# التحليل البَعدي لثبات درجات اختبار ويسكونسن لتصنيف البطاقات

# The Meta-Analysis of Wisconsin Card Sorting Test Reliability

**نائلة الخروصي[1،*]، وحسين الخروصي[1]، و مي حلمي[1،2]**

**Naila Al Kharusi[1,*], Hussain Alkharusi[1] & Mai Helmy[1,2]**

[1]قسم علم النفس، كلية التربية، جامعة السلطان قابوس، سلطنة عُمان. [2]قسم علم النفس، كلية الآداب، جامعة المنوفية، مصر.

[1]Department of Psychology, College of Education, Sultan Qaboos University, Sultanate of Oman. [2]Department of Psychology, Faculty of Arts, Menoufia University, Egypt.

[*]الباحث المراسل: naila.al.kharusi@gmail.com

تاريخ التسليم: (2023/2/5)، تاريخ القبول: (2023/6/22)، تاريخ النشر: (2024/5/1)

DOI: 10.35552/0247.38.5.2188

**ملخص**

هدفت الدراسة إلى التحليل البعدي لمعاملات ثبات درجات اختبار ويسكونسن لتصنيف البطاقات لمعرفة القيمة التقديرية القياسية لمعامل ثبات درجات الاختبار، والكشف عن العوامل المختلفة التي قد تساهم في ارتفاع أو انخفاض قيمة معامل الثبات. ولقد خضعت (36) دراسة للتحليل البعدي لثبات، حيث بلغ عدد معاملات الثبات المحسوبة (206) معامل ثبات من (16) مؤشر من مؤشرات اختبار ويسكونسن لتصنيف البطاقات. وركزت الدراسة الحالية على (98) معامل ثبات محسوب بطريقة ثبات الإعادة لستة مؤشرات. بينت النتائج أن معاملات الثبات تراوحت بصفة عامة بين (0.3) و(0.7)، وأظهرت النتائج بأنه لا توجد فروق ذات دلالة إحصائية في معاملات ثبات الإعادة تعزى إلى المتغيرات المستقلة. وخلصت الدراسة إلى الحاجة إلى استخدام اختبارات أخرى إلى جانب اختبار ويسكونسن لتصنيف البطاقات للأغراض البحثية أو في العملية التشخيصية للكشف عن قصور الوظائف التنفيذية للدماغ.

**الكلمات المفتاحية:** التحليل البعدي، الثبات، اختبار ويسكونسن لتصنيف البطاقات.

**Abstract**

The study aimed at meta-analysis of the reliability coefficients of Wisconsin Card Sorting Test scores to find out the typical estimate value of the reliability coefficient of test scores, and to reveal the various factors that may contribute to the high or low value of the reliability coefficient. A total of (36) studies were subjected to meta-analysis of reliability, where the number of reliability coefficients calculated were (206) reliability coefficients out of (16) indicators of the Wisconsin Card Sorting Test. The current study focused on (98) reliability coefficients calculated by the method of test-retest reliability for six indicators. The results showed that the reliability coefficients generally ranged between (0.3) and (0.7). Also, the results showed that there were no statistically significant differences in the reliability coefficients attributed to the independent variables. The study concluded that other tests besides the Wisconsin Card Sorting Test should be used for research purposes or in the diagnostic process to detect insufficiency of the executive functions of the brain.

**Keywords:** Meta-Analysis, Reliability, Wisconsin Card Sorting Test

**المقدمة**

الإنسان كائن غامض ومتعدد القدرات والإمكانيات خلقه الله عز وجل في كون متميز، وميزه بالعقل فيه من قدرات تفوق في حجمها وإمكانياتها ما لدى الكائنات الأخرى جميعا. إذ يتميز هذا الجزء من الجسم الذي يقع داخل جمجمة محصنة بقدرات هائلة تمكن الإنسان من التفكير والتخطيط والإنجاز في مختلف مجالات الحياة. يحتوي عقل الإنسان على مجموعة من الوظائف المعرفية الأساسية التي تمكنه من مواجهة صعوبات الحياة المختلفة، والتكيف معها وتحقيق غايته في الحياة. وتعتبر الوظائف التنفيذية المعرفية لدماغ الجزء المتحكم بسلوك الإنسان السوي لأداء الأنشطة الحياتية من أجل تحقيق الهدف، وتشمل عمليات التحكم المجالات المعرفية، والانفعالية، والاجتماعية (Malika, 2010).

ويعَّرف مارتن (Martin, 2016) الوظائف التنفيذية على أنها العمليات المعرفية المسؤولة عن التعَّرف على المواقف ودلالاتها وتقييم أهميتها والقيام بخطط بديلة وسريعة للتعامل مع المتغيرات الطارئة. وتعَّرف أيضاً على أنها مجموعة من القدرات التي تمكِّن الفرد من الانخراط في سلوك فعَّال وغرضي يخدم الذات بنجاح (Abdul Qawi, 2017)، حيث أن الوظائف التنفيذية تشمل مجموعة من القدرات المعرفية العليا يتم استخدامها لتمثيل مجموعة من الآليات

العصبية المسؤولة عن توجيه، وتنسيق جوانب متعددة من الإدراك والعاطفة والسلوك. وبالتالي فإن حدوث خلل في هذه الوظائف قد يتسبب في العديد من الصعوبات في تنظيم سلوكه وتنظيم العواطف والخطط لإنجاز المهام والالتزام بها. أن الأثر المدمرة الذي قد يسببه القصور في هذه الوظائف دفعت علماء النفس والأعصاب والتربية إلى البحث والتحقق عن مدى ارتباطه بالاضطرابات النفسية والعصبية المختلفة وبإصابات المخ كذلك.

وكل ما وجد إنما وجد بمقدار وما وجد بمقدار يمكن قياسه. لذلك ابتكر عدد من الباحثين اختبارات لهذه الجوانب من القدرات، ومن أشهرها اختبار ويسكونسن لتصنيف البطاقات (Berg, 1948). ويعتبر اختبار ويسكونسن لتصنيف البطاقات من أبرز الاختبارات استخداماً في تقييم المرونة المعرفية، والكشف عن قصور في الوظائف التنفيذية ( Abdul Qawi, 2017). ويتم تطبيق الاختبار على المفحوص من خلال تصنيف البطاقات على أساس اللون أو الرقم أو الشكل. وأثناء قيام المفحوص بتصنيف البطاقات وبمجرد أن يقوم المفحوص بعددٍ معين من الردود الصحيحة المتتالية، يقوم الفاحص بدون تفسير بتغيير أساس التصنيف "الصحيح". حيث لا يجب على المفحوص أن يتعلم كيفية المحافظة على قاعدة المطابقة الصحيحة مع منع المحفزات غير ذات الصلة فقط، بل عليه أيضاً إظهار المرونة المعرفية في اكتشاف متى تغيرت القاعدة وتكييف اختياراته وفقاً لذلك (Grant & Berg, 1948). ويتضمن الملحق (1) موجز تعريفي لمؤشرات الاختبار.

وهناك دراسات أولت اهتماماً بالخصائص السيكومترية لاختبار ويسكونسن لتصنيف البطاقات عند تطبيقه على عينات مختلفة في الحجم والخصائص، كدراسة أمير (Amir, 2018)، ودراسة مريندا وزملائه (Miranda, *et al.* 2019)، ودراسة ستينك وزملاءه (Steinke, *et al.* 2021)، ودراسة تشيو وزملائه (Chiu, *et al.* 2021). وقد أظهرت نتائج تحليل ثبات درجات الاختبار العديد من التناقضات، حيث تراوحت معاملات الثبات لدرجات اختبار ويسكونسن لتصنيف البطاقات بين (0.02) و(0.96)، مما قد يشكل تحدياً لمستخدمي الاختبار حول درجة دقة القرارات المتعلقة بالنتائج، ومع صعوبة استخلاص نتيجة مشتركة بين الدراسات السابقة فيما يخص خصائص اختبار ويسكونسن لتصنيف البطاقات كان لزاماً البحث عن نهج يقوم بتحليل وتقييم نتائج الدراسات السابقة بأسلوب منظم وعلمي والخروج بتقديرات أكثر دقة لاختبار ويسكونسن لتصنيف البطاقات وذلك عن طريق التحليل البَعدي (Mikolajewicz & Komarova, 2019).

ويعّرف التحليل البَعدي على أنه مدخل لتركيب نتائج البحوث ويعتمد على التحليل الإحصائي في معالجة البيانات ليطبق على نتائج الدراسات الكمية والفردية ( Al Obeidi, 2019). ويسمح التحليل البَعدي للباحثين بدمج نتائج دراسات سابقة متعددة إحصائيا والناتجة عن تطبيق اختبار معين على عينات وسياقات مختلفة (Chertok & Haile, 2018). ولقد اهتمت الدراسة الحالية بإجراء تحليلٍ بَعدي لتقديرات ثبات درجات اختبار ويسكونسن لتصنيف البطاقات في الدراسات السابقة، حيث اتبعت الدراسة طريقة فاشا-هاس كأسلوب لتحليل البَعدي

للثبات وذلك لدراسة مصادر التباين في معاملات الثبات، حيث تعتبر قيم معاملات الثبات متغيرات تابعة، بينما تعتبر الخصائص المنهجية للدراسات السابقة المتعلقة بالعينة والأداة وطرائق تقدير الثبات وظروف التطبيق متغيرات مستقلة (Vacha-Haase, 1998).

**الدراسات السابقة**

لا توجد دراسات سابقة – في حدود علم الباحثين – قامت بتقييم الخصائص السيكومترية لاختبار ويسكونسن لتصنيف البطاقات سوى دراسة لكلميا وزملاءه (Calamia, *et al*. 2013) والتي هدفت إلى تقديم ملخص كمي باستخدام التحليل البَعدي لثبات الإعادة على مجموعة من الاختبارات النفسية-العصبية في الولايات المتحدة، حيث خضعت (9) اختبارات للتحليل البَعدي وهي: اختبار فرز البطاقات لويسكونسن، واختبار بوسطن للتسمية، واختبار الاحتفاظ البصري لبنتون (الجزء أ)، واختبار الشكل المركب، واختبار رابطة الكلمات الشفوية، واختبار التعلم اللفظي لكاليفورنيا، واختبار راي التعلم اللفظي السمعي، واختبار توصيل الدوائر، ومقياس ذكاء واكسلر للكبار. وتراوحت القيمة التقديرية لمعاملات ثبات الإعادة من متوسط إلى مرتفع ( r≥ 0.7). وكان التأثير المقدر لطول الفترة الزمنية بين الاختبار وإعادة الاختبار سالباً، وارتبط استخدام نموذج بديل عند إعادة الاختبار بانخفاض في قيمة معامل الثبات، كما كانت معاملات الثبات في العينات الإكلينيكية منخفضة مقارنة بالأصحاء.

أما بالنسبة لأمثلة الدراسات السابقة التي وظفت التحليل البعدي لتقييم ثبات المقاييس، فهناك دراسة لكاروسو (Caruso, 2000) باستخدام التحليل البَعدي للتوصل إلى القيمة التقديرية القياسية لمعامل ثبات درجات مقياس الشخصية نيو. تكونت العينة من (51) دراسة استخدمت إصدارات مختلفة لمقياس الشخصية نيو. اعتمدت درجات الثبات لعينة الدراسة على المجالات الخمسة لمقياس الشخصية، وأظهرت الدراسة قدراً كبيراً من التباين في الثبات بين مجالات الشخصيات المختلفة. كما أظهرت النتائج بأن درجات مجال التوافق كانت الأضعف، لاسيما في العينات السريرية، والعينات المخصصة للذكور فقط. كما هدفت دراسة فان وين ( Fan & Yin, 2003) إلى معرفة القيمة التقديرية القياسية لمعامل ثبات درجات قائمة بيك للاكتئاب، ومعرفة العوامل التي تؤثر على معامل الثبات. استخدم الباحثين (7.5%) فقط من الدراسات التي تمت مراجعتها عن تقدير الثبات، وكشفت الدراسة بأن خطأ القياس في تقدير ثبات إعادة الاختبار كان أكبر بكثير من خطأ قياس تقدير ثبات الاتساق الداخلي لعدم تجانس العينة، كما أن تقديرات الثبات التي تشمل مدمني المواد المخدرة كانت أقل من تقديرات الثبات المحسوبة من عينات أشخاص عاديين، وربما يرجع ذلك إلى تقييد مشاكل النطاق. وكشفت تحليلات الارتباط أن الأخطاء القياسية للقياس لم تكن مرتبطة بتقديرات الثبات، ولكنها كانت مرتبطة إلى حد كبير بالانحرافات المعيارية لدرجات الاختبار، مما يشير إلى أنه ينبغي النظر في الخطأ المعياري للقياس وتقديرات الثبات معاً عند تفسير درجات الاختبار الفردية.

وأجرى الهنائي وآخرون (Al-Hinai, *et al*. 2019) دراسة تهدف إلى استخراج القيمة التقديرية القياسية لمعاملات ثبات درجات اختبار رايفن للمصفوفات المتتابعة والأخطاء

المعيارية للقياس المناظرة لها، حيث طبق التحليل البَعدي على (56) دراسة شملت اختبار رايفن بإصداراته المختلفة. ولقد بلغت القيمة التقديرية لمعامل ثبات درجات اختبار رايفن للمصفوفات المتتابعة (0.85) بطريقة ألفا كرونباخ بخطأ معياري قدره (2.51). كما بلغت القيمة التقديرية لمعامل ثبات الإعادة (0.76) بخطأ معياري قدره (3.43). كما هدفت دراسة لسانتوس وآخرون (Santos, *et al*. 2022) إلى تقييم الخصائص السيكومترية لمقياس سوء استخدام المخدرات باستخدام التحليل البَعدي. وقد كشفت نتائج (387) دراسة أن (65%) من تقديرات معامل ثبات ألفا كرونباخ كانت بين المتوسط إلى المرتفع. كما ظهر (44%) من نتائج تقديرات كابا للثبات بين المقيمين في نطاق بين المتوسط إلى المرتفع. بالإضافة إلى ذلك، كانت (69%) و(97%) و(37%) و(96%) من التقديرات المركبة على أبعاد المقياس المتمثلة في الحساسية والنوعية والقيمة التنبؤية الإيجابية والقيمة التنبؤية السلبية على التوالي في النطاق بين المتوسط والمرتفع.

**مشكلة الدراسة**

تنبثق مشكلة الدراسة الحالية من خلال وجود تباين في معاملات الثبات للقياس بين الدراسات السابقة والتي طبقت اختبار ويسكونسن لتصنيف البطاقات، ومما لا شك فيه أن هذا التباين الكبير في معامل الثبات بين الدراسات السابقة قد يشكك من دقة تقدير الأداء على الاختبار، وبالتالي صعوبة الوثوق في القرارات المتعلقة بنتائج الاختبار، وقد اختار فريق البحث اختبار ويسكونسن لتصنيف البطاقات لكونه من أكثر اختبارات الوظائف التنفيذية شيوعاً واستخداماً في المجالات النفسية، وقد تم تطبيقه في دول عديدة على عينات متباينة، ولذا فإن التحليل البَعدي للثبات في الدراسة الحالية قد يساعد على دعم القرارات المبنية على نتائج اختبار ويسكونسن لتصنيف البطاقات، كما أن قلة استخدام أسلوب التحليل البَعدي في دراسات الثبات في البيئة العربية، والذي يقابله اهتمام كبير في البيئات الغربية، كان محفزاً لفريق البحث من أجل الاستفادة من هذا الاتجاه في مجال القياس.

**أسئلة الدراسة**

هدفت الدراسة الحالية إلى الإجابة على الأسئلة التالية:

1. ما القيمة التقديرية القياسية لمعامل ثبات درجات اختبار ويسكونسن لتصنيف البطاقات؟
2. ما نسبة إسهام خصائص الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات (جنس العينة، طبيعة العينة، الفئة العمرية للعينة، إصدار الاختبار، الفترة الزمنية بين مرتي التطبيق) في تباين معاملات ثبات درجات الاختبار؟

**أهمية الدراسة**

يمكن أن نجمل أهمية هذه الدراسة في الجوانب الآتية:

1. الأهمية النظرية، وتتمثل في إنها تتناول موضوعًا حيويا يتعلق بالتحليل البعدي لثبات الاختبارات، فهي تسعى إلى إظهار العوامل المؤثرة في معامل ثبات درجات الاختبار،

والبحث عن الطريقة المناسبة في تقدير ثبات درجات الاختبار، ومعرفة مؤشرات الاختبار الأكثر حساسية للسمة المقاسة. وبالتالي، فأنه من المتوقع أن تسهم الدراسة الحالية في فتح باب للدراسات في هذا المجال وفق لما تتوصل إليه من نتائج.

2. الأهمية التطبيقية، وتتمثل في إن نتائج الدراسة الحالية قد توضح مدى أهمية استخدام اختبار ويسكونسن لتصنيف البطاقات كمحك خارجي لكثير من الاختبارات المتعلقة بالوظائف التنفيذية في التأكد من ثباتها، وقد تمكن نتائج الدراسة من معرفة الظروف المعيارية للتطبيق التي تسهم في ارتفاع أو انخفاض معامل ثبات درجات الاختبار، مما قد يساعد في التحكم بها أثناء تطبيق الاختبار، وتقديم مؤشرات استرشادية لمعامل الثبات لمستخدمي اختبار ويسكونسن لتصنيف البطاقات.

**أهداف الدراسة**

هدفت الدراسة الحالية إلى:

1. تحديد القيمة التقديرية القياسية لمعامل ثبات درجات اختبار ويسكونسن لتصنيف البطاقات.
2. معرفة نسبة إسهام خصائص الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات في تباين معاملات ثبات درجات الاختبار.

**مصطلحات الدراسة**

**الثبات:** يعرف علام (Allam, 2000) ثبات درجات الاختبار من خلال اتساق الدرجات أثناء قياس سمة معينة في ظروف متباينة ومدى خلوها من الأخطاء غير المنتظمة التي تشوب القياس، أي قدرة الاختبار على قياس المقدار الحقيقي لسمة التي يهدف لقياسها. وإجرائيًا تعّرف الدراسة الحالية الثبات بأنه الاتساق في درجات اختبار ويسكونسن لتصنيف البطاقات عند إعادة تطبيقه على نفس المفحوصين بعد فترات زمنية مختلفة.

**التحليل البَعدي:** يعرف ميكولايفيتش وكوماروفا (Mikolahewicz & Komarova, 2019) التحليل البعدي بأنه تحليل إحصائي للبيانات التي تم جمعها من دراسات أولية سابقة والتي تركز الإجابة على نفس السؤال البحثي بهدف الحصول على قيمة تقديرية قياسية للظاهرة المدروسة. وإجرائيًا تركز الدراسة الحالية على التحليل البعدي لثبات درجات اختبار ويسكونسن لتصنيف البطاقات من خلال التحليل الإحصائي لمعاملات الثبات المتعددة الناتجة عن تطبيق الاختبار على عينات وسياقات مختلفة في الدراسات السابقة بهدف تقدير متوسط معاملات الثبات وتحليل التباين بينها.

**اختبار ويسكونسن لتصنيف البطاقات:** يعد اختبار ويسكونسن من أكثر الاختبارات فائدة وحساسية في مجال قياس وظائف المنطقة الأمامية من الفص الجبهي لدماغ. ويقيس الاختبار الوظائف التنفيذية للدماغ من خلال قياس قدرة المفحوص على التصنيف على أساس معين، كما

يقيس القدرة على استخدام المفاهيم ومنع أو كف الاستجابة والتفكير المجرد وتسلسل السلوك بالإضافة إلى المرونة المعرفية (Abdul Qawi, 2017).

**حدود الدراسة**

**الحدود الموضوعية:** من حيث الموضوع، فإن الدراسة الحالية اقتصرت على التحليل البعدي لثبات درجات اختبار ويسكونسن لتصنيف البطاقات بجميع إصداراته المختلفة (النسخة القياسية، والنسخة المختصرة، والنسخة المعدلة). ومن حيث الأداة، فإن عملية جمع البيانات من الدراسات السابقة والترميز تمت من خلال الاستعانة باستمارة لتفريغ البيانات وترميزها من إعداد فريق البحث وفق الخطوات التي اقترحها هينسون (Henson, 2001) لإعداد استمارة التحليل البعدي للثبات.

**الحدود المكانية:** اقتصر البحث على الدراسات المتوافرة في قواعد البيانات البحثية بجامعة السلطان قابوس، وقواعد بيانات أخرى مثل Wiley Online Library، و Taylor & Francis Online journal، و Springer E-References collection، و SAGE journals، و Oxford University Press journals، و Dissertations database، وThe BMJ medical journal، وباستخدام الكلمات المفتاحية " Wisconsin Card Sorting Test" و"اختبارات ويسكونسن لتصنيف البطاقات" إضافة إلى الدراسات المنشورة في الدوريات العربية، وتلك الواردة في قوائم مراجع الدراسات المتوافرة.

**الحدود الزمانية:** اقتصرت الدراسة الحالية على الدراسات المنشورة باللغتين العربية والاجنبية في الفترة ما بين (1993-2021).

**منهج الدراسة**

اعتمدت الدراسة الحالية على المنهج الوصفي باستخدام التحليل البعدي (Meta-Analysis) وتم اتباع طريقة فاشا-هاس (Vacha–Haase, 1998) للتحليل البعدي لثبات درجات اختبارات ويسكونسن لتصنيف البطاقات.

**مجتمع الدراسة وعيّنتها**

يشمل مجتمع الدراسة الحالية جميع الدراسات باللغتين العربية والانجليزية في الفترة ما بين 1993-2021 وذلك لتشمل الدراسة أكبر عدد من الدراسات التي استخدمت اختبارات ويسكونسن لتصنيف البطاقات بجميع إصداراته (النسخة الأصلية، النسخة المحوسبة من الإصدار الأصلي النسخة المختصرة، والنسخة المعدلة) والمتوافرة في قواعد البيانات البحثية بجامعة السلطان قابوس، وقواعد بيانات أخرى مثل Wiley Online Library، و Taylor & Francis Online journal، و Springer E-References collection، و SAGE journals، و Oxford University Press journals، و Dissertations database، وThe BMJ medical journal، باستخدام الكلمات المفتاحية "Wisconsin Card

"Sorting Test" و"اختبارات ويسكونسن لتصنيف البطاقات" إضافة إلى الدراسات المنشورة في الدوريات العربية، وتلك الواردة في قوائم مراجع الدراسات المتوافرة.

وقد أسفر هذا البحث عن (4657) دراسة، وبلغ عدد الدراسات التي تم الاطلاع عليها (417) دراسة، ثم تم تقسيم الدراسات إلى أربع مجموعات وفق ما يلي:

1. الدراسات النظرية التي ذكرت اختبار ويسكونسن لتصنيف البطاقات، ولم تستخدمه، وعددها (104) دراسة، وتمثل (25%) من الدراسات التي تم الاطلاع عليها.
2. الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات، ولم تذكر أي معلومات عن الثبات، ولم تشر إليه من دراسات أخرى، وعددها (161) دراسة، وتمثل (38.6%) من الدراسات التي تم الاطلاع عليها.
3. الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات، وتحدثت عن الثبات من خلال معلومات مستقاة من دراسات سابقة، وعددها (116) دراسة، وتمثل (27.8%) من الدراسات التي تم الاطلاع عليها.
4. الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات، وذكرت معلومات عن الثبات من البيانات المتاحة في الدراسة نفسها، وعددها (36) دراسة، وتمثل (8.6%) من الدراسات التي تم الاطلاع عليها.

وعليه، بلغ عدد الدراسات التي خضعت للتحليل البعدي (36) دراسة، ويمكن للقارئ الحصول على قائمة الدراسات التي خضعت للتحليل البعدي من فريق البحث. ويلخص الجدول 2 خصائص تلك الدراسات وفق المتغيرات التصنيفية للدراسة. وقد تم استخراج معاملات الثبات الواردة في تلك الدراسات لكل مؤشر من مؤشرات الاختبار، ويلخص الجدول 3 ذلك.

**جدول (1):** خصائص الدراسات التي خضعت للتحليل البعدي لثبات درجات اختبار ويسكونسن لتصنيف البطاقات.

| المتغير | الفئات | العدد | النسبة المئوية |
|---|---|---|---|
| نوع معامل الثبات | معامل ألفا | 8 | %22.22 |
| | التجزئة النصفية | 2 | %5.56 |
| | إعادة الاختبار | 15 | %41.67 |
| | الصور المتوازية | 0 | %0 |
| | معاملات الارتباط بين الفئات | 3 | %8.33 |
| | إمكانية التعميم | 6 | %16.66 |
| | أكثر من نوع من معامل الثبات | 2 | %5.56 |
| جنس العينة | ذكور | 2 | %5.56 |
| | إناث | 2 | %5.56 |
| | غير محدد | 32 | %88.88 |

... تابع جدول رقم (1)

| المتغير | الفئات | العدد | النسبة المئوية |
|---|---|---|---|
| طبيعة العينة | فئة العاديين (الأصحاء) | 18 | %50 |
| | الفئة الخاصة (غير الأصحاء) | 11 | %30.56 |
| | غير محدد | 7 | %19.44 |
| الفئة العمرية | 6-11 سنة | 1 | %2.76 |
| | 12-22 سنة | 3 | %8.34 |
| | 23 سنة فأكثر | 13 | %36.12 |
| | غير محدد | 19 | %52.78 |
| نوع الاختبار | الإصدار التقليدي | 24 | %61.11 |
| | إصدار الحاسوب | 4 | %11.11 |
| | الإصدار المختصر | 3 | %8.33 |
| | الإصدار المعدل | 5 | %13.89 |
| | أكثر من إصدار | 2 | %5.56 |
| الفترة الزمنية | 1993-2002 | 15 | %41.66 |
| | 2003-2012 | 10 | %27.78 |
| | 2013-2021 | 11 | %30.56 |
| بيئة التطبيق | عربية | 5 | %13.89 |
| | غير عربية | 31 | %86.11 |

يوضح الجدول (1) أن أكثر من نصف الدراسات 61.11% طبقت الإصدار الأصلي من اختبار ويسكونسن لتصنيف البطاقات، وأن 41.67% من الدراسات استخدمت طريقة إعادة الاختبار، و22.22% من الدراسات طبقت طريقة ألفا و5.56% من الدراسات طبقت أكثر من طريقة لحساب الثبات. وتبين بأن جل الدراسات طبقت على بيئة غير عربية (31 دراسة، بنسبة 86.11%)، كما أن أكثر 36.12% من الدراسات طبقت الاختبار على فئة الراشدين ولم يركز. وأيضا لوحظ بأن معظم الدراسات شملت الجنسين معا و5.56% فقط ركزت على الذكور و5.56% فقط ركزت على الإناث، ونصف الدراسات 50% من الدراسات ركزت على الأصحاء. وقد بلغ مجموع حجوم العينات في الدراسات (17423) مشاركاً، حيث تفاوت حجم العينة في الدراسات ما بين (10) مشارك إلى (752) مشارك بمتوسط (78.13) وانحراف معياري (114.56). وبالنسبة للفترة الزمنية بين مرتي التطبيق لثبات إعادة الاختبار، فقد تراوحت بين يوم واحد و(900) يوم بمتوسط (3.54.5) يوم تقريبا وبانحراف معياري (3.65.81).

**جدول (2):** عدد معاملات الثبات المحسوبة لكل مؤشر من مؤشرات اختبار ويسكونسن لتصنيف البطاقات.

| المؤشر | معامل ألفا | التجزئة النصفية | إعادة الاختبار | الصور المتوازية | إمكانية التعميم | معاملات ارتباط بين الفئات | المجموع |
|---|---|---|---|---|---|---|---|
| مجموع الاستجابات | | | 3 | | | | 3 |
| الاستجابات الصحيحة | | | 7 | | | | 7 |
| الاستجابات الخاطئة | | | 16 | | 3 | | 19 |
| نسبة الاستجابات الخاطئة | | | 2 | 1 | 1 | | 4 |
| الاستجابات المثابرة | | | 16 | | 3 | 7 | 26 |
| نسبة استجابات المثابرة | | | | | 1 | | 1 |
| أخطاء المثابرة | 3 | 5 | 21 | 1 | 1 | 6 | 37 |
| نسبة أخطاء المثابرة | | | 2 | | 1 | | 3 |
| الأخطاء التي ليس لها مثابرة | 1 | 1 | 16 | | 1 | 6 | 25 |
| نسبة الأخطاء التي ليس بها مثابرة | | | | | 1 | | 1 |
| استجابات المستوى المفاهيمي | | | 5 | | | | 5 |
| نسبة استجابات المستوى المفاهيمي | | | 7 | | 1 | | 8 |
| الفئات المكتملة | 3 | 2 | 17 | 1 | | | 23 |
| استجابات إكمال المجموعة الأولى | | | 2 | | | | 2 |
| الفشل في الحفاظ على المجموعة | | 2 | 12 | 1 | | | 15 |
| القدرة على التعلم | | | 4 | | | 1 | 5 |
| الاختبار ككل | 22 | | | | | | 22 |
| المجموع | 29 | 10 | 130 | 4 | 13 | 20 | 206 |

يوضح الجدول (2) أن مؤشر استجابات أخطاء المثابرة حاز على اهتمام معظم الباحثين الذين قاموا بتطبيق اختبار ويسكونسن لتصنيف البطاقات في الدراسات التي أجروها، حيث بلغ عدد معاملات الثبات المحسوب للمؤشر 37 معامل ثبات، وتليها استجابات المثابرة والاستجابات التي ليس بها مثابرة والفئات المكتملة والاستجابات الخاطئة والفشل في الحفاظ على مجموعة حيث بلغ عدد معاملات الثبات المحسوبة لكل منهما 26، و25، و23، و19، و15 معامل ثبات على التوالي. أما بالنسبة للمؤشرات الأخرى فقد حازت على اهتمام أقل من الباحثين حيث تراوحت عدد معاملات الثبات المحسوب لبقية المؤشرات بين 1 إلى 8 معامل ثبات. كما يوضح

الجدول أن أعلى عدد لمعاملات الثبات المحسوبة كانت عن طريق ثبات الإعادة إذ بلغ عددها 130 معامل ثبات وهذا العدد يعتبر مرتفعا مقارنة بمعاملات الثبات المستخرجة بالطرق الأخرى والتي تراوح بين 4 إلى29 معامل ثبات.

في ضوء ملخص البيانات في الجدول (2)، فأن الدراسة الحالية ركزت على التحليل البعدي لمعاملات ثبات الإعادة لستة مؤشرات من مؤشرات اختبار ويسكونسن لتصنيف البطاقات وهي: الاستجابات الخاطئة، والاستجابات المثابرة، وأخطاء المثابرة، والأخطاء التي ليس لها مثابرة، والفئات المكتملة، والفشل في الحفاظ على المجموعة؛ نظرا لقلة عدد معاملات الثبات المحسوبة (أقل من 10) بالطرق الأخرى وقلة عدد معاملات الثبات لمؤشرات الاختبار الأخرى مما يضعف من القوة الإحصائية (Statistical Power) للتحليل (Valentine, *et al*. 2010).

**أداة الدراسة**

لتحقيق أهداف الدراسة، قام فريق البحث بإعداد **استمارة ترميز (Coding Form)** كأداة لجمع البيانات من الدراسات، وقد اشتملت الاستمارة على ثلاثة أجزاء، تناول الجزء الأول بيانات عامة عن الدراسة من حيث رقم الدراسة، وأسماء المؤلفين، وعنوان الدراسة، وسنة النشر. واشتمل الجزء الثاني على ترميز متغيرات الدراسة ومستويات كل متغير من المتغيرات التصنيفية المتمثلة في بيئة الدراسة، ونوع إصدار اختبار ويسكونسن لتصنيف البطاقات، ومؤشرات الاختبار، والمرحلة العمرية، وجنس العينة، وطبيعة العينة، ونوع معامل الثبات، هذا بالإضافة إلى بيانات المتغيرات الكمية وهي الفترة الزمنية لثبات الإعادة، وحجم العينة. وتضمن الجزء الثالث تسجيل قيم معامل الثبات المحسوبة لكل مؤشر من مؤشرات الاختبار وفق الطرق المختلفة للثبات. وقد اتبع فريق البحث الخطوات التي اقترحها هينسون (Henson, 2001) لإعداد استمارة الترميز من حيث الاستنارة بالدراسات المشابهة لموضوع التحليل البعدي للثبات (مثل: Graham & Christiansen, 2009; Lopez-Nicolas, *et al.* 2021 Nunez-Nunez, *et al.* 2022)، ثم تحليل الدراسات التي طبقت اختبار ويسكونسن لتصنيف البطاقات وحصر جميع المتغيرات التي تضمنتها.

**صدق أداة الدراسة**

للتحقق من الصدق الظاهري لأداة الدراسة، عرض فريق البحث الاستمارة في صورتها الأولية على (6) محكمين من أعضاء هيئة التدريس المتخصصين في القياس والتقويم، وقد تم الحصول على ردود من (3) محكمين، حيث طلب من المحكمين إبداء الرأي حول مدى وضوح وملائمة نظام الترميز، وشمولية المتغيرات ومستوياتها. وقد أبدى المحكمون مناسبة نظام الترميز ووضوحه لأغراض الدراسة، مع ضرورة الأخذ في الاعتبار نوع الدرجة لكل مؤشر ونوع معادلة الثبات المستخدمة. وقد قام فريق البحث بالتعديل على استمارة الترميز.

**ثبات أداة الدراسة**

للتأكد من ثبات الأداة قام فريق البحث بتحليل عينة عشوائية قدرها (5) دراسات من عينة الدراسة، وتم الاستعانة بمحللتين (ماجستير في تخصص علم نفس)، لتحليل نفس العينة المختارة باستخدام استمارة الترميز، وذلك لحساب ثبات المقدرين (Inter-rater reliability). ولزيادة درجة الاتساق بين المحللتين، عقد فريق البحث اجتماعاً مع المحللتين لغرض التعريف بآلية التحليل والتدريب عليها. وبعد انتهاء عملية التحليل قام فريق البحث بحساب بنود الاتفاق وبنود الاختلاف، ومن ثم حساب معامل الثبات باستخدام معامل ثبات هولستي (Holsti, 1969)، وقد تراوحت معاملات الاتفاق بين (0.8) و(1) بمتوسط (0.93)، وهو يعد معامل ثبات مقبول لأغراض الدراسة (Mouter & Noordegraaf, 2012). ويوضح الجدول (3) معاملات ثبات المقدرين لأداة الدراسة.

**جدول (3):** معاملات الثبات المقدرين لأداة الدراسة.

| متغيرات الدراسة | العدد الكلي للدراسات | الدراسات المتفق عليها | الدراسات التي لم يتم الاتفاق عليها | معامل الثبات |
|---|---|---|---|---|
| بيئة التطبيق | 5 | 5 | 0 | 1 |
| إصدار الاختبار | 5 | 4 | 1 | 0.8 |
| الفئة العمرية | 5 | 5 | 0 | 1 |
| نوع العينة | 5 | 4 | 1 | 0.8 |
| حجم العينة | 5 | 5 | 0 | 1 |
| نوع معامل الثبات | 5 | 5 | 0 | 1 |

**إجراءات الدراسة**

قام فريق البحث بدراسة التحليل البعدي للثبات وفق الخطوات التالية:

1. جمع الدراسات الصالحة للتحليل وهي الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات، وذكرت معلومات عن الثبات من البيانات المتاحة في الدراسة نفسها، وعددها (36) دراسة.
2. إعداد استمارة الترميز لجمع البيانات من الدراسات، والتحقق من الصدق الظاهري والثبات لنظام الترميز.
3. تصميم جدول لتفريغ البيانات في برنامج أكسل.
4. معالجة البيانات إحصائيا باستخدام برنامج SPSS وMedCalc.

**تحليل البيانات**

تم إجراء التحليلات الآتية:

1. تم تحويل معاملات ثبات الإعادة المحسوبة بمعامل كاندل (Kendall's τ) إلى معامل ارتباط بيرسون (*r*) باستخدام المعادلة [$r = sin\ (0.5\pi\tau)$] (Walker, *et al.* 2003)؛ بحيث تكون جميع معاملات الثبات وفق معامل ارتباط بيرسون.
2. تحويل معاملات الارتباط إلى قيم زائية (*z*) باستخدام فيشر (Fisher's *r* to *z*)؛ وذلك للحصول على التوزيع الاعتدالي للمعاملات (Lipsey & Wilson, 2001)، ثم التحويل العكسي من (*z*) إلى (*r*) عند وصف الإحصاءات الوصفية لمعاملات الثبات.
3. استخدام نموذج الآثار العشوائية (Random-effects model) لحساب متوسط معاملات الثبات واستخراج فترة الثقة عند مستوى 95%، حيث يفترض نموذج الآثار العشوائية أن حجم الأثر يختلف من دراسة إلى أخرى في مجتمع الدراسات، بينما يفترض نموذج الآثار الثابتة (Fixed-effects model) أن كل الدراسات في المجتمع لها نفس حجم الأثر (Borenstein, *et al.* 2009).
4. فحص الاختلاف بين نتائج الدراسات الفردية (Heterogeneity) باستخدام إحصاء (*Q*)، وتحديد نسبة الاختلاف بينها باستخدام إحصاء ($I^2$) (Cooper, *et al.* 2019)، حيث تمثل النسب 25%، 50%، و75% مقدار اختلاف صغير، ومتوسط، وكبير على التوالي. وفي حالة وجود اختلاف دال إحصائيا بمقدار ($I^2 > 25\%$)، يتم إجراء تحليل لمعرفة مصدر الاختلاف وفق المتغيرات المستقلة في الدراسة (Higgins & Thompson, 2002). حيث تمثل معاملات الثبات المتغير التابع، بينما تمثل المتغيرات (حجم العينة، طبيعة العينة، الفئة العمرية، الفترة الزمنية بين مرتي التطبيق) متغيرات مستقلة.

**نتائج الدراسة ومناقشتها**

**نتائج السؤال الأول ومناقشتها: ما القيمة التقديرية القياسية (Typical Estimate) لمعامل ثبات درجات اختبار ويسكونسن لتصنيف البطاقات؟**

يلخص الجدول (4) متوسط معامل ثبات الإعادة، وفترة الثقة (95%) لمتوسط معامل الثبات، ونتائج فحص الاختلاف بين الدراسات وفق إحصاء (*Q*) وإحصاء ($I^2$) لكل مؤشر من مؤشرات اختبار ويسكونسن لتصنيف البطاقات.

**جدول (4):** الإحصاءات الوصفية لمعامل ثبات الإعادة لكل مؤشر من مؤشرات اختبار ويسكونس لتصنيف البطاقات.

| المؤشر | العدد | المتوسط | فترة الثقة (95%) | | قيمة ($Q$) | قيمة ($I^2$) |
|---|---|---|---|---|---|---|
| | | | الحد الأدنى | الحد الأعلى | | |
| الاستجابات الخاطئة | 16 | 0.61 | 0.49 | 0.71 | 69.94** | 78.55% |
| الاستجابات المثابرة | 16 | 0.7 | 0.61 | 0.77 | 59.04** | 74.59% |
| أخطاء المثابرة | 21 | 0.6 | 0.49 | 0.69 | 85.99** | 77.9% |
| الأخطاء التي ليس لها مثابرة | 16 | 0.55 | 0.46 | 0.63 | 37.37** | 59.86% |
| الفئات المكتملة | 17 | 0.64 | 0.52 | 0.73 | 92.46** | 82.7% |
| الفشل في الحفاظ على المجموعة | 12 | 0.3 | 0.16 | 0.42 | 21.95* | 49.88% |

*دال إحصائيا عند قيمة احتمالية < من 0.05. **دال إحصائيا عند قيمة احتمالية < من 0.01.

يتضح من الجدول (4) أن مؤشر استجابات المثابرة كان له أعلى معامل ثبات في المتوسط بمقدار (0.70) بينما جاء مؤشر الفشل في الحفاظ على المجموعة بأقل معامل ثبات في المتوسط بمقدار (0.3). وبصفة عامة، يلاحظ أن معاملات ثبات الإعادة جاءت في المتوسط أقل من (0.80) وهو المستوى المقبول للثبات للأغراض البحثية وأقل من (0.90) وهو المستوى المقبول للثبات للأغراض الإكلينيكية (Charter, 2003; Nunnally & Bernstein, 1994).

كما أظهرت نتائج فحص عدم التجانس (Heterogenity Test) وجود اختلاف بدرجة ما بين المتوسطة والكبيرة ($I^2 > 25\%$) دال إحصائيا بين الدراسات السابقة في معاملات ثبات الإعادة لجميع مؤشرات اختبار ويسكونسن لتصنيف البطاقات. وعليه، فأن السؤال الثاني من الدراسة يحاول معرفة العوامل التي ساهمت في تباين معاملات ثبات الإعادة لدرجات اختبار ويسكونسن لتصنيف البطاقات على مؤشرات الاستجابات الخاطئة، والاستجابات المثابرة، وأخطاء المثابرة، والأخطاء التي ليس لها مثابرة، والفئات المكتملة، والفشل في الحفاظ على مجموعة.

إن من أهم النتائج التي يسعى الباحث للتوصل إليها في دراسة التحليل البعدي للثبات هي إيجاد القيمة التقديرية القياسية لمعامل ثبات درجات الاختبار ومعرفة العوامل التي تسهم في انخفاض او ارتفاع قيمة معامل الثبات (Vacha-Haase, 1998)، وقد بينت نتائج الدراسة الحالية ان معاملات ثبات درجات اختبار ويسكونسن لتصنيف البطاقات بطريقة ثبات الإعادة تراوحت بصفة عامة بين (0.3) و(0.7)، وهي أقل من المستوى المقبول عادة للثبات للأغراض البحثية والإكلينيكية (Charter, 2003; Nunnally & Bernstein, 1994)، وهذا يدعو إلى الحاجة إلى استخدام اختبارات أخرى إلى جانب اختبار ويسكونسن لتصنيف البطاقات في العملية التشخيصية للكشف عن أعراض اضطراب الوظائف التنفيذية، ومن الناحية السيكومترية فأن

انخفاض معامل ثبات الإعادة قد يعزى الى أن طريقة إعادة التطبيق تضع في الاعتبار الأخطاء العشوائية الناجمة عن ظروف التطبيق والتغيرات التي تطرأ على المفحوصين، مما قد يجعل القيمة التقديرية منخفضة (Allam, 2006).

أن انخفاض قيمة معامل ثبات الإعادة لدرجات اختبار ويكونسن قد يشير إلى ضعف الاستقرار النسبي للسمة المقاسة وهي الوظائف التنفيذية للدماغ، حيث تأكد بعض الدراسات على العينات السريرية التغيرات الفعلية التي قد تطرأ على العينة والتي قد تأثر على استقرار إعادة الاختبار. فبرغم من أن دراسة إنجرام وآخرون (Ingram, *et al.* 1999) والتي أبلغت عن استقرار مرتفع لثبات الإعادة أعلى من 0.7 في 5 مؤشرات من أصل 11 مؤشر. ومع ذلك، فإن مرضى انقطاع التنفس الانسدادي يظهرون تغيرا في قدراتهم المعرفية مع ازدياد أو تراجع نوبات انقطاع التنفس (Greve, 2002)، وبالتالي لا يمكن اعتبار عينة ذات عجز معرفي مستقر. كما تشير دراسة تاتي وآخرون (Tate, *et al.* 1998) والتي هدفت إلى فحص الاستقرار الزمني للاختبار في 23 شخصا مصابا بإصابات الدماغ، كانت معاملات الاستقرار لهذه المجموعة السريرية في نطاق 0.3 إلى 0.4؛ مما لا شك فيه أن هذا يرجع إلى حقيقة أن هؤلاء المرضى ما زالوا يخضعون للشفاء وبالتالي من المتوقع أن يتحسنوا مع مرور الوقت. حيث 56٪ من مرضى إصابات الدماغ الرضحية قد تحسنوا بشكل ملحوظ في التطبيق الثاني لاختبار ويسكونسن، ودراسة جريف وزملائه (Greve, 2002) تأكد هذه النتيجة. كما أن الفترة الزمنية الفاصلة بين مرتي التطبيق قد تفسر أيضا ذلك الانخفاض إذا ما علمنا ان متوسط الفترة الزمنية بين مرتي التطبيق في الدراسات التي خضعت للتحليل بلغ (354.5) يوم تقريبا، وهي فترة زمنية طويلة، هذا بالإضافة إلى الظروف المتعلقة بالتطبيق في المرتين والعوامل التي ربما قد تكون أدت دوراً خلال الفترة الفاصلة مما يؤثر في القيمة التقديرية لمعامل ثبات الإعادة. وهكذا، بالإضافة إلى آثار عدم ثبات الاختبار وممارسته، فإن معاملات الثبات التي أبلغ عنها (Tate, *et al.* 1998; Ingram, *et al.* 1999; Greve, 2002) تعكس أيضا آثار التغييرات في القدرة الفعلية للعينة.

ولقد اقترح (paolo, *et al.* 1996; Greve, *et al.* 2002; & Duff, 2012) ثلاثة طرق إحصائية تُستخدم لمساعدة الفاحص في تقييم الفروق في النتيجة عند إعادة الاختبار ويسكونسن لتصنيف البطاقات، كالخطأ المعياري للقياس، والخطأ المعياري لتنبؤ مع الأخذ في الاعتبار عدم ثبات السمة في الإختبار. كما يشترط التوزيع الطبيعي وبالتالي لا يمكن استخدامها إلا في الدرجات المعيارية ويجب أن يكون فاصل الثقة متماثل. وأيضا طريقة ماترازو وهيرمان (1984) ويسمى "درجة التناقض بين الاختبار وإعادة الاختبار" ( Abnormal test-retest discrepancy score)، ويستخدم لتقييم التغيير الكبير ما يشار إليه على أنه "غير طبيعي". تنتج هذه الطريقة فترات ثقة عن طريق فحص التوزيع الفعلي لدرجات الفروق (التطبيق الثاني ناقص التطبيق الأول). حدود فاصل الثقة هي تلك الدرجات التي يتجاوز بعدها عدد معين من الدرجات فقط. في هذه الحالة، نستخدم فاصل ثقة بنسبة 90٪، مما يعني أن ما يقرب من 5٪ من العينة الحالية يقع خارج الحدود العليا والسفلى. هذه الطريقة في تقييم التغيير لها ميزتان على

طريقة الخطأ المعياري للقياس، والخطأ المعياري لتنبؤ، أولاً، لا يفترض التوزيع الطبيعي وبالتالي يمكن استخدامه مع الدرجات ذات التوزيعات المنحرفة. يتيح ذلك استخدامه مع جميع نتائج اختبار ويسكونسن الأولية، بالإضافة إلى درجات المعيارية التي تم إنشاؤها بواسطة اختبار ويسكونسن. ثانيا، يسمح بفواصل الثقة غير المتكافئة.

وبرغم من ذلك يظهر من خلال الدراسة الحالية أن 41.67% من الباحثين قاموا باستخدام الاستقرار الزمني كوسيلة لتحقق من ثبات اختبار ويسكونسن لتصنيف البطاقات. وقد يعود ذلك إلى شيوع التقييمات المتكررة نسبيا في المجال الإكلينيكي. إذ يمكن إعادة التقييم عدة مرات لمتابعة التقدم الطبيعي لحالة ما، مثل إعادة تقييم مرضى الخرف لمتابعة تطرو الحالة المرضية، ولتتبع التعافي بعد الإصابات الدماغية، وتقييم فعالية التدخل (على سبيل المثال، متابعة الحالة بعد استئصال الورم، أو إعادة التأهيل المعرفي) (Duff, 2012)، مما يجعل طريقة ثبات الإعادة أكثر صلة من الطرق الأخرى (على سبيل المثال: الاتساق الداخلي، والأشكال البينية، والأشكال المتوازية).

كما اتضح من خلال الدراسة الحالية بأن 67% من الدراسات التي تم الحصول عليها إما أنها استندت على معاملات ثبات من الدراسات السابقة أو أنها تغافلت عن ذكر الثبات في الدراسة وهو مشابه للنتيجة التي توصلت إليها دراسة (Nunez-Nunez, *et al.* 2022)، وهذا من شأنه التأثير على مدى موثوقية نتائج الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات خلال العقود الثلاثة الماضية.

فتقدير ثبات الدرجات مهم بشكل أساسي ليس فقط للدراسات السيكومترية، بل لجميع الدراسات الإمبريقية (Empirical Studies). فبغض النظر عن مدى رصانة التصميم البحثي للدراسة، وبغض النظر عن حجم الأثر الذي قد يتم الحصول عليها من النتائج، فإن الدراسة التي تعتمد على درجات لم يتم التحقق من ثباتها فمن غير المرجح أن تسفر عن نتائج ذات معنى (Graham & Christiansen, 2009). وقد يخطئ بعض الباحثين في أن الثبات هو خاصية للاختبار، أي أن الاختبار نفسه ثابت أو غير ثابت، والصواب أن خصائص الاختبار نفسه من حيث طوله ومحتواه قد تؤثر على مقدار معامل الثبات (Thompson, 1994). وبالإضافة إلى ذلك، فإن خصائص العينة التي يجرى عليها الاختبار وظروف تطبيق الاختبار قد تؤثر ايضا على معامل ثبات الدرجات (Vacha-Haase, 1998). وبالتالي، فإن الممارسة الشائعة المتمثلة في الاعتماد على تقديرات الثبات المحسوبة من بيانات سابقة تمثل مشكلة بحاجة إلى الوقوف عليها وتصحيحها، وقد وصفت فاشا-هاس وزملائها هذه الممارسة باستقراء الثبات (reliability induction)، حيث يتم تقدير الثبات باستخدام بيانات عينة ما وتفترض أنها تنطبق على جميع البيانات التي يتم جمعها باستخدام نفس الأداة من عينات أخرى ( & Graham Christiansen, 2009; Vacha-Haase, *et al.* 2000). وعليه، نظرا لأن ثبات درجات الاختبار ليست خاصية متأصلة في الاختبار ذاته، فمن المهم تقدير ثبات درجات الاختبار من بيانات عينة الدراسة موضوع التحليل، وهذا يتماشى مع توصيات المجلات العلمية والمبادئ

التوجيهية للنشر العلمي الصادرة من مختلف المنظمات الدولية، مثل جمعية علم النفس الأمريكية (APA)، والجمعية الأمريكية للبحوث التربوية (AERA) (Lopez-Nicolas1, 2021).

**نتائج السؤال الثاني ومناقشتها: ما نسبة إسهام خصائص الدراسات التي استخدمت اختبار ويسكونسن لتصنيف البطاقات (طبيعة العينة، الفئة العمرية للعينة، إصدار الاختبار، الفترة الزمنية بين مرتي التطبيق) في تباين معاملات ثبات درجات الاختبار؟**

للإجابة عن هذا السؤال، تم دمج فئات بعض المتغيرات المستقلة كطبيعة العينة، والفئة العمرية، وإصدار الاختبار نظرا لصغر حجوم العينات ولغرض المقارنة مع الفئة الأكثر شيوعاً للمتغير المستقل. فبالنسبة لمتغير طبيعة العينة، تم تقسيم معاملات الثبات إلى معاملات ثبات درجات الأشخاص الأصحاء العاديين مقابل معاملات ثبات فئات الأشخاص الأخرى (غير الأصحاء أو من لم يتم تحديدهم أصحاء أو غير أصحاء). وبالنسبة لمتغير الفئة العمرية، فقد تم تقسيم معاملات الثبات إلى معاملات ثبات درجات الفئة العمرية (23 سنة فأكثر) مقابل معاملات ثبات درجات الفئات العمرية الأخرى. وبالنسبة لمتغير إصدار الاختبار، فقد تم تقسيم معاملات الثبات إلى معاملات ثبات درجات الاختبار إصدار 1993م مقابل معاملات ثبات درجات الإصدارات الأخرى. كما تم استبعاد متغير جنس العينة لكون أغلبية الدراسات (88.9%) لم تذكر معامل الثبات وفق الجنس، ومتغير بيئة التطبيق لكون معظم الدراسات (86.1%) التي تمت في بيئات غير عربية. ويلخص الجدول (5) توزيع معاملات ثبات الإعادة لدرجات اختبار ويسكونسن لتصنيف البطاقات لمؤشرات الاستجابات الخاطئة، والاستجابات المثابرة، وأخطاء المثابرة، والأخطاء التي ليس لها مثابرة، والفئات المكتملة، والفشل في الحفاظ على مجموعة لكل متغير من المتغيرات المستقلة التصنيفية وهي طبيعة العينة، والفئة العمرية، وإصدار الاختبار.

**جدول (5):** توزيع معاملات ثبات الإعادة لدرجات مؤشرات اختبار ويسكونسن لتصنيف البطاقات لكل متغير من المتغيرات المستقلة التصنيفية.

| مؤشرات الاختبار | طبيعة العينة | | الفئة العمرية | | إصدار الاختبار | |
|---|---|---|---|---|---|---|
| | عاديين | أخرى | 23 سنة فأكثر | أخرى | إصدار 1993 | أخرى |
| الاستجابات الخاطئة | 7 (43.8%) | 9 (56.2%) | 10 (62.5%) | 6 (37.5%) | 10 (62.5%) | 6 (37.5%) |
| الاستجابات المثابرة | 6 (37.5%) | 10 (62.5%) | 11 (68.8%) | 5 (31.2%) | 11 (68.8%) | 5 (31.2%) |
| أخطاء المثابرة | 10 (47.6%) | 11 (52.4%) | 13 (61.9%) | 8 (38.1%) | 12 (57.1%) | 9 (42.9%) |
| الأخطاء التي ليس لها مثابرة | 7 (43.8%) | 9 (56.2%) | 11 (68.8%) | 5 (31.2%) | 10 (62.5%) | 6 (37.5%) |
| الفئات المكتملة | 10 (58.8%) | 7 (41.2%) | 9 (52.9%) | 8 (47.1%) | 10 (58.8%) | 7 (41.2%) |
| الفشل في الحفاظ على مجموعة | 6 (50%) | 6 (50%) | 6 (50%) | 6 (50%) | 8 (66.7%) | 4 (33.3%) |

ولمعرفة إسهام كل متغير من المتغيرات المستقلة (طبيعة العينة، الفئة العمرية للعينة، اصدار الاختبار، الفترة الزمنية بين مرتي التطبيق) في تباين معاملات ثبات الإعادة لدرجات مؤشرات الاختبار، تم استخراج معاملات ارتباط بوينت بايسيريال بين معاملات الثبات وكل متغير من المتغيرات التصنيفية الثنائية وهي طبيعة العينة والفئة العمرية وإصدار الاختبار، كما تم استخراج معامل ارتباط بيرسون بين معاملات الثبات والفترة الزمنية بين مرتي التطبيق، وحجم العينة، والجدول (6) يلخص نتائج ذلك. ولوصف مستوى الاسهام النسبي للمتغيرات المستقلة في تباين معاملات الثبات، تم حساب معامل التحديد (مربع معامل الارتباط)، واعتماد معيار كوهين (Cohen, 1982) الذي يحدد الارتباطات (0.10-0.29)، و(0.30-0.49)، و(0.50-1) على أنها ارتباطات ضعيفة، ومتوسطة، وكبيرة على التوالي.

**جدول (6):** معاملات الارتباط و(معاملات التحديد) بين معاملات ثبات الإعادة لدرجات مؤشرات اختبار ويسكونسن لتصنيف البطاقات والمتغيرات المستقلة.

| مؤشرات الاختبار | طبيعة العينة | الفئة العمرية | إصدار الاختبار | حجم العينة | الفترة الزمنية لمرتي التطبيق |
|---|---|---|---|---|---|
| الاستجابات الخاطئة | 0.43 (18.5%) | -0.28 (7.8%) | -0.12 (1.4%) | -0.23 (5.3%) | 0.22 (4.8%) |
| الاستجابات المثابرة | 0.3 (9%) | -0.15 (2.3%) | -0.21 (4.4%) | -0.23 (5.3%) | -0.06 (0.36%) |
| أخطاء المثابرة | 0.34 (11.6%) | -0.2 (4%) | -0.14 (2%) | -0.24 (5.8%) | -0.05 (0.3%) |
| الأخطاء التي ليس لها مثابرة | -0.11 (1.2%) | 0.03 (0.09%) | 0.03 (0.9%) | 0.08 (0.6%) | -0.22 (4.8%) |
| الفئات المكتملة | 0.43 (18.5%) | -0.31 (9.6%) | -0.22 (4.8%) | 0.04 (0.2%) | 0.14 (2%) |
| الفشل في الحفاظ على مجموعة | 0.50 (25%) | -0.16 (2.6%) | 0.06 (0.4%) | -0.23 (5.3%) | -0.37 (13.7%) |

بالرغم من عدم وجود دلالة إحصائية عند مستوى (0.05) لمعاملات الارتباط بين معاملات ثبات الإعادة لدرجات مؤشرات اختبار ويسكونسن لتصنيف البطاقات والمتغيرات المستقلة، إلا أنه يلاحظ من الجدول (6) أن متغير طبيعة العينة أسهم بدرجة كبيرة (معامل الارتباط > 0.49) في تباين معاملات ثبات الإعادة لدرجات مؤشر الفشل في الحفاظ على مجموعة، وبدرجة متوسطة (0.30 ≤ معامل الارتباط ≤ 0.49) في تباين معاملات ثبات الإعادة لدرجات مؤشرات الاستجابات الخاطئة، واستجابات المثابرة، وأخطاء المثابرة، والفئات المكتملة، وبدرجة ضعيفة (معامل الارتباط < 0.30) في تباين معاملات ثبات الإعادة لدرجات مؤشر الأخطاء التي ليس بها مثابرة. كما أن لمتغير طبيعة العينة علاقة موجبة مع معاملات ثبات الإعادة لدرجات مؤشرات الاستجابات الخاطئة، والاستجابات المثابرة، وأخطاء المثابرة، والفئات المكتملة، والفشل في الحفاظ على مجموعة، وعلاقة سالبة مع معامل ثبات الإعادة لدرجات مؤشر الأخطاء التي ليس لها مثابرة. وهذا يعني أن معاملات ثبات الإعادة لدرجات فئة الأشخاص الأصحاء العاديين أعلى في المتوسط من معاملات ثبات الإعادة لدرجات فئات الأشخاص الأخرى على مؤشرات الاستجابات الخاطئة، والاستجابات المثابرة، وأخطاء المثابرة، والفئات المكتملة، والفشل في الحفاظ على مجموعة، والعكس صحيح بالنسبة على مؤشر الأخطاء التي ليس لها مثابرة.

وبالنسبة لمتغير الفئة العمرية، فقد أسهم بدرجة متوسطة (0.30 ≤ معامل الارتباط ≤ 0.49) في تباين معاملات ثبات الإعادة لدرجات مؤشرات الفئات المكتملة، وبدرجة ضعيفة (معامل الارتباط < 0.30) في تباين معاملات ثبات الإعادة لدرجات المؤشرات الأخرى. وقد جاءت معاملات الارتباط سالبة لمؤشرات الاستجابات الخاطئة، واستجابات المثابرة، وأخطاء المثابرة، والفئات المكتملة، والفشل في الحفاظ على المجموعة، وسالبة في مؤشر الأخطاء التي ليس لها مثابرة. وهذا يعني أن معاملات ثبات الإعادة لدرجات الفئة العمرية (23 سنة فأكثر) أقل في المتوسط من معاملات الثبات درجات الفئات العمرية الأخرى على مؤشرات الاستجابات الخاطئة، واستجابات المثابرة، وأخطاء المثابرة، والفئات المكتملة والعكس صحيح على مؤشر الأخطاء التي ليس لها مثابرة. وجاء إسهام متغير الفئة العمرية بنسبة ضعيفة جداً (أقل من 1%) في تباين معاملات ثبات الإعادة لدرجات مؤشر الأخطاء التي ليس لها مثابرة.

وبالنسبة لمتغيرات إصدار الاختبار، وحجم العينة، والفترة الزمنية لمرتي التطبيق، فقد أسهمت المتغيرات بدرجة ضعيفة في تباين معاملات ثبات الإعادة لدرجات جميع مؤشرات الاختبار التي خضعت للتحليل البعدي في الدراسة الحالية باستثناء وجود إسهام بدرجة متوسطة (0.30 ≤ معامل الارتباط ≤ 0.49) لمتغير الفترة الزمنية لمرتي التطبيق في تباين معاملات ثبات الإعادة لدرجات مؤشر الفشل في الحفاظ على مجموعة.

أن قلة عدد الدراسات ومعاملات الثبات التي خضعت للتحليل البعدي في الدراسة الحالية أدى إلى ضعف القوة الإحصائية في إيجاد دلالة إحصائية عند مستوى (0.05) لمعاملات الارتباط بين معاملات الثبات والمتغيرات المستقلة في الدراسة الحالية، إلا أن الاهتمام في دراسات التحليل البعدي يركز على حجم الإسهام النسبي لخصائص الدراسات في تباين معاملات الثبات (Vacha–Haase, 1998)، وقد كشفت نتائج الدراسة الحالية عن وجود إسهام بدرجة تراوحت بين الضعيفة والمتوسطة لبعض خصائص الدراسات في تباين معاملات ثبات الإعادة لدرجات مؤشرات اختبار ويسكونسن لتصنيف البطاقات.

ومن أبرز خصائص الدراسات التي ساهمت في تباين معاملات الثبات متغير طبيعة العينة من حيث كونها أصحاء عاديين أو فئات أخرى، حيث أظهرت النتائج أن معاملات ثبات الإعادة لدرجات فئة الأشخاص الأصحاء العاديين أعلى في المتوسط من معاملات ثبات الإعادة لدرجات فئات الأشخاص الأخرى على مؤشرات الاستجابات الخاطئة، والاستجابات المثابرة، وأخطاء المثابرة، والفئات المكتملة، والفشل في الحفاظ على مجموعة وهذا بسبب عدم استقرار السمة عند الفئات الخاصة، حيث بينة نتائج دراسة كالاميا وزملائها (Calamia, *et al.* 2012) بأن مرضى الفصام حققوا درجات نتائج مختلفة عند إعادة تطبيق اختبار ويسكونسن لتصنيف البطاقات في مؤشر الاستجابات الخاطئة ومؤشر أخطاء المثابرة مقارنة بالأصحاء، كما أظهرت دراسة برانكو وزملائه (Branco, *et al.* 2021) بأن هناك ارتباطات دالة إحصائيا بين عدد الأخطاء المثابرة ونسبة الأخطاء في تمييز خلل الفص الجبهي، مما يؤكد حساسية الاختبار ويسكونسن لتصنيف البطاقات على تمييز مرضى خلل الفص الجبهي من الأصحاء. وقد بينت نتائج الدراسة الحالية بأن متغير طبيعة العينة فسر (25%) من التباين في معاملات ثبات الإعادة

بين الدراسات التي تناولت مؤشر الفشل في الحفاظ على مجموعة، و (18.5%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الاستجابات الخاطئة، و(11.6%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر أخطاء المثابرة، و(1.2%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الأخطاء التي ليس لها مثابرة، و(18.5%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الفئات المكتملة، و(9%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الاستجابات المثابرة. وعليه، فأن هذه النتائج تقود إلى التساؤل فيما إذا كانت الخصائص السيكومترية لأداة القياس تكون عند أفضل مستوياتها في بيانات العينات التي تم إعداد أداة القياس لأجلها كما هو الحال في اختبار ويسكونسن لتصنيف البطاقات بهدف قياس الاضطراب الوظيفي عند مرضى خلل الفص الجبهي.

وبالنسبة لمتغير الفئة العمرية، فقد تبين أن (7.8%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الاستجابات الخاطئة يفسره الاختلاف في الفئات العمرية، و أن (2.6%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر أخطاء المثابرة يفسره الاختلاف في الفشل في الحفاظ على مجموعة، وأن (4%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر أخطاء المثابرة يفسره الاختلاف في الفئات العمرية، كما فسرت الفئة العمرية (0.09%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر لأخطاء التي ليس لها مثابرة، و(9.6%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الفئات المكتملة، و(2.3%) من التباين في معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشر الاستجابات المثابرة.

وتتفق نتائج بعض الدراسات مع النتيجة الحالية (Calamia, *et al.* 2012; Calamia, *et al.* 2013; Al-Ghatani, 2011) والتي أشارت إلى أن زيادة العمر يؤدي إلى انخفاض الدرجات المقدرة عند إعادة الاختبار في مؤشر الاستجابات الخاطئة ومؤشر أخطاء المثابرة. كما تؤكد نتائج دراسة (Arango-Lasprilla, *et al.* 2015) التي بينت أن درجات اختبار ويسكونسن ساءت مع زيادة العمر في مؤشرات عدد الأخطاء المثابرة وعدد الأخطاء الإجمالية. وتبين دراسة (Scharfen, *et al.* 2018) بأن المشاركين الأكبر سنا أظهروا تأثيرات أصغر قليلا في درجات إعادة الاختبار. ويمكن تفسير ذلك من خلال قيام المشاركين الأصغر سنا بوضع استراتيجيات خاصة بالاختبار بسرعة أكبر من المشاركين الأكبر سنا. حيث تبين أن المشاركين الأصغر سنا يتمتعون بقدرة أعلى على الحفاظ على المعلومات وتحديثها (Braver & Barch, 2002)، ويمكنهم توقع التعلم من تجربة الاختبار السابقة أكثر من المشاركين الأكبر سنا (Van Iddekinge, *et al.* 2011). وتبين حديثاً دراسة (Marquine, *et al.* 2021) بأنه لم يتم العثور على اختلافات ملحوظة بين المجموعات بعد عمر 30 عاما.

وبالنسبة لمتغير إصدار الاختبار، فقد أظهرت النتائج بصفة عامة إسهام بدرجة ضعيفة في تباين معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشرات الاستجابات الخاطئة وأخطاء المثابرة والأخطاء التي ليس لها مثابرة والفئات المكتملة والاستجابات المثابرة، والفشل في

الحفاظ على مجموعة، مما يعني عدم تأثير إصدار الاختبار المطبق على مقدار ثبات الاستقرار للسمة المقاسة، وهذا يتفق مع دراسة (Shan, *et al.* 2008) بأنه لا يوجد دليل على وجود فرق كبير بين الإصدارات المختلفة لاختبار ويسكونسن لتصنيف البطاقات على نتائج المؤشرات.

كما اتضح من نتائج الدراسة الحالية أيضاً أنه لا توجد علاقة دالة إحصائيا عند مستوى (0.05) بين معامل ثبات الإعادة وحجم العينة لدرجات مؤشرات اختبار ويسكونسن لتصنيف البطاقات. فقد تراوحت مستويات إسهام متغير حجم العينة في تباين معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشرات الاستجابات الخاطئة وأخطاء المثابرة والأخطاء التي ليس لها مثابرة والفئات المكتملة والاستجابات المثابرة، والفشل في الحفاظ على مجموعة بين (0.2%) و(5.8%).

كما اتضح من نتائج الدراسة الحالية أيضاً أنه لا توجد علاقة دالة إحصائيا عند مستوى (0.05) بين معامل ثبات الإعادة وطول الفترة الزمنية بين مرتي التطبيق لدرجات مؤشرات اختبار ويسكونسن لتصنيف البطاقات. فقد تراوحت مستويات إسهام متغير الفترة الزمنية في تباين معاملات ثبات الإعادة بين الدراسات التي تناولت مؤشرات الاستجابات الخاطئة وأخطاء المثابرة والأخطاء التي ليس لها مثابرة والفئات المكتملة والاستجابات المثابرة بين (0.3%) و(13.7%). وتتعارض هذه النتيجة مع دراسة ( Calamia, *et al.* 2012; Calamia, *et al.* 2013 ) والتي تشير إلى انخفاض الارتباطات بين الدرجات المقدرة للاختبار وإعادة الاختبار عند زيادة المدة الفاصلة بين مرتي التطبيق في مؤشر الاستجابات الخاطئة ومؤشر أخطاء المثابرة، كما أشارت نتائج دراسة شارفن (Scharfen, *et al.* 2018) والتي هدفت إلى تحليل بُعدي لمعاملات ثبات الإعادة للاختبارات المعرفية، بأن زيادة طول الفترة الزمنية بين الاختبار وإعادة الاختبار بمقدار 50 أسبوعا من شأنه أن يؤدي الى انخفاض معامل الاستقرار بمقدار (0.04) انحراف معياري، ولعل السبب نتيجة الدراسة الحالية هو قلة عدد الدراسات ومعاملات الثبات التي خضعت للتحليل البعدي لكل مؤشر. يتأثر ثبات الإعادة بالفاصل الزمني بين التقييمات الأولية والتقييمات المتكررة. تؤدي فترات إعادة الاختبار الأقصر إلى معاملات موثوقية أعلى، وتؤدي فترات إعادة الاختبار الأطول إلى قيم ثبات أقل. ليس من المستغرب أن تشير معظم كتيبات الاختبار إلى ارتباطات الاختبار - إعادة الاختبار عبر فترات إعادة الاختبار لا تتعدى أيام إلى أسابيع؛ ولكنها فترات أقصر بكثير من معظم الممارسة الفعلية لإعادة الاختبار السريرية من أشهر إلى سنوات (Duff, 2012).

يختلف اتجاه التغير في فرق الدرجات على اختبار ويسكونسن لتصنيف البطاقات عند تكراره على المفحوص، يمكن أن يشير هذا إما إلى التحسن أو الانخفاض. بالنسبة للاستجابات الخاطئة، واستجابات المثابرة، وأخطاء المثابرة، وسيرتبط التحسن بدرجات أقل عند إعادة الاختبار، بينما بالنسبة مؤشر الفئات المكتملة، سيرتبط التحسن بدرجات أعلى عند إعادة الاختبار. بالنسبة لمؤشر الفشل في الحفاظ على مجموعة والاستجابات التي ليس لها مثابرة، يعتمد اتجاه التغيير على المستوى العام للمفحوص من ضعف في التطبيق الأول. وتظهر عدة دراسات أن مؤشر الاستجابات الخاطئة، وعدد الفئات المكتملة، وأخطاء المثابرة أكثر حساسية

اتجاه العجز في القدرة التنفيذية من مؤشرين أخطاء التي ليس لها مثابرة، والفشل في الحفاظ على مجموعة مرضى الفصام، والأفراد المصابين بالعديد من الحالات العصبية الأخرى، بما في ذلك مرض باركنسون، وإدمان الكحول المزمن، والخرف الوعائي تحت القشري، والضعف الإدراكي المعتدل. واظهر العجز التنفيذي في الأفراد الذين يعانون من اضطراب ثنائي القطب والاكتئاب الشديد والأفراد المعرضين وراثيا لخطر الإصابة بالفصام.

**التوصيات**

بناءً على ما توصلت اليه الدراسة الحالية من نتائج فإنها توصي بما يلي:

1. استخراج معامل الثبات من العينة التي طبقت عليها الدراسة، مع ذكر خصائص العينة والمتوسط الحسابي والانحراف المعياري لدرجاتها، حيث أن هذه القيم هي قيم نسبية وليست مطلقة، ومرتبطة بالطريقة المستخدمة في تقدير الثبات، وخصائص العينة.
2. يصعب الوثوق بمعاملات الثبات الاتساق الداخلي لبيانات اختبار ويسكونسن لتصنيف البطاقات في الدراسات السابقة، حيث أن الاختبار يتكون من استجابات غير مستقلة ولا يمكن الوثوق بمعامل الثبات العام ويمكن حل هذه المشكلة من خلال دمج بعض مؤشرات الاختبار بطريقة منطقية كمؤشر الاستجابات الصحيحة والخاطئة، أو مؤشر الاستجابات التي ليس بها مثابرة ومؤشر استجابات المثابرة واستخراج ثبات الاتساق الداخلي من خلالها.
3. استخراج الثبات من عينة استطلاعية قبل تقدير الثبات من عينة الدراسة لكي تتم معرفة مدى صلاحية أداة الدراسة لبيئة التطبيق ودقتها في قياس السمة المراد قياسها.
4. ذكر الفترة الزمنية عند استخدام إعادة التطبيق في معامل ثبات الاستقرار للدرجات، حيث توفر معلومات للباحثين يستطيعون من خلالها اتخاذ قرارات بشأن دقية معامل الثبات.
5. الاهتمام بتقدير الثبات لكل مجموعة عمرية ولكل من الذكور والإناث بصفة مستقلة، نظراً للفروق بين المجموعات.

**المقترحات**

وبناءً على ما توصلت اليه الدراسة الحالية من نتائج فإنها تقترح ما يلي:

1. إجراء المزيد من دراسات التحليل البعدي تهدف الى معرفة العوامل التي قد تؤثر على الأنواع الأخرى لمعاملات الثبات للدرجات في اختبار ويسكونسن لتصنيف البطاقات.
2. إجراء دراسات تحليل بعدي تهدف الى معرفة العوامل المؤثرة على صدق تفسير نتائج اختبار ويسكونسن لتصنيف البطاقات.
3. يمكن التطرق إلى دراسة طول فترة المرض، ومدة فترة العلاج، المرحلة التعليمية، والمستوى الاقتصادي والتحقق من تأثيرها على معاملات ثبات اختبار ويسكونسن.

4. إنشاء قاعدة معلومات عربية شاملة للمقاييس والاختبارات التربوية والنفسية، والدراسات المرتبطة بها حتى تساعد الباحثين الذين يرغبون في استخدام التحليل البعدي في الحصول على البيانات اللازمة.
5. إجراء المزيد من دراسات تقنين اختبار ويسكونسن لتصنيف البطاقات على الأفراد في البيئة العربية.

**المراجع العربية**


- عبد القوي، سامي. (2017). *علم النفس العصبي الأسس وطرق التقييم*. (ط 2). مكتبة الأنجلو المصرية.
- العبيدي، عبدالله أحمد خلف. (2019). حجم *الأثر وتطبيقاته في منهجية البحوث التربوية والنفسية والطبية*. دار وائل لنشر والتوزيع.
- علام، صلاح الدين محمود. (2000). *القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجيهاته المعاصرة*. دار الفكر العربي.
- علام، صلاح الدين محمود. (2006). *الاختبارات والمقاييس التربوية والنفسية*. دار الفكر للنشر والتوزيع.
- مليكة، لويس كامل. (2010). *التقييم النيوروسيكولوجي*. دار الفكر ناشرون وموزعون.


**الموافقة الأخلاقية والموافقة على المشاركة:** تم حسب الأصول المتعارف عليها.

**توافر البيانات والمواد:** تم جمعها بطرق ممنهجة وصحيحة.

**مساهمة المؤلفين:** ساهم جميع المؤلفين في اثراء البحث العلمي.

**تضارب المصالح:** لا يوجد تضارب للمصالح.

**التمويل:** لا يوجد أي تمويل أو دعم مالي من أي جهة.

**شكر وتقدير:** الشكر موصول لجامعة السلطان قابوس.

## References (Arabic & English)


- Abdul Qawi, S. (2017). *Neuropsychology Principles and methods of evaluation* (2 ed.). The Anglo Egyptian Bookshop.
- AERA, APA, & NCME. (2014). Standards for Educational and Psychological Testing: National Council on Measurement in Education. *American Educational Research Association*.
- Al Obeidi, A. A. K. (2019). *The size of the impact and its applications in the methodology of educational, psychological and medical research*. Wael Publishing and Distribution center.
- Al-Ghatani, A. M. Obonsawin, M. C. Binshaig, B. A. & Al-Moutaery K. R. (2011). Saudi normative data for the Wisconsin Card Sorting test, Stroop test, Test of Non-verbal Intelligence-3, Picture Completion and Vocabulary (subtest of the Wechsler Adult Intelligence Scale-Revised). *Neurosciences*, 16 (1). 29-41.
- Al-Hinai, S. Alkharusi, H. & Ibrahim, M. (2019). A meta-analysis of reliability of scores on Raven's Progressive Matrices Tests. *Perdana: International Journal of Academic Research*, 6(1). 51-68.
- Allam, S. M. (2006). *Educational and psychological tests and measures*. Dar Al-Fiker for Publishing & Distribution.
- Allam, S. M. (2000). *Educational and psychological measurement and evaluation: its basics, applications and contemporary guidance*. Dar Al-Fiker for Publishing & Distribution.
- Amir, N. (2018). Validity and Reliability Test of Wisconsin Card Sorting Test (WCST) in Indonesian version on Schizophrenia Group and Healthy Group. Faculty of Medicine.
- Arango-Lasprilla, J. C. Rivera, D. Longoni, M. Saracho, C. P. Garza, M. T. Aliaga, A. & Perrin, P. B. (2015). Modified Wisconsin card sorting test (M-WCST): normative data for the Latin American

Spanish speaking adult population. *NeuroRehabilitation,* 37(4). 563-590.

- Bangert-Drowns, R. L. & Rudner, L. M. (1990). Meta-analysis in educational research. *Practical Assessment, Research, and Evaluation,* 2(1). 8.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39*(1). 15-22.
- Borenstein, M. Hedges, L. V. Hedges, L. V. Higgins, J. P. T. Rothstein, H. Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Branco, L. D. Cotrena, C. Shansis, F. M. & Fonseca, R. P. (2021). Cognitive abilities underlying performance on the modified card sorting test: novel and traditional scores. *Applied Neuropsychology: Adult,* 28(5). 544-555.
- Braver, T. S. & Barch, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience & Biobehavioral Reviews,* 26(7). 809-817.
- Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, 26(4). 543-570.
- Calamia, M. Markon, K. & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test–retest correlations. *The Clinical Neuropsychologist.* 27(7). 1077-1105.
- Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement*, 60(2). 236-254.

- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *The Journal of General Psychology*. 130(3). 290–304.
- Chertok I. R. A. & Haile Z. T. (2018). Meta-Analysis. *SAGE Journals*, 34(3). 420-423.
- Chiu E. Wu, W. Hung, J. & Tseg, Y. (2017). Validity of the Wisconsin Card Sorting Test in patients with stroke. *National Library of Medicine*. 40(16). 1967-1971.
- Cohen, L. & Holliday, M. (1982). Statistics for social scientists: an introductory text with computer programs in basic. *Harper and Row*.
- Cooper, H. Hedges, L. V. & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
- Diamond A. (2013) Executive functions. *Annu. Rev. Psychol*, 64(135)–168.
- Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Archives of Clinical Neuropsychology*, 27(3). 248-261.
- Fan, X. & Yin, P. (2003). Examinee characteristics and score reliability: An empirical investigation. *Educational and Psychological Measurement*, 63(3). 357-368.
- Graham, J. M. & Christiansen, K. (2009). The reliability of romantic love: A reliability generalization meta-analysis. *Personal Relationships*, 16(1). 49-66.
- Grant D. A. & Berg E. (1948). Reliability Generalization: exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*. 58(1).
- Greve, K.W. Love, J. M. Sherwin, E. Mathias, C. W. Houston, R. J. & Brennan, A. (2002). Temporal Stability of the Wisconsin Card-

Sorting Test in a Chronic Traumatic Brain Injury Sample. *Assessment*, 9(3). 271–277.

- Heaton, R. K. Chelune, G. J. Talley, J. L. Kay, G. G. & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual REVISED AND EXPANDED*. Psychological Assessment Resources, Inc. PAR.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34(3). 177-189.
- Higgins, J. P. T. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–58.
- Holsti, O. R. (1969). Content analysis for the social sciences and humanities. *Reading. MA: Addison-Wesley (content analysis)*.
- Ingram, F. Greve, K.W. Ingram, P.T.F. & Soukup V. M. (1999). Temporal stability of the Wisconsin Card Sorting Test in an untreated patient sample. *Br. J. Clin. Psychol.* 38(2). 209–211.
- Kopp, B. Maldonado, N. Scheffels, J. F. Hendel, M. & Lange, F. (2019). A meta-analysis of relationships between measures of Wisconsin card sorting and intelligence. *Brain Sciences*, 9(12). 349.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.
- Lopez Nicolas, R. Rubio Aparicio, M. Lopez Ibanez, C. & Sanchez Meca, J. (2021). A reliability generalization meta-analysis of the Dimensional Obsessive-Compulsive Scale. *Psicothema*.
- Lung, F. W. Chen, P. F. & Shu, B. C. (2018). Performance of Wisconsin Card Sorting Test in five-year-old children in Taiwan: Relationship to intelligence and cognitive development. *PloS one*, 13(8).

- Malika, L. K. (2010). *Neuropsychological assessment*. Dar Al Fiker Publishers & Distributors.
- Marquine, M. J. Morlett Paredes, A. Madriaga, C. Blumstein, Y. Umlauf, A. Kamalyan, L. & Cherner, M. (2021). Demographically adjusted norms for selected tests of verbal fluency: Results from the Neuropsychological Norms for the US-Mexico Border Region in Spanish (NP-NUMBRS) project. *The Clinical Neuropsychologist,* 35(2). 269-292.
- Martin, N. G. (2016). *Human Neuropsychology*. Published by Pearson
- Mikolajewicz N. & Komarova, S. V. (2019). Meta-Analytic Methodology for Basic Research: A Practical Guide. *Front. Physiol, 10*(203).
- Miranda, A. R. Sierra, J. F. Roulet, A. M. Rivadero, L. Serra, S. V. & Soria, E. A. (2019). Age, education and gender effects on Wisconsin card sorting test: standardization, reliability and validity in healthy Argentinian adults. *Aging, Neuropsychology, and Cognition A Journal on Normal and Dysfunctional Development*, 27(6). 807-825.
- Mouter, N. & Vonk Noordegraaf, D. M. (2012). Intercoder reliability for qualitative research: You win some, but do you lose some as well? In *Proceedings of the 12th TRAIL congress, 30-31 oktober 2012, Rotterdam, Nederland*. TRAIL Research School.
- Nunez-Nunez, R. M. Rubio-Aparicio, M. Marin-Martinez, F. Sanchez-Meca, J. Lopez-Pina, J. A. & Lopez-Lopez, J. A. (2022). A Reliability Generalization Meta-analysis of the Padua Inventory-Revised (PI-R). *International Journal of Clinical and Health Psychology,* 22(1). 100-277.
- Nunnally, J.C. & Bernstein, I. H. (1994). The Assessment of Reliability. *Psychometric Theory,* 3, 248-292.

- Paolo, A. M. Axelrod, B. N. & Troster, A. I. (1996). Test-Retest Stability of the Wisconsin Card Sorting Test. *Assessment,* 3(2), 137–143.
- Scharfen, J. Peters, J. M. & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67. 44-66.
- Shan, I. K. Chen, Y. S. Lee, Y. C. & Su, T. P. (2008). Adult normative data of the Wisconsin card sorting test in Taiwan. *Journal of the Chinese Medical Association,* 71(10). 517-522.
- Steinke, A. Kopp, B. & Lange, F. (2021) The Wisconsin Card Sorting Test: Split-half reliability estimates for a self-administered computerized variant. *Brain Sci.* 11(5). 529. 10.3390/brainsci11050529
- Tate, R. L. Perdices, M. & Maggiotto, S. (1998). Stability of the Wisconsin Card Sorting Test and the Determination of Reliability of Change in Scores. *Clin. Neuropsychol,* 12(3). 348–357.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement,* 54. 837–847.
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*. 58(1). 6-20.
- Vacha-Haase, T. Kogan, L. R. & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement,* 60. 509–522.
- Valentine, J. C. Pigott, T. D. & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics,* 35(2). 215-247.
- Van Iddekinge, C. H. Morgeson, F. P. Schleicher, D. J. & Campion, M. A. (2011). Can I retake it? Exploring subgroup differences and

criterion-related validity in promotion retesting. *Journal of Applied Psychology*, 96(5). 941–955.

- Walker, W. E. Harremoes, P. Rotmans, J. Van Der Sluijs, J. P. Van Asselt, M. B. Janssen, P. & Krayer von Krauss, M. P. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1). 5-17.