.

Thompson, B., & Haase, T. (2004). How to Estimate and Interpret

Various Effect Sizes. *Journal of Counseling Psychology*, *51*(4), 473–481.

Thompson, B. (2002). "Statistical," "Practical," and "Clinical": How Many Kinds of Significance Do Counselors Need to Consider? *Journal of Counseling and Development*, Vol. *80*, 64–71.

Thompson, B. (1999). If Statistical Significance Tests Are Broken/Misused, What Practices Should Supplement or Replace Them? *Theory & Psychology*, *9*(2), 165–181.

Tomczak [1], M., & Tomczak[2], E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size, *Trends in Sport Sciences*,*1*(21), 19-25.

Wise, E. (2004). Methods for Analyzing Psychotherapy Outcomes: A Review of Clinical Significance, Reliable Change, and Recommendations for Future Directions, *Journal of Personality Assessment*, *82*(1), 50–59.

Xitao, F. (1999). Statistical Significance, and effect size: two sides of a coin, *APaper presented at the Annual Meeting of the American Evaluation Association, Utah State University*.

Zahra, D., & Hedge, C. (2010). The Reliable Change Index: Why isn't it more popular in academic psychology? *The British Psychological Society*, Issue *76*, 14 – 19.

الملاحق

ملحق (1)

أسماء الخبراء المحكمين

	اسم المحكم	الصفة العلمية	الاختصاص العلمي
1	أ.د. محمد وحيد صيام	الأستاذ في قسم المناهج وطرائق التدريس	تقنيات التعليم
2	أ.د. أمينة رزق	الأستاذة في قسم علم النفس	علم النفس التربوي
3	أ.د. علي النحيلي	الأستاذ في قسم علم النفس	علم النفس التربوي
4	أ.د. طاهر سلوم	الأستاذ في قسم المناهج وطرائق التدريس	طرائق تدريس الدراسات الاجتماعية
5	أ.د. رمضان درويش	الأستاذ في قسم القياس والتقويم	إحصاء تطبيقي
6	د .منذر بوبو	الأستاذ المساعد في قسم القياس والتقويم في كلية التربية – جامعة تشرين	الإحصاء والمعلوماتية ونظرية اتخاذ القرار
7	د. محمد شيخو	الأستاذ المساعد في قسم المناهج وطرائق التدريس	طرائق تدريس الزراعة
8	د. عزيزة رحمة	الأستاذة المساعدة في قسم القياس والتقويم	الإحصاء في التربية وعلم النفس
9	د. منال الشيخ	الأستاذة المساعدة في قسم الإرشاد النفسي	الإرشاد النفسي
10	د. عالية الرفاعي	الأستاذة المساعدة في قسم التربية الخاصة	التربية الخاصة (الإعاقة السمعية)
11	د. منار طومان	الأستاذة المساعدة في قسم القياس والتقويم في جامعة حلب	التقويم والقياس والإحصاء في التربية وعلم النفس
12	د. ياسر جاموس	المدرّس في قسم القياس والتقويم	القياس والتقويم التربوي والنفسي
13	د. رنا قوشحة	المُدرّسة في قسم القياس والتقويم	القدرات العقلية
14	د. اعتدال عبد الله	المُدرّسة في قسم القياس والتقويم	القياس والتقويم التربوي والنفسي
15	د. خلود الجزائري	المُدرّسة في قسم المناهج وطرائق التدريس	مناهج وطرائق التدريس

ملحق (2)

نموذج جدول التحليل

الدلالة المعيارية	قيمة مؤشر (RCI)												
	قانون الدلالة المعيارية المستخدم												
نوع قانون حجم الأثر وتصنيفه	مرتفع												
	متوسط												
	ضعيف												
الدلالة الإحصائية	غير دال												
	دال												
الاختبار الإحصائي المستخدم													
فرضيات الدراسة													
عدد أفراد العينة													
القسم													
العام													
نوع البحث	دكتوراه												
	ماجستير												

ملحق (3)
نسب الأبحاث التجريبية

العام	نوع المنهج		المناهج		علم النفس		الإرشاد النفسي		التربية الخاصة		المجموع
			ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	
من 2000 إلى 2006	تجريبي	العدد	15	14	3	6	1	1	1	1	42
	وصفي تحليلي	العدد	36	8	18	10	8	0	2	1	83
	المجموع		51	22	21	16	9	1	3	2	125
	النسبة المئوية للأبحاث التجريبية		%29.41	%63.64	%14.29	%37.5	%11.11	%100	33.33 %	%50	%33.6
2007	تجريبي	العدد	4	3	-	-	1	-	6	0	14
	وصفي تحليلي	العدد	4	6	-	1	2	-	19	0	32
	المجموع		8	9	-	1	3	-	25	0	46
	النسبة المئوية للأبحاث التجريبية		%50	%33.33	-	-	%33.33	-	%24	-	30.43 %
2008	تجريبي	العدد	5	1	-	1	1	0	0	1	9
	وصفي تحليلي	العدد	10	2	-	3	4	1	5	2	27
	المجموع		15	3	-	4	5	1	5	3	36
	النسبة المئوية للأبحاث التجريبية		%33.33	%33.33	-	%25	%20	-	-	33.33 %	%25
2009	تجريبي	العدد	4	6	1	0	1	1	2	0	15
	وصفي تحليلي	العدد	12	3	8	5	2	0	3	0	33
	المجموع		16	9	9	5	3	1	5	0	48
	النسبة المئوية للأبحاث التجريبية		%25	%66.66	%11.11	-	%33.33	%100	%40	-	31.25 %

العام	نوع المنهج		المناهج		علم النفس		الإرشاد النفسي		التربية الخاصة		المجموع
			ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	
2010	تجريبي	العدد	7	5	0	3	0	4	6	1	26
	وصفي تحليلي	العدد	15	2	5	1	10	1	10	2	46
	المجموع		22	7	5	4	10	5	16	3	72
	النسبة المئوية للأبحاث التجريبية		%31.82	%71.43	-	%75	-	%80	%37.5	33.33 %	36.11 %
2011	تجريبي	العدد	14	17	1	0	1	2	6	2	43
	وصفي تحليلي	العدد	8	2	5	4	10	1	9	0	39
	المجموع		22	19	6	4	11	3	15	2	82
	النسبة المئوية للأبحاث التجريبية		%63.64	%89.47	%16.67	-	%9.1	66.67 %	%40	%100	52.44 %
2012	تجريبي	العدد	14	12	1	2	0	4	23	2	58
	وصفي تحليلي	العدد	9	2	10	4	7	0	16	4	52
	المجموع		23	14	11	6	7	4	39	6	110
	النسبة المئوية للأبحاث التجريبية		%60.87	%85.71	%9.1	%33.33	-	%100	58.97 %	33.33 %	52.72 %
2013	تجريبي	العدد	12	11	0	0	0	1	5	4	33
	وصفي تحليلي	العدد	17	1	17	3	13	3	5	0	59
	المجموع		29	12	17	3	13	4	10	4	92
	النسبة المئوية للأبحاث التجريبية		%41.38	%91.67	-	-	-	%25	%50	%100	35.87 %

العام	نوع المنهج		المناهج		علم النفس		الإرشاد النفسي		التربية الخاصة		المجموع
			ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	
2014	تجريبي	العدد	11	11	1	1	-	4	10	4	42
	وصفي تحليلي	العدد	7	2	16	4	4	3	1	0	37
	المجموع		18	13	17	5	4	7	11	4	79
	النسبة المئوية للأبحاث التجريبية		%61.11	%84.61	%5.88	%20	-	57.14 %	90.91 %	%100	53.16 %
2015	تجريبي	العدد	13	14	1	-	1	3	9	2	43
	وصفي تحليلي	العدد	5	3	15	-	11	4	7	1	46
	المجموع		18	17	16	-	12	7	16	3	89
	النسبة المئوية للأبحاث التجريبية		%72.22	%82.35	%6.25	-	%8.33	42.86 %	56.25 %	66.67 %	48.31 %
2016	تجريبي	العدد	5	9	2	3	0	1	9	4	33
	وصفي تحليلي	العدد	3	2	17	2	12	2	6	1	45
	المجموع		8	11	19	5	12	3	15	5	78
	النسبة المئوية للأبحاث التجريبية		%62.5	%81.82	%10.52	%60	-	33.33 %	%60	%80	42.31 %
2017	تجريبي	العدد	13	13	1	3	1	2	4	2	39
	وصفي تحليلي	العدد	4	1	13	4	3	0	1	0	26
	المجموع		17	14	14	7	4	2	5	2	65
	النسبة المئوية		%76.47	%92.86	%7.14	%42.86	%25	%100	%80	%100	%60

العام	نوع المنهج		المناهج		علم النفس		الإرشاد النفسي		التربية الخاصة		المجموع
			ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	ماجستير	دكتوراه	
2018	تجريبي	العدد	18	15	1	4	1	2	4	5	50
	وصفي تحليلي	العدد	10	2	9	7	7	1	0	1	37
	المجموع		28	17	10	11	8	3	4	6	87
	النسبة المئوية للأبحاث التجريبية		%64.28	%88.23	%10	%36.36	%12.5	66.67 %	%100	83.33 %	57.47 %
2019	تجريبي	العدد	7	19	3	4	2	3	5	4	47
	وصفي تحليلي	العدد	3	2	5	6	5	1	0	1	23
	المجموع		10	21	8	10	7	4	5	5	70
	النسبة المئوية للأبحاث التجريبية		%70	%90.47	%37.5	%40	%28.57	%75	%100	%80	67.14 %
2020	تجريبي	العدد	10	17	0	3	1	2	3	4	40
	وصفي تحليلي	العدد	2	0	4	3	5	2	2	0	18
	المجموع		12	17	4	6	6	4	5	4	58
	النسبة المئوية للأبحاث التجريبية		%83.33	%100	-	%50	%16.67	%50	%60	%100	68.96 %
المجموع	تجريبي	العدد	152	167	15	30	11	30	93	36	534
	وصفي تحليلي	العدد	145	38	142	57	103	19	86	13	603
	المجموع		297	205	157	87	114	49	179	49	1137
	النسبة المئوية للأبحاث التجريبية		%51.18	%81.46	%9.55	%34.48	%9.65	61.22 %	51.95 %	73.47 %	46.96 %
	مجموع الأبحاث التجريبية		319		45		41		129		534
	مجموع الأبحاث الوصفية التحليلية		183		199		122		99		603
	المجموع الكلي		502		244		163		228		1137
	النسبة المئوية للأبحاث التجريبية		%63.54		%18.44		%25.15		%56.58		46.96 %

ملحق (4)
نسب قوانين حجم الأثر

		المناهج وطرائق التدريس			علم النفس			الإرشاد النفسي			التربية الخاصة		
		العدد الكلي للاختبارات الإحصائية	عدد الاختبارات التي جرى التحقق من دلالتها العملية	نسبتها المئوية	العدد الكلي للاختبارات الإحصائية	عدد الاختبارات التي جرى التحقق من دلالتها العملية	نسبتها المئوية	العدد الكلي للاختبارات الإحصائية	عدد الاختبارات التي جرى التحقق من دلالتها العملية	نسبتها المئوية	العدد الكلي للاختبارات الإحصائية	عدد الاختبارات التي جرى التحقق من دلالتها العملية	نسبتها المئوية
2006-2000	ماجستير	121	16	13.2 %	30	0	-	5	0	0	13	0	-
	دكتوراه	143	3	2.09 %	76	0	-	8	1	12.5 %	6	0	-
	المجموع	264	19	7.19 %	106	0	-	13	1	7.69 %	19	0	-
2007	ماجستير	32	3	9.38 %	-	-	-	12	-	-	20	0	-
	دكتوراه	34	0	-	-	-	-	-	-	-	-	-	-
	المجموع	66	3	4.54 %	-	-	-	12	-	-	20	0	-
2008	ماجستير	54	4	7.41 %	-	-	-	7	0		0	0	-
	دكتوراه	28	0	-	8	2	%25	-	-	-	2	0	-
	المجموع	82	4	4.88 %	8	2	%25	7	-	-	2	0	-
2009	ماجستير	45	4	8.89 %	10	0	-	4	0	-	23	1	4.35 %
	دكتوراه	117	0	-	-	-	-	5	0	-	0	0	-
	المجموع	162	4	2.46 %	10	0	-	9	0	-	23	1	4.35 %
2010	ماجستير	94	11	11.7 %0	-	-	-	-	-	-	34	0	-
	دكتوراه	72	25	34.7 %2	49	5	10.2 %	48	10	20.8 %3	9	0	-
	المجموع	166	36	21.6 %9	49	5	10.2 %	48	10	20.8 %3	43	0	-
2011	ماجستير	161	24	14.9 %1	6	2	33.3 %3	16	0	-	27	5	18.5 %2
	دكتوراه	301	75	24.9 %2	-	-	-	11	3	27.2 %7	17	0	-
	المجموع	462	99	21.4 %3	6	2	33.3 %	27	3	11.1 %1	44	5	11.3 %6

		المناهج وطرائق التدريس			علم النفس			الإرشاد النفسي			التربية الخاصة		
		عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية	عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية	عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية	عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية
2012	ماجستير	183	57	31.1%5	8	3	37.5%	-	-	-	120	22	18.3%3
	دكتوراه	264	70	26.5%2	24	6	%25	73	8	10.9%6	23	2	8.69%
	المجموع	447	127	28.4%1	32	9	28.1%2	73	8	10.9%6	143	24	16.7%8
2013	ماجستير	151	4	2.65%	-	-	-	-	-	-	22	5	22.7%3
	دكتوراه	232	104	44.8%3	-	-	-	9	3	33.3%3	19	7	36.8%4
	المجموع	383	108	28.2%0	-	-	-	9	3	33.3%3	41	12	29.2%7
2014	ماجستير	108	32	29.6%3	8	0	-	-	-	-	43	13	30.2%3
	دكتوراه	123	44	35.7%7	11	5	45.4%5	35	10	28.5%7	20	8	%40
	المجموع	231	76	32.9%0	19	5	26.3%2	35	10	28.5%7	63	21	33.3%3
2015	ماجستير	87	39	44.8%3	30	0	-	4	1	%25	29	14	48.2%8
	دكتوراه	159	42	26.4%2	-	-	-	25	6	%24	12	3	%25
	المجموع	246	81	32.9%3	30	0	-	29	7	24.1%4	41	17	41.4%6
2016	ماجستير	30	13	43.3%3	40	12	%30	-	-	-	37	19	51.3%5
	دكتوراه	128	48	37.5%	24	8	33.3%3	3	1	33.3%3	22	4	18.1%8
	المجموع	158	61	38.6%1	64	20	31.2%5	3	1	33.3%3	59	23	38.9%8
2017	ماجستير	74	33	44.5%9	8	2	%25	12	4	33.3%3	19	4	21.0%5
	دكتوراه	133	49	36.8%4	36	13	36.1%	11	6	54.5%4	10	6	%60
	المجموع	207	82	39.61	44	15	34.1%	23	10	43.4%8	29	10	34.4%8

		المناهج وطرائق التدريس			علم النفس			الإرشاد النفسي			التربية الخاصة		
		عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية	عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية	عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية	عدد الاختبارات الإحصائية المحسوبة	عدد الاختبارات التي جرى حساب حجم الأثر لها	نسبتها المئوية
2018	ماجستير	123	60	48.78%	6	4	66.6%	12	5	41.67%	15	8	53.33%
	دكتوراه	183	67	36.61%	40	16	%40	15	11	73.33%	26	14	53.85%
	المجموع	306	127	41.50%	46	20	43.48%	27	16	59.26%	41	22	53.66%
2019	ماجستير	53	34	64.15%	36	18	%50	23	9	39.13%	20	11	%55
	دكتوراه	237	128	54.01%	65	45	69.23%	32	14	43.75%	27	15	55.56%
	المجموع	290	162	55.86%	101	63	62.38%	55	23	41.82%	47	26	55.32%
2020	ماجستير	55	37	67.27%	-	-	-	2	1	%50	12	7	58.33%
	دكتوراه	158	81	51.27%	27	17	62.96%	20	15	%75	15	9	%60
	المجموع	213	118	55.40%	27	17	62.96%	22	16	72.73%	27	16	59.26%
المجموع	ماجستير	1371	371	27.06%	182	41	22.53%	97	20	20.62%	434	109	25.11%
	دكتوراه	2312	736	31.83%	360	117	35.5%	295	88	29.83%	208	68	32.69%
	المجموع	3683	1107	30.05%	542	158	29.15%	392	108	27.55%	642	177	27.57%

ملحق (5)

مثال لحساب مؤشر التغيُّر الموثوق

يُوضح هذا الملحق مثالاً محسوباً لدرجات فردٍ واحدٍ على مقياس المخاوف النوعية، ويبيّن بالتفصيل طريقة حساب مؤشر التغيُّر الموثوق وفق الصيغ الرياضية الأربع لطرائق حساب الدلالة العيادية:

المعلومات ذات الصلة المتصلة بدرجات الفرد والعينة التجريبية التي سُحب منها:

درجة الفرد في الاختبار القبلي= 125

درجة الفرد في الاختبار البعدي "بعد المعالجة التجريبية"= 97

المتوسط القبلي للمجموعة التجريبية = 102.47

المتوسط البعدي للمجموعة التجريبية = 96.69

الانحراف المعياري القبلي= 9.85

الانحراف المعياري البعدي= 13.39

معامل ثبات ألفا كرونباخ =0.632

الخطأ المعياري للقياس (SE) = 5.91

معامل الارتباط بين الدرجات القبلية والدرجات البعدية = 0.48

Jacobson – Truax Method (JT):

$$\mathbf{RCI} = \frac{(X_2 - X_1)}{\sqrt{2S_E^2}} = \frac{97-125}{\sqrt{2\times 5.91^2}} = \frac{-28}{8.36} = -3.34$$

يُصنَّف الفرد بأنه متحسناَ تحسناً دالاً عيادياً ، لأن قيمة مؤشر التغيُّر الموثوق لديه أدنى من الدرجة الحدية (1.96−)

Gulliksen- Lord- Novick Method (GLN):

$$\mathbf{RCI} = \frac{[X_{post} - M_{pop}] - r_{xx}[X_{pre} - M_{pop}]}{SD_{pop}\sqrt{1-r_{xx}^2}} = \frac{(97-102.47) - 0.632\ (125-102.47)}{9.85\sqrt{1-0.632^2}} = \text{-2.56}$$

يُصنَّف الفرد بأنه متحسناً تحسناً دالاً عيادياً ، لأن قيمة مؤشر التغيُّر الموثوق لديه أدنى من الدرجة الحدية (1.96−)

Edwards- Nunnally Method (EN):

$$\text{RCI}= \left[r_{xx}\left(X_{pre} - M_{pre}\right) + M_{pre}\right] \pm 2SD_{pre}\sqrt{1 - r_{xx}}$$

$$= [0.632(125 - 102.47) + 102.47] \pm 2 \times 9.85\sqrt{1 - 0.632}$$

حيث: الحد الأعلى = 116.7 + 11.82 = 128.52

والحد الأدنى = 116.7 + 11.82 = 104.88

وبذلك يُصنَّف الفرد بأنه متحسناً تحسناً دالاً عيادياً ، لأن درجته في الاختبار البعدي (97) أدنى من الحد الأدنى لفترة الثقة المحسوبة.

Hageman-Arrindell Method (HA):

$$\text{RCI}= \frac{(X_{post} - X_{pre})r_{dd} + (M_{post} - M_{pre})(1 - r_{dd})}{(\sqrt{r_{dd}})\left(\sqrt{2S_E^2}\right)} =$$

$$= \frac{(97-125)0.52+(96.69-102.47)(1-0.52)}{\sqrt{0.52}\sqrt{2\times 5.91^2}} = \frac{-17.33}{6.01} = -2.88$$

يُصنَّف الفرد بأنه متحسناً تحسناً دالاً عيادياً ، لأن قيمة مؤشر التغيُّر الموثوق لديه أدنى من الدرجة الحدية (1.65−)

$$r_{dd} = \frac{SD_{pre}^2 \times r_{pre} + SD_{post}^2 \times r_{post} - 2SD_{pre}SD_{post}r_{pre*post}}{SD_{pre}^2 + SD_{post}^2 - 2SD_{pre}SD_{post}r_{pre*post}}$$

$$r_{dd} = \frac{9.85^2 \times 0.64 + 13.39^2 \times 0.80 - 2 \times 9.85 \times 13.39 \times 0.48}{9.85^2 + 13.39^2 - 2 \times 9.85 \times 13.39 \times 0.48} = 0.52$$

$$r_{pre} = \frac{SD_{pre}^2 - S_E^2}{SD_{pre}^2} = \frac{9.85^2 - 5.91^2}{9.85^2} = 0.64$$

$$r_{post} = \frac{SD_{post}^2 - S_E^2}{SD_{post}^2} = \frac{13.39^2 - 5.91^2}{13.39^2} = 0.80$$

ملحق (6)

قيم الدلالة العيادية للأبحاث المُجمّعة

رقم البحث	القسم	نوع البحث	نوع المجموعة	عدد أفرادها	الخطأ المعياري للفرق بين الدرجات القبلية والبعدية	نقطة القطع
1	المناهج وطرائق التدريس	دكتوراه	تجريبية	66	2.39	34.74
			ضابطة	66		
2	المناهج وطرائق التدريس	ماجستير	تجريبية1	33	4.61	30.44
			تجريبية2	30		
			تجريبية3	51		
			ضابطة	33		
3	الإرشاد النفسي	دكتوراه	تجريبية	15	10.40	144.97
			ضابطة	15		
4	المناهج وطرائق التدريس	ماجستير	تجريبية	50	12.40	35.4
			ضابطة	50		
5	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	41	5.81	60.18
			تجريبية 2	41	16.49	171.08
			تجريبية 3	41	7.37	83.49
6	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	60	8.31	54.29
			ضابطة	60		
7	تربية الطفل	ماجستير	فقط تجريبية	30	2.74	16.7
8	المناهج وطرائق التدريس	ماجستير	تجريبية	35	3.95	27.41
			ضابطة	39		
9	المناهج وطرائق التدريس	ماجستير	تجريبية 1	65	3.03	21.17
			ضابطة 1	37		
			تجريبية 2	60	3.35	16.35
			ضابطة 2	27		
10	رياض الأطفال	ماجستير	تجريبية	40	6.93	33.25
11	المناهج وطرائق التدريس	دكتوراه	تجريبية	70	10.1	44.42
			ضابطة	70		
12	المناهج وطرائق التدريس	دكتوراه	تجريبية	64	7.17	76.39
			ضابطة	61		

الرقم	القسم	نوع البحث	نوع المجموعة	عدد أفرادها	الخطأ المعياري للفرق بين الدرجات القبلية والبعدية	نقطة القطع
13	المناهج وطرائق التدريس	ماجستير	تجريبية	10	4.88	38.84
			ضابطة	10		
14	المناهج وطرائق التدريس	دكتوراه	تجريبية	66	1.17	2.52
			ضابطة	66		
15	المناهج وطرائق التدريس	دكتوراه	تجريبية	123	1.19	5.49
			ضابطة	124		
16	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	81	2.84	16.3
			تجريبية 2	160		
			ضابطة	106		
17	المناهج وطرائق التدريس	دكتوراه	تجريبية	50	4.91	21.64
			ضابطة	50		
18	المناهج وطرائق التدريس	دكتوراه	تجريبية	68	6.80	29.9
			ضابطة	68		
19	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	122	1.12	8.56
			ضابطة 1	100		
			تجريبية 2	18	7.58	68.61
			ضابطة 2	17		
20	المناهج وطرائق التدريس	دكتوراه	تجريبية	185	4.04	20.38
			ضابطة	171		
21	التربية الخاصة	دكتوراه	تجريبية	10	6.02	47.82
			ضابطة	10		
22	المناهج وطرائق التدريس	دكتوراه	تجريبية	158	11.84	54.89
			ضابطة	138		
23	علم النفس	ماجستير	تجريبية 1	11	11.84	2.73
			تجريبية 2	12		
			تجريبية 3	13		
			ضابطة	11		
			تجريبية 1	11	11.41	3.03
			تجريبية 2	12		
			تجريبية 3	13		
			ضابطة	11		

الرقم	القسم	نوع البحث	نوع المجموعة	عدد أفرادها	الخطأ المعياري للفرق بين الدرجات القبلية والبعدية	نقطة القطع
24	المناهج وطرائق التدريس	ماجستير	تجريبية 1	28	5.66	18.78
			تجريبية 2	28	1.73	8.72
25	المناهج وطرائق التدريس	ماجستير	تجريبية 1	31	2.89	12.44
			تجريبية 2	31	3.33	12.98
26	الإرشاد النفسي	ماجستير	تجريبية	18	8.45	122.17
			ضابطة	18		
27	المناهج وطرائق التدريس	ماجستير	تجريبية	33	3.98	18.77
			ضابطة	29		
28	المناهج وطرائق التدريس	ماجستير	تجريبية	68	1.75	8.092
			ضابطة	65		
29	المناهج وطرائق التدريس	ماجستير	تجريبية 1	38	2.49	20.93
			تجريبية 2	37		
			ضابطة	39		
30	المناهج وطرائق التدريس	ماجستير	تجريبية 1	20	2.89	18.96
			تجريبية 2	20	1.81	17.82
31	المناهج وطرائق التدريس	دكتوراه	تجريبية	43	3.21	8.95
			ضابطة	43		
32	المناهج وطرائق التدريس	ماجستير	تجريبية	72	2.60	15.8
			ضابطة	69		
33	التربية الخاصة	ماجستير	تجريبية	12	4.51	36.4
			ضابطة	12		
34	المناهج وطرائق التدريس	ماجستير	تجريبية	32	2.40	12.32
			ضابطة	33		
35	المناهج وطرائق التدريس	ماجستير	تجريبية	61	4.7	23.58
			ضابطة	61		
36	التربة الخاصة	ماجستير	تجريبية	5	3.34	78.12
			ضابطة	5		
37	المناهج وطرائق التدريس	ماجستير	تجريبية	40	2.16	15.96
			ضابطة	39		

الرقم	القسم	نوع البحث	نوع المجموعة	عدد أفرادها	الخطأ المعياري للفرق بين الدرجات القبلية والبعدية	نقطة القطع
38	التربة الخاصة	ماجستير	تجريبية 1	6	6.19	51.29
			ضابطة 1	6		
			تجريبية 2	6	2.65	11.33
			ضابطة 2	6		
39	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	82	6.97	26.63
			تجريبية 2	82		
			ضابطة	49		
			تجريبية 1	82	6.82	41.23
			تجريبية 2	82		
			ضابطة	49		
40	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	42	6.71	42.55
			ضابطة 1	40		
			تجريبية 2	42	3.25	33.97
			ضابطة 2	40		
41	التربية الخاصة	ماجستير	تجريبية	6	2.22	14.04
			ضابطة	6		
42	المناهج وطرائق التدريس	دكتوراه	تجريبية	33	6.82	37.07
43	التربية الخاصة	دكتوراه	تجريبية	5	12.85	33.7
			ضابطة	5		
44	التربية الخاصة	ماجستير	تجريبية	5	7.66	90.58
			ضابطة	5		
45	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	20	7.04	44.52
			تجريبية 2	20	1.43	25.6
46	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	14	5.20	31.51
			تجريبية 2	14	5.22	24.62
			تجريبية 3	14	4.84	14.27
			تجريبية 4	14	2.88	7.68
47	المناهج وطرائق التدريس	دكتوراه	تجريبية	100	4.64	20.92
			ضابطة	100		
48	المناهج وطرائق التدريس	دكتوراه	تجريبية	30	3.53	48.98
			ضابطة	17		
49	المناهج وطرائق التدريس	دكتوراه	تجريبية	30	5.88	20.55

الرقم	القسم	نوع البحث	نوع المجموعة	عدد أفرادها	الخطأ المعياري للفرق بين الدرجات القبلية والبعدية	نقطة القطع
50	المناهج وطرائق التدريس	ماجستير	تجريبية 1	64	1.87	5.53
			تجريبية 2	62		
			ضابطة	65		
51	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	46	4.38	24.1
			تجريبية 2	46		
			ضابطة	46		
52	المناهج وطرائق التدريس	دكتوراه	تجريبية	34	3.24	14
			ضابطة	31		
53	المناهج وطرائق التدريس	دكتوراه	تجريبية	30	2.11	12.19
			ضابطة	30		
54	المناهج وطرائق التدريس	دكتوراه	تجريبية	39	9.09	51.74
			ضابطة	38		
55	المناهج وطرائق التدريس	ماجستير	تجريبية	35	7.42	36.84
			ضابطة	35		
56	المناهج وطرائق التدريس	دكتوراه	تجريبية	40	4.71	15.91
			ضابطة	40		
57	المناهج وطرائق التدريس	ماجستير	تجريبية	32	4.55	24.58
			ضابطة	32		
58	المناهج وطرائق التدريس	دكتوراه	تجريبية 1	34	9.66	34.32
			تجريبية 2	32		
			تجريبية 3	37		
			ضابطة	35		
59	التربية الخاصة	دكتوراه	تجريبية	15	12.95	159.73
			ضابطة	12		
60	المناهج وطرائق التدريس	ماجستير	تجريبية	40	3.80	18.89
			ضابطة	40		
			تجريبية	40	6.96	88.13
			ضابطة	40		
61	المناهج وطرائق التدريس	دكتوراه	تجريبية	73	7.18	26.22
			ضابطة	66		
62	علم النفس	دكتوراه	تجريبية	14	3.87	37.45
			ضابطة	14		
63	المناهج وطرائق التدريس	ماجستير	تجريبية	42	7.75	36.38
			ضابطة	41		
64	المناهج وطرائق التدريس	ماجستير	تجريبية 1	30	7.71	27.84
			تجريبية 2	30		
			ضابطة	30		
65	المناهج وطرائق التدريس	دكتوراه	تجريبية	40	12.7	42.03
			ضابطة	40		

الرقم	القسم	نوع البحث	نوع المجموعة	عدد أفرادها	الخطأ المعياري للفرق بين الدرجات القبلية والبعدية	نقطة القطع
66	المناهج وطرائق التدريس	دكتوراه	تجريبية	43	1.5	15.86
			ضابطة	43		
			تجريبية	43	1.47	16.29
			ضابطة	43		
67	التربية الخاصة	ماجستير	تجريبية	9	7.76	113.6
			ضابطة	9		
68	الإرشاد النفسي	دكتوراه	تجريبية	13	8.47	65.23
			ضابطة	13		
69	المناهج وطرائق التدريس	ماجستير	تجريبية	34	2.36	14.46
			ضابطة	39		
70	المناهج وطرائق التدريس	دكتوراه	تجريبية	30	1.29	15.02
			تجريبية	30		
			ضابطة	30		
71	المناهج وطرائق التدريس	دكتوراه	تجريبية	21	2.135	12.58
			تجريبية	21	8.47	39.04
72	علم النفس	ماجستير	تجريبية	20	7	58.83
			ضابطة	20		
73	المناهج وطرائق التدريس	ماجستير	تجريبية	62	2.11	13.31
			ضابطة	62		
74	المناهج وطرائق التدريس	دكتوراه	تجريبية	38	11.3	40.55
			ضابطة	38		
			تجريبية	38	1.15	3.17
			ضابطة	38		
75	المناهج وطرائق التدريس	ماجستير	تجريبية تحصيل	60	5.58	43.45
			ضابطة تحصيل	59		
			تجريبية	60	3.51	19.57
			ضابطة	59		
76	المناهج وطرائق التدريس	دكتوراه	تجريبية	40	2.11	10.03
77	المناهج وطرائق التدريس	دكتوراه	تجريبية1	33	7.16	43
			تجريبية2	33		
			ضابطة	36		
78	التربية الخاصة	ماجستير	تجريبية	9	1.1	8.72
			ضابطة	9		
79	الإرشاد النفسي	ماجستير	تجريبية	20	4.1	88.48
80	المناهج وطرائق التدريس	دكتوراه	تجريبية	30	6.21	26.73

ملحق (7) قيم الدلالات العلمية للأبحاث المُجمّعة

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
1	تجريبية	66	20.26 0.00	0.86	مرتفع	40	60.6 %	17	25.8 %	9	13.6 %	-	-
	ضابطة	66	5.5 0.00	0.31	مرتفع	2	%3	2	%3	62	%94	-	-
2	تجريبية1	33	14.75 0.00	0.72	مرتفع	31	93.9 %	-	-	2	%6.1	-	-
	تجريبية2	30	4.78- 0.00	0.87	مرتفع	30	%100	-	-	-	-	-	-
	تجريبية3	51	25 0.00	0.93	مرتفع	50	%98	-	-	1	%2	-	-
	ضابطة	33	14.07 0.00	0.86	مرتفع	29	%88	-	-	4	%12	-	-
3	تجريبية	15	3.41 0.00	0.88	مرتفع	15	%100	-	-	-	-	-	-
	ضابطة	15	3.41 0.00	0.88	مرتفع	15	%100	-	-	-	-	-	-
4	تجريبية	50	22.75 0.00	0.91	مرتفع	47	%94	-	-	3	%6		
	ضابطة	50	10.99 0.00	0.71	مرتفع	25	%50	2	%4	22	%44	1	%2
5	تجريبية معرفي	41	38.67 0.00	0.97	مرتفع	41	%100	-	-	-	-	-	-
	تجريبية مهاري	41	24.6 0.00	0.94	مرتفع	41	%100						
	تجريبية وجداني	41	19.1 0.00	0.90	مرتفع	33	80.5 %	7	17.1 %	1	%2.4	-	-
6	تجريبية	60	4.31 0.00	0.24	مرتفع	7	11.67 %	6	%10	46	76.67 %	1	1.66 %
	ضابطة	60	19.31 0.00	0.86	مرتفع	-	-	-	-	60	%100	-	-
7	تجريبية	30	4.79 0.00	0.87	مرتفع	30	%100	-	-	-	-	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
8	تجريبية	35	31.32 0.00	0.97	مرتفع	35	%100	-	-	-	-	-	-
	ضابطة	39	12.27 0.00	0.80	مرتفع	20	51.29 %	1	2.56 %	18	46.15 %	-	-
9	تجريبية 1	65	50.99 0.00	0.98	مرتفع	65	%100	-	-	-	-	-	-
	ضابطة1	37	44.46 0.00	0.98	مرتفع	37	%100	-	-	-	-	-	-
	تجريبية2	60	38.57 0.00	0.96	مرتفع	60	%100	-	-	-	-	-	-
	ضابطة2	27	4.55 0.00	0.88	مرتفع	27	%100	-	-	-	-	-	-
10	تجريبية فقط	40	10.17 0.00	0.73	مرتفع	4	%10	11	27.5 %	24	%60	1	2.5 %
11	تجريبية	70	54.28 0.00	0.98	مرتفع	70	%100	-	-	-	-	-	-
	ضابطة	70	26.34 0.00	0.91	مرتفع	69	98.6 %	-	-	1	%1.4	-	-
12	تجريبية	64	20.28 0.00	0.88	مرتفع	19	29.69 %	33	51.56 %	12	18.75 %	-	-
	ضابطة	61	11.36 0.00	0.68	مرتفع	7	11.4 %	27	44.3 %	27	44.3 %	-	-
13	تجريبية	10	2.81 0.005	0.89	مرتفع	8	%80	2	%20	-	-	-	-
	ضابطة	10	2.67 0.008	0.84	مرتفع	3	%30	-	-	7	%70	-	-
14	تجريبية	66	71.79 0.00	0.99	مرتفع	66	%100	-	-	-	-	-	-
	ضابطة	66	36.40 0.00	0.95	مرتفع	66	%100	-	-	-	-	-	-
15	تجريبية	123	124.15 0.00	0.99	مرتفع	123	%100	-	-	-	-	-	-
	ضابطة	124	65.07 0.00	0.97	مرتفع	124	%100	-	-	-	-	-	-
16	تجريبية1	81	14.53 0.00	0.73	مرتفع	17	%21	12	14.8 %	52	64.2 %	-	-
	تجريبية2	160	23.44 0.00	0.78	مرتفع	46	28.8 %	37	23.1 %	77	48.1 %	-	-
	ضابطة	106	13.96 0.00	0.65	مرتفع	8	%7.5	25	23.6 %	73	68.9 %	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
17	تجريبية	50	27.61 0.00	0.94	مرتفع	48	%96	1	%2	1	%2	-	-
	ضابطة	50	13.74 0.00	0.79	مرتفع	43	%86	6	%12	1	%2	-	-
18	تجريبية	68	60.50 0.00	0.98	مرتفع	68	%100	-	-	-	-	-	-
	ضابطة	68	44.38 0.00	0.97	مرتفع	68	%100	-	-	-	-	-	-
19	تجريبية طلاب	122	29.51 0.00	0.87	مرتفع	85	69.7 %	7	%5.7	30	24.6 %	-	-
	ضابطة طلاب	100	13.05 0.00	0.63	مرتفع	-	-	7	%7	93	%93	-	-
	تجريبية دبلوم	18	3.73 0.00	0.88	مرتفع	16	88.88 %	1	5.56 %	1	5.56 %		
20	تجريبية	185	94.66 0.00	0.98	مرتفع	185	%100	-	-	-	-	-	-
	ضابطة	171	47.32 0.00	0.93	مرتفع	158	92.4 %	1	%0.6	12	%7	-	-
21	تجريبية	10	2.81 0.005	0.89	مرتفع	10	%100	-	-	-	-	-	-
	ضابطة	10	2.7 0.007	0.85	مرتفع	-	-	-	-	10	%100	-	-
22	تجريبية	158	81.31 0.00	0.98	مرتفع	158	%100	-	-	-	-	-	-
	ضابطة	138	26.47 0.00	0.84	مرتفع	60	43.5 %	26	18.8 %	52	37.7 %	-	-
23	تجريبية1	11	2.94 0.003	0.89	مرتفع	11	%100	-	-	-	-	-	-
	تجريبية2	12	3.06 0.002	0.88	مرتفع	12	%100	-	-	-	-	-	-
	تجريبية3	13	3.18 0.001	0.88	مرتفع	13	%100	-	-	-	-	-	-
	ضابطة	11	2.93 0.003	0.88	مرتفع	11	%100	-	-	-	-	-	-
	تجريبية1	11	2.94 0.003	0.89	مرتفع	11	%100	-	-	-	-	-	-
	تجريبية2	12	3.07 0.002	0.89	مرتفع	12	%100	-	-	-	-	-	-
	تجريبية3	13	3.18 0.001	0.88	مرتفع	13	%100	-	-	-	-	-	-
	ضابطة	11	2.94 0.003	0.89	مرتفع	11	%100	-	-	-	-	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
24	تجريبية 1	28	4.63 0.00	0.88	مرتفع	24	85.7 %	-	-	4	14.3 %	-	-
	تجريبية 2	28	4.64 0.00	0.88	مرتفع	28	%100	-	-	-	-	-	-
25	تجريبية 1	31	59.03 0.00	0.99	مرتفع	31	%100	-	-	-	-	-	-
	تجريبية 2	31	23.30 0.00	0.69	مرتفع	29	93.5 %	2	%6.5	-	-	-	-
26	تجريبية	18	3.07 0.002	0.72	مرتفع	5	27.8 %	-	-	13	72.2 %	-	-
	ضابطة	18	1.06 0.289	0.25	ضعيف	-	-	-	-	18	%100	-	-
27	تجريبية	33	24.76 0.00	0.95	مرتفع	32	%97	-	-	1	%3	-	-
	ضابطة	29	0.96 0.336	0.18	منخفض	1	%3.4	1	%3.4	26	89.8 %	1	3.4 %
28	تجريبية	68	55.39 0.00	0.98	مرتفع	68	%100	-	-	-	-	-	-
	ضابطة	65	17.12 0.00	0.82	مرتفع	53	81.5 %	12	18.5 %	-	-	-	-
29	تجريبية1	38	35.22 0.00	0.97	مرتفع	33	86.8 %	5	13.2 -%	-	-	-	-
	تجريبية2	37	22.46 0.00	0.93	مرتفع	35	94.6 %	-	-	2	%5.4	-	-
	ضابطة	39	10.14 0.00	0.73	مرتفع	1	2.56 %	9	23.08 %	29	74.36 %	-	-
30	تجريبية 1	20	3.93 0.00	0.88	مرتفع	7	%35	1	%5	12	%60	-	-
	تجريبية 2	20	3.84 0.00	0.86	مرتفع	2	%10	6	%30	12	%60	-	-
31	تجريبية	43	22.63 0.00	0.92	مرتفع	43	%100	-	-	-	-	-	-
	ضابطة	43	12.61 0.00	0.79	مرتفع	39	90.7 %	-	-	4	%9.3	-	-
32	تجريبية	72	43.74 0.00	0.96	مرتفع	72	%100	-	-	-	-	-	-
	ضابطة	69	30.30 0.00	0.93	مرتفع	69	%100	-	-	-	-	-	-
33	تجريبية	12	3.065 0.002	0.89	مرتفع	12	%100	-	-	-	-	-	-
	ضابطة	12	3.07 0.002	0.89	مرتفع	12	%100	-	-	-	-	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
34	تجريبية	32	46.05 0.00	0.98	مرتفع	32	%100	-	-	-	-	-	-
	ضابطة	33	19.83 0.00	0.92	مرتفع	31	93.9 %	-	-	2	%6.1	-	-
35	تجريبية	61	19.92 0.00	0.87	مرتفع	40	65.6 %	5	%8.2	16	26.2 %	-	-
	ضابطة	61	8.59 0.00	0.55	مرتفع	-	-	-	-	61	%100	-	-
36	تجريبية	5	2.03 0.04	0.91	مرتفع	5	%100	-	-	-	-	-	-
	ضابطة	5	2.06 0.03	0.92	مرتفع	5	%100	-	-	-	-	-	-
37	تجريبية	40	27.26 0.00	0.95	مرتفع	33	82.5 %	7	17.5 %	-	-	-	-
	ضابطة	39	19.35 0.00	0.91	مرتفع	19	48.7 %	17	43.6 %	3	%7.7	-	-
38	تجريبية 1	6	2.23 0.026	0.91	مرتفع	6	%100	-	-	-	-	-	-
	ضابطة 1	6	1.63 0.102	0.67	مرتفع	-	-	-	-	6	%100	-	-
	تجريبية2	6	2.21 0.027	0.90	مرتفع	6	%100	-	-	-	-	-	-
	ضابطة2	6	0	0	-	-	-	-	-	6	%100	-	-
39	تجريبية1	82	24.80 0.00	0.79	مرتفع	59	%72	-	-	23	%28	-	-
	تجريبية2	82	25.52 0.00	0.89	مرتفع	61	74.4 %	-	-	21	25.6 %	-	-
	ضابطة	49	13.21 0.00	0.68	مرتفع	22	44.9 %	-	-	27	55.1 %	-	-
	تجريبية1	82	18.61 0.00	0.81	مرتفع	53	64.6 %	3	%3.7	26	31.7 %	-	-
	تجريبية2	82	25.65 0.00	0.89	مرتفع	64	%78	-	-	18	%22	-	-
40	تجريبية1	42	18.15 0.00	0.89	مرتفع	29	%69	6	14.3 %	7	16.7 %	-	-
	ضابطة1	40	10.43 0.00	0.73	مرتفع	3	%7.5	8	%20	29	72.5 %	-	-
	تجريبية2	42	17.05 0.00	0.88	مرتفع	29	%69	7	16.7 %	6	14.3 %	-	-
	ضابطة2	40	7.78 0.00	0.61	مرتفع	11	27.5 %	5	12.5 %	24	%60	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
41	تجريبية	6	2.21 0.027	0.90	مرتفع	6	%100	-	-	-	-	-	-
	ضابطة	6	0.337 0.75	0.022	ضعيف	1	16.67 %	-	-	4	66.66 %	1	16.67%
42	مجموعة تجريبية واحدة	33	25.94 0.00	0.97	مرتفع	33	%100	-	-	-	-	-	-
43	تجريبية	5	2.023 0.043	0.90	مرتفع	4	%80	1	%20	-	-	-	-
	ضابطة	5	0.73 0.465	0.30	متوسط	-	-	-	-	5	%100	-	-
44	تجريبية	5	2.023 0.043	0.90	مرتفع	5	%100	-	-	-	-	-	-
	ضابطة	5	0.95 0.343	0.42	متوسط	-	-	-	-	5	%100	-	-
45	تجريبية1	20	3.92 0.00	0.88	مرتفع	20	%100	-	-	-	-	-	-
	تجريبية 2	20	3.93 0.00	0.88	مرتفع	20	%100	-	-	-	-	-	-
46	تجريبية1	14	3.3 0.001	0.88	مرتفع	14	%100	-	-	-	-	-	-
	تجريبية2	14	3.3 0.001	0.88	مرتفع	14	%100	-	-	-	-	-	-
	تجريبية 3	14	3.3 0.001	0.88	مرتفع	10	71.4 %	-	-	4	28.6 %	-	-
	تجريبية 4	14	3.31 0.001	0.89	مرتفع	13	92.9 %	-	-	1	7.1	-	-
47	تجريبية	100	146.57 0.00	0.99	مرتفع	100	%100	-	-	-	-	-	-
	ضابطة	100	111.74 0.00	0.99	مرتفع	100	%100	-	-	-	-	-	-
48	تجريبية	30	4.813 0.00	0.88	مرتفع	30	%100	-	-	-	-	-	-
	ضابطة	17	1.28 0.199	0.23	متوسط	-	-	-	-	17	%100	-	-
49	تجريبية مدرسين	30	4.79 0.00	0.87	مرتفع	27	%90	-	-	3	%10	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
50	تجريبية 1	64	57.44 0.00	0.98	مرتفع	64	%100	-	-	-	-	-	-
	تجريبية 2	62	68.82 0.00	0.99	مرتفع	62	%100	-	-	-	-	-	-
	ضابطة	65	27.66 0.00	0.92	مرتفع	65	%100	-	-	-	-	-	-
51	تجريبية 1	46	47.60 0.00	0.98	مرتفع	46	%100	-	-	-	-	-	-
	تجريبية 2	46	44.22 0.00	0.98	مرتفع	46	%100	-	-	-	-	-	-
	ضابطة	46	8.79 0.00	0.63	مرتفع	35	76.1 %	-	-	11	23.9 %	-	-
52	تجريبية	34	80.40 0.00	0.99	مرتفع	34	%100	-	-	-	-	-	-
	ضابطة	31	19.36 0.00	0.93	مرتفع	31	%100	-	-	-	-	-	-
53	تجريبية	30	4.6 0.00	0.84	مرتفع	30	%100	-	-	-	-	-	-
	ضابطة	30	4.8 0.00	0.88	مرتفع	4	13.3 %	1	%3.3	25	83.4 %	-	-
54	تجريبية	39	12.14 0.00	0.79	مرتفع	26	66.7 %	2	%5.1	11	28.2 %	-	-
	ضابطة	38	2.92 0.006	0.19	مرتفع	4	10.53 %	1	2.63 %	33	86.84 %	-	-
55	تجريبية	35	50.71 0.00	0.99	مرتفع	35	%100	-	-	-	-	-	-
	ضابطة	35	26 0.00	0.95	مرتفع	-	-	-	-	35	%100	-	-
56	تجريبية	40	32.62 0.00	0.96	مرتفع	40	%100	-	-	-	-	-	-
	ضابطة	40	1.25 0.217	0.04	منخفض	-	-	-	-	40	%100	-	-
57	تجريبية	32	32.77 0.00	0.97	مرتفع	32	%100	-	-	-	-	-	-
	ضابطة	32	24.48 0.00	0.95	مرتفع	3	%9.4	9	28.1 %	20	62.5 %	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
58	تجريبية1	34	11.82 0.00	0.81	مرتفع	7	20.6 %	4	11.8 %	23	67.6 %	-	-
	تجريبية2	32	7.44 0.00	0.64	مرتفع	2	%6.3	4	12.5 %	26	81.2 %	-	-
	تجريبية3	37	11.18 0.00	0.78	مرتفع	10	%27	1	%2.7	26	70.3 %	-	-
	ضابطة	35	5.08 0.00	0.43	مرتفع	4	11.4 %	1	%2.9	30	85.7 %	-	-
59	تجريبية	15	3.41 0.001	0.88	مرتفع	15	%100	-	-	-	-	-	-
	ضابطة	12	0.582 0.56	0.17	متوسط	-	-	-	-	12	%100	-	-
60	تجريبية تحصيل	40	38.11 0.00	0.97	مرتفع	40	%100	-	-	-	-	-	-
	ضابطة تحصيل	40	17.92 0.00	0.89	مرتفع	4	%10	4	%10	32	%80	-	-
	تجريبية العادات العقلية	40	3.25 0.00	0.21	مرتفع	40	%100	-	-	-	-	-	-
	ضابطة العادات العقلية	40	7.95 0.00	0.62	مرتفع	-	-	1	%2.5	39	97.5 %	-	-
61	تجريبية	73	27.57 0.00	0.91	مرتفع	68	93.2 %	-	-	5	%6.8	-	-
	ضابطة	66	13.08 0.00	0.72	مرتفع	23	34.9 %	2	%3	41	62.1 %	-	-
62	تجريبية	14	3.3 0.001	0.88	مرتفع	14	%100	-	-	-	-	-	-
	ضابطة	14	2.87 0.002	0.77	مرتفع	-	-	-	-	14	%100	-	-
63	تجريبية	42	36.74 0.00	0.97	مرتفع	40	%100	-	-	-	-	-	-
	ضابطة	41	19.38 0.00	0.90	مرتفع	35	85.4 %	-	-	6	14.6 %	-	-
64	تجريبية 1	30	4.71 0.00	0.86	مرتفع	18	%60	-	-	12	%40		
	تجريبية 2	30	4.785 0.00	0.87	مرتفع	19	63.3 %	-	-	11	36.7 %	-	-
	ضابطة	30	4.6 0.00	0.84	مرتفع	9	%30	-	-	21	%70	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
65-	تجريبية	40	28.78 0.00	0.95	مرتفع	40	%100	-	-	-	-	-	-
	ضابطة	40	17.12 0.00	0.88	مرتفع	40	%100	-	-	--	-	-	-
66	تجريبية	43	23.76 0.00	0.93	مرتفع	26	60.5 %	17	39.5 %	-	-	-	-
	ضابطة	43	7.02 0.00	0.54	مرتفع	6	%14	2	%4.7	35	81.3 %	-	-
	تجريبية	43	40.33 0.00	0.97	مرتفع	35	81.4 %	8	18.6 %	-	-	-	-
	ضابطة	43	16.79 0.00	0.87	مرتفع	24	55.8 %	16	37.2 %	3	%7	-	-
67	تجريبية	9	2.67 0.008	0.89	مرتفع	9	%100	-	-	-	-	-	-
	ضابطة	9	2.45 0.014	0.82	مرتفع	-	-	-	-	9	%100	-	-
68	تجريبية	13	3.185 0.00	0.88	مرتفع	4	30.8 %	9	69.2 %	-	-	-	-
	ضابطة	13	المتوسطان متساويان	-	-	-	-	-	-	13	%100	-	-
69	تجريبية	39	56.64 0.00	0.99	مرتفع	39	%100	-	-	-	-	-	-
	ضابطة	39	32.86 0.00	0.97	مرتفع	39	%100	-	-	-	-	-	-
70	تجريبية1	30	4.81 0.00	0.88	مرتفع	28	93.3 %	2	%6.7	-	-	-	-
	تجريبية2	30	4.79 0.00	0.87	مرتفع	30	%100	-	-	-	-	-	-
	ضابطة	30	4.79 0.00	0.87	مرتفع	28	93.3 %	2	%6.7	-	-	-	-
71	تجريبية أدائي	21	4.023 0.00	0.88	مرتفع	21	%100	-	-	-	-	-	-
	تجريبية معرفي	21	4.017 0.00	0.88	مرتفع		%100	-	-	-	-	-	-
72	تجريبية	20	3.93 0.00	0.88	مرتفع	10	%50	3	%15	7	%35	-	-
	ضابطة	20	0.785 0.433	0.18	متوسط	-	-	-	-	20	%100	-	-

الرقم	نوع المجموعة	عدد أفرادها	الدلالة الإحصائية وقيمة t	قيمة حجم التأثير	تفسير حجم التأثير	النسبة المئوية لفئة الأفراد المُعافين		النسبة المئوية لفئة الأفراد الذين تحسنوا		النسبة المئوية لفئة الأفراد الذين لم يتغيروا		النسبة المئوية لفئة الأفراد الذين تراجعوا	
						العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية	العدد	النسبة المئوية
73	تجريبية	62	57.20 0.00	0.98	مرتفع	62	%100	-	-	-	-	-	-
	ضابطة	62	28.29 0.00	0.93	مرتفع	51	82.3 %	-	-	11	17.7 %	-	-
74	تجريبية إبداعي	38	6.53 0.00	0.54	مرتفع	15	39.5 %	-	-	23	60.5 %	-	-
	ضابطة إبداعي	38	3.48 0.001	0.25	مرتفع	13	34.2 %	-	-	25	65.8 %	-	-
	تجريبية مفاهيم	38	45.85 0.00	0.98	مرتفع	38	%100	-	-	-	-	-	-
	ضابطة مفاهيم	38	16.20 0.00	0.88	مرتفع	37	97.4 %	-	-	1	%2.6	-	-
75	تجريبية تحصيل	60	12.70 0.00	0.73	مرتفع	17	28.3 %	10	16.7 %	33	%55	-	-
	ضابطة تحصيل	59	9.93 0.00	0.63	مرتفع	1	%1.7	8	13.6 %	50	84.7 %	-	-
	تجريبية حل مشكلات	60	15.49 0.00	0.80	مرتفع	26	43.33 %	8	13.33 %	26	43.34 %	-	-
	ضابطة حل مشكلات	59	4.72 0.00	0.28	مرتفع	4	6.79 %	5	8.47 %	49	83.05 %	1	1.69 %
76	تجريبية أدائي	40	37.32 0.00	0.97	مرتفع	40	%100	-	-	-	-	-	-
77	تجريبية1	33	16.47 0.00	0.89	مرتفع	27	81.8 %	-	-	6	18.2 %	-	-
	تجريبية2	33	14.84 0.00	0.87	مرتفع	25	75.8 %	-	-	8	24.2 %	-	-
	ضابطة	36	7.88 0.00	0.64	مرتفع	7	19.45 %	3	8.33 %	26	72.22 -%	-	-
78	تجريبية	9	2.699 0.007	0.90	مرتفع	6	66.7 %	-	-	3	33.3 %	-	-
	ضابطة	9	2.236 0.025	0.73	مرتفع	-	-	-	-	9	%100	-	-
79	تجريبية	20	3.93 0.00	0.88	مرتفع	-	-	16	%80	4	%20	-	-
80	تجريبية	30	42.3 0.00 0.12	0.98	مرتفع	30	%100	-	-	-	-	-	-

ملحق (8)

معاملات سهولة مصفوفات اختبار الذكاء

المجموعة (A)				المجموعة (B)			
رقم المصفوفة	عدد الإجابات الصحيحة	العدد الكلي	قيمة معامل السهولة	رقم المصفوفة	عدد الإجابات الصحيحة	العدد الكلي	قيمة معامل السهولة
1	125	125	**1**	**1**	124	125	**0.992**
2	125	125	**1**	**2**	121	125	**0.968**
3	125	125	**1**	**3**	119	125	**0.952**
4	124	125	**0.992**	**4**	112	125	**0.896**
5	124	125	**0.992**	**5**	103	125	**0.824**
6	123	125	**0.984**	**6**	103	125	**0.824**
7	120	125	**0.96**	**7**	102	125	**0.816**
8	112	125	**0.896**	**8**	106	125	**0.848**
9	112	125	**0.896**	**9**	100	125	**0.80**
10	110	125	**0.88**	**10**	93	125	**0.744**
11	99	125	**0.792**	**11**	87	125	**0.696**
12	86	125	**0.688**	**12**	75	125	**0.60**

المجموعة (C)				المجموعة (D)				المجموعة (E)			
رقم المصفوفة	عدد الإجابات الصحيحة	العدد الكلي	قيمة معامل السهولة	رقم المصفوفة	عدد الإجابات الصحيحة	العدد الكلي	قيمة معامل السهولة	رقم المصفوفة	عدد الإجابات الصحيحة	العدد الكلي	قيمة معامل السهولة
1	122	125	**0.976**	**1**	118	125	**0.944**	**1**	103	125	**0.824**
2	118	125	**0.944**	**2**	113	125	**0.904**	**2**	89	125	**0.712**
3	116	125	**0.928**	**3**	104	125	**0.832**	**3**	77	125	**0.616**
4	109	125	**0.872**	**4**	103	125	**0.824**	**4**	66	125	**0.528**
5	100	125	**0.80**	**5**	94	125	**0.752**	**5**	64	125	**0.512**
6	100	125	**0.80**	**6**	90	125	**0.72**	**6**	54	125	**0.432**
7	98	125	**0.784**	**7**	89	125	**0.712**	**7**	28	125	**0.224**
8	84	125	**0.672**	**8**	80	125	**0.64**	**8**	25	125	**0.20**
9	86	125	**0.688**	**9**	72	125	**0.576**	**9**	25	125	**0.20**
10	71	125	**0.568**	**10**	42	125	**0.336**	**10**	12	125	**0.096**
11	62	125	**0.496**	**11**	39	125	**0.312**	**11**	11	125	**0.088**
12	27	125	**0.216**	**12**	11	125	**0.088**	**12**	6	125	**0.048**

ملحق (9)

الدرجات على اختباري SPM-PV و Toni -4

ملحق (9) درجات أفراد عينة البحث التجريبية على اختباري الذكاء SPM-PV و Toni -4

رقم الفرد	SPM-PV القبلي	SPM-PV البعدي المباشر	SPM-PV البعدي المؤجل	Toni -4 قبلي	Toni -4 بعدي مباشر
1	49	54	50	38	50
2	39	49	48	37	45
3	36	51	43	40	48
4	39	48	45	35	47
5	42	53	54	39	50
6	41	51	50	32	39
7	36	47	46	37	47
8	47	57	50	42	48
9	37	51	50	32	45
10	44	55	51	36	48
11	34	51	49	31	42
12	35	51	50	30	41
13	45	54	51	30	37
14	47	56	52	42	51
15	48	57	56	46	51
16	39	52	51	44	43
17	54	58	57	41	52
18	39	52	48	32	41
19	28	45	40	30	38
20	51	54	55	46	50
21	41	53	52	43	49
22	38	49	42	23	35
23	41	53	50	34	45
24	42	55	49	30	43
25	39	52	50	21	39
26	38	51	44	40	43
27	44	54	51	23	38
28	37	48	50	28	43
29	44	56	51	41	47
30	38	51	41	41	47

43	36	46	52	42	31
48	45	55	56	51	32
49	43	50	53	44	33
42	35	51	53	40	34
38	31	44	47	46	35
37	30	38	48	31	36
40	26	50	52	41	37
52	37	53	60	48	38
44	38	45	55	49	39
50	46	49	57	45	40
44	37	44	52	39	41
41	34	40	51	33	42
49	43	55	56	44	43
42	35	45	53	39	44
50	48	50	52	43	45
40	36	46	50	38	46
41	38	49	51	39	47
38	38	48	46	42	48
35	36	44	47	38	49
40	32	53	50	41	50
50	44	49	57	47	51
47	41	52	55	44	52
46	39	52	50	37	53
40	35	50	53	44	54
41	37	49	50	32	55
47	38	48	57	45	56
43	35	50	51	39	57
47	41	51	54	47	58
52	47	46	53	41	59
38	37	50	49	40	60
47	44	56	57	50	61
46	40	54	53	42	62
45	38	51	50	43	63

43	40	52	54	45	64
50	43	54	54	48	65
40	41	46	48	37	66
41	40	51	50	35	67
36	36	49	46	34	68
49	40	52	54	41	69
47	38	55	56	50	70
43	40	48	52	40	71
47	39	51	56	45	72
45	37	50	51	46	73
38	36	44	52	39	74
48	41	51	54	47	75
52	42	50	58	47	76
37	32	49	53	41	77
50	40	49	54	42	78
50	46	48	56	52	79
42	37	51	53	40	80
49	41	48	55	44	81
49	40	52	54	46	82
54	45	56	58	53	83
45	38	53	50	40	84
48	39	46	53	40	85
39	31	47	48	33	86
51	43	55	57	50	87
43	44	51	49	37	88
42	31	53	50	39	89
48	41	50	53	42	90
41	32	52	55	46	91
31	29	49	49	40	92
37	33	41	46	32	93
39	35	52	53	46	94
50	40	51	55	42	95
43	36	48	52	40	96

41	32	41	50	35	97
40	25	44	50	37	98
48	41	51	53	44	99
47	37	52	54	46	100
53	39	58	59	55	101
47	40	49	52	42	102
47	39	54	55	45	103
46	40	56	57	52	104
41	33	44	50	31	105
43	33	48	55	44	106
42	35	49	48	41	107
50	39	52	58	47	108
42	34	52	49	46	109
41	35	47	49	36	110
50	48	48	51	43	111
52	43	50	52	43	112
46	38	54	55	53	113
44	33	52	52	40	114
47	40	51	56	47	115
42	35	50	51	39	116
42	30	47	52	38	117
47	40	49	53	43	118
39	28	40	49	35	119
44	44	51	49	45	120
42	31	43	52	35	121
39	32	48	49	33	122
50	39	51	54	43	123
45	36	47	50	39	124
40	34	50	52	39	125

Summary

The Reality of Scientific Significances in Experimental Research and the Effect of Training in them and in the Psychometric Properties of Tests

The importance of scientific significances, which has been increasingly emphasized on the importance of its account in the research literature, publication evidence, and scientific associations, prompted a large number of researchers to subject it to study and research, whether by distinguishing between its different types or in terms of monitoring the reality of the use of some of them, but none of them sought to analyze the reality of their use, or study the relationships between their indicators together, and the adequacy of each of them in reporting the results of experimental research, and none of them aimed to study the extent to which their results are affected by factors related to the test or training situation to which the examinees are subjected. Hence, the current research aims to analyze the reality of the use of the three significances in experimental research, and to study the relationships between their indicators, and the adequacy of each of them in the accuracy of the statistical decision. It also seeks to study how the training variable can affect its results, and raise or lower the indicators of the Psychometric Properties of the test.

The research questions centered on analyzing the reality of using these significances in experimental research, and studying their adequacy in reporting their results, and the predictive ability between their coefficients on the one hand, and studying the impact of the training variable on their results and the Psychometric Properties of the test on the other hand. In addition to the comparison between the methods used to calculate the clinical significance to determine the most appropriate methods for the experimental programs prepared by the researchers.

Three basic approaches were adopted in answering the research questions: descriptive analytical, experimental, and meta – analysis approach. The research community was limited to one statistical community associated with the descriptive analytical, and meta- analysis method, and represented all experimental researches in the following sections: Curricula and Teaching Methods, Psychological Counseling, Psychology and Special Education, between 2000-2020, and the first semester of the academic year 2020-2021.

As for the research sample, it consisted of two basic samples. The first is the academic research sample: It consisted of all experimental master's Theses and doctoral Dissertations in the same departments of the original community, that is, the sample size was equal to the size of the original community for the target .time period and for the four departments, with a representation rate of 100%.

The second is the experimental research sample: it is an occasional available sample consisted of students in the second year of Psychology, and the number of its members reached (125) male and female students.

The research reached the following results:

- The percentage of statistical tests whose practical significance was verified in the results of experimental research was (29.47%) only, and this is not proportional to the number of total statistical significance tests calculated in them based on the results of the chi-square test, and none of the clinical significance tests were calculated in sample research analysis at all.
- Both types of decision based on statistical significance have been related to the classifications of practical and clinical significance, all to varying degrees, which indicates that the statistical decision alone is unable to accurately report on the actual and real impact of the experimental programs on the one hand, and on the actual changes that occur at the level and direction of individuals on the other hand.
- The different classifications of effect sizes, according to the results of the meta-analysis, coincided with all classifications of clinical significance, except for the low effect size that was not associated with the improved classification, which indicates the inadequacy of the different classifications of effect sizes in reporting the results of experimental research.
- The presence of a predictive ability for the statistical significance coefficients with the classifications of clinically significant Recovered and Unchanged, and the presence of a predictive ability for the practical significance coefficients by classifications: Recovered, Unchanged and Deteriorated clinically significant, but both of these types of significance could not predict the values and classifications of all clinical significance. Thus, calculating either of the two types of statistical or practical

significance, or even calculating both, is not sufficient to extract the results of experimental research.

- The training variable effected the scientific significance values of the experimental group's scores of the in the direct post- test with a statistically, practical, and clinical significant, without in fact affecting the mental abilities of the research sample individuals and their basic levels of intelligence, as the scores of a good proportion of the sample members in the post-delayed test declined and regressed to return to their basic rates as in the pre-test.
- The presence of very high statistically significant agreement rates between the values and classifications extracted from the three methods (JT), (GLN) and (EN). As for the (HA) method, its results showed very high rates of clinically significant improvement ratings. Perhaps the reason for this is due to the statistical complexities in the numerator of the mathematical formula of this method, which made it more suited to the nature of clinical therapeutic intervention programs.
- The training variable affected the increase of the criterion validity coefficient of the intelligence test with a statistically significant increase, while it did not affect the discriminatory validity coefficients of the test with a statistical and practical significant, but it affected some values of clinical significance for the individuals of the upper and lower categories.
- The training variable affected the increase of Cronbach's alpha and split – half reliability coefficients a statistically significant increase in the direct post-test, and perhaps this indicates its importance and its role in improving the Psychometric Properties of the test.

Based on the previous results, a number of proposals were presented, the most important of which are:

- Working on the adoption of the three scientific significances in experimental research of its various types, and the adoption of the statistical and practical significances in descriptive analytical research.
- Conducting studies and research aimed at analyzing the reality of the use of practical significance as well as statistical in descriptive analytical research.
- Empowering postgraduate students in various departments with the necessary knowledge and skills to calculate the three types of scientific significances, and to interpret their results.

- Adding Test Taking Strategies as an integral part of the curriculum in the different educational stages, because of its importance in raising the learners' abilities to deal with tests in light of the lack of interest in this variable locally, and increasing interest in it and focusing on its importance globally.
- Conducting training programs on Test Taking Strategies in psychological and other educational tests, such as achievement tests and personality tests.

شكر وتقدير

بعد رحلة طويلةٍ من الجد والعمل تكللت بخروج هذا البحث الذي أتمنى أن يكون على القدر المرجو من المنفعة العلمية إلى حيز الوجود، وبفضلٍ وتوفيقٍ من الله تعالى، يطيب لي أن أتقدم بجزيل الشكر والامتنان والعرفان بالجميل إلى مشرفتي الموقرة، التي سيبقى اسمها وساماً يتوج دربي ما حييت

الأستاذة الدكتورة إيمان عز

أَخَذَت بيدي لإنجاز هذا البحث، ومنحتني الاهتمام والتوجيه والمتابعة الدائمة حتى في أصعب الظروف

جزاها الله عني كل خير . . وأدام عليها نعمة الفكر المتوقد والروح المعطاءة

كما أتقدم بخالص شكري وامتناني إلى السادة الأفاضل أعضاء لجنة الحكم لتفضلهم بقبول تحكيم ومناقشة ما قدمتُه من عمل، وعلى الوقت والجهد الذي بذلوه في تصويبه وتقويمه

لهم مني فائق الاحترام والتقدير

ويطيب لي أن أتقدم بأبلغ عبارات التقدير إلى عمادة كلية التربية الفاضلة متمثلةً بعميدتها، وأعضاء الهيئة التدريسية الأفاضل، وإلى جميع أساتذتي الكرام الذين لم يبخلوا يوماً بعلمٍ أو جهدٍ أو معرفة.

ولا يفوتني أن أشكر الدكتورة رندة العمري التي تكرمت بتدقيق البحث لغوياً، فلها مني خالص الاحترام، والشكر موصولٌ لكل من ساعدني في إتمام هذا البحث.

أهلي . . أصدقائي . . أحبتي

عائلتي الثانية إخوتي وزملائي الأوفياء في وزارة التربية، ومركز القياس والتقويم التربوي

جزاكم الله تعالى عني كل خير، ولكم مني كل المحبة والامتنان.

فهرس المحتويات

الفصل الأول: الإطار المنهجي للبحث	رقم الصفحة 1- 12
مقدمة	2
1. مشكلة البحث	3- 7
2. أهمية البحث	8
3. أهداف البحث	8
4. أسئلة البحث	8- 9
5. منهج البحث	9
6. مجتمع البحث وعينته	9
7. متغيرات البحث	9
8. أدوات البحث	10
9. مصطلحات البحث العلمية	10- 12
خلاصة	12
الفصل الثاني: أبحاثٌ ودراساتٌ سابقة	**رقم الصفحة 13- 28**
تمهيد	14
1. أبحاث الدلالات العلمية	14- 18
2. أبحاث الدلالة العيادية	18- 20
3. أبحاث التدريب والخصائص القياسية	21- 22
4. تعقيبٌ ومناقشةٌ	22- 28
5. موقع البحث الحالي من الأبحاث السابقة	28
خلاصة	28
الفصل الثالث: الإطار النظري للبحث	**رقم الصفحة 29- 84**
تمهيد	30
أولاً: أنواع الدلالات العلمية	30- 67
1. الدلالة الإحصائية	30- 36

– مستوى الدلالة والقيمة الاحتمالية	31– 32
– أنواع الفروض	32–33
– الأخطاء المرتبطة باختبارات الفروض	33 – 34
– العوامل المؤثرة في اختبارات الدلالة الإحصائية	34 – 36
2. الدلالة العملية وحجم الأثر	37– 52
– مفهوم الدلالة العملية وحجم الأثر	37– 38
– فوائد حساب حجم الأثر	38
– أنواع مقاييس حجم الأثر	39– 49
– مفهوم قوة الاختبار الإحصائي	50
– استخدامات حجم الأثر	50
– الانتقادات الموجهة إلى الدلالة العملية	50– 52
3. الدلالة العيادية ومؤشر التغيُّر الموثوق	52– 67
– مفهوم الدلالة العيادية	52– 54
– مفهوم مؤشر التغيُّر الموثوق	54– 56
– طرائق حساب الدلالة العيادية	56– 65
• طريقة جاكبسون وترواكس	56– 60
• طريقة جليكسين لورد نوفيك	60– 61
• طريقة إدواردز ونانللي	61
• طريقة هاجمان وآرينديل	61–62
• الطريقة الخطية الهرمية	62
– مجالات توظيف الدلالة العيادية	66
– الانتقادات الموجهة إلى الدلالة العيادية	66– 67
ثانياً: التدريب وأثره في الخصائص القياسية	68– 84
1. مفهوم الخصائص القياسية	68– 69
2. أثر التدريب في الخصائص القياسية	70– 71
3. مفهوم إستراتيجيات أداء الاختبار	71– 72
4. التداخل بين الإستراتيجيات والحكمة الاختبارية	72 – 74

5. تصنيفات الإستراتيجيات كما وردت في الأبحاث	74 – 75
6. الإستراتيجيات المُستخدمة في البحث الحالي	75 – 82
• إدارة وقت الاختبار	75- 76
• تحليل التعليمات والأسئلة الاختبارية	76- 77
• التعامل مع ورقة الإجابة	77- 78
• الاستدلال الاستنباطي	78- 79
• التخمين	79- 80
• الضبط الذاتي	80- 81
• المراجعة	81- 82
7. أهمية إستراتيجيات أداء الاختبار	82- 83
خلاصة	84
الفصل الرابع: منهج البحث وإجراءاته	**رقم الصفحة 85- 109**
تمهيد	86
1. منهج البحث	86
2. مجتمع البحث وعينته	86- 89
3. متغيرات البحث	89
4. أدوات البحث	90- 102
5. إجراءات البحث	102 – 108
• إجراءات المنهج الوصفي التحليلي	102- 104
• إجراءات منهج التحليل البعدي	104
• إجراءات المنهج التجريبي	105 – 108
6. المعالجات الإحصائية	109
الفصل الخامس: عرض النتائج وتحليلها	**رقم الصفحة 110- 136**
تمهيد	111
1. الإجابة عن السؤال الأول	111 – 113
2. الإجابة عن السؤال الثاني	113- 114
3. الإجابة عن السؤال الثالث	115

4. الإجابة عن السؤال الرابع	115- 118
5. الإجابة عن السؤال الخامس	118- 126
6. الإجابة عن السؤال السادس	126- 131
7. الإجابة عن السؤال السابع	132- 136
خلاصة	136
الفصل السادس: مناقشة النتائج والمقترحات	**رقم الصفحة 137- 150**
تمهيد	138
1. مناقشة نتائج البحث	138- 148
2. قوة البحث وحدوده	148- 149
3. الصعوبات المتعلقة بإجراء البحث	149- 150
4. مقترحات البحث	150
خلاصة واستنتاج	150
ملخص البحث باللغة العربية	**151 – 153**
قائمة المراجع	**رقم الصفحة** 154- 164
الملاحق	**رقم الصفحة**
	165- 199
ملخص البحث باللغة الانكليزية	**رقم الصفحة** i– iv

فهرس الجداول

رقم الجدول	عنوان الجدول	رقم الصفحة
1	ملخص الأبحاث السابقة	25 -28
2	مقاييس حجم الأثر لفروق المتوسطات	41
3	مقاييس حجم الأثر للتباينات غير المصححة	44
4	مقاييس حجم الأثر للتباينات المصححة	46
5	مقاييس حجم الأثر للاختبارات اللامعلمية	48
6	محكات حجوم الأثر حسب درجات حرية	49
7	قوانين حساب الدلالة العيادية	63
8	مهارات إستراتيجية تحليل التعليمات والأسئلة	77
9	الخصائص الديموغرافية للعينة	89
10	الخصائص الديموغرافية لعينة Toni-4	91
11	فرق متوسطي الفئتين على اختبار TONI-4	92
12	فرق متوسطي الفئتين على الاختبار المحكّي	92
13	معاملات سهولة مصفوفات اختبار الذكاء	97
14	معاملات سهولة المصفوفات وفق المستويات	98
15	متوسط زمن الإجابة عن المصفوفات	99
16	جلسات البرنامج في التجربة الاستطلاعية	101- 102
17	عدد الأبحاث ونسبها المئوية	103
18	جلسات البرنامج في التجربة النهائية	106-107
19	نسب الدلالة العملية إلى جانب الإحصائية	111
20	نسب حجوم الأثر	112
21	كاي تربيع للدلالتين الإحصائية والعملية	113
22	تصنيفات حجوم الأثر والقرار الإحصائي	113
23	تصنيفات الدلالة العيادية والقرار الإحصائي	114

24	تصنيفات الدلالة العيادية وحجوم الأثر	115
25	تحليل الانحدار للدلالتين الإحصائية والعملية	116
26	تحليل الانحدار للدلالتين الإحصائية والعيادية	116
27	تحليل الانحدار للدلالتين العملية والعيادية	117
28	الدلالتان الإحصائية والعملية لأنواع تطبيق الاختبار	119
29	مؤشرات التغيُّر الموثوق لأنواع تطبيق الاختبار	123-119
30	التصنيفات العيادية لمراحل التطبيق	125
31	قيم الدلالة العيادية وفق طرائق حسابها	130-126
32	معامل كابا كوهين لتصنيفات الدلالة العيادية	131
33	نتائج (V) لدلالة الفرق بين معاملي الصدق المحكي	132
34	فروق متوسطات الفئتين العليا والدنيا في الاختبارين	132
35	توزع أفراد الفئتين العليا والدنيا في الاختبارين	134-133
36	مؤشر التغيُّر الموثوق للأفراد الذين تغيرت فئاتهم	135
37	فروق معاملي ثبات التطبيقين القبلي والبعدي المباشر	136
38	تزامُن نتائج الدلالات العلمية	141

فهرس الاشكال والرسوم البيانية

رقم الشكل	عنوان الشكل أو الرسم البياني	رقم الصفحة
1	هرم الدلالات العلمية	54
2	منحنى بياني لطرائق تحديد نقطة القطع	58
3	إستراتيجيات أداء الاختبار	83
4	تدفُّق عينة الرسائل الأكاديمية	88
5	أعمدة بيانية لتوزع أعمار أفراد العينة	89
6	النموذج العام لتصميم التعليم	93
7	هرم مستويات معاملات السهولة	98
8	خط بياني للنسب المئوية للأبحاث التجريبية	103
9	التصميم التجريبي للبحث	108
10	خط بياني لنسب تزايد حجوم الأثر	112
11	تمثيل بياني لتوزع الدرجات في الاختبار القبلي	124
12	تمثيل بياني لتوزع الدرجات في الاختبار البعدي المباشر	124
13	تمثيل بياني لتوزع الدرجات في الاختبار البعدي المؤجل	125
14	تمثيل بياني للتصنيفات العيادية وفق مراحل تطبيق الاختبار	126
15	تمثيل بياني للتصنيفات العيادية وفق طرائق حسابها	130

فهرس الملاحق

رقم الملحق	عنوان الملحق	رقم الصفحة
1	أسماء الخبراء المحكمين	166
2	نموذج جدول التحليل	167
3	نسب الأبحاث التجريبية	171-168
4	نسب قوانين حجم الأثر	174-172
5	مثال لحساب مؤشر التغيُّر الموثوق	177-175
6	قيم الدلالة العيادية للأبحاث المُجمّعة	183-178
7	قيم الدلالات العلمية للأبحاث المُجمّعة	193-184
8	معاملات سهولة مصفوفات اختبار الذكاء	194
9	الدرجات على اختباري SPM-PV وToni -4	199-195

قائمة الاختصارات

	الرمز أو الاختصار	المصطلح الاجنبي المقابل له	المصطلح العربي المقابل له
1	SS	Scientific Significances	الدلالات العلمية
2	Ss	Statistical Significance	الدلالة الإحصائية
3	PS	Practical Significance	الدلالة العملية
4	CS	Clinical Significance	الدلالة العيادية
5	ES	Effect Size	حجم الأثر
6	$p_{calculated}$	Calculated Probare Value	القيمة الاحتمالية المحسوبة
7	$p_{critical}$	Critical Probare Value	القيمة الاحتمالية النظرية
8	SPM-PV	Standard Progressive Matrices- Parallel Version	المصفوفات المتتابعة المعياري - النسخة الموازية
9	JT	Jacobson – Truax Method	طريقة جاكبسون وترواكس لحساب الدلالة العيادية
10	GLN	Gulliksen- Lord- Novick Method	طريقة جليكسين لورد نوفيك لحساب الدلالة العيادية
11	EN	Edwards- Nunnally Method	طريقة إدواردز ونانللي لحساب الدلالة العيادية
12	HA	Hageman-Arrindell Method	طريقة هاجمان وآريندیل لحساب الدلالة العيادية
13	HLM	Hierarchical Liner Method	الطريقة الخطية الهرمية
14	RCI	Reliable Change Index	مؤشر التغيُّر الموثوق
15	SE	Standard error off measurement	الخطأ المعياري للقياس
16	ADDIE	Analysis- Design- Development- Implementation- Evaluation	النموذج العام للتصميم التعليمي
17	APA	American Psychological Association	الجمعية النفسية الأمريكية
18	ASA	American Statistics Association	الجمعية الأمريكية للإحصاء