الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة بليدة 2

كلية العلوم الإنسانية والاجتماعية

قسم العلوم الاجتماعية

دراسة تقييميه لطرق إعادة التحقق من صلاحية الاختبارات النفسية المطبقة في عينة من الدراسات الأكاديمية

أطروحة مقدمة لنيل شهادة الدكتوراه في: علم النفس

تخصص: القياس في علم النفس وعلوم التربية

إعداد الطالبة:

تشلابي سكينة

إشراف الأستاذ:

بوسالم عبد العزيز

السنة الجامعية: 2022/2021

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة بليدة 2

كلية العلوم الإنسانية والاجتماعية

قسم العلوم الاجتماعية

دراسة تقييميه لطرق إعادة التحقق من صلاحية الاختبارات النفسية المطبقة في عينة من الدراسات الأكاديمية

أطروحة مقدمة لنيل شهادة الدكتوراه في: علم النفس

تخصص: القياس في علم النفس وعلوم التربية

إعداد الطالبة:

تشلابي سكينة

إشراف الأستاذ:

بوسالم عبد العزيز

السنة الجامعية: 2022/2021

شكر وتقدير

بسم الله الرحمن الرحمان الرحيم

"الحمد لله الذي هدانا لهذا وما كنا لنهتدي لولا أن هدانا الله"

سورة الأعراف الآية 43

انطلاقًا من مبدأ أنه لا يشكر الله من لا يشكر الناس، فإننا نتوجه بالشكر الجزيل للأستاذ الدكتور **(بوسالم عبد العزيز)** الذي رافقنا في مسيرتنا لإنجاز هذا العمل وكانت له بصمات واضحة من خلال توجيهاته وانتقاداته البناءة والدعم الأكاديمي.

كذلك أتقدم بالشكر للأساتذة الكرام الذين وافقوا على مناقشة هذه الرسالة، باذلين في ذلك جهودا من أجل قراءتها وإبداء ملاحظاتهم وإضافة لمساتهم العلمية. **(لجنة المناقشة)**

أيضا أتقدم بالشكر لكل **(أساتذتي)** الذين كان لي شرف الدراسة عندهم وحتى إلى من لم أحظى بذلك، وأخص بالذكر الأستاذة **(اليازيدي فاطمة الزهراء)** لإيمانها بي وتشجيعها الدائم لخوض مسار البحث العلمي.

ولا يسعني إلا أن أشكر **(زوجي)** الذي لولاه لما وصلت إلا هنا.

كما لا أنسى أن أتقدم بالشكر إلى من يسر لي هذا العمل ومد يد العون من إداريين وزملاء وغيرهم. وأخص بالذكر **كاتيا وزهراء**.

الإهداء

إلى الزوج الذي كان نعم الشريك ونعم السند والمعين، إلى من خفف وحمل عني الكثير، إلى الذي بفضل الله ثم بفضله وصلت إلى هنا، إلى الروح التي سكنت روحي حفظك الله لي...**زوجي الغالي وسندي**

إلى الأم المعجزة التي ما زالت تحطم كل الأرقام بصبرها وتضحياتها، إلى التي من أجلها كافحت وسأظل أكافح، أهدي لك بعض عطائك عله يكون فخرا لك ...**أمي أغلى البشر**

إلى أمي الثانية والتي كسرت كل صور الحماة، ودعمتني في كل خطوة ولو بدعاء، حفظك الله وحقق كل أمانيك...**أم أعز الناس**

إلى الأب الذي ليس كمثله أب، إلى من أرفع رأسي عاليا افتخارا واعتزازا به، إلى من لن يضاهيه أحد ...**أبي أروع أب**

إلى أبي الثاني الرجل المجاهد، والذي مازال يجاهد إلى الآن لإسعادنا وإسعاد أحفاده، إلى قدوة أطفالي وعمود العائلة...**أب زوجي وأبي الثاني**

إلى عيون أولادي الأجمل، إلى الذين أطمح أن يسيروا على خطاي، إلى الذين أدعوا الله أن يرزقني برهم..**أولادي قطعة من قلبي**

إلى أخواتي الغاليات رفيقات الدرب وحبيبات القلب، إلى ضلعي الثابت، من أنجبتهم أمي **أخواتي** ومن لم تفعل **أخت زوجي وزوجة أخيه**

إلى من هم عزوتي ومخبئي من شدائد السنين وظهري الذي احتمي وراءه **إخوتي وإخوة زوجي وأزواج أخواتي**

إلى رياحين الحياة وأمل العائلة إلى كل **أبناء وبنات أخواتي وإخوتي**

وأخيرا أهديه إلى **صديقاتي العزيزات** وإلى **جدتي** رحمها الله جمعني الله وإياك في جنات النعيم.

ملخص الدراسة

يعتبر الصدق والثبات أهم دليلين على صلاحية الاختبار النفسي للاستخدام المرجو منه، لذلك هدفت هذه الدراسة إلى تقييم الطرق التي يتم بها إعادة التحقق من صدق وثبات الاختبارات النفسية، وقد تم الاعتماد في ذلك على قائمة المعايير الدولية COSMIN، والشبكة الخاصة بها، من أجل تقييم إجراءات إعادة التحقق من صدق وثبات 171 اختبار نفسي تم استخدامه في عينة من الدراسات الأكاديمية العربية والأجنبية، وذلك بإتباع المنهج الوصفي، وقد أظهرت النتائج أن ثبات الاتساق الداخلي هو الإجراء الأكثر استخداما بـ152 تكرار يليه الصدق الهيكلي الذي تم استخدامه في 97 اختبار، أما الأقل استخداما فهو الصدق المحكي الذي استخدم مرتين فقط، في حين أنه ولا دراسة عمدت إلى التحقق من خطأ القياس أو الاستجابة. وفيما يخص جودة هذه الأدلة فقد تراوحت بين منخفضة ومرتفعة، وقد احتل صدق المحتوى المرتبة الأخيرة، أما الإجراء الذي تصدر قائمة الجودة فهو ثبات الاتساق الداخلي بـ97 اختبار ذو جودة مرتفعة. وفي الأخير فإن الباحثين في جميع الدراسات ما عادا دراسة واحدة اختاروا الاعتماد على الاختبار والعمل به رغم عدم صلاحيته في العديد من الدراسات.

Abstract

Validity and reliability are the two most important indicators of the validity of a psychological test for its intended use, so this study aimed to evaluate the ways in which the validity and reliability of psychological tests, It has been relied Standards on the international COSMIN and its

own checklist, For an evaluation of the re-verified the validity and reliability of 171 psychological tests that were used in a sample of Arab and foreign academic studies, by following the descriptive approach. The results showed that the internal consistency is the most used procedure with 152 tests, followed by the structural validity that was used in 97 tests. The least used is the Criterion validity that was used only twice, while no study aimed to verify the measurement or response error. Regarding the quality of evidence, it ranged between low and high, and the validity of the content ranked last. As for the procedure that topped the quality list, it was the consistency of internal consistency with 97 high-quality tests. Finally, researchers in all studies except for one study chose to rely on the test and work with it despite its invalidity in many studies.

فهرس الموضوعات

شكر وتقدير

الإهداء

ملخص الدراسة

فهرس الموضوعات

فهرس الجداول

فهرس الأشكال

فهرس الملاحق

مقدمة ---------- **أ-ب-ج**

الفصل الأول مدخل نظري عام للدراسة

1-الإشكالية ---------- **6**

2-أهداف الدراسة ---------- **12**

3-أهمية الدراسة---------- **13**

4-تحديد المفاهيم إجرائيا ---------- **14**

4-1-دراسة تقييميه ---------- 14

4-2-طرق إعادة التحقق ---------- 14

4-3-صلاحية الاختبارات النفسية ---------- 14

5-الدراسات السابقة ---------- **15**

الفصل الثاني صلاحية الاختبارات النفسية

1-مفهوم صلاحية الاختبارات النفسية ---------- **37**

2-دوافع إعادة التحقق من صلاحية الاختبارات النفسية ---------- **38**

2-1-الغرض من الاختبار ---------- 39

2-2-مجتمع وعينة الاختبار ---------- 41
2-3-عامل الزمن ---------- 43
2-4-عامل الثقافة---------- 44
2-5-الاختبار المعدّل---------- 45
2-6-عامل استمرارية الصلاحية---------- 47
3-أهم مهددات صلاحية الاختبارات النفسية---------- 48
4-مفهوم الاختبارات النفسية ---------- 56
5-شروط الاختبار الجيّد---------- 58
6-أنواع الاختبارات النفسية ---------- 67
7-أهمية الاختبارات النفسية ---------- 81
8- خطوات بناء الاختبارات النفسية ---------- 82
9-الفرق بين الاختبار والاستبيان والمقياس ---------- 87
الفصل الثالث الصدق في الاختبارات النفسية
1-مفهوم الصدق ---------- 91
2-أنواع الصدق:---------- 95
2-1-صدق المحتوى ---------- 98
2-2-الصدق المرتبط بالمحك:---------- 99
2-2-1-الصدق التلازمي---------- 102
2-2-2-الصدق التنبؤي---------- 102
2-3-صدق التكوين الفرضي---------- 104
2-4-أنواع الصدق المذكورة في الدراسات ---------- 105
2-4-1-الصدق الظاهري وصدق المحكمين---------- 105
2-4-2-الصدق الذاتي---------- 106
2-4-3-الصدق التطابقي---------- 107
3-إجراءات أنواع الصدق ---------- 108
4-أدلة الصدق ---------- 117
5-تعقيب على أنواع الصدق وأدلته ---------- 128

6-العوامل المؤثرة على صدق الاختبار ---------- 129

7-الصدق وفق دراسة كوسمين COSMIN ---------- 132

8-معايير الصدق حسب COSMIN ---------- 134

الفصل الرابع الثبات في الاختبارات النفسية

1-مفهوم الثبات ---------- 140

2-أنواع الثبات ---------- 144

2-1-ثبات إعادة الاختبار---------- 144

2-2-ثبات الصور المتكافئة ---------- 149

2-3-ثبات الاتساق الداخلي ---------- 151

3-الخطأ المعياري للقياس ---------- 162

4-العوامل المؤثرة في الثبات ---------- 165

5-الثبات وفق دراسة كوسمين COSMIN ---------- 169

6-معايير الثبات حسب COSMIN ---------- 171

7-العلاقة بين الصدق والثبات ---------- 174

الجانب التطبيقي

الفصل الخامس الإطار المنهجي للدراسة

1-منهج الدراسة ---------- 180

2-عينة الدراسة ---------- 180

2-1-إستراتيجية جمع العينة ---------- 180

2-1-1-معايير القبول والاستبعاد ---------- 181

2-1-2-عملية فرز الدراسات ---------- 182

2-2-خصائص عينة الدراسة---------- 185

3-أداة الدراسة ---------- 192

3-1-نبذة عن مبادرة COSMIN ---------- 192

3-2- قائمة المعايير الأصلية COSMIN ---------- 193

3-3-قائمة المعايير COSMIN المستخدمة في الدراسة الحالية ---------- 195

4-الأساليب الإحصائية المستخدمة في الدراسة ---------- 197

الفصل السادس عرض ومناقشة النتائج

1-عرض نتائج الدراسة ---------- 200

2-مناقشة النتائج ---------- 264

الخاتمة ---------- 282

الاقتراحات ---------- 283

المراجع ---------- 286

الملاحق ---------- 315

فهرس الجداول

الجدول 01: توزيع العينة حسب متغير الدراسة---------- 185

الجدول 02: نوع الدراسات المدمجة في العينة ---------- 186

الجدول 03: توزيع العينة حسب النسخة المستخدمة من الاختبار---------- 189

الجدول 04: توزيع الاختبارات حسب التعديل---------- 190

الجدول 05: طرق الصدق والثبات المستخدمة في إعادة التحقق من صلاحية الاختبارات النفسية---------- 203

الجدول 06: سؤال العينة عن الملائمة---------- 206

الجدول 07: سؤال العينة عن الشمولية---------- 209

الجدول 08: سؤال العينة عن فهم واستيعاب الاختبار ---------- 212

الجدول 09: سؤال المحكمين عن الملائمة---------- 215

الجدول 10: سؤال المحكمين عن الشمولية---------- 217

الجدول 11: الكشف عن أحادية البعد---------- 220

الجدول 12: تقييم الصدق الهيكلي---------- 221

الجدول 13: ثبات الاتساق الداخلي---------- 225

الجدول 14: الجدول المتقاطع بين ثبات الاتساق الداخلي والكشف عن بنية الاختبار ---------- 229

الجدول 15: صدق الاختبارات المنقولة ثقافيا ---------- 231

الجدول 16: ثبات التطبيق وإعادة التطبيق---------- 235

الجدول 17: أخطاء القياس---------- 239

الجدول 18: تقييم الصدق المحكي---------- 241

الجدول 19: تقييم الصدق التقاربي والتباعدي---------- 243

الجدول 20: تقييم الصدق التمييزي بين المجموعات ---------- 245

الجدول 21: تقييم استجابة المحك---------- 248
الجدول 22: تقييم الاستجابة مع الاختبار التقاربي أو التباعدي---------- 249
الجدول 23: تقييم الاستجابة التمييزية---------- 251
الجدول 24: استجابة التدخل---------- 252
الجدول 25: جودة صدق المحتوى---------- 253
الجدول 26: جودة الصدق الهيكلي---------- 255
الجدول 27: جودة ثبات الاتساق الداخلي---------- 256
الجدول 28: جودة الصدق الثقافي---------- 257
الجدول 29: جودة التطبيق وإعادة التطبيق---------- 258
الجدول 30: جودة الصدق المحكي---------- 260
الجدول 31: جودة الصدق التقاربي والتباعدي ---------- 261
الجدول 32: جودة الصدق التمييزي---------- 262

فهرس الأشكال

الشكل 01: أنواع الاختبارات حسب GREGORY (2014)---------- 1

الشكل 02: أنواع الاختبارات حسب GAY (1980) وأكابور، أهلاوات وآخرون (1985)---------- 80

الشكل 03: طريقة فرز الدراسات حسب منهجية PRISMA---------- 183

الشكل 04: توزيع العينة حسب البلد الذي تنتمي إليه الدراسة---------- 187

الشكل 05: توزيع العينة حسب سنة المناقشة أو النشر ---------- 188

فهرس الملاحق

الملحق 01: شبكة COSMIN للدراسة الحالية---------- 316

الملحق 02: الشبكة COSMIN الأصلية---------- 321

الملحق 03: جداول ملخصات المعايير ---------- 329

الملحق 04: منهجية PRISMA الأصلية---------- 333

الجدول 05: جدول مواصفات عينة الدراسة---------- 334

الملحق 06: نموذج شامل لخصائص القياس---------- 399

مقدمة

مقدمة

إن استخدام أي اختبار نفسي، يفرض على صاحبه مجموعة من الشروط أو المعايير الواجب التقيد بها، فمن أجل ضمان جودة الدراسة يجب أن يتم جمع البيانات باستخدام اختبارات تتمتع بصدق وثبات كافيين على الأقل، وذلك انطلاقا من أن جودة النتائج من جودة الاختبارات المستخدمة فيها، وجودة الاختبارات من جودة الخصائص السيكومترية لها.

ويتم التحقق من جودة الاختبارات النفسية من خلال جمع أدلة ملائمة وكافية حول صدقها وثباتها، وفي بعض الأحيان يجب إضافة أدلة على عدم تحيزها أيضا، وكل ذلك باستخدام أساليب متعددة وردت في الكثير من أدبيات القياس. فالتحقق من الصدق والثبات سواء في مرحلة البناء أو عند استخدام اختبارا جاهزا، مرحلة حرجة يتخذ فيها الباحث قرارات مهمة حول صلاحية الأداة، خاصة عند استخدامها في اتخاذ وإصدار أحكام متعلقة بمصير الأفراد، سواء في البحوث أو الممارسات السيكولوجية لذلك يجب تحديد مستوى الثقة في نتائجها.

وبما أن الصدق والثبات من الخصائص الجوهرية في القياس، ومن أهم معايير تقييم جودة الاختبارات النفسية والتربوية، لذلك وجب على الباحث قبل الجزم بصلاحية الاختبار أن ينتبه إلى أن الصدق خاصية نسبية شأنه شأن الثبات، فالصدق والثبات ليستا مطلقتان بل تختلفان من اختبار لآخر.

ولهذا يضطر الباحث أو مستخدم الاختبار بصفة عامة إلى إعادة التحقق من صدق وثبات الاختبار النفسي الذي يريد الاعتماد عليه خاصة إذا كان هذا الأخير يحتاج إلى بعض التعديلات من حذف للبنود أو إضافة، أو تم تغيير الغرض الذي بني من أجله الاختبار، أو العينة وغيرها من الدوافع، التي تجعل من الباحث مجبرا على إعادة التحقق من صلاحيته لذلك الاستخدام.

ولكن ما كشفت عنه الدراسات والأدبيات من تجاوزات وانتهاكات في الواقع البحثي، لا يعكس ذلك فنجد العديد من الأخطاء عند استخدام الاختبارات النفسية، سواء من حيث تجاهل إعادة التحقق من صلاحية الاختبار أصلا، أو القرارات التي يتخذها الباحثون بعد إعادة التحقق والتي غالبا ما تكون الاحتفاظ بجميع البنود والأبعاد رغم النتائج التي تدل على عدم صلاحيته، أو الانتهاكات التي تحدث عند تقدير الصدق والثبات. والعديد من التجاوزات الأخرى التي أصبحت شائعة في الدراسات الأكاديمية، الأمر الذي يستلزم التأمل قبل الوثوق بالنتائج الجديدة؛ وتحديد الدافع أو الداعي إلى إعادة التحقق من صلاحية الاختبارات النفسية، ومحاولة ضبط المعايير التي نستند عليها في اختيار أدلة الصدق والثبات وتفسير النتائج. وهذا ما هدفت إليه هذه الدراسة من خلال الفصول التالية:

الفصل الأول والذي يهدف إلى عرض إشكالية الدراسة وتوضيح تساؤلاتها، ومن ثم الكشف عن أهدافها، وإظهار نقاط قوة هذه الدراسة من خلال عرض أهميتها، لينتهي هذا الفصل بتحديد المفاهيم إجرائيا حتى لا يقع أي لبس عند للقارئ.

أما الفصل الثاني والذي تم التركيز فيه على صلاحية الاختبارات النفسية، انطلاقا من البحث في الأسباب والدوافع التي تجعل إعادة التحقق من صلاحية الاختبارات النفسية ضرورة لا بد منها، مرورا بالكشف عن أهم مهددات صلاحية الاختبارات النفسية، لتنتقل بعدها الباحثة إلى التفصيل في أنواع الاختبارات النفسية وكيفية بناءها وغيرها من التفاصيل التي يجب أن يلم بها أي باحث قبل استخدامه للاختبار النفسي.

في حين يهدف الفصل الثالث إلى الكشف عن أنواع الصدق في الاختبارات النفسية، وكيف أن هنالك العديد من الألقاب التي تطلق على النوع الواحد، مما تسبب في ارتباك واضح في

البحوث والدراسات، وهو ما دفع بنا إلى محاولة الفصل في هذا الخلط الواقع لدى الباحثين، من خلال عرض كل نوع بشكل منفصل والتفصيل في الإجراءات المرتبطة به، وارتأت الباحثة أنه يجب عرض النظرية الحديثة للصدق، ليستطيع الباحث بعدها، تحديد إجراءاته انطلاقا من إلمامه بجميع أنواع الصدق مهما اختلفت تسمياتها وإجراءاتها. وقد تم اختتام هذا الجزء بعرض الصدق حسب قائمة المعايير كوسمين COSMIN، والإشارة إلى بعض المعايير المتعلقة به والتي تم اعتمادها في الحكم على الصدق أثناء عملية التقييم.

وغير بعيد عن نفس الفكرة فقد هدف الفصل الرابع إلى عرض جميع الجوانب المتعلقة بالثبات، من خلال التعريف به وذكر أنواعه وإجراءاته، لتعرج بعدها الباحثة إلى العوامل المؤثرة فيه، بغية أن يأخذها الباحث في الحسبان عند التحقق من ثبات اختباره، وفي الأخير تطرقت الباحثة إلى الثبات وفق قائمة كوسمين COSMIN، وعرض بعض المعايير التي تم اعتمادها في تقييم طرق إعادة التحقق من الثبات في عينة الدراسة الحالية.

أما في الفصل الخامس فقد ركزت الباحثة على عرض المنهج والمجتمع الذي سحبت منه عينة الدراسة، لتفصل بعدها في كيفية سحب العينة مع شرح منهجية برزما PRISMA التي اتبعتها الباحثة في حصر تلك العينة، لتتطرق بعدها إلى أداة الدراسة المتمثلة في قائمة أو شبكة COSMIN، بعد التعريف بها، ومن ثم عرض التعديلات التي ألحقتها بها، لتعرج الباحثة بعد ذلك للأساليب الإحصائية المستخدمة في تحليل بيانات هذه الدراسة.

وأخيرا الفصل السادس والذي خصص لعرض النتائج ومناقشتها ومن ثم استخلاص أهم الاستنتاجات، لتختم الباحثة بعد ذلك هذه الدراسة مقدمةً بعضا من الاقتراحات التي ترى أنها مهمة لمواصلة تطوير هذا البحث وتحسينه.

الجانب النظري

الفصل الأول

مدخل نظري عام للدراسة

1-الإشكالية

تحتل الاختبارات النفسية دوراً مهماً في البحوث والدراسات الأكاديمية، سواء كان الهدف منها التشخيص أو التقييم أو حتى التنبؤ، فإن على الباحث أن يقوم ببنائها أو اختيارها بعناية، وعادة ما يلجأ الباحثون إلى استخدام الاختبارات والمقاييس المعدة مسبقا والمتوافرة عالميا، لأنه كما هو معلوم فإن عملية إعداد أو بناء الاختبارات النفسية عملية شاقة وصعبة تحتاج إلى أن تتضافر فيها جهود العديد من الخبراء والمختصين في القياس وفي مجالات أخرى.

وفي نفس السياق فإن عملية اختيار الاختبار الجاهز ليست عشوائية، فمن الضروري دراسة وتقييم جميع تفاصيل الاختبار قبل استخدامه، لأن الباحث غير قادر على تمييز جودة الاختبار وجاهزيته للاستخدام الحالي؛ فقط لأنه يدرس نفس السمة التي يود الباحث دراستها، ولذلك اقترح كل من سوزا، أليكسندر وغويراردلو Souza, Alexandre, & Guirardello (2017) من أجل تقييم ملائمة الاختبار، أن يتم التحقق من الغرض الذي بني من أجله، المجالات التي يستخدم فيها، المجتمع الذي بني فيه، البيئة التي يصلح فيها، وخصوصًا خصائصه السيكومترية.[1] كل هذا قبل اختيار الاختبار وذلك من أجل ضمان جودة النتائج.

ولابد من القول بأن هناك صفات أساسية تمثل الملامح الهامة للحكم على الاختبار بأنه جيد، بعيدا عن مدى ملاءمته للاستخدام المرجو، وتتمثل هذه الصفات في الصدق والثبات بالإضافة إلى وجود صفات ثانوية ممثلة بسهولة التطبيق وسهولة التصحيح، والتكلفة المادية وغيرها. ولكن كما

[1] Souza, A., Alexandre, N. M., & Guirardello : Edinêis de BritoPropriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. Epidemiologia e Serviços de Saúde. 26 (3). 2017, pp 649-659.

أكد سلموند Salmond (2008) بأن الحكم على أداء الاختبار النفسي مرتبط أكثر بصدقه وثباته[1].

ويتفق العديد من خبراء القياس على اعتبار الصدق والثبات أهم الخصائص التي يجب توفرهما في الاختبارات النفسية.[2]

وترجع تلك الأهمية لكون الخصائص السيكومترية للاختبارات النفسية، هي التي تحدد مدى مصداقية الأداة وموضوعيتها، وتحدد إمكانية الاعتماد على نتائجها وتعميمها، فحسب بودلال و موراس (2021)[3]. تعد من الإجراءات المنهجية الأكثر أهمية في البحوث النفسية والتربوية، ذلك أن ضعف صدق وثبات الأدوات البحثية يؤدي بالضرورة إلى ضعف صحة وسلامة نتائج الدراسة بأكملها، لذلك يحرص الباحث على اختيار الأداة ذات الصدق والثبات العاليين. ولكن بسبب ازدياد عدد الاختبارات النفسية المتاحة بشكل كبير خلال العقود الماضية، فإن اختيار الاختبار الذي سيتم استخدامه أصبح يمثل صعوبة كبيرة. فرغم أن العديد من الاختبارات قد بنيت ووضعت موضع التطبيق، إلاَ أن الكثير منها لم يتم إعادة التحقق من صلاحيتها بصورة موضوعية.[4]

---

[1]Salmond, S : Evaluating the reliability and validity of measurement instruments. Orthop Nurs.27. 2008, pp 19-30.

[2] Mokkink, L., de Vet, H., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., et al : COSMIN risk of bias checklist for assessing the methodological quality of studies on the measurement properties of Patient-Reported. epub ahead of print . 2017, pp. 1171-1179.

3 فتيحة بودلال، و محبوبة موراس: دراسة تقييميه لقياس الخصائص السيكومترية (الصدق والثبات) في البحوث العلمية. *مجلة ضياء للبحوث النفسية والتربوية . 1* (2). 2021، ص 111-131.

[4] Chen, X., & Gabrenya, W. K: In search of cross-cultural competence: A comprehensive review of five measurement instruments. International Journal of Intercultural Relations . 82.2021, pp 33-55.

بحيث نجد العديد من الاختبارات ذات الجودة الرديئة، والتي لم يتم فيها احترام معايير ومبادئ بناء الاختبارات، وحتى خصائصها السيكومترية والتي من المفروض أن تدل على صلاحيتها ومستوى الثقة فيها، نجدها أصبحت تتعرض للعديد من التجاوزات والأخطاء سواء في اختيار الطرق والأدلة أو في كيفية استخدامها أو في اختيار الأساليب الإحصائية المناسبة.

وغير بعيد عن نفس الفكرة السابقة فحتى الاختبار الجيد الموثوق في صلاحيته والمعروف بجودته، لم يسلم من الانتهاكات، ولكن هذه المرة ليس من مطور الاختبار نفسه إنما من مستخدمه، فيقوم بعض الباحثين حسب ظاظا (2011) بإضافة أو تعديل أو حذف فقرات من الاختبار دون الأخذ بعين الاعتبار التغير الذي قد يحدث في دلالات صدقه وثباته نتيجة تغيير في بنية فقراته الأصلية ومعانيها ومدلولاتها ومدى مناسبتها للفئات العمرية أو المجموعات المختلفة[1]. وهناك من الباحثين من يكتفون بالخصائص السيكومترية الواردة للاختبارات في البيئات الأصلية لها، رغم تغير العينة والغرض من الاختبار وغيرها، في حين يعمل البعض الآخر على محاولة استخراج دلالات صدق وثبات جديدة للاختبارات كإجراء إحصائي محض لإعادة التحقق من صلاحية الاختبارات.[2] في حين أن البعض الآخر من يكتفي بعرض الاختبار على مجموعة من المحكمين، ورغم اقتراحهم إضافة عدد من البنود أو إزالتها وغيرها من التعديلات، إلا أن الباحثين يقررون أن الاختبار بعد عرضه على المحكمين وتعديل ما طلب منهم أصبح جاهز

---

[1] حيدر إبراهيم ظاظ: درجة توافق دلالات صدق وثبات الاختبارات المقننة المستخدمة في رسائل الماجستير المقدمة في كليات التربية في الجامعات الأردنية مع دلالات صورها الأصلية. *دراسات العلوم التربوية ، 38* (2). 2011، ص2400-2416.

[2] Beichi, A., Tchalabi, S., & Alsalhi, N: Internal Consistency Between Validity and Reliability in. *Journal of Statistics Applications & Probability .11* (1). 2022, pp 207-216.

للاستخدام، متجاهلين في ذلك إخضاع البنود الجديدة التي تمت إضافتها أو تعديلها وفقًا لآراء المحكمين، للتجريب والتحليل الإحصائي للتحقق من صلاحية الاختبار المعدل للبيئة الجديدة.

ومن أجل هذه التجاوزات وغيرها نشهد مؤخرًا نشر عدد كبير من الدراسات التقييميه للاختبارات المتاحة، على سبيل المثال دراسة (Ong, Krafft, Levin, & Twohig, 2021) و (El-Den, Chen, Gan, Wong, & O'Reilly, 2018) ودراسة (Polit, 2015). وهدفت جميع هذه المراجعات المنهجية إلى نفس الهدف وهو تقييم جودة الاختبارات النفسية المنشورة في البحوث والمقالات، وتوصلت جميعها إلى أن جودة الاختبارات النفسية ليست في المستوى المطلوب، وأن غالبية الباحثين يميلون إلى التركيز على طرق معينة في تقدير الصدق والثبات أكثر من غيرها مثل ألفا كرونباخ، كذلك اتفقت الدراسات الثلاثة السابقة على أن عملية التحقق من صلاحية الاختبارات النفسية سواء للاختبارات الجاهزة أو التي يتم بناؤها غالبا ما تعتمد على النظرية التقليدية للقياس.

وعادة ما تهدف الدراسات العربية التقييميه للاختبارات، إلى الكشف عن الأخطاء التي يقع فيها الباحثون، سواء عند استخدام الاختبار أو بناءه أو غيرها مُحاولةً تقويم هذه الانتهاكات، في حين نجد أن الدراسات الأجنبية تميل إلى المراجعات المنهجية من أجل الحكم على جودة الاختبار في حد ذاته، بهدف نشر تلك الدراسات في مواقع خاصة بالمراجعات المنهجية مثل موقع PRISMA، ومن ثم تمكن الباحثين من اختيار الاختبار المناسب بناءا على جودته.

وفي ظل الانتشار الكبير للمصطلحات والتعاريف المتعددة لخصائص القياس، أصبح الباحث يجد صعوبة في التمييز بين تلك الطرق والأنواع لكثرتها وتداخل إجراءاتها حتى صارت الإجراءات أنواع والأنواع إجراءات، وهذا ما أكدته نتائج دراسة كل من (Tian, Cao, & Feng, 2019) و

(Zeilinger, et al., 2022) بحيث توصلت إلى أن هنالك تداخل واضح بين أدلة الصدق وإجراءاته خاصة، أكثر من أدلة الثبات فهي أكثر وضوحا عند الباحثين. وهذا راجع غالبا إلى تعدد مسميات الطريقة الواحدة فنجد مثلا الصدق الظاهري، صدق المحكمين، صدق الخبراء، صدق الوجه وغيرها من التسميات التي كلها في الأصل تدل على نوع واحد وهو صدق المحتوى، وتعتبر هذه التسميات في الأصل من إجراءات التحقق منه.

ولكي يستطيع الباحث الاختيار فإنه يستند إلى الخصائص السيكومترية للاختبارات المتاحة، ومن هنا ظهرت الحاجة إلى معايير موحدة للحكم على الجودة المنهجية للخصائص السيكومترية. وحسب تروي وآخرون Terwee et al (2007) فإن العديد من المقالات تقدم معايير لتقييم الاختبارات، ومع ذلك فإن ما تفتقر إليه هذه المعايير غالبًا؛ هو الوضوح لما يشكل خصائص قياس جيدة، على سبيل المثال لم يتم تحديد معايير لحجم العينة المطلوب للدراسات التي تقيم الخصائص السيكومترية. ولهذا السبب عمل تروي وآخرون Terwee et al (2007) على وضع بعض المعايير التي يمكن أن يستند الباحث إليها عند البناء أو التكييف أو إعادة التحقق من صلاحية الاختبارات، إلا أنها دائما ما كانت تحتاج إلى تعديل وتطوير.[1]

هذا ما دفع بمجموعة من الباحثين إلى تأسيس مبادرة COSMIN –اختصار COnsensus –based Standards for the selection of health Measurement Instruments التي عملت على تبني هذه الفكرة وتطويرها وجعلها أكثر شرعية من خلال الاعتماد على طريقة

[1] Terwee, C. B., Bot, S. D., de Boer, M. R., Van der windt, D. A., Knol, D. L., Dekker, J., et al: Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. *60* (1). 2007, pp 34-42.

دلفي (Delphi Method)، بهدف التوصل إلى توافق في الآراء بين فريق دولي من خبراء القياس والإحصاء وغيرها، كل هذا من أجل الخروج في الأخير بقائمة المعايير COSMIN[1]، والتي تهدف إلى توحيد مفاهيم القياس وتوحيد المعايير التي يستند إليها في اختيار أدلة الصدق والثبات، وإنشاء قوائم مرجعية لتقييم الاختبارات في البحوث، وتوحيد القيم التي من خلالها يتم الحكم على جودة الاختبارات، والعديد من الأهداف الأخرى. الأمر الذي دفع بالباحثة إلى ترجمة هذه القائمة الى اللغة العربية واعتمادها، من أجل الكشف عن مدى التزام الباحثين بالمعايير المطلوبة، عند إعادة تحققهم من صلاحية الاختبارات النفسية قبل تطبيقها، وذلك من خلال الإجابة عن التساؤلات التالية والتي تعبر عن مختلف الإجراءات المطلوب من الباحثين القيام بها للتحقق من صلاحية اختباراتهم.

- ما هي الطرق الأكثر استخداما من طرف الباحثين في إعادة تحققهم من صدق وثبات الاختبارات النفسية في عينة من الدراسات الأكاديمية العربية والأجنبية؟
- هل صدق المحتوى الذي تحقق منه هؤلاء الباحثين تم بطرق سليمة موافقة لمعايير COSMIN؟
- ما نسبة الدراسات التي تحققت من الصدق الهيكلي للاختبار النفسي بالشكل المطلوب؟
- هل تم التحقق من ثبات الاتساق الداخلي للاختبار بجميع أشكاله باستخدام الإجراءات الصحيحة؟

[1] Terwee, C., Prinsen, C., Chiarotto, A., Westerman, M., Patrick, D., Alonso, J., et al: COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*. *27* (5).2018, pp 1159-1170.

- هل تحققت الدراسات التي استخدمت ثبات الاتساق الداخلي من أحادية البعد بالنسبة للسمة قبل استخدامه؟
- هل تم التحقق من صلاحية الاختبارات المنقولة ثقافيا باستخدام الخطوات المحددة؟
- هل قام الباحثون بالكشف عن أخطاء القياس؟
- إلى أي مدى التزم الباحثون بالمعايير الموضوعة عند استخدامهم التطبيق وإعادة التطبيق وكذا الصور المتكافئة في التحقق من ثبات الاختبار؟
- ما مدى استخدام الباحثين للصدق المحكي والتزامهم بخطوات التحقق منه؟
- هل تم التحقق من صدق الفرضيات بالشكل المطلوب حسب معايير COSMIN؟
- هل تم تقييم الاستجابة للاختبارات النفسية بالطريقة الصحيحة حسب المعايير التي وضعتها COSMIN؟
- هل تعتبر الخصائص السيكومترية التي تم استخدامها في إعادة التحقق من صلاحية الاختبارات النفسية ذات جودة مقبولة؟

وللإشارة فإن الإجابة على السؤال الواحد تستدعي الكثير من الأسئلة الدقيقة ليتم في الأخير حصر جميع الإجابات، في محاولة للخروج بنتيجة إجمالية كإجابة على التساؤل الواحد من التساؤلات السابقة وهو ما سيتضح جليا في الجانب التطبيقي.

2-أهداف الدراسة

تهدف هذه الدراسة الحالية إلى:

- التعرف على أكثر أنواع الصدق والثبات استخداما في إعادة التحقق من صلاحية الاختبارات النفسية المطبقة في عينة من الدراسات الأكاديمية العربية والأجنبية.

- الكشف عن العيوب التي يخلفها الباحثون عند إعادة التحقق من صلاحية الاختبارات النفسية.
- تقييم جودة الخصائص السيكومترية المستخدمة في إعادة التحقق من صدق وثبات الاختبارات النفسية في عينة من الدراسات الأكاديمية العربية والأجنبية.
- تسليط الضوء على معايير القياس الحديثة في تقييم الخصائص السيكومترية للاختبارات النفسية.

3-أهمية الدراسة

تنبع أهمية الدراسة الحالية من أهمية الصدق والثبات بالنسبة للاختبارات والمقاييس النفسية، والتي تعتبر كمعايير للحكم على جودتها، ومنه دقة النتائج وإمكانية الوثوق بها.

وتضيف الدراسة الحالية دليلًا مفيدًا للدراسات العربية، حيث أنها من الدراسات السباقة التي استخدمت أداة بحث دولية متفق عليها COSMIN في الدراسات العربية-حسب الاطلاع الموسع الذي قامت به الباحثة في هذا المجال-وذلك من أجل التحقق من جودة الخصائص السيكومترية للاختبارات النفسية.

كما تكمن أهميتها في مساعدة الباحثين على اختيار الطريقة المناسبة لإعادة التحقق من صدق وثبات أدوات القياس من خلال الكشف عن الأخطاء التي وقع فيها غيرهم عند إعادة التحقق من صلاحية أدوات القياس.

وتبرز أهمية الدراسة الحالية في تركيزها على استخدام أحدث المعايير الصادرة في تقييم جودة الخصائص السيكومترية للاختبارات النفسية، والعمل على ترجمة هذه المعايير الى اللغة العربية من أجل توفيرها للباحثين. وتعمل الباحثة حاليا بالتعاون مع مبادرة COSMIN على ترجمة قائمة

المعايير وشبكة التحليل وكذلك دليل الاستخدام الذي يتكون من عدة أجزاء إلى اللغة العربية، وإطلاق تلك الترجمة في الموقع الرسمي للمبادرة. https://www.cosmin.nl/

4-تحديد المفاهيم إجرائيا

سوف يتم عرض المفاهيم حسب ترتيب ورودها في عنوان الدراسة، وتعريفها إجرائيا كما يلي:

4-1-دراسة تقييميه

ويقصد بها في هذه الدراسة، القيام بعملية جمع البيانات وتحليلها وتفسيرها من أجل تقييمها والحكم على جودتها من خلال مدى التزام الباحثين في عينة الدراسة الحالية بالمعايير الواردة في قائمة COSMIN.

4-2-طرق إعادة التحقق

والمقصود بها هو الكشف عن الخصائص السيكومترية للاختبارات النفسية التي تم التحقق من صدقها وثباتها في استخدام سابق، وبالضبط هي جميع أنواع الصدق والثبات التي تم استخدامها في عينة الدراسة الحالية. والمقصود بالثبات والصدق هنا هي جميع الإجراءات التي استخدمها الباحثون لإثبات أن الاختبار موثوق ويقيس ما يراد به.

4-3-صلاحية الاختبارات النفسية

الاختبار الصالح هو الاختبار الموثوق في نتائجه أي هو الاختبار الصادق والثابت، وبالتالي فإن المقصود بصلاحية الاختبارات النفسية هنا أي جاهزيتها للاستخدام المرجو.

5-الدراسات السابقة

لا يكاد يخلو أي بحث في الدراسات النفسية الأكاديمية من استخدام الاختبارات النفسية، لما لها من أهمية، ولكن هذا الإفراط في الاستخدام، جعل بعض الباحثين يقعون في أخطاء عديدة، انطلاقا من اختيار الاختبار وصولا إلى تحليل نتائجه، الأمر الذي دفع المختصين في القياس والباحثين الآخرين إلى محاولة ترشيد هذا الاستهلاك، وذلك بتقييم الدراسات التي اعتمدت على الاختبارات النفسية، وقد اختارت الباحثة عرض بعض الدراسات الحديثة التي تمس موضوعها، من حيث تقييم طرق التحقق من صلاحية الاختبارات النفسية في الدراسات الأكاديمية سواء العربية أو الأجنبية.

5-1-دراسات اهتمت بتقييم قياس صلاحية الاختبارات النفسية في الدراسات الأكاديمية

ستعرض هنا الباحثة أهم الدراسات التي استفادت منها في رسم خطة واضحة للدراسة الحالية:

5-1-1-دراسة حيدر ظاظا (2011) بعنوان درجة توافق دلالات صدق وثبات الاختبارات المقننة المستخدمة في رسائل الماجستير المقدمة في كليات التربية في الجامعات الأردنية مع دلالات صورها الأصلية، وهدفت الدراسة إلى التحقق من درجة توافق دلالات صدق وثبات الاختبارات المقننة المستخدمة في رسائل الماجستير مع دلالات صورها في البيئة الأصلية، ولتحقيق غرض الدراسة تمت مراجعة (27) رسالة ماجستير أجريت في الفترة 1980 -2006 وأظهرت النتائج وجود اتفاق بين الباحثين في الإجراءات المتبعة لتقنين الاختبارات للبيئة الجديدة بلغت 97 %. كما تبين أن أكثر دلالات الصدق استخدامًا في البيئتين :صدق المحك التلازمي، والارتباطات بين الفقرات أو الأبعاد، في وطريقة الإعادة على الترتيب. في حين كانت أكثر دلالات

الثبات استخدامًا في البيئتين معادلة كرونباخ ألفا وطريقة الإعادة على الترتيب، وأظهرت النتائج وجود فروق ذات دلالة إحصائية في تكرار استخدام صدق المحكمين بين البيئتين ولصالح البيئة الجديدة، وفي تكرار استخدام صدق المحك التنبؤي واستخدام صدق المحتوى بين البيئتين، ولصالح البيئة الأصلية .أما فيما يتعلق بالثبات فلم تظهر النتائج فروق ذات دلالة إحصائية في تكرار استخدام جميع دلالات الثبات بين البيئتين. كذلك أظهرت نتائج مربع كاي2 وجود توافق في استخدام دلالات الصدق والثبات في البيئتين؛ إلاَّ أن هذا التوافق كان ضعيفًا حسب معيار كوهن. وفي ضوء ذلك أوصت الدراسة ضرورة إجراء الترجمة العكسية من قبل متخصصين بمجال الاختبار واللغة التي كتب فيها الاختبار في بيئته الأصلية ولغة البيئة الجديدة، وذكر قيم وأنواع دلالات ثبات وصدق الاختبارات عند تقنينها لبيئات أو ثقافات جديدة.

5-1-2-دراسة همام السردي (2012) بعنوان تقويم أدوات البحث المستخدمة في البحوث التربوية لدى طلبة الدراسات العليا في الجامعة الإسلامية بغزة. وهدفت هذه الدراسة لتقويم أدوات البحث المستخدمة في البحوث التربوية لدى طلبة الدراسات العليا في الجامعة الإسلامية بغزة، واستخدم الباحث في دراسته المنهج الوصفي التحليلي على جميع رسائل الماجستير المقدمة لكلية التربية، وبلغ عددها 404 رسالة، واستخدم الباحث بطاقة تحليل كأداة لدراسته وذلك بعد إعداد قائمة بالمعايير وكانت النتائج كما يلي: استخدمت 93.32% من الرسائل إحدى أدوات جمع البيانات، حيث استخدمت الإستبانة 48.51%، والاختبار30.53% وتحليل المحتوى 13.37% والملاحظة 6.60%، وأخيراً المقابلة استخدمت بنسبة 0.99% من الدراسات المستخدمة للأدوات. أما بالنسبة لمعاملات الصدق فكانت أكثرها استخداماً هو صدق المحكمين بنسبة 51.60%يليه صدق الاتساق الداخلي بنسبة 84.41%، ثم الصدق التمييزي بنسبة 5.56% وأخيرًا الصدق العاملي بنسبة 0.91%، كان الاستخدام صحيحاً فيها كما يلي: صدق المحكمين 98.76%،

الاتساق الداخليٍ99.78%، الصدق العاملي 100%، الصدق التمييزي9.84%، من إجمالي الاستخدام لكل معامل منها. وفي ما يخص معاملات الثبات فكان أكثرها استخداماً هو معامل التجزئة النصفية بنسبة 41.75% كانت صحيحة 80.24 %منها، يليه معامل كرونباخ ألفا بنسبة استخدام 35.54%، الصحيح منها نسبته 74.79%، ثم معامل كودر ريتشاردسون بنسبة استخدام قدرها 8.45% كانت صحيحة جميعها، ثم معامل هوليستي بنسبة استخدام 6.62% كانت صائبة جميعها، كذلك معامل الاتفاق استخدم بنسبة 3.56% جميعها استخدمت بشكل سليم، يأتي بعد ذلك طريقة إعادة التطبيق والتي استخدمت بنسبة 2.55% بنسبة صواب 36%، وأخيراً استخدم معامل كوبر بنسبة 1.5% جميعها صائبة.

5-1-3-دراسة عبد العزيز بوسالم (د س) تحت عنوان منهجية مقترحة لإعادة التحقق من صدق وثبات الاختبارات النفسية، وهدفت الدراسة إلى تتبع واقع تطبيق الاختبارات النفسية غير المكيفة في الدراسات الأكاديمية المطبقة في البيئة الجزائرية من خلال تناول 82 مذكرة ماجستير وأطروحة دكتوراه (أنجزت في سبعة جامعات وطنية) في تخصصات علم النفس وعلوم التربية كمحاولة لمعرفة مدى موضوعية الإجراءات التي يقوم بها أصحاب الدراسات من خلال محاولتهم إعادة التحقق من صدق وثبات مختلف المقاييس والاختبارات التي طبقوها في دراساتهم بحجة النسبية الثقافية، وقد أشارت الدراسة التحليلية إلى أن جميع هذه الدراسات- باستثناء دراستين- توصلت تحاليلها الإحصائية إلى قبول قيم صدق وثبات هذه الاختبارات، وبالتالي تم تطبيقها كما هي وبدون أي تعديل. وقد توصلت الدراسة إلى أن الباحثين قاموا بإعادة التحقق من الصدق والثبات بطرق غير مناسبة تماما لطبيعة السمات التي يقيسها الاختبار، كما أنهم لم يقدموا التبرير العلمي لاختيارهم أنواع معينة من الصدق والثبات دون الأنواع الأخرى. أيضا من بين النتائج التي أسفرت عنها هذه الدراسة أن نوع الصدق الأكثر استعمالا هو صدق المحكمين بتكرار 42 مرة أي

بنسبة 56% من المجموع الكلي، وفي الرتبة الثانية صدق المحتوى بمجموع 21 مرة أي بنسبة 28%، كما أستعمل بعض الباحثين صدق المحكات لإعادة التحقق من درجة صدق اختباراتهم بتكرار 09 مرات وبنسبة 12%، وأخيرا صدق المفهوم بتكرار 03 مرات 04% وهي أصغر نسبة، أي اقل أنواع الصدق استعمالا.

5-1-4-دراسة لوتزين وآخرون Lotzin, et al (2015) تحت عنوان مراجعة تقييميه لأدوات المراقبة لقياس التفاعل بين الوالدين والرضيع. وتهدف هذه الدراسة إلى تقييم ووصف أدوات المراقبة المستخدمة لتقييم التفاعل بين الوالدين والرضيع وذلك من حيث سلامتها السيكومترية، وقد تم البحث في عشرين قاعدة بيانات إلكترونية من يوم انطلاقها حتى جوان 2013. ليتم تضمين 24 أداة من بين 23961 مقال، وقد قيمت الدراسة الصدق في خمسة مجالات (محتوى الاختبار، وعملية الاستجابة، والهيكل الداخلي، والعلاقات مع المتغيرات الأخرى)، وقد أظهرت النتائج أن معظم الأدوات تم الاعتماد في الحكم على صلاحيتها من خلال صدق المحكمين، والصدق التمييزي، ثبات إعادة التطبيق والصيغ المتكافئة، وذلك بناءً على دراسات ذات جودة موثوقة. في حين تفتقر هذه الأدوات إلى الصدق العاملي والصدق التنبؤي، والمعايير الموحدة. وبالتالي هناك حاجة إلى مزيد من الصقل للأدوات الموجودة، خاصة في مجالات صدق المحتوى والصدق التنبؤي. قد ترشد أدلة الصلاحية هذه وأوصاف الأدوات المذكورة في هذه الدراسة الباحثين في اختيار الأداة المناسبة.

5-1-5-دراسة عبد الرحمان بن صافي (2017) تحت عنوان دراسة تقييميه لطرق تقدير صدق وثبات أدوات جمع المعطيات في أطروحات دكتوراه علم النفس وعلوم التربية بجامعة الجزائر 2، وهدفت الدراسة إلى التعرف على مدى جودة طرق تقدير صدق وثبات أدوات جمع المعطيات

في أطروحات دكتوراه علم النفس وعلوم التربية بجامعة بليدة 2. ولتحقيق تلك الأهداف قام الباحث ببناء بطاقة لتقييم طرق تقدير الصدق والثبات، وتمثلت عينة الدراسة في 346 أداة جمع المعطيات مستخدمة في 112 أطروحة دكتوراه في علم النفس وعلوم التربية بجامعة الجزائر 2. وقد أسفرت النتائج على أنه تم التحقق من صدق أدوات جمع المعطيات بنسبة 72%، أما الثبات فبنسبة 69.90%. وأكثر طرق التحقق من الصدق شيوعا هي طريقة صدق المحكمين بنسبة 43%، أما بالنسبة للثبات فكانت طريقة معامل ألفا كرونباخ بنسبة 39.57%. كذلك لوحظ أن استخدامات طرق التحقق من الصدق كانت جيدة بنسبة 93.38، أما الثبات فبنسبة 83.09%. ومن بين النتائج التي وردت أيضا أنه من الأخطاء التي وقع فيها الباحثون عند استخدامهم لطرق التحقق من الصدق: عدم كفاية طرق التحقق بنسبة66.67%، وعدم التحقق من الصدق على عينة ممثلة بمجتمع الدراسة بنسبة 53.61%، أيضا عدم التحقق من الصدق على عينة مناسبة الحجم بنسبة 07.83%، أما الأخطاء المتعلقة بالثبات فقد كانت عدم التحقق من الثبات على عينة ممثلة لمجتمع الدراسة بنسبة 54.55%، كذلك عدم التحقق من الثبات على عينة مناسبة الحجم بنسبة 10.46%، كما لم يكن معامل الثبات مقبولا بنسبة 09.06%.

5-1-6-دراسة ويتكوفسكي، جاريت، كالام وويسبرغ Wittkowski, Garrett, Calam, & Weisberg (2017) بعنوان مقاييس التقرير الذاتي للفعالية الذاتية للوالدين: مراجعة منهجية للأدبيات الحالية، وكان الغرض من هذه المراجعة المنهجية هو فحص وتقييم الخصائص السيكومترية لمقاييس الفعالية الذاتية المتاحة، وقد تم البحث في إحدى عشرة قاعدة بيانات إلكترونية. وتم تضمين المقالات التي تحققت من صلاحية مقياس جديد أو إعادة التحقق من صلاحية مقياس جاهز، وكان العدد النهائي 34 مقياسًا. وبشكل عام كانت جودة التدابير

المتاحة متنوعة، فكما يتضح تشمل الخصائص السيكومترية التي تم فحصها بدقة في الغالب صدق المحتوى (حصل 34/18 الأكثر استخداما)، صدق المحكمين (34/16)، والاتساق الداخلي (34/14)، وصدق البناء (34/13) وقابلية التفسير (34/12). في المقابل تم إيلاء اهتمام أقل لتقييم الثبات، والاستجابة والصدق المحكي. وقد لوحظ أنه من بين المقاييس الـ 18 التي أظهرت صلاحية ممتازة للمحتوى ، تم الإبلاغ عن جميع الـ 18 مقياسًا في عدة دراسات لتقييم الخصائص السيكومترية.

5-1-7-دراسة إلدين، شين، غان، ونغ وأريلي El-Den, Chen, Gan, Wong, & O'Reilly (2018) تحت عنوان الخصائص السيكومترية لأدوات فحص الاكتئاب في أماكن الرعاية الصحية الأولية: مراجعة منهجية. وقد هدفت الدراسة إلى تقييم الخصائص السيكومترية لأدوات فحص الاكتئاب، وقد تم إجراء بحث منهجي في الأدبيات يتكون من أربعة مصطلحات (الفحص والقياس النفسي والاكتئاب والرعاية الصحية الأولية) في PubMedو EMBASE و PsycINFO و MEDLINE، بين يناير 1995 حتى أكتوبر 2015. وتم تضمين 60 منشورًا، لتقييم الخصائص السيكومترية لـ 55 أداة أو تكييفها، أبلغت النتائج عن قيم الحساسية والنوعية تراوحت بين من 28٪ إلى 100٪ و 43٪ إلى 100٪ على التوالي، وتراوحت قيم كرونباخ ألفا من 0.56 إلى 0.94، وكانت الأشكال الأخرى من اختبار الثبات والصدق أقل اتساقًا واستخداما.

5-1-8-دراسة بودلال ومورواس (2021) بعنوان: دراسة تقييميه لقياس الخصائص السيكومترية (الصدق والثبات) في البحوث العلمية. وهدفت الدراسة إلى تقييم قياس الخصائص السيكومترية من صدق وثبات للأدوات البحثية المطبقة في رسائل الدكتوراه، ولأجل ذلك تم استخدام

المنهج الوصفي من خلال تحليل محتوى (21) رسالة دكتوراه، تم اختيارها بطريقة عشوائية من بين (206) أطروحة أنجزت في قسنطينة2 في الفترة الممتدة بين (2011/2010) إلى غاية (2020/2019)، وذلك باستعمال شبكة لتقييم قياس الخصائص السيكومترية للأدوات البحثية صممت من طرف الباحثين، وقد توصلت النتائج إلى: أن الصدق الأكثر استعمالا هو صدق المحكمين 76.2%، يليه الاتساق الداخلي 19% ثم المقارنة الطرفية14.3% أما الأقل استعمالا هو الصدق الذاتي 4.8%، أما الصدق العاملي فلم يستعمل بتاتا، هذا في ما يخص الصدق أما في ما يخص الثبات فالإجراء الأكثر استخداما تمثل في التجزئة النصفية وألفا كرونباخ بنسبة 38.1%، واستعمل أسلوب إعادة الاختبار بنسبة قدرت بـ 23.8%، أما أسلوب الصور المتكافئة وكيودر–ريتشاردسون فلم يطبقا في أي رسالة، وفي الأخير توصلت هذه الدراسة إلى أنه لم يتم التحقق من صدق وثبات الأدوات البحثية في رسائل الدكتوراه (عينة الدراسة) بطريقة صحيحة.

5-1-9-دراسة بايشي، تشلابي والصالحي (2022) Beichi, Tchalabi, Alsalhi تحت عنوان الاتساق الداخلي بين الصدق والثبات: دراسة تقييميه. حيث هدفت هذه الدراسة إلى مناقشة جدلية الاتساق الداخلي بين الصدق والثبات، ومن ثم تقييم عينة من الدراسات العربية والأجنبية التي عمدت إلى التحقق من صلاحية الاختبارات النفسية قبل استخدامها، وتكونت العينة من (83) اختبار موزع على رسائل الدكتوراه والماجستير والمقالات، في أكثر من (20) بلد، وقد تم استخدام المنهج الوصفي التقويمي لملاءمته لطبيعة الدراسة. وتم الاعتماد على شبكة تقييم أعدَها فريق من الباحثين، وقد أظهرت النتائج أن الاتساق الداخلي أقرب في معناه إلى الثبات أكثر منه إلى الصدق. وحتى وإن تم اعتباره جزافا صدقا، لا يمكن أبدا الاكتفاء به كدليل على صدق الاختبار بل وجب دائما مصاحبته بأدلة أخرى. وأسفرت النتائج أيضا على أن أكثر أدلة الصدق

استخداما في الدراسات الأكاديمية العربية والأجنبية هي صدق الاتساق الداخلي بـ37 اختبار من مجموع 83 اختبار، وصدق المحكمين بـ13 اختبار. مع الإشارة إلى أن العديد من معاملات الارتباط المقبولة في الاتساق الداخلي ضعيفة جدا وأحيانا غير دالة. كذلك من بين النتائج استخدام الاتساق الداخلي كدليل صدق وثبات في آن واحد في 22 اختبار من أصل 83. وأما في ما يخص الثبات فألفا كرونباخ أكثر أنواع الثبات استخداما في الدراسات الأكاديمية بـ64 اختبار من أصل 83.

5-2-دراسات اعتمدت على قائمة المعايير COSMIN في تقييم قياس صلاحية الاختبارات النفسية

بعد عرض الباحثة للدراسات التي استفادت منها في تقييم طرق التحقق من صلاحية الاختبارات النفسية، ارتأت أنه من المهم أيضا عرض الدراسات التي تشترك مع دراستها، من حيث اعتمادها على قائمة المعايير الدولية COSMIN في تقييم جودة الخصائص السيكومترية للاختبارات النفسية في الدراسات الأكاديمية.

5-2-1-دراسة دينيس بوليت Denise Polit (2015) بعنوان تقييم القياس في مجال الصحة: ما وراء الثبات والصدق. والغرض من هذه الدراسة هو وصف الأفكار الناشئة المتعلقة بوضع واختبار تدابير جديدة في المجالات الصحية، وتقديم تصنيف قياس منقح يستند إلى COSMIN، واستكشاف مدى دور مفاهيم القياس الجديدة في التقييمات النفسية في التمريض. وقد أجري تحليل وصفي لعينة من الأوراق النفسية التي نشرت في ثلاث مجلات تمريضية رئيسية. وتم استعراض عينة من 105 دراسة، تمثل عينة متتالية من الدراسات النفسية المنشورة في المجلة الدولية للدراسات التمريضية، وأبحاث التمريض، والبحوث في التمريض والصحة بين عامي 2010

و2014. على أن تتضمن الدراسات لتكون ضمن العينة أحد الأمور التالية: التطوير والتقييم النفسي لأداة جديدة؛ ترجمة أداة إلى لغة أخرى وتقييم النسخة المترجمة؛ تقييم سيكومتري لأداة موجودة لمجموعة سكانية جديدة أو غرض مختلف أو غيرها من الأسباب؛ أو تقييم تكييف الأداة. وذلك من أجل التقييم السيكومتري لأدوات التمريض النفسي على التصنيف الجديد للخصائص السيكومترية COSMIN. وكانت النتائج أنه تم حساب معامل ألفا في جميع الدراسات 105 باستثناء 3 دراسات فقط. وتم تقييم ثبات إعادة الاختبار في 46 دراسة أي 43.8٪، وذلك باستخدام معامل الارتباط بيرسون في 18 من هذه الدراسات الـ 46 (39.1٪)، وفشلت دراستان في تحديد إحصائية إعادة الاختبار التي تم استخدامها. واشتملت 4 دراسات فقط على مقياس الملاحظة في تقييمات الثبات بين المقيمين. كذلك تم الإبلاغ عن إحصائيات خطأ القياس في ورقتين من الأوراق الثلاث التي تتضمن نموذج نظرية استجابة راش أو البند. وفيما يخص الصدق كان صدق المحتوى الأكثر استخداما في 48 دراسة. كذلك في عينة من 105 أوراق تمريض نفسي، 19 دراسة (18.1٪) لم تتضمن التحقق من صدق بناء اختبار الفرضيات أو التحقق من صدق المعيار. في معظم هذه الأوراق الـ 19، اعتمد الباحثون بدلاً من ذلك على تحليل العوامل كنهجهم لتقييم صلاحية البناء، بحيث تم إجراء تحليل العوامل في 85 دراسة (81.0٪)، واستخدم تحليل العوامل الاستكشافية حصريًا في 51 منها (60.0٪). أما فيما يخص الصدق عبر الثقافات اشتملت 36 دراسة (34.3٪) منها على جهود لتقييم الصلاحية عبر الثقافات لمقياس مترجم. وفي الأخير أجرت 8 فرق بحثية فقط (7.6٪) تقييمات للاستجابة على النحو المحدد بواسطة COSMIN لمقاييسهم . إذن معظم الدراسات التمريضية التي تمت مراجعتها التزمت بالمفاهيم التقليدية للقياس النفسي، والتي تركز على الثبات والصدق فقط. ونادراً ما شملت الدراسات التي

أجريت في العينة تقييمات لجوانب القياس الطولي، أي ثبات وصدق درجات التغيير (الاستجابة أو المآل).

5-2-2-دراسة كاسيدي، برادلي، بوين، ويغام، ورودجرز Cassidy, Bradley, & Rodgers, Bowen, Wigham, (2018) تحت عنوان قياس خصائص الأدوات المستخدمة لتقييم الانتحار لدى البالغين المصابين بالتوحد وعامة السكان: مراجعة منهجية. وهدفت إلى تحديد الأدوات المستخدمة لتقييم الانتحار لدى البالغين المصابين بالتوحد وعامة السكان، وتقييم طرق التحقق من الخصائص السيكومترية، وتقديم توصيات للاختيار المناسب لأدوات تقييم الانتحار في البحث والممارسة السريرية. وتم البحث في ثلاث قواعد بيانات (PsycInfo, Web of Knowledge, Medline) ليتم تحديد أربع أدوات لتقييم الانتحار مستخدمة بشكل متكرر، وتم تصنيفها لاحقًا من حيث جودة الأدلة لدعم خصائص القياس الخاصة بهم باستخدام قائمة مراجعة (COSMIN) على الرغم من الدراسات التي استكشفت الانتحار لدى البالغين المصابين بالتوحد، لم يستخدم أي منهم أداة تم التحقق من صحتها بجودة عالية. فمثلا تشير النتائج إلى أن الدراسات التي تقيم صلاحية المعيار على أساس التمييز بين المجموعات الفرعية بالغت في تقدير الدقة التشخيصية للأداة. وفي الأخير كان هناك دليل قوي يدعم الاتساق الداخلي والصلاحية الهيكلية، وأدلة المعتدلة لدعم صحة المعيار، وأدلة ضعيفة في دعم اختبار الفرضيات.

5-2-3-دراسة تيان، كاو، وفانغ Tian, Cao, & Feng, (2019) بعنوان تقييم الخصائص السيكومترية لأدوات تقييم الاحتياجات النفسية لمرضى السرطان: مراجعة منهجية للدراسات. قد أجريت هذه المراجعة المنهجية بهدف تقييم جودة الخصائص السيكومترية لأدوات تقييم الاحتياجات النفسية بين مرضى السرطان من أجل مساعدة أخصائيي الرعاية الصحية على

اختيار أنسب أدوات تقييم الاحتياجات في الممارسة السريرية الروتينية. وأجريت عمليات بحث في قواعد البيانات الإلكترونية لـ PUBMED من عام 1966، وCINAHL من عام 1960، وEMBASE من عام 1980، وPsychINFO من عام 1967، فضلا عن 11 مصدر إضافي. وتم تقييم نوعية الخصائص السيكومترية للأدوات التي تم توظيفها باستخدام معايير الجودة المتفق عليها لقياس خصائص اختبارات الحالة الصحية كوسمين COSMIN. وتم تحديد 37 دراسة قيمت الخصائص النفسية لـ 20 أداة لتقييم الاحتياجات. تم اختبار الاتساق الداخلي في 32 دراسة، مع 9 دراسات تشير إلى تصنيف سلبي، و 4 دراسات تصنيف متوسط. وقيّم أقل من نصف الدراسات أي 13 دراسة ثبات الاختبار، وأفادت 4 دراسات فقط عن تصنيف إيجابي. وكان صدق المحتوى هو أكثر الخصائص النفسية اختبارًا التي تم تقييمها في 33 دراسة، وأشارت إلى تقييم إيجابي في جميع الدراسات التي تم تقييمها. وتم تقييم الصدق البنائي بشكل كاف في 28 دراسة مع 23 دراسة تفيد بتقدير متوسط. وأكثر من نصف الدراسات 29 دراسة. أما اختبار الفرضيات فتم التحقق منها في 13 اختبار وتم تقييمها إيجابية. وفي ما يخص الصدق عبر الثقافات فقد تم تقييمه في 13 دراسة و7 دراسات منها أظهرت تصنيف سلبي. ولم تتوفر بيانات عن خطأ القياس وصدق المعيار. وقد قيمت دراسة واحدة فقط مدى الاستجابة وأظهرت تصنيفاً وسيطاً.

5-2-4-دراسة ماكندروز، ريتشاردسون، ستوبا McAndrews,Richardson, & Stopa (2019)

بعنوان الخصائص السيكومترية لمقاييس القبول النفسي: مراجعة منهجية. وكان الهدف من الدراسة هو التقييم المنهجي للخصائص السيكومترية للأدوات المصممة لقياس القبول النفسي المبلغ عنه ذاتيًا في الدراسات. وتم اختيار المقالات إذا كان الهدف الأساسي هو تطوير أو تقييم خصائص القياس (الصدق، الثبات، الاستجابة) لمقياس قبول التقرير الذاتي أو

النطاق الفرعي. لتقييم الخصائص السيكومترية للدراسات المشمولة فقد اعتمدت الدراسة على المعايير المستندة إلى قائمة التحقق من أدوات قياس الحالة الصحية (COSMIN). وحددت إستراتيجية البحث 3097 مقالة. استوفى عشرون مقالاً من 32 دراسة على معايير الاشتمال. وأظهرت النتائج أنه لم يتم تقييم أي من الأدوات في جميع مجالات الخصائص السيكومترية. ولم تحقق أي دراسات في خطأ القياس أو الصدق عبر الثقافات. وكان الاتساق الداخلي هو الخاصية الأكثر تقييمًا، كذلك كان مقبولًا بشكل عام عبر جميع المستويات. أيضا من بين النتائج أن عدم مشاركة السكان المستهدفين أدى إلى ضعف صدق المحتوى لمعظم المقاييس. وكان ثبات المقيمين في اختيار الدراسة والتقييم ممتاز. كذلك أظهرت النتائج بأن هناك اختلافات مفاهيمية مهمة عبر مقاييس القبول الحالية، ولا يمكن التوصية بأي من المقاييس التي تم تقييمها على أنها تتمتع بخصائص سيكومترية فائقة. يجب معالجة القيود المهمة في صدق المحتوى، مع مشاركة أكبر للمجتمع المستهدف.

5-2-5-دراسة صيني، هوفمان، بانتيليس، إيفرال، وبوسمان ,Saini, Hoffmann, Pantelis, Everall, & Bousman, (2019) بعنوان مراجعة منهجية وتقييم نقدي لأدوات قياس إساءة معاملة الأطفال. وهدفت هذه المراجعة إلى التقييم النقدي للجودة المنهجية وخصائص القياس لتدابير إساءة معاملة الأطفال المنشورة، وفحصت قوة الأدلة على هذه الأدوات للاستخدام البحثي باستخدام المعايير المستندة إلى قائمة مراجعة القياس الصحي(COSMIN). حدد البحث المنهجي لقواعد البيانات الإلكترونية 52 دراسة مؤهلة من 2095 دراسة.أي فقط 15٪ وعدد الاختبارات التي تحويها الدراسات هي 8 اختبارات، وأسفرت النتائج عن مستويات قوية إلى متوسطة من الأدلة لثلاثة أو أكثر من معايير COSMIN التسعة.

بحيث لم يكن لدى معظم الأدوات معلومات تتعلق باتساقها الداخلي (58٪)، والثبات (56٪)، وخطأ القياس (98٪)، وصدق المحتوى (62٪)، والصدق الهيكلي (90٪)، واختبار الفرضيات (56 ٪)، الصدق عبر الثقافات (94٪)، صدق المعيار (85٪)، وأخيرا الاستجابة (100٪) لم يتم التحقق منها. أي أنه لا يوجد اختبار يحتوي على مستويات كافية من الأدلة لجميع المعايير، ولم يتم استيفاء أي معايير من قبل جميع الأدوات. وتشير النتائج أيضا إلى عدم وجود أداة واحدة متفوقة على جميع الأدوات الأخرى عبر المجموعات والسكان. توافر الاختبارات القادرة على التقاط آثار إساءة معاملة الأطفال على نمو الدماغ والأنماط السلوكية المرتبطة به محدودة .هناك ما يبرر الأدوات المصقولة مع التركيز على التقاط أحداث الإساءة أثناء التطوير بالإضافة إلى مزيد من التقييم للخصائص السيكومترية لهذه الأدوات.

5-2-6-دراسة جيويل وآخرون.Jewell, et al (2019) تحت عنوان تدابير التعلق في مرحلة الطفولة المتوسطة والمراهقة: مراجعة منهجية لخصائص القياس. وهدفت هذه الدراسة إلى تقييم الخصائص السيكومترية لاختبارات التعلق في مرحلة الطفولة المتوسطة والمراهقة في مراجعة منهجية. وذلك باستخدام معايير(COSMIN).وقد بحث اثنان من الباحثين بشكل مستقل في قواعد بيانات MEDLINE وPsychINFO وEmbase عن المقالات ذات الصلة. وشُمِلَت 54 دراسة في المراجعة. وأسفرت نتائج الجودة السيكومترية للدراسات بأنها تراوحت بين متوسطة أو ضعيفة، مع تصنيف عدد قليل فقط من الدراسات على أنها ذات جودة جيدة أو ممتازة .وبشكل عام، تضمنت الأدلة الخاصة بصلاحية الاختبارات نتائج إيجابية وسلبية تتعلق بكفاية خصائص القياس. بحيث لم تقدم أي من المقالات معلومات بخصوص جميع معايير COSMINالثمانية. على وجه الخصوص، لم تقم أي دراسة بتقييم ثبات إعادة الاختبار، أو خطأ القياس، أو صدق المحتوى، أو الاستجابة، وقامت دراسة واحدة فقط بتقييم صحة المعيار. وركزت

معظم الدراسات على الاتساق الداخلي (80٪) والصدق البنائي (70٪). وأجرى نصف الدراسات المشمولة أشكالًا مختلفة من اختبار الفرضيات؛ ومع ذلك فإن العديد منهم لم يذكر بوضوح فرضيات مسبقة. وتم اختبار ثلاث دراسات لقياس ثوابت القياس، والتي تم اعتبار واحدة منها فقط إجراءً جيدًا لثوابت القياس. وأخيرًا قدمت دراسة واحدة فقط معلومات بشأن صلاحية المعيار.

5-2-7-دراسة ويتكوفسكي، فاتر، موهيني، غاريت، وهندرسون Wittkowski, Vatter, Muhinyi, Garrett & Henderson (2020) بعنوان قياس الترابط أو الارتباط في العلاقة بين الوالدين والرضيع: مراجعة منهجية لتدابير تقييم تقرير الوالدين وخصائصها السيكومترية وفائدتها السريرية. لقد هدفت هذه الدراسة إلى تحديد جميع مقاييس تقرير الوالدين المتاحة للعلاقة بين الوالدين والرضيع أو الرابطة ومن ثم تقييم خصائصها السيكومترية والقياسية. وتم إجراء المراجعة المنهجية باستخدام معايير (COSMIN) لعام 2018. وذلك من خلال البحث في ثماني قواعد بيانات إلكترونية. وتم تضمين الأوراق التي تصف تطوير مقاييس التقرير الذاتي للروابط بين الوالدين والرضيع أو التعلق أو العلاقة من الحمل حتى عامين بعد الولادة أو تقييم خصائصهم السيكومترية. وقد تم تحديد ومراجعة خمسة وستين مقالة تقيم 17 مقياسًا أصليًا و 13 نسخة جاهزة. بعد التقييم السيكومتري لجميع المقاييس، تم تصنيف ثلاثة مقاييس فقط على أنها "كافية" لصدق المحتوى بشكل عام مع تلقي الاختبارات المتبقية تقييمات "غير متسقة ."ومع ذلك، كانت جودة الأدلة المتعلقة باختبارين منخفضة جدًا. في حين أنه تم اختبار الصدق الهيكلي لغالبية الاختبارات المشمولة (27 من 29؛ 93٪). كذلك 28 من الاختبارات الثلاثين (93.3٪) لديها دراسات أبلغت عن الاتساق الداخلي، و11 من 30 مقياسًا (36.7٪) لديها دراسات أبلغت عن ثبات إعادة الاختبار. ولم تقم أي من الدراسات المشمولة بقياس خطأ القياس والاستجابة على

النحو المحدد في معايير(COSMIN) وبالتالي تم حذفها. إذن تفاوتت الجودة المنهجية للدراسات (خطر التحيز) بين "جيد جدًا" و "غير ملائم" اعتمادًا على خاصية القياس التي تم تقييمها؛ ومع ذلك، كانت دراسات التطوير في الغالب ذات جودة "غير كافية ."على الرغم من أن معظم الاختبارات كانت لها فائدة سريرية جيدة، إلا أن التقييم السيكومتري لخصائصها كان ضعيفًا إلى حد كبير. في حين أن النسخ الأصلية أو المعدلة من اختبار ارتباط ما بعد الولادة بشكل جماعي حاز على أقوى تقييمات في الخصائص السيكومترية، مع جودة عالية من الأدلة.

5-2-8-دراسة جورجيفا، توماس ونافارو-بيريز Georgieva, Tomas, & Navarro-Pérez(2021) بعنوان مراجعة منهجية وتقييم نقدي لاختبار صدمة الطفولة - نموذج قصير(CTQ-SF) - وقد هدفت هذه الدراسة تقييم نقدي ومقارنة وتلخيص الجودة المنهجية والخصائص السيكومترية للمقالات البحثية المنشورة للتحقق من صلاحية اختبار صدمات الطفولة - نموذج قصير- في السنوات العشر الماضية في قواعد بيانات Scopus و Web of Scienceو ProQuest ، والتي تم تحليلها بشكل مباشر أو غير مباشر للخصائص السيكومترية للاختبار وتم فحصها وتقييمها باستخدام قائمة مراجعة (COSMIN). وقد بلغ إجمالي الدراسات 153 دراسة تم اختيار 10 منها لهذه المراجعة. وأشارت النتائج الرئيسية إلى وجود نمط عام لتقييم نفس الخصائص السيكومترية الثلاثة (الاتساق الداخلي، والصلاحية الهيكلية، واختبار الفرضيات) في مجموعة متنوعة من العينات، حيث أبلغت جميع الدراسات تقريبًا عن معلومات تتعلق بالاتساق الداخلي والصلاحية الهيكلية (90٪ من الدراسات)، وجميعها تقريبًا قامت باختبار الفرضيات (70٪). مع ترك باقي الخصائص التي تم تصنيفها بواسطة قائمة مراجعة (COSMIN) دون تقييم. بالإضافة إلى ذلك، هناك بعض المشاكل في الاتساق الداخلي بالضبط التحليل العاملي.

كذلك هناك فجوة علمية كبيرة في المعلومات المتعلقة ببعض الخصائص السيكومترية. ويُقترح أن تتناول الأبحاث المستقبلية الخصائص السيكومترية المتبقية مثل الثبات، وخطأ القياس، وصدق المحتوى، والصدق عبر الثقافات وصدق المعيار، وكذلك إعادة فحص الاتساق الداخلي لبعض الأبعاد، من أجل التقدم في المعرفة حول تقييم سوء المعاملة في مرحلة الطفولة.

5-2-9-دراسة أونغ، كرافت، ليفين، وتوهيغ Ong, Krafft, Levin, & Twohig (2021) تحت عنوان مراجعة منهجية وتقييم نفسي لتدابير التقرير الذاتي لاضطراب الاكتناز. وهدفت الدراسة إلى تقييم الخصائص السيكومترية لاختبارات التقرير الذاتي لاضطراب الاكتناز، باستخدام معايير التصنيف التي صاغتها المعايير القائمة على التوافق لاختيار مجموعة أدوات القياس الصحية (COSMIN). وتم إدراج 17 أداة لتقييم شدة الاكتناز عند البالغين والأطفال. وكانت النتائج بأن الخصائص سيكومترية بشكل عام ذات نوعية رديئة في جميع المقاييس والنتائج أشارت إلى أداء غير مرضٍ للقياسات. فقد تم تصنيف جودة تطوير PROM بأنها غير كافية بناءً على معايير COSMIN لجميع الاختبارات باستثناء اختبار واحد، والتي تم تصنيفها مشكوك فيها، وفيما يتعلق بالخصائص السيكومترية، كان أداء SI-R الأفضل بين الاختبارات التي تم فحصها رغم أنه أظهر صدق هيكلي غير متسق. كذلك أظهرت النتائج بأن قاعدة الأدلة لثبات وصدق HRS-SR ضعيفة، كما أن الخصائص النفسية الحاسمة الأخرى مثل الاتساق الداخلي والصدق الهيكلي واستجابة العلاج لم يتم تقييمها رسميًا بعد. أما اختبار OCI-HD فقد أظهر صدق بناء كافي في دراسة واحدة فقط. ومن بين مقاييس العملية الأخرى أظهر AAQHفقط صدق بناء كافي. ولم يكن هناك دليل يدعم صدق بناء كافي لـ MOMS أو RSI في الدراسة. أما بالنسبة لاختبار CSI فقد كان لها صدق البناء غير متسق، وبالتالي قد لا يكون مقياسًا مفيدًا لاكتناز عند الأطفال.

5-2-10-دراسة ستالوود وآخرون .Stallwood, et al (2021) بعنوان مراجعة منهجية: للخصائص السيكومترية لمقاييس تصنيف اكتئاب الأطفال تمت مراجعته عند المراهقين المصابين باضطراب اكتئابي كبير. وهدفت الدراسة لتقييم الأدلة الموجودة بشكل منهجي لخصائص القياس لمقياس تصنيف اكتئاب الأطفال – المنقح لدى المراهقين المصابين باضطراب اكتئابي كبير. تم البحث في سبع قواعد بيانات للدراسات التي قيمت خصائص قياس الاكتئاب لدى المراهقين (الذين تتراوح أعمارهم بين 12 و 18 عامًا) من بين 65 دراسة تم فحصها بالنص الكامل، تم تضمين 6. وقد تم تقييم خصائص القياس باستخدام المعايير المستندة إلى إرشادات أدوات القياس الصحية (COSMIN). وأسفرت النتائج على أنه لم تقم أي دراسة بتقييم صلاحية المحتوى أو الصلاحية عبر الثقافات / ثبات القياس أو خطأ قياس الاختبار لدى المراهقين المصابين بالاكتئاب. وتم العثور على أدلة منخفضة الجودة لصحة البناء الكافية (ن = 4 دراسات) والاستجابة (ن = دراستين) التي تم تقييمها من خلال أدوات المقارنة. كذلك أظهرت النتائج أنه تم العثور على أدلة منخفضة الجودة للغاية لثبات المقيمين (ن=دراستين)، وكانت نتائج الصلاحية البنيوية كافية في (ن = 3 دراسات) والاتساق الداخلي (ن = 5 دراسات) غير حاسمة.

5-2-11-دراسة زيلينجر وآخرون .Zeilinger, et al (2022) تحت عنوان أدوات التقييم المستندة إلى المعلومات للخرف لدى الأشخاص ذوي الإعاقة الذهنية: مراجعة منهجية وتقييم موحد. هدفت الدراسة إلى تقييم الأدوات التي تقيم الخرف القائمة على المعلومات. وتم إجراء مراجعة منهجية عبر عشر قواعد بيانات إلكترونية دولية .تم *تطبيق المعايير القائمة على التعداد لاختيار إرشادات أدوات القياس النفسي* (COSMIN). بما في ذلك تقييم مخاطر التحيز، لاستخراج المعلومات وتقييم خصائص القياس وجودة الأدلة المتاحة. وقد تم اعتماد 42 دراسة،

تندرج فيها 18 أداة التقييم القائم على المعلومات. وأسفرت النتائج عن مستوى متراوح في صدق المعيار بين منخفض جدا ومتوسط، أما صدق المحتوى فصنفت أغلب الاختبارات في خانة منخفضة ومنخفضة جدا مع تصنيف اختبار واحد متوسط، كذلك صدق البناء صنف في الثلاث اختبارات التي اعتمدته بمنخفض جدا، في حين أن الاتساق الداخلي تراوح تصنيفه بين منخفض جدا ومعتدل، والأمر نفسه في تقدير الثبات، غير أن الصدق التقاربي والتمايزي كان له نصيب في التصنيف المرتفع في اختبارين وتصنيف منخفض ومعتدل في أربع اختبارات، وأما تقييم الاستجابة فتم التحقق منه في ثلاث اختبارات فقط، ليبقى في الأخير تقييم خطأ القياس الذي لم يتم التحقق منه في أي اختبار.

5-3-التعقيب على الدراسات السابقة

حاولت الباحثة اعتماد وعرض الدراسات الحديثة في هذا الموضوع، وذلك لما يحدث من تجديد وتطوير في أساليب التحقق من صلاحية الاختبارات النفسية، فنجد أن أقدم دراسة تم إدراجها كدراسة سابقة هي دراسة حيدر ظاظا (2011)، كذلك نلاحظ أن أغلبية الدراسات في حدود (13) دراسة نوقشت أو نشرت في الخمس سنوات الأخيرة.

لقد ركز الباحثون في المحور الأول للدراسات السابقة خاصة على ذكر وتعداد أنواع الصدق والثبات المستخدمة في التحقق من صلاحية الاختبارات النفسية في الدراسات الأكاديمية، مثل دراسة لوتزين وآخرون Lotzin, et al. (2015) ودراسة ويتوفسكي، جاريت، كالام، وويسبرغ Wittkowski, Garrett, Calam, & Weisberg (2017)، بدل تركيزهم على تقييم ونقد تلك الإجراءات، فكانت الدراسات تميل إلى الوصف أكثر منها إلى التقييم، ماعدا بعض الدراسات مثل دراسة عبد الرحمان بن صافي (2017) والتي حاول فيها وضع بعض المعايير والاستناد إليها في تقييم إجراءات التحقق من صلاحية الاختبارات النفسية، وقد اتفقت الدراسة الحالية مع هذه الدراسة

من حيث فكرة التقييم، ولكن اختلفت معها من حيث وضع المعايير، فالباحثة هنا فضلت استخدام قائمة المعايير الدولية COSMIN والتي تعتبر أدق وأشمل، خاصة وأن هذه القائمة تمت بالاعتماد على طريقة دلفي بين العديد من الخبراء الدوليين المختصين في القياس النفسي وغيرها من مجالات الرعاية الصحية.

كذلك اختلفت الباحثة مع الدراسات السابقة من حيث العينة، فقد اعتمدت أغلبية الدراسات على تقييم رسائل الماجستير والدكتوراه، مثل دراسة همام السردي (2012) ودراسة بودلال ومووراس (2021). أوتقييم المقالات المنشورة مثل دراسة إلدين، شين، ونغ، وأوريلي El–Den, Chen, Gan, Wong, & O'Reilly (2018) ودراسة زيلينجر وآخرون .Zeilinger, et al (2022)، في حين حاولت الباحثة في هذه الدراسة الجمع بين النمطين السابقين أي توسيع مجتمعها بين رسائل الماجستير والدكتوراه بالإضافة إلى المقالات الأكاديمية، وهو ما اتفقت معه الدراسة الوحيدة من بين جميع الدراسات السابقة التي عملت على إدماج المقالات وهي دراسة بايشي وتشلابي وصالحي (2022).

أيضا اختلفت الدراسة الحالية مع أغلبية الدراسات السابقة فيما يخص العينة أيضا، من حيث اللغة فقد تم إدراج الدراسات باللغتين العربية والأجنبية. في حين أن الدراسات السابقة العربية اختارت العمل على تحليل الدراسات العربية فقط، والدراسات السابقة الأجنبية اختارت تحليل الدراسات الأجنبية فقط، ما عدا دراسة بايشي وتشلابي والصالحي (2022) التي عملت على تقييم دراسات باللغتين العربية والأجنبية. كما هو الحال في هذه الدراسة.

لقد استفادت الباحثة من الدراسات السابقة خاصة المحور الأول في وضع خطة للأساس النظري الذي تنطلق منه، وكذلك اعتمدت على نتائج تلك الدراسات في تحليل ومناقشة نتائج الدراسة الحالية.

اتفقت الدراسات بمجملها على وجود أخطاء في كيفية التحقق من صلاحية الاختبارات النفسية، واختلفت في تركيزها على نوعية هذه الأخطاء فمنها من ركزت على أنواع الصدق والثبات المستخدمة وصلاحيتها لذلك الاستخدام مثل دراسة عبد العزيز بوسالم (د س) ودراسة تيان، كاو، وفانغ Tian, Cao, & Feng, (2019)، ومنها من ركزت على القيم المقبولة في الصدق والثبات حسب معايير خاصة وضعتها كل دراسة، واستندت إلى تلك المعايير في الحكم على قيم الصدق والثبات مثل دراسة حيدر ظاظا (2011) ودراسة بودلال ومووراس (2021)، والجدير بالذكر هنا أن المعايير اختلفت من دراسة إلى أخرى فهناك من مثلا يرى أن نسبة 0.70 كافية لقبول الثبات، مثل دراسة عبد الرحمان بن صافي (2017) ومنها من يرى أن النسبة يجب أن تصل إلى 0.90 مثل دراسة دينيس بوليت Polit, (2015)، وهذا ما دفع بالدراسة الحالية إلى تفضيل الاعتماد على قائمة المعايير الدولية من أجل توحيد المعايير ومن ثم إمكانية المقارنة السليمة بين نتائج الدراسات والاختبارات النفسية.

أما في المحور الثاني للدراسات السابقة والذي يعتبر أقرب إلى الدراسة الحالية، بحيث اتفقت الباحثة مع جميع الدراسات السابقة في هذا المحور، من حيث استنادها على نفس قائمة المعايير، ونفس الشبكة لتحليل وتقييم عينة البحث، إلا أنها اختلفت معها في طريقة استخدام تلك القائمة والشبكة وغيرها، فالباحثة اعتمدت في هذه الدراسة على استخدام القائمة لتحليل عينة من الاختبارات النفسية بأنواعها المستخدمة في الدراسات الأكاديمية، على عكس جميع الدراسات في المحور الثاني فقد عملت على تقييم كيفية التحقق من صلاحية الاختبارات النفسية التي تخص اضطراب واحد في المقالات الأكاديمية المنشورة.

استفادت الدراسة الحالية من الدراسات السابقة خاصة في المحور الثاني؛ فالعمل على تحليل وتقييم الاختبارات النفسية بالاعتماد على قائمة المعايير يعد معقدا إلى حد ما، وبالتالي فقد اعتمدت الباحثة على الدراسات السابقة كثيرا في فهم كيفية عمل الشبكة وتلخيصها وغيرها من النقاط.

إن هدف الدراسة الحالية هو تقييم طرق إعادة التحقق من صلاحية الاختبارات النفسية، وبالتالي فقد اختلفت مع جميع دراسات المحور الأول ماعدا دراسة بوسالم (د س)، واتفقت إلى حد ما مع دراسات المحور الثاني، لأن الدراسات القائمة على معايير COSMIN الدولية دائما ما تشرك الدراسات التي يتم فيها إعادة التحقق من صلاحية الاختبارات النفسية، على عكس الدراسات في المحور الأول فغالبا ما كان يتم التركيز فيها على تقييم طرق التحقق من صلاحية الاختبارات التي تم تطويرها وبناءها أو التي تم تكييفها وتقنينها. وهذا ما دفع الباحثة إلى التركيز على أوجه القصور في الدراسات السابقة ومحاولة تدارك ما تم التغاضي عنه، خاصة في الدراسات العربية، بالإضافة إلى توسيع مجال التقييم في عدة أنواع من الاختبارات النفسية بدل حصر المجال في اضطراب واحد.

وفي الأخير تضيف الدراسة الحالية دليلًا مفيدًا للدراسات العربية، حيث أنها أول من استخدم أداة بحث دولية متفق عليها COSMIN في الدراسات العربية–حسب الاطلاع الموسع الذي قامت به الباحثة في هذا المجال– وذلك من أجل التحقق من جودة الأدلة لمجموعة من خصائص القياس في الاختبارات النفسية. وتعمل الباحثة حاليا بالتعاون مع مبادرة COSMIN على ترجمة قائمة المعايير وشبكة التحليل وكذلك دليل الاستخدام الذي يتكون من عدة أجزاء إلى اللغة العربية، وإطلاق تلك الترجمة في الموقع الرسمي للمبادرة. https://www.cosmin.nl/

الفصل الثاني

صلاحية

الاختبارات النفسية

تمهيد

لا شك أنّه عندما يتعلق الأمر بأدوات القياس في علم النفس، فإنّه تقودنا الأفكار غالبا إلى الاختبارات النفسية، سواء كانت تقوم بقياس الذكاء أو القدرات العقلية الأخرى، أو غيرها من الخصائص النفسية، فالاختبارات النفسية تعتبر من أكثر الأدوات استخداما لجمع البيانات النفسية، فيستخدم الاختبار بهدف التشخيص والتصنيف والتقييم والتنبؤ وغيرها، لهذا أضحت هذه الاختبارات تحتل مكانة مهمة في البحوث والدراسات، ولكن هذا الاحتلال صاحبته بعض التوقعات غير الواقعية وسوء الاستخدام، مما يفرض على المشتغلين في مجال الاختبارات التحقق من صلاحيتها، ومعرفة كيفية استخدامها، وطرق تفسير نتائجها، وهذا ما سنحاول التّطرق إليه في هذا الفصل بشكل من التفصيل من حيث صلاحية الاختبارات النفسية، والدوافع وراء إعادة التحقق منها، والتعرف على الاختبارات النفسية وأنواعها وشروطها والخطوات العلمية المعتمدة لبنائها وغيرها من النقاط.

1-مفهوم صلاحية الاختبارات النفسية

إن مصطلح الصلاحية كما ورد في معجم لسان العرب:" من فعل صلَحَ / صلَحَ لـ يَصلَح ويَصلُح، صَلاحًا وصَلاحِيةً وصُلُوحًا، فهو صالِح ، والمفعول مصلوحٌ له. وعند القول صَلَحَ أَمْرُهُ أَوْ حَالُهُ: صَارَ حَسَناً وَزَالَ عَنْهُ الْفَسَادُ، وصَلَحَ الشيءُ: أي كان نافِعاً أو مُناسِباً." [1]

[1] محمد ابن منظور: *لسان العرب*. ط6. بيروت: دار صادر. 1994، ص 4022

إن مصطلح الصلاحية باللغة الانجليزية (Validity) وبالتالي فإن المصطلح يحمل المعنيين الصدق والصلاحية، وبما أن الاختبار الصالح هو اختبار صادق وثابت، والاختبار الصادق بالضرورة ثابت، فإن الاختبار الصالح هو الاختبار الصادق، أي أن كلمة صلاحية تحمل نفس المعنى لمرادفين مختلفين، فقط مصطلح الصلاحية أعم من الصدق.

وصلاحية الاختبارات النفسية تعتبر مصطلحا واسعا يراد به أحيانا صدق الاختبار وأحيانا أخرى كفاءة الاختبار، في حين أن الباحثة هنا تقصد بمصطلح صلاحية الاختبارات ما هو أبعد من ذلك وأشمل، فالمقصود في هذه الدراسة أن الاختبار الصالح هو الاختبار الصادق والثابت والغير متحيز لذلك الاستخدام ولتلك العينة والظروف وغيرها من المتغيرات، فنقصد هنا بصلاحية الاختبار جاهزيته للاستخدام المرجو منه. فالاختبار الصالح هو الاختبار الموثوق في نتائجه.

وفي الأخير فإن الاختبار يجب أن يرتكز على مستويات مقبولة من الصدق والثبات، باعتبارهما أهم الخصائص السيكومترية لمنحه صفة الصلاحية. وغالبا ما يراد بالصلاحية في هذه الدراسة صدق وثبات الاختبار.

2-دوافع إعادة التحقق من صلاحية الاختبارات النفسية

عادة ما يلجأ معظم الباحثين والمختصين إلى استخدام الاختبارات النفسية الجاهزة بدل بنائها، وذلك توفيرا للجهد والوقت والمال ناهيك عن الصعوبات التي تصاحب عملية بناء الاختبارات النفسية. فيقوم الباحثون بالاستعانة بدل ذلك بالاختبارات النفسية الجاهزة. إلا أن الإشكال الواقع هنا هو استخدام هذه الاختبارات كما هي دون إعادة التحقق من صلاحيتها متجاهلين كل العوامل والمتغيرات التي قد تؤثر فيها، وكأن مستخدم الاختبار جازم بصلاحيته لكل الأزمنة والعينات، وأن الصدق والثبات خصائص مطلقة للاختبار وليست نسبية. وهو ما أكدت عليه دراسة كوسمين

COSMIN (2010) فكما ورد فيها "رغم أن صلاحية الاختبار قد تكون جيدة في المجموعة المستهدفة التي تم تطوير الاختبار من أجلها، إلا أنها قد تكون أقل جودة عند استخدامها مع مجموعة أخرى من المفحوصين، أو مع استخدامات جديدة غير التي صممت من أجلها. ولهذا يتطلب كل استخدام أدلة داعمة جديدة". [1] مما دفع بنا إلى محاولة تحديد أهم العوامل التي تدفع الباحث أو مستخدم الاختبار بصفة عامة إلى إعادة التحقق من صلاحية اختباره قبل استخدامه.

2-1-الغرض من الاختبار

إن اسم الاختبار لا يدل أبدا على الغرض من بنائه كما يتوهم الكثيرون، فاختبار الذكاء مثلا قد يمكننا من تشخيص صعوبات التعلم، ويمدنا بدرجة كمية لقدرة الذكاء لدى الفرد، كما يمكننا استخدامه في التنبؤ والانتقاء وغيرها، لذلك وجب على مستخدم الاختبار أن يتحقق من مطابقة الغرض الذي بني الاختبار من أجله مع غرض دراسته، أي أن بينات صلاحية الاختبار مناسبة للغرض الذي يريد الباحث استخدامه فيه، وبصيغة أخرى فإن على الباحث قبل استخدام أي اختبار التوجه إلى دليله لمعرفة الهدف الذي بني الاختبار من أجله وهل هو نفس هدف بحثه أم لا، وقد أشار تروي وآخرون (Terwee, et al.(2007 [2] إلى ذلك من خلال تأكيده على ضرورة مطابقة الهدف من الاستخدام الحالي للاختبار مع هدف الاختبار والغرض منه. وفي نفس السياق تؤكد كل من كروكر و ألجينا (2017) "بأنه عند اختيار اختبار لغرض معين، على مستخدمه مسؤولية واضحة لضمان أن له أدلة صلاحية مناسبة للاستخدام المقصود في هذا الموقف. وعندما لا يقدم

---

[1] Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., et al:The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement istrument: an international Delphi study. Qual Life Res. 19 (4). 2010, 539-549.

2 Ibid, 35

مؤلف الاختبار أدلة مفصلة عن ذلك، ولكن يعتقد مستخدمه أن الاختبار يمكن أن يكون مفيدا، فعلى هذا الأخير إعادة التحقق لضمان ذلك"[1]. أي أن مستخدم الاختبار لا يكتفي بصلاحية الاختبار المقدم من طرف مطوره عندما يختلف الغرض أو الهدف منه، بل وجب عليه إعادة التحقق من صلاحيته واختيار الإجراءات المناسبة حسب الغرض الذي يريد الباحث استعماله فيه. وهذا ما أكده هوبيت وكرامر Howeet & cramer (2016) بقولهما "أن الصدق والثبات ليست خصائص ذاتية للمقاييس السيكولوجية، فهما سوف يتباينان وفقا لسياق القياس والغرض منه". [2] . وبالتالي فإن الاختبار ليس صالح كما يعتقد البعض لجميع الاستخدامات المرتبطة بالسمة المقاسة، فحسب علام (2000) فإن الاختبار أو الأداة يمكننا استخدامها لأغراض مختلفة ومتعددة حسب المواقف لذلك وجب أن يتم التحقق من صدق الأداة لكل غرض منها. ولكي نحدد مدى ملاءمة الاختبار لاستخدامات معينة ينبغي جمع معلومات مناسبة تتعلق بصدق الاختبار، وتعتمد هذه المعلومات على أهداف عملية القياس وليس على نوع الاختبار أو المقياس.[3]

وفي نفس السياق أوضحت أنستازي وأوربينا Urbina & Anastasi (2015) "بأن اختيار إجراء معين للتحقق من الصدق يعتمد على الاستخدام المرجو من درجات الاختبار. فالاختبار نفسه عندما يستخدم لأغراض مختلفة ينبغي أن يتم التحقق من صدقه بطرق مختلفة".[4] فمثلا إذا

---

[1] ليندة كروكر، و جايمس ألجينا: تأليف مدخل إلى نظرية القياس التقليدية والمعاصرة. ترجمه: هند عبد المجيد الحموري، و زينات يوسف دعنا. دار الفكر ناشرون وموزعون. عمان، الاردن. 2017، ص 304

[2] دينيس هوبيت، و دونكان كرامر: *مقدمة لطرائق البحث في علم النفس*. ترجمه: صلاح الدين محمود علام. دار الفكر. عمّان، الأردن. ط1. 2016، ص 393

[3] صلاح الدين محمود علام: *القياس والتقويم التربوي والنفسي ،أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة: دار الفكر العربي. 2000، ص 190

[4] آن انستازي، و سوزان أوربينا: القياس النفسي. ترجمه: صلاح الدين محمود علام. دار الفكر العربي ناشرون وموزعون. عمان، الأردن. ط1. 2015.

استخدم مثلا اختبار تحصيلي معين للتنبؤ بالأداء التالي على مستوى التعليم العالي كما الحال عند انتقاء طلبة المدارس الثانوية للقبول بالجامعة فإنه يحتاج إلى تقييمه في ضوء محك الأداء التالي في الجامعة وليس من حيث صدق محتواه.

لذلك، بناءً على تنوع استخدامات الاختبار النفسي والنتائج المستخلصة منه، فإن له عدة أدلة وإثباتات على صلاحيته، وعليه فإن الاختبار الذي يتمتع بدرجة مقبولة من الصلاحية لغرض ما قد لا يكون هو الحال بالنسبة للآخر. نتيجة لذلك، يرى فرج (2017)[1] أنه سواء كان ذلك لأغراض البحث أو لأغراض عملية أو تطبيقية، يجب أولاً الاعتماد على منهجية محددة للتحقق من صلاحية الاختبار، ثم الاعتماد على النتائج التي تم الحصول عليها من الاختبار.

2-2-مجتمع وعينة الاختبار

عند تصميم أي اختبار نفسي يعمد ناشره ومصممه على وصف العينة التي وضع الاختبار من أجلها وتحديد المجتمع الذي سحبت منه، وهذا ليس عبثا وإنما ليعرف مستخدم الاختبار الحدود المسموح بها عند استخدامه لذلك الاختبار، فلا يمكنه استخدام اختبار مصمم للبالغين وتطبيقه على الأطفال، فذلك من شأنه أن يقدم نتائج مغلوطة وتفسيرات خاطئة لها، وبالتالي يعتبر الاختبار هنا غير صالح لهذا الاستخدام، حتى وان كان ناشر الاختبار قد قدم أدلة على صلاحية اختباره، وبالتالي بمجرد تغيير المجموعة تتغير معه خصائص الاختبار فلا يمكننا الجزم بصدقه أو صلاحيته إلا بالتحقق من ذلك، وهذا ما يدعوا إلى ضرورة تحديد العينة التي استخدمت عند استخراج معامل الصدق والثبات وغيرها، فحسب أنستازي وأوربينا Urbina & Anastasi (2015) ربما يقيس نفس الاختبار أغراض مختلفة عندما يُطبق على أفراد مختلفين في السن أو

[1] صفوت فرج: القياس النفسي. ط1. مصر: مكتبة الأنجلو المصرية. 2017، ص 237

مستوى الدراسي أو الوظيفة وبالتالي يمكن أن يتميز الاختبار بصلاحية مرتفعة في عينة ما، وصلاحية منخفضة أو منعدمة في عينة مختلفة.[1] والاختبارات التي تصمم للاستخدام في عدة مجتمعات ينبغي أن تشير في دليلها إلى ذلك والتفصيل في المجتمعات التي يناسبها هذا الاختبار. وعليه يتأثر صدق الاختبار بالخصائص النوعية للعينة كما تتدخل الفروق الحضارية بين مجتمع وآخر وما تؤدي له من اختلاف في دلالات السلوك إلى التأثير في صدق الاختبار.[2] وبالتالي التأثير على صلاحية الاختبار ومصداقيته.

وللإشارة هنا فإن التغيير الكلي للمجتمع الذي طور الاختبار من أجله، يستدعي أكثر من إعادة التحقق من صلاحية الاختبار، مثل استخدام اختبار مخصص للأطفال العاديين مع الأطفال من ذوي التأخر العقلي، فالإجراء هنا يستلزم التكييف أو التقنين وغيرها. في حين أن الاحتفاظ بنفس الميزات العامة للمجتمع مع تغيير بعض الخصائص المرتبطة بالعينة، مثل استخدام اختبار طُور لأطفال متلازمة داون مع أطفال التأخر العقلي، فإنه وجب علينا هنا إعادة التحقق من صلاحية الاختبار النفسي لهذه العينة مع التأكيد على ضرورة استخدام صدق المحتوى. وأبعد من ذلك ما ذهب إليه كل من جاي، ميلز وايرسيان Airasian & Mills & Gay (2012) فالصدق والثبات حسبهم مرتبط بالتفسير المطلوب والعينة المدروسة. وبتعبير آخر: لا يفترض أن نقول "الاختبار صالح بصفة مطلقة"، وإنما ينبغي أن نقول: "هذا الاختبار صالح لهذا التفسير، وهذه

---

[1] انستازي و أوربينا، المصدر المذكور سابقا، ص 189.

[2] فرج، المصدر المذكور سابقا، 265

العينة".[1] وبالتالي أصباح على الباحث لزاما إعادة التحقق من صلاحية الاختبار النفسي في حالة تغييره لمجتمع الدراسة.

2-3-عامل الزمن

قد يختلف صدق وثبات الاختبار أيضاً بين فترة زمنية وأخرى، فنتيجة للتقدم في التعليم والتطور الاجتماعي والتكنولوجي، تتغير المؤشرات الدالة على سمة ما، لأنها تمثل السلوك المقاس فعلا، والذي بدوره يتأثر بالتغيرات الثقافية والاجتماعية والتكنولوجيا، وقد ذكر فرج (2017)[2] بأن "البنود التي كانت صادقة في قياس الذكاء في عشرينيات وثلاثينات القرن الماضي لا تصلح الآن ولا يمكن استمرار وجودها في مقياس صادق للذكاء، في القرن الواحد والعشرين". وعلى هذا الأساس يُتطلب التحقق من صلاحية الاختبار قبل استخدامه وفحص مضمونه فحصاً متأنياً وذلك للاعتبارات الحضارية التي تضفي على بنوده دلالة معينة، كما يتعين أن يضع الباحث في اعتباره عند استخدام اختبار ما أن الأمر يتطلب مراجعة صدقه وثباته من حيث تقدير تأثير المتغيرات الزمنية، التي تؤثر بشكل مباشر في المفاهيم التي يتضمنها الاختبار.

ولهذا يوصي تروي Terwee (2007) بضرورة تحديد الفترة الزمنية التي تشير إليها الأسئلة بوضوح وتبريرها. وذكر مدى تأثر الاختبار بعامل الزمن. [3]

وذلك لكون الاختبار الصالح في فترة معينة قد لا يكون صالحا في الفترات التالية. وقد أشار بن صافي (2017) في دراسته إلى وجوب إعادة التحقق من صلاحية الاختبار بعد مرور 10

---

[1] ل جاي، ج ميلز، و ب ايرسيان: البحث التربوي كفايات للتحليل والتطبيق. ترجمه: صلاح الدين محمود علام. دار الفكر ناشرون وموزعون.الأردن.2012، ص 262

[2] صفوت فرج، المصدر المذكور سابقا، ص 238

[3] Terwee, 40

سنوات على تطويره.[1] إلا أن الباحثة ترى أن تحديد المدة الزمنية هو أمر خاطئ، لأن عامل الزمن هنا هو عامل نسبي فهنالك سمات تتميز بالاستقرار إلى حد ما وبالتالي البنود التي تقيسها تتميز هي الأخرى بنوع من الاستقرار، في حين توجد بعض السمات أو بالضبط البنود التي تقيسها يعرف عنها أنها غير مستقرة نسبيا. فمثلا في الفترة الأخيرة وبالضبط بعد تعرض العالم لجائحة الكوفيد 19، تغيرت العديد من السلوكيات وخاصة المرتبطة بالعلاقات الاجتماعية، فنرى أن العديد من الأداءات اختفت لتحل محلها سلوكات مغايرة لها تماما، فأصبحت العزلة كمثال مؤشرا على الوعي بعد أن كانت مؤشرا على العديد من الاضطرابات مثل الاكتئاب وغيره، فلذلك تحديد الفترة الزمنية حسب الباحثة هو أمر مرتبط بالسمة المدروسة والبنود التي تقيسها، وبالتالي فمستخدم الاختبار هو من يقرر إذا كان هذا الأخير قد تأثر بعامل الزمن أم لا. وذلك بمساعدة المختصين في المجال المدروس وفي القياس، وصدق المحتوى و هنا على الباحث التفكير في الانطلاق منها.

2-4-عامل الثقافة

يبدو عامل الثقافة على درجة كبيرة من الأهمية عند نقل أو ترجمة اختبار من حضارة ذات خصائص معينة إلى حضارة أخرى تتسم بخصائص نوعية واضحة تنعكس على القيم وأساليب السلوك السائدة ودلالتها الاجتماعية. [2]

لذلك وجب التفطن لهذا العامل عند استخدام الاختبارات النفسية التي تم ترجمتها وتكييفها في دولة عربية ما، ويتم تداولها على أنها صالحة لكل الدول العربية، وكأنه لا توجد حدود ثقافية بين الدول

---

[1]عبد الرحمان بن صافي: دراسة تقييمية لطرق تقدير صدق وثبات أداوت جمع المعطيات . أطروحة دكتوراه. كلية العلوم الانسانية والاجتماعية: جامعة مستغانم. 2017، ص 130.

[2] صفوت فرج، المرجع المذكور سابقا، ص 238

العربية وهو بالطبع أمر خاطئ رغم أنه شائع، بل أبعد من ذلك فقد نبهت دراسة COSMIN (2018) إلى أن التحقق من الصلاحية عبر الثقافات للاختبارات ليست فقط للمجتمعات المختلفة ثقافيًا، مثل المختلفة دينيا أو لغويا والتي تحتاج إلى تكييف، ولكن أيضًا للمجموعات الأخرى مثل الجنس أو الفئات العمرية أو العينات المختلفة. إذ يجب أن يتم تقييم الصدق الثقافي معها، وذلك من خلال تقييم ما إذا كان المقياس ثابتًا في القياس أو ما إذا كانت وظيفة العنصر التفاضلي- Functioning *Differential* Item ويرمز له اختصارا (DIF) – تحدث أم لا.[1] ويؤكد كل من هامبلتون، ميرندا، وسبيلبيرغر Hambleton, Merenda & Spielberger (2006) أنه في حالة تم نقل الاختبار ثقافيا فإنه يتطلب وجود إثبات على أن أعضاء المجموعتين لهما الكفاءة الواحدة باستخدام العنصر التفاضلي، وعندما لا يكون الأداء متساويا يجب أن يكون هناك سبب وجيه أو يلغى البند من الاختبار.[2]

وبالتالي فإن الاختبار الذي يتم نقله من لغة مختلفة وثقافة مختلفة تماما، يحتاج إلى أكثر من إعادة التحقق، فهو يحتاج إلى تكييفه في البيئة الجديدة، في حين يتم الاقتصار على إعادة التحقق من صدق وثبات الاختبار إذا شكك الباحث بأن هناك عامل ثقافي ما قد يؤدي إلى تحيز بنوده، حتى لو تم اختيار الاختبار من نفس البيئة التي سيستخدم فيها.

2-5-الاختبار المعدَل

يعتبر الاختبار المعدل هو الاختبار الذي تم تغيير بعض أجزائه، ونقصد بذلك تغيير عدد

---

[1] Mokkink, L., Vet, H., Prinsen, C., Patrick, D., Alonso, J., & Bouter, L: COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res , 27* (5). 2018, p 58.

[2] هامبلتون، ميريندا، و سبيلبيرغر: *تكييف الاختبارات التربوية والنفسية للتقييم عبر الثقافات*. ترجمه: هالة برمدا. مكتبة العبيكان. الرياض. ط1. 2006، ص 56.

البنود أو خيارات الاستجابة أو عدد الأبعاد وغيرها من أشكال التغيير التي نستطيع إدخالها على الاختبار، ومن شأنها أن تؤثر على بنية الاختبار. وبالتالي هذا يدفعنا حسب Mokkink, et al., (2018)[1] إلى التعامل مع الاختبار المعدل باعتباره اختبارًا جديدًا. ولذلك وجب إعادة التحقق من صلاحيته، لكن ليس بمعزل عن المعلومات المتعلقة بتطوير الإصدار الأصلي للاختبار وكذلك دراسات الصلاحية الإضافية يمكن أن تكون مفيدة." فإذا كان الاختبار المعدل عبارة عن نسخة مختصرة من اختبار تم تطويره مسبقًا، فيمكن الحصول على معلومات حول ملاءمة العناصر وإمكانية فهمها من الدراسات في الاختبار الأصلي. ومع ذلك، يجب أن يتم التحقق حول شمولية الاختبار المختصر من دراسة جديدة يقوم بها مستخدم الاختبار المختصر. ويجب على الباحثين تحديد المعلومات التي يمكن استخدامها من الاختبار الأصلي والمعلومات يجب أن يضيفوها باستخدام الاختبار المعدل. فعند قيام مستخدم الاختبار بتغيير أساسي في بنية الاختبار أو طريقة الاستخدام أو التعليمات، أو اللغة أو المحتوى، يجب عليه حسب هامبلتون، ميريندا، و سبيلبيرغر، (2006) [2] إعادة التحقق من صلاحية الاختبار حسب التغييرات التي أحدثها مع عرض أسباب منطقية تدعم مصداقية الاختبار. كذلك عندما تعديل الاختبار من لهجة إلى أخرى لاستخدامها لاختبار مجموعات ذات لغة واحدة يجب أيضا التحقق من مصداقيتها وصلاحيتها.

يجب التعامل مع الاختبار المعدل باعتباره اختبارًا جديدًا. ومع ذلك إذا كان الاختبار المعدل عبارة عن نسخة مختصرة من اختبار تم تطويره مسبقًا.

---

[1] Ibid, p 1173

[2] هامبلتون، ميريندا، و سبيلبيرغر، المرجع المذكور سابقا، ص 23

فيمكن حسب Dobson, Choi, Hall, & Hinman (2012)[1] الحصول على معلومات حول ملاءمة العناصر وإمكانية فهمها من الدراسات في الاختبار الأصلي. في حين يجب أن تأتي المعلومات حول شمولية الاختبار المختصر من خلال إعادة التحقق من صدق محتوى الاختبار باستخدام النسخة المختصرة. ووجب التنبيه إلى ضرورة الحصول على إذن مسبق من مطوَر الاختبار قبل استخدامه فما بالك بتعديله.

2-6-عامل استمرارية الصلاحية

إن صلاحية الاختبارات هي عملية مستمرة ومتجددة، فحسب رينولدز و ليفينجستون(2013)[2] تُعد الصلاحية عملية تتضمن جهدا متواصلاً وديناميكيا، لتجميع أدلة لأساس علمي منطقي يستند إليه في التفسيرات المقترحة لدرجات الاختبار. لذلك يتواصل التحقق من الصلاحية في الأدبيات العلمية. كذلك لا ينبغي أن يتقيّد من يستخدم الاختبارات بالقيود الضرورية التي تفرض على المعلومات في دليل الاختبار، بل ينبغي الاطلاع على الأدبيات المتراكمة عن العملية المستمرة المتعلقة بالتحقق من الصلاحية. وهذا ما أيدته وثيقة المعايير المشتركة AREA, APA, & NCM (2014)[3] حيث ذكر فيها أن "التحقق من الصلاحية مسؤولية مشتركة بين مؤلف الاختبار

---

[1] Dobson, F., Choi, Y., Hall, M., & Hinman, R: Clinimetric properties of observer-assessed impairment tests used to evaluate hip and groin impairments: a systematic review. *Arthritis Care Res (Hoboken) , 64* (10).2012, p1565-1575.

[2] رونالد ليفينجستون، و سيسيل رينولدز: إتقان القياس النفسي الحديث النظريات والطرق. ترجمه: صلاح الدين محمود علام. دار الفكر ناشرون وموزعون. عمان، الاردن. 2013، ص 217

[3] AREA, A. E., APA, A. P., & NCME, N. C: *Standards for Educational and Psychological Testing.* Washington: American Educational Research Association. 2014, p12

ومستخدمه". وبالتالي حتى لو قدم مؤلفو الاختبارات أدلة عن صلاحية اختباراتهم إلا أنه يتوقع من الباحثين تقديم أدلة صلاحية جديدة، أي أنه يتوقع من مستخدمي الاختبارات أن يقيموا أدلة الصلاحية ويصدروا أحكامهم الشخصية حول ملائمة الاختبار لدراستهم أو استخداماتهم، أي إصدار الحكم النهائي لمناسبة هذا الاختبار. وهو أكثر الأدوار مسؤولية في هذه العملية.

3-أهم مهددات صلاحية الاختبارات النفسية

لاحظت الباحثة عند تقييم الدراسات في الجانب التطبيقي، أن هنالك العديد من المفاهيم التي كانت غائبة أو تم استخدامها بشكل خاطئ مما أدى إلى المساس بكينونة الاختبارات النفسية وتهديد صلاحيتها في كثير من الأحيان، ومن أهم هذه المهددات، هو الخلط الواضح في المفاهيم خاصة بين التكييف والترجمة وإعادة التحقق، فيكتفي بعض من الباحثين بالترجمة على أساس أن الاختبار قد أصبح جاهزا ولا يحتاج إلى باقي الإجراءات ليصبح مكيفا لتلك البيئة فعلا، بل أكثر من ذلك أن الترجمة في حد ذاتها لا تكون وفق الخطوات الواجب إتباعها والتي أوصت بها اللجنة الدولية للاختبارات- (International Test Commission) والتي يرمز لها اختصارا (ITC)-، مثل الترجمة والترجمة الراجعة وغيرها، كما لاحظت الباحثة تعديل الباحثين للاختبار بحذف بعض من البنود مثلا، ثم التحقق من صدق وثبات الاختبار كإجراء إحصائي محض في أحسن الأحوال، واعتبار الاختبار في الأخير بأنه أصبح جاهزا. هذا ما دفع بالباحثة إلى محاولة ضبط المصطلحات المرتبطة بصلاحية الاختبارات النفسية والفصل فيها وتوضيح البعض من الإجراءات المتعلقة بها. كذلك التطرق إلى عنصر مهم يعتبر من أهم مهددات صلاحية الاختبارات النفسية والتي يغفل بعض الباحثين -حسب عينة الدراسة- من التحقق منه ألا وهو دراسة الأداء التفاضلي للبند ومن ثم الجزم بخلو الاختبار من أي تحيز في بنوده وبالتالي التأكيد على صلاحية هذا

الاختبار للاستخدام المرجو. فمثلا لم يتم استخدامه في أي دراسة من عينة الدراسة الحالية رغم أن بعض الدراسات قامت بترجمة الاختبارات، كما قامت دراسات أخرى بتغيير مجتمع الدراسة وغيرها من العوامل التي كانت تستدعي استخدامه.

3-1-الترجمة دون التكييف

بداية يجب التمييز بين تكييف الاختبارات وترجمتها، فمصطلح "تكييف الاختبارات" أوسع من مصطلح "ترجمة الاختبارات" وأكثر انعكاسا على ما يجب القيام بع في الواقع، فيتم تكييف الاختبارات النفسية في أحد الظروف التالية: عند الاستخدام في بلدان غير تلك التي جرى تطويرها فيها، للاستخدام في البلدان التي جرى تطويرها فيها مع تكييفها لتستخدم مع أشخاص يختلفون في اللغة، الثقافة، أو صفات مهمة أخرى، كذلك يتم تكييفها للاستخدام في بلدين أو أكثر حيث تجري تطبيقات على قوميات متعددة. أما الترجمة هي عند استخدام اختبار تم إعداده للاستخدام في لغة وثقافة واحدة وتطبيقه في لغة وثقافة أخرى. هامبلتون، ميريندا، و سبيلبيرغر(2006)[1] بينما يتضمن تكييف الاختبار كل الأنشطة بدءا من تقرير عما إذا كان باستطاعة الاختبار الاحتفاظ بتركيبته ذاتها في لغة وثقافة أخرى، واختيار المترجمين، وتقرير التكييف المناسب الذي يجب القيام به لإعداد الاختبار للاستعمال في لغة ثانية، إلى تكييف الاختبار والتأكد من تطابقه مع الشكل المكيف، وبالتالي فإن ترجمة الاختبار هي خطوة واحدة من عملية تكييف الاختبار. [2]

وحتى في تلك الحالة مصطلح التكييف مناسب أكثر من مصطلح الترجمة لوصف العملية الحقيقية التي تجرى، ذلك لأن المترجمين يحاولون الحصول على مفاهيم، مفردات وتعابير متعادلة

---

[1] هامبلتون، ميريندا، و سبيلبيرغر، المرجع المذكور سابقا، ص 22

[2] Ibid, p22

ثقافيا، نفسيا ولغويا للغة والثقافة الأخرى، بذلك تأخذ المهمة أبعادا أكثر من ترجمة محتويات الاختبار حرفيا. [1]

ولضمان صلاحية الاختبارات النفسية وسلامة إجراءات الترجمة والتكييف يتم التركيز على الاختيار الصحيح للمترجم والذي يعرف هنا بـ "الخبير" وهو شخص أو فريق لديه معرفة مشتركة كافية بـ: اللغات المعنية، بالثقافات المعنية، بمحتوى الاختبار المعني، وبالمبادئ العامة للاختبار، والهدف من ذلك هو إنتاج ترجمة وتكييف ذا جودة عالية. كذلك يمكن الاعتماد على عدة نماذج للترجمة وتعتبر الترجمة الأكثر استخداما لتحقيق التكافؤ اللغوي الوظيفي، بدلاً من التكافؤ الحرفي، هي الترجمات أحادية الاتجاه والترجمات ثنائية الاتجاه. [2]

وأرادت الباحثة هنا أي بهذا العنصر؛ توجيه انتباه الباحثين إلى ضرورة الاعتماد على الخبير عند ترجمة الاختبارات وأنها ليست أمرا يعتمد فيه الباحث على نفسه، إلا في حالة توفر السمات المطلوبة في المترجم في شخص الباحث، كذلك التنويه إلى ضرورة الاعتماد على النماذج التي أطلقتها أو فسرتها اللجنة الدولية للاختبارات (ITC) وغيرها من التفاصيل الواجب الاهتمام بها عند ترجمة الاختبارات، فهي ليست عملية تقنية بحتة وإنما تستند إلى خطوات وشروط عملية فصلت فيها اللجنة الدولية للاختبارات (ITC). كذلك هنالك نقطة مهمة وجب توضيحها هنا وهي الخلط الذي التمسناه بين الترجمة والتكييف في العديد من البحوث، والأسوء من ذلك الاعتماد على الاختبارات النفسية بعد ترجمتها ترجمة حرفية فقط، ومن ثم إعادة التحقق من صدقها وثباتها في أحسن الأحوال وكأنها بذلك أصبحت جاهزة للاستخدام، متغاضيين عن ضرورة التكييف من أجل

---

[1] هامبلتون، ميريندا، و سبيلبيرغر، المرجع المذكور سابقا، ص 22

[2] International Test Commission: The ITC Guidelines For Translating and Adapting Test. N. E. Boudouda, A. A. Bensouilah, M. E. Seridi, M. Elafri, S. Ben Youssef, & K. Gana, Trans. 2ed. 2017, p 13

الصلاحية الفعلية لتلك الاختبارات. وبالتالي تؤكد الباحثة هنا على عدم الاكتفاء بالترجمة في أي حال من الأحوال بل وجب دائما تكييف الاختبارات الأجنبية، والاهتمام بشروط الترجمة كخطوة من خطوات التكييف، وأن إعادة التحقق من صدق وثبات الاختبار بعد ترجمته لا يدل على صلاحيته فعلا للاستخدام حتى وإن كانت الأرقام تشير إلى ذلك. لأنه وجب دعمه بأدلة تظهر عدم تحيزه مثل الأداء التفاضلي للبند والذي سيتم تناوله تاليا.

3-2-الأداء التفاضلي للبند:

عادةً ما يتم التحقق من صلاحية الاختبارات النفسية عن طريق الصدق والثبات، وهذا بالطبع ضروري ولكنه غير كاف لدعم الجودة السيكومترية للاختبار، وإمكانية إجراء استنتاجات صحيحة من الدرجات التي تحصل عليها الأفراد، وهو هدف أي تقييم غالبا، ولذلك يجب أن يخضع أي اختبار يتم استخدامه لهذا الغرض، لإجراء تقييم مناسب للبنية التي يتم قياسها، من أجل الوصول إلى الجزم بأن التباين الحاصل في نتائج هذه الاختبارات راجع للفرق في امتلاك السمة وليس لأمر أخر. ولذلك فإن أحد الجوانب المهمة للتحقق من صلاحية الاختبار هو فحص مصادر التحيز، وقد أكد Messick (1995)[1] هذه الفكرة بقوله "أن أحد العوامل التي قد تهدد صلاحية الاختبار هو التباين غير ذي الصلة بالبناء، والذي يشير إلى مدى تأثر درجات الاختبار بالمتغيرات التي تكون غريبة عن البنية". ولذلك يعد الأداء التفاضلي للعنصر (DIF) مصدرًا مهمًا للمعلومات التي

[1] Messick, S.: Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning.50. American Psychologist.1995,p 741-749.

تؤثر على جميع مصادر أدلة الصلاحية المتفق عليها في معايير الاختبارات التربوية والنفسية. في هذا الصدد جادل كل من Gomez-Benito, Sireci, Padilla, Hidalgo, & Benitez, (2018).[1] بأن دراسة الأداء التفاضلي هي بحكم تعريفها دراسة للصدق لأنها يمكن أن توفر معلومات حول جميع المصادر الخمسة لأدلة الصدق. على سبيل المثال: ستحلل دراسة الأداء التفاضلي ما إذا كان مرتبطًا بعوامل ذات صلة أو غير ذات صلة بالبناء، أو ما إذا كانت هناك عواقب سلبية مرتبطة بوجوده وما إلى ذلك.

ولهذا نجد أن مبادرة COSMIN قد أدرجته في أدلة التحقق من الصدق بمسمى الصدق عبر الثقافات أو الصدق الثقافي، تحت نوع صدق الفرضيات (أنظر الملحق رقم:04) وهو ما تبنته الباحثة في هذه الدراسة، من حيث اعتبار الأداء التفاضلي دليل من أدلة صدق الاختبار.

وقبل الغوص أكثر في هذا الموضوع يجب توضيح معنى التحيز بطريقة مبسطة كما عرفته مبادرة كوكرين Cochrane (2009): "فالتحيز هو خطأ منهجي أو انحراف عن الحقيقة، سواء في النتائج أو في الاستنتاجات. يمكن أن تعمل التحيزات في أي اتجاه، مما يؤدي إلى المبالغة في تقدير دقة الاختبار أو التقليل من شأنها. من المستحيل أن نعرف على ما إذا كانت الدراسة متحيزة، أو اتجاه أو حجم التحيز. ومع ذلك، عند تحديد المصدر يمكن إصدار أحكام بشأن خطر التحيز.[2]

---

[1] Gomez-Benito, J., Sireci, S., Padilla, J., Hidalgo, I., & Benitez, I: Differential Item Functioning. Beyond validity evidence based on internal structure. Psicothema.30 (1) 2018, p 104-109.

[2] Reitsma, J., Rutjes, A., Whiting, P., Vlassov, W., Leeflang, M., & Deeks, J :Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. (2009) Retrieved from The Cochrane Collaboration: http://srdta.cochrane.org/.

ومن أجل توضيح المعنى أكثر فإن التحيزات يمكن أن تنشأ من خلال مشاكل في تصميم أو تنفيذ الدراسة أو من خلال تجنيد المشاركين الخطأ وغيرها. والمقصود بالتحيز كما فسره كل من Ary, Jacobs, Sorensen, & Walker (2014) [1]أنه إذا حدث انخفاض أو تضخم في درجات الأفراد لمجموعة من المتقدمين للاختبار؛ بسبب أنه يشتمل على بعض العوامل الدخيلة، فإن صحة الاستنتاجات التي تم إجراؤها حول تلك المجموعة مشكوك فيها بسبب تباين غير ذي صلة بالبناء والذي يشار إليه باسم تحيز الاختبار. لذلك من أجل تجنب مثل هذه الآثار السيئة، يجب بذل الجهود لتحديد مصادر الخطأ والتحيز بحيث يمكن الحصول على أقصى قدر من العدالة والإنصاف لجميع المتقدمين.

وقد تم اقتراح طرق إحصائية مختلفة لتحديد وإزالة المصادر المحتملة للتحيز في القياس، وتقدم نظرية استجابة للفقرة، طريقة قوية لاستكشاف تحيز العنصر، والذي يشار إليه أيضًا باسم الأداء التفاضلي للعنصر (DIF)، والذي قدمه Holland & Thayer (1986)[2] بعدما كان يطلق عليه تحيز الفقرة، ويعتبر تحليل الأداء التفاضلي حاليا أحد أكثر الطرق شيوعًا للكشف عن التحيز في عناصر الاختبار. وتتيح هذه الطريقة تحديد ما إذا كانت العناصر تعمل لصالح مجموعة فرعية من المتقدمين للاختبار، وتساعد في تحديد عناصر الاختبار التي تعمل بشكل مختلف لمجموعتين من المتقدمين للاختبار. قد يؤدي وجود الأداء التفاضلي إلى نتائج اختبار غير

---

[1] Ary, D., Jacobs, L., Sorensen, C., & Walker, D: Introduction to research in education . 9 éd. Belmont: Wadsworth.2014, p243

[2] Holland, P., & Thayer, D: Differential item performance and the Mantel-Haenszel procedure. New Jersey: Educational Testing Service. 1986, p 1

صالحة، مما يمنع المقارنات بين السكان وبالتالي يثبت ذلك أن الأداة لا تعتبر مقياسًا مناسبًا للبنية المستهدفة. [1]

وتشير مؤشرات الأداء التفاضلي الكبيرة وذات الدلالة الإحصائية إلى أن عناصر الاختبار لا تعكس فقط متغير السمات الكامنة المستهدفة؛ بل بدلاً من ذلك تقيس تركيبات أخرى إلى جانب تلك التي تم تصميمها من أجلها، وتظهر وجود تباين غير ذي صلة بالبناء ناتج عن بُعد ثانوي غير مقصود، الذي يتسبب في أداء العناصر بشكل مختلف من مجموعة إلى أخرى. [2]

يمكن للباحث تحليل التحيز من خلال فحص الأسباب التي تجعل العنصر يؤدي إلى استجابات مختلفة في الممتحنين أو المجموعات المختلفة. يتم ذلك من خلال مراعاة الغرض من الاختبار، وتجربة المستجيبين، والمعلومات الإحصائية المتعلقة بتحليل العناصر أو فئات الاستجابة.[3]

---

[1] Adkins, j: An Examination of Bias in Oral Reading Fluency: Differential Effects across Race, Gender, and Socioeconomic Status. Arizona State University Unpublished doctoral dissertion.2013, p22

[2] Gierl, M: Using a multidimensionality-based DIF analysis to identify and interpret constructs that elicit group differences. Educational Measurement Issues and Practice , 24 (1). 2005, p3-14.

[3] Holland, P., & Wainer, H: Differential item functioning, Lawrence Erlbaum. *Associates Inc* .1993.

وخلال أكثر من 50 عامًا من البحث في تحليل الأداء التفاضلي، تم اقتراح العديد من الطرق لاكتشافه. فقد سعت الطرق المبكرة إلى تحديد العناصر المتحيزة عن طريق تحليل التباين.[1] ولكن سرعان ما حلت محلها تقنيات أخرى أكثر دقة أو قوة. فقد صنف Camili & Shepard (1994) [2] طرق الكشف عن الأداء التفاضلي في ثلاث فئات: تلك القائمة على تحليل التباين ونظرية الاختبار الكلاسيكية، وتلك القائمة على نظرية الاستجابة للعنصر(IRT)، وتلك القائمة على تحليل جداول الاحتمالات.

وعلى العموم فإن الأداء التفاضلي للبند يعتبر ضروري للدلالة على تحيز البند، حسب بوسالم وكريش (2016) ولكنه غير كاف للجزم بذلك، فعند القيام بالتحليل ولم تحتوي أي من البنود على الأداء التفاضلي للبند دل هذا على عدم وجود تحيز البند، أما في حالة احتواء أحد البنود على الأداء التفاضلي للبند فهذا يعتبر غير كافي للحكم على البند بأنه متحيز، بل على الباحث مواصلة التحليل، بالقيام بتحليل المحتوى والتقييم التجريبي، لتحديد إذا ما كان حقيقة البند متحيز، والاستعانة بالمتخصصين (الخبراء) في ذلك. [3]

في الأخير فإن الأداء التفاضلي للبند، يعتبر خطوة مهمة وضرورية للتحقق من صلاحية الاختبار، شأنه شأن الصدق والثبات، سواء عند بناء الاختبار أو استخدامه جاهزا، فإن هذا الإجراء لا يقل أهمية عنهما للدلالة على صلاحية الاختبار، بل ويعتبر إجراء ضروري من إجراءات

---

[1] Cleary, T., & Hilton, T: An investigation of item bias. *Educational and Psychological Measurement , 28*. 1968, p 61-75.

[2] Camili, G., & Shepard, L: Methods of identifying biased test item. *Sage Publications* . 1994.

[3] العزيز بوسالم، و أحمد كريش: الأداء التفاضلي للبند في الاختبارات النفسية، مصادره وتفسيره. *مجلة العلوم* الاجتماعية .2016، ص 127-142.

الصدق، وهو ما يغفل عنه الكثير من الباحثين فكما هو ملاحظ في الجانب التطبيقي، لم يتم الكشف عن الأداء التفاضلي لإثبات صلاحية الاختبار للاستخدام المراد، بل يتم فقط التحقق من الصدق والثبات في أحسن الأحوال. وبالتالي أرادت الباحثة الإشارة إلى هذا الخطأ الشائع في هذا العنصر الذي يعتبر من مهددات صلاحية الاختبارات النفسية.

4-مفهوم الاختبارات النفسية

لقد تعددت تعاريف الاختبارات النفسية وتنوعت، ولكن في الحقيقة تشير أغلبها إلى معنى متقارب إلى حد ما، فيتم تعريفها غالبا على أنها الإجراء التقييمي، الذي يتم من خلاله الحصول على عينة من سلوك المفحوص في مجال معين، وتقييمها لاحقا وتسجيلها باستخدام عملية موحدة.

وقد عرّف أدوقا Adoga (2020) الاختبارات النفسية: "بأنّها الأداة التّي يتم من خلالها إجراء المثير النفسي، ويوصف الاختبار النفسي بأنّه الأداة أو الحافز الذي يطبق على الفرد، ومن ثم تسجيل الردود وإبلاغها إلى ذلك الفرد.[1] وفي نفس السياق عرّفتها كامبل Campbell وآخرون (2020) بأنّها "أداة قياس مستمدة من الناحية النفسية، وتقيّم البنى النفسية، في عينة منظمة من سلوك المفحوص في موقف معين".[2] فيتضح مما سبق بأنّ الاختبارات النفسية أداة تتميز باستخدام عينات من السلوك، من أجل تقييم التركيبات النفسية لأداء فرد ما، مثل الذكاء أو الشخصية أو أي جانب آخر من الجوانب المعرفية والعاطفية وغيرها.

---

[1] Adoga, R. : psychological test in counselling. prestige journal of conselling psychology. 2020 , p. 248.

[2] Campbell, L. F., Stanford, L. D., Alfonso, V. C., Cuttler, M. J., Demers, S. T., Hass, G. A., et al :APA Guidelines for psychological Assessment and Evaluation. *American Psychological Association* . 2020, p 1-27.

وأضاف رحيم (2008) بأن الاختبارات النفسية "عبارة عن مجموعة مقننة من المنبهات التي تمثل عينة من السلوك المراد قياسه وتقيس قدرة أو سمة شخصية معينة لدى الفرد، ويعبر عن النتائج في صورة كمية أو كيفية أو كليهما معا". [1] وهو ما ذهب إليه غرغوري Gregory (2014) أيضا بأنّ "إجراء معياري مدروس لأخذ عينات من السلوك ووصفها في درجات". (Gregory,[2] 2014, p. 23)

أما حسب (فرج 2017) فعرّف الاختبارات على أنها "أي محك أو عملية يمكن استخدامها بهدف تحديد حقائق معينة أو تحديد معايير الصواب أو الدقة أو الصحة".[3] أيضا أضاف Rolin (2018) ميزة إمكانية المقارنة التي توفرها الاختبارات في تعرفيه لها بحيث ذكر بأنّها "الأدوات المعيارية التي تسمح بتحديد موقع الفرد بالنسبة للآخرين على أن تقام تحت نفس الشروط والظروف".[4] ليتضح لنا في الأخير بأنّ الاختبار النفسي هو أداة موضوعية تستخدم من خلال تحديد عينات السلوك بدقة، لتمثيل السمة التي سيتم اختبارها ودراستها، وفي هذه الحالة يطلب من المفحوص تنفيذ إجراء معين، ثم يتم تقدير النتيجة بناءا على صحة الاستجابة ومقدارها ووقت الاستخدام وغيرها.

---

[1]. عبد القادر رحيم: اتجاهات معاصرة في كيفية استخدام أدوات القياس النفسي. مجلة أبحاث البصرة (العلوم الانسانية) ، 01 (22). 2008، ص 195-205.

[2] Gregory, R: Psychological testing histiory principles and applications. 7 ed. united states of America: Peason Education. 2014, p 23

[3] صفوت فرج، المرجع مذكور سابقا، ص 99

[4] Roulin, J. L: Savoir, Comprendre, Aprendre Leçons de Psychométrie. 3ed. Suisse: Université Savoie Mont Blanc.2018, p36

5-شروط الاختبار الجيّد

تعتبر الاختبارات والمقاييس النفسية من أهم وسائل جمع المعلومات في أي بحث نفسي ككل، ولها شروط أساسية يجب أن تتوفر فيها، ومن أهمها ما يلي:

5-1-الموضوعية

يقصد بالموضوعية في الاختبارات النفسية خلو هذه الأخيرة من أي ميول وذاتية سواء في الاختبار في حد ذاته أو في مطبقه. وهذا بالضبط ما ذهب إليه إبراهيم (2013) في تعريفه بأنّها "عدم تأثر نتائج الاختبار بالآراء الشخصية والتحيزات لمن يقوم بتطبيق المقياس.[1] وأضاف بن عرعور وعمارة (2019) بأن الموضوعية يجب تكون في البحث والموضوعية في عرض النتائج، فالبحث العلمي يجب أن يكون منزها عن الهوى الذاتي، وأن تكون غايته الأولى الدخول إلى الحقيقة واكتسابها سواء اتفقت مع ميول الباحث أو لا تتفق. [2]

أيضا عرّفها Mawardin (2020) بأنّها غياب العناصر الشخصية التّي تؤثر على الاختبار، بمعنى أن الاختبار يكون موضوعيا إذا لم تكن هنالك عوامل ذاتية تؤثر بشكل أساسي في نظام الدرجات عند إجراء الاختبار. [3]

---

[1] سليمان عبد الواحد يوسف ابراهيم: علم النفس التعليمي .ط1. عمان: دار أسامة للنشر والتوزيع. 2013، ص 215

[2] بن عرعور، و عمارة: مواصفات الباحث والبحث العلمي ملتقى وطني حول آداب وآخلاقيات مهنة التدريس الجامعي والبحث العلمي. *مجلة الباحث للعلوم الرياضية والاجتماعية . 2* (1). 2019، ص 373-387.

[3] Mawardin, M. R: Educational Calendar: Specification of a good test. Jurnal Pendidikan Bahasa dan sastra Arab.2 (2). 2020, p 161-175.

وفي معنى غير بعيد عنه عرفها عيواج، عثمان، وبلبكاي (2020) "هي عدم تأثر نتائج الاختبار بصورة جوهرية بذاتية الفاحص، ولو تغير الفاحص ستكون النتائج نفسها كما في حالة وجوده. [1] وفي الأخير يمكن القول أنّ الموضوعية هي شرط أساسي من شروط الاختبار الجيد، والاختبار الموضوعي هو الذي يتسم بعدم التحيّز أو الذاتية سواء في عملية التطبيق أو التصحيح أو التفسير وغيرها، وبالتالي لابد أن تكون نتائجها بشكل عام موثوقة.

5-2-الصدق

بما أن الباحثة قد خصصت فصلا كاملا للصدق وآخر للثبات، لذاك فضلت هنا الإشارة فقط إلى تعريفين أو ثلاث للصدق، وبإمكان القارئ الإطلاع على المزيد من التفاصيل في الجزء المتعلق به، ومن بين التعريفات التي اختارتها الباحثة تعريف AREA, APA, & NCME, (2014)[2] "بأنّ الصدق هو الأدلة والتفسيرات الداعمة لتفسير نتيجة الاختبار ذات الصلة بغرض استخدام الاختبار". وعليه فالصدق مرتبط بالمجال الذّي يتم قياسه، وبالأداة المستخدمة وبدرجة نتيجة القياس.

ويذكر Cronbach (1955) الصدق بأنّه قدرة الاختبار على قياس البناء الذي من المفترض أن يقيسه، إنّها الدرجة التّي ينتج عنها مقياس معين لبناء الدرجات التّي تؤكد التوقعات الواردة في

---

[1] عيواج، عثمان، و بلبكاي: البحث العلمي- أسسه وأدواته: الاختبارات والمقاييس النفسية نموذجا. *مجلة الباحث للعلوم الرياضية والاجتماعية*. *3* (1). 2020، ص 316-327.

[2] AREA, A. E., APA, A. P., & NCME, N. C: Standards for Educational and Psychological Testing. Washington: American Educational Research Association . 2014, p 14

النظرية وهكذا يأخذ البناء معناه من النظرية التّي تحدد طبيعة البناء، ومظاهره وتأكيد العلاقات المفترضة للمقياس مع التركيبات الأخرى.[1]

أما حسب Bringle, Mindy, & Hudson (2004) فإنّه يقول أنّ الصدق ليس له تعريف واحد متفق عليه ولكنّها تشير دوما إلى مدى وجود مفهوم أو استنتاج أو قياس يقوم على أسس سليمة ويتطابق بدقة مع العالم الحقيقي وتعتبر الصلاحية أداة القياس (على سبيل المثال: اختبار في التعليم) هي الدرجة التّي تقيس بها هذه الأداة ما تريد قياسه.[2]

وعليه يمكن القول أنّ الصدق يشير إلى صحة الاختبار، ويهدف إلى وصف ما يتم قياسه، ولتوضيح ذلك هناك سؤالان حول صدق الاختبار هما: ما الذي يقيسه الاختبار، وهل يقيسه بنجاح؟ وليس لذلك علاقة مع اسم الاختبار، ولكن العلاقة مع مضمونه، وتتلخص إجراءات تحديد صدق الاختبار في فحص العلاقات بين الأداء على الاختبار وحقائق أخرى مستقلة قابلة للملاحظة على خصائص السلوك المقصود.

5-3-الثبات

إن الثبات حسب Mohajan (2017) يشير إلى اتساق النتائج واستقرارها وإمكانية تكرارها للنتائج، أي أنّ نتيجة الباحث تعتبر موثوقة إذا تّم الحصول على نتائج متسقة في مواقف متطابقة لكن في ظروف مختلفة، ولكّن يشار إليها في البحوث النوعية إلى أنّها: عندما يكون نهج الباحث

---

[1] Cronbach, L., & Meehl, P: Construct validity in psychological tests. *Psychological Bulletin .52* (4)1955, p. 281-302.

[2] Bringle, R., Mindy, P., & Hudson: The Measure of Service learning. 2 ed. Washington: American Psychological Association. 2004, p 10

متسقا بين مختلف الباحثين والمشاريع المختلفة. [1] وكما هو ملاحظ من هذا التعريف فإنه تم التركيز فيه على إجراء إعادة تطبيق الاختبار في حين ركز كل من Terwee, Mokkink, Knol, Ostelo, Bouter, & de Vet (2012) [2] على أن التعريف الأساسي لثبات الاختبار هو درجة خلو الأداة من الأخطاء العشوائية

وعليه الثبات بشكل عام يعني أننا نحصل على نفس النتائج أو بالأصح نتائج مشابهة، إذا ما تم تطبيق الاختبار مرة أخرى على نفس العينة وفي نفس الظروف، فإنّه يكون كذلك اتساق نتائج الاختبار عند تطبيقه من وقت لآخر، بمعنى آخر يمكن أن نقول بأنّ الثبات هو اتساق القياس في المرة الواحدة أو عبر العديد من القياسات.

5-4-التمييز

إن من أهم صفات الاختبار الجيّد أن تكمن فيه القدرة على التمييز بين مختلف المفحوصين من حيث الأداء، وهذا ما أكده بوداري وبوعزة (2019) في تعريفه للتمييز "بأن يكون الاختبار أو المقياس قادرا على إظهار الفروق الفردية بين الأفراد ويتطلب ذلك تنوع الأسئلة، والوحدات بين السهولة والصعوبة وتدرجها بحيث يظهر الذّين يكونون أعلى أو أقل من العاديين". [3] وفي نفس السياق يعرفها بلعالية (2021) على أنّها قدرة الاختبار على كشف الفروق الفردية بين المفحوصين، لذلك ينبغي أن تكون جميع الأسئلة التّي يشملها الاختبار مميزة، فالاختبار السهل

---

[1] Mohajan, H. K:Two criteria for good Measurements in Reserch: validity and Reliability. Annals of spiru Haret university.17 (3). 2017,p 58-82.

[2] Terwee, C., Mokkink, L., Knol, D., Ostelo, R., Bouter, L., & de Vet, H: Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res. 21* (4). 2012, p 651-657.

[3] بوداري، و بوعزة: أهمية الاختبارات والمقاييس النفسية في البحوث النفسية. *مجلة الوقاية و الأرغنوميا*، جامعة الجزائر 2 .7 (1). 2019، ص 70-86.

الذّي يتحصل فيه كلّ مفحوص على علامات عالية، أو الاختبار الصعب الذّي يتحصل فيه كلّ المفحوصين على درجات منخفضة لا يحقق شرط التمييز، فالاختبار الجيد هو الذي يحتوي على أسئلة ذات مستويات متنوعة من السهولة والصعوبة. [1]

ولكي يفرق الاختبار بين الفئات، فإنّ أنوقيراه Anugerah (2017) يؤكد على واضعي الاختبارات أن يتوخوا الدقة والاعتدال في توزيع مدى سهولة الأسئلة وصعوبتها، بحيث لا تكون كلها صعبة يبرز فيها المتفوقون فقط، أو متوسطة يجيب عنها المتفوقون والمتوسطون دون الضعاف، أو سهلة كلّها بحيث لا تفرق بين الجميع. [2]

وعليه فالاختبار الجيد هو ذلك الاختبار الذي يكون قادرا على تمييز طرفي السمة، وفي نفس الوقت يكون لديه قدرة الكشف عن الفروق الفردية لدى العينة المطبق فيها، إضافة إلى باقي النقاط التي تم ذكرها سابقا مثل الصدق والموضوعية والتي سيتم ذكرها لاحقا مثل الشمولية.

5-5-الشمولية:

إن عنصر الشمولية يمكن أن يتضح فقط من خلال عنوان، فكما يرى زايدي وسهل (2020) أنّ الشمولية هي أن يشمل الاختبار الأهداف المسطرة من أجله، وكذلك أن يشمل الاختبار جميع

---

[1] محمد بلعالية: الاختبارات التحصيلية كأسلوب فعال للتقويم التربوي. *مجلة الخلدونية للعلوم الانسانية والاجتماعية* .13 (1). 2021، ص 53-61.

[2] Anugerah, R: Evaluation of question items in the National Comprehensive Arabic Language Test at the level of the Islamic secondary school for the academic year 2014-2015. *Jurnal pendidikan Bahasa Arab dan kebahasaar aban* . 5 (2). 2017, p 150-175.

العناصر المكونة لموضوع الاختبار، حتى تأتي نتائجها صادقة يجب أن تمثل فقراته للمحتوى الدراسي.[1]

يرى غليون (2009) أنّ الشمولية يقصد به شمول أسئلة الورقة الاختبارية لموضوعات المنهج، على أساس أنّ الشمول يعد أحد المؤشرات الرئيسية لصدق محتوى الاختبار التحصيلي.[2] (غليون، 2009، ص 186)

كما هو واضح من التعريفات السابقة فإن أغلبية الباحثين يعتقدون بأن عنصر الشمولية مرتبط فقط بالاختبارات التحصيلية، ولكن في الحقيقة قد أضافت دراسة COSMIN شرط الشمولية عند التحقق من صدق المحتوى لجميع أنواع الاختبارات، بالإضافة إلى شرط الملائمة، وركزت على أن يشرك الباحثين أفراد عينة البحث عند التحقق من الشمولية والملائمة، فيجب أن يُسأل المفحوصين صراحة عما إذا كانت العناصر معًا تغطي بشكل شامل بنية الاختبار (أو المقياس الفرعي) التي يعتزم قياسها. [3] ولكن حتى بعد التطبيق يمكن لمعد أو مطبق الاختبار التحقق من شموليته وبالتالي التحقق من صلاحيته. إذا فشمولية الاختبار تعني أن تغطي بنود الاختبار جميع جوانب السمة أو البرنامج أو الأهداف المتوقع من الاختبار أن يقيسها، ويفترض أن يشارك كل من الخبراء وأفراد العينة في تحديد مدى شمولية الاختبار.

---

[1] زايدي وسهل: الاختبارات ودورها في تحقيق مبدا الجودة التعليمية. 22(49). 2020، ص 269-296.

[2] أزهار محمد غليون: تقويم الاختبارات النهائية للدراسات العليا بكلية التربية صنعاء في ضوء معايير الاختبار الجيد. *المجلة العلمية لكلية التربية .1* (8). 2009، ص 176-200.

[3] Mokkink, L., de Vet, H., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., et al:Des 19). COSMIN risk of bias checklist for assessing the methodological quality of studies on the measurement properties of Patient-Reported. epub ahead of print. 2017. p. 1171-1179.

5-6-التقنين

إن التقنين يتمثل في التحديد الدقيق لإجراءات التطبيق والتصحيح والتفسير، أي أنّ شروط الاختبار هي نفسها بالنسبة لجميع العينات، مما يسمح بأن يستخدم الاختبار ويعطي الدرجة ذاتها من الكفاءة مهما اختلفت نتائج العينة. وهي حسب عباس و مهلل (2016) [1] العملية التي يتم من خلالها التحكم في العوامل غير مناسبة التّي يمكن أن تؤثر في عملية القياس وذلك من خلال تخفيض أخطاء القياس إلى حدها الأدنى عن طريق اختبار عينة ممثلة لمجتمع الدراسة يطبق عليها اختبار، ومن ثم توحيد فقراته وإجراءات تطبيقه وتصحيحه بشكل يوفر للاختبار خصائص سيكومترية تتفق مع خصائص الاختبار ومن ثم توفير المعايير المناسبة لتفسير الدرجات الخام.

في نفس السياق يرى الزاملي (2017) أنّ التقنين هو رسم خطة شاملة وواضحة ومحددة لجميع خطوات الاختبار وإجراءاته وطريقة تطبيقه وتفسير درجاته وتحديد السلوك المطلوب من الفرد والشروط المحيطة به، في أثناء تطبيق الاختبار بالإضافة إلى وجود معايير لتفسير الدرجات. والباحثون يستخدمون الاختبارات المقننة للأسباب الآتية:

أ. توفر عليهم بذل الجهد والوقت لوضع اختبارات قد تحقق أو لا تحقق الهدف الذّي يسعون إليه.

ب. تمكنهم من إجراء مقارنة بين أداء الأفراد عندهم مع أفراد آخرين طبق عليهم الاختبار نفسه.[2]

إذا ما نستطيع قوله حول شرط التقنين أنه يمكننا من توحيد وتطبيق إجراءات الاختبار وتصحيحها وكذا تفسير نتائجها بطريقة موحدة، حتى لو اختلف الفاحصون وتعدد المفحوصين وتنوعت مواقف الاختبار وتكررت مرات القياس.

---

[1] عبد الرحمان عباس، و زينة مهلل: النقل الثقافي للاختبارات النفسية ذات المنشأ الغربي في البيئة الجزائرية. المجلة الجزائرية للطفولة والتربية . 4 (1). 2016، ص 231-260.

[2] علي حسين هاشم الزاملي: *بناء وتقنين المقاييس النفسية*. جامعة القادسية: دار الكتب في بغداد.2017، ص 79

5-7-القابلية للاستعمال

بعد أن يضع الباحث الاختبار في صورته النهائية التّي تقدم للمفحوصين، هنا عليه أن ينتبه للقابلية للاستعمال، فحسب الكحلوت (2016) تتضح القابلية للاستعمال من خلال مجموعة من الاعتبارات العملية التّي يأخذ بها مستخدم الاختبار لتقويم ملاءمة الاختبار لأغراضه الخاصة في خطة التقويم، أو البحث التّي يعمل بها كذلك تشمل هذه الاعتبارات عناصر الاقتصاد بالكلفة، والجهد والوقت في ضوء تقدير الأهمية للقرارات، والفوائد المتحققة من استخدام الاختبار إضافة إلى ذلك يحتاج مستخدم الاختبار أن يتأكد من وجود تعليمات محددة وواضحة لتطبيق الاختبار. [1]

أما حسب إساعدي وعزي (2015) يصبح الاختبار قابلا للاستعمال إذا ما كان عمليا من حيث:

5-7-1-سهولة التطبيق: فنجد أنّه كلما كان تطبيق الاختبار سهلا كلما كان ذلك أفضل لأنّ صعوبة التطبيق قد تكون عائق يمنع من تحقيق صدق الاختبار وثباته وموضوعيته، وتشمل سهولة التطبيق وضوح الأسئلة وتعليمات الاختبار، فالتعليمات توجه الطالب وترشده في أداء الاختبار، كونها يخبره بما ينبغي أن يفعله بالضبط، وتوضح الوقت المخصص لحل أسئلة الاختبار، وهذه التعليمات يجب صياغتها بلغة بسيطة وواضحة توضح الهدف من الاختبار، وطريقة تسجيل الاجابات، وتوضح كذلك للطالب أن يتجنب التخمين في حال ارتيابه في تحديد الإجابة الصائبة. [2]

---

[1] عماد حنون الكحلوت.: فاعلية برنامج تدريبي لتنمية مهارات إعداد الاختبارات النفسية وتطبيقها وتفسير نتائجها لدى الطلبة المرشدين النفسيين. *مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية . 4* (14). 2016، ص 133-172.

[2] فارس إساعدي، و ايمان عزي: مواصفات الاختبار التحصيلي الجيّد. *مجلة الدراسات والبحوث الاجتماعية. 3* (2). 2015، ص 189-200.

أما زايدي وسهل (2020) يرى أنّه أي أسئلة واضحة وينطبق عليها خصائص الاختبار الجيد التي ذكرناها سابقا وكذلك التعليمات التّي يتم وضعها يجب أن تكون واضحة وسهلة. [1]

5-7-2-سهولة تصحيحه: ويرى ملحم (2002) أنّ تصحيح الاختبار يتوقف على نوع الاختبار المستخدم وعلى الإجراءات التّي أعدت لأغراض التصحيح، وأي كانت نوعية الاختبار، فإنّه يجب على المصحح مراعاة الموضوعية من خلال توفير مفتاح تصحيح الاختبار الذّي يعده. حسب [2] .

أما عدائكة وبن موسى (2017) يرى أنّه تختلف عملية التصحيح بالنسبة للاختبارات المقالية والاختبارات الموضوعية في أنّ تصحيح اختبارات المقال معقدة مقارنة بالاختبارات الموضوعية حيث تسمح بتدخل العوامل الأخرى في عملية التصحيح كالخدعة من قبل الطالب والعوامل الذاتية من قبل المعلم مما يجعل من الصعب التقيد بدليل التصحيح وتتمثل سهولة التصحيح في أن تكون الأسئلة واضحة ومحددة مع إمكانية وضع سلم للتصحيح. [3] من خلال ما سبق نلاحظ أنّه من العوامل الهامة المحددة لاختيار أي أداة للقياس هي مدى توفرها على خاصية قابلية الاستخدام في المواقف المطلوبة والفعلية، ومدى ما تتطلبه من خبرة أو تدريب على تطبيقها وكذلك سهولة استخدامها وتصحيحها وتفسير درجاتها.

---

[1] زايدي وسهل، مرجع مذكور سابقا، ص 283.

[2] محمد عرفات جخراب، ربيعة جعفور، و دنيا عدائكة: كفايات إعداد الاختبارات التحصيلية الجيدة لدى أساتذة التعليم الثانوي. *مجلة العلوم الانسانية . 26* (4). 2015، ص 515-540.

[3] دنيا عدائكة، و يمينة بن موسى: معايير الجودة في بناء الاختبارات التحصيلية الجيّدة. *مجلة السراج في التربية وقضايا المجتمع. 1* (1). 2017، ص 216-227.

6-أنواع الاختبارات النفسية

هناك أنواع عديدة من الاختبارات النفسية، وتصنف إلى عدّة أصناف وطبعا لا يمكن حصرها جميعا لكثرتها، لذلك ستحاول الباحثة تلخيصها قدر الإمكان وعرض جميع ما وقع تحت يدها ولكن باختصار مدروس. وتم عرض الأنواع في تصنيفين الأول حسب Gregor (2014) والثاني حسب (Gay, 1980 وأكابور، أهلاوات وآخرون 1985) كما ورد في ملحم (2000) وقد اختارت الباحثة هذان التصنيفان لكونهما مكملين لبعضهم البعض، فلم يتقاطع التصنيفين إلا في بعض الاختبارات، فما ورد في التصنيف الأول لم يرد في التصنيف الثاني إلا قليلا.

6-1-التصنيف حسب غريغوري Gregory

يرى Gregory (2014) [1] بأنّه يمكن تصنيف الاختبارات إلى عدة تصنيفات أهمها:

6-1-1-اختبارات فردية مقابل اختبارات جماعية: من الواضح من خلال العنوان أن أهم فرق بين هذين النوعين من الاختبارات هو عدد المفحوصين، فإذا كان فردا واحدا فعدد المفحوصين، فإذا كان فردا واحدا فإن الاختبار يعتبر فرديا، أما إذا طبق الاختبار على مجموعة فالاختبار هنا هو جماعي، ولكن هناك عدة اختبارات أخرى وفروق تلعب دورا أكبر فحسب امطانيوس (2016) من الفروق المهمة بين الاختبارات الفردية والاختبارات الجماعية أنّ التعليمات في هذه الأخيرة تكون أبسط منها في الأولى كما لا تتطلب الاختبارات الجمعية خبرة أو مهارة خاصة من جانب الفاحص، وهي أقل كلفة وأكثر انتشارا من الاختبارات الفردية، غير أنّ الاختبارات الفردية تعطي الفرصة للتواصل الحي بين الفاحص والمفحوص مما يفسح المجال

[1] Gregory, R: *Psychological testing histiory principles and applications* .7 ed. united states of America: Peason Education. 2014, p 26-28

لاستثارة دافعية المفحوص وضمان ملاحظة الطرائق التّي يلجأ إليها المفحوص في أدائه الاختباري وأساليب حله للمشكلات والتّعرف على نوع الأخطاء التّي يرتكبها. [1] وفي الأخير لا يمكن القول أن هذا النوع أحسن أو أسهل في تطبيقه وغيرها وإنما لكل نوع مميزات خاصة به. فتعتبر الاختبارات جماعية عندما تكون مناسبة لاختبار مجموعات كبيرة من الأشخاص في نفس الوقت، أما الاختبارات الفردية هي الأدوات التّي من خلالها يقوم التصميم وعلى أساس واحد.

6-1-2-اختبارات الذكاء: تّم تصميم اختبارات الذّكاء في الأصل لأخذ عينة من مجموعة واسعة من المهارات من أجل تقدير المستوى الفكري العام للفرد. وهذا ذهب إليه الكافي (2001) في أنّ قياس الذكاء ما هو إلا عملية تحويل المعطيات المختلفة للنمو العقلي والذكاء إلى أرقام و كميات، ومعرفة مدى تناسب تلك الأرقام مع عمر الطفل وللحصول على عمر الطفل الحقيقي وللحصول على مستوى الذكاء عند الطفل ومقارنته بالعمر ذاته لدى الأطفال الآخرين في نفس المستوى تقريبا. [2] أما أكسلسون Axelsson (2016) فيعتقد أن اختبارات الذكاء صممت أساسا للكشف عن التأخر العقلي فلقد صمم بنيه اختبارات الذكاء لتحديد التلاميذ المتأخرين في نموهم العقلي.[3] وهذا في الحقيقة الأساس الذي انطلقت منه اختبارات الذكاء، فهو الهدف الأول وراء بناء اختبارات الذكاء في البداية لكنه لم يعد الأمر الآن مرتبطا فقط بهذا الهدف، بل أصبحت اختبارات الذكاء تستخدم مع العديد من الأهداف الأخرى مثل التوظيف والقبول الجامعي وغيرها.

---

[1] نايف مخائيل امطانيوس: *بناء الاختبارات والمقاييس النفسية والتربوية وتقنيتها* (ط1). عمان: دار الاعصار العلمي للنشر والتوزيع.2016، ص 36

[2] إسماعيل عبد الفتاح الكافي: *اختبارات الذكاء والشخصية*. مركز الاسكندرية للكتاب.2001، ص 7

[3] Axelsson, T:i ntelligence testing: ethnicity, and construction of the deviant child: foucauld and special education in sweden. *Nordic journal of social research. 7* (1). 2016, p. 22-35.

6-1-3-اختبارات القدرات: تعتبر اختبارات القدرات من أفضل أدوات التشخيص فقد تعددت استخداماتها وتطبيقاتها بشكل كبير في العديد من مجالات الحياة، خاصة أنها تجمع العديد من القدرات تحت مظلة واحدة، وهذا واضح في تعريف معمرية (2012) لها بأنها تقيس مدى قدرة الفرد على التعلم والتدريب على مهنة أو مهارة معينة كالقدرة الميكانيكية والقدرة الموسيقية، والقدرة الحسابية والقدرة اللفظية وغيرها، وعادة ما يجمع هذا النوع من الاختبارات في اختبار واحد ويسمى بطارية، وتقيس البطارية عددا من القدرات الخاصة في نفس الوقت. [1] . وفي معنى لا يختلف كثيرا عن المعنى السابق يرى ستيواتي Setiwati (2020) على أنّها تتضمن بطاريات متعددة القدرات التّي يمكن استخدامها لقياس مختلف القدرات كالقدرات العددية والقدرات الميكانيكية والرؤية والتفكير العددي والسرعة والدقة في الإدراك. [2] أما العتيبي (2018) يرى بأنّها ترتكز على قياس قدرة الطالب ومهاراته على التفكير المنطقي السليم أكثر من التركيز على قياس قدرة الطالب في تذكر المعارف والحقائق التّي تعلمها في مراحله الدراسية المعينة واسترجاعها عند الحاجة. [3] ومن خلال جميع ما سبق نستخلص أن اختبار القدرات هو اختبار يقيس بوضوح أجزاء محددة متجانسة نسبيا من القدرة، وتأتي مثل هذه الاختبارات في نوعين هما الكفاءة الفردية، وبطاريات اختبارات القدرات المتعددة، ومن الواضح أنّ تقييم اختبار الكفاءة هو تقييم واحد فقط أي يرتكز على قدرة معينة، في حين توّفر بطارية اختبارات القدرات المتعددة لمحة لعدد من القدرات.

---

[1] بشير معمرية: *أساسيات القياس النفسي وتصميم أدواته*. جامعة الحاج لخضر، باتنة: دار الخلدونية للنشر والتوزيع. 2012، ص 111

[2] Setiwati, F. A: Aptitude test's presdictive Ability for Academic success in psychology student. Psychological Reaserch Student . 3 (1).2020, p 1-12.

[3] خالد العتيبي: مستويات أداء الطلبة الثانوية العامة في المملكة العربية السعودية في اختبار القدرات العامة. *مجلة العلوم التربوية والنفسية*. *19* (1). 2018، ص 613-638.

6-1-4-اختبارات الإبداع: تعتبر اختبارات الإبداع من الاختبارات المنتشرة بكثرة، فلقد استخدمت لأغراض عديدة من بينها الكشف عن الطلبة الموهوبين واختيارهم للالتحاق ببرامج تربوية خاصة، كما استخدمت في إجراء البحوث والدراسات الجامعية وغيرها، بالإضافة إلى أنه يمكن تطبيق الاختبارات الإبداعية بصورة فردية أو جماعية. ويرى كروبلي cropley (2000) بأنّ اختبارات الإبداع تعتمد على عمليات معرفية متعددة كالتفكير بشكل متباين، أو تكوين الجمعيات أو بناء ودمج فئات واسعة، أو العمل على العديد من الأفكار، كما أنّها تقيس أيضا الجوانب غير المعرفية للإبداع مثل التحيز على سبيل المثال (التعبير عن الاندفاع والرغبة في التجديد والمخاطرة) والخصائص الشخصية كالمرونة والتسامح ، أو المواقف الايجابية تجاه الاختلاف، وكما ترتبط الاختبارات أيضا بدرجة معقولة بمعايير مختلفة للإبداع مثل تقييمات المعلم وهي تنبؤات مفيدة لسلوك البالغين وبالتالي فهي مفيدة في كل من البحث والتعليم، ومع ذلك فإن مفهوم الإبداع متعدد يشير إلى أنّ التقييمات يجب أن تستند إلى عدة اختبارات بدلا من الاعتماد على درجة واحدة. [1] وهو الأمر الذي أكده Tanhen banm (1985) كما ورد في حزيمة (2016) فالاختبار الإبداعي يستخدم للكشف عن الطلبة الذّين يتمتعون بقدرة إبداعية في كثير من البرامج الخاصة لتعليم الموهوبين والمتفوقين، ولا سيما في ذلك النوع من البرامج التّي تركز على تقديم خبرات لتنمية الإبداع والتفكير الإبداعي لدى الطلبة وقد تكون هذه الخبرات مرتبطة بالمناهج المدرسية وقد تكون مستقلة عنها تماما، وتتطلب أسئلة اختبارات الإبداع والتفكير الإبداعي الخصائص السيكومترية التي تتمتع بها اختبارات الذكاء الفردية المعروفة، من حيث الصدق

[1] cropley, A: Defining and Measuring creativity: Are creativity tests worth using? *Journal of Educational Measurement .23* (2). 2000, p. 72-79.

والثبات، لذا لا ينصح باستخدامها منفردة في الكشف عن الطلبة الموهوبين، بل تكون مصدر إضافي أو ثانوي. [1] إذاً تقيّم اختبارات الإبداع قدرة الموضوع على إنتاج أفكار أو رؤى أو إبداعات فنية جديدة يتم قبولها على أنّها اجتماعية أو جمالية أو علمية القيمة.

6-1-5-اختبارات التحصيل: تعتبر اختبارات الذكاء وسيلة منظمة تهدف إلى قياس كمية المعلومات التّي اكتسبها التلميذ أو الطالب، كما تشير أيضا إلى قدرته على فهمها واستخدامها في المواقف المختلفة. وهو ما أكده كوهن وسويدريك Cohen & Swerdik (2010) فحسبه تم تصميم اختبارات التحصيل لقياس درجة التعلم الذّي حدث نتيجة التعرض لتجربة تعليمية محددة نسبيا.[2] وكما يرى رولين Rolin (2018) بأنّها اختبارات تجعل من الممكن تقييم المعرفة المكتسبة كاختبارات المعرفة الأكاديمية والمهنية.[3] وفي نفس السياق يعرفها ميهرن وليمن Mehrens & Lehman (1975) بأنّه اختبار يقاس به مدى تحصيل الفرد لشيء من الأشياء أو اكتسابه لمعرفة محددة أو مهارات معينة وفي الغالب تتم هذه الأشياء نتيجة تعليم أو تدريب مخطط منه. أي بأنّه اختبار صمم ووضع خصيصا لقياس التعلم لمكتسبات معينة أو التعلم النظامي الرسمي. كذلك يعرفها ألرفوه Alrfooh (2012) على أنّها الطريقة المنظمة لتحديد مستوى تحصيل الطالب للمعلومات والمهارات المتضمنة في موضوع الدراسة المستفادة من خلال الإجابة على عينة

---

[1] كمال عبد المجيد حزيمة.: الكشف عن فاعلية أداتي لكشف الموهوبين من وجهة نظر مدريسهم. مجلة كلية التربية. 1 (27). 2016، ص 175-194.

[2] Cohen, R. J., & Swerdik, M: *Psychological testing and Assessment.* ed 9. united stastes of America: MC Graw- Hill Education. 2018, p 20

[3] Roulin, J. L: *Savoir, Comprendre, Aprendre Leçons de Psychométrie* . ed 3. Suisse: Université Savoie Mont Blanc. 2018, p 45

من الأسئلة (الفقرات) التّي تمثل محتويات المادة. [1] وأما شافيزان shafizan (2013) يرى بأنّه الغرض من اختبارات التحصيل هو قياس ما تّم تعلمه بدلا من ذلك من توقع الأداء المستقبلي. [2] وفي الأخير فإنّ الهدف من إجراء الاختبارات التحصيلية هو التّعرف على مختلف قدرات الطالب من حيث مدى استيعابه للمادة العلمية المقدمة له وإمكانية توظيفه لها، والاختبار التحصيلي لا يعبر فقط عن مستوى الطالب وإنّما يعبر أيضا عن مختلف قدرات الأستاذ في نقل وتوصيل المعارف إلى الطالب من جهة، وفي اختياره للاختبار المناسب لطبيعة المعلومات المقدمة والموافق أيضا لقدرات الطالب من جهة أخرى. فالاختبار التحصيلي هو قياس درجة الشخص في التّعلم أو الانجاز أو تحقيق الأهداف في منهاج ما أو موضوع أو مادة تعليمية معينة.

6-1-6-اختبارات الشخصية: تتعلق اختبارات الشخصية عادة بقياس السمات والصفات أو السلوكيات التّي تحدد فردية الشخص، وكما تساعد هذه المعلومات في توقع السلوك المستقبلي. وقد عرّفها رولين Roulin (2018) بأنها الاختبارات التي تقيّم جميع الجوانب غير المعرفية للشخصية منها الاهتمامات والمواقف والقيّم الشخصية والسمات الشخصية، وتركز بعض اختبارات الشخصية في بعض الأحيان على جانب واحد فقط.[3] أي أن هنالك بعض اختبارات الشخصية التي تقيس العديد من الجوانب الشخصية، في حين توجد بعض اختبارات الشخصية التي تعمل على قياس جانب واحد فقط من الجوانب الشخصية. أمّا Adoga أدوقا (2020) فيرى بأنّها هي أدوات

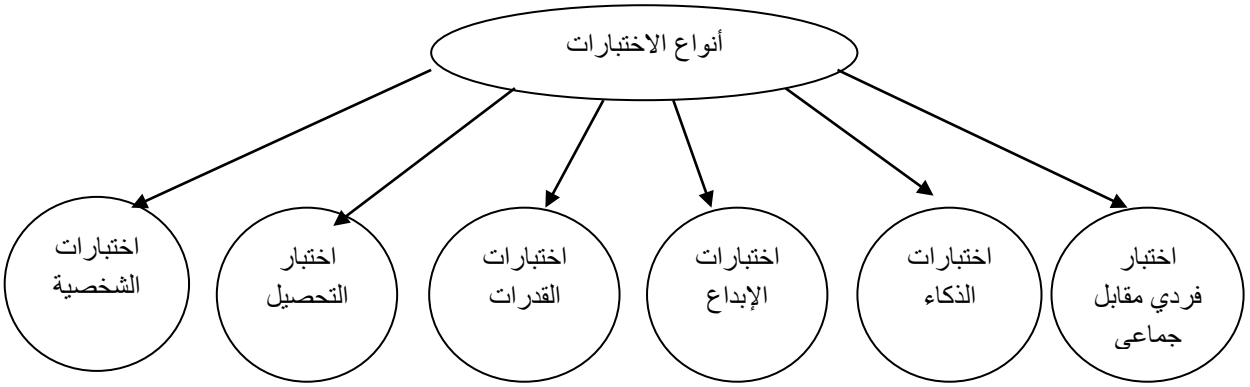
[1] Alrfooh, A. E: type of avheivement tests which are preferred by outstanding students at al-hussein bin talal university. *Journal of Education and practice . 3* (13)2012 ., p 149-156

[2] shafizan, s.: Item analysis of student comprehensive test for research in teaching beginner string ensemble usingmodel based teaching among music students in public universities. *1* (12). 2013, p 1-14.

[3] Ibid, p 45

لقياس الجوانب العاطفية أو غير الفكرية للسلوك. ويمكن استخدامها لقياس جوانب الشخصية كالاستقرار العاطفي، الود، الدافع، الاهتمامات والتواصل الاجتماعي والانطواء وكذلك تساعد هذه الاختبارات لمعرفة نوع السمات الشخصية التي من المرجح أن يصور الفرد فيها حالة الحياة الفعلية. [1] وفي الأخير فإن مجال قياس الشخصية يعتبر أحد المجالات الهامة في مضمار القياس السيكولوجي، ذلك أنّها تتناول خصائص الفرد وسماته وتهتم مقاييس الشخصية أساسا بالمظاهر الوجدانية، ومصطلح الشخصية يستخدم للإشارة إلى كل سمات الفرد مثل التوافق الوجداني والانطواء، الانبساط القلق والعلاقات الشخصية والدوافع والميول والاتجاهات وغيرها.

الشكل 01: أنواع الاختبارات حسب Gregory **(2014)**



المصدر: من إعداد الباحثة اعتمادا على [2](Gregory, 2014)

يعتبر هذا النموذج هو اختصار لجميع ما تم ذكره سابقا، وكما هو واضح فقد اعتمد Gregory (2014) في تصنيفه للاختبارات النفسية على السمات، فقد تم تقسيم الاختبارات على أساس السمة التي يقيسها كل اختبار، مثل التحصيل أو الإبداع وغيرها، ماعدا النوع الأول والذي هو الاختبارات

[2] Adoga, R.: psychological test in counselling. prestige journal of conselling psychology. 2020, p. 248.

[2] Ibid, p 26-28

الفردية مقابل الجماعية فهو النوع الوحيد الذي يعتمد على كيفية تطبيق الاختبار وليس السمة التي يقيسها. وهذا على عكس ما سنراه في التصنيف التالي.

6-2-التصنيف حسب ملحم:أضاف كل من (Gay, 1980 وأكابور، أهلاوات وآخرون 1985) كما ورد في ملحم (2000) [1] بعض التصنيفات التي لم يتم ذكرها سابقا، ولذلك عمدت الباحثة على إضافتها أما الأنواع أو التصنيفات التي تم ذكرها سابقا فتم تجاهلها لتفادي التكرار.

6-2-1-الاختبارات الموضوعية مقابل الاختبارات المقالية: حيث يكون الاختبار موضوعيا إذا كان إعطاء العلامة للسؤال أو الاختبار بعيد عن الذاتية، أي أن الحكم على الإجابة يكون بطريقة مدروسة ومقننة وبعيدة عن الآراء والأفكار الشخصية للمصحح، وهذا مرتبط بخصائص وقواعد يعبر عنها الاختبار الموضوعي. وقد عرّفها حسن (2018) بأنها الاختبارات التي تمكن المعلم من تكوين أحكام موضوعية على إجابة الطالب، والموضوعية هنا يقصد به البعد عن الذاتية في تقدير الدرجة على الإجابة، أي أن الدرجة التي يحصل عليها الطالب لا تختلف باختلاف المصحح، وبذلك لا تتأثر هذه الاختبارات بالعوامل الذاتية. [2]. أما الاختبار المقالي فهو الاختبار الذي يتضمن أسئلة المقال كالإنتاجيات الفنية أو الأعمال الخلاقة، حيث تدخل الذاتية عند إعطاء العلامة، فإذا صحح اثنان أو أكثر مقالة أو رسما أو تلوينه، فإنّهم سيختلفون في وضع العلامة بل قد يقدم المصحح نفسه علامتين مختلفتين في مناسبتين مختلفتين، وهذا نوع مرتبط

---

[1] سامي محمد ملحم: *القياس والتقويم في التربية وعلم النفس*. ط 1. عمان: دار المسيرة للنشر والتوزيع والطباعة. 2000، ص 55

[2]علي صلاح عبد المحسن حسن: الأخطاء الشائعة في بناء الاختبارات الموضوعية وتأثيرها علي الخصائص السيكومترية للاختبارات التحصيلية لدى طلاب كلية التربية بجامعة أسيوط. *المجلة العلمية لكلية التربية*. *34* (2). 2018، ص 322-402.

أيضا بخصائص وقواعد ومزايا يعبر عنها الاختبار المقالي أو الذاتي. وكما أشار زايدي وسهل (2020) في تعريفه للاختبارات المقالية بأنها القائمة على طلب المعلم من المتعلم كتابة مقال أو نص يقرب في مواصفاته من المقال معبرا فيه عن أفكاره وآرائه، وتعتبر الاختبارات المقالية هي أكثر رواجا في المدارس لسهولة إعدادها من قبل المعلمين وللحرية الكبيرة، المتروكة لهم في تصحيحها، أي أنّ هذا النوع من الاختبارات يعتمد على تثبيت وكتابة السؤال والجواب في الورقة. [1] إن الحديث عن ذاتية المصحح في هذا النوع من الاختبارات لا ينقص من قيمة الاختبار المقالي، وكذلك لا يرفع من قيمة الاختبار الموضوعي فلكل منهما مزاياه وعيوبه، ولكل منهما مواقف تستدعي نوعا معينا دون الآخر، ولكن ما يفضل الاختبار الموضوعي لسهولة تصحيحه والأهم من ذلك لإمكانية وضع إجابة نموذجية يتفق عليها المصححين.

6-2-2-الاختبار المقنن مقابل الاختبار الغير مقنن: ويقصد بالاختبار المقنن هو الذي يعد من قبل مختصين في القياس وبناء الاختبارات، بطرق وخطوات متفق عليها ومدروسة، عند البناء وعند إدارة الاختبار وحتى عند تصحيحه. كما أنّ هذا النوع ينتجه الناشرون والتجاريون عادة بإشراف المختصين في القياس طبعا مع مراعاة إدارتها وتصحيحها وتفسيرها تحت ظروف معيرة ومحددة. وقد عرّف مهلل وعباس (2016) الاختبار المقنن بأنه الاختبار الذي يخضع للعملية التّي يتم من خلالها التّحكم في العوامل غير المناسبة التّي يمكن أن تؤثر في عملية القياس وذلك من خلال تخفيض أخطاء القياس إلى حدها الأدنى عن طريق اختبار عينة ممثلة لمجتمع الدراسة يطبق عليها اختبار، ومن ثّم توحيد فقراته وإجراءات تطبيقية وتصحيحية بشكل يوفر للاختبار خصائص سيكومترية تتفق مع خصائص الاختبار الجيّد، ومن ثم توفير المعايير المناسبة لتفسير

[1] زايدي وسهل، مرجع مذكور سابقا، ص 269- 296.

الدرجات الخام. [1] بأنّها تلك الاختبارات التّي يتم إعدادها بمعرفة من الباحثين المتخصصين في مجالات القياس التربوي والنفسي أو الرياضي، وتتميز الاختبارات المقننة بأنّها مصممة بعناية فائقة وأنّه قد تم تجريبها مرات كثيرة للتحقق من مدى صلاحيتها. أما الاختبارات الغير مقننة فهي تمثل غالبية الاختبارات في مجال التربية والتعليم. وهي اختبارات تصمم من قبل المعلمين وتتميز بأساليب غير معيرة لكنّه يمكن مقارنة نتائجها مع مجموعة أخرى من الاختبارات، وتمتلك عدة خصائص ومزايا تتلاءم وأهداف المعلم المحددة لعينة من الطلاب الذّين يقوم المعلم بقياس قدراتهم في موضوع دراسي ما. في حين عرّفها إبراهيم (2013) بأنّها الاختبارات الغير الرسمية وهي اختبارات تكون من صنع المعلم وغير مقننة ولا يعتمد على نتائجها. [2] وللإشارة هنا لا تنحصر الاختبارات الغير مقننة في الواقع فقط على الاختبارات في مجال التربية والتعليم، فلقد لاحظت الباحثة العديد من الاختبارات الموجودة فعلا في ميدان علم النفس وهي غير مقننة، رغم أن الباحثين يعتبرونها غير ذلك، فهنالك خلط واضح بين التحقق من صلاحية الاختبار وتقنينه.

6-2-3-اختبارات لفظية مقابل اختبارات أدائية: إن الاختبار اللفظي كما هو واضح من اسمه هو الاختبار الذي يعتمد على الألفاظ والكلمات والتعبيرات المنطوقة في استخدامه، وقد عرّفه قصاب (2014) بأنّه يعتمد على اللغة والألفاظ في تكوينه، ومن أشهر هذا النوع من الاختبارات اختبار ستانفورد بنيه واختبار وكسلر للذكاء، إذ تقدم المفردات في صورة لفظية، أي عبارات لها معنى، ولقدرة الفرد المشارك على استخدام الكلمات وفهمها دور مهم في تحديده فيجب أن يكون

---

[1] عبد الرحمان عباس، و زينة مهلل: النقل الثقافي للاختبارات النفسية ذات المنشأ الغربي في البيئة الجزائرية. *المجلة الجزائرية للطفولة والتربية . 4* (1). 2016، ص 231-260.

[2] ابراهيم، المرجع مذكور سابقا، ص 217

المشارك قادرا على إصدار الاستجابة من عدمها. [1] وفي المقابل نجد الاختبارات الأدائية وهي الاختبارات التي تعتمد على الأداء كما هو واضح وليس على اللفظ كسابقه، وتصمم لمشاهدة الاستجابات غير اللفظية مثل أداء مهمة أو تجميع وتفكيك جزء من آلة وغيرها من النشاطات النفسية الحركية للمفحوصين.

وحسب فرج (2007) فإنّ كل أنواع الاختبارات تتطلب أداء بشكل من الأشكال، إلا أنّ تعبير اختبار أدائي يستخدم عادة في الإشارة إلى الاختبارات التي تتطلب استجابة غير لفظية، وإنّ كانت لا تقتصر على مجرد الإعفاء من استخدام اللغة، بل تتطلب تعاملا مع مواد معينة، سواء في شكل ترتيب أو تصنيف أو تركيب لأشكال أو خامات أو نماذج مختلفة. وتستخدم الاختبارات الأدائية في مجال معين، وصنّف الذكاء في فترة مبكرة من بحوث الذكاء العام إلى ذكاء لفظي وذكاء أدائي، وقد اهتم الباحثون بالتعرف على الارتباطات بين كلّ من العاملين اللفظي والأدائي، ودلالة الفروق بين هذين المجالين المميزين من مجالات الأداء النفسي. [2] والاختبار اللفظي لا يعتبر أهم من الاختبار الأدائي أو العكس فكلاهما مكمل للآخر، ولكل منهما خصائص تميزه وتجعله صالحا لمواقف دون الأخرى. وكما تم الإشارة سابقا فالاختبار الأدائي يعتمد بكثرة في المرحلة التي يكون فيها الفرد عاجزا عن التعبير اللفظي سواء لصغر سنه أو لمرض أصابه أو لمشكلة في التعبير وغيرها، ولكن لا يقتصر على هذا النوع من الاستخدام.

6-2-4-اختبارات السرعة مقابل اختبارات الدقة: فاختبار السرعة يكون التركيز فيه على سرعة الإجابة وضبطها أكثر من التركيز على المحتوى أو المعرفة. حيث يتمكن الطلاب من

---

[1] ريم نصر قصاب: تقنين اختبار أوتيس– لينون للقدرة المدرسية على عينة من طلاب التعليم الأساسي في مدينة حماة. رسالة دكتوراه غير منشورة . كلية التربية: لجامعة دمشق. 2014، ص 41

[2] صفوت فرج: *القياس النفسي*. ط 6. مصر: مكتبة الأنجلو المصرية. 2007، ص 115.

خلالها في أداء مهمة بدرجة معينة من الضبط والتّي تمثل الهدف الحقيقي للقياس، وعليه ففي هذا النوع من الاختبار فإنّ السمة التّي يجب قياسها هي سرعة الأداء.

وحسب معمري (2018) تكون سرعة الإجابة هي العامل الحاسم في تحديد الأداء.[1] في حين أشار بركات (2018) بأنّها اختبارات التّي يعتمد تقويم المفحوص عليها وفق سرعته في الإجابة، والسرعة بهذا المعنى هي المقياس الحاسم كمحك للإتقان، وتكون اختبارات السرعة عادة محددة الزمن، وهذه الاختبارات مرتبطة باختبارات الأداء الأقصى و المتمثلة باختبارات التحصيل.[2] أما اختبارات الدقة فهي التي تستخدم بهدف تقدير حدود القدرة أو العمق أو التوسع المعرفي في مجال موضوع معين. وكما أشار زمزمي (2008) على أنّها تهدف إلى قياس دقة الأداء باستخدام عدد من الفقرات المتساوية في الصعوبة، وتحسب دقة الأداء بالنسبة المئوية من الأخطاء التي يقع فيها المفحوص. [3] (زمزمي، 2008، ص 38). أي أن الهدف هنا هو قياس دقة الإجابة ومدى تمكن المفحوص أو الطالب من الإجابة بطريقة صحيحة وسليمة دون إعطاء أهمية للوقت الذي استغرقه في استرجاع المعلومة أو حلها، فالهدف هنا هو قياس دقة الأداء وليس سرعته.

6-2-5-اختبارات استدعاء الاستجابة مقابل اختيارها: فاستدعاء الاستجابة يهدف إلى اختبار الذاكرة بحيث يقدم الشخص الإجابة من ذاكرته ويستدعي الاستجابة للسؤال.

---

[1] أحلام معمري: أسس وضع اختبارات تعليم اللغة العربية للناطقين بغيرها. *مجلة الباحث في العلوم الانسانية والاجتماعية* .(35). 2018، ص 595-606.

[2] زياد بركات: *القياس والتقويم النفسي والتربوي بين النظرية والتطبيق.* جامعة القدس: عمادة البحث العلمي للنشر والتوزيع. 2018، ص 181

[3] عبد الرحمن بن معتوق بن عبد الرحمن زمزمي: تقنين اختبار تورانس للتفكير الابتكاري الشكل (ب) على الطلاب الصم وضعاف السمع في المرحلة المتوسطة بمنطقة مكة المكرمة. رسالة دكتوراه غير منشورة. كلية التربية: جامعة أم القرى، مكة المكرمة. 2008، ص 38

فالاختبار الذي يعتمد على استدعاء الاستجابة حسب الجراح و أبو أحمد (2015) يختص بتذكر الأحداث والخبرات التّي يتم تعلمها في الماضي دون وجود المثيرات أو المواقف التّي أدت لحدوث التعلم . [1] في حين أن اختبار اختيار الاستجابة يقدم الاستجابة أو الخيارات الصحيحة والخاطئة لكلّ فقرة. وما على الطالب إلا أن يحدد أو يختار الاستجابة من عدة بدائل معطاة مثل الاختبارات من نوع الصح والخطأ، اختبارات المزاوجة، أسئلة الاختبار من متعدد.

6-2-6-اختبارات معيارية المرجع مقابل محكية المرجع: يصنف الاختبار بأنّه معياري المرجع إذا صمم ليزود الفاحص بعلامات تفسر بمقارنتها مع علامات مجموعة محددة جيدا في اختبار معين، وعلى هذا الأساس يمكن تفسير علامات الفرد باتخاذ مجموعته معيارا، فيمكن مقارنتها مع علامة كلّ فرد لتحديد موقعها النسبي أو رتبتها المئينية بين المجموعة كلها. وقد عرّفها أنانثي Ananthi (2020) بأنّه اختبار أولي لمقارنة انجاز المفحوص بمجموعة كبيرة من المفحوصين في نفس الصف، وتعرف المجموعة التمثيلية باسم المجموعة المعيارية، والاختبار المعياري المرجعي هو اختبار صميم لتوفير مقياس الأداء الذي يمكن تفسيره من حيث نسبة الفرد وموقعه في مجموعة معروفة ومحددة مسبقا. [2] (Ananthi, 2020, p. 606) (. أما الاختبار المحكي المرجع فيفسر في ضوء أداء محكي مستقل، وفي هذه الحالة يكون مستوى الأداء محدد وموصوفا بوضوح. وهذا المستوى من الأداء يعمل كمرجع مقابل كلّ علامة اختبار ستقارن به.

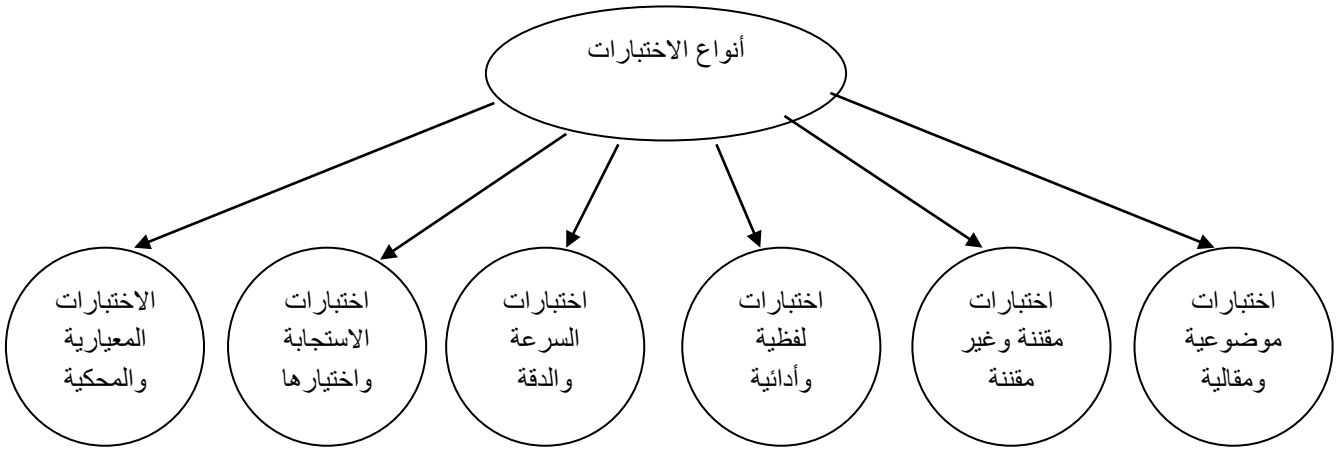
---

[1] عبد الناصر الجراح، و ميساء أبو أحمد: أثر طريقة تقديم المعلومات ونوعها وفترات الاحتفاظ في القدرة على التعرف والاستدعاء لدى طلبة جامعة اليرموك. المجلة الأردنية في العلوم التربوية . 11 (1). 2015، ص 45-60.

[2] Ananthi, B: Non- Referenced test, criterion- Referenced test and self- Referenced test Frameworks. journal of information and computation of sciences. 10 (5). 2020, p. 605-618.

ويعرف popham (1978) الاختبار المحكي بأنّه الاختبار الذّي يستخدم لتقدير أداء الفرد في نطاق سلوكي محدد تحديدا دقيقا. أما Gray (1978) فيعرفه بأنّه ذلك الاختبار الذّي يعدّ بهدف إحداث قياسات يمكن تفسيرها بشكل مباشر بناء على معايير (مستويات) أداء محددة. حسب ماورد في [1] . واختصارا لكل ما سبق فالاختبار المعياري المرجع هو الذي يكون الهدف منه مقارنة سلوك أو أداء الفرد بالمجموعة التي ينتمي إليها أو التي يحددها مطبق الاختبار، في حين أن الاختبار المحكي المرجع هو الذي يهدف إلى مقارنة الأداء أو سلوك الفرد بأداء أو معيار أو محك يكون قد حدده مطبق الاختبار مسبقا ويعتمد فيه أن يكون المحك دقيقا.

الشكل 02: أنواع الاختبارات حسب Gay (1980) **وأكابور، أهلاوات وآخرون (1985)**



المصدر: من إعداد الباحثة بالاعتماد على (ملحم، 2000) [2]

إن الهدف من عرض كلا النموذجين هو محاولة اختصار أنواع الاختبارات وتوضيحها بطريقة مبسطة، وكما هو واضح من النموذج أعلاه فإن أنواع الاختبارات في هذا التصنيف تختلف كليا

[1] ايمان عبد الكريم ذيب: أثر طول الاختبار على تحديد درجة القطع لاختبار تحصيلي محكي المرجع في مادة علم نفس الشخصية. *مجلة مداد الآداب. 1* (6). 2020، ص 330-408.

[2] سامي محمد ملحم، المرجع مذكور سابقا، ص 55.

عن أنواع الاختبارات في النموذج السابق، وهذا يدل على كثرة أنواع الاختبارات الأمر الذي يدفع بالباحثين إلى الاختلاف في التصنيف، فكل باحث أو مختص يعتمد في تصنيفه على أساس ما، فقد اعتمد Gregory (2014) في تقسيمه لأنواع الاختبارات على السمات التي يقيسها الاختبار، في حين اعتمد كل من Gay(1980) وأكابور، أهلاوات وآخرون (1985) في تقسيمهم على خصائص الاختبار من بنائه إلى كيفية تطبيقه وتصحيحه، وهذا لا يعني أن أحد النموذجين على خطأ، بل كل منهما يكمل الآخر، والغرض هنا هو محاولة عرض أكبر عدد من الأنواع بطريقة مختصرة. كذلك للتنويه هنالك العديد من التصنيفات الأخرى مثلا تصنيف أحمد (1960)[1] وتصنيف مجيد (2014) [2] كذلك تصنيف معمرية (2012) [3] التي لم تتطرق إليها الباحثة وبالطبع لكثرتها، فقد ارتأت أن الجمع بين هاذين النموذجين سيكون شاملا وكافيا.

7-أهمية الاختبارات النفسية

يعتبر الاختبار النفسي أداة أساسية من أدوات التشخيص النفسي ويستهدف الكشف عن قدرات وإمكانيات الفرد العقلية وكذلك ميوله واتجاهاته المختلفة. ومن هنا تبرز أهمية الاختبارات النفسية فحسب الخطيب و الخطيب (2010) فيريان بأنّها تبرز في نقاط تالية:*القياس النفسي هو أداة للحصول على عينة من سلوك الفرد في موقف معين وبذلك يمكن جمع بيانات عن هذا السلوك في

---

[1] محمد عبد السلام أحمد: القياس النفسي والتربوي . ط 1. القاهرة: ملتزمة الطبع والنشر مكتبة النهضة المصرية. 1960، ص 92

[2] سوسن شاكر مجيد: اسس بناء الاختبارات والمقاييس النفسية والتربوية. ط1. جامعة بغداد: مركز ديبونو لتعليم التفكير. 2014، ص 39

[3] بشير معمرية: *أساسيات القياس النفسي وتصميم أدواته*. جامعة الحاج لخضر، باتنة: دار الخلدونية للنشر والتوزيع. 2012، ص 112

أسلوب منظم وهذه البيانات لن يكون لها قيمة إذا اقتصرت دلالاتها على سلوك في الموقف المعين دون إمكانية التنبؤ عن سلوك الفرد.

* تمدنا الاختبارات بمعطيات قد لا يكون المفحوص واعيا بها أو غير قادر على التعبير عنها.

* تقييد البيانات الكمية والنوعية في تحديد درجة الفرد أو مستواه العقلي بالنسبة للمجموعة التّي ينتمي إليها كما أنّها تسمح للفاحص في تحديد الإطار المرجعي للمفحوص ومعرفة قدراته وحاجاته وأهدافه.

* تعتبر الاختبارات بأنواعها المختلفة والمتنوعة من أهم الأدوات المستخدمة في علم النفس للتوصل إلى فهم أعمق ومعرفة أكثر شمولا لشخصية الفرد. [1]

وعليه يمكن القول بأنّه تكتسي الاختبارات والمقاييس النفسية أهمية بالغة في التعبير عن القدرات الموجودة عند الأفراد، وفي عمليتي تصنيف المعلومات وتقييم البرامج المستخدمة في مجال التعليم أو غيرها. كذلك تساعد على التشخيص والتقييم والتنبؤ وغيرها من المجالات الواسعة التي يمكن للاختبارات المساعدة فيها.

8- خطوات بناء الاختبارات النفسية

حسب مجيد (2014) [2] تكمن خطة تصميم الاختبارات النفسية فيما يلي:

---

[1] محمد أحمد الخطيب، و أحمد حامد الخطيب: الاختبارات والمقاييس النفسية. عمان: دار ومكتبة الحامد للنشر والتوزيع. 2010، ص 19

[2] سوسن اكر مجيد، المرجع مذكور سابقا، ص 69

8-1-تحديد الهدف أو الأهداف الرئيسية من الاختبار

إن تحديد الهدف من الاختبار من حيث ميادين استخدامه والمجتمع المراد تطبيقه عليه، على جانب كبير من الأهمية فإنّ كان الهدف من استخدام الاختبار في الحصول على بيانات دقيقة كان لابد من استخدام الطرق الإحصائية والفنية التّي تحقق هذا الغرض، أما إذا كان المطلوب جمع بيانات سريعة أو إذا كان المفروض أن يعهد بالاختبار إلى فئة من المعلمين غير المدربين تدريبا كافيا على استخدام الاختبارات المقننة روعي في تصميم الاختبار سهولة تفسير معاييره حتى ولو أدت هذه السهولة إلى التضحية ببعض الدقة ذلك لأنّ تعقيد الاختبار من ناحية التعليمات أو التصحيح سيؤدي حتما إلى أخطاء أكبر. ويرى معمرية (2012) في هذا الصدد بأنّ بعد أن يحدد الأخصائي في القياس الخاصية التّي يريد قياسها، يحدد الهدف الذّي يسعى إلى تحقيقه من عملية القياس، ويعتبر تحديد الهدف الخطوة الحاسمة التي ينبغي الاهتمام بها عند التفكير في بناء وتصميم أداة قياس. [1] إذن تحديد الهدف هو أولى خطوات تصميم الاختبار إذ على الباحث أن يحدد أولا الخاصية التّي يريد قياسها من خلال الاختبار وهذا يتعلق بمشكلة البحث، ومن جهة ثانية أنّ الباحث لا يتجه نحو تصميم اختبار معيّن بدون أسباب قوية تدعوه لذلك فإن توفرت اختبارات تقيس السمة المراد قياسها ومكيفة بما يتوافق خصائص مجتمع البحث وتتوافر لها درجات عالية من الصلاحية فلا داعي لتضييع الوقت والجهد في بناء اختبارات جديدة.

8-2-تقرير محك أو معيار الدرجة

تفرض أهداف الاختبار نوع الإطار المرجعي الذّي تفسر في ضوئه درجة الفرد وما إذا كان يعد جيدا الأداء أم لا ويمكن التمييز هنا بين نوعين من المحكات:

---

[1] بشير معمرية، المرجع مذكور سابقا، ص 144

8-2-1-محكي المرجع: المحك المرجعي حسب أنانثي Ananthi (2020) يتم وفقه وصف أداء الطالب من حيث إتقانه للمحتوى، في مجال محدد جيدًا أو إنجازه فيما يتعلق بهدف تعليمي واضح من خلال مقارنة درجاته في الاختبار ببعض المعايير المحددة مسبقًا للنجاح.[1]. ويستخدم غالبا في اختبارات التحصيل والقدرات والمهارات، ويتقرر نجاح الفرد في الاختبار من خلال مقارنة الأداء وفق معيار محدد سابقا.

8-2-2-معياري المرجع: إن المحك المعياري المرجع يأخذ الصيغة الكمية في أغلب الأحوال مثل المتوسط الحسابي لدرجات عينة التقنين كمعيار لوصف الأداء العادي في الاختبار، وفي ضوئه تتحدد الأوضاع النسبية للأفراد فنقول أعلى من المتوسط أو أقل من المتوسط أو المتوسط. فالاختبارات ذات المرجع المعياري تختلف عن الاختبارات ذات المرجع المحكي، فحسب العساف والوزان (2014) فهي تلك الاختبارات التي تقوّم أداء الأفراد في ضوء معايير معينة بحيث تسمح بمقارنة أداء الفرد بأداء غيره من الأفراد في المستوى نفسه.[2] أي أن الدرجة الخاصة بالفرد لا تقارن وفق حجم سابق التحديد للمطلب موضوع الاختبار، وإنّما نقارن بالأداء الخاص ببقية الأفراد. والاختبار بعد هذا في حاجة للتجريب حتى تتلافى المشكلات التي يمكن أن تقع في مجال التجريب مثل تعديل بعض الأسئلة أو حذفها أو إضافة أخرى جديدة. وبالتالي فالدرجة على المقياس في حد ذاته لا معنى لها على الإطلاق إلا إذا نسبت إلى إحدى المحكين: المحك المرجعي بأن تنسب إلى الأداء نفسه بوصفه مرجعا للدرجة، أو المحك المعياري بأن تنسب إلى درجات بقية الأفراد بوصفهم معيارا لهذه الدرجة.

---

[1] Ibid, p 610

[2] نادية مصطفى العساف، و ختام محمد الوزان: أسس تصميم اختبارات اللغة العربية للناطقين بغيرها. مجلة العلوم الانسانية والاجتماعية .41 (1). 2014، ص 173-183.

8-3-ترجمة المفاهيم والأهداف إلى خصائص محددة

إن السمات المدروسة كالذكاء والانبساط وغيرها، هي تجريديات لخصائص مفترضة في الأفراد ونستدل على هذه الخصائص من وقائع سلوكية محددة سواء في شكل أفكار معبر عنها أو حلول لمشكلات أو استجابة لمنبهات محددة أو أساليب تعامل مع البيئة أو صفات شخصية تتسم بالاستقرار. ونقوم بترجمة هذه المفاهيم إلى خصائص محددة بصورة تسمح بصياغتها في تعاريف إجرائية مضبوطة. وحسب مورفي Murphy (1937) كما ورد في فرج (2017) أنّه يتيح هذا التعريف الإجرائي وضع منبهات لاستدعاء استجابات معينة هي التّي يقاس بواسطة كمها واستيفائها لمحك الدلالة طلاقة المفحوص، ولم يكن متاحا بالمثل قياس مفهوم الذكاء دون توافر التعريف الإجرائي الذّي وضعه بينية وترجمته إلى أعمال محددة يقوم بها الطفل. [1] إذن نحن بحاجة إلى ترجمة هذه المفاهيم لخصائص محددة تسمح بصياغتها بطريقة يمكننا بها إخضاعها لعملية القياس، بمعنى تحويل التعريف الإجرائي للسمة المراد قياسها إلى جملة من الواجبات والمتطلبات، بمعنى أنّ الباحث في هذه الخطوة يقوم بتحديد الأبعاد والمحاور الأساسية للمفهوم الذّي هو في صدد دراسته، وبعد ذلك ينتقل إلى الخطوة الموالية التّي تتطلب أكثر تفصيل وتحديد.

8-4-إعداد الاختبار للاستخدام

بعد أن ينتهي الباحث من جميع المراحل الأساسية لتصميم الاختبار يصبح من الضروري أن يضع الاختبار في صورته النهائية التّي سيقدم بها إلى المفحوصين، من ذلك ما إذا كانت فقراته ستقدم في شكل بطاقات مستقلة تتضمن كلّ بطاقة فقرة واحدة أو ستوضع الفقرات في كتيب مستقل على أن يرفق بها صحيفة منفصلة للإجابة كما يجب وضع تعليمات مناسبة للاختبار قبل تجربة

[1] صفوت فرج، المرجع مذكور سابقا، ص 137.

الاختبار. وفيما يلي بعض الأمور الواجب إتباعها عند وضع التعليمات للاختبار وتجربته حسب جلبي (2005):

– يجب شرح فكرة الاختبار شرحا دقيقا حتى لا يكون عدم فهم الأفراد لفكرة الاختبار سببا في الإخفاق فيه.

– يجب أن تبرز التعليمات الهامة وأن تكتب بخط واضح ومن المستحسن أن تكون حروفها أكبر من بقية التعليمات العادية.

– من الأفضل أن تحتوي التعليمات على مثال أو نوع من التمرين للمختبر قبل كل اختبار.

– يحسن أن تكون تعليمات الأجزاء المختلفة للاختبار متسقة أو موحدة إنّ أمكن فقد ثبت بالتجربة أن توحيد التعليمات يؤدي إلى دقة أكبر وإلى ألفة المختبر بالتعليمات وتمسكه بها، بينما اختلافها قد يؤدي إلى بعض الأخطاء أو إلى التضحية بالدقة.

– بالإضافة إلى أن التعليمات يجب أن تحتوي على طريقة الإجابة ومكانها وأسلوبها ويجب أن تحتوي أيضا على الزمن المطلوب مراعاته عند الإجابة إذا كان الاختبار يتطلب ذلك.

– من الأفضل إجراء تجربة استطلاعية على عينة صغيرة لغرض التّعرف على مدى الاختبار ومدى فهم العينة لهذه التعليمات ويقصد بمدى ملائمة الاختبار أي مدى وضوح العبارات والمدة اللازمة للإجابة على الاختبار ومن ثم تعديل فقرات الاختبار في ضوء التجربة الاستطلاعية.[1]

---

[1] سوسن شاكر الجلبي.: *أساسيات بناء الاختبارات والمقاييس النفسية والتربوية. ط1.* دمشق: مؤسسة علاء الدين للطباعة والتوزيع. 2005، ص 72

8-5-تقنين الاختبار

إن المرحلة الأخيرة من مراحل تصميم الاختبار هي التقنين وتتطلب عمليات التقنين إجراء بعض الدراسات الاستطلاعية على عينات محددة، بهدف الوصول إلى مؤشرات الصدق والثبات وغيرها. وقد يقوم الباحث بأكثر من تجربة استطلاعية لدراسة الصلاحية. والتقنين حسب امطانيوس (2016) يرى بأنّه يشير إلى توحيد الشروط المحيطة بالمفحوصين كافة وضبط العوامل والمتغيرات جميعها التي يمكن أن تؤثر في أدائهم الاختباري وتحييد أثرها إلى الدرجة القصوى، وذلك بهدف قياس المتغير الذّي صمّم الاختبار لقياسه بحجمه الحقيقي لدى تلك العينة من الأفراد التّي سيطبق عليها ودون أن يكون لأي من المتغيرات الأخرى أثره في أداء أفراد تلك العينة. [1] فالمقصود هنا بالتقنين جميع الخطوات الواجب إتباعها لضمان طرق توحيد وتصحيح الاختبار من جهة، والتحقق من صلاحيته من جهة أخرى.

9-الفرق بين الاختبار والاستبيان والمقياس

إن الاختبار حسب أوربينا urbina (2004) هو إجراء منهجي للحصول على عينات من السلوك ذات الصلة بالأداء المعرفي أو العاطفي، ولتسجيل وتقييم تلك العينات حسب المعايير.[2]

---

[1] نايف مخائيل امطانيوس، المرجع مذكور سابقا، ص 86.

[2] Urbina, S: *E ssentials of psychological testing.* canada: John Wiley & Sons,Inc. 2004, p 2

في حين يعرفه كوهين وسورديك Cohen & Swerdik (2018) بأنّه عملية القياس للمتغيرات المتعلقة بعلم النفس عن طريق الأجهزة أو الإجراءات المصممة للحصول على عينة من السلوك.[1]

أما المقياس حسب بوعطيط وبوحارة (2016) هي طريقة تسعى إلى الفهم ومعرفة شخصية الفرد، أو معرفة الإنسان لبعض الجوانب من شخصيته وسلوكه وقدراته العقلية وذلك من خلال الدلالات الكمية المعبرة عنها، بالإضافة إلى التحليل الكيفي لنتائج المقياس.[2] لهذا يرى تايلر(1988) بأنّه حتى الآن لم ننظر ما إذا كان هناك أي فرق بين مصطلح الاختبار ومصطلح المقياس. وعلى الرغم أنّ هناك تداخلا في المعنى إلا أنّهما ليسا مترادفين تماما. فمصطلح المقياس يستخدم في ميادين كثيرة من ميادين البحوث النفسية حيث لا يكون مصطلح الاختبار مناسبا للاستخدام. والاختبار يمكن أن يعرف على أنّه موقف مقنن صمم خصيصا للحصول على عينة من سلوك الفرد، وعندما يعبر عن هذه العينة بالأرقام، فمن الممكن أن نستخدم كلمة اختبار أو مقياس. وعلى ذلك فإنه بالرغم من أن التداخل بين هذين المفهومين ليس كاملا، فإنّه لا يزال في إمكاننا أن نقول معظم الاختبارات هي وسائل قياس، وأنّ معظم المقاييس النفسية يمكن أن تستخدم كاختبارات. [3]

أما الاستبيان حسب Good (1954) هو قائمة من الأسئلة تعد بشكل جيّد لمعرفة آراء ومعتقدات واتجاهات الآخرين نحو موضوع معين. في حين يرى فادالين (2003) بأنه من الوسائل

---

[1] Cohen, R. J., & Swerdik, M. (2018). Psychological testing and Assessment. Ed 9 united stastes of America: MC Graw- Hill Education. 2018, p 33

[2] جلال الدين بوعطيط، و هناء بوحارة: الاختبارات النفسية ودورها في تفعيل العمليات المهنية بمؤسسات العمل الواقع والأهمية. *المجلة العربية نفسانيات*. (50-51). 2016، ص 37-47.

[3] ليونا تايلر: الاختبارات والمقاييس. ط 3. بيروت: دار الشروق للنشر والتوزيع. 1988، ص 47.

العلمية الميسرة لتعريض المستجيبين لمثيرات واحدة ومختارة ومرتبة بعناية بقصد جمع البيانات اللازمة لإثبات صدق فرض ما أو رفضه. كما ورد في [1].

وفي الأخير فإن الفرق بين الاختبار والمقياس يصعب جدا تحديده أو الفصل فيه من حيث طريقة إعدادهما وتقنينهما وصلاحيتهما وغيرها، إلا أنه غالبا يعتبر المقياس أشمل وأوسع، في حين أن الاستبيان يعتبر استمارة مدروسة بعناية، إلا أنه لا يرقى إلى مستوى الاختبار والمقياس من حيث جودته وصلاحيته والثقة في نتائجه.

خلاصة

في هذا الفصل تّم التركيز على صلاحية الاختبارات النفسية، من حيث ماهيتها والدافع وراء إعادة التحقق منها ومهدداتها، وكذلك حاولت الباحثة عرض الاختبارات النفسية انطلاقا من التعريف بها والتفصيل في أنواعها وطرق إعدادها وصولا إلى الفرق بينها وبين المقياس والاستبيان، وهذا لما للاختبارات النفسية من دور بارز في القياس النفسي من حيث أنّها تعد من أهم وسائل جمع البيانات والمعلومات حول الأفراد، سواء بهدف تشخيصهم وتصنيفهم أو إصدار حكم على مستوى تقدمهم ومكانتهم النسبية في الجماعة المعيارية وغيرها.

---

[1] أحمد محمد سيد: طيف الاختبارات النفسية وإشكاليات القياس في الوطن العربي. مجلة دراسات نفسية وتربوية . 1 (1). 2008، ص 81-131.

الفصل الثالث

الصدق في

الاختبارات النفسية

تمهيد

لإعادة التحقق من صلاحية الاختبارات النفسية علينا الالتزام بإجراءات معينة ومحددة حسب نوع كل اختبار والغرض منه وغيرها، وهذا ما سنحاول التطرق إليه في هذا الفصل من خلال التفصيل في الصدق، وأهم أنواعه وإجراءاته، وكذا المعايير اللازم الرجوع إليها لاختيار النوع أو الدلالة المعتمدة لإثبات صلاحية الاختبار للغرض المرجو منه وغيرها.

1-مفهوم الصدق

إنّ التعريف القاموسي لمصطلح "الصدق" كما ورد في المعجم الوسيط هو: "الأمر الصالح لا شية فيه من نقص أو كذب" [1] أما كما ورد عند هوبيت وكرامر Howitt & Cramer (2016) فهو: "ما إذا كان شيء ما متأسس جيدا، أو منطقي أو يمكن تبريره" [2] وانطلاقا من هذا التعريف وغير بعيد عن نفس المعنى والفكرة أورد المختصين في القياس عدة تعريفات لصدق الاختبارات النفسية نذكر منها:

تعريف كرونباخ Cronbach (1971) للصدق على أنه "العملية التي عن طريقها يجمع مستخدم الاختبار أو مطوره؛ الدلائل من درجات الاختبار والتي يدعم من خلالها استدلالاته.

---

[1] إبراهيم أنيس، عبد الحليم منتصر، عطية الصوالحي، و محمد خلف الله: *المعجم الوسيط* . ط1. القاهرة: مجمع اللغة العربية-مكتبة الشروق الدولية.2004، ص 510.

[2] دينيس هوبيت، و دونكان كرامر: *مقدمة لطرائق البحث في علم النفس*. ترجمه: صلاح الدين محمود علام. دار الفكر. عمّان، الأردن. 2016، ص 400.

وعند التحقق من الصدق أو الصلاحية، يجب تحديد الاستدلال المطلوب، لكي يتم جمع الأدلة المناسبة لهذا النوع من الاستدلال " [1] إذن هذا يعني أنه يتم ضمان صدق الاختبار أثناء بنائه وتطويره، فمطور الاختبار يعتمد على جمع الاستدلالات التي تدل أن الاختبار يقيس ما يريده أن يقيسه وهو في صدد البناء، فالصدق حسب كرونباخ هو عملية تبدأ من أول خطوات بناء الاختبار ولا تنتهي بانتهاء ذلك البناء بحيث تظل تلك العملية حتى بعد استخلاص درجات الاختبار، أي أنه يجب تحديد نوع الصدق المرغوب استخدامه ومن ثم جمع الاستدلالات اللازمة له.

كما عرّفه ميسك Messick (1995) بأنه "حكم تقييمي شامل للدرجة التي يُتحصل عليها من الاختبار والتي تؤيّدها الأدلة التجريبية والنظرية، أي من حيث ملائمة الأدلة وكفايتها، اعتمادا إلى درجات الاختبارات وغيرها من التقييمات" [2] . وعرّف رينولدز Reynolds (1998) صدق درجات الاختبارات بنفس الطريقة، وجادل بأن "الصدق يشير إلى ملاءمة ودقة تفسير الأداء على اختبار معين، ويعبّر عادة عن ذلك الأداء بدرجة الاختبار". وقد اتفق المختصين في القياس على أن الصدق لم يعد يتعلق بالاختبار، بل أصبح مرتبط بمدى ملاءمة ودقة تفسير درجة الاختبار، لذلك فالتحقق من صدق الاختبار يعد مسؤولية مشتركة بين مطور الاختبار ومستخدمه[3].

---

[1] Cronbach, L : Test Validation. *Educational Measurement* . 1971.p, 443-507.

[2] Messick, S: Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist . 50*.1995, p 741-749.

[3] سيسيل رينولدز، و رونالد ليفينجستون: تأليف *إتقان القياس النفسي الحديث النظريات والطرق*. ترجمه: صلاح الدين محمود علام. دار الفكر ناشرون وموزعون. عمان، الأردن. 2013، ص 216

كذلك أضاف عبد الخالق (2000) في تعريفه للصدق بأنه "يشير إلى مدى صلاحية الاختبار وصحته في قياس ما يعلن أنه يقيسه فيدلنا صدق الاختبار إذن عن أمرين أساسيين هما: ما الذي يقيسه الاختبار؟ وكيف ينجح في قياسه؟ وليس لذلك علاقة باسم الاختبار بل بمضمونه". [1] . وعليه فالصدق حسب عبد الخالق هو دليل على صحة الاختبار في أنه يقيس ماوضع لقياسه، غير أنه أضاف في التعريف تلميحا إلى عدم الاكتفاء بالصدق الظاهري بل وجب الذهاب إلى ماهو أعمق من ذلك. أماّ حسب علام (2000) فهو يشير إلى الاستدلالات الخاصة التي تعطينا إياها درجات الاختبار من حيث مناسبتها ومعناها وفائدتها وتحقيق صدق اختبار معين معناه قيام الباحث بتجميع الأدلة التي تؤيد هذه الاستدلالات، ولذلك يشير الصدق في إحدى معانيه إلى مدى صلاحية استخدام درجات الاختبار في القيام بتفسيرات معينة. [2] وعليه فالصدق من أهم الخصائص السيكومترية للاختبار لأنه لا يمكن الوثوق في نتائج اختبار مشكوك في أنه يقيس السمة التي وضع لقياسها، فالصدق هو تحقيق الاختبار للغرض الذي وضع لأجله.

وغير بعيد عن التعريفات السابقة عرّفت الجمعية الأمريكية للبحث التربوي American Educational Research Association (2006) والتي يرمز لها اختصارا "AERA" الصدق بأنه "مدى دعم الأدلة النظرية في تفسير درجات الاختبار، والتي تتحدد وتختلف وفق استخدامات معينة للاختبار". [3] كما هو واضح من خلال مجموعة التعريفات السابقة فإن الصدق لم يعد يتعلق

---

[1] أحمد محمد عبد الخالق: *استخبارات الشخصية* . ط3. القاهرة: دار المعرفة. 2000، ص 131

[2] صلاح الدين محمود علام: *القياس والتقويم التربوي والنفسي ،أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة. دار الفكر العربي. 2000، ص 187.

[3] AERA :Standards for Reporting on Empirical Social Science Research in AERA Publications .American Educational Research Association. Educational Researcher. 35 (6).2006, p 33-40.

بالدرجة الكمية الممنوحة لصدق الاختبار بل أصبح يرتبط بمجموعة من الأحكام والاستدلالات التي من خلالها يمكن استنباط مدى دقة التفسيرات المقدمة لنتائج الاختبار، فالصدق خاصية تعنى بتفسير وتبرير دقة وملائمة النتائج أكثر مما هي مرتبطة بالاختبار في حد ذاته ويمكن التعبير عن تلك التفسيرات والتبريرات بدرجة كمية أو بأدلة نظرية والأفضل اعتماد كلا منهما.

هذا وقد لخص معمرية (2012) مفهوم الصدق في ثلاثة نقاط أساسية وهي:

*أن يكون الاختبار قادرا على قياس ما وضع لقياسه.

*أن يكون الاختبار قادرا على قياس ما وضع لقياسه وفقط.

*أن يكون الاختبار قادرا على التمييز بين طرفي الخاصية. [1] والواضح من تعريف معمرية هو أنه أضاف إلى التعريفات السابقة نقطة جديدة وهي قدرة الاختبار على التمييز بين طرفي السمة، وكما هو معلوم فهي طريقة من طرق استخراج الصدق إلا أنه وجب الاشارة هنا إلى عدم الاكتفاء بهذه الطريقة لتعليل صدق الاختبار، فغالبا ماتعتبر هذه الطريقة تحصيل حاصل، وبالتالي وجب إثبات صدق الاختبار بطرق أخرى إضافة إلى هذه الطريقة.

وأضاف كل من المجلس الوطني للقياس في التربية (National Council on Measurement in Education) والتي تكتب اختصارا **NCME** والجمعية الأمريكية لعلم النفس(American Psychological Association) واختصارها **APA** والجمعية الأمريكية للبحث التربوي (American Educational Research Association) واختصارها **AREA**

[1] بشير معمرية: *أساسيات القياس النفسي وتصميم أدواته*. جامعة الحاج لخضر، باتنة. دار الخلدونية للنشر والتوزيع. 2012، ص 181.

وثيقة أو دليل يتضمن معايير الاختبارات النفسية والتربوية وأطلق عليها "وثيقة المعايير" ويؤكد هذا الدليل بأنه تختلف طرق التحقق من الصدق باختلاف أغراض الاختبار. ومن أهم هذه الأغراض:

*تشخيص أداء الفرد الحالي على الاختبار الذي يتميز بأنه شامل للنطاق السلوكي للسمة المراد قياسها.

*التنبؤ بأداء الفرد التالي أي أداءه في المستقبل سواء في السمة التي يقيسها الاختبار أو سمة مغايرة لها.

*تحديد درجة تملك الفرد للسمة المراد قياسها. [1] وفي الأخير فالصدق كما عرفته مبادرة كوسمين COSMIN بأنه الدرجة التي يقيس بها الاختبار البنية أو التركيبات التي يزعم قياسها. [2]

إذًا فالصدق في الأخير هو العملية التي يقوم من خلالها صانع الاختبار أو مطوره وحتى مستخدمه من إثبات صحة نتائج هذا الاختبار وصلاحيتها لذلك الاستخدام وحتى درجة الثقة فيها، فالصدق هو دليل على صلاحية هذا الاختبار ودليل على أنه يقيس السمة التي وضع لقياسها بالاضافة الى قياس هذه السمة وفقط.

2-أنواع الصدق

إن خاصية الصدق في الاختبارات النفسية تعتمد على أساس منطقي للبرهنة عليها أكثر من اعتمادها على الإحصائيات إن صح التعبير، على عكس الثبات الذي يعتبر ذو أساس إحصائي محض. ولنتمكن من التعرف على مدى صدق الاختبارات نحن مجبرون على جمع مجموعة من

---

[1] AREA, A. E., APA, A. P., & NCME, N: C*Standards for Educational and Psychological Testing.* Washington: American Educational Research Association.2014, p 13

[2] Ibid, p 1178

الأدلة والبراهين المنطقية والتي تؤكد لنا مدى تمكن الاختبار من قياس السمة التي وضع لقياسها والتمييز بين طرفي السمة وغيرها، وهذه الأدلة هي ما أطلق عليها فيما سبق أنواع الصدق إلى حين ظهور النظرية الحديثة للصدق والتي جمعت تلك الأنواع المتعددة تحت سقف واحد أطلقت عليه أدلة الصدق بعدما تمايزت أنواعه وتعددت أسماؤه في نظرية الصدق الكلاسيكية إن صح التعبير، حتى أصبح يصعب جدا معرفة ما الهدف من كل نوع وما أهميته وما الدافع إلى استخدام هذا النوع من الصدق دون الآخر. فهنالك العديد من أنواع الصدق المذكورة في الدراسات، وعلى سبيل المثال لا الحصر سنحاول ذكر البعض منها مثل: الصدق الظاهري، صدق المحتوى، صدق المحك–أي الصدق التلازمي والصدق التنبؤي– صدق التكوين الفرضي، الصدق التباعدي، الصدق التقاربي، الصدق الداخلي، الصدق الخارجي، الصدق التمييزي، الصدق العاملي، الصدق المنطقي وغيرها الكثير الكثير من التسميات، لدرجة أصبح التمييز بين هذه الأنواع المختلفة للصدق عبر الأعوام غامضا في المراجع والدراسات.

ولهذا ذكر تيغزة (2008) كيف أن الدليل الإرشادي الأول لخص هذه الأنماط في أربع أنواع من الصدق: صدق المحتوى والصدق التنبؤي والصدق التلازمي، وأخيرا صدق المفهوم أو التكوين الفرضي أو البنائي، وقلص الدليل الإرشادي الصادر الثاني هذا التصنيف الرباعي إلى تصنيف ثلاثي للصدق إذ أبقي على صدق المحتوى، وصدق المفهوم، أما الصدق التنبؤي والصدق التلازمي فأدرجا تحت مسمى :الصدق المعياري، كما اتسمت أيضا بالتركيز على التمايز الوظيفي لأنواع الصدق، أي أن لكل نوع وظيفة خاصة به تختلف عن وظائف الأنواع الأخرى من الصدق. وترتب عن ذلك أن كل نوع يناسب بعض أنواع الاختبارات والمقاييس ولا يناسب الأنواع الأخرى. فصدق المضمون يناسب اختبارات التحصيل، والصدق المحكي يناسب اختبارات القدرات

والاستعدادات، أما صدق المفهوم يناسب اختبارات الشخصية.[1] وهذا ما دفع بنظرية الصدق الحديثة إن صح التعبير إلى توحيد هذه الأنواع تحت نوع واحد والذي يتضمن ويمثل جميع الأنواع السابقة، والتي اعتبرتها وثيقة المعايير أدلة وبينات صدق المفهوم أو ما يسمى أيضا بصدق التكوين الفرضي.[2]

إلا أن النقاش الذي مازال قائما إلى الآن والذي ينادي به أصحاب النظرية التقليدية للصدق ويطالبون به هو: ما الجديد من الناحية التقنية والإجرائية الذي قدمته نظرية الصدق الحديثة؟ غير توحيد المفاهيم الموجودة سابقا وإعادة بلورتها تحت أشكال تبدو في مظهرها جديدة إلا ان لبّها لايختلف عن سابقه، وهذا ماسنحاول التفصيل فيه وتوضيحه من خلال عرض أنواع الصدق في مقابل أدلة الصدق أو البينات. وقد تم محاولة الفصل بينهما بالاعتماد على التاريخ الزمني، وذلك انطلاقا من التاريخ الذي أعلنت فيه "وثيقة المعايير الخامسة" عن وجود نوع واحد من أنواع الصدق وأن البقية هي بينات لهذا النوع وحددت تلك البينات، وكان ذلك سنة 1999. مع الإشارة إلى أنه في الوثيقة السابقة أي الرابعة سنة 1985 كانت قد عمدت فيها إلى تسمية أنواع الصدق بجوانب الصدق وذكرت فيها أن الصدق هو وحدة واحدة، إلا أن تلك الوثيقة تعتبر غير واضحة ولم تفصل في الرؤية الجديدة التي تتبناها وثيقة المعايير حاليا.

وسنتطرق هنا بشيء من التفصيل إلى ثلاث أنواع متمايزة من الصدق قبل التطرق إلى إجراءتها، وسنحاول بعد ذلك التعريج على عدة أنواع أخرى ذكرت في الدراسات والتي في الحقيقة إما أنها إجراءات للأنواع الأولى أو أنها هي نفسها فقط إلتبست على الباحثين من كثرة اختلاف

---

[1] امحمد تيغزة: توجهات حديثة في تقدير صدق وثبات درجات أدوات القياس: تحليل نظري تقويمي تطبيقي. مجلة العلوم النفسية والتربوية .4(1) .2017، ص7–29.

[2] Ibid, p 10

ألقابها، والدافع وراء هذا التعريج هو كثرة استخدام الباحثين لهذه الألقاب دون وعي منهم بماهيتها وإجراءتها. بالإضافة إلى وجود هذه التسميات في العديد من الرسائل التي تم تقييمها في هذه الدراسة وبالتالي وجب علينا التعريف بها والفصل فيها.

2-1-صدق المحتوى

يطلق على صدق المحتوى حسب هيلمستادتر Helmstadter (1966) اسم صدق المضمون أو الصدق بحكم التعريف أو صدق عينة الاختبار، وهذا المعنى الأخير هو أقرب المعاني المقصودة، فالاهتمام الأساسي ينصب هنا على ما إذا كان مجال سلوكي معين ومحدد بشكل دقيق ممثلا في شكل مجموعة من البنود بصورة مناسبة أم لا [1].

فصدق المحتوى الجيد حسب هوييت وكرامر Howitt & Cramer (2016) يعتمد على البناء الجيد لنطاق متسع من المفردات. وتجميع هذه المفردات معا بعناية كي تعكس أنواعا متعددة من أوجه المفهوم المراد قياسه. واستخدام وسائل متنوعة لتكوين مفردات يمكن تضمينها يعد أمرا مهما. وهذه الوسائل ربما تشتمل على أدبيات البحوث، والمقابلات مع أناس مماثلين للمشاركين المحتملين، ونظرية متأسسة في المجال وغير ذلك. وبالبحث عن مفردات من نطاق متسع، يمكن أن يثري صدق محتوى المقياس. [2].

وأضاف كل من جاي، ميلز وإيرسيان Gay, Mills & Airasian (2012) بأن الصدق "يتطلب كلا من صدق المفردة وصدق المعاينة، ويتعلق صدق المفردة بما إذا كانت مفردات

---

[1] صفوت فرج، المرجعمذكور سابقا، ص 266.

[2] هويت وكرامر، المرجع مذكور سابقا، ص 402.

الاختبار موائمة لقياس مجال المحتوى المرجو. ويتعلق صدق المعاينة بجودة معاينة الاختبار للمجال الكلي للمحتوى موضع القياس". [1]

غير أن المطّلع على الدراسات والرسائل الأكاديمية –على سبيل المثال لا الحصر دراسة عائشة الشامسي (2005) ودراسة ديانا ابزاخ (2019)– يجد صيغ وإجراءات مختلفة نوعا ما لما يشكله صدق المحتوى، فقد تم حصر مفهوم صدق المحتوى في زاوية مرتبطة بصدق المفردة والمحكمين، بحيث يقدم الخبراء آراءهم حول المفردات إن كانت تمثل المجال المراد قياسه وقادرة على تغطيته. وهذا يعد نظرة قاصرة، فهو يركز على جانب من جوانب صدق المحتوى وأحد مظاهره وسبله فقط، بل هكذا أصبح هذا الصدق وكأنه الصدق الظاهري، فمن الخطأ التركيز على جانب واحد فقط في حين إهمال عدة إجراءات تقودنا إلى المعنى الحقيقي لصدق المحتوى.

2-2-الصدق المرتبط بالمحك

يتمثل الصدق المرتبط بالمحك أو الصدق المحكي أو المعياري وغيرها من الألقاب في شكلين أساسيين وهما الصدق التنبؤي والصدق التلازمي. ويعرف بصفة عامة حسب بوليت polit (2015) بأنه "اختبار ما إذا كان القياس متوافقا مع المعيار الذهبي الذي تم تطبيقه في نفس الوقت أو في وقت لاحق". [2] كذلك هو "الاستدلال على صدق أداة مجهولة الصدق بمقارنتها بأداة أخرى معلومة الصدق، ويتم ذلك عندما يكون هنا كعلاقة قوية بين نتائج استخدام الأداتين". [3]

---

[1] ل جاي، ج ميلز، و ب ايرسيان: البحث التربوي كفايات للتحليل والتطبيق. ترجمه: صلاح الدين محمود علام. دار الفكر ناشرون وموزعون. الأدرن. 2012، ص 252.

[2] Polit, D., & Yang, F: *Measurement and the measurement of change: a primer for the health professions.* Philadelphia: Wolters Kluwer.2016, p 1748

[3] إحسان الأغا: *البحث العلمي"منهجه-عناصره-أدواته"*. غزة: مطبعة مقداد. 1997، ص 104.

إذن في هذا النوع من الصدق يتم الاستعانة بمحك خارجي، مستقل عن الاختبار، يستخدم في تقويم صدق ذلك الاختبار. وقد يكون هذا المحك اختبارا آخر للخاصية التي أعد الاختبار الجديد لقياسها، وقد يكون دلائل كمية أخرى يمكن الاعتماد عليها مثل تقديرات الآخرين للخاصية. والمحك بصفة عامة عبارة عن عامل مستقل عن الاختبار، نعلم أو نفترض بأنه مرتبط بالخاصية التي نريد قياسها بالاختبار.

وكما أشار العجيلي (2017) بأن هذا النوع من الصدق يعتمد على الإجراءات العملية في تطبيق الاختبار مع المحك، فإنه يعتمد كذلك على مؤشرات إحصائية يعبّر عنها بمعامل الارتباط بين علامات الاختبار وعلامات المحك المستخدم.[1] (العجيلي، 2017، ص 255)

فالصدق المرتبط بالمحك كما هو واضح وجلي من خلال تسميته، هو صدق يعتمد أو مرتبط بمحك آخر، سواءً كان هذا المحك اختبار أو خبير أو إحصائيات وغيرها من الأمور التي قد تعتبر محكا، كذلك يشترط أن يكون هذا المحك صادق هو الآخر وثابت، أيضا وجب الإشارة إلى أن هذا النوع من الصدق رغم أن ظاهره سهل بسيط في تطبيقه وإجراءاته، إلا أن الباحث قد يقع في عدة أخطاء عند استخدامه له دون الانتباه إليها، وأهم هذه الأخطاء استخدام اختبارات تقيس نفس السمة واعتبارها محكا في حين يشترط هذا النوع من الصدق أن يكون المحك معيارا ذهبيا، وليس أي معيار يقيس سمة مشابهة أو معاكسة للسمة المراد دراستها لأن ذلك الإجراء مرتبط بصدق المفهوم وهو ما يعرف بالصدق التقاربي والصدق التباعدي. والمعيار الذهبي حسب بوليت polit (2015)"هو المعيار الأمثل الذي يمكن أن نقيم على أساسه الدرجات في الصدق المحكي" ومع ذلك، زعم العديد من الباحثين الذين يفتقرون إلى المعيار الذهبي بشكل غير ملائم أنهم قاموا بتقييم

---

[1] صباح حسين العجيلي: *اساسيات في القياس والتقويم* . ط 1. عمان، الاردن: دار وائل للنشر والتوزيع.2017، ص 255

الصدق المحكي. فعندما يدرس الباحثون العلاقة بين الدرجات على مقياس محوري وعشرات من الاختبارات الأخرى التي تعتبر ذات صلة، فإنها تقيم الصدق التقاربي وليس الصدق المحكي. [1]

وهذا ما أشارت إليه العديد من الدراسات فمثلا قد أعرب باحثون من مجموعة COSMIN عن اعتقادهم بأن النتائج التي يبلغ عنها مستخدم الاختبار عند التحقق من الصدق المحكي "تفتقر غالبا إلى المعيار الذهبي" [2] ومع ذلك، يبقى تحديد المعيار الذهبي ممكن غالبًا رغم صعوبة الأمر، فبعض النتائج المبلغ عنها من قبل مستخدم الاختبار كمثال في دراسة بوليت ويانغ Polit & Yang (2016) عند التحقق من صدق مقياس التقرير الذاتي في التفكير الانتحاري، تم استخدام التشخيص من طبيب كمعيار ذهبي في تقييم الصدق التلازمي. كما تم تقييم مستوى نية الشخص في التصرف بطريقة معينة (على سبيل المثال، الإقلاع عن التدخين) مقابل السلوك اللاحق الفعلي (مثل التدخين) في تقييم الصدق التنبؤي. [3]

ولهذا النوع من الصدق نوعان فرعيان هما الصدق التلازمي والصدق التنبؤي. سنوضح كل منهما.

---

[1] Polit: Assessing measurement in health: Beyond reliability and validity,International. Journal of Nursing Studies .52. 2015. p 1746-1753.

[2] Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., et al: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of Clinical Epidemiology. (7).2010 ,p 737-745.

[3] Polit, D., & Yang, F: Measurement and the measurement of change: a primer for the health professions. Philadelphia: Wolters Kluwer.2016,p 25

2-2-1-الصدق التلازمي

ويعني كشف العلاقة بين الاختبار ومحك تجمع البيانات عليه أثناء أو قبل إجراء الاختبار، أي أننا نقارن بين درجات الأفراد على الاختبار ودرجاتهم على مقياس موضوعي آخر بحسب مركزهم فيما يقيسه الاختبار، على أن يعطى الأفراد درجات على المحك في نفس الوقت الذي نطبق فيه الاختبار.

وهو حسب علام (2006) "يرتبط بدرجة اقتران بين تباين درجات اختبار ما بتباين درجات اختبار آخر يطبق في الوقت نفسه تقريبا، أي يهتم بالوصف، ويستخدم الصدق التلازمي بدرجة أكبر في التدريس ومجالات الصناعة وانتقاء الأفراد والتشخيص الإكلينيكي." [1]

التلازمي هو الصدق الذي يهتم بالعلاقة بين الاختبار والمحك أي كان نوعه، والمهم هنا أنه في هذا النوع من الصدق يطبق الاختبار والمحك في وقت واحد أي لايفصل بينهما فاصل زمني طويل إن صح التعبير فهما متلازمان، على عكس ماسنلاحظه في الصدق التنبؤي.

2-2-2-الصدق التنبؤي

يجدر بنا الإشارة أنه في هذا النوع من الصدق نحتاج إلى فاصل زمني بين تطبيق الاختبار والمحك على عكس الصدق التلازمي، كذلك هنالك أمر آخر نفرق به بين الصدق التلازمي والتنبؤي عدا عامل الزمن وهو الهدف من تطبيق الاختبار فكما هو واضح يشترك كل من الصدق التنبؤي والتلازمي في ارتباطهما مع المحك لكنهما يختلفان في عامل الزمن والهدف من الاختبار. فالصدق التلازمي يستخدم عادة لغرض التقييم الحالي والتشخيص وغيرها في حين أن الصدق التنبؤي كما هو جلي في اسمه يستخدم للتنبؤ غالبا. وهذا بالضبط ما أوضحتاه أنستازي وأوربينا

---

[1] صلاح الدين محمود علام: *الاختبارات والمقاييس التربوية والنفسية* (ط1). الأردن: دار الفكر.2006، ص 208

Anastasi & Uraina (2015) بأن اختيار إجراء معين للتحقق من الصدق يعتمد على الاستخدام المرجو من درجات الاختبار. فالاختبار نفسه عندما يستخدم لأغراض مختلفة ينبغي أن يتم التحقق من صدقه بطرق مختلفة. [1] ولهذا عرف معمرية الصدق التنبؤي (2012) على أنه "مدى قدرة الاختبار أو الاستبيان على التنبؤ بنجاح الفرد في مهنة معينة أو في أداء معين، كمؤشر على صدق الاختبار"[2]

فالاختبار الواحد إذن تتعدد أدلة وبينات صدقه بتعدد أغراض استخداماته والاستنتاجات المرجوة منه، وبالتالي فالاختبار الذي يتصف بدرجة مقبولة من الصدق لغرض ما ربما لا يكون كذلك لغرض آخر. كذلك استخدام هذا النوع من الصدق المرتبط بمحك لا يخلو من بعض المشكلات والأخطاء الشائعة التي يقع فيها الباحثين وأغلب هذه الأخطاء تتمحور حول: تحيزات المحك التي تؤثر في صلاحيته وحجم العينة المرتبط بالمحك.

وملخص القول هنا لا يمكن العثور على "المعيار الذهبي" بسهولة في جميع السمات. كما أن التغلب على توقع اختبار يعتبر معيارا ذهبيا يشكل تحدياً. ويتوقع الباحث على الأقل أداة لها بعض المزايا على المعيار المختار، إما لأنها أسهل في الاستخدام، أو أن لديها وقت إدارة أقل أو حتى لأنها خفضت التكلفة. كذلك ليس كل اختبار يقيس نفس السمة معناه يصلح ليكون "المعيار الذهبي" وبالطبع الاختبارات التي تقيس سمات مغايرة لا تصلح لهذا النوع من الصدق، فلا يجب الخلط بين الصدق المحكي والصدق التباعدي والتقاربي.

---

[1] أنستازي واوربينا، المرجع مذكور سابقا، ص 178.

[2] بشير معمرية ، المرجع مذكور سابقا، ص 215.

2-3-صدق التكوين الفرضي

إن فكرة صدق التكوين الفرضي تعتبر حديثة إذ ما قارنها بالأنواع الأخرى للصدق وحسب كرونباخ وميهل Meehl & Cronbach (1955) يعتبر صدق البناء للاختبار بأنه الدرجة التي يعمل الاختبار على قياس خاصية أو سمة فرضية صمم أساسا لقياسها. [1] والتكوينات الفرضية هي أقسام متسعة تشتق من المظاهر المشتركة التي تتقاسمها متغيرات سلوكية قابلة للملاحظة المباشرة. وكينونات نظرية ليست قابلة للملاحظة بذاتها والاهتمام بالتكوينات الفرضية أدى إلى تقديم ما اعتبر في البداية نوعا ثالثا من صدق الاختبارات والذي أطلق عليه صدق التكوين الفرضي. (أنستازي و أوربينا، 2015، ص 150) والمعروف أن صدق البناء كباقي أنواع الصدق هو مجمل إجراءات مستمرة ولا تنتهي. وتضع لوفينجر Loevinger (1979) كما ورد في النبهان (2013) صدق البناء في إطارين هما:

-إلى أي حد يقيس الاختبار خاصية أو سمة لها وجود فعلي؟

-إلى أي حد يكون هناك تناظر بين التفسير المقترح للسمة أو للخاصية وما يقيسه الاختبار فعلا؟

ويتصل الإطار الأول بصدق الاختبار كما تبين سابقا، في حين يعرف الثاني بصدق التفسير. ولا شك أن التساؤلين يكمل بعضهما الآخر إذ يقدم الإطار الأول أدلة حول مقدار ونوع الارتباطات بين فقرات الاختبار، وهذا يضفي معنى سيكومتري على الاختبار. في حين يزود الإطار الثاني معنى عملي أو مهني (سيكولوجيا كان أو تربويا) للاختبار. [2]

---

[1] Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin . 52* (4). 1955, p 281-302.

[2] موسى النبهان: *أساسيات القياس في العلوم السلوكية* (الإصدار ط2). عمان، الأردن: دار الشروق للنشر والتوزيع.2013 ، ص 358

ويجدر بنا الإشارة إلى ما قام به فريق COSMIN في بالتمييز بين جوانب من صدق البناء، وهي: الصدق الهيكلي، واختبار الفرضيات، والصدق عبر الثقافات. على أن يقيَّم الصدق الهيكلي (اختبار بنية الاختبار مثل الاتساق الداخلي والتحليل العاملي) فقط بالنسبة للاختبارات ذات الأبعاد المتعددة، في حين أن الجوانب الأخرى لصدق البناء مطلوبة بالنسبة لجميع الاختبارات. أي أنها تستخدم لتحديد أو تأكيد الجداول الفرعية القائمة، التي تستخدم فيما بعد في الفرضيات التي يجري اختبارها؛ وينبغي أن يتم تقييم الصدق عبر الثقافات فقط للأدوات المترجمة.[1] وهذه الجملة الأخيرة لا يمكن قبولها دون التعقيب عليها فربط الصدق عبر الثقافات فقط باللغة هو أمر قاصر فقد أهملت هنا دراسة COSMIN العديد من الجوانب الأخرى التي قد تؤثر على هذا النوع من الصدق ماعدا اللغة مثل الدين والعادات وغيرها، وبالتالي فهذا النوع من الصدق مرتبط بجميع الاختبارات التي يمكن أو يشك الباحث في أنها تحتمل أن تكون متحيزة.

2-4-أنواع الصدق المذكورة في الدراسات

عمدت الباحثة ذكر هذه الأنواع من الصدق كمحاولة للفصل وتوضيح الخلط الواقع بين أنواع الصدق الفعلية وهذه التسميات. والجدير بالذكر أن هذه التسميات لا يمكن اعتبارها من أنواع الصدق إنما هي إجراء من إجراءاته في أحسن الأحوال، ولا يمكن الاكتفاء بها وحدها كإجراء بل غالبا ما تستلزم معها إجراءات أخرى عند التحقق من الصدق.

2-4-1-الصدق الظاهري وصدق المحكمين

إن ما يطلق عليه الصدق الظاهري هو في الحقيقة قراءة ظاهرية للاختبار من ناحيته الشكلية فقط، بحيث يقوم مستخدم الاختبار بفحص مفردات الاختبار من أجل تقييم ما إذا كانت تقيس ما

---

[1] Ibid, p 740.

وضع لقياسه. وبالطبع فهو لا يصلح مع الاختبارات التي مفرداتها ضمنية ولا تعكس ما يقيسه الاختبار مثل اختبارات الشخصية. كذلك وجب التنويه إلى أن الصدق الظاهري هو إجراء من إجراءات صدق المحتوى، فهذا الأخير يتم الحصول عليه من خلال تحليل منظم مدروس وفني لمحتوى الاختبار، إذ حسب رينولدز وليفينجستون Reynolds & Livingston (2013) ربما يبدو "اختبار معين صادقا ظاهرياّ لعامة الناس ولكنه لا يكون صادقا عند فحصه فحصا منظما يتضمن تحليلا فنيا لمحتواه". [1]

وعلينا أن نضع في اعتبارنا في نهاية الأمر أن الصدق الظاهري إجراء ضروري في بعض الاختبارات، لكنه أبدا لا يكفي بمفرده لتقدير صدق الاختبار. ووجب التأكيد على عدم الخلط بين صدق المحتوى والصدق الظاهري فالصدق الظاهري لا يعد صدقا بالمعنى التقني، فهو لا يدل على ما يقيسه الاختبار حقاً، وإنما يدل على ما يبدو ظاهريا أنه يقيسه.

وهو الأمر نفسه الذي ينطبق على صدق المحكمين فيعتبر هو الآخر صدق ظاهري لأنه يعتمد نفس المبدأ مع إجراء إضافي وهو حساب نسبة الاتفاق بين المحكمين، ورغم أن هذا الصدق أو بالأصح الإجراء أكثر دقة من الصدق الظاهري إلا أنه لا يزال قاصرا عن تحقيق هدف الصدق وبالتالي لا يرقى إلى أنواع الصدق بل هو إجراء من إجراءات صدق المحتوى مثله مثل الصدق الظاهري وبالتالي لا يجب الاعتماد عليه دون إضافة إجراءات أخرى في التحقق من الصدق.

2-4-2-الصدق الذاتي

يعرف الصدق الذاتي بأنه الجذر التربيعي لمعامل الثبات، أي أن هذا النوع من الصدق يعتمد اعتمادا كليا على معامل الثبات، لذلك أشار فرج (2017) بأنه يترتب على هذا مبدأ خطير وهو أن معامل صدق أي اختبار أكبر دائما من معامل ثباته وهو أمر غير مبرر. وعلى هذا فإن استخدام

[1] رينولد و ليفينجستون، المرجع مذكور سابقا، ص 226

مفهوم الصدق الذاتي لا يتعارض فقط مع المبادئ السيكومترية الأساسية بل يعد تشويها مباشرا لمعنى الصدق، ولا يمكن قبوله وفق أي تبرير. [1] فهو يتعارض مع مبدأ أساسي في القياس ألا وهو مبدأ الصدق يوجب الثبات والعكس غير صحيح وبالتالي فإن مستخدم هذا النوع من الصدق يخالف بذلك منطق القياس.

2-4-3-الصدق التطابقي

وهو ما يعرف أيضا بالصدق التقاربي والتباعدي، ويعتبر كذلك إجراء من إجراءات صدق البناء وليس نوع مستقل بذاته، ويعتمد هذا الإجراء حسب Engellant, Holland, & Piper, (2016) [2] على دراسة العلاقة بين الاختبار المراد التحقق من صدقه واختبارات أخرى إما أنها تدرس نفس السمة فنقول تم التحقق من الصدق التقاربي أو تدرس سمة مناقضة فنقول تم التحقق من الصدق التباعدي.

إن هذا التصنيف والتسميات السابقة هو مايطلق عليه غالبا بالنظرية الكلاسيكية أو التقليدية للصدق غير أنه مؤخرا بدأ خبراء القياس في التحول نحو تصور فكري للصدق كمفهوم تكاملي أي أنه بينما تحدثنا سابقا عن أنواع مختلفة من الصدق (أي صدق المحتوى والصدق المرتبط بمحك وصدق التكوين الفرضي) إلا أن هذه الأنواع تمثل فقط في الحقيقة طرقا مختلفة لجمع أدلة تدعم صدق تفسيرات الأداء على اختبار ما. وليس الغرض من عرض هذه التصنيفات هو استعراض المعلومات أو حشو الصفحات، بل الغرض من هذا العرض والتصنيف لأنواع وأدلة الصدق

---

[1] صفوت فرج، المرجع مذكور سابقا، ص 288

[2] Engellant, A., Holland, D., & Piper, T: Assessing convergent and discriminant validity of the motivation construct for the technology integration education (TIE) model. *16* (1). 2016, p 3-16.

وإجراءتهما هو تفهم أصلها ونشأتها. فهذا التأثير التراكمي للصدق يثبت بأنه ليس مفهوما مطلقا إما موجوداً أو غير موجود، أي أنه لا يمكننا القول بأن الاختبار صادق أو غير صادق بصفة قطعية، بل يوجد على متصل، ونحاول أن نحدد درجاته على ذلك المتصل بالطريقة الكلاسيكية أو نحدد مدى صدق تفسير درجات معينة على ذلك المتصل بالطريقة الحديثة. كذلك للإشارة فإن هنالك العديد من الألقاب الأخرى التي قد يصادفها القارئ والتي في الحقيقة هي عبارة عن إجراء من إجراءات أنواع الصدق الثلاثة المذكورة سابقا، مثل الصدق العاملي وصدق المقارنة الطرفية والصدق التمييزي وغيرها من الإجراءات، وقد حاولت الباحثة هنا ذكر أهمها أو الأصح أكثرها شيوعا.

3-إجراءات أنواع الصدق

وللتذكير مرة أخرى أن التقسيم بين التقليدي والحديث في الصدق ما هو إلا اجتهاد بعض الباحثين كمحاولة للفصل وتوضيح الخلط القائم بين تسميات وإجراءات وتصنيفات الصدق وغيرها، وقد اعتمدت الباحثة الفصل بناءً على وثيقة المعايير الصادرة عن كل من NCME و APA وAREA سنة 1999 رغم إصدار أو بالأحرى تحديث هذه الأخيرة سنة 2014 إلا أنه تم الأخذ بالوثيقة الخامسة لأنها تعتبر الوثيقة الرسمية الأولى التي خطت هذا التوجه الحديث في الصدق بشكل واضح وصريح. فالوثيقة الرابعة الصادرة 1985 مهدت لهذا الطرح فقط.

أيضا وجب الإشارة إلى أن هذه الإجراءات ليست ثابتة بل هي إجراءات مرنة يمكن التغيير والتعديل فيها وفق كل بحث وحاجة، فهي تعتبر خطوات عامة وكأنها إرشادات تساعد الباحث للوصول إلى هدفه.

3-1-إجراءات صدق المحتوى:

يتضمن إجراء صدق المحتوى بشكل أساسي الفحص المنهجي لمحتوى الاختبار لتحديد ما إذا كان يتضمن عينة تمثيلية من نطاق السلوك المراد قياسه. ويتطلب صدق المحتوى في الحد الأدنى الخطوات الآتية حسب كلين وكوسكوف Klein & Kosecoff (1975) :

*تطوير قائمة الأهداف.

*تقييم الأهمية النسبية لكل هدف.

*إنشاء جدولاً به صف واحد لكل محكم أو مختص وعمود واحد لكل هدف من أجل تسجيل التصنيفات ذات الأهمية.

*وضع كل فقرة من فقرات الاختبار في بطاقة.

* القيام بتعيين لكل فقرة الأهداف التي تقيسها.

*تسجيل مهمة تخصيص البند-الهدف ويتم ذلك عن طريق: أ-تقنية الحكم التعسفي. ب- تقنية الترجيح النسبي.

*حساب الارتباط بين الأهمية النسبية لكل هدف وعدد الفقرات المخصصة له.

*حساب عدد الفقرات التي لم يتم قياسها بأي فقرة من الاختبار. [1] وقد اقترح إيبل Ebel (1956) أن يقوم المختص الذي يعمل على التحقق من صدق المحتوى بالإطلاع على كل من الفقرات ويحدد الإجابة الصحيحة تماما كما يمكن أن يفعل المفحوص.[2]

---

[1] Klein, S. P., & Kosecoff, J. P: Determining how well a test measures your objectives (CSE Report 94). Los Angeles, University of California Los Angeles: Center for the Study of Evaluation. 1975, p 2-7

[2] Ebel, R: Must all tests be valid? *American Psychologist.16* (10). 1961, p 640-647.

كما هو واضح فإن الخطوات الفعلية للتحقق من صدق المحتوى ليس ما يعتقده الكثير من الباحثين مجرد عرض الاختبار على الأساتذة أو المختصين بل يتم وفق خطوات مدروسة وإجراءات لابد منها رغم أنها مرنة إلا أننا نحتاج دائما إلى بضع خطوات مدروسة، وعرض الاختبار على المحكمين ما هو إلا خطوة من خطوات صدق المحتوى لا تصح هذه الخطوة بمعزل عن باقي الخطوات ولا يمكن أن نطلق عليها صدق المحتوى إنما هي مجرد صدق ظاهري لا يمكن الاكتفاء به.

كذلك أضاف هامبلتون وروفينيل Hambleton & Rovinelli (1977) بعض الشروط عند استخدام الخطوات السابقة من بينها: أنه يجب أن تطابق الفقرة هدفا واحدا وواحدا فقط من مجموعة الأهداف. وتعكس إجراءات جمع البيانات هذا الافتراض، وذلك لأنه تم توجيه المحكمين لربط الفقرة بكل هدف يتطابق معها و أن يعطوا قيمة (1) فيما لو كان هناك تطابق بينهما، والقيمة صفر في حالة الشك في التطابق والقيمة (−1) فيما لو لم تتطابق الفقرة مع الهدف بوضوح.[1]

إن الشرط السابق يعرج بنا إلى مشكلة مهمة في صدق المحتوى قد يقع فيها مطور الاختبار أو حتى مستخدمه فعند التعريف بالمجال وتحديد الأهداف وبناء الفقرات إلى آخره، لم يتم التطرق أبدا في تلك الخطوات السابقة إلى خصائص المفحوصين فمثلا إذا تم تقديم في مواصفات مجال الاختبار على أنه يقيس القدرة على حل المسائل اللفظية في الحساب وتم عرض الفقرات اللفظية باللغة الفرنسية لأطفال متحدثين باللغة العربية، عندها يتم المطابقة بين الأهداف والفقرات على المحكمين دون دراية بهذه النقطة، وقد ويتم الوصول إلى أن الاختبار ذا صدق محتوى جيد، لكن

---

[1] Ronna C, T., & Laurie, C: Indexes of Item-Objective Congruence for Multidimensional Items. *International Journal of Testing .2003 , p* 163-171.

الاختبار في الأصل لا يعكس قدرة الأطفال حقا حتى وإن كانت الفقرات تمثل الأهداف ونسبة المطابقة عالية جدا. لذلك وجب تضمين سمات المفحوصين وكل الخصائص المطلوبة التي من شأنها التأثير على استجاباتهم في الأوصاف المنشورة للمجال السلوكي الذي تم معاينته بالاختبار.

كذلك من بين أهم المشكلات التي تصادف صدق المحتوى هي عدم تمثيل الأهداف للمجال السلوكي، أي أنه برغم مطابقة الفقرات بشكل جيد للأهداف التي تم تسطيرها، إلا أن تلك الأهداف قد لا تمثل المجال السلوكي الذي يرغب مستخدم الاختبار التحقق من صدقه. أي أنها تمثله بشكل خاطئ أو بشكل ناقص.

3-2-إجراءات الصدق المرتبط بمحك:

ترتبط إجراءات التحقق من الصدق المحكي عادة بالخطوات الآتية:

*تحديد المجال السلوكي المراد قياسه.

*تحديد عينة مناسبة ممثلة من المفحوصين المستهدفين.

*تحديد المحك المراد استخدامه وإثبات أنه يعد معيارا ذهبيا في حالة تم الاعتماد على اختبار آخر، كما يمكن الاعتماد على محكات أخرى عدا الاختبارات كبيانات أو رأي خبير أو غيرها.

*تطبيق الاختبار المراد التحقق من صدقه.

*تطبيق المحك أو جمع البيانات والآراء وغيرها من المحكات التي تم الاعتماد عليها، في نفس الوقت إذا كان الغرض استخدام الصدق التلازمي.

*تحديد قوة العلاقة بين درجات الاختبار والأداء على المحك. [1]

[1] Ebel, R. (1961). Must all tests be valid? American Psychologist. 16 (10) .1961, p 640-647.

كذلك من ضمن الخطوات التي لم يتم ذكرها أعلاه وتعتبر من الخطوات المهمة وأكثرها إهمالا من طرف الباحثين هي تقديم وصف شامل للمحك المستخدم وتقديم دليل صلاحيته لهذا الموقف الاختباري، كذلك تبرير استخدام هذا المحك بالضبط دون غيره.

3-3-إجراءات صدق البناء:

على الرغم من أن أدلة صدق البناء يتم تجميعها عادة من خلال سلسلة دراسات إلا أن هذه العملية بشكل عام تتضمن الخطوات الآتية حسب كروكر وألجينا Crocker & Algina (2017):

*صياغة فرضية أو أكثر، ويجب أن تعتمد هذه الفرضيات على نظرية صريحة صياغتها واضحة تكمن خلف البناء المستهدف وتزود بتعريفه اللغوي.

*اختيار أداة قياس تتألف من فقرات تمثل سلوكات محددة وأدلة ملموسة للبناء.

*جمع بيانات تجريبية تمكن من اختبار العلاقات المفترضة.

*تحديد ما إذا كانت البيانات متسقة مع الفرضيات، مع الأخذ بعين الاعتبار الدرجة التي وفقها يمكن تفسير النتائج الملاحظة بواسطة النظريات المنافسة أو التفسيرات البديلة.[1] وللإشارة ويتطلب صدق البناء تجميع أنواع عديدة من الأدلة أكثر من أي أسلوب صدق آخر.

للتحقق من الخطوات السابقة الذكر هنالك العديد من الإجراءات التي يمكن اعتمادها إلا أن كرونباخ وميهل Cronbach & Meehl (1955) حددا خمسة إجراءات يمكن اعتمادها للوصول إلى هذا النوع من الصدق وهي:

---

[1] ليندة كروكر، و جايمس ألجينا: تأليف *مدخل إلى نظرية القياس التقليدية والمعاصرة*. ترجمه: هند عبد المجيدالحموري، و زينات يوسف دعنا. دار الفكر ناشرون وموزعون. عمان، الاردن. 2017، ص 315.

3-3-1-الفروق في الجماعة: ما يطلق عليه أيضا بالصدق التمييزي بحيث يعتمد هذا الإجراء على التمييز بين حدي السمة في الجماعة الواحدة. [1] وهو ما يثبت أن الاختبار قادر على تمييز طرفي السمة في العينة انطلاقا من مسلمة القياس وهي الفروق الفردية.

3-3-2-مصفوفات الارتباط والتحليل العاملي: والذي يصمم للكشف والتعرف على التباينات المشتركة والارتباطات والسمات الافتراضية بين الاختبارات أفضل بكثير من الارتباط بين اختبارين بوصفه معامل لصدق التكوين الفرضي. [2] والتحليل العاملي يعد بدرجة أساسية أسلوبا إحصائيا لتحليل العلاقات البينية للبيانات السلوكية. وفي عملية التحليل العاملي يختزل عدد المتغيرات أو الأقسام من حيث ما يمكن بها وصف أداء كل فرد من العدد الأصلي للاختبارات إلى عدد أقل نسبيا من العوامل أو السمات المشتركة. وأضاف كامبل وفيسك Campbell& Fiske (1959) أسلوبا آخر للتحقق من صدق التكوين الفرضي يغطي نقائص معامل الارتباط والتحليل العاملي أيضا، وقد أطلقا عليه اسم "مصفوفة السمات المتعددة"[3] ويعتمد هذا الأسلوب في تحليله على التحقق وفحص معاملات الارتباط في عدة اختبارات أي أنه لا يكتفي بدراسة معاملات الارتباط مع الاختبارات التي تقيس نفس السمة فقط بل يتعداها إلى الاختبارات التي تقيس سمات مغايرة.

3-3-3-دراسة الفروق عبر الزمن: وهو ما يعرف بثبات درجات الاختبار ("إعادة الاختبار، التكافؤ ، "تقنية N من Cattell")، يمكننا أن نعتبر استقرار درجات الاختبار عبر الزمن

---

[1] Ibid, p 287

[2] Ibid, p 647

[3] Campbell, L. F., Stanford, L. D., Alfonso, V. C., Cuttler, M. J., Demers, S. T., Hass, G. A., et al: APA Guidelines for psychological Assessment and Evaluation. American Psychological Association. 2020, p 1-27.

عند تطبيقه على نفس العينة في نفس الظروف أو ظروف مشابهة تماما؛ كدليل على صدق الاختبار إذا كان الباحث افترض استقرار هذه السمة عبر الزمن. [1] ولكن يمكن العكس كذلك هنا بمعنى أن يكون أداء الاختبار غير مستقر أي أنه توجد فروق بين التطبيق الأول والثاني ونعتبره هو الآخر دليل على الصدق عند افتراض الباحث أن السمة التي يقيسها سمة متحركة وغير مستقرة. فالتفسير المقترح يعتمد بناءً على النظرية التي تم البناء وفقها.

3-3-4-الاتساق الداخلي: رغم أن الاتساق الداخلي يستخدم في العديد من البحوث على أنه صدق، إلا أن هذا يعتبر خطأ شائعا فالاتساق الداخلي بجميع أساليبه يقع تحت إجراءات الثبات، بحيث أن الاتساق الداخلي يعني أن البنود جميعها تقيس شيئا مشتركا أو تقيس ما يقيسه البعد أو المحور بصفة عامة، لكن هذا لا يعني أن البنود تقيس ما وضعت لقياسه، ولا يدل كذلك على أن البنود تقيس أمرا إضافيا. وحتى يمكن فهم أسلوب تقدير الاتساق الداخلي للاختبار في علاقته الضعيفة بمفهوم الصدق يتعين التمييز بين طريقتين حسب أنستازي وأوربينا Anastasi & Uraina (2015): فالطريقة الأولى هي الارتباط بين البند والدرجة الكلية. وتفسر معاملات الارتباط الناتجة عن هذا الأسلوب إلى أن البند يقيس الشيء نفسه الذي يقيسه الاختبار ككل ولكنه لا يتضمن معلومة محددة حول ما هو هذا الشيء نفسه الذي يقيسه الاختبار ككل. أما الطريقة الثانية فهي الارتباطات الداخلية بين البنود وفي هذه الطريقة نحسب الارتباطات لا بين كل بند والدرجة الكلية، بل الارتباطات فيما بين البنود وعادة ما يوفر لنا هذا الإجراء حزمة من المعلومات المهمة من ذلك إمكانية حساب ثبات الاختبار من خلال وسيط الارتباطات بين البنود، أما الاستخدام الأكثر أهمية فهو أنه يساعد على فهم القدرة التمييزية للبنود، وفي كلتا الحالتين لا يمكن

---

[1] Ibid, p 289

الخروج بدلالات حول صدق ما يقيسه الاختبار من خلال المعلومات المتعلقة بالارتباط بين البنود والدرجة الكلية أو من الارتباطات البينية بين البنود. [1]

ومن الواضح حسب المرجع السابق نفسه، أنه على الرغم من ذلك فإن الإضافة التي توفرها بيانات الاتساق الداخلي لصدق الاختبار محدودة ففي ضوء غياب بيانات من خارج الاختبار نفسه فلن نعرف إلا القليل عن ما يقيسه الاختبار. وعلى هذا فمن الأفضل الاكتفاء باستخدام أساليب الاتساق الداخلي في مجالي تحليل البنود لأغراض التمييز والصعوبة أو لحساب ثبات الاختبار.

وغير بعيد عن هذه الفكرة عبرت كورتينا Cortina (1993) أنه ليس من المناسب اعتبار الاتساق الداخلي مقياساً للصدق لأنه لا يخبرك ما إذا كانت الفقرات تقيس البناء الذي يراد قياسه. بل إنه يخبرك فقط ما إذا كان الاختبار يقيس شيئا مشتركا، ولكن ما هو ذلك الشيء؛ لا يزال غير معروف. [2]

إذن واختصارا للنقاش السابق فإن الاتساق الداخلي للبنود لا يعتبر صدقا، ولكن هذا لا يمنع استغلاله في الصدق لإثبات التكوين الفرضي للبنود وبأنها تقيس شيئا مشتركا ولكن لا يمكن الاكتفاء به أبدا للدلالة على صدق الاختبار بل وجب دائما أن تصاحبه أدلة أخرى تثبت صدق الاختبار.

3-3-5-دراسة ميكانيزمات الأداء على الاختبار: إن مراقبة عملية أداء الفرد، واحدة من أفضل طرق تحديد ما يفسر التباين في الاختبار بشكل غير رسمي.[3] وتعتمد هذه الطريقة على

---

[1] أنستازي و اوربينا، المرجع مذكور سابقا، ص 168

[2] Cortina, J: What is coefficient alpha? An examination of theory and applications. J Apple Psychol.78. 1993, p 98-104.

[3] Ibid, p 290

دراسة كيفية الاستجابة على الاختبار، ويمكن للفاحص أو مستخدم الاختبار تسجيل ملاحظاته عن طريقة استجابة المفحوص أو بإمكانه إعداد استمارة أو استبيان أو حتى أداة يستخدمها لجمع الملاحظات الخاصة بموقف الاختبار ثم يتم حساب الارتباط بينهما.

هذه النقاط الخمسة السابقة في ما يتعلق بالتحقق من صدق التكوين الفرضي حسب كرونباخ وميهل Cronbach & Meehl (1955) ولكن أنستازي وأوربينا Anastasi & Uraina (2015) أضافتا عنصرين آخرين ألا وهما:

3-3-6-التدخلات التجريبية: أحد المصادر المتعلقة بالتحقق من صدق التكوين الفرضي قدمتها التجارب الخاصة بأثر متغيرات منتقاة على درجات الاختبارات. مثلا المقارنة بين درجات اختبار قبلي واختبار بعدي، والوجهة المثالية ينبغي ألا تجيب النسبة الأكبر من المختبرين على مفردة معينة في الاختبار القبلي إجابة صحيحة وتجيب عليها إجابة صحيحة في الاختبار البعدي.[1] أي أن الاختبار قادر على تمييز التغيرات التي حدثت في السمة المراد قياسها وبالتالي انطلاقا من ذلك يمكننا اعتبار أن الاختبار صادق من حيث قدرته على كشف مستوى السمة والتغيرات التي حدثت لها انطلاقا من التدخلات التجريبية.

3-3-7-نمذجة المعادلة البنائية: الخطوة الأساسية في نمذجة المعادلة البنائية هو تصميم العلاقات السببية الافتراضية المراد التحقق منها، ومن المهم أن تكون العلاقات المفترضة لها تبرير نظري منطقي ويتم التحقق الفعلي من النموذج بحل مجموعة من معادلات الانحدار الخطي الآنية.[2] إذن النمذجة بالمعادلات البنائية تعتمد على التبرير النظري المنطقي في بنائها وتفسيرها

---

[1] انستازي و أوربينا، المرجع مذكور سابقا، ص 169.

[2] أنستازي وأوربينا، المرجع مذكور سابقا، ص 170.

وتحليلها للنموذج مما جعلها خطوة أو إجراء مهم تم إضافته إلى باقي إجراءات التكوين الفرضي وأصبح الآن من أهم الخطوات المعتمدة فيه.

4-أدلة الصدق

سنتطرق في هذا العنصر إلى مصادر الأدلة المختلفة التي يمكن الاعتماد عليها لإثبات صلاحية الاختبار لاستخدام ما. وقد تم الاعتماد في هذا التقسيم كما وقد تم الإشارة إليه سابقا على وثيقة المعايير الصادرة سنة (1999) وتم اعتماد هذا الإصدار- الخامس- بالضبط دون الإصدار السابق الصادر سنة (1985) رغم أنه يطابقه من الناحية الفكرية إلا أن الإصدار الرابع لم يفصل في الأدلة بل أشار فقط إلى فكرة توحيد أنواع الصدق، في حين أن الإصدار الخامس فصل فيه وبين ما هي هذه الأدلة وضحد فكرة أنواع الصدق كليا واعتمد على هذه البينات لإثبات صلاحية الاختبارات وبالتالي لاحظنا أنها لم تستخدم أبدا المصطلحات التقليدية للصدق مثل صدق المحتوى أو المحكي وغيرها. وكذلك صدرت وثيقة معايير جديدة سنة (2014) إلا أنها لم تغير الكثير في ما يخص موضوع أدلة وبينات صلاحية الاختبارات فقط أضافت بعض التعديلات هنا وهناك وقد تم الاعتماد عليها هي الأخرى.

وقبل عرض الأدلة المقدمة أدناه نبهت وثيقة المعايير (2014) أنه ليس من المطلوب تجميع كل الأدلة المعروضة أدناه، أو أن كل دليل يجب أن يقدم في جميع الظروف، بل بدلا من ذلك اقترحت تقديم تبريرات وتفسيرات لاستخدام أدلة معينة دون الأخرى، لأن الصدق مفهوم تكاملي فهو الدرجة التي تدعم بها جميع الأدلة المجتمعة التفسير المستهدف لدرجات الاختبارات للأغراض المقترحة. AREA, APA, & NCME (2014)[1]

[1] Ibid, p 14

ويشير الجزء الأول من تعريف الصدق حسب فور وباشاراش Furr & Bacharach (2014) إلى أن الدرجة ليست إما موجودة أو غير موجودة، ولكنها قد توجد بكميات متفاوتة. يمكن أن تختلف صلاحية درجة تقييم معينة من سياق إلى آخر ومن فرد إلى آخر. الجزء الثاني، "الدليل والنظرية"، يشير إلى أن دعم المصداقية يتكون من كل من البيانات التجريبية والنظرية العلمية. توفر الإحصائيات السيكومترية شكلاً من أشكال الدعم، وتوفر النظريات المستخدمة لإنشاء التقييم شكلاً آخر .الجزء الثالث من التعريف هو "تفسيرات درجات الاختبار" مما يعني أن فعل إرفاق المعنى بدرجات التقييم ينطوي على الاستدلال .وبالتالي فإن صحة هذه الدرجة هي في حد ذاتها نظرية علمية مفتوحة للاختبار والمراجعة .أخيرًا تشير عبارة "للاستخدامات المقترحة" إلى أن السياق مهم.[1]

وقد حددت وثيقة معايير عام (2014) أقسام الأدلة الخمسة الآتية المتعلقة بصدق تفسيرات درجات الاختبارات وقد اعتمدت الباحثة في عرض هذه الأدلة على هذه الوثيقة، والوثيقة الصادرة قبلها باعتبارهما المصدر الرسمي والأساسي لهذه الأدلة:

4-1-أدلة تعتمد على محتوى الاختبار

يمكن الحصول على دليل صدق مهم من خلال تحليل العلاقة بين محتوى الاختبار والبنية التي يراد قياسها. ويشير محتوى الاختبار إلى الصياغة وتنسيق الفقرات أو المهام أو الأسئلة في الاختبار وكيفية ارتباطها بالسمات المدروسة. وكذلك كيفية إدارة البنود وتسجيلها وترتيبها قد تكون أيضًا ذات صلة بالأدلة القائمة على المحتوى. وحسب فام pham (2015) يمكن أن تتضمن الأدلة المستندة إلى محتوى الاختبار تحليلات منطقية أو تجريبية حسب مجال المحتوى وأهمية

[1] Furr, R., & Bacharach, V. (2014). *Psychometrics: An Introduction* (2e ed.). London: Sage.2014, p 199.

مجال المحتوى للتفسير المقترح لنتائج الاختبار. كما يمكن أن تستند أيضا على أحكام الخبراء للحكم على العلاقة بين أجزاء الاختبار والبناء. ويسمى هذا الإجراء منفردا بالصدق الظاهري وهو مصدر صلاحية وثيق الصلة بصدق المحتوى وهو جزء لا يتجزأ منه. [1] والصدق الظاهري أو ما يسمى أيضا بصلاحية الوجه يشير إلى الدرجة التي ينظر بها المتقدمون للاختبار إلى العناصر الموجودة في الأداة باعتبارها ذات صلة بالخاصية التي يتم تقييمها. [2] وعلى الرغم من أنها لا تشكل صدق حقيقي من الناحية السيكومترية. [3] ، إلا أن الصدق الظاهري قد يؤثر على الصدق الكلي للأداة بطرق مختلفة. ويمكن جمع الأدلة المتعلقة بالصدق الظاهري من خلال المقابلات أو الاستبيانات عن طريق مطالبة المستجيبين بتقييم مقدار كل عنصر يبدو أنه يمثل البنية التي يتم قياسها.[4]

إذن كما أشار سانتي، مارزالك وهاردينجر Santee, Marszalek, & Hardinger (2019) فإن الأدلة المستندة إلى محتوى الاختبار تعنى بالعلاقة بين محتوى الاختبار والبنية المراد قياسها. والسؤال الرئيسي المطروح هو: إلى أي مدى يمثل محتوى الاختبار البنية؟ وغالبًا ما تأتي الأدلة المستخدمة لمعالجة هذا السؤال في شكل مراجعات من قبل خبراء الموضوع أو المحكمين، ولكن قد تتكون الأدلة أيضًا من عمليات مرتبطة بتخطيط وبناء جيد للمجال السلوكي لسمات

---

[1] Pham, J: A student's perspective on assessment and the affective domain. Am J Pharm Educ.79 (8) .2015, p 315

[2] Sartori, R :Face validity in personality tests: Psychometric instruments and projective techniques in comparison. *Quality & Quantity. 44* (4). 2010, p 749-759.

[3] Ibid, p14

[4] Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., et al : COSMIN guideline for systematic reviews of patient-reported outcome maesures. Quality of Life Research. 27 (5). 2018, p 1147-1157.

محتوى الاختبار والمظهر الخارجي الذي يشير إلى وجود مستوى جيد من البناء. [1] وهنالك أربعة متطلبات لضمان عملية التطوير والبناء الجيد وهي: (1) تحديد البنية بوضوح بناءً على بحث شامل في الأدبيات، (2) إنشاء الفقرات اعتمادا على الأسباب المنطقية، (3) مشاركة خبراء مؤهلون في تطوير المقياس، و (4) على الإجراء أن يكون خاليًا من تدخل محتوى غير ذي صلة.[2]

يمكن القيام بتحديد ما إذا كان محتوى المقياس مناسب أو غير مناسب لتعريف المفهوم عن طريق الخطوات التالية:

-تقدير مدى قدرة فقرات المقياس على تمثيل مجال المفهوم أو نطاقه.

-القيام بمراجعة أو التدقيق في الفقرات من حيث تلوثها أو خلوها من التحيز باللجوء إلى محكمين أو خبراء والتدقيق في إنصاف المقياس للمفحوصين.

-تقويم إجراءات التطبيق والتصحيح للإطمئنان مثلا عن وضوح تعليمات الاختبار وكفايتها وتحديد مدة الوقت المناسب للإجابة، وتدريب القائمين بالتقدير إذا استعملت مقاييس أو سلالم التقدير وغيرها. [3] (تيغزة ا.، 2008، الصفحات 21-22) ويستحسن هنا عند استخدام هذا النوع من

---

[1] Santee, J. A., Marszalek, J. M., & Hardinger, K. L : A critique of validity analysis from instruments assessing the affective domain. Currents in Pharmacy Teaching and Learning , 11 (2). 2019, 222

[2] Chen, X., & Gabrenya, W. K: In search of cross-cultural competence: A comprehensive review of five measurement instruments. *International Journal of Intercultural Relations .82*. 2021, p33-55.

[3] امحمد بوزيان تيغزة: نظرية الصدق الحديثة ومتضمناتها التطويرية لواقع القياس. *جامعة الملك سعود، كلية التربية قسم علم النفس*. 2008، ص 21

الأدلة التركيز على أمر مهم وعدم إغفاله لأنه في غاية الأهمية وهو شمولية الفقرات لجميع جوانب السمة المراد قياسها، فلا يجب التركيز فقط على مواءمة البنود للسمة والعينة والغرض إلى آخره بل أيضا وجب التركيز على شموليتها أيضا.

4-2-أدلة تعتمد على عمليات الاستجابة

يمكن أن توفر التحليلات النظرية والتجريبية لعمليات استجابة المتقدمين للاختبار أدلة تتعلق بالتوافق بين التركيب والطبيعة التفصيلية للأداء أو الاستجابة التي يشارك فيها بالفعل المتقدمون للاختبار. كما يمكن أن يساهم الدليل الذي يعتمد على عمليات الاستجابة في الإجابة عن أسئلة حول الاختلافات في المعنى أو تفسير درجات الاختبار عبر المجموعات الفرعية ذات الصلة من المتقدمين للاختبار [1]. وتستند الأدلة المعتمدة على عمليات تحليل الاستجابة، بشكل كبير إلى مدى اعتماد مطور الاختبار عند بنائه وحتى عند التحقق من صلاحية اختباره على إشراك المفحوصين، وذلك من خلال تحليل الاستجابات الفردية، مثل إجراء مقابلة مع إحدى المفحوصات حول أفكارها وهي تجيب على عناصر الاختبار، أو بعد الانتهاء من الإجابة ومن ثم مناقشة أفكارها وفهمها عن كل بند من بنود الاختبار. [2] ولا تقتصر دراسات عمليات الاستجابة على المتقدم للاختبار. غالبًا ما تعتمد التقييمات على مراقبين أو حكام لتسجيل و / أو تقييم أداء أو منتجات المتقدمين للاختبار. في مثل هذه الحالات، تشمل أدلة الصدق ذات الصلة مدى توافق عمليات المراقبين أو الخبراء مع التفسير المقصود للدرجات. قد يشمل التحقق من الصحة دراسات تجريبية لكيفية تسجيل

---

[1] AREA, APA, & NCME, p 15

[2] Santee, J. A., Marszalek, J. M., & Hardinger, K. L : A critique of validity analysis from instruments assessing the affective domain. *Currents in Pharmacy Teaching and Learning .11* (2). 2019, p 224.

المراقبين أو الخبراء وتقييم البيانات جنبًا إلى جنب مع تحليلات مدى ملاءمة هذه العمليات للتفسير المقصود أو تعريف البناء.[1]

وقد لخص تيغزة (2008) إجراءات الحصول على هذا النوع من الأدلة في بضع خطوات:

–تحليل استجابات المفحوصين بطريقة التحليل "البروتوكولي"، حيث يطلب من المفحوصين أن يفكروا بصوت مسموع خلال قيامهم بحل المشكلة، أو يطلب منهم عقب انتهائهم من حل المشكلة أن يصفوا الطرق والاستراتيجيات والعمليات التي استعملوها للوصول إلى الحل.

–فحص المنطق التي قامت عليه إجابات الأفراد، وأيضا كيفيات أو طرق إجابتهم (بتحليل أسبابها ودواعيها) على المقياس؛ وذلك بتحليل أوجه التأويل؛ والتعليلات التي يحررها المفحوصون أو يدلون بها شفويا لتبرير المنطق والطرق التي استعملوها وتعليلها.

–طرق تحليل أخطاء الأداء في الموقف التقويمي بحيث يستطيع المقوم أن يصل إلى بعض الاستنتاجات والتأويلات عند تحليله للعمليات الناجمة عن أخطاء في الإجراءات، أو أخطاء تتعلق بالتصور أو التمثل الذهني للمشكلة من طرف المفحوص.

–بالنسبة للأسئلة غير الموضوعية، أو بالنسبة للتقويم البديل الذي يتبنى فقرات أو مهمات أو أنشطة تقويمية واقعية، فلا بد من الإطمئنان بأن المقدرين (أو المصححين) يطبقون محكات تقدير الدرجات بطريقة دقيقة، وبأن تقديراتهم لأداء المفحوصين لم تتأثر بعوامل خارجية كالتحيز أو تأويل محكات التصحيح تأويلا خاطئا. [2] (تيغزة ا.، 2008، ص 24). إن هذا النوع من الأدلة يعتبر مفهوما حديثا للصدق إلى حد ما، ولهذا ربما هو نادر الاستخدام في الدراسات كما هو واضح في الجانب التطبيقي.

---

[1] AREA, APA, & NCME, p 15.

[2] تيغزة، المرجع مذكور سابقا، ص 24

4-3-أدلة تعتمد على البنية الداخلية

يمكن أن تشير تحليلات البنية الداخلية للاختبار إلى الدرجة التي تتوافق بها العلاقات بين عناصر الاختبار ومكونات الاختبار مع البنية التي تستند إليها تفسيرات درجات الاختبار المقترحة. قد يتضمن الإطار المفاهيمي للاختبار بُعدًا واحدًا للسلوك، أو قد يفرض عدة مكونات يُتوقع أن يكون كل منها متجانسة، ولكنها تختلف أيضًا عن بعضها البعض. [1]

إذن الهدف من هذه الأدلة هو الإجابة عن التساؤل التالي: هل يعكس الهيكل الداخلي للاختبار الفهم النظري للبناء؟ لذلك تتكون الأدلة في الغالب من الدرجة التي تتوافق بها العلاقات بين الأجزاء المختلفة للاختبار مع النظرية. ويمكن أن يشير مصطلح "الأجزاء" إلى المكونات / العوامل / الأبعاد أو العناصر الفردية أو كليهما. والطرق الأكثر شيوعًا التي تستخدم كأدلة البنية الداخلية هما تحليل العوامل وموثوقية الاتساق الداخلي. [2] وتفترض الاستنتاجات المستندة إلى تحليل الاتساق الداخلي أن الأداة المعنية تقيس بنية أحادية البعد ومع ذلك، يمكن للأداة قياس عدد من التركيبات أو البنيات الفرعية. على سبيل المثال قد تقوم الأداة بتقييم ثلاثة عوامل مختلفة تمثل السمات العاطفية التالية: الإيثار والنزاهة والاحترام. قد تقيس بعض العناصر الإيثار، وقد تقيس بعض العناصر النزاهة، وقد تقيس العناصر المتبقية الاحترام .سيكون تقييم قيمة واحدة فقط لألفا كرونباخ غير مناسب إذا كان للأداة أكثر من عامل واحد (ما لم يكن بالإمكان إثبات تجريبياً أن عوامل الترتيب الأول المتعددة هي في حد ذاتها أبعاد عامل واحد. [3]

---

[1] AREA, APA, & NCME, p 16

[2] Ibid, p 45

[3] Furr, R., & Bacharach, V :*Psychometrics: An Introduction* .2 ed. London: Sage2014, p 209.

يمكن الحصول على بيانات الصدق التي تركز على البنية الداخلية لأدوات القياس حسب تيغزة (2008) من الإجراءات التالية:

-تقدير معامل الصعوبة، ومعامل التمييز للفقرات ويعرف هذا الإجراء كلاسيكيا بتحليل فقرات المقياس.

-استعمال منحنى تحليل استجابة الفقرة Item Response Approach للكشف عن خصائص منحنى الفقرة والمقياس Item/test characteristic curves.

-حساب معاملات الارتباط بين فقرات المقياس للوقوف على مستوى الارتباطات البينية للفقرات وعلى طبيعتها أو اتجاهها.

-حساب معاملات الارتباط بين كل فقرة بالمقياس ككل، للوقوف على مستوى هذه العلاقات الارتباطية واتجاهها. ويعرف هذا الإجراء في نظرية الصدق الكلاسيكية بصدق الاتساق الداخلي.

-تقدير ثبات المقياس (الإعادة لتقدير الاستقرار، الصور المتكافئة بفاصل زمني بين تطبيق الصورتين لتقدير الاستقرار والتكافؤ، الصور المتكافئة بتطبيق الصورتين في الجلسة ذاتها لتقدير التكافؤ، التجزئة النصفية لتقدير الاتساق، التجانس الداخلي باستعمال المعادلة العامة ألفا لكرونباخ لتقدير الاتساق والتجانس).[1] -التحليلي العاملي ويمكن تقسيم التحليل العاملي إلى ثلاثة مناهج مختلفة: تحليل المكونات الرئيسية(PCA) والتحليل العاملي الاستكشاف (EFA) ، والتحليل العاملي التوكيدي (CFA) [2] ويعتبر كل من EFA و PCA نهجين استقرائيين يعتمدان على البيانات ويتطلبان في بعض الأحيان إعادة تفسير البيانات في شكل تناوب العوامل .يعتمد CFA - أيضًا على نموذج العامل المشترك، ولكنه نهج استنتاجي مدفوع بالنظرية .يقوم مطور الاختبار،

[1] تيغزة، المرجع مذكور سابقا، ص 25.

[2] Ibid, p 210

بدلاً من السماح للبيانات باقتراح عدد من العوامل أو الأبعاد، بتحديدها مسبقًا في نموذجCFA ، ثم يختبر مدى ملاءمة النموذج للبيانات [1] إذا تعتمد هذه الأدلة على كل ما هو متعلق بالاختبار في حد ذاته، والمقصود هنا وحدة بنية الاختبار، واتساق فقراته وتجانسه وذلك من خلال العديد من الإجراءات المتاحة.

4-4-أدلة تعتمد على العلاقة بمتغيرات أخرى

في كثير من الحالات، يعني التفسير المقصود لاستخدام معين أن البناء يجب أن يكون مرتبطًا ببعض المتغيرات الأخرى، ونتيجة لذلك، توفر تحليلات العلاقة بين درجات الاختبار والمتغيرات الخارجية للاختبار مصدرًا مهمًا آخر لأدلة الصدق. ويتم استخراج هذا النوع من الأدلة عن طريق:

-الدراسات البعدية Meta-analysis studies. فقد تكون الملخصات الإحصائية لدراسات التحقق السابقة في مواقف مماثلة مفيدة في تقدير علاقات الاختبار والمعيار في موقف جديد. ويشار إلى هذه الممارسة بدراسة تعميم الصدق. [2].

-الدراسة الارتباطية لمدى قدرة المقياس الذي يراد تقدير صدقه على التنبؤ بالمحكات الخارجية (أو بالمحك اختصارا) التي تقاس في وقت لاحق.

-الدراسات التجريبية لاختبار ما إذا كان أثر المعالجات التي أدخلت على المجموعة أو المجموعات التجريبية مقارنة بالضابطة، أو على مجموعات المقارنة تنسجم وتوقعات الفرضية، مما يدل على ان الاختبار المستعمل لقياس الأثر يعكس المفهوم المقاس.

---

[1] Ibid, p 11

[2] AREA, APA, & NCME, p 18

–دراسة ما إذا كان اداء المجموعات التي تتمايز في الواقع في السمة المقاسة على الاختبار متمايزا أو متفاوتا.[1]

–إن مصفوفة السمات المتعددة–الطرق المتعددة تسمح لنا بتحديد ما الذي يرتبط به الاختبار وما الذي لا يرتبط به وكيف تؤثر طريقة القياس في هذه العلاقات. وهذا المدخل له قبول كبير من الوجهة الفنية والنظرية، غير أن صعوبة تنفيذه وتفسيره حدت بدرجة ما من تطبيقه في الوقت الحاضر. [2]

–الأدلة التي تستند على العلاقات مع المتغيرات الأخرى، أي مدى وقوة العلاقات التي يجب أن يمتلكها البناء مع المتغيرات الأخرى. وتشمل الأمثلة العلاقات بين درجة الاختبار والتدابير الأخرى التي تهدف إلى تقييم نفس البنية (دليل متقارب)، أو بنيات مختلفة (دليل متمايز).[3] والحكم على الأدلة المتقاربة أنها جيدة إذا كانت ارتباطاتها مع مقاييس التي تشترك معها في السمة معتدلة أو عالية، أما الأدلة المتمايزة أو التباعدية فعندما تكون الارتباطات مع مقاييس البناء غير ذات الصلة أي سمة مختلفة، منخفضة حينها نقول أنها جيدة [4] إذا قد تتضمن المتغيرات الخارجية مقاييس لبعض المعايير التي من المتوقع أن يتنبأ بها الاختبار، وكذلك العلاقات بالاختبارات الأخرى المفترضة لقياس نفس التركيبات، والاختبارات التي تقيس التركيبات ذات الصلة بالاختبار أو المختلفة عنه.

4-5-أدلة تعتمد على عواقب العملية الاختبارية

[1] تيغزة ، المرجع موجود سابقا، ص 28

[2] رينولدز و ليفينجستون، ص 237

[3] Santee, Marszalek, & Hardinger, p 232

[4] Ibid, p 40

تأتي بعض نتائج استخدام الاختبار مباشرة من تفسير درجات الاختبار للاستخدامات التي قصدها مطور الاختبار. تتضمن عملية التحقق من الصدق هنا جمع الأدلة لتقييم سلامة هذه التفسيرات المقترحة للاستخدامات المقصودة. [1]

وقد أيّد بعض المؤلفين هذا التصوّر الأكثر اتساعاً للصدق، بحيث يتضمن قضايا اجتماعية وقيمية فمثلا، اقترح ميسك Messick (1994) في أحد الفصول المسهمة أن تصور الصدق ينبغي أن يتسع بحيث "يقدم بشكل رسمي تضمينات الاعتبارات القيمية والعواقب الاجتماعية في الإطار المرجعي للصدق". [2] وقد انتقد بعض خبراء القياس الآخرين هذا التوجه. فمثلا اقترح بوفام Popham (2000) أن تضمين العواقب الاجتماعية في تعريف الصدق يؤدي إلى تشتيت الانتباه عن وضوح المفهوم. وجادل بأن الصدق يعرّف بوضوح بأنه "دقة الاستدلالات التي تعتمد على الدرجات وأن تضمين القضايا الاجتماعية يعقد الصدق دون داع لذلك. [3] . وعلى كل فالهدف من هذا الدليل هو معرفة هل تتوافق تفسيرات درجة الاختبار واستخداماتها مع نوايا مطوري الاختبار؟ وأحد الأمثلة على ذلك هو تحقيق بعض الفوائد المقصودة من تفسير درجات الاختبار واستخدامها، مثل منع شخص غير مؤهل من دخول مهنة. [4] . إن هذا الدليل الذي استحدث في الصدق والذي يعتبر غير موجود بأي شكل من الأشكال في النظرية التقليدية للصدق، يعتبر جد مهم خاصة أن الصدق لا يرتبط بالاختبار في حد ذاته بل بنتائجه، وكذلك لأهمية مآل الاختبار كدليل على صدقه فعليا.

---

[1] AREA, APA, & NCME, p 19.

[2] Messick, S :Validity of psychological assessment: validation of inferences from persons responses and performances as scientific inquiry into score meaning. New jersey: Educational Testing Service.1994, p 14.

[3] Ibid, p 13.

[4] Santee, Marszalek, & Hardinger, 223.

5-تعقيب على أنواع الصدق وأدلته

من الواضح بشكل عام أن عملية التحقق من الصدق لا تنتهي أبدًا، حيث توجد دائمًا معلومات إضافية يمكن جمعها لفهم الاختبار بشكل كامل وفهم الاستنتاجات التي يمكن استخلاصها منه. ولذلك فإن تفسير الاختبار لاستخدام معين يعتمد على أدلة متعلقة ومرتبطة بذلك الاستخدام. ويرتبط اختيار مجموعة من الأدلة دون غيرها بالعديد من النقاط أهمها هي البناء الذي يدرسه الاختبار والهدف منه والمجموعة التي سيطبق معها إلى آخره. ووجود أنواع عدة من الصدق، جعل النقاش مرتبط دائما بأي نوع من أدلة الصدق هو الأكثر أهمية وأساسية لمدة نصف قرن تقريبا. والموقف الذي تتبناه الباحثة هنا هو أن نوع الصدق الأكثر أهمية مرتبط بالنتائج المراد استخلاصها من درجات الاختبار. فيجب على صانعي الاختبارات ومستخدميها أن يفكروا بتحديد أي أنواع الأدلة الأكثر فائدة لدعم الاستنتاجات التي من المحتمل استخلاصها من درجات الاختبار. وبعدها يجب تخطيط دراسات الصدق لتوفير مثل هذه الأدلة. والأكثر تفضيلا تضمين هذه الدراسات عناصر من أنواع الصدق الثلاث أو أدلته.

إن الغرض من هذا العرض لأنواع وأدلة الصدق ليس حشو الصفحات أو التفصيل الغير مدروس، إنما أرادت الباحثة بهذا الطرح أن تنبه إلى أمرين الأول: أن أغلبية الدراسات سواء العربية أو الأجنبية كما هو واضح في الجانب التطبيقي لهذه الدراسة قد استعانت بأنواع الصدق بتسميتها التقليدية، والأمر الثاني أن العديد من الدوريات والمجلات المهمة في مجال القياس والاختبارات مثل: دورية القياس التربوي والنفسي، ودورية التقييم النفسي. كما أشار رينولدز و ليفينجستون(2013) [1] تلزم الناشرين بتوصيات وثيقة المعايير، والتي بدورها تصر على الطرح الجديد في ما يتعلق بأدلة الصدق، رغم أن الباحثة حقيقة لا ترى أن هناك اختلافات جذرية أو

[1] رينولدز و ليفينجستون، ص 223

جوهرية من الناحية الإحصائية والإجرائية، بل هذا التحديث في أدلة الصدق هو عبارة عن تغيير فلسفي إلى حد ما، فالمطلع على الخطوات والإجراءات التي ذكرناها سابقا لكل من الأنواع والأدلة يلاحظ أن الأدلة كأنها الأنواع بثوب جديد، ولكن هذا لا يحط من قيمة التعديلات التي أدخلتها الوثيقة المعيارية والتي تعتبر جد مهمة مثل صدق المآل والاستجابة والتي اعتمدتها بدورها دراسة كوسمين COSMIN في وضعها لقائمة المعايير التي استندت إليها الباحثة في تحليلها لعينة الدراسة الحالية.

وفي الأخير هدفت الباحثة بعرضها لكل من الأنواع والأدلة إلى تنبيه القارئ إلى ما قد يصادفه سواء في الجانب التطبيقي لهذه الدراسة أو الأدبيات في القياس أو في الدراسات الأكاديمية، من تسميات تطلق على نفس المعنى والمطلوب من الصدق، وأن الهدف الحقيقي ليس التركيز على استخدام نوع دون الآخر أو تفضيل دليل على دليل، بل الهدف هو أن يحسن الباحثون اختيار النوع أو الدليل الذي يناسب استخدامهم، وكذلك يحسنون إتباع الخطوات و الإجراءات المطلوبة في التحقق من ذلك الدليل على أكمل وجه، ومرة أخرى نركز على أنه من الأحسن أن يستخدم الباحث في تحققه من صدق الاختبار أكثر من نوع ودليل.

6-العوامل المؤثرة على صدق الاختبار

هنالك عدة عوامل تؤثر على الصدق أو مايطلق عليها أيضا مهددات الصدق وسنحاول عرضها على شكل نقاط:

6-1-طول الاختبار: يزداد صدق الاختبار بزيادة عدد بنوده لأن ذلك يؤدي إلى شمول الاختبار لأكبر عدد ممكن من مؤشرات السمة أو ابعاد المحتوى. [1] أي أنه كلما زاد عدد الأسئلة المدروسة والتي تمثل فعلا المجال السلوكي وتضعف من تأثير الشوائب وأخطاء القياس يزداد صدق الاختبار، كذلك بما أن الصدق يعتمد على الثبات ويتأثر به والثبات يعتمد على طول الاختبار، لذا فإن الصدق يعتمد أيضا على طول الاختبار.

6-2-ثبات الاختبار: الصدق دالة لمعامل الثبات، ويصل ثبات الاختبار إلى أقصاه حين يزداد طوله ويؤثر ثبات الاختبار في صدقه، فانخفاض معامل الثبات يدل على وجود خلل في الاختبار وفي نفس الوقت مؤشر لانخفاض صدقه، و أن الثبات العالي ليس مؤشرا لصدق الاختبار. [2]

إذا يتأثر صدق الاختبار بالقيمة العددية لمعامل ثباته تأثرا مباشراً فيزداد الصدق كلما زاد ثبات الاختبار.

6-3-تباين العينة: بما أن الصدق صورة من صور الارتباط القائم بين الاختبار والميزان، وحيث إن الارتباط يتاثر بالفروق داخل العينة، فإن الصدق يتأثر بتلك الفروق الفردية وهكذا نجد أن التباين المنخفض (التجانس) يقلل من الصدق، وأن التباين المرتفع يزيد من القيمة العددية لذلك الارتباط [3] ، لذلك نجد أن زيادة تجانس العينة (أي قلة التباين بين أفرادها) تقلل من الصدق و أن العينة غير المتجانسة (أي زيادة التباين بين أفراد العينة) تزيد من القيمة العددية للارتباط بين الاختبار والميزان وبالتالي تزيد من صدق الاختبار.

---

[1] عبد العزيز بوسالم: *القياس في علم النفس وعلوم التربية الأسس النظرية والمبادئ التطبيقية*. الجزائر: دار قرطبة للنشر والتوزيع. 2014، ص 74.

[2] مصطفى محمود الامام، أنور حسين عبد الرحمان، و صباح حسين العجيلي. (2016). *التقويم والقياس*. الأردن: دار الأيام. 2016، ص 132

[3] العدل، المرجع مذكور سابقا، ص 40

6-4-الإطار النظري: نوع التحليل النظري الذي يسبق بناء الاختبار فالتحليل النظري المنظم لمفهوم الخاصية أو القدرة المراد قياسها، والمستند إلى ما كتب حول موضوع الخاصية ومكوناتها الأساسية والمظاهر السلوكية التي تدل عليها، وهو في حد ذاته بيّنة منطقية عن صدق الاختبار. فعندما يخفق مصمم الاختبار في تحقيق هذه الخطوة بشكل جيد فإن ذلك يؤثر سلباً في صدق الاختبار. [1] وقد لخص ميسك , .Messick S(1987)[2] كل ما سبق في تهديدين رئيسين للصدق هما: أولا عندما يقيس الاختبار سمة أو مهارة أو صفة غير مرتبطة بالتكوين الفرضي المتعلق بالاختبار. وثانياعندما يقيس الاختبار ما يقل أو يزيد عن التكوين الفرضي الذي صمم الاختبار من أجل قياسه.

6-5-خصائص المختبرين: عدم الدقة في إجراءات التطبيق والتصحيح، مثل مكان وبيئة التطبيق كالتهوية والإضاءة والضوضاء وارتفاع درجة الحرارة أو انخفاضها بشكل ملموس. عدم جدية المفحوص في الاستجابة لفقرات الاختبار، خصوصاً عندما يواجهها باستهتار أو باللامبالاة. التوتر والقلق والخوف والارتباك الذي ينتاب المفحوص أثناء الإجابة فيحصل على نتيجة لا تمثل قدرته الحقيقة، إضافة إلى العادات السيئة في الإجابة كالتخمين أو الغش. [3] فأي عوامل شخصية تغير أو تقيد استجابات المختبرين في الموقف الاختباري يمكن أن تنقص من قيمة الصدق.

6-6-التعليم والتدريب الخاص: بالإضافة إلى محتوى الاختبار الذي يؤثر في الصدق فإن التعليم القبلي أو التدريب الخاص يمكن أيضا أن يؤثر في الصدق. [4] ويحدث ذلك عادة بغير قصد أو بقصد فأحيانا يعمد بعض المحامين مثلا إلى تدريب عملائهم على الاختبارات النفسية وضبط

---

[1] الامام، عبد الرحمان، و العجيلي، المرجع مذكور سابقا، ص 259

[2] Ibid, p 161

[3] أبو عقيل، المرجع مذكور سابقا، ص 119

[4] رينولدز و ليفينجستون، المرجع مذكور سابقا، ص 219

استجاباتهم وفق المتطلعات التي يريدون أن يظهر بها هؤلاء العملاء، وبالتالي يحدث تشويه في الاستجابات ويزيد هذا التشويه في القيمة الفعلية للصدق.

7-الصدق وفق دراسة كوسمين COSMIN

تم تطوير قائمة المعايير الدولية كوسمين COSMIN (2018) باستخدام طريقة دلفي عام (2010) وتحديثها سنة (2016) و(2018) كما سيتم التفصيل فيها أكثر في الجانب التطبيقي. وبما أن الباحثة اعتمدت في تقييمها للدراسات سواء العربية أو الأجنبية على هذه قائمة، فلذلك وجب علينا أن نشير إلى الصدق حسب هذه الدراسة أنه سيتم الاعتماد على معاييرها في التقييم والحكم على عينة الدراسة الحالية.

في تقييم جودة الاختبار تميز COSMIN ثلاثة مجالات، الثبات والصدق والاستجابة. ويحتوي كل مجال على واحد أو أكثر من خصائص القياس، أي جوانب الجودة لأدوات القياس. يحتوي الثبات على ثلاث خصائص قياس: ثبات الاتساق الداخلي، ثبات الإعادة، وخطأ القياس. ويحتوي الصدق أيضًا على ثلاث خصائص قياس: صدق المحتوى (للعينة والمحكمين)، والصدق البناء (بما في ذلك الصدق الهيكلي واختبار الفرضيات والصدق الثقافي)، وأخيرا الصدق المعياري. والاستجابة.

7-1-صدق المحتوى: الدرجة التي يكون بها محتوى الاختبار انعكاسًا مناسبًا للبناء المراد قياسه. أما الصدق الظاهري والذي يعتبر خطوة من خطوات صدق المحتوى فتعرفه COSMIN بأنه الدرجة التي تبدو بها عناصر الاختبار بالفعل كما لو كانت انعكاسًا مناسبًا للبنية المراد قياسها. ويتم تقييمها من خلال الشمولية والملاءمة والاستيعاب عند العينة من جهة والشمولية

والملاءمة عند المحكمين من جهة أخرى [1]

7-2-صدق البناء : الدرجة التي تتوافق بها درجات الاختبار مع الفرضيات (على سبيل المثال فيما يتعلق بالعلاقات الداخلية، والعلاقات مع عشرات الأدوات الأخرى، أو الاختلافات بين المجموعات ذات الصلة) بناءً على افتراض أن الاختبار يقيس بشكل صحيح البنية المراد قياسها. وتم الاتفاق بين الخبراء على تقسيمه إلى ثلاثة أقسام:

7-2-1-الصدق الهيكلي: الدرجة التي تكون فيها درجات الاختبار انعكاسًا مناسبًا لأبعاد البناء المراد قياسه.

7-2-2-صدق اختبار الفرضيات: الدرجة التي تتوافق بها درجات أداة القياس مع الفرضيات القائمة على افتراض أن أداة القياس تقيس بشكل صحيح البنية المراد قياسها.

7-2-3-الصدق الثقافي: الدرجة التي يكون بها أداء العناصر في الاختبار المترجم أو المعدّل ثقافيًا انعكاسًا مناسبًا لأداء عناصر الإصدار الأصلي للاختبار. [2] (Terwee, et al., 2012, p. 651). ويتم استخدام هذا النوع من الصدق مع الاختبارات التي ستطبق مع عينة مختلفة ثقافيا عن العينة التي صمم الاختبار من أجلها، وليس المقصود هنا بالاختلاف الثقافي فقط اللغة أو الديانة، وإنما يشمل على أي اختلاف قد يكون مصدرا لتحيز الاختبار، فمثلا الجنس أو الفئة العمرية أو اضطراب معين إلى آخره. [3] (Mokkink, et al., 2018, p. 58) وذلك لأن

---

[1] Mokkink, et al, p 1171

[2] Terwee, C., Mokkink, L., Knol, D., Ostelo, R., Bouter, L., & de Vet, H: Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res. 21* (4). 2012, p 651-657.

[3] Mokkink, et al, p 58

صانع الاختبار عادة ما يحدد في دليل الاختبار المتغيرات والشروط التي يصلح فيها تطبيق الاختبار، وبالتالي فإن تغيير تلك الضوابط يستلزم التحقق من الصدق الثقافي.

7-3-الصدق المعياري: الدرجة التي تكون فيها درجات الاختبار انعكاسًا مناسبًا "للمعيار الذهبي". [1]

8-معايير الصدق حسب COSMIN

لا يمكن التفصيل وذكر جميع المعايير التي تم ذكرها في دراسة COSMIN لعدة أسباب أهمها أن ذكر جميع المعايير تستلزم دراسة خاصة بها لكثرة محتواها وتفاصيلها، كذلك تطرقت الباحثة إلى أهم المعايير في الفصل التطبيقي. ولذلك ستكتفي الباحثة باختصار بعض المعايير هنا وعرض باقي المعايير باللغة الإنجليزية في الملحق (03)

8-1-بعض معايير صدق المحتوى

-يعد صدق المحتوى أهم خصائص القياس للاختبار والأكثر صعوبة في التقييم.

-يتم تمييز ثلاثة جوانب من صدق المحتوى: الملاءمة: "يجب أن تكون جميع العناصر في الاختبار ذات صلة بالسمة المراد دراستها ضمن مجتمع محدد وغرض الاستخدام واضح"، الشمولية: "لا ينبغي أن تكون الجوانب الرئيسية للبناء مفقودة" ، وقابلية الفهم: "يجب أن يفهم المفحوص العناصر على النحو المنشود". [2] -يجب إعادة التحقق من صدق المحتوى من خلال سؤال كل من المفحوصين والمحكمين حول ملاءمة العناصر وشمولها وقابليتها للفهم.[3]

-يجب على مؤلفي الاختبارات تحديد نطاق مراجعتهم بوضوح. نعني بالنطاق: البنية أو السمة،

---

[1] Ibid, p 542

[2] Terwee C. , et al, p1136.

[3] Mokkink, et al, p 740

المجتمع المستهدف والغرض من الاستخدام (مثل الغرض التمييزي أو التقييمي أو التنبؤي). يجب أن يكون هذا النطاق هو النقطة المرجعية لتقييم صدق المحتوى الاختبارات.

– يختلف صدق المحتوى للاختبار عند استخدام الاختبار لقياس التركيبات المختلفة، في مجموعات سكانية مختلفة، أو في سياقات استخدام مختلفة. وعادة لا يقوم الباحثون بالتحقق من صدق الاختبار بل تطبيقه. في حين قد يكون صدق المحتوى للاختبار جيد في المجموعة المستهدفة التي تم تطوير الاختبار من أجلها في الأصل، ولكنها أقل جودة عند استخدامها مع مجموعة أخرى من المفحوصين. قد تكون بعض الاختبارات صالحة لمجموعة واسعة من الاستخدامات في مجموعات سكانية مختلفة، ولكن قد يتطلب كل استخدام أدلة داعمة جديدة. [1]

– إذا كان الاختبار المعدل عبارة عن نسخة مختصرة من اختبار تم تطويره مسبقًا، فيمكن الحصول على معلومات حول ملاءمة العناصر وإمكانية فهمها من الدراسات في الاختبار الأصلي. ومع ذلك، يجب أن تأتي المعلومات حول شمولية الاختبار المختصر من دراسة جديدة باستخدام الاختبار المختصر. يجب على المراجعين تحديد المعلومات التي يمكن استخدامها من الدراسات في الاختبار الأصلي وأي المعلومات يجب أن تأتي من الدراسات الجديدة، باستخدام الاختبار المعدل [2]

8-2-بعض معايير صدق البناء

وتتضمن هذه المعايير كل من معايير الصدق الهيكلي وصدق اختبار الفرضيات والصدق الثقافي.

---

[1] Mokkink, et al, p 733

[2] Dobson, F., Choi, Y., Hall, M., & Hinman, R: Clinimetric properties of observer-assessed impairment tests used to evaluate hip and groin impairments: a systematic review. *Arthritis Care Res (Hoboken) .64* (10)..20121565-1575.

– يشير الصدق الهيكلي إلى كيفية ارتباط العناصر المختلفة في الاختبار، وتتعلق هذه الخطوة بتقييم الصدق الهيكلي (بما في ذلك أحا[1]دية البعد) باستخدام التحليل العاملي أو تحليلات راش.[2]

– الصدق الهيكلي والصدق الثقافي والأشكال الأخرى من صدق البناء تتم باستخدام تحليلات أداء العناصر التفاضلية (DIF) أو تحليلات العامل التأكيدي ومتعدد المجموعات(MGCFA) ، نحن هنا نشير إلى استخدامها في إعادة التحقق من الاختبارات الجاهزة وليس تطوير اختبارات جديدة.

– تركز خصائص الثلاثة هذه– الصدق الهيكلي وصدق اختبار الفرضيات والصدق الثقافي – على جودة العناصر والعلاقات بينها على عكس خصائص القياس المتبقية، والتي تركز بشكل أساسي على جودة المقاييس. [3]

–نوصي بتقييم الصدق الهيكلي مباشرة بعد تقييم صلاحية المحتوى للاختبار.

–نوصي أولاً بتقييم صدق البناء، ليتبعها ثبات الاتساق الداخلي والصدق الثقافي.

– عند تقييم صدق البناء باستخدام اختبارات تقيس سمة مشتركة يجب أن تكون الارتباطات مع الأدوات التي تقيس التركيبات المتشابهة أكبر أو تساوي 0.50.

– يجب أن تكون الارتباطات مع أدوات القياس تقيس نفس السمة، ولكن غير متشابهة، بين 0.50–0.30

– يجب أن تكون الارتباطات مع الأدوات التي تقيس تركيبات مغايرة أصغر من 0.30. [1]

---

[1] Mokkink, et al, p 720

[3] Saether, R., Helbostad, J., Riphagen, I., & Vik, T: Clinical tools to assess balance in children and adults with cerebral palsy: a systematic review. *Dev Med Child Neurol*. *55* (11).2013, p 988–999.

8-3-بعض معايير الصدق المعياري

-يجب أن يقرر فريق المراجعة ما إذا كان هناك معيار ذهبي متاح لقياس السمة المراد دراستها وفي المجتمع المستهدف.

-من حيث المبدأ، لا يوجد معيار ذهبي للاختبار، يمكن اعتبار الإصدار الطويل من الاختبار المختصر فقط معيارًا ذهبيًا.

-إذا تم اعتبار أداة قياس النتائج معيارًا ذهبيًا، فإن الدراسات التي تقارن الاختبار بهذه الأداة المعينة تعتبر دليلاً على صحة المعيار.

-عند دراسة العلاقة بين الاختبار الحالي واختبارات أخرى تحمل نفس السمة، لا تعتبر بالضرورة تلك الاختبارات معيارا ذهبيا، وبالتالي لا نعتبر انه تم التحقق من الصدق المعياري بل يعتبر أنه تم التحقق من صدق البناء. [2]

وفي الأخير تعتبر هذه المعايير هي عبارة عن نبذة فقط عما هو مفصل فيه في أدلة قائمة المعايير COSMIN، والتي تعمل الباحثة حاليا على ترجمتها إلى اللغة العربية بصفة رسمية ليتم وضعها قيد تصرف الباحثين في الموقع الرسمي لدراسة COSMIN، وبما أن الباحثة لم تتحصل

---

[1] de Vet, H., Terwee, C., Mokkink, L., & Knol, D. (2011). Measurement in Medicine: a practical guide. *Cambridge University Press* .2011, p 45

[2] Mokkink, L., de Vet, H., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., et al: Des 19). COSMIN risk of bias checklist for assessing the methodological quality of studies on the measurement properties of Patient-Reported. *epub ahead of print* .2017, p. 1171-1179.

على حق النشر إلا بعد أن تتم الموافقة من طرفهم ونشرها على الموقع، أي تم تقييد استخدام هذه الأدلة بغرض الدراسة فقط وباللغة الإنجليزية، فلذلك تم عرض الأدلة مفصلة باللغة الإنجليزية في الملاحق، لمن أراد المزيد من التفاصيل.

خلاصة

غالبا ما يتم الاعتماد على الاختبارات كأداة لقياس الظواهر والسمات في العلوم الإنسانية والاجتماعية، وسواء كانت هذه الاختبارات جاهزة أو يتم بناءها فإنه لزاما على الباحث أي صانع الاختبار أو مستخدمه أن يتحقق من صلاحية هذا الاختبار للاستخدام المرجو منه، وعادة ما ترتبط الصلاحية بالصدق، وذلك لكون الصدق يوجب الثبات، وقد حاولت الباحثة في هذا الفصل عرض مفهوم الصدق وأنواعه مع إجراءات كل نوع، ونفس الأمر في ما يتعلق بالأدلة فقد تم التطرق إلى إجراءاتها هي الأخرى، كذلك عرجت الباحثة إلى العوامل المؤثرة على الصدق، لتختم في الأخير بالمعايير التي تستند إليها في الحكم على طرق تقييم صدق الاختبارات في عينة الدراسة الحالية.

الفصل الرابع

الثبات في

الاختبارات النفسية

تمهيد:

يجب أن تقدم الأدوات، قبل اعتبارها مناسبة، بيانات دقيقة وصحيحة وقابلة للتفسير للأغراض التي توفرها. وعلاوة على ذلك، من المفترض أن توفر الاختبارات نتائج قوية علميا. ويأتي أداء نتائج هذه الاختبارات من صدقها وثباتها. ويجمع الباحثون على اعتبار الثبات والصدق كخصائص القياس الرئيسية للأدوات. وعلى هذا الأساس عملت الباحثة جاهدة على ذكر الصدق وأهم تفاصيله في الفصل السابق، لتركز في هذا الفصل على الجانب الآخر لصلاحية الاختبارات النفسية ألا وهو الثبات.

1-مفهوم الثبات

إن مصطلح ثابت هو ترجمة للمصطلح Reliable بالانجليزية، وفي حقيقة الأمر هي ترجمة مضللة نوعا ما، فالأصل الأجنبي للكلمة يدل على أن هذا الاختبار "موثوق فيه ويعتمد عليه" وهذا هو المقصود غالبا بكلمة ثابت هنا. فالكلمة وإن كانت توحي إلى تكرار نفس النتائج، إلا أن المقصود بثبات الاختبار هو دقته وكذلك مستوى الثقة في نتائجه وهو أيضا ما يدعمه التعريف القاموسي لمصطلح ثبات في المعجم الجامع[1] .بحيث ورد فيه أن ثبات الاختبار يعنى به دقة القياس. وهو ما يتفق معه المختصون في القياس في تعريفهم للثبات كخاصية من خصائص القياس كما هو واضح في التعريفات التالية:

فحسب كرونباخ Cronbach (1974) يشير الثبات إلى درجة التباين الثابتة في سمة ما بعد تطبيق الاختبار. وأيضا هو درجة الاختلافات الثابتة الناجمة عن عوامل معينة يحددها الاختبار.

[1] مروان العطية: *المعجم الجامع* (ط1). الأردن: دار الغيداء للنشر والتوزيع.2018، ص 217

كذلك هو الدرجة التي تشير بها درجة الاختبار إلى الاتساق الداخلي، وعليه يمكن القول أنّ لمعامل الثبات أهمية بالغة في بناء الاختبارات النفسية، يمدنا بافتراض بسيط يقرر أنّ درجة الاختبار مركبة من مكونين أساسيين منها الدرجة الحقيقية ودرجة الخطأ. [1] ويتضح لنا من خلال هذا التعريف أن كرونباخ حاول الجمع بين العديد من طرق التحقق من ثبات الاختبار في تعريف واحد، ليجعل هذا الأخير شاملا قدر الإمكان. أمَا ليمك وويرسما Lemke & Weirsma (1976) فقد عرفا الثبات بأنه نسبة التباين الحقيقي إلى التباين الكلي لدرجات المستجيبين، وكلما ارتفع التباين الحقيقي زادت درجات ثبات المقياس وقل الخطأ المعياري. [2] وتم التركيز في هذا التعريف على لب الثبات ومعناه الدقيق بغض النظر عن الطريقة المستخدمة للتحقق منه. وفي نفس السياق تقريبا ركز كل من نانلي وبرنستين Nunnally & Bernstein (1994) على أنّ الثبات هو خلو أو تحرر الدرجة من الأخطاء العشوائية أي الحصول على نفس النتائج عند إعادة تطبيقها على نفس الأفراد وفي نفس الظروف المماثلة لها. [3]. فلقد اعتمد كلا من التعريفين السابقين على نقطة مشتركة وهي أن الخطأ العشوائي والثبات بينهما علاقة عكسية، فكلما كانت قيمة الخطأ العشوائي كبيرة كانت درجة ثبات الاختبار صغيرة، وكلما تم ضبط الأخطاء العشوائية والتقليل من قدرها كلما ارتفعت قيمة الثبات. فالاختبار الثابت هو ذلك الاختبار الذي تم التحكم في مصادر أخطائه قدر الإمكان.

---

[1] Cronbach, L. J. (1974). TESTN RELIABILITY: ITS MEANING AND DETERMINQTION. *PSYCHOMETRIKA .12* (1). 1974, p 1−16.

[2] Lemke, E., & Wiersa, w :*Principales of psychologycal Mesurement.* chicago.1976, p 71

[3] Nunnally, J., & Bernstein, I: *Psychometric theory* (3rd ed ed.). New York: MC Graw-Hill. 1994, p 220

كذلك وجب بنا الإشارة إلى أن الثبات لا يتعلق كما هو شائع بالاختبار بل هو مرتبط بنتائج الاختبار والمقصود هنا أنه عند قولنا أن الاختبار ثابت فإننا بذلك نقصد بأن النتائج التي نتحصل عليها من هذا الاختبار ثابتة وهذا ما أكده ملحم (2005) فقد أشار إلى أنّ الثبات ليس صفة للاختبار في ذاته، بل صفة تتعلق بالدرجات التّي نتحصل عليها بعد تطبيق الاختبار على مجموعة من الأفراد. [1] . وذهب ماير Mayer (2010) في تعريفه للثبات بأنه يهتم باتساق القياسات عبر مختلف الظروف التّي تطبق وتصحح وفقها أدوات القياس. [2] أي أنه هنا ركز على طريقة معينة في استخراج درجة ثبات الاختبار وهي الاتساق عبر الزمن أو ما يسمى بإعادة التطبيق. وهو ما ذهب إليه أيضا كل من نيكتو وبروك هارت Nictko & Brook hart (2018) وكوبيزين وبوريش Kubiszyn & Borich (2013) في أنّ الثبات يقصد به الاتساق الناتج عن تطبيق الاختبار عدة مرات على نفس الأفراد، وبالتالي يمكن القول أنّ الاختبار ثابت إذا تحصلنا على نفس النتائج تقريبا في التطبيقين. أي أنّ مفهوم الثبات هو درجة بقاء نتائج الاختبار مستقرة، أو ثابتة عبر عدة مرات من التطبيق، بمعنى حصول الفرد على الدرجة نفسها إذا أدى الاختبار في وقتين مختلفين، أو صححت من قبل مصححين أو عندما تكون هناك صور متكافئة. [3] .

وفي نفس السياق عرَف النبهان (2013) الثبات على أنّه" درجة الاتساق أو التجانس بين نتائج مقياسين في تقدير صفة أو سلوك ما، إلا أنّ هناك تفاوت في درجة الثبات أداة القياس باختلاف

---

[1] سامي محمد ملحم: . *القياس والتقويم في التربية وعلم النفس*. ط 1. عمان: دار المسيرة للنشر والتوزيع.2005، ص 250

[2] Mayer, J :Reliability. London: Oxford university press.2010, p 75

[3] Nictko, A., & Brook hart, S :Educational assessment of students englercood chiffs. NJ Mersill.2018, p 201.

المجال الذّي يعمل على قياسه" [1] . أي أنه أضاف فكرة اختلاف درجة الثبات باختلاف الغرض من الاستخدام، وهذا يدل على وجوب إعادة التحقق من ثبات الاختبار عند تغيير الهدف من الاستخدام، ولقد تم التركيز في التعريفات السابقة الذكر على إجراءات الحصول على الثبات أكثر من التركيز على المعنى الحقيقي والأدق له، على عكس ما أشار إليه محاسنة (2013) إلى أنّ الثبات هو دقة الاختبار وبالتالي دقة القرارات المتخذة اعتمادا على نتائجه، وبالتالي فإنّ الدقة ليست للاختبار بحد ذاته بقدر ما هي نتائجه، أي ثبات درجات الاختبار أو نتائجه. [2] وهو ما ذهبت إليه أوربينا Urbina (2014) أيضا بحيث ترى "أنّ الثبات يشير إلى درجات الاختبار ذات قيمة عالية من الثقة والدقة التّي تسمح بأن نتخذ على أساسها قرارات وأحكام عن المستجيبين الذّين طبق عليهم الاختبار". [3] أي أنها تركز في تعريفها للثبات على معناه الحقيقي والشامل لجميع إجراءات استخراجه، بحيث بغض النظر عن طريقة التحقق منه فإن الثبات هو درجة الدقة والثقة في النتائج. وهذا هو لب مفهوم الثبات، فعند تعريفنا للثبات باستخدام أحد إجراءاته فهذا يجعل من التعريف ناقصا لأنه حصر المفهوم الواسع الذي يحمله الثبات بطريقة أو طريقتين من طرق التحقق منه، بالإضافة إلى كونه غير شامل فهو أيضا خاطئ إلى حد ما، فتلك التعريفات شرحت طريقة

---

[1] موسى النبهان: *أساسيات القياس في العلوم السلوكي. ط2* . عمان، الأردن: دار الشروق للنشر والتوزيع. 2013، ص 276.

[2] ابراهيم محمد محاسنة. (2013). القياس النفسي في ظل النظرية التقليدية و النظرية الحديثة. عمان: دار جرير للنشر والتوزيع.2013، ص 117.

[3] Urbina, S: *Essentials of psychological testing* . 2 ed .United states of America: Wiley. 2014, p 117.

الحصول عليه متجاهلة ماهيته أصلا. وفي الأخير فإنّ الثبات يعني اتساق درجات الاختبار ودقة نتائجه وتحرره من أثر الخطأ.

2-أنواع الثبات

توجد عدة أنواع أو طرق لحساب وتقدير الثبات، ويتم اختيار طريقة دون الأخرى حسب خطأ القياس المراد ضبطه والغرض من الاختبار وغيرها، فحسب علام (2000) [1] تختلف هذه الطرق باختلاف مصدر الأخطاء العشوائية التي نود تقدير تأثيرها في الدرجات الملاحظة، وهذا يعتمد على طبيعة الاختبار وفيم تستخدم نتائجه. وهنالك عدة تصنيفات لهذه الأنواع إلا أنه تم الاعتماد هنا على تصنيف وثيقة المعايير للقياس النفسي (AERA, APA, & NCME, 2014) والتي صنفت الثبات كالآتي:

-1-ثبات إعادة الاختبار تعتمد هذه الطريقة بكل بساطة على تطبيق الاختبار وإعادة تطبيقه مرة أخرى بعد فترة زمنية معينة، وينتج عن هذه الطريقة معامل تقدير يطلق عليه "معامل الاستقرار"، ويطلق عليه أيضا حسب Gronlund (1981) بمعامل التجانس عبر الزمن من خلال تطبيق الاختبار على عينة ثم إعادة تطبيقه عليها مرة أخرى وفي نفس الظروف، بعد فاصل زمني يقدر بقدر طبيعة السمة المقاسة وطبيعة أفراد العينة [2]. وحسب Hogan (2013) أنّ هذا النوع من الثبات يعتبر من أسهل الطرق لحساب ثبات الاختبار، ويمكن أن نتحصل عليه من خلال إجراء اختبار واحد على نفس الأفراد في مناسبتين مختلفتين، ومن ثم حساب معامل الارتباط

---

[1] علام، مرجع مذكور سابقا، ص 143.

[2] Gronlund, N. E. (1981). *Measurement and evaluation in teaching.* 4 ed. Macmillan Pub.1981, p 69.

بين درجات الاختبار الأول والثاني. وهو ما ذهب إليه أيضا Drost (2011) في تعريفه لهذا النوع من الثبات بأنه إجراء الاختبار لمجموعة من المستجيبين ثم إعادة إجراءه مرة أخرى على نفس المستجيبين في وقت لاحق وبعدها حساب العلاقة بين درجات الاختبار الأول والثاني. [1] وينبغي دائما عند استخدام هذه الطريقة في تقدير الثبات أن يتضمن دليل الاختبار الفترة الزمنية بين التطبيقين، وأعمار المفحوصين، والطبيعة النظرية للسمة المقاسة وغيرها من التفاصيل المهمة التي يتأثر بها صدق وثبات الاختبار.

فعند تقييم الثبات بطريقة إعادة الاختبار هناك اعتبار مهم يجب أن نتطرق إليه وهو المدة الزمنية الفاصلة بين التطبيقين، فيجب أن لا تكون المدة الفاصلة بين التطبيقين قصيرة جدا لاجتناب التعب والتذكر أو التعلم وأن لا تكون طويلة جدا لتجنب التغير في قدرات الشخص. وبالتالي فإن المدة لا يمكن تحديدها وإنما فقط يمكن لمطور الاختبار أو مستخدمه تقدير وتبرير الوقت اللازم بين التطبيقين وذلك حسب خصائص العينة والغرض من الاختبار وغيرها. فالفترة بين تطبيقي الاختبار إذا كانت قصيرة (أي ساعات أو أيام). إن تقدير الثبات يتضخم اصطناعيا نتيجة التذكر والتعود من التطبيق الأول. وإذ كانت الفترة طويلة فإن معامل الثبات ربما ينخفض ليس فقط بسبب عدم استقرار الدرجات، وإنما أيضا بسبب التغيرات الحقيقية التي حدثت في هذه الفترة عند المختبرين[2] . وباختصار فإنه لا توجد فترة زمنية مثالية فكما سبق وأشرنا فالمدة تحدد حسب الغرض من الاختبار، فمدة الاختبار الذي سوف يستخدم في التنبؤ ليست نفسها للاختبار الذي يستخدم في التشخيص، كذلك حسب التكوين الفرضي أيضا فهنالك سمات مستقرة نسبيا مقارنة بسمات أخرى.

---

[1] Drost, E : Validity and Reliability in social science Research. Research Educational Resarch and perspectives .38 (1). 2011, p. 110.

[2] ليفينجستون و رينولدز، المرجع المذكور سابقا، ص 170

وتعتمد هذه الطريقة عادة وبشكل رئيسي على استخدام معامل الارتباط بيرسون بين التطبيقين.[1] إلا أنه تم توجيه العديد من الانتقادات في ما يخص معامل ارتباط بيرسون فهو يعتبر غير كافٍ، وذلك لأن الفروق المنهجية لا تؤخذ في الاعتبار، والمقصود هنا بالفروق المنهجية هي الفروق بين الأفراد في التطبيقين، ولذلك ينصح كل من ستريتر، نورمان وكرني Cairney & Streiner, Norman (2015) باستخدام معامل الارتباط البين صفي أو مايسمى أيضا معامل الارتباط داخل الصف (intraclass correlation coefficients) ويرمز لها اختصارا (ICC) أما بالنسبة للقياسات الترتيبية، يجب استخدام معامل كابا الموزونة. ويعتبر الثبات مقبولا عندما يكون معامل الارتباط داخل الصف أو كابا الموزونة 0.70 على الأقل في حجم عينة لا يقل عن 50 فردا.[2] وسنحاول التفصيل أكثر في ما يخص معامل الارتباط بيرسون أو الصفي في الجانب التطبيقي.

اتفقت كل من (Urbina, 2014, p. 135) و (محاسنة، 2013، ص 126) و (بوسالم، 2014، ص 86) إلى أنّ هذه الطريقة تبدو سهلة وبسيطة، لكن في الحقيقة لها بعض الصعوبات التّي تجعلها منها معقدة بعض الشيء، ومن بين هذه الصعوبات التّي من شأنّها أن تؤثر على أداة الفرد وبالتالي في قيمة ثبات درجات الاختبار نذكر منها:

2-1-1-الفترة الزمنية بين التطبيقين: إن ثبات إعادة التطبيق يعني إجراء اختبار مرتين تفصل بينهما فترة زمنية هذه المدة أو الفاصل الزمني إذا طال فإنّ إجابات المفحوصين تتأثر بالعوامل النضج أو النمو وإذا قصرت فإنّها أيضا تتأثر بالعامل الذاكرة وبالتالي يتذكر المستجيبين

[1] Kubisyzn & Borich, p 338.

[2] Streiner, D., Norman, G., & Cairney, J: *Health measurement scales A practical guide to their development and use* . 5 ed. New York: Oxford University Press. 2015, p 178.

أو المفحوصين إجاباتهم في المرة الأولى لهذا السبب لم يتم تحديد الفاصل الزمني بين التطبيقين بدقة. [1] ولكن هنالك بعض المحاولات لتحديد المجال الزمني التقريبي المناسب فمثلا يرى دومين ودومين Domin & Domin (2006) أنّ أحسن مدة زمنية فاصلة بين التطبيقين تكون بضعة أيام أو أسابيع أو أشهر مع ضرورة ذكر هذه المدة في عرض نتائج معامل الارتباط بين درجات التطبيقين . [2] ، كذلك ذهب محاسنة (2013) إلى تقدير تلك المدة مابين يوم واحد إلى خمسة عشر يوما. [3] . كذلك يجب الإشارة إلى أمر مهم وهو أن معاملات ثبات الإعادة تختلف بطبيعة الحال باختلاف الفترة الزمنية بين التطبيقين. وأن الباحث أو مستخدم الاختبار هو المسئول الأول والأخير عن المدة المناسبة بين التطبيقين، وذلك حسب ما يراه مناسبا للسمة التي يقوم بدراستها والعينة التي سيطبق عليها وغيرها من الظروف التي تجعله يحدد الفترة الزمنية المناسبة. غالبًا ما يكون أسبوعًا أو أسبوعين مناسبًا، لكن كما ذكرنا سابقا لا نحدد مدى ملاءمة الفترة الزمنية، ولكننا نطلب فقط وصف هذه الفترة الزمنية وتبريرها.

2-1-2-التغيرات النفسية والصحية والاجتماعية للمستجيبين مابين التطبيقين: كلما طالت الفترة بين التطبيقين كلما طرأت العديد من التغيرات سواء النفسية أو الاجتماعية وغيرها والتي قد تؤثر بطريقة أو بأخرى على المستجيبين، الأمر الذي من شأنه أن يؤثر على ثبات الاختبار، وقد أشار معمرية (2007) إلى أنّ خصائص القائم بتطبيق الاختبار وطريقته في إلقاء التعليمات وقدرته على ضبط الموافق أثناء الاختبار، كلها أسباب أيضا قد تنبه المفحوصين لدوافع

---

[1] Kubisyzn & Borich, p336

[2] Domino, E., & Domino, L:*Psychological testing an introduction*. 2 ed. cambridge: cambridge university press.2006 , p44

[3] محاسنة، المرجع المذكور سابقا، ص 136.

أخرى كالدافع للتعاون، التزييف أو الإهمال وهذه كلها مسببات في شخصية المجرب وسلوكه اتجاه موقف القياس. [1]

2-1-3-التغيرات الناتجة عن التمرين والتعلم أو التدريب أو الذاكرة: ويعتبر من أهم الصعوبات أيضا التي تواجه هذا النوع من الثبات، وتشير Urbina (2015) إلى أنّ الثبات إعادة الاختبار لا توجد فيه مدة زمنية محددة يمكن التوصية بها في جميع الاختبارات وعليه فالمدة الزمنية القصيرة تجعل المستجيبين يتذكرون إجاباتهم في التطبيق الثاني وبالتالي عامل التذكر على درجات المستجيبين يجعلهم يغيرون إجاباتهم، ومنه يحدث تغير في درجات التطبيق الثاني. [2] وعليه فالفترة الزمنية تؤثر في الجزئيتين السابقتين، فإذا طالت الفترة بين التطبيقين قد تطرأ تغيرات على المستجيبين وبالتالي يؤثر الأمر على الثبات، وإذا قصرت الفترة بين التطبيقين قد يتدخل عامل التذكر والتدريب وغيرها. كذلك هنالك العديد من الصعوبات مثل الوصول لنفس الأفراد في التطبيق الثاني، والتكاليف المادية والمعنوية وكذا الوقت الذّي يتطلبه التطبيقين.

كذلك قد أشار بوسالم (2014) إلى أنّ هذه الطريقة للتحقق من ثبات درجات الاختبار لا تصلح في الاختبارات التّي تقيس سمات سريعة التّغير أو تعتمد على الذاكرة، وعليه فإنّ هذا النوع يفضل استخدامه في حالة الاختبارات النفسية والعقلية وخاصة عند تقييم الاتجاهات والميول والقيم وبعض القدرات والاستعدادات. [3] وعليه فطريقة إعادة الاختبار رغم أنها تبدو سهلة في ظاهرها، حيث يتم تطبيق الاختبار على مجموعة من الأفراد وبعد فترة زمنية يحددها الباحث، يعود ليطبق الاختبار نفسه مرة أخرى على نفس العيّنة وفي نفس الظروف والإجراءات وحساب معامل الارتباط

---

[1] معمرية، المرجع مذكور سابقا، ص 206.

[2] Urbina, p 124

[3] بوسالم، المرجع مذكور سابقا، ص 86

بين درجات التطبيقين، إلا أنها في حقيقة الأمر أصعب مما تبدو وتتخللها العديد من المصاعب والتعقيدات.

2-2-ثبات الصور المتكافئة

إنّ هذه الطريقة تعتمد على بناء صورتين متكافئتين من حيث المحتوى والسمة المقاسة وغيرها من الشروط، ويطبق على العيّنة بفاصل زمني قصير ويحسب معامل الارتباط ين درجات الأفراد. ويشير دروست Drost (2011) إلى أنّ هذه الطريقة تشبه إلى حد كبير الطريقة السابقة (إعادة الاختبار) باستثناء بعض الإجراءات، فهذه الأخيرة يتّم تطبيق اختبار واحد في أوقات مختلفة أما الثانية (الصور المتكافئة) يتم تطبيق صورتين أو شكلين مختلفين مع بعض أو في وقت لاحقا، وحساب الارتباط بينهما، وإذا كان الارتباط منخفض بينهما فإنّه يدل على وجود خطأ في القياس، وبالتالي فإنّ اختبارين لا يقيسان نفس السمة المقاسة. [1] وتساعدنا هذه الطريقة في التّخلص من عامل الذاكرة والممارسة والتدريب الذّي يواجه النوع الأول. على عكس ليمن وميهرن Lehmnn & Mehrens (1973) فيريان أن هذه الطريقة عكس طريقة ثبات إعادة الاختبار، يتم تقدير ثبات الصور المتكافئة بواسطة صورتين أو شكلين من نفس (المحتوى/تعليمات/البنود) من اختبار لنفس المجموعة في نفس اليوم ومن ثم حساب الترابط بين درجات الاختبارين.[2] (Lehmnn & Mehrens, 1973) وأضافت أوربينا Urbina (2014) أنه يمكن تطبيق شكلين مختلفين أو أكثر من الاختبار متطابقان في الغرض ولكنهما يختلفان في محتوى معين، يتم إجراءهما على نفس

---

[1] Drost , p 110

[2] Lehmnn, I., & Mehrens, W: *Measurement and Evaluation In Education and Psychology.* New York: Rinehart and winston.1973.

المجموعة من الأفراد، ثم الربط بين درجات الصورتين حتى يتم الحصول على معامل ثبات التكافؤ، ويهدف هذا الثبات إلى تقدير مقدار الخطأ في درجات الاختبار الذّي يعزى إلى خطأ في عينات المحتوى. [1]

ويذكر لناعلام (2000) [2] أنّ من محاسن هذه الطريقة أنّها تلغي لنا عامل الألفة وأثر التعلم الذي نواجه في ثبات إعادة الاختبار في مقابل نجد صعوبة كبيرة في بناء صورتين متكافئتين لقياس سمة معيّنة تتوفر على كلّ الشروط. أي أننا نواجه هنا تحديا آخر وهو التكافؤ بين الاختبارين، وقد أشار معمرية (2012) إلى أن التكافؤ في الاختبارين يعني أن:

* يكون الاختباران يقيسان نفس الخاصية.

* أن يكون عدد البنود متساويا في كلا الصورتين.

* أن يكون معامل السهولة والصعوبة لكلّ بند من البنود الاختبار متساوي.

* تساوى المتوسطات الحسابية والانحرافات المعيارية لكلا الصورتين.

* أن يصل كلا الصورتين نفس التعليمات.

* التشابه التام في أمثلة التوضيحية.

* تساوي الوقت المخصص لكلّ صورة. [3]

إن ثبات الصور المتكافئة يختلف باختلاف كيفية تطبيق الصورتين فعند تطبيق الباحث للصورتين في نفس الجلسة أو الوقت، فإننا هنا نطلق على معامل الارتباط بين الصيغتين المتكافئتين بمعامل التكافؤ. في حين يطلق على معامل الارتباط بين الصيغتين المتكافئتين التي

---

[1] Urbina, p 137.

[2] علام، مرجع مذكور سابقا، ص 467.

[3] معمرية، مرجع مذكور سابقا، ص 269.

تطبق وبينهما فارق زمني أي يطبق الباحث الصورة الأولى وبعد فترة زمنية معيّنة يطبق الصورة الثانية، بمعامل التكافؤ والاستقرار. وهنا الفترة الزمنية يحددها الباحث ويتم ذكرها في التقرير.

2-3-ثبات الاتساق الداخلي

إن ثبات الاختبار يهتم بتجانس البنود أي ارتباط عناصر الاختبار فيما بينها، وقد أكد passmore (2008) هذا في تعريفه للاتساق الداخلي بأنه "مدى ارتباط عناصر الاختبار التّي تقيس نفس السمة المرتبطة ببعضها البعض". [1] . فالاتساق الداخلي يهتم بتجانس البنود فيما بينها من جهة وتجانسها مع البعد أو المحور التي تنتمي إليه من جهة أخرى. فيتعلق الاتساق الداخلي بثبات مكونات الاختبار، وحسب دروستي Drosrt (2011) هو اتساق داخل البند أو العنصر مع باقي العناصر التّي تقيس سلوكا معينا أو خاصية معينة، أي هو مدى ارتباط بنود أو عناصر الاختبار مع بعضها البعض. [2] وهو ما أكده أيضا النبهان (2013) في تعريفه للاتساق الداخلي بأنّه تجانس فقرات الاختبار، أي مدى ترابط الاستجابات على الفقرة الواحدة مع الدرجة الكلية للاختبار.[3]

ويعتبر ثبات الاتساق الداخلي من أكثر الطرق استخداما وشيوعا في التحقق من ثبات الاختبار حسب [4] ولهذه النوع ثلاثة طرق يتم استخدامها في البحوث النفسية والتربوية هي: التجزئة النصفية، كيودر ريتشاردسون، وألفا كرونباخ، كلّ هذه الطرق تعمل على قياس الاتساق الداخلي.

---

[1] Passmore, J :psychometrics in coaching using psychological and psychometric tools for development. London and Philadelphia: kogan page limited.2008, p 12.

[2] Drost, p 111.

[3] النبهان، مرجع مذكور سابقا، ص 293.

[4] Hogan, p 129.

2-3-1-التجزئة النصفية

هذه الطريقة شبيهة بثبات الصور المتكافئة التّي تعتمد أساسا على استخدام نموذجين أو اختبارين، طريقة التجزئة النصفية قريبة جدا من ثبات التكافؤ، غير أنّ هذه الأخيرة تعتمد على وجود اختبار واحد يتم تقسيمه إلى جزأين أو نصفين، وبالتالي يكون لنا اختبارين ومن ثم حساب الارتباط بين نصفي الاختبار، يجب أن نشير إلى نقطة مهمة جدا في تجزئة الاختبار فهناك تقسيمين حسب هوجان hogan (2013) :

* تقسيم الاختبار إلى بنود فردية وبنود زوجية وهي الأكثر شيوعا.

* تقسيم الاختبار إلى بنود أكثر سهولة وهذا التقسيم له تأثير كبير في ثبات درجات الاختبار.

hogan, (2013) [1]

يجب التنويه هنا بأن درجة ثبات الاختبار تتأثر بطوله، وبالتالي طول الاختبار في الاتساق الداخلي يقلل من قيمة ثبات الاختبار، وعليه فإنّ معامل الارتباط الذّي تحصلنا عليه هو ثبات نصف الاختبار وليس كله، لهذا وجب تصحيحه، وعادة ما يتم استخدام معادلة سبيرمان براون.

2-3-1-1- معادلة سبيرمان براون

التّي نتحصل من خلالها على ثبات كلّ الاختبار. وتكتب معادلتها بالشكل التالي: $\mathrm{P}xx' = \frac{2\rho AB}{1+\rho AB}$

ρAB: معامل الارتباط المعدل للاختبار الكلي

$\mathrm{P}xx'$: معامل ارتباط نصفي الاختبار

[1] Ibid, p 131.

وللإشارة فإنّ قيمة ثبات الاختبار المصحح تكون دائما أكبر من ارتباط النصفين. وما يميز هذه الطريقة أنّها تتطلب إجراء اختبار واحد وبالتالي تستبعد تأثير العامل الذاكرة أو الممارسة.[1] (ولكن ما يعاب على معادلة سبيرمان براون أنها تعتمد بالأساس على افتراض أن نصفي الاختبار متوازيان تماما. وكلما تم انتهاك هذا الافتراض، كانت النتائج أقل دقة. وهذا ما دفع برولون (Rulon) إلى اقتراح طريقة بديلة لتقدير الثبات من درجات نصفي الاختبار، بحيث لا تتطلب هذه الطريقة معادلة سبيرمان براون التصحيحية، كما سيتم توضيحه في العنصر التالي.

2-3-1-2-معادلة رولن

تعتمد الطريقة على تباين درجات نصفي الاختبار عوضا عن معامل الارتباط بيرسون، فهي طريقة تتميز بالسهولة والسرعة في تقدير معامل الثبات، كما أنها طريقة بديلة تتطلب استخدام فرق الدرجات بين نصفي الاختبار أي D = A – B حيث تمثل A درجة الفرد على نصف الاختبار الأول وB درجة الفرد على نصف الاختبار الثاني. ويستخدم تباين فروق الدرجات على أنه تباين الخطأ في الصيغة المحددة لمعامل الثبات: $\acute{P}xx" = 1 - \frac{\sigma 2d}{\sigma 2x}$

$\sigma 2d$: تباين الفرق بين الدرجات الفردية والدرجات الزوجية.

$\sigma 2x$: تباين الدرجات الكلية للاختبار. [2]

ومع أن طريقة رولون هي الطريقة الأسهل استخداما في الحسابات اليدوية، فهي تكافئ معادلة أخرى لتقدير ثبات التجزئة النصفية التي تعود إلى جوتمان (Guttman).

---

[1] Kubisyzn & Borich, p 341.

[2] ألجينا و كركور، مرجع مذكور سابقا، ص 194.

2-3-1-3-معادلة جوتمان

وهي أحد الأساليب التي يستخرج بتطبيقه معامل ثبات التجانس الداخلي، وتستعمل حين يكون التباين دال إحصائيا بين الفقرات الفردية والزوجية كما ذكرنا في طريقة التجزئة النصفية، وتعتمد في حسابها على استخراج قياس تباين الفقرات الفردية وتباين الفقرات الزوجية والتباين الكلي لمجموع درجات المستجيبين على الفقرات. وتتمثل معادلة جوتمان في: $pxx' = 2\left[1 - \frac{(\sigma 2x + \sigma 2X')}{\sigma 2r}\right]$

$\sigma 2x$: ترمز لتباين الفقرات الفردية

$\sigma 2X'$: تباين الفقرات الزوجية

$\sigma 2r$: التباين الكلي للدرجات

ومعامل الثبات (1) [1]. وتستخدم الطريقة أيضا في تقدير معامل ثبات الاختبار بواسطة التجزئة النصفية بين الاختبار الفرعي الأول والاختبار الفرعي الثاني، كما أنها تصلح عندما لا تتساوى الانحرافات المعيارية لنصفي الاختبار.

2-3-2-معادلة التجانس لكيودر وريتشاردسون

طور هذه الطريقة كلّ من جي كيودر وإماديليو ريتشادرسون J. Kewder & Amadelio Richardson (1937)، وسميت نسبة إليهما، وتتطلب هذه الطريقة تحقيق شرط تساوي درجات كل من نصفي الاختبار كما هو الحال في طريقة التجزئة النصفية، وهذه الطريقة كيودر وريتشاردسون جاءت كبديل لمعالجة القصور التّي تخلفها طريقة التجزئة النصفية. وتهدف حسب علام (1999) للتوصل إلى قيمة تقديرية لمعامل ثبات الاختبارات غير الموقوتة أي اختبارات القوة

[1] ياسر خلف الشجيري، و عبد الكريم حيدر الزهيري: *اتجاهات حديثة في القياس والتقويم النفسي والتربوي* . ط 1. الأردن: مكتبة المجتمع العربي ودار الإعصار العلمي.2022، ص 286.

والتّي تكون مفرداتها ثنائية الدرجات، أي أن تعطي درجة(1) للعبارات الصحيحة والدرجة (0) للعبارات الخاطئة. [1] وتعددت صيغ هذه الطريقة إلا أن هناك صياغتين الأكثر شيوعا وهما KR21 – KR20

2-3-2-1-الصيغة (20) لكيودر وريتشاردسون KR20

حسب hogan (2013) تعتبر هذه الطريقة الأكثر استخداما، تستخدم عندما تكون درجات البنود ثابتة (0-1) فقط، صيغتها على النحو التالي:

$$kR=\frac{n}{n-1}\left(\frac{sx2-\sum_p q}{s\times L}\right)$$

حيث: n = عدد البنود للاختبار.

P = نسبة الأفراد الذين أجابوا عن مفردة إجابة صحيحة.

Sx2 = تباين درجات الاختبار (مربع الانحراف المعياري).

q = نسبة عدد الأفراد الذّين أجابوا عن مفردة إجابة خاطئة.

وكما تم الإشارة إليه فإن هذه الصيغة أكثر استخداما من الصيغة الثانية أي كيودر ريتشاردسون 21 سواء في البحوث والدراسات الأكاديمية العربية أو الأجنبية.

2-3-2-2- الصيغة (21) لكيودر وريتشاردسون KR21

تتطلب هذه الطريقة تساوي بنود الاختبار في درجات الصعوبة أو قربها من متوسطها، مع توفر شروط الصيغة 20 من كيودر وريتشاردسون . [2] ولهذا يرى بوسالم (2014) أنّه من النادر

[1] علام، مرجع مذكور سابقا، ص 160.

[2] hogan, p 131.

توفر هذا الشرط وهو تساوي درجة الصعوبة في الاختبارات، لهذا يلجأ الباحثين والمختصين في القياس إلى استخدام الصيغة KR20 وهي الأكثر ملائمة للاختبارات.[1]

صيغة كيودر وريتشاردسون 21 هي كالتالي:

$$rkR21 = \left(\frac{n}{n-1}\right)1-\frac{p(n-p)}{n(\sum_s x2)}$$

n هو عدد بنود الاختبار.

p هو متوسط الدرجات الكلية للاختبار.

Sx2 هو تباين درجات الاختبار.

نستنتج من خلال ما سبق أنّ صيغة كيودر وريتشاردسون KR20 أسهل في استخدامها من صيغة KR21 وذلك لشرط تساوي البنود في درجة الصعوبة والذي يصعب تحقيقه في الصيغة الثانية، وتتفق الصيغتين في باقي الشروط، مثل استخدامهما مع المفردات ثنائية الدرجة، وبالتالي عند استخدام سلم ليكرت الثلاثي أو الخماسي وغيرها، فهنالك طرق أخرى منها طريقة ألفا كرونباخ لتقدير ثبات المقياس غير ثنائي المفردات.

2-3-3-معامل ثبات ألفا كرونباخ (α)

كما تم الإشارة إليه سابقا فإنّ معادلة كيودر وريتشاردسون تستعمل مع البنود ثنائية الدرجات، وبالتالي في حالة استخدام بدائل أخرى مثل مقياس ليكرت الثلاثي أو الخماسي فإننا عادة نستخدم البديل ألفا كرونباخ. وقد تّم اشتقاق هذا المعامل من صاحبه كرونباخ [2],وهو من أهم معاملات

---

[1] بوسالم، مرجع مذكور سابقا، ص 96.

[2] Cronbach , p 132.

الاتساق الداخلي للاختبار، ويعتمد على متوسط ارتباط بنود الاختبار، أي مدى قدرة البنود على قياس وتمثيل بنية السمة المقاسة، ويطلق عليه معامل التجانس. وتعتبر طريقة ألفا كرونباخ حسب بوسالم (2014) بأنَها تعميم لصيغة كودر وريتشاردسون الصبغة 20 لتقدير ثبات درجات أنواع الاختبارات والمقاييس. [1] وصيغة ألفا كرونباخ الإحصائية كالتالي:

$$\alpha=(\frac{n}{n-1})1-(\frac{\sum n2}{sx2})$$

n هو بنود الاختبار.

sn2 هو تباين درجات كلّ بند من بنود الاختبار.

Sx2 هو تباين درجات كلّ الاختبار.

إن الدرجة المحصل عليها من معامل ألفا كرونباخ تساوي متوسط القيم التقديرية لمعامل ثبات كل من نصفي الاختبار لجميع طرق التجزئة النصفية الممكنة كما هو الحال مع معامل التجانس لكيودر وريتشاردسون، وبعبارة أخرى فإنّ معامل ألفا كرونباخ يعطي قيمة تقديرية للارتباط بين عينتين عشوائيتين من المفردات المستمدة من نطاق شامل للمفردات التّي تماثل مفردات كلّ من العينتين. [2]) . ويجدر الإشارة هنا أن معادلة ألفا كرونباخ تستخدم عندما تكون تباينات الفقرات المكونة للاختبار متكافئة، ففي حالة لم تتوفر ظروف التكافؤ للفقرات أو الأجزاء فإن معامل الثبات حسب هذه المعادلة سيكون أقل من قيمة معامل الثبات الحقيقية. [3] (محاسنة، 2013، ص 134) لذلك ينصح تيغزة (2017) باستخدام معادلة ألفا الطبقية Stratified Alpha في تقدير الثبات على مستوى المقاييس المتعددة الأبعاد، بحيث تعطي نتائج أدق بكثير من معامل ألفا العادية كذلك أشار إلى الارتفاع الكبير في معامل ألفا كأن يتجاوز 0.85 أو0.90 قد يندر بوجود

---

[1] بوسالم، مرجع مذكور سابقا، ص 98.

[2] علام، مرجع مذكور سابقا، ص 166.

[3] محاسنة، مرجع مذكور سابقا، ص 134.

تأثير عامل آخر، ولا يدل على كمية تجانس المقياس. [1] أي المقصود هنا أن البنود وجب أن تكون شاملة لكل أبعاد السمة المدروسة، وبالتالي بقدر ما زاد اتساقها دل ذلك على أنها تقيس شيئا مشتركا، أي أنها قد تكون أهملت الجوانب الأخرى وركزت على جانب معين من السمة.

2-3-4-معامل ثبات هويت بتحليل التباين

استخدم هويت أسلوبا جديدا يختلف عن الأساليب السابقة المذكورة في تقدير معامل الثبات يؤدي إلى نتائج متماثلة لتلك الناتجة من معامل ألفا كرونباخ، وتعد هذه الطريقة من طرائق استخراج الاتساق الداخلي، فقد أكد كل من [2] بأن ثبات هويت يعتمد على تحليل التباين وذلك بمعالجة الأفراد والبنود على أنها مصادر للتباين، وتتطلب طريقة "هويت" معرفة طرق إحصائية أكثر تعقيدا من الطرق الأخرى لتقدير معامل الثبات، وباستخدام رموز تحليل التباين المعياري فإن معامل الثبات يحدد على النحو التالي:

$$Ṕxx" = \frac{MSp - MSr}{MSp}$$

MS persons: متوسط مربعات الأفراد المأخوذة من جدول تحليل التباين.

MS residual: متوسط مربعات البواقي المأخوذة من جدول تحليل التباين.

وقد ربط هويت هذه الصيغة بالتعريف النظري لمعامل الثبات، وذلك بملاحظة أن متوسط مجموع المربعات للأفراد MS persons تمثل تباين الدرجة الملاحظة ومتوسط مجموع مربعات البواقي MS residual تمثل تباين الخطأ في التعريف النظري للثبات.

$$Ṕxx" = \frac{\sigma 2x + \sigma 2e}{\sigma 2x}$$

---

[1] تيغزة، مرجع مذكور سابقا، ص 23.

[2] الشجيري و الزهيري، المرجع مذكور سابقا، ص 291.

كما هو ملاحظ من خلال العمليات المستخدمة فإن معادلة هوييت للتحليل التباين رغم أنها قريبة من مفهوم الاتساق الداخلي، إلا أنها في الحقيقة تختلف عنه [1] ، بافتراض أن تباين الخطأ الخاص بكل مكون يتوزع إعتداليا وأن تباين الخطأ يتساوى بالنسبة لكل سؤال من أسئلة الاختبار، كذلك معامل الارتباط بين الأسئلة أي أن معامل الارتباط بين السؤال الأول والثاني يساوي معامل الارتباط بين السؤال الثاني والثالث وهكذا. وهذه الافتراضات بالطبع تعتبر صعبة التحقيق مما يجعل هذه المعادلة غير صالحة في أغلب الاختبارات ونادرة الاستخدام.

2-3-5-التحليل العاملي

تعتمد طريقة التحليل العاملي على اختزال عدد من العوامل أو التكوينات الفرضية التي تفسر معاملات الارتباط بين مفردات الأداة، وتصنيفها في مجموعات متجانسة تقيس كل واحدة منها سمة معينة فإذا تم إيجاد عامل مشترك فإن هذا يعني أن مفردات الأداة متجانسة فيما تقيسه. [2] ويستخدم التحليل العاملي للكشف عن العوامل التي تتكون منها الأداة، فعندما تكون الأداة عبارة عن اختبار متعدد الأبعاد أو يقيس أكثر من سمة فإن تحليل العوامل يكشف عن ذلك. ويفضل أن يكون عدد أفراد العينة من خمسة إلى عشرة أضعاف عدد مفردات الأداة على الأقل، فكلما كان التباين بين أفراد العينة أكبر كانت العوامل الناتجة أكثر وضوحا ودلالة. [3] وقد فصلت الباحثة أكثر في ما يخص التحليل العاملي في الجزء المتعلق بالصدق، وذلك لكونه أكثر استخداما في الصدق منه في الثبات في الدراسات، ومع أن الاتساق الداخلي سواء بالتحليل العاملي أو بغيره،

---

[1] الامام، عبد الرحمان، و العجيلي، ص 161.

[2] علام، ص 167.

[3] عبد الرحمان بن صافي، و علي طاجين: تقييم طرق تقدير صدق أدوات القياس المستخدمة في البحوث النفسية والتربوية. *سلوك مخبر تحليل المعطيات الكمية والكيفية* (3). 2016، ص 45.

يكشف عن تناسق البنود وتجانسها وبالتالي ثباتها ودقتها، إلا أنه يستخدم في الصدق بكثرة رغم أنه لا يكشف الكثير عن ماهية هذه البنود ومحتواها.

2-3-6-ثبات تقديرات المحكمين (ثبات المقدَرين)

يعتبر هذا النوع من الثبات شكلا من أشكال الاتساق الداخلي، وهو يعتمد في الأساس على الأحكام الذاتية للمحكمين شأنه شأن صدق المحكمين، أي يتم استخدام محكمين أو مقيمين لقياس السلوك ويتم تقييم الثبات من خلال أحكامهم أو الاتساق الداخلي المشترك للأحكام، ومن ثم حساب مدى ارتباطهما.وقد أكد كوالي coaley (2010) بأن ثبات المحكمين قائم على التقييمات الذاتية من طرف المقيم أو المصحح أو المحكم اتجاه الاختبارات كالمقابلات و الملاحظات حيث يقيم المحكمين أو للمقيمين للأفراد بشكل مستقل بعد خضوعهم إلى التدريب حتى يضمنوا نسبة من الاتساق والثبات العالي، ومن ثم الاعتماد على مقياس واحد متفق عليه لتقييم الأفراد وربط بين درجات المحكمين أو المصححين عن طريق معامل الارتباط أو نسبة الاتفاق. [1] وعليه هذا النوع يعتمد على الذاتية وفي المقابل الأنواع السابقة للثبات تعتمد على التقييمات الموضوعية البعيدة عن ذاتية الباحث.

ويشير كوهين وسوردليك Cohen & swerdlik (2009) إلى ثبات المحكمين بأنه درجة الاتساق أو الاتفاق بين اثنين من المقيمين أو أكثر يتعلق بمقياس معين، بصفة أيضا بثبات

[1] Coaley, K. :*An introduction to psychological Assessment and psychometrics.* London: Sage Publication.2010, 110

المراقبين، وحتى يتم تحديد درجة الاتفاق بين المصححين يتم حساب معامل الارتباط عليه بمعامل ثبات بين المصححين. [1]

ويعرف بأنه تقييم الاتساق لدرجات الخاصة بالأداء من قبل مقيمين مختلفين وشكل مستقل ثمّ الربط بين درجاتهم من خلال وضع عدد فترات التّي يتلقى فيها أداء الفرد نفس الدرجات من المقيمين، والنسبة التي تقع ضمن نقطة واحدة لبعضهم البعض. ويعتبر هذا النوع من الثبات ضروري في الاختبارات الإسقاطية والإبداعية والأخلاقية التّي تترك عادة جانب من التقدير للفاحص أو المصحح، وهذا الأخير يعد عامل رئيسي في تقدير ثبات هذه الاختبارات الذّي يعتمد على ذاتية الفاحصين.[2] ولهذا أشار معمرية (2012) إلى نقطة مهمة في ما يتعلق بثبات المصحح، وهو مدى تغيير النتائج في الدقة و الثقة، إذا ما تغير القائم بالتقدير والتصحيح.[3]

ويمكن حساب تقديرات المحكمين من خلال تطبيق معادلات خاصة بالاتفاق، ومن أهمها المعادلات التالية:

$$\text{معادلة هولستي Holsyi} = \frac{\text{القيم المشتركة} \times \text{عدد المحللين}}{\text{إجمالي قيم المحلل الأول} + \text{إجمالي قيم المحلل الثاني}}$$

$$\text{معادلة كوبر Coper} = \frac{\text{عدد مرات الاتفاق}}{\text{عدد مرات الاتفاق} + \text{عدد مرات عدم الاتفاق}} \times 100$$

ويفضل كوبر أن لا يقل معامل الاتفاق أو الثبات 75%

$$\text{معادلة سكوت Scott:} \frac{\text{مجموع الاتفاق بين الملاحظين} - \text{الاتفاق الناجم عن الصدق}}{\text{أكبر اتفاق ممكن} - \text{الاتفاق ناجم عن الصدق}}$$

[1] Cohen, R., & Swerdlik, M. *Psychological testing and assessment:An introduction to tests and measurement* .ed 7.new york: Mc Gruw-Hill.2009, p 150.

[2] Gregory, p 109.

[3] معمرية، مرجع مذكور سابقا، ص 266.

وهذه المعادلة أي معادلة سكوت، مناسبة حينما يكون عدد الملاحظين لا يقل عن ثلاثة، أما في حالة ملاحظين اثنين تستعمل معادلة كوبر أو معادلة هولستي. [1] إن لكل معادلة ميزة خاصة بها تميزها عن باقي المعادلات وإن صح التعبير شروط خاصة بها فكما أوضحنا سابقا مثلا يشترط كوبر أن لا يقل معامل الاتفاق أو الثبات عن 75% وإن قل عن ذلك فهذا يعني أن الاختبار غير ثابت أي أنه صالح.

3-الخطأ المعياري للقياس

نظرا لأن القياس النفسي يتأثر بمصادر أخطاء متعددة، فإنه لا يوجد اختبار يقيس سمة معينة لدى الفرد قياسا دقيقا تاما، لذلك ينبغي تقدير قيمة هذه الأخطاء، وأخذها بعين الاعتبار عند تفسير درجات الاختبار. فإذا استطعنا تقدير الدرجات الحقيقية للأفراد في الاختبار، فإن الانحراف المعياري لتوزيع هذه الدرجات يطلق عليه الخطأ المعياري للقياس Standard Error of Measurement ويختصر بـSEM، والذي يعرفه علام (2018) بأنه مؤشر لمدى اختلاف الدرجات الملاحظة للأفراد في الاختبار عن درجاتهم الحقيقية. ويفيد هذا المؤشر في تقدير قيمة الخطأ المعياري للقياس لكل فرد في الاختبار والخطأ المعياري للقياس هو الانحراف المعياري للخطأ، ويزودنا بقيمة مطلقة وليست نسبية لمدى انطباق الدرجات الملاحظة والدرجات الحقيقية. فمعاملات الثبات تعتمد على تباين درجات مجموعة الأفراد، بينما الخطأ المعياري للقياس لا يتأثر بهذا التباين. وتقدير الخطأ المعياري للقياس لا يتطلب معرفة الدرجات الحقيقية. [2] وفي تعريف آخر يعتبر الخطأ المعياري للقياس هو الانحراف المعياري لتوزيع الدرجات التي سوف يحصل عليها أحد الأشخاص إذا اختبر بعدد لا نهائي من صيغ متوازية للاختبار الذي يشتمل على

---

[1] الشجيري و الزهيري، المرجع مذكور سابقا، ص 393.

[2] علام، مرجع مذكور سابقا، ص 101

مفردات تمت معاينتها عشوائيا من نفس نطاق المحتوى. فوجود أخطاء القياس سوف يمنع الشخص من الحصول على نفس الدرجة في كل مرة. وعلى الرغم من أن كل اختبار ربما يمثل نطاق المحتوى بنفس القدر، إلا أن الفرد المختبر ربما يكون أداؤه أفضل في بعض الاختبارات، وأسوأ في اختبارات أخرى نظرا للأخطاء العشوائية، وبأخذ جميع الدرجات التي تم الحصول عليها من جميع هذه الاختبارات، فإنه ينتج توزيعا للدرجات. ومتوسط هذا التوزيع هو الدرجة الحقيقية للفرد، والخطأ المعياري للقياس وهو الانحراف المعياري لهذا التوزيع المتعلق بأخطاء الدرجات. [1]

3-1-تقييم الخطأ المعياري للقياس

إن الخطأ المعياري للقياس حسب دراسة كوسمين COSMIN هو الخطأ المنهجي لدرجة المفحوص التي لا تعزى إلى التغييرات الحقيقية في البنية المراد قياسها [2] ,، فالخطأ المعياري للقياس هو دالة للثبات (رس س)، والانحراف المعياري لدرجات الاختبار. وعند حساب الخطأ المعياري للقياس، يأخذ معامل الثبات بعين الاعتبار أخطاء القياس التي توجد في درجات الاختبار، والانحراف المعياري، وهو يعكس تباين الدرجات في التوزيع. ويتم تقدير الخطأ المعياري للقياس باستخدام الصيغة:

الخطأ المعياري للقياس =ع× $\sqrt{1 - \text{رس س}}$

ع= الانحراف المعياري للدرجات الملاحظة

رس س= ثبات درجات الاختبار

---

[1] ليفينجستون و رينولدز، المرجع مذكور سابقا، ص 194.

[2] Mokkink, et al, p 742.

وفي الأخير يجب ملاحظة أنه كلما انخفضت قيمة ثبات درجات الاختبار، ازدادت قيمة الخطأ المعياري للقياس. ونظرا لأن معامل الثبات يعكس الجزء من تباين الدرجة الملاحظة الذي يعزى إلى تباين الدرجة الحقيقية، والخطأ المعياري للقياس هو تقدير لمقدار الخطأ في درجات الاختبار، فهذه العلاقة العكسية هي ما نتوقعه. [1] . أي كلما زادت قيمة ثبات درجات الاختبار، انخفضت بالطبع قيمة الخطأ المعياري للقياس، وبالتالي ازدادت ثقتنا في دقة نتائج الاختبار. وكلما انخفضت قيمة الثبات، ازدادت قيمة الخطأ المعياري للقياس، وقلت ثقتنا في دقة نتائج الاختبار.

3-2-المصادر الرئيسية من خطأ التباين المقترن بأنواع الثبات

إن الظروف المختلفة تستدعي تقديرات ثبات مختلفة، لذلك ينبغي أن نراعي عدة عوامل عند اختيار نوع معين من الثبات دون الآخر، فطبيعة التكوين الفرضي المدروس والغرض من تطبيق الاختبار والأهم من كل ذلك الخطأ المراد ضبطه، فحسب [2] (ليفينجستون و رينولدز، 2013، ص 184) إذا صمم الاختبار كي يطبق أكثر من مرة واحدة على نفس الأفراد، فإن ثبات إعادة الاختبار وثبات الصيغ المتكافئة بتطبيق مؤجل يكونان مناسبين، لأنهما أكثر حساسية لأخطاء القياس الناجمة عن معاينات الوقت، في حين يكون تقدير الاتساق الداخلي مناسبا عندما يطبق مرة واحدة فقط، وثبات التجزئة النصفية يقدَر ثبات تباين الخطأ الناجم عن معاينات المحتوى، بينما معامل ألفا والمعامل KR-20 يقدَران تباين الخطأ الذي يعزى إلى معاينات الوقت وعدم تجانس المحتوى، ونظرا لكونهما حسَاسين لعدم تجانس المحتوى فإنه يمكن تطبيقهما عندما يقيس الاختبار نطاق متجانس أو خاصية أحادية البعد، على عكس المحتوى الغير متجانس يفضل استخدام طريقة التجزئة النصفية.

---

[1] ليفينجستون و رينولد، المرجع المذكور سابقا، ص194.

[2] ليفينجستون و رينولد، المرجع المذكور سابقا، ص 195.

4-العوامل المؤثرة في الثبات

هنالك العديد من العوامل التي قد تؤثر في قيمة الثبات، وستحاول الباحثة ذكر أهمها فيما يلي:

4-1-عدد الأسئلة (طول الاختبار): إن طول الاختبار يؤثر بطريقة مباشرة في ثباته وحسب هوجان hogan (2013) أنّه كلما زاد عدد بنود الاختبار كلما زاد ثباته، وكلما زاد المتوسط يزداد الارتباط بين البنود ويزيد ثباته، وعلاوة على ذلك فإنّه عندما يكون هناك عدد قليل من البنود، يكون الثبات منخفض جدا، وبالتالي الارتباطات بين البنود منخفضة والعكس صحيح، وبذلك نستنتج أنّ معامل ألفا يعتمد على متوسط الارتباط بين البنود، ويشير إلى المدى الذّي تقيس به بنود الاختبار نفس البنية أو السمات ويطلق عليه تجانس البنود. [1] أي أنّ القيمة العددية لمعامل الثبات تزداد تبعا لزيادة عدد الأسئلة، وهذا يعني أن معامل ثبات الاختبار الطويل أكبر من معامل ثبات الاختبار الذي تنقص عدد أسئلته حيث إنّ زيادة عدد الأسئلة تزيد من احتمال تمثيل العينة السلوكية بالاختبار، وبالتالي يزداد ثباته، أي أنّه كلما زاد عدد بنود الاختبار، كلما زاد احتمال تمثيل العينة السلوكية بالاختبار، وعلى هذا يكون من الطبيعي زيادة عدد أسئلة الاختبار إذا أردنا زيادة معامل الثبات إلى قيمة معينة.

ويرى أبو عقيل (2017) أنّ طول الاختبار هو أهم عامل يؤثر في ثبات الاختبار، ولهذا عند حساب ثبات الاختبار عن طريق التجزئة النصفية فإن معامل الثبات يكون منخفض نوعا ما وعند استخدام معادلة سبيرمان- براون يرتفع معامل الثبات، وهذا يعني أننا إذا أردنا تحديد معامل ثبات مرغوب فيه لأي اختبار فيمكن أن تحصل عليه من خلال إضافة عدد آخر من البنود إلى

---

[1] Ibid, p 133.

الاختبار مما يؤدي إلى رفع ثباته.[1] ولكن على الباحث أو مطور الاختبار الانتباه إلى أن هذه الزيادة يجب أن تكون مدروسة، فالعبرة ليست بعدد البنود وحسب بل بشموليتها للسمة المدروسة وملاءمتها أيضا، فلا يجب التغاضي عن التمثيل الصحيح للمجال السلوكي والركض وراء الأرقام فقط.

4-2-زمن الاختبار: يتأثر ثبات الاختبارات الموقوتة بالزمن المحدد لها، حيث يزداد الثبات كلما زاد الزمن المخصص للاختبار حتى يصل الزمن إلى الحد المناسب ويصل الثبات إلى نهايته العظمى، ثم يقل الثبات بعد ذلك كلما زاد الزمن عن ذلك الحد، أي أن معامل الثبات يزداد بزيادة الوقت الذي يستغرقه إجراء الاختبار إلى درجة معينة، ثم ينخفض الثبات إذا كان الزمن أكبر من اللازم، كذلك فإنّ الاختبارات المحددة بزمن معين تمثل بأن تكون معاملات ثباتها مرتفعة أكثر من الاختبارات التي تمنح متسعا من الوقت. [2] وعليه ينبغي تحديد الزمن المطلوب للاختبار وفقا للجهد المطلوب للإجابة عليه ومدى قدرة الفرد على الاستمرار في الإجابة دون تعب أو جهد. وأيضا غير بعيد عن مفهوم الزمن فإن الاختبارات التي يعتمد على استخراج ثباتها من خلال إعادة التطبيق أي سواء باستخدام معامل الاستقرار أو التكافؤ، تتأثر بعامل الزمن أيضا فكلما زادت المدة الفاصلة بين التطبيقين، انخفضت قيمة معامل الثبات.

4-3-استقلالية بنود الاختبار: يجب أن تكون البنود مستقلة حتى ولو كانت تقيس خاصية واحدة، وحسب معمرية (2012) يعني الاستقلال من إجابة معينة على بند إلى إجابة على بند آخر، مثل أن تؤدي إجابة النفي أو الإثبات على بند معين إلى الإجابة بنفس الطريقة على بنود أخرى، لأنّ هذا يؤدي ضمنيا إلى انخفاض عدد البنود، ويقلل من الفروق بين الأفراد، مما يؤدي

[1] أبو عقيل، مرجع مذكور سابقا، ص 135.

[2] الامام، عبد الرحمان، و العجيلي، المرجع مذكور سابقا، ص 164.

إلى انخفاض معامل الثبات. [1] لذلك وجب أن تكون البنود مستقلة عن بعضها البعض من أجل الرفع من قيمة الثبات وأيضا من أجل تمثيل السمة المراد قياسها عن أكمل وجه دون الحاجة إلى تكرار البنود بدون هدف واضح.

4-4-ظروف إجراء الاختبار: إن ظروف تطبيق الاختبارات يجب دائما أن تكون مدروسة ومقننة، أي أنها تستند إلى خطوات ومعايير واضحة يضعها صاحب الاختبار بما يخدم الهدف من تطبيق الاختبار. ولهذا يرى الشجيري و الزهيري (2022) بأنّه أي تغيير في الظروف الخاصة بتطبيق الأداة من وقت إلى آخر يؤثر على نتائج الأداة، وذلك من خلال تأثيره على عوامل الخطأ، والذّي بدوره يؤثر في ثبات الأداة. وذلك لأن معامل الثبات يتأثر بتوحيد التعليمات، وتهيئة ظروف مشابهة لإجراء الاختبار في المرات المتتالية، وأي تغيير في هذه الظروف من تطبيق لآخر يؤثر في نتائجه ويعتبر من عوامل الخطأ التّي تقلل من معامل ثبات الاختبار.

4-5-تجانس العينة: إن ثبات الاختبار يرتبط ارتباطا وثيقا بتباين درجاته وكلما زاد تباين الدرجات زاد ثبات الاختبار وكلما قل تباين الدرجات نقص ثبات الاختبار. وحسب أبو عقيل (2017) يؤدي التجانس الشديد في العينة إلى انخفاض ملموس في معامل الثبات، وذلك لأن التباين داخل هذه العينة المتجانسة يكون منخفضا، وهنا كلما كبر حجم العينة وتزايد عدم تجانسها يؤدي هذا إلى تباين خطأ صغير، وبالتالي تؤدي إلى ارتفاع ثبات الاختبار. واختصر ميهرن وليمن Mehrens & Lehmann (1991) كلما سبق في جملة واحدة " كلما كانت المجموعة غير متجانسة كلما زادت ثبات الاختبار المطبق والعكس صحيح". [2] إذا فكلما كان الأفراد الذين يطبق

---

[1] بشير معمرية، مرجع مذكور سابقا، ص 230

[2] Lehmnn, I., & Mehrens, W: Measurement and Evaluation In Education and Psychology. New York: Rinehart and winston.1973, p 253.

عليهم الاختبار أقل تجانسا ومن مستويات مختلفة وتباينت درجاتهم على الاختبار فإن ذلك يزيد من معامل الثبات، أما إذا كان المختبرون أكثر تجانسا ومتقاربين في مستوياتهم ودرجاتهم على الاختبار فإن ذلك يقلل من معامل الثبات.

4-6-التخمين: صعوبة أسئلة الاختبار، وغموض تعليماته وغيرها قد تجعل بعض الأفراد وبخاصة ضعيف الاستعداد يعتمدون على حل الأسئلة عن طريق التخمين. ويرى محاسنة (2013) أنّ كلما خمن المفحوصين في الإجابة يقل ثبات الاختبار فقد تختلف الإجابة والتخمين في مرتين للتطبيق. [1] وهذا يظهر خاصة في الاختبارات الموضوعية (الاختبار من متعدد)، فيلجأ بعض من المفحوصين في حالة عدم تأكدهم من الإجابة الصحيحة إلى التخمين مما يؤدي إلى خفض ثبات الأداة فزيادة أثر التخمين يقود إلى نقص الثبات. [2] إذا التخمين يقلل من معامل ثبات الاختبار، لأن الإجابة التي اعتمدت على التخمين في المرة الثانية لإجراء ذلك الاختبار على نفس المجموعة، قد لا تكون هي نفسها التي خمنت في المرة الأولى، وبذلك تضعف الصلة بين نتائج المرتين، وتنخفض تبعا لذلك القيمة العددية لمعامل الثبات، فهي تجعل الأفراد الذين لا يعرفون الإجابة يعتمدون في حل الاختبار على التخمين، وإذا حدث ذلك زاد تباين الخطأ وقل ثبات الاختبار.

4-7-مستوى صعوبة بنود الاختبار: زيادة صعوبة الاختبار، وصعوبة مفرداته يجعل الأفراد يلجأون إلى التخمين العشوائي في إجابتهم. [3] وهو ما فصلنا فيه في النقطة السابقة، فهذه تجر لتلك، ولهذا يرى معمرية (2012) بأن هذا العامل يتعين أن يكون مستوى صعوبة البنود كلها

---

[1] محاسنة، المرجع مذكور سابقا، ص 141.

[2] الشجيري و الزهيري، المرجع مذكور سابقا، 296.

[3] صلاح الدين محمود علام : الإختبارات والمقاييس التربوية والنفسية.ط 6. عمان، الأردن: دار الفكر ناشرون وموزعون.2018، ص 104.

متقاربة فإذا كانت البنود تتراوح صعوبتها بين 40 % إلى 60 % فإن معامل الثبات يكون مرتفع أما إذا كان بعضها سهل جدا، وبعضها الآخر صعب جدا، فإنّها تؤدي إلى انخفاض معامل الثبات. [1] فإذا كان مدى الصعوبة ضيقا بين أسهل بند وأصعب بند، فإن ثبات الاختبار يصبح مرتفعا، لأن البنود شديدة الصعوبة أو السهولة لا تميز بين الأفراد، وبالتالي تزيد من تجانسهم، وتخفض من قيمة الثبات، ولذلك نولي أهمية لتضييق مدى الصعوبة، إلا أن هذا التضييق يتطلب تدرج البنود في الصعوبة داخل هذا المدى.

4-8-موضوعية التصحيح: تعد موضوعية التصحيح من بين العوامل المؤثرة في ثبات الاختبار وعلى الأخص في الاختبارات التي تعتمد على تقدير المصحح كاختبارات المقال التحصيلية، اختبارات الإبداع والاختبارات الإسقاطية، حيث أن تباين التصحيح يؤثر في زيادة تباين الخطأ وبالتالي إلى نقصان معامل الثبات. [2] . إذا ترفع موضوعية الاختبار من معامل ثباته، ولكي تتحقق الموضوعية، من الضروري التحقق من ثبات المختبرين، وثبات المصححين، وثبات التعليمات ووضوحها.

5-الثبات وفق دراسة كوسمين COSMIN

إن الثبات حسب دراسة COSMIN يتمثل في عدة تعريفات تكمل بعضها بعضا فهو درجة الترابط بين العناصر، وكذلك درجة خلو القياس من خطأ القياس، ويعبر عنه أيضا بمدى استقرار درجات المفحوصين الذين لم يتغيروا في القياس المتكرر بموجب عدة شروط طبعا، وهو أيضا

---

[1] بشير معمرية، مرجع مذكور سابقا، ص 290.

[2] الامام، عبد الرحمان، و العجيلي، مرجع مذكور سابقا، ص 165.

نسبة التباين الكلي في القياسات التي ترجع إلى اختلافات بين المفحوصين. [1]., إذا الثبات يشير إلى مدى استقرار أو اتساق أو دقة أداة ما. وقد يختلف اختيار الاختبارات الإحصائية المستخدمة لتقييم الثبات، تبعاً لما يعتزم قياسه. ومن المهم التذكير أن الثبات ليست خاصية ثابتة للاختبار. بل على العكس من ذلك، يعتمد الثبات على وظيفة الاختبار والغرض منه، وعلى المجتمع الذي يستخدم فيه، في السياق، وفي الظروف؛ وعلى أساس ما هو موجود في بنود هذا الاختبار؛ أي أنه لا يمكن اعتبار نفس الاختبار ثابتا وموثوقاً به في ظل ظروف مختلفة. [2] وتتأثر تقديرات الثبات بعدة جوانب أهمها خصائص العينة، نوع الأداة، طريقة الإدارة والطريقة الإحصائية المستخدمة. وبالتالي؛ لا يمكن تفسير نتائج البحوث التي تستخدم أدوات القياس إلا عندما تعرض بوضوح شروط التقييم والنهج الإحصائي المستخدم. أي أنه لا يستعمل منفردا [3]

وتصنف COSMIN الثبات في ثلاثة خصائص، ويرد أدناه وصف لمعايير هذه الخصائص التي تعتبر ذات أهمية كبيرة للباحثين ألا وهي: إعادة التحقق أي معامل الاستقرار، الاتساق الداخلي، والصور المتكافئة أي معامل التكافؤ. كما سنصف الطرق الإحصائية الأكثر استخداماً لتقييم كل جانب من جوانب هذه الخصائص.

---

[1] Mokkink, et al, p 540.

[2] Keszei, A., Novak, M., & Streiner, D: Introduction to health measurement scales. *J Psychosom Res .68* (4). 2010, p 319-323.

[3] Terwee, C. B., Schellingerhout, J. M., Verhagen, A. P., Koes, B. W., & de Vet, H. C :Methodological Quality of Studies on the Measurement Properties of Neck Pain and Disability Questionnaires: A Systematic Review. *Journal of Manipulative and Physiological .34* (4). 2011, p261-272.

6-معايير الثبات حسب COSMIN

وللإشارة هنا فإن الباحثة لا يمكنها عرض جميع المعايير بالتفصيل لعدة أسباب سبق وتم ذكرها في فصل الصدق، ولذلك سنشير إلى بعضها فقط هنا. وللاطلاع على جميع المعايير يرجى النظر في الملحق رقم (03).

6-1-بعض معايير إعادة الاختبار والصور المتكافئة

يقيس الاستقرار مدى تشابه النتائج عند قياسها في وقتين مختلفين، أي أنه يقدر اتساق تكرار القياس. ويمكن إجراء تقييم الاستقرار باستخدام أسلوب اختبار إعادة الاختبار. ويتمثل الإجراء في تطبيق نفس المقياس في وقتين مختلفين. ويتطلب استخدام هذه الطريقة أن يظل العامل الذي يجب أن يكون مقاييس هو نفسه في كل من لحظات الاختبار وأي تغيير في النتيجة يمكن أن يكون سببه أخطاء عشوائية. [1]

يجب أن يكون المفحوصين مستقرين فيما يتعلق بالبنية المراد قياسها بين التطبيقات. ويعتمد ذلك على البنية المراد قياسها والعينة المستهدفة. فعندما يتم تدخل أي متغيرات في الفترة المؤقتة بين التطبيقين يمكن للباحث عندها أن يفترض أن (العديد من) المفحوصين قد تغيروا في البنية المراد قياسها. في هذه الحالة لا ينصح باستخدام الطرق التي تعتمد على تطبيق الاختبار مرتين.

كذلك وجب أن تكون الفترة الزمنية بين التطبيقات مناسبة. فيجب أن يكون الفاصل الزمني طويلًا بما يكفي لمنع تحيز الاسترجاع، وقصيرًا بدرجة كافية لضمان عدم تغير في البنية المراد قياسها لدى المفحوصين. فالفاصل الزمني المناسب يعتمد على البناء المراد قياسه والعينة

[1] de Vet, Terwee, Mokkink, & Knol, p 115.

المستهدفة أيضا. وغالبًا ما يُعتبر الفاصل الزمني الذي يبلغ حوالي أسبوعين مناسبًا في تقييم الاختبارات.

كذلك ظروف الاختبار يجب أن تكون متشابهة، تشير شروط الاختبار إلى كيفية التطبيق، والإعداد الذي تم فيه استخدام الأداة، والتعليمات المقدمة إلى آخره. فقد تؤثر ظروف الاختبار هذه على استجابات المفحوصين. قد يتم التقليل من قيمة الثبات إذا كانت ظروف الاختبار غير متشابهة. [1]

وبالطبع تؤثر خيارات الاستجابة على نوع الثبات المستخدم، فللدرجات المستمرة يفضل استخذام معامل الارتباط داخل الصف intraclass correlation coefficient (ICC)، لتقييم ثبات الاختبار، وتعتبر معاملات ارتباط بيرسون وسبيرمان ليست مناسبة لهذا النوع من الثبات، ومشكوكًا في نتائجها خاصة عندما لا يكون من الواضح ما إذا كانت هناك اختلافات منهجية حدثت، لأن هذه الارتباطات لا تأخذ الخطأ المنهجي والفروق بين الأفراد في الاعتبار. إذن فمعلمة القياس المناسبة لتقدير ثبات إعادة الاختبار مع القياسات المستمرة (درجات المقياس) هو معامل الارتباط داخل الصف (ICC)، وليس معامل الارتباط بيرسون، وهناك فئتان أساسيتان من معاملات الارتباط داخل الصف؛ واحد للاتساق والثاني للاتفاق، فيجب على الباحثين الإبلاغ عن معامل الارتباط داخل الفئة التي تم استخدامها. [2]

بالنسبة للدرجات ثنائية التفرع أو الدرجات الاسمية، فإن كابا كوهين هي الطريقة الإحصائية المفضلة أما بالنسبة للمقاييس الترتيبية فيفضل استخدام كابا مرجح. وسواء تم استخدام معامل الارتباط داخل الصف أو كابا المرجحة، فإن كلاهما يجب أن يتجاوز 0.70 عند دراسة الثبات،

---

[1] Mokkink, et al, p 545.

[2] de Vet, Terwee, Mokkink, & Knol, p 517.

وفي حجم عينة تتكون من 50 فردا فما فوق.[1]

6-2-بعض معايير الاتساق الداخلي

الاتساق الداخلي هو قياس لمدى ارتباط العناصر الموجودة في مقياس فرعي ومتجانس، وبالتالي قياس نفس المفهوم، فالاتساق الداخلي هو خاصية قياس مهمة للاختبارات التي تهدف إلى قياس مفهوم أساسي واحد (بناء) باستخدام عناصر متعددة. وفي المقابل بالنسبة للاختبارات التي تكون العناصر فيها مجرد جوانب مختلفة لظاهرة إكلينيكية معقدة لا يجب أن تكون مترابطة، كما هو الحال في مقياس أبغار، فإن الاتساق الداخلي يعتبر غير مناسب. لأنه يقيس عدة سمات[2] أي أنه أي قد يشير تقدير الاتساق الداخلي المنخفض إلى أن البنود تقيس بنيات مختلفة أو أن الإجابات على أسئلة الاختبار غير متسقة.

يتم تحقيق مقياس متسق داخليًا (متجانس أو أحادي البعد) من خلال تعريفات البنية الجيدة، والعناصر الجيدة، ثم تحليل المكون الرئيسي أو تحليل العامل الاستكشافي، متبوعًا بتحليل العاملي التوكيدي. عندما يكون الاتساق الداخلي مناسبًا، يجب تطبيق التحليل العاملي الاستكشافي لتحديد ما إذا كانت العناصر تشكل مقياسًا كليًا واحدًا (بعدًا) أو أكثر من مقياس واحد، في حالة عدم وجود فرضية مسبقة فيما يتعلق بأبعاد الاختبار، ويمكن تطبيق تحليل العاملي الاستكشافي. ولكن إذا كانت هناك فرضية واضحة تتعلق ببنية العامل، فيجب استخدام التحليل العاملي التوكيدي.[3]

بعد تحديد عدد المقاييس (المتجانسة) (الفرعية)، يجب حساب ألفا كرونباخ لكل مقياس (فرعي) على حدة. يصعب تفسير ألفا كرونباخ على أساس مقياس ليس أحادي البعد. لذلك توصي

---

[1] Terwee, et al, p 1165.

[2] Mokkink, et al, p 739.

[3] Mokkink, et al, p 1174

COSMIN بعدم استخدامه للمقاييس متعددة الأبعاد أي عندما يكون هناك دليل على أن هذه المقاييس ليست أحادية البعد. فللحصول على تفسير مناسب لمعامل الاتساق الداخلي، ويجب حساب معامل الاتساق الداخلي لكل مقياس فرعي أحادي البعد على حدة. ويمكن العثور على معلومات حول أحادية البعد إما في نفس الدراسة أو من دراسات أخرى.[1] وتشير ألفا كرونباخ المنخفضة إلى عدم وجود ارتباط بين العناصر في المقياس، مما يجعل تلخيص العناصر غير مبرر. أما ألفا كرونباخ العالية جدًا فتدل على ارتباطات عالية بين العناصر الموجودة في المقياس، أي التكرار في عنصر واحد أو أكثر. علاوة على ذلك، عادةً ما يتم العثور على ألفا كرونباخ المرتفع جدًا للمقاييس التي تحتوي على عدد كبير من البنود، أيضا وجب الإشارة إلى أن ألفا كرونباخ لا يعطي أي معلومات عن عدد المقاييس الفرعية في الاختبار، لأن ألفا يمكن أن تكون عالية عندما يتم الجمع بين مقياسين فرعيين أو أكثر. وتقترح COSMIN معيارًا يتراوح بين 0.70 و0.90 كمقياس للاتساق الداخلي الجيد. [2]

7-العلاقة بين الصدق والثبات

يرتبط الصدق والثبات ارتباطا وثيقا، فكلاهما وجهان لشيء واحد، هو مدى صلاحية ذلك المقياس في أن يقيس ما وضع لقياسه وفي إعطائه نتائج متماثلة، حيث أن الثبات يبحث في مدى اتساق مفردات الأداة، بينما الصدق يتعلق بالغرض الذي بنيت من أجله هذه الأداة، لذا ينبغي في

---

[1] Mokkink, Vet, Prinsen, Patrick, Alonso, & Bouter, p 1173.

[2] Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., et al: COSMIN guideline for systematic reviews of patient-reported outcome maesures. Quality of Life Research. 27 (5).p 1147-1157.

أداة القياس أن تكون صادقة وثابتة في آن واحد، فالعلاقة بينهما يفترض أن تكون إرتباطية عالية. [1]. فالصدق يتضمن الثبات، وهو مظهر من مظاهره، ولكن الصدق أعم وأشمل من الثبات، إذ أن ثبات الاختبار ودقته لا تدل على صدقه، بل من الممكن أن يكون الاختبار ثابتا ودقيقا، وأن درجة الفرد عليه لا تتغير كثيرا من إجراء إلى آخر، ولكنه ليس صادقا، بمعنى أن يقيس دقة عامل آخر وثباته غير العامل الذي صمم الاختبار لقياسه، ومن جهة أخرى فإن كل اختبار صادق هو ثابت بالضرورة ولكن لا يمكن القول أن كل اختبار ثابت هو صادق بالضرورة. [2] (الشجيري و الزهيري، 2022، ص 313) وذلك لأن الاختبار الثابت يدل على اتساق البنود في بينها وتجانسها وحتى على دقة الاختبار في ما يقيسه، لكنه في الوقت نفسه قد تكون تلك البنود أو الاختبار بصفة عامة يقيس سمة غير التي يراد به أن يقيسها، ولذلك نقول أن الاختبار الثابت ليس بالضرورة هو اختبار صادق، على عكس الاختبار الذي تكون بنوده صادقة، فهي مدروسة ومختارة بعناية وتمثل السمة التي يراد دراستها وبالتالي يتطلب ذلك اتساق البنود في ما بينها ودقة ما تقيسه وعليه فلا يمكن أن يكون الاختبار صادقا وهو غير ثابت.

على الرغم من أن الثبات المنخفض يحد من الصدق، إلا أن الثبات المرتفع لا يضمن الصدق. فمن الممكن دائما أن ينتج اختبار درجات ثابتة، ولكن الاستدلالات اعتمادا على درجات الاختبار يمكن أن تكون غير صادقة مطلقا. وعليه فالثبات ضروري ولكنه شرط غير كاف للصدق. [3] . وهو ما ذهب إليه أيضا بوليت Polit (2015) بقوله أنه وجب دائما التحقق من

---

[1] بن صافي، المرجع مذكور سابقا، ص 50.

[2] الشجيري و الزهيري، المرجع مذكور سابقا، ص 313

[3] ليفينجستون و رينولد، المرجع مذكور سابقا، ص 220

الصدق ومن الثبات أيضا، فقيمة الثبات المرتفعة لا تضمن أبدا الصدق، حتى لو كانت من أهم شروطه [1]

وفي الاتجاه الآخر يرى كل من هوييت وكرامر (2016) أن الثبات والصدق متمايزان مفاهميا إلى حد ما، ولا توجد حاجة لأن تكون هناك علاقة ضرورية بينهما. ولذلك، يجب الحيطة جيدا عندما يقال بأن الاختبار أو المقياس الصادق ينبغي أن يتسم بالثبات. [2] والمقصود هنا ليس نفي العلاقة تماما بين الصدق والثبات ولكن نفي التلازم التام بينهما، فمثلا هناك سمات نفسية غير مستقرة نسبيا بطبيعتها، واذا تم اعتماد مقياس يقيسها فعلا ويقيسها هي وفقط أي أن المقياس صادق حقا، وبذلك نفترض حسب مبدأ الصدق يوجب الثبات أن المقياس ثابت أيضا، ولكن الحقيقة أن المقياس رغم كونه صادق إلا أنه يحتمل ويفترض منطقيا أن يكون غير ثابت، خاصة إذا استخدم في التحقق منه طريقة إعادة التحقق، لأن السمة في حد ذاتها غير ثابتة وإذا كان كذلك وكانت نسبة الثبات مرتفعة فهذا يدفعنا للشك في صدقه أيضا، وهو الأمر الذي يدفعنا إلى التحقق من الثبات وعدم الاكتفاء بالصدق، بالاضافة لمعرفة نسبة أو درجة ثبات الاختبار.

---

[1] Polit, D.: Assessing measurement in health: Beyond reliability and validity,International. *Journal of Nursing Studies .52*. 2015, p 1746-1753.

[2] هويت وكرامر، المرجع مذكور سابقا، ص 400.

خلاصة

دائما ما تهدف البحوث والدراسات إلى تحقيق نتائج دقيقة وموثوق فيها، ويتم من أجل ذلك الاستعانة بجمع البيانات باستخدام أدوات تتمتع بصدق وثبات كافيين، وذلك على اعتبار أن جودة النتائج تخضع إلى الشروط السيكومترية للأدوات. أي جمع أدلة كافية وموثوقة حول صدقها وثباتها، ورغم أنّ الصدق يعتبر أكثر أهمية، إلا أن ذلك لا ينقص من قيمة الثبات، فهو دليل على درجة الثقة في النتائج، وقد ناقشت الباحثة في هذا الفصل الجوانب الرئيسية في تقدير الثبات، انطلاقا من التعريف به، معرجة على أنواعه وإجراءاته والعوامل المؤثرة فيه وكذلك المعايير التي استندت عليها الباحثة في تقييم عينة الدراسة الحالية، وصولا إلى العلاقة بينه وبين الصدق. في محاولة منها إلى تسليط الضوء على أهمية الثبات من خلال التطرق إلى أغلبية تفاصيله المهمة.

الجانب التطبيقي

الفصل الخامس

الإطار المنهجي للدراسة

1-منهج الدراسة

اعتمدت الدراسة الحالية على المنهج الوصفي والذي يسعى إلى تحديد الوضع الحالي لظاهرة معينة ومن ثم وصفها، فهو يعتمد على دراسة الظاهرة كما توجد في الواقع ويهتم بوصفه. وقد وقع الاختيار على هذا النوع من المناهج باعتباره الأكثر ملاءمة لأغراض وأهداف الدراسة، حيث يقوم على جمع البيانات وتحليلها وتفسيرها للوصول إلى حل مناسب للمشكلة، وهو بالضبط ما تهدف إليه الدراسة الحالية من محاولة جمع البيانات وتحليلها من أجل الوصول إلى حل للأخطاء التي يقع فيها الباحثون عند إعادة التحقق من صدق وثبات الاختبارات النفسية قبل تطبيقها.

2-عينة الدراسة

بما أنه يستحيل تقييم جميع مفردات المجتمع، المتمثلة في الدراسات الأكاديمية الموجودة سواء في البوابات المذكورة سابقا أو غيرها من البوابات والمكتبات لكثرتها، فإن الباحثة اعتمدت على سحب العينة بالطريقة القصدية فاختارت الدراسات التي تم الاعتماد فيها على إعادة التحقق من صدق وثبات اختبار جاهز أي لم يتم بناءه في تلك الدراسة، وكما اتضح من الإجراءات السابقة أصبح العدد النهائي للدراسات وبالضبط للاختبارات المقبولة لتقييمها 171 اختبار، وللتعريف بعينة هذه الدراسة وجب ذكر خصائصها التالية:

2-1-إستراتيجية جمع العينة

أجري البحث عن عينة الدراسة من خلال قواعد البيانات الإلكترونية، وتم إجراء عملية البحث وجمع الدراسات بين الفترة 1 أكتوبر 2020 إلى غاية 4 فيفري 2022، وقد شملت العينة عينة من الدراسات العربية والأجنبية. بالنسبة للدراسات العربية تم اعتماد بوابة **AskZad** لجمع الاطروحات وقاعدة البيانات **المعرفة** لجمع **الأطروحات** والمقالات، وأما في ما يخص الدراسات الأجنبية فقد تم الاعتماد على قاعدة **Science Direct** للمقالات وبوابة **OATD** للأطروحات، وذلك باستخدام

مزيج من عناوين الموضوعات التي تستخدم فيها عادة الاختبارات النفسية ومن هذه العناوين مثلا: (الوسواس القهري، الاكتئاب، الاحتراق النفسي، القلق، الأفكار اللاعقلانية، الذكاء.....) وذلك فقط بالنسبة لبوابة **OATD** لأن باقي البوابات قد تم تضمين والبحث في جميع الدراسات الموجودة على البوابات، بحيث أن البوابات الثلاثة الأخرى تسمح بتحديد ميادين علم النفس وعلوم التربية وغيرها، وكذلك إمكانية تحديد المجال الزمني، وبعد ذلك تم فرز عناوين المقالات وملخصاتها والتحقق منها. واعتماد الدراسات التي تستوفي الشروط فقط، أما التي لا تستوفي معايير الاشتمال فقد استبعدت استناداً إلى عناوينها أو ملخصاتها أو حتى مضمونها.

2-1-1-معايير القبول والاستبعاد

وضعت الباحثة مجموعة من المعايير التي من خلالها يتم اعتماد الدراسة وإدراجها في العينة أو عدم اعتمادها من الدراسة الحالية، وهذه المعايير كالآتي:

-يجب أن تعتمد الدراسة على اختبار نفسي.

-أن يكون هذا الاختبار جاهز أي تم بناءه في دراسة أخرى غير الدراسة التي يتم تقييمه فيها.

-يجب أن يتم إعادة التحقق من صلاحية هذا الاختبار (صدقه وثباته) في الدراسة.

-أن تقع الدراسة ضمن المجال الزمني الذي حددته الباحثة بين 2012-2022.

-استبعاد الدراسات التي اعتمدت على بناء أو تكييف أو تقنين تلك الاختبارات.

-استبعاد الدراسات التي ذكرت فقط أنها قامت بإعادة التحقق من صلاحية الاختبارات النفسية دون التطرق لأية تفاصيل عن كيفية القيام بذلك.

-استبعاد الدراسات التي ذكرت أنواع الصدق والثبات التي تم استخدامها لكن بدون تفاصيل يمكن تقييمها من خلالها.

-استبعاد الدراسات التي تحققت من صلاحية الاختبار فقط باستخدام الصدق الظاهري، أي أن الباحث صاحب الدراسة تحقق بنفسه من صلاحية الاختبار عن طريق قراءة البنود فقط.

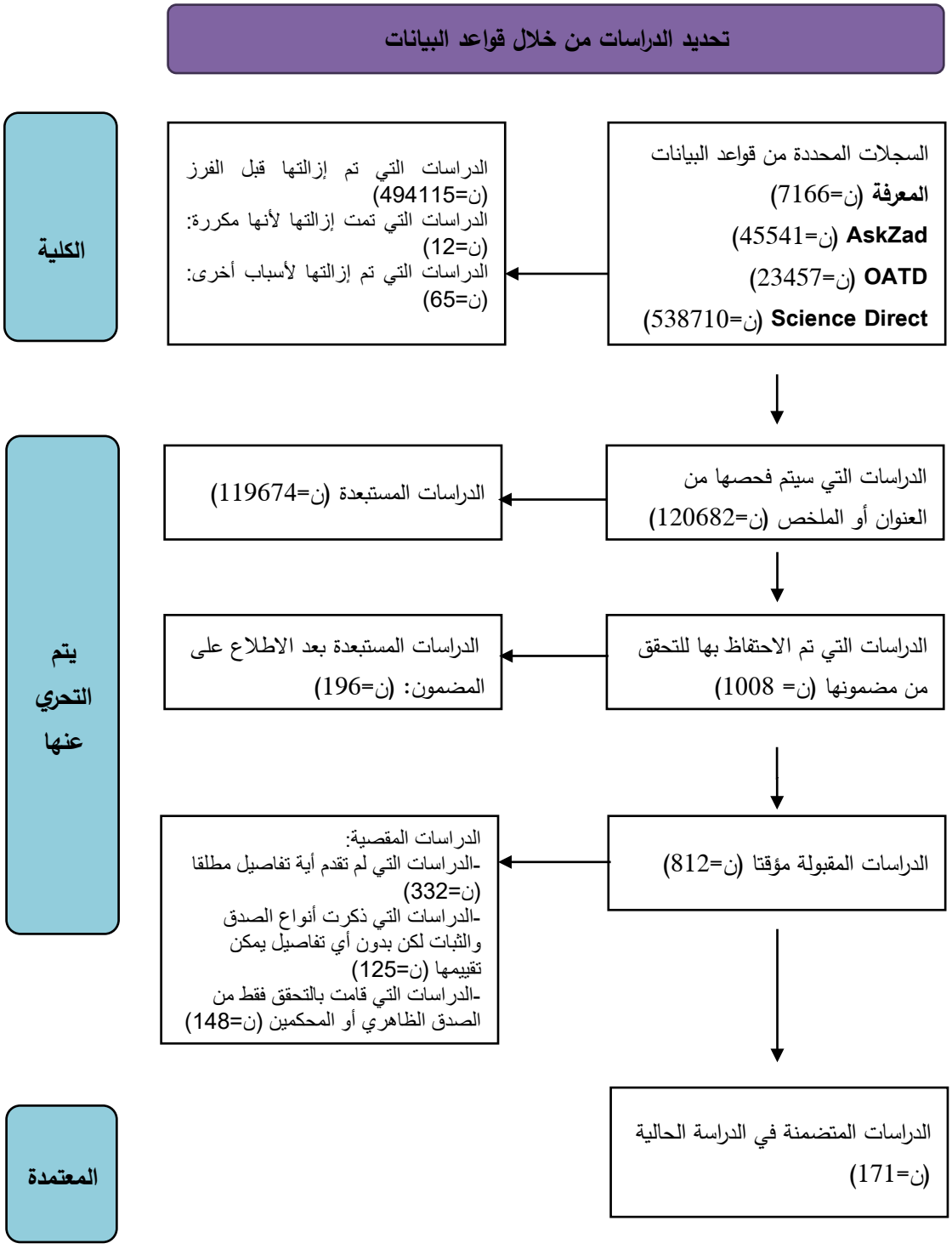
-استبعاد الدراسات التي اعتمدت فقط على صدق المحكمين وفي نفس الوقت لم تذكر أي تفاصيل عن عددهم أو من هم هؤلاء المحكمين أو كيف تم تحكيم الاختبار وغيرها من التفاصيل المهمة.

2-1-2-عملية فرز الدراسات

لقد قامت الباحثة بإتباع منهجية **PRISMA** (McKenzie, Bossuyt, Boutron, Hoffmann, & Mulrow, 2020)-وهي اختصار لـ(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) والتي تعني (تجهيز المفردات المناسبة لإعداد المراجعات المنهجية والتحليلات الوصفية)- في خطواتها (انظر الملحق 04)؛ ليس بحذافيرها ولكن بفكرتها الأصلية وذلك لأن أغلب المنهجيات الدولية لاختيار الأبحاث، في المراجعات المنهجية (الدراسات التقييميه) مثل PRISMA أو Cochrane، تعتمد فقط على قواعد البيانات الدولية مثل: PsychINFO, CINAHL, PubMid Science Direct, EMBASE وغيرها، والتي بدورها تعتمد على اللغة الانجليزية فقط، مما يقصي الدراسات العربية من عملية اختيار عينة الدراسة، وبالتالي اعتمدت الباحثة كما ذكر سابقا على منهجية PRISMA في طريقة اختيار العينة متجاهلة ضرورة اختيارها فقط من قواعد البيانات السابقة الذكر-للإشارة فإن الغرض من تحديد PRISMA لقواعد البيانات السابقة الذكر هو أنها تحتوي فقط على البحوث ذات الجودة العالية- وقد أضافت الباحثة قاعدتين عربيتين وهما المعرفة وAskZad. كذلك حددت الباحثة المجال الزمني في البحث عن العينة بين سنة 2012 و 2022 من أجل حصر عملية البحث

وحصر الكم الهائل من الدراسات سواء العربية أو الأجنبية، الذي يحتاج إلى الكثير من الوقت، كما سيتضح ذلك من خلال الأعداد الهائلة الموضحة في الشكل التالي:

الشكل 03: طريقة فرز الدراسات حسب منهجية PRISMA



المصدر: من إعداد الباحثة بالاعتماد على نموذج PRISMA

إن إجمالي الدراسات في قواعد البيانات الأربع التي اعتمدتها الباحثة والمذكورة سابقا هو 614874 دراسة، وقد تم استبعاد 494115 دراسة منها وذلك لأنها صدرت قبل التاريخ الزمني الذي حددته الباحثة أي قبل سنة 2012، كذلك قد تم استبعاد 12 دراسة لأنها مكررة في كل من البوابتين المعرفة وAskZad، وأيضا تم إزالة 65 دراسة لأن الوصول إليها كان غير مفتوح وبالتالي لا يمكن رؤية وقراءة المقال كاملا إنما فقط الملخص لذلك قامت الباحثة باستبعاد تلك المقالات وكان هذا في بوابة Science Direct، ليبقى 120682 دراسة سيتم فحصها من خلال العنوان والملخص، وتم في هذه المرحلة استبعاد 119674 دراسة فقط من خلال عناوينها أو ملخصاتها وفقا لشروط الاستبعاد التي وضعت سابقا فالدراسات في هذه المرحلة إما أنها لم تعتمد على الاختبارات النفسية أو أنها اعتمدت على بناء الاختبار أو تكييفه أو تقنينه وغيرها من المعايير التي ذكرتها الباحثة سابقا، ليتبقى للباحثة 1008 دراسة وجب عليها التحقق من مضمونها بعدما قبلت مبدئيا من خلال العناوين والملخصات، والتي استبعدت منها 196 دراسة لعدم وجود التفاصيل التي تم ذكرها في الملخصات، وبالتالي أصبح عدد الدراسات 812 دراسة مقبولة مؤقتا، ولاختيار الدراسات التي ستصبح مقبولة بصفة نهائية تم فحص مضمون 812 دراسة تلك، لتقصي الباحثة منها 322 دراسة التي لم تقدم أية تفاصيل مطلقا فقط اكتفت بذكر أنها أعادت التحقق من صلاحية الاختبار النفسي، و125 دراسة التي ذكرت أنواع الصدق والثبات المستخدمة في إعادة التحقق ولكنها بدون تفاصيل تمكن الباحثة من تقييم تلك الطرق، كذلك تم إقصاء 148 دراسة قامت بالتحقق فقط من الصدق الظاهري أو صدق المحكمين حسب معيار الاستبعاد المذكور سابقا. ليبقى في الأخير 171 دراسة تم تضمينها في عينة الدراسة الحالية، أي تم تقييم طرق إعادة التحقق من صلاحية الاختبارات النفسية فيها.

2-2-خصائص عينة الدراسة

إن الغرض من عرض خصائص عينة الدراسة هنا هو محاولة وصف أهم مميزات هذا الدراسات.

2-2-1- توزيع العينة حسب متغير الدراسة:

في هذا العنصر سيتم عرض عدد كل من رسائل الدكتوراه والماجستير وكذلك المقالات التي أدرجت في عينة الدراسة من خلال الجدول والشكل التاليين:

الجدول 01: توزيع العينة حسب متغير الدراسة

	التكرارات	النسب المئوية
دكتوراه	40	%23.4
ماجستير	55	%32.2
مقال	76	%44.4
الكلية	171	%100

المصدر: من إعداد الباحثة

كما يتضح من الجدول أعلاه فإن عدد المقالات المدرجة ضمن العينة كان هو الأكثر تكرارا ب76 مقالا ما نسبته 44.4%، وبالمقابل نجد 55 رسالة ماجستير بنسبة 32.2% و40 أطروحة دكتوراه بنسبة 23.4%، وتعتبر هذه الأعداد متقاربة إلى حد ما. وكانت هذه النتائج عكس توقعات الباحثة فقد توقعت الباحثة أثناء عملية فرز وجمع العينة أن عدد الأطروحات سيكون أكثر، وذلك لأن غالبا في المقالات المنشورة حتى وإن تم إعادة التحقق من صلاحية الاختبار، فإن الباحثين يتغاضون عن ذكر الإجراءات المستخدمة في ذلك، والدليل أن الباحثة استبعدت الكثير

من المقالات التي ذكرت أنها أعادت التحقق من صلاحية الاختبار النفسي لكن دون ذكر أي تفصيل يسمح بتقييم طرق إعادة التحقق.

2-2-2- توزيع العينة حسب نوع الدراسات:

تم سحب عينة الدراسة من البيئة العربية والأجنبية معا، وذلك بغرض توسيع دائرة التقييم، وقد كان عدد الدراسات العربية والأجنبية متقارب إلى حد كبير كما هو موضح في الجدول أسفله:

الجدول 02: نوع الدراسات المدمجة في العينة

	التكرارات	النسب المئوية
الدراسات العربية	87	%50.9
الدراسات الأجنبية	84	%49.1
الكلية	171	%100

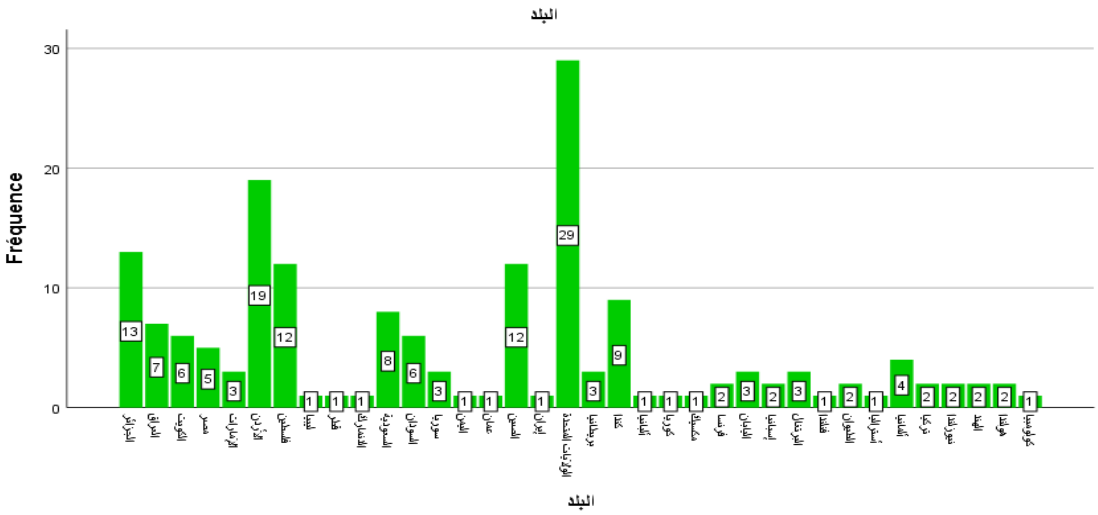
المصدر: من إعداد الباحثة

كما هو متوقع فإن عدد الدراسات العربية والأجنبية كان جد متقارب، وذلك دون قصد من الباحثة، وقد قدر عدد الدراسات العربية بـ87 دراسة ما نسبته 50.9%، أما عدد الدراسات الأجنبية فقد كان 84 دراسة بنسبة 49.1%. وللإشارة فإن الباحثة لم تقم بدراسة مقارنة بين الدراسات العربية والأجنبية لأن ذلك لم يكن من أهداف الباحثة أولا، وثانيا لأن بيانات البيئتين ليست منفصلة في هذه الدراسة، لكنها ستقوم بفصل البيانات والعمل مقارنة نتائج الجودة لكل دراسة على حدى والمقارنة بينها في بحث آخر.

2-2-3-توزيع عينة الدراسة حسب البلد:

سيتم هنا عرض مختلف البلدان التي تنتمي إليها الدراسات، التي كانت ضمن عينة الدراسة الحالية واختارت الباحثة عرض التوزيع على شكل رسم بياني لأنه مختصر وواضح أكثر من الجدول:

الشكل 04: توزيع العينة حسب البلد الذي تنتمي إليه الدراسة



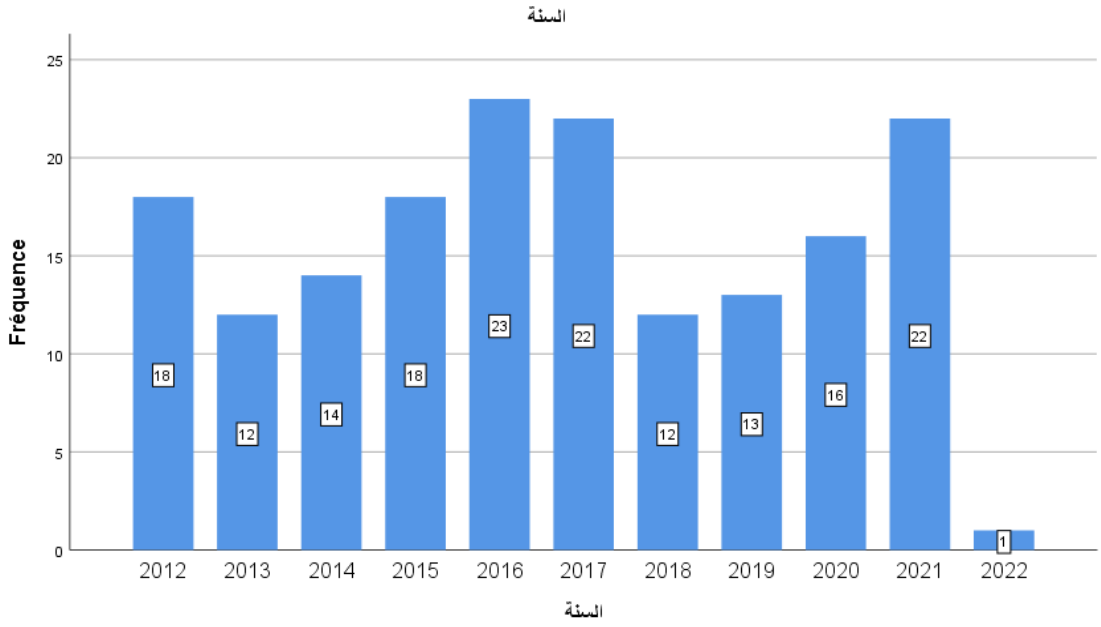
المصدر: من إعداد الباحثة

يتضح من الشكل السابق أن العينة توزعت على 36 بلد مختلف، وكانت أعلى نسبة من الدراسات بتكرار 29 دراسة من الولايات الأميركية المتحدة، وتليها 19 دراسة من الأردن لتأتي بعدها 13 دراسة من الجزائر، تليها 12 دراسة للصين وكذلك نفس العدد لفلسطين، ثم 9 دراسات لكندا، لتتوزع باقي الدراسات بتكرارات متراوحة بين تكرار واحد و8 تكرارات على 30 بلد مختلف (كولومبيا، هولندا، الهند، نيوزلندا، تركيا، ألمانيا، أستراليا، الطيوان، فنلندا، البرتغال، إسبانيا، اليابان، فرنسا، المكسيك، كوريا، ألبانيا، بريطانيا، إيران، عمان، اليمن، سوريا، السودان، السعودية، الدنمارك، قطر، ليبيا، الإمارات، مصر، الكويت، العراق).

2-2-4-توزيع العينة حسب سنة المناقشة أو النشر:

أيضا فضلت الباحثة هنا عرض السنوات على شكل الرسم البياني التالي:

الشكل 05: توزيع العينة حسب سنة المناقشة أو النشر



المصدر: من إعداد الباحثة

كما تم ذكره سابقا فقد حددت الباحثة الفترة الزمنية بين 2012 و2022، من أجل حصر مجتمع البحث، وبالتالي يتضح من الرسم البياني أن السنوات 2016 و2017 و2021 تقاربت نسبة الدراسات المسحوبة منها بتكرار 23 دراسة لسنة 2016 و22 دراسة لكل من 2017 و2021، أما باقي السنوات فتراوحت التكرارات فيها بين 12 و18 دراسة لكل سنة، ماعدا سنة 2022 فقد أخذت منها دراسة واحدة فقط وذلك راجع لكون الباحثة توقفت نهائيا عن عملية جمع وفرز العينة في الشهر الثاني من سنة 2022، وبالتالي قلت نسبة عدد الدراسات المسحوبة في هذه السنة.

2-2-5-توزيع العينة حسب النسخة المستخدمة من الاختبار:

يتم هنا عرض النسخة التي استخدمها الباحثين من الاختبار في دراساتهم وقد لخصتها الباحثة في كل من الجدول والشكلين التاليين:

الجدول 03: توزيع العينة حسب النسخة المستخدمة من الاختبار

	التكرارات	النسب المئوية
المترجمة	64	%37.4
الأصلية	98	%57.3
المكيفة	3	%1.8
المستحدثة	6	%3.5
الكلية	171	%100

المصدر: من إعداد الباحثة

يتبين لنا من خلال الجدول والشكل السابقين أن أكثر من نصف الدراسات اعتمدت على النسخ الأصلية للاختبارات النفسية بنسبة 57% أي بـ98 تكرار من أصل 171 دراسة، في حين أن 64 دراسة بنسبة 37.4% فضلت استخدام النسخ المترجمة للاختبار، وللإشارة فإن ضمن 64 دراسة تلك توجد 13 دراسة فضل أصحابها ترجمة الاختبار بأنفسهم (انظر الملحق 05)، كذلك جميع هذه الدراسات 64 اختبارها الأصلي باللغة الانجليزية (انظر الملحق 05)، أما الاختبارات المكيفة المستخدمة في الدراسات فقد قدر عددها بـ3 دراسات فقط بنسبة 1.8%، ليبقى في الأخير 6 دراسات استخدمت النسخ الحديثة من الاختبارات النفسية، أي أن صاحب الاختبار هو من قام بتحسين وتعديل النسخة الأولى ليخرج بنسخة حديثة محسنة وهي التي تم استخدامها في الدراسات

بنسبة 3.5%. وللعلم فإن الباحثة صنفت النسخ إلى هذا التصنيف وفق ما أشار إليه الباحثين في تعريفهم للاختبار المستخدم. كذلك تعيب الباحثة ترجمة الاختبار دون تكييفه أو حتى تقنينه وهو ما ستتحدث عنه بشيء من التفصيل في مناقشة النتائج.

2-2-6-توزيع العينة حسب تعديل الاختبار:

لقد فضلت الباحثة عرض هذا الجزء المتعلق بتعديل الاختبار هنا بدل عرضه في الجزء المتعلق بعرض نتائج الدراسة، لأنه لا يدخل ضمن الإجابة عن تساؤلات الدراسة، وكذلك لا يدخل ضمن هدف الدراسة والذي يتمحور حول تقييم طرق إعادة التحقق من صدق وثبات الاختبارات النفسية، ولذلك فإنه قبل تقييم تلك الطرق هنالك بعض الخصائص المتعلقة بالاختبار وجب إدراكها فمثلا الاختبار الذي تم تعديله حسب مبادرة أو دراسة COSMIN، يعتبر كأنه نسخة جديدة أو حتى حديثة من الاختبار، ولذلك فعند إعادة التحقق من صدقه وثباته وجب أن يراعى في ذلك هذا الأمر. ولذلك قررت الباحثة عرض الاختبارات التي تم تعديلها، وأيضا عرض الاختبارات التي تم إعادة التحقق من صدقها وثباتها بعد التعديل وكانت النتائج كما يلي:

الجدول 04: توزيع الاختبارات حسب التعديل

الاختبارات	التكرارات
لم يتم تعديلها	111
تم تعديلها	60
تم إعادة التحقق منها بعد التعديل	36
الكلية	171

المصدر: من إعداد الباحثة بالاعتماد على مخرجات spss

من خلال الجدول الظاهر أمامنا يمكن استنتاج أهم نقطة موجودة فيه، وهو أن الاختبارات التي تم تعديلها ولكن لم يتم تقييمها بعد التعديل قدرت بـ36 اختبار، وهو عدد كبير مقارنة بأنه يشكل أكثر من نصف عدد الاختبارات التي تم تعديلها والتي قدرت بـ60 اختبار، ويشمل هذا التعديل ما ذكر في الجانب النظري في الفصل المتعلق بصلاحية الاختبارات، بالضبط في جزئية الدوافع وراء إعادة التحقق من صلاحية الاختبارات النفسية، فقد ذكرت الباحثة استنادا إلى أراء المختصين والخبراء في هذا؛ بأن تغيير شكل الاختبار مثل ترجمته، تعديل صياغة العديد من بنوده، أو تغيير بنية الاختبار كحذف أحد أبعاده، أو حذف العديد من بنوده وغيرها من أشكال التغيير، فمثلا فقط من خلال جدول المواصفات (الملحق 05) نلاحظ أن 25 اختبار تم تغيير بنيته من خلال حذف أو إضافة عدد البنود. وبالتالي فإن عدم إعادة التحقق من صدق وثبات 36 اختبارا بعد التعديل يعتبرا خطأ منهجيا. لأنه وكما ذكرنا سابقا فإن الاختبار المعدل هو عبارة عن نسخة جديدة للاختبار، وبالتالي وجب التحقق من خصائصه السيكومترية (الصدق، الثبات والتحيز) قبل تطبيقه. وفي الأخير فإن عدد الاختبارات التي تم تطبيقها في شكلها الأصلي فقد قدر بـ111 اختبار لم يتم تعديله.

لقد لفت انتباه الباحثة أثناء محاولتها الإجابة على تساؤلات هذه الدراسة؛ وبالضبط أثناء ملئ جدول المواصفات (الملحق 05)، العديد من الملاحظات التي لم تكن في الأصل من أهداف الدراسة؛ ولكن لا يمكن المرور دون الإشارة إليها، ومن بين هذه الملاحظات ما يلي:

- أغلبية المحكمين هم من أساتذة الجامعة.
- الدراسات الأجنبية تقريبا كلها اعتمدت على العينة الأصلية للدراسة في التحقق من صلاحية الاختبارات النفسية، في المقابل اعتمدت الدراسات العربية على العينة

الاستطلاعية، لذلك عند النظر إلى الملحق رقم (05) تجد أن حجم العينة في الدراسات الأجنبية كبير مقارنة بالدراسات العربية.

- ليكرت الخماسي هو الخيار الأكثر استخداما في الاختبارات النفسية.
- الطلبة والتلاميذ هم المجتمع الأكثر استخداما في البحوث والدراسات النفسية.

3-أداة الدراسة

اعتمدت الباحثة في تقييمها لجودة الخصائص السيكومترية المستخدمة في إعادة التحقق من صلاحية الاختبارات النفسية، على شبكة أو قائمة مراجعة COSMIN (Mokkink, Vet, Prinsen, Patrick, Alonso, & Bouter, 2018)[1] وهي أداة تم تطويرها من قبل فريق من الخبراء باستخدام دراسة دلفي الدولية وسيتم في النقاط التالية تقديم لمحة عن الأداة ومن طورها وكيفية العمل بها ومعاييرها وغيرها.

3-1-نبذة عن مبادرة COSMIN

عمل فريق مبادرة COSMIN والتي هي اختصار لـ COnsensus-based Standards for the selection of health Measurement Instruments والتي تعني المعايير المستندة إلى التعداد لاختيار أدوات قياس الصحة، والتي تأسست سنة 2005، لتطوير قائمة مراجعة تحتوي على معايير لتقييم الجودة المنهجية لخصائص القياس في الدراسات المرتبطة بالصحة الجسدية والنفسية، وتمكن الفريق سنة 2010 من إطلاق النسخة الأولى لها، أما النسخة الأخيرة المعتمدة في هذه الدراسة فقد تم تحديثها سنة 2018، وكان السبب الرئيسي وراء هذه الدراسة هو عدم وجود اتفاق واضح حول تعريفات الخصائص السيكومترية (مثل الصدق والثبات)

[1] Mokkink, Vet, Prinsen, Patrick, Alonso, & Bouter, p 1179.

من قبل، بل أبعد من ذلك وجود عدة تسميات لمصطلح واحد، وقد يشترك نوعين مختلفين من الخصائص في تسمية واحدة (مثل الاتساق الداخلي)، ومن أهم الأسباب أيضا عدم وجود معايير موحدة للحكم على جودة الاختبار من خلال خصائصه السيكومترية، ما دفع بمجموعة من الباحثين المؤسسين لهذا المبادرة والذي كان عددهم 8 باحثين من مجالات مختلفة –عمل هؤلاء الباحثين كلجنة توجيهية ولم يتم إشراكهم ضمن دراسة دلفي– إلى العمل على توحيد مفاهيم الخصائص السيكومترية وتوضيحها ووضع معايير واضحة من حيث الطرق الإحصائية المستخدمة فيها، وقد تم ذلك من خلال دراسة دلفي الدولية التي كانت على شكل 4 جولات تم فيها استدعاء 251 خبير في الجولة الأولى، ليصل عدد الخبراء في الجولة الأخيرة إلى 91 خبير من مختلف التخصصات (القياس النفسي، علم النفس، الإحصاء، الطب السريري، علم الأوبئة وغيرها) وقد تم اختيارهم وفق شروط معينة مثل أن يمتلك الخبير ما لا يقل عن 5 مقالات منشورة في PubMed، وينتمي إلى أحد المنظمات الدولية للأبحاث الصحية وغيرها، وقد تم تحديد نسبة 67% اتفاق بين الخبراء كمعيار لقبول التعريف أو تحديد المفهوم أو غيرها، بحيث طلب من أعضاء الفريق تقييم اتفاقهم أو عدمه مع كل اقتراح على مقياس من 5 نقاط (من موافق بشدة إلى غير موافق بشدة). ويمكن الاعتماد على المرجع التالي (Mokkink, Vet, Prinsen, Patrick, Alonso, & Bouter, 2018) للإطلاع على تفاصيل أكثر حول دراسة دلفي.

3-2- قائمة المعايير الأصلية COSMIN

تم تحديث الأداة آخر مرة سنة 2018 وتم تغيير اسمها من قائمة التحقق COSMIN Checklist، إلى قائمة COSMIN لمراجعة مخاطر التحيز The COSMIN Risk of Bias checklist، وربما يتساءل القارئ هنا عن العلاقة بين مخاطر التحيز وتقييم الخصائص السيكومترية، لذلك تريد أن تنبه الباحثة إلى أن المبادرة قد أشارت في موقعها الإلكتروني

https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/ وفي الأدلة المنشورة في المجلات العلمية. إلى تبنيها نفس تعريف التحيز الذي أطلقته مبادرة كوكرين Cochrane- وهي مبادرة تحمل نفس أهداف مبادرة COSMIN وأقدم منها ولكن فقط في تخصص الطب- فالتحيز يشير إلى ما إذا كان يمكنك الوثوق بالنتائج بناء على الجودة المنهجية (صدق، ثبات، استجابة) للدراسة.

وتتكون القائمة الحالية من عشرة مجالات كل منها يقيم جانب معين من خاصية قياس مختلفة، وذلك عن طريق مجموعة من البنود مجتمعة في مربعات مختلفة: فالمربع الأول يقيم منهجية بناء وتطوير الاختبار، أما المربع الثاني فيقيم صدق المحتوى، في حين أن المربع الثالث يقيم صدق البناء، أما المربع الرابع فيحتوي على بنود تقيم ثبات الاتساق الداخلي، والمربع الخامس يقيم الصدق عبر الثقافات، أما المربع السادس فيقيم ثبات الإعادة والتكافؤ، في حين يقيم المربع السابع خطأ القياس، والمربع الثامن الصدق المحكي، ليبقى المربع التاسع لتقييم صدق الفرضيات، وأخيرا المربع العاشر لتقييم الاستجابة. وتجتمع هذه المجالات العشر تحت ثلاث مجموعات رئيسية: الصدق، الثبات والاستجابة. كما هو واضح من خلال النموذج الذي قدمته مبادرة COSMIN الموجود في الملحق (06)

عند استخدام قائمة مراجعة COSMIN، يتم إكمال المربعات المتعلقة بخاصية القياس التي تم استخدامها فقط (أي ليس من الضروري إكمال كل مربع لكل اختبار)؛ وبالتالي سيختلف عدد تقييمات الجودة المتاحة لكل اختبار، يحتوي كل مجال على ما بين 3 و35 عنصرًا ، بإجمالي 116 عنصرًا. ويتم تصنيف العناصر الموجودة في كل مربع على أنها "جيدة جدًا" أو "مناسبة" أو

"مشكوك فيها" أو "غير كافية" ويتم استخدام أدنى تصنيف لأي مربع لتحديد التصنيف الإجمالي لكل خاصية قياس.

وفي الأخير يتم تصنيف جودة صدق المحتوى في ص جديدة من حيث الملائمة (خمسة معايير)، والشمولية (معيار واحد)، والفهم (أربعة معايير). ويتم تصنيف جودة خصائص القياس الأخرى في ص أخرى على أنها "كافية (+)" أو "غير كافية (−)" أو "غير محددة (؟)" بناءً على معايير خصائص القياس الجيدة. وفي ص أخرى يتم تصنيف جودة الأدلة لكل خاصية قياس على أنها "مرتفع" أو "متوسط" أو "منخفض" أو "منخفض جدًا" باستخدام نسخة معدلة من نهج تصنيف التوصيات وتطويرها وتقييمها(GRADE) . ويمكن الاطلاع على النسخة الأصلية للقائمة في الملحق (02)

3-3-قائمة المعايير COSMIN المستخدمة في الدراسة الحالية

اعتمدت الباحثة على قائمة المعايير COSMIN، وقامت كأول خطوة بترجمتها من أجل عرضها في الدراسة الحالية، في حين مازالت تعمل على ترجمة الدليل المرافق الذي يساعد الباحثين على فهم كيفية عمل القائمة واستخدام المعايير الموجودة في الدليل للحكم على كل بند فيها، وسيتم نشر العمل كاملا بعد إتمامه-بمساعدة فريق خاص لهذا العمل والتقيد بالشروط اللازمة لكل ترجمة رسمية- في موقع COSMIN الرسمي (**https://www.cosmin.nl/**) مع إجراء بعض التعديلات التي تتطلبها الدراسة الحالية، وتتمثل التعديلات في النقاط التالية:

- حذف المربع الأول الذي يتم فيه تقييم طريقة بناء الاختبار. (**Prom Design**) لأن هذه الدراسة تركز على تقييم دراسات التي قامت بإعادة التحقق من الاختبار وليس البناء أو التطوير.

وبالتالي انطلقت الباحثة من المربع الثاني الذي يقيم صدق المحتوى. وهو ليس بالأمر المستحدث فقد قامت دراسة (Haile, et al., 2022) بنفس الأمر ولنفس السبب.

-احتفظت الباحثة بالمربع الثاني دون أي حذف أو تعديل أو إضافة.

-احتفظت الباحثة بالمربع الثالث كما هو مع تعديل في معيار الحكم، فقد أضافت الباحثة الاتساق الداخلي (العلاقة بين البنود والمحور، أو العلاقة بين المحاور والدرجة الكلية)، أي في حالة قام الباحث بالتحقق منه سيتم الحكم على جودته بأنه مناسب، ففي قائمة COSMIN الأصلية يتم الحكم على الاختبار بأن جودته جيدة إذا تم التحقق منه بالتحليل العاملي التوكيدي، وأنه مناسب إذ تم استخدام التحليل العاملي الاستكشافي، ولم يتم أبدا التطرق إلى الاتساق الداخلي، وللأسف يعتبر من أكثر الطرق استخداما في الدراسات العربية **وحتى الأجنبية** لذلك وبعد تشاور الباحثة مع المشرف قررت إدراجه وتقييمه بأنه مناسب، مع وضع الضوابط الخاصة به مثل الا يقل حجم العينة عن 50 فردا وقيمته لا تقل عن 0.70 وبالطبع دال. وقد تم استنباط هذين الشرطين من دراسة (Terwee, et al., 2007).

-احتفظت الباحثة بالمربع الرابع كما هو مع إضافة التجزئة النصفية واعتمدت على استنباط المعايير المتعلقة بمعامل الارتباط من قائمة COSMIN، أما في وضع المعايير المتعلقة بالمعادلات التصحيحية وكيفية التقسيم فقد اعتمدت على المعايير المذكورة في الجانب النظري بالضبط الجزء المتعلق بالتجزئة النصفية في فصل الثبات.

-احتفظت الباحثة بباقي المربعات الأخرى كما هي دون أي حذف أو تعديل أو إضافة.

-تم حذف الص المتعلقة بجودة صدق المحتوى لأنها تتطلب تقييمات مراجعين اثنين لكل اختبار وأيضا تجميع صلاحيات المحتوى لنفس الاختبار من دراسات أخرى وغيرها من المتطلبات، التي

صعب على الباحثة الالتزام بها لخصوصية هذه الدراسة وأهدافها. والإبقاء على الجدول 2 و3 اللذان يهدفان إلى تقييم الجودة لكل المربعات العشرة.

-تم حذف الص المتعلقة بجودة المربعات الباقية لأنها مكررة في الجدول رقم 3.

-تم حذف الجدول 1 لأنه مرتبط بجودة المربع الأول المتعلق ببناء وتصميم الاختبار.

-تم حذف معيار الحد الأدنى الذي تعتمده القائمة في الحكم على المربع ككل، أي أن قائمة معايير **COSMIN** تستند على أدنى تقييم للبنود في الحكم على جودة المربع الذي يدرس خاصية معينة، ومن ثم الحكم على جودة الاختبار ككل من خلال جمع تقييمات المربعات، ولأن هدف الباحثة هو تقييم الخصائص السيكومترية للاختبار، وليس الحكم أو تقييم جودة الاختبار الشاملة. فلم تستند إلى هذا الحكم واكتفت بعرض تقييمات الخصائص السيكومترية فقط أي عرض تقييمات المربعات دون الاختبار ككل. وهو ما أيدته دراسة (Jewell, et al., 2019) بحيث انتقد استخدام قاعدة "الدرجة الأدنى" لأن عيبًا واحدًا فقط في الدراسة سيؤدي إلى تقييم سلبي، كذلك سلط الضوء على الطبيعة التعسفية للدرجات الفاصلة التي تحدد ما إذا كانت خاصية القياس "مناسبة" أو "غير مناسبة"، لأنه في بعض الحالات كانت القيم الإحصائية التي تشير إلى التصنيف السلبي قريبة جدًا من القيم التي تشير إلى التصنيف الإيجابي.

وفي الأخير يبدوا أن قائمة معايير COSMIN الحالية لعام 2018 هي الأكثر صرامة وتعقيدًا للتطبيق بسبب أن العملية متعددة الخطوات. يمكن الاطلاع على النسخة المعربة المستخدمة في هذه الدراسة من خلال الملحق (01).

4-الأساليب الإحصائية المستخدمة في الدراسة

بعد تجميع عينة الدراسة، من أجل تقييم طرق إعادة التحقق من صدق وثبات الاختبارات النفسية، باستخدام أداة الدراسة المتمثلة في قائمة أو شبكة COSMIN، وللإجابة عن أسئلة الدراسة تم استخدام برنامج الحزم الإحصائية للعلوم الاجتماعية (SPSS V 26) من أجل تسهيل عملية تحليل البيانات وتنظيم المخرجات، وقد تم ذلك باستخدام:

-النسب المئوية.

-التكرارات.

-الجداول المتقاطعة.

خلاصة

حاولت الباحثة في هذا الفصل عرض المنهج أو الطريق الذي سلكته للوصول إلى نتائج الدراسة التي سيتم عرضها ومناقشتها في الفصل القادم، بعد أن ركزت في هذا الفصل على شرح الطريقة التي استخدمتها وما الذي جعلها ملائما لهذه الدراسة، وأيضا مجتمع البحث الذي سحبت منه العينة مع التفصيل في كيفية استخراج العينة باستخدام منهجية PRIMSA ووضع بعض القيود التي فرضتها الباحثة انطلاقا من أهداف الدراسة، وأيضا تطرقت الباحثة إلى التعريف بأداة البحث المستخدمة المتمثلة في قائمة COSMIN مع عرض التعديلات التي أدخلتها على الأداة، لتختم في الأخير بالأساليب الإحصائية التي سيتم استخدامها؛ للإجابة على تساؤلات الدراسة.

الفصل السادس

عرض ومناقشة النتائج

1-عرض نتائج الدراسة

بعد فرز وجمع عينة الدراسة وتقييمها باستخدام منهجية PRISMA، عمدت الباحثة إلى تلخيص النتائج حسب التقسيم الذي ورد في القائمة وعرضها على شكل جداول منفصلة، بحيث يتم تقييم الخصائص السيكومترية أو صلاحية المقياس حسب دراسة COSMIN من خلال ثلاث جوانب رئيسية وهي: **الصدق، الثبات والاستجابة.**

وينقسم الصدق إلى ثلاث طرق لتقديره هي: **صدق المحتوى** والذي بدوره يتم تقييمه من خلال 5 جوانب (الملائمة، الشمولية، الفهم للعينة. والملائمة، الشمولية للمحكمين)، **صدق البناء** والذي بدوره ينقسم إلى (1:الصدق الهيكلي، 2:صدق الفرضيات- والذي يحتوي على الصدق التقاربي والتباعدي وصدق المقارنات الطرفية- 3:الصدق الثقافي)، وأخيرا **الصدق المحكي.**

أما الثبات فيضم **ثبات الاتساق الداخلي** (بجميع أنواعه ألفا كرونباخ، تجزئة نصفية وغيرهما)، **ثبات الإعادة والصور المتكافئة** والذي يضم أيضا ثبات المقيمين أي باختصار يضم جميع الإجراءات التي تتطلب تطبيق الاختبار أكثر من مرة واحدة، وليبقى **خطأ القياس** الإجراء الأخير للتحقق من الثبات.

وأخيرا **الاستجابة**، وهي تشير إلى قدرة الاختبار على اكتشاف التغيير بمرور الوقت في البنية المراد قياسها، رغم أنها تعتبر خاصية قياس منفصلة عن الصدق، إلا أن الاختلاف الوحيد هو أن الصدق يشير إلى صحة درجة النتائج، والاستجابة تشير إلى صحة درجة التغيير، لذلك معايير الاستجابة مماثلة لمعايير البناء والصدق المحكي (انظر الملحق 01). ويمكن الاطلاع على النموذج الذي يلخص كل ما سبق باختصار في الملحق (06). وتم دراسة كل تلك الخصائص السابقة الذكر في بنود موجودة في مربعات، وكل مربع يدرس خاصية معينة كما وتم شرحه سابقا

في فصل الإطار المنهجي، في الجزء المتعلق بأداة الدراسة. وقد قامت الباحثة بعرض جميع المربعات من أجل الإجابة على التساؤلات التالية:

- ما هي الطرق الأكثر استخداما من طرف الباحثين في إعادة تحققهم من صدق وثبات الاختبارات النفسية في عينة من الدراسات الأكاديمية العربية والأجنبية؟
- هل صدق المحتوى الذي تحقق منه هؤلاء الباحثين تم بطرق سليمة موافقة لمعايير COSMIN؟
- ما نسبة الدراسات التي تحققت من الصدق الهيكلي للاختبار النفسي بالشكل المطلوب؟
- هل تم التحقق من ثبات الاتساق الداخلي للاختبار بجميع أشكاله باستخدام الإجراءات الصحيحة؟
- هل تحققت الدراسات التي استخدمت ثبات الاتساق الداخلي من أحادية البعد بالنسبة للسمة قبل استخدامه؟
- هل تم التحقق من صلاحية الاختبارات المنقولة ثقافيا باستخدام الخطوات المحددة؟
- هل قام الباحثون بالكشف عن أخطاء القياس؟
- إلى أي مدى التزم الباحثون بالمعايير الموضوعة عند استخدامهم التطبيق وإعادة التطبيق وكذا الصور المتكافئة في التحقق من ثبات الاختبار؟
- ما مدى استخدام الباحثين للصدق المحكي والتزامهم بخطوات التحقق منه؟
- هل تم التحقق من صدق الفرضيات بالشكل المطلوب حسب معايير COSMIN؟
- هل تم تقييم الاستجابة للاختبارات النفسية بالطريقة الصحيحة حسب المعايير التي وضعتها COSMIN؟

- هل تعتبر الخصائص السيكومترية التي تم استخدامها في إعادة التحقق من صلاحية الاختبارات النفسية ذات جودة مقبولة؟

وللإجابة عن كل سؤال كان ولابد من طرح العديد من الأسئلة الفرعية التي تساعد على تحليل وتقييم كل طريقة تم استخدامها في إعادة التحقق من صدق وثبات الاختبارات النفسية، وفضلت الباحثة عرض كل رزمة من البنود التي تقيس شيئا مشتركا، في جدول واحد، وذلك من أجل الإجابة على كل سؤال فرعي. أي تم عرض البنود حسب ورودها في قائمة COSMIN (انظر الملحق01)

1-1-الإجابة عن السؤال الأول: ما هي الطرق الأكثر استخداما من طرف الباحثين في إعادة تحققهم من صدق وثبات الاختبارات النفسية في عينة من الدراسات الأكاديمية العربية والأجنبية؟

للإجابة على هذا السؤال لخصت الباحثة تكرارات كل طريقة تم استخدامها لإعادة التحقق من صلاحية الاختبارات النفسية في عينة الدراسة الحالية، وقد تم جمعها وتقييمها باستخدام قائمة COSMIN، ولذلك سنجد طرق تقدير الصدق والثبات حسب تصنيف COSMIN.

الجدول 05: طرق الصدق والثبات المستخدمة في إعادة التحقق من صلاحية الاختبارات النفسية

الصدق والثبات	التكرارات	القيم المفقودة
الملائمة للعينة	5	166
الشمولية للعينة	5	166
الفهم للعينة	5	166
الملائمة للمختصين	59	112
الشمولية للمختصين	59	112
الصدق الهيكلي	97	74
ثبات الاتساق الداخلي	152	19
الصلاحية عبر الثقافات	4	167
إعادة التطبيق والصور المتكافئة	50	121
الصدق المحكي	2	169
الصدق التقاربي والتباعدي	16	155
الصدق التمييزي	20	151

المصدر: من إعداد الباحثة

كما يتضح من الجدول أعلاه فإن أكثر أنواع أو طرق تقدير صدق وثبات الاختبارات النفسية هي ثبات الاتساق الداخلي (ألفا كرونباخ – تجزئة نصفية – كابا كوهين) بتكرار 152 مرة من أصل 171 اختبار، وسيتم في الجداول اللاحقة التفصيل في كل إجراء لوحده مثلا كم مرة استخدم ألفا كرونباخ وكم مرة استخدمت التجزئة النصفية والتحليل العاملي وغيرها من الإجراءات أثناء

تقييمنا لكل إجراء. يلي ثبات الاتساق الداخلي الصدق الهيكلي و الذي هو عبارة أيضا عن اتساق داخلي (الارتباط بين البنود والمحور أو ارتباط المحاور مع الدرجة الكلية، التحليل العاملي الاستكشافي، والتحليل العاملي التوكيدي) بتكرار 97 مرة، وللتذكير فإن الاتساق بين البنود والمحور أو المحاور والدرجة الكلية هو إجراء أضافته الباحثة كما سيتضح في الجدول المتعلق بالصدق الهيكلي. ليليها إجراء صدق المحتوى للمحكمين بتكرار 59 مرة، وبعدها ثبات التطبيق وإعادة التطبيق والصور المتكافئة بتكرار 50 مرة، أما إجراء صدق الصدق التمييزي فقد تم استخدامها في 20 دراسة وإجراء الصدق التقاربي والتباعدي في 16 دراسة، في حين تم التحقق من الصلاحية باستخدام إجراء الصدق الثقافي في 4 دراسات، ليأتي في المرتبة ما قبل الأخيرة إجراء صدق المحتوى للعينة بـ5 تكرارات، أما في المرتبة الأخيرة فالصدق المحكي بتكرارين فقط. وذلك لصعوبة الحصول على المعيار الذهبي وبالتالي يلجأ الباحثون إلى استخدام الصدق التقاربي بدلا من ذلك، وقد تغاضت الباحثة في هذا الجدول عن الأسماء التي أطلقها الباحثون في تقييمهم للصدق والثبات، ووضعت كل إجراء في مكانه المفروض، فمثلا إذا تم استخدام اختبار يقيس نفس السمة التي يقيسها الاختبار قيد المراجعة ولكنه ليس اختبار ذهبي–كما سيتم التفصيل فيه في معايير السؤال المتعلق بالصدق المحكي– فإن الباحثة ستعتبره صدق تقاربي حتى وإن أطلق عليه الباحث الصدق المحكي.

كذلك فضلت الباحثة الالتزام بالتقسيم الذي اعتمدته قائمة COSMIN في عرضها تكرارات الإجراءات بالتفصيل بدل عرض تكرار النوع المستخدم فقط فمثلا بدل عرض العدد الكلي لصدق المحتوى فضلت الباحثة عرض كل إجراء منفصلا عن غيره، لأن الدراسات التي قامت بالتحقق من صدق الملائمة والشمولية عند المحكمين ليست هي نفسها التي قامت بالتحقق من الملائمة

والشمولية والفهم حسب العينة، وعلى العموم فقد استخدم صدق المحتوى ككل في التحقق من 62 اختبار.

في حين اكتفت الباحثة بعرض الأرقام الإجمالية للنوع المستخدم عندما يتخالف ذلك مع التقسيم الذي اعتمدته القائمة COSMIN، مثل الصدق الهيكلي والذي يحمل في طياته 3 إجراءات في التحقق من صدق الاتساق الداخلي كما ذكرنا سابقا (العلاقة بين البنود والمحور والعلاقة بين المحاور والدرجة الكلية، التحليل العاملي الاستكشافي، التحليل العاملي التوكيدي) ولكن ذلك لا يمنع من التفصيل في كل إجراء وعدد تكراراته في الجدول المتعلق به، كما سيتضح من خلال الجداول الآتية التي سنجيب من خلالها على باقي الأسئلة ونفصل أكثر في الإجراءات المستخدمة وجودتها.

1-2-الإجابة عن السؤال الثاني: هل صدق المحتوى الذي تحقق منه هؤلاء الباحثين تم بطرق سليمة موافقة لمعايير COSMIN؟

اقترحت قائمة COSMIN مجموعة من المعايير تم صياغتها على شكل بنود لتقييم خمسة جوانب مرتبطة بصدق المحتوى، ثلاثة منها تخص العينة والجانبين المتبقيين يخصان المحكمين أو فريق الخبراء، وتمثلت الجوانب الخمسة في : الملائمة، الشمولية والفهم (بالنسبة للعينة)، والملائمة والشمولية (بالنسبة للمحكمين)، وللإجابة على هذا التساؤل الثاني تم طرح خمسة أسئلة فرعية لكل جانب من جوانب صدق المحتوى، وكل سؤال فرعي يتم تقييمه من قبل الباحثة بالإجابة على مجموعة من الأسئلة باستخدام مقياس ليكرت الرباعي الذي وضعته دراسة COSMIN والمتمثل في: **V** Very good= جيد جدا، **A** Adequate = مناسب، **D** Doubtful = مشكوك فيه، **I** Inadequate= غير كاف أو غير مناسب، N Not applicable= لم يطبق. ويتم تقييم كل

بند بالاعتماد على معايير COSMIN التي ستختصرها الباحثة تحت كل جدول، ويمكن إيجاد النسخة الأصلية في الملحق (03) أو في الرابط التالي: https://www.cosmin.nl/ .

1-2-1-السؤال الفرعي الأول: هل تم تقييم ملائمة بنود الاختبار بشكل مناسب مع العينة؟

للإجابة على هذا السؤال تم تقييم الاختبارات التي تم التحقق من صدق محتواها بالاعتماد على العينة.

الجدول 06: سؤال العينة عن الملائمة

سؤال العينة عن الملائمة	V	A	D	I	N
هل تم استخدام طريقة مناسبة لسؤال العينة عما إذا كان كل بند ملائمًا مع حالتهم؟	2	2	1	0	166
هل تم اختبار كل بند في عدد مناسب من العينة؟	5	0	0	/	166
هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة مع العينة؟	0	3	2	/	166
هل كانت الاجتماعات مع العينة أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟	0	2	3	/	166
هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟	0	1	4	0	166
هل تم استخدام منهج مناسب لتحليل البيانات؟	0	2	3	0	166
هل شارك باحثان على الأقل في التحليل؟	0	0	0	/	171

المصدر: من إعداد الباحثة

كما هو واضح من خلال النظرة العامة على الجدول أعلاه أن 5 دراسات فقط من قامت بالتحقق من ملائمة الاختبار للعينة المستهدفة، وقد اختلف تقييم الباحثة لكل خطوة من خطوات التحقق من الملائمة كما هو واضح؛ فالبند الأول الذي كانت صياغته كالآتي: هل تم استخدام طريقة مناسبة لسؤال العينة عما إذا كان كل بند ملائما مع حالتهم؟ تحصل على تقييمين جيدين جدا V، وتقيمين مناسبين A، في حين أن دراسة واحدة تحصلت على تقييم مشكوك فيه أو غير مناسبD. وكانت معايير هذا السؤال كالآتي: يمكن الحصول على التقييم **V** إذا تم التواصل

المباشر مع العينة وسؤال أفرادها عن وجهة نظرهم حول ملائمة الاختبار وأهمية كل بند بالنسبة لهم، ويحبذ أن تكون المقابلات فردية. أما الدراسات التي تم تقييمها بـ**A** فذلك استنادا إلى المعيار الخاص بها الذي ينص على: افتراض أن طريقة التحقق كانت مناسبة ولكن لم يتم وصفها بوضوح، في حين أن التقييم **D** يتم اعتماده في حالة كانت الطريقة غير واضحة أو مشكوك فيها بحيث لم يتم وصف كيف تم تقييم أهمية العناصر أو ملاءمتها بالنسبة للعينة. في حين أن التقييم **I** يتم اعتماده إذا كانت الطريقة غير مناسبة أو لم تسأل العينة عن الملائمة، أو لم يتم سؤال جميع أفراد العينة.

أما البند الثاني هل تم اختبار كل بند في عدد مناسب من العينة؟ فجميع الدراسات تحصلت على تقييم V أي جيد جدا، وذلك بالاعتماد على المعايير التالية: **V** 7 أفراد أو أكثر. **A** من 4 إلى 6 أفراد. **D** أقل من 4 أفراد. **I** لا يمكن الحصول على هذا التقييم أي يعتبر التقييم D هو أدنى تقييم يمكن الحصول عليه في هذا السؤال.

والبند الثالث الذي كانت صيغته كما يلي: هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة مع العينة؟ فقد تحصلت الدراسات على 3 تقييمات A و تقييمين لـ D وهذا يعني حسب المعايير التالية: **V** استخدام أشخاص ماهرين وذوي خبرة كبيرة. **A** أشخاص لديهم خبرة محدودة أو تم تدريبهم خصيصا لهذه الدراسة. **D** ليس واضح من هم الأشخاص أو مشكوك في أن هؤلاء الأشخاص ذوي خبرة كافية. **I** لا يمكن الحصول على هذا التقييم.

أما تقييمات البند الرابع هل كانت الاجتماعات مع العينة أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟ فقد انحصرت بين تقييمين لـ A و3 تقييمات لـ D. وتم الاعتماد على المعايير التالية: **V** تم الاعتماد على دليل المقابلة أو موضوع واضح ومدروس. **A** يفترض أن

المقابلة قامت على أساس موضوع مدروس مسبقا. **D** ليس واضح ما هو الموضوع الذي تم القيام بالمقابلة على أساسه أو إذا كان هنالك موضوع أصلا. **I** لا يمكن الحصول على هذا التقييم.

في حين أن البند الخامس هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟ دراسة واحدة فقط تم تقييمها بـ A في حين أن 4 دراسات الباقية تم تقييمها بـD. وكانت معايير هذا البند كالآتي: **V** تم تسجيل المقابلات والاجتماعات سواء باستخدام الكاميرا وغيرها من الوسائل أو عن طريق تسجيلها كتابيا ولكن على شكل نسخ حرفي. **A** يفترض أنه تم تسجيل المقابلات أو الاجتماعات، لكن لم يتم وصف ذلك بوضوح **D** مشكوك فيه إذا تم تسجيل المقابلات أو الاجتماعات حرفيا أو غير واضح ما الذي تم تسجيله. **I** لم يتم تسجيل المقابلات أو حتى تسجيل بعض النقاط.

في البند السادس الذي كانت صياغته: هل تم استخدام منهج مناسب لتحليل البيانات؟ تم تقييم دراستين بـ A و3 دراسات بـ D وكانت معايير هذا البند هي: **V** تم ترميز البيانات وإدخالها في البرامج المعروفة وحساب نسبة الاتفاق بين فرقة البحث وغيرها من الإجراءات التي تعمل على تحويل بيانات المقابلات إلى بيانات كمية. **A** لم يتم وصف التحليل بوضوح، إلا أنه تم الإشارة إلى استخدام البرامج الخاصة بتحليل البيانات النوعية، لذلك فمن المفترض أن النهج كان مناسبًا. **D** المنهج المستخدم مشكوك فيه أو غير واضح ما إذا كان تم تحليل البيانات بالطريقة المطلوبة. **I** المنهج المستخدم غير مناسب.

ليبقى البند السابع: هل شارك باحثين على الأقل في تحليل البيانات؟ بدون أي تقييم لأنه ولا دراسة تم فيها مشاركة باحثين أو أكثر لذلك لم يتم تقييم هذا البند. أما في ما يخص معاييره فهي كالآتي: **V** شارك باحثين أو أكثر **A** يفترض أنه شارك باحثين **D** ليس واضح من شارك في تحليل

البيانات. **I** لا يمكن الحصول على هذا التقييم. أي يعتبر التقييم D هو أدنى تقييم يمكن الحصول عليه في هذا السؤال.

وفي الأخير فإن المعيار **N** في جميع البنود السابقة وحتى اللاحقة، يعني دائما أن هذا البند لم يتم تقييمه من طرف الباحثة، وذلك لأنه لم يتم التحقق منه من قبل الباحثين، أي أن البند لم يطبق.

1-2-2-السؤال الفرعي الثاني: هل تم تقييم شمولية عناصر الاختبار بالطريقة المناسبة مع العينة؟

للإجابة على هذا السؤال يتم اعتماد نفس المقياس السابق V، A، D، I، N، وتقريبا نفس المعايير مع بعض الاختلافات التي ستوضحها الباحثة أسفل الجدول كما في السؤال السابق، وكانت النتائج كما هي موضحة في الجدول التالي:

الجدول 07: سؤال العينة عن الشمولية

سؤال العينة عن الشمولية	V	A	D	I	N
هل تم استخدام طريقة مناسبة لتقييم شمولية الاختبار؟	2	2	1	0	166
هل تم اختبار كل بند في عدد مناسب من العينة؟	5	0	0	/	166
هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة مع العينة؟	0	3	2	/	166
هل كانت الاجتماعات أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟	0	2	3	/	166
هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟	0	0	5	0	166
هل تم استخدام منهج مناسب لتحليل البيانات؟	0	1	4	0	166
هل شارك باحثان على الأقل في التحليل؟	0	0	0	/	171

المصدر: من إعداد الباحثة

يتضح من الجدول أن 5 دراسات من أصل 171 دراسة تحققت من صدق المحتوى عند عينة الدراسة. وكانت نتائج كل بند على حدى كما يلي: البند الأول هل تم استخدام طريقة مناسبة لتقييم شمولية الاختبار؟ تحصل على تقييم واحد V، وتقيم واحد مناسب A ، في حين أن 3 دراسات تحصلت على تقييم مشكوك فيه أو غير مناسب D . وكانت معايير هذا السؤال كالآتي: يمكن الحصول على التقييم **V** إذا تم التواصل المباشر مع العينة وسؤال أفرادها عن المفاهيم المفقودة. أما المعيار **A** فيتم اعتماده إذا طلب من العينة تعديل أو إضافة أو حذف أي بنود، أو تم الإشارة ضمنيا إلى إضافة بنود اقترحها أفراد العينة. في حين أن التقييم **D** يتم اعتماده في حالة كانت الطريقة غير واضحة أو مشكوك فيها بحيث لم يتم وصف كيف تم تقييم شمولية الاختبار بالنسبة للعينة. في حين أن التقييم **I** يتم اعتماده إذا كانت الطريقة غير مناسبة أو لم تسأل العينة عن الشمولية، أو لم يتم سؤال جميع أفراد العينة.

وبخصوص البند الثاني هل تم اختبار كل بند في عدد مناسب من أفراد العينة؟ فجميع الدراسات تحصلت على تقييم V أي جيد جدا، وذلك بالاعتماد على المعايير السابقة الذكر : **V** 7 أفراد أو أكثر. **A** من 4 إلى 6 أفراد. **D** أقل من 4 أفراد. **I** لا يمكن الحصول على هذا التقييم.

والبند الثالث الذي كانت صيغته كما يلي: هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة مع العينة؟ فقد تحصلت الدراسات على 3 تقييمات A و تقييمين لـ D وهذا يعني حسب المعايير التالية: **V** استخدام أشخاص ماهرين وذوي خبرة كبيرة. **A** أشخاص لديهم خبرة محدودة أو تم تدريبهم خصيصا لهذه الدراسة. **D** ليس واضح من هم الأشخاص أو مشكوك في أن هؤلاء الأشخاص ذوي خبرة كافية. **I** لا يمكن الحصول على هذا التقييم.

أما تقييمات البند الرابع هل كانت الاجتماعات مع العينة أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟ فقد انحصرت بين تقييمين لـ A و3 تقييمات لـ D. وتم الاعتماد على

المعايير التالية: **V** تم الاعتماد على دليل المقابلة أو موضوع واضح ومدروس. **A** يفترض أن المقابلة قامت على أساس موضوع مدروس مسبقا. **D** ليس واضح ما هو الموضوع الذي تم القيام بالمقابلة على أساسه أو إذا كان هنالك موضوع أصلا. **I** لا يمكن الحصول على هذا التقييم.

في حين أن البند الخامس هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟ جميع الدراسات 5 التي تحققت من الشمولية تم تقييمها بـD. وكانت معايير هذا البند كالآتي: **V** تم تسجيل المقابلات والاجتماعات سواء باستخدام الكاميرا وغيرها من الوسائل أو عن طريق تسجيلها كتابيا ولكن على شكل نسخ حرفي. **A** يفترض أنه تم تسجيل المقابلات أو الاجتماعات، لكن لم يتم وصف ذلك بوضوح **D** مشكوك فيه إذا تم تسجيل المقابلات أو الاجتماعات حرفيا أو غير واضح ما الذي تم تسجيله. **I** لم يتم تسجيل المقابلات أو حتى تسجيل بعض النقاط.

في البند السادس الذي كانت صياغته: هل تم استخدام منهج مناسب لتحليل البيانات؟ تم تقييم دراسة واحدة بـ A و4 دراسات بـ D وكانت معايير هذا البند هي: **V** تم ترميز البيانات وإدخالها في البرامج المعروفة وحساب نسبة الاتفاق بين فرقة البحث وغيرها من الإجراءات التي تعمل على تحويل بيانات المقابلات إلى بيانات كمية. **A** لم يتم وصف التحليل بوضوح، إلا أنه تم الإشارة إلى استخدام البرامج الخاصة بتحليل البيانات النوعية، لذلك فمن المفترض أن النهج كان مناسبًا. **D** المنهج المستخدم مشكوك فيه أو غير واضح ما إذا كان تم تحليل البيانات بالطريقة المطلوبة. **I** المنهج المستخدم غير مناسب.

ليبقى البند السابع: هل شارك باحثين على الأقل في تحليل البيانات؟ بدون أي تقييم لأنه ولا دراسة تم فيها مشاركة باحثين أو أكثر لذلك لم يتم تقييم هذا البند. أما في ما يخص معاييره فهي كالآتي: **V** شارك باحثين أو أكثر **A** يفترض أنه شارك باحثين **D** ليس واضح من شارك في تحليل

البيانات. **I** لا يمكن الحصول على هذا التقييم. أي يعتبر التقييم D هو أدنى تقييم يمكن الحصول عليه في هذا السؤال.

بالنسبة للمعايير المكررة في هذا الجدول والجدول السابق سيتم تكررها مرات أخرى في الأسئلة المتعلقة بصدق المحتوى ولتجنب إعادة كتباتها وتكرارها في كل مرة ستشير الباحثة إلى رقم البند ومعياره ورقم الجدول، ومن خلال ذلك يمكن للقارئ العودة إلى هذا الجدول أو الذي يسبقه وقراءة المعايير الموجودة أسفله حسب ترقيم البند.

1-2-3-السؤال الفرعي الثالث: هل تم تقييم فهم واستيعاب العينة لكافة بنود الاختبار؟

سيتم هنا عرض النتائج المرتبطة بتقييم الدراسات التي تحققت من فهم واستيعاب العينة.

الجدول 08: سؤال العينة عن فهم واستيعاب الاختبار

سؤال العينة عن فهم واستيعاب الاختبار	V	A	D	I	N
هل تم استخدام طريقة مناسبة لتقييم مدى فهم تعليمات الاختبار والبنود وخيارات الاستجابة وفترة الاسترجاع؟	3	2	0	0	166
هل تم اختبار كل بند في عدد مناسب من العينة؟	5	0	0	/	166
هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة مع العينة؟	0	4	1	/	166
هل كانت الاجتماعات أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟	0	2	3	/	166
هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟	0	2	3	0	166
هل تم استخدام منهج مناسب لتحليل البيانات؟	0	2	3	0	166
هل شارك باحثان على الأقل في التحليل؟	0	0	0	/	171

المصدر: من إعداد الباحثة

أيضا يتضح من خلال الجدول دائما أنه تم تقييم الفهم والاستيعاب في 5 دراسات فقط، وعلى العموم فإن التحقق من صدق المحتوى عند العينة بصفة عامة قامت به 5 دراسات فقط، ومن خلالها تم التحقق من كل من الملائمة، الشمولية والفهم. وفي خصوص البند الأول: هل تم استخدام طريقة نوعية مناسبة لتقييم مدى فهم تعليمات الاختبار والبنود وخيارات الاستجابة وفترة الاسترجاع؟ تحصل على 3 تقييمات V، وتقييمين مناسب A. وفي ما يتعلق بمعايير هذا السؤال فهي كالآتي: يمكن الحصول على التقييم **V** إذا تم التواصل المباشر مع العينة وسؤال أفرادها عن فهم بنود الاختبار وخيارات الاستجابة ومناسبة فترات الاسترجاع. أما المعيار**A** فيتم اختياره إذا افترضنا أنه تم سؤال أفراد العينة بطريقة مناسبة عن كل ما هو غير مفهوم ولكن لم يتم توضيح الطريقة. في حين أن التقييم **D** يتم اعتماده في حالة تم استخدام الأساليب الكمية فقط أو الطريقة غير واضحة أو مشكوك فيها بحيث لم يتم وصف كيف تم تقييم فهم أفراد العينة للاختبار. في حين أن التقييم **I** يتم اعتماده إذا كانت الطريقة غير مناسبة تماما.

وبخصوص البند الثاني هل تم اختبار كل بند في عدد مناسب من أفراد العينة؟ فجميع الدراسات 5 تحصلت على تقييم V أي جيد جدا، وذلك بالاعتماد على المعايير السابقة الذكر في البند الثاني من (الجدول رقم07).

أما البند الثالث والذي كانت صيغته كما يلي: هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة مع العينة؟ فقد تحصلت الدراسات على 4 تقييمات A و تقييم واحد لـ D وذلك حسب المعايير المذكورة في البند الثالث في (الجدول رقم07).

وفي ما يخص تقييمات البند الرابع هل كانت الاجتماعات مع العينة أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟ فقد انحصرت بين تقييمين لـ A و3 تقييمات لـ D. وتم الاعتماد على المعايير المتعلقة بالبند الرابع الموجودة تحت (الجدول رقم07).

في حين أن البند الخامس هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟ فقد تم تقييم A في دراستين والثلاث دراسات الأخرى تم تقييمها D. وبالنسبة لمعايير هذا البند فهي متواجدة تحت (الجدول رقم07) للبند الخامس.

في البند السادس الذي كانت صياغته: هل تم استخدام منهج مناسب لتحليل البيانات؟ تم تقييم دراستين بـ A و3 دراسات بـ D وكانت معايير هذا البند هي: **V** تم ترميز الاستجابات واستخدام جداول المخرجات وتحديد الاتجاهات وغيرها. **A** لم يتم وصف التحليل بوضوح، لذلك من المفترض أن المنهج كان مناسبًا. **D** المنهج المستخدم مشكوك فيه أو غير واضح ذكر أنه تم تطبيق الاختبار على العينة من أجل معرفة إذا كان الاختبار مفهوم لكن لم يتم ذكر كيف تم التعامل مع النتائج . **I** المنهج المستخدم غير مناسب.

وفي الأخير البند السابع: هل شارك باحثين على الأقل في تحليل البيانات؟ كالعادة بدون أي تقييم لأنه ولا دراسة تم فيها مشاركة باحثين أو أكثر لذلك لم يتم تقييم هذا البند. أما في ما يخص معاييره فهي موجودة أسفل (الجدول رقم07) في معايير البند السابع.

1-2-4-السؤال الفرعي الرابع: هل تم تقييم ملائمة بنود الاختبار بالطريقة المناسبة مع المحكمين والخبراء؟

بعد عرض نتائج تقييم الدراسات التي تحققت صدق المحتوى عند أفراد العينة، سيتم هنا عرض نتائج تقييم الدراسات التي تحققت من صدق المحتوى عند الخبراء والمحكمين.

الجدول 09: سؤال المحكمين عن الملائمة

سؤال المحكمين عن الملائمة	V	A	D	I	N
هل تم استخدام طريقة مناسبة لسؤال المحكمين عما إذا كان كل عنصر مناسبًا لبناء الاهتمام؟	18	15	23	3	112
هل تم تضمين المحكمين من جميع التخصصات ذات الصلة بالسمة المدروسة؟	11	24	24	/	112
هل تم اختبار كل بند في عدد مناسب من المحكمين؟	42	4	13	/	112
هل تم استخدام منهج مناسب لتحليل البيانات؟	13	14	32	0	112
هل شارك باحثان على الأقل في التحليل؟	4	1	0	/	166

المصدر: من إعداد الباحثة

في هذا الجدول تم التحقق من صدق ملاءمة المحتوى عند الخبراء والمحكمين في 59 دراسة. وكانت نتائج البنود مفصلة كما يلي: بداية بالبند الأول هل تم استخدام طريقة مناسبة لسؤال المحكمين عما إذا كان كل عنصر مناسبًا لبناء الاهتمام؟ تحصل 18 اختبار على تقييم جيد جدا V، و15 اختبار على تقييم مناسب A ، في حين أن 23 اختبار تحصل على تقييم مشكوك فيه D . لتبقى 3 اختبارات تحصلت على تقييم **I**، وكانت معايير هذا السؤال كالآتي: يمكن الحصول على التقييم **V** إذا تم تزويد المحكمين بقائمة الأهداف والموافقة وعناصر الاختبار، واستنباط ملاحظتهم بطريقة مدروسة وموحدة. **A** يفترض أنه تم سؤال المحكمين عن ملاءمة البنود ووضوحها إلى انه لم يتم وصف الطريقة بوضوح. **D** ليس من الواضح ما إذا تم سؤال المحكمين عما إذا كان كل عنصر ذي صلة أم لا، أو إذا كان هناك شك بأن الطريقة مناسبة. في حين أن التقييم **I** يتم اعتماده إذا كانت الطريقة غير مناسبة أو لم يسأل المحكمون عن مدى ملاءمة جميع البنود.

وبخصوص البند الثاني هل تم تضمين المحكمين من جميع التخصصات ذات الصلة بالسمة المدروسة؟ فكانت النتائج بأن 11 اختبار تحصل على تقييم V أي جيد جدا، و24 اختبار تحصل

على تقييم مناسب **A.** ونفس العدد بالنسبة للاختبارات التي تحصلت على تقييم **D.** وكان ذلك استنادا إلى المعايير التالية: **V** تم تضمين خبراء من جميع التخصصات المطلوبة مثل الباحثين، الأطباء، العاملين في مجال الرعاية الصحية وغيرها. **A** يفترض أن تم تضمين محكمين وخبراء من عدة تخصصات لكن لم يتم وصفهم بوضوح. **D** ليس واضح من هم الأشخاص أو مشكوك في أن هؤلاء الأشخاص من عدة تخصصات. **I** لا يمكن الحصول على هذا التقييم.

وبالنسبة للبند الثالث هل تم اختبار كل بند في عدد مناسب من المحكمين؟ فقد تحصل 42 اختبار على تقييم V . وتحصلت 4 اختبارات على تقييم A .في حين تحصل 13 اختبار الباقي على تقييم D وهذا حسب المعايير التالية: **V** 7 أفراد أو أكثر. **A** من 4 إلى 6 أفراد. **D** أقل من 4 أفراد. **I** لا يمكن الحصول على هذا التقييم.

أما تقييمات البند الرابع الذي كانت صياغته: هل تم استخدام منهج مناسب لتحليل البيانات؟ تم تقييم 13 اختبار بـV أي جيد جدا. و14 اختبار بـ A أي مناسب. و32 اختبار بـ D أي مشكوك فيه. في حين لم يقيم ولا اختبار بـ I أي غير مناسب. وكانت معايير هذا البند هي: **V** مثلا طلب من المحكمين إكمال استبيان يقيمون فيه الاختبار من خلال الرد على الأسئلة الواردة في الاستبيان، لمعرفة نسبة الاتفاق بينهم كدليل على أهمية كل بند . **A** يفترض أنه مناسب لكن لم يتم وصفه بوضوح. **D** ليس واضح ماهو المنهج المتبع أو مشكوك فيه. **I** المنهج المستخدم غير مناسب.

ليبقى في الأخير البند الخامس: هل شارك باحثين على الأقل في تحليل البيانات؟ تم تقييم 4 اختبارات بـV أي جيد جدا. و اختبار واحد بـ A أي مناسب. في حين لم يقيم ولا اختبار بـ D أو I في أي دراسة. أما في ما يخص معاييره فهي كالآتي: **V** شارك باحثين أو أكثر **A** يفترض أنه

شارك باحثين لكن لم يتم وصفهم بوضوح **D** ليس واضح إذا كان 2 أو واحد فقط. **I** لا يمكن الحصول على هذا التقييم. أي يعتبر التقييم D هو أدنى تقييم يمكن الحصول عليه في هذا السؤال.

1-2-5-السؤال الفرعي الخامس: هل تم تقييم شمولية بنود الاختبار بالطريقة المناسبة مع المحكمين والخبراء؟

للإجابة على هذا السؤال تم تخصيص بضعة بنود للكشف عن مدى تقييم الباحثين لشمولية الاختبار لجميع جوانب السمة المدروسة، وكانت النتائج كما هو موضح في الجدول التالي:

الجدول 10: سؤال المحكمين عن الشمولية

سؤال المحكمين عن الشمولية	V	A	D	I	N
هل تم استخدام طريقة مناسبة لتقييم شمولية الاختبار مع المحكمين والخبراء؟	7	3	41	8	112
هل تم إدراج محكمين من جميع التخصصات ذات الصلة؟	9	19	31	/	112
هل تم اختبار كل بند في عدد مناسب من المحكمين؟	42	3	14	/	112
هل تم استخدام منهج مناسب لتحليل البيانات؟	11	11	37	0	112
هل شارك باحثان على الأقل في التحليل؟	4	1	1	/	165

المصدر: من إعداد الباحثة

كما هو متوقع فإنه تم التحقق من صدق شمولية المحتوى عند الخبراء والمحكمين في 59 دراسة أيضا مثل صدق ملاءمة المحتوى. أما نتائج البنود بالتفصيل فهي كالآتي: بالنسبة للبند الأول: هل تم استخدام طريقة مناسبة لتقييم شمولية الاختبار مع المحكمين والخبراء؟ تحصلت 7 اختبارات على تقييم جيد جدا V، و3 اختبارات على تقييم مناسب A ، في الوقت الذي تحصل في 41 اختبار على تقييم مشكوك فيه D. وأخيرا 8 اختبارات تحصلت على تقييم **I**، وكانت معايير هذا السؤال كالآتي: يتم الحصول على التقييم **V** إذا سئل المحكمون عما إذا كان قد تم طرح أي

قضايا مهمة محذوفة، أو تم إهمال أي جوانب من السمة وإغفالها. **A** يفترض أنه تم سؤال المحكمين عن شمولية البنود إلى انه لم يتم وصف الطريقة بوضوح، مثل ذكر الباحثين في النتائج أنه تم إضافة عامل جديد بناءا على رأي المحكمين، لكن لم يتم وصف الطريقة بوضوح. **D** ليس من الواضح ما إذا تم سؤال المحكمين عما إذا كان هنالك عناصر أو جوانب مفقودة وتم إغفالها في الدراسة. في حين أن التقييم **I** يتم اعتماده إذا كانت الطريقة غير مناسبة أو لم يتم التحقق من شمولية الاختبار.

وبخصوص البند الثاني هل تم إدراج المحكمين من جميع التخصصات ذات الصلة بالسمة المدروسة؟ فكانت النتائج بأن 9 اختبارات تحصلت على تقييم V أي جيد جدا، و19 اختبار تحصل على تقييم مناسب **A**. في حين أن العدد الأكبر للاختبارات بتكرار 31 اختبار تحصلت على تقييم **D**. وكان ذلك استنادا إلى المعايير التالية: **V** تم تضمين خبراء من جميع التخصصات المطلوبة مثل الباحثين، الأطباء، العاملين في مجال الرعاية الصحية وغيرها. **A** يفترض أن تم تضمين محكمين وخبراء من عدة تخصصات لكن لم يتم وصفهم بوضوح. **D** ليس واضح من هم الأشخاص أو مشكوك في أن هؤلاء الأشخاص من عدة تخصصات. **I** لا يمكن الحصول على هذا التقييم.

وبالنسبة للبند الثالث هل تم اختبار كل بند في عدد مناسب من المحكمين؟ فقد تحصل 42 اختبار على تقييم V . وتحصلت 3 اختبارات على تقييم A .في حين تحصل 14 اختبار الباقي على تقييم D وهذا حسب المعايير التالية: **V** 7 أفراد أو أكثر. **A** من 4 إلى 6 أفراد. **D** أقل من 4 أفراد. **I** لا يمكن الحصول على هذا التقييم.

أما تقييمات البند الرابع الذي كانت صياغته: هل تم استخدام منهج مناسب لتحليل البيانات؟ تم تقييم 11 اختبار بـV أي جيد جدا. و11 اختبار بـ A أي مناسب. و37 اختبار بـ D أي مشكوك

فيه. في حين لم يقيم ولا اختبار بـ I أي غير مناسب. وكانت معايير هذا البند هي: **V** مثلا طلب من المحكمين إكمال استبيان يقيمون فيه الاختبار من خلال الرد على الأسئلة الواردة في الاستبيان، لمعرفة نسبة الاتفاق بينهم كدليل على أهمية كل بند . **A** يفترض أنه مناسب لكن لم يتم وصفه بوضوح. **D** ليس واضح ما هو المنهج المتبع أو مشكوك فيه. **I** المنهج المستخدم غير مناسب.

ليبقى في الأخير البند الخامس: هل شارك باحثين على الأقل في تحليل البيانات؟ تم تقييم 4 اختبارات بـV أي جيد جدا. و اختبار واحد بـ A أي مناسب. وأيضا تقييم اختبار واحد بـ D. ولم يتم تقييم I في أي دراسة. أما في ما يخص معاييره فهي كالآتي: **V** شارك باحثين أو أكثر **A** يفترض أنه شارك باحثين لكن لم يتم وصفهم بوضوح **D** ليس واضح إذا كان 2 أو واحد فقط. **I** لا يمكن الحصول على هذا التقييم. أي يعتبر التقييم D هو أدنى تقييم يمكن الحصول عليه في هذا السؤال.

عند استخدام صدق المحتوى يفترض أن يتم التحقق من صلاحية الاختبار من خلال العينة والخبراء معا، على عكس ما وجدناه في عينة الدراسة الحالية، بحيث من أصل 171 دراسة فقط دراستين من قامت بالتحقق من صدق المحتوى عند العينة والمحكمين مع بعض في حين أن 3 دراسات الباقية من أصل 5 دراسات تحققت من صدق المحتوى عند العينة فقط، وفي المقابل تحققت 57 دراسة من صدق المحتوى عند المحكمين فقط.

1-3-عرض نتائج السؤال الثالث: ما نسبة الدراسات التي تحققت من الصدق الهيكلي للاختبار النفسي بالشكل المطلوب؟

يهدف الصدق الهيكلي إلى الكشف عن أحادية البعد كهدف رئيسي، لذلك قبل الإجابة عن هذا السؤال كان لابد من معرفة الاختبار هل هو أحادي البعد أو يحتوي على العديد من المحاور،

لنتمكن بعد ذلك من تقييم الصدق الهيكلي وثبات الاتساق الداخلي وغيرها من الإجراءات على هذا الأساس، فالاختبار الذي يقيس عدة أبعاد، لا يمكن استعمال ألفا لدرجته الكلية مثلا. لهذا تم تفريع السؤال الثالث إلى سؤالين:

1-3-1-السؤال الفرعي الأول: هل الاختبار أحادي البعد؟

والمقصود هنا بأحادية البعد أي أن الاختبار عبارة عن مجموعة بنود تشكل معا محورا واحدا فقط، في حين الاختبار التي تحتوي على عدة أبعاد تكون على شكل اختبار يحتوي على عدة عوامل بحيث تشكل كل مجموعة من البنود عامل ما، لتشكل تلك العوامل الاختبار الكلي. ويتم من خلال الإجابة على هذا السؤال معرفة الاختبار الذي يحتوي على العديد من الأبعاد والاختبار الذي هو عبارة عن بعد واحد، ليتم بعد ذلك تقييم باقي الإجراءات على هذا الأساس.

الجدول 11: الكشف عن أحادية البعد

هل الاختبار أحادي البعد؟	التكرارات	النسبة المئوية
نعم	63	%36.8
لا	34	%19.9
N	74	%43.3
الكلية	171	%100

المصدر: من إعداد الباحثة

من النظر إلى الجدول يظهر لنا جليا أن عدد الاختبارات التي تحتوي على الأبعاد أكثر من الاختبارات التي تنطوي على بعد واحد، فمقابل 63 اختبار نفسي يحتوي على أبعاد نجد 34 اختبار أحادي البعد، أما في ما يخص الاختبارات 74 الباقية فهي تدل على الدراسات التي لم يتم

فيها التحقق من صدق الاتساق الداخلي، أي لم يتم فيها الكشف عن أحادية البعد للاختبار أو استكشاف أو تأكيد العوامل التي يحتويها.

1-3-2-السؤال الفرعي الثاني: هل تم التحقق من الصدق الهيكلي للاختبار النفسي بالشكل المطلوب؟

وللإجابة على هذا السؤال عدلت الباحثة في خصوص معيار البند الأول كما سيتضح:

الجدول 12: تقييم الصدق الهيكلي

N	I	D	A	V	الصدق الهيكلي
74	2	/	70	25	بالنسبة إلى النظرية التقليدية CTT[299]: هل تم إجراء الاتساق الداخلي، التحليل العاملي الاستكشافي أو التوكيدي؟
171	0	0	0	0	بالنسبة للنظرية الحديثة Rasch[**]/IRT[***]: هل يتناسب النموذج المختار مع سؤال البحث؟
74	5	38	11	43	هل كان حجم العينة المشمول في التحليل كافيا؟
74	7	68	/	22	هل كان هناك أي عيوب أخرى مهمة في طريقة الاستخدام أو الأساليب الإحصائية الأخرى؟

المصدر: من إعداد الباحثة

يمكننا من خلال الجدول استنتاج أن عدد الدراسات التي تحققت من الصدق الهيكلي قدرت بـ97 دراسة وهو عدد لا بأس، ويعتبر ثاني أكثر إجراء استخداما في إعادة التحقق من صلاحية الاختبارات النفسية في هذه الدراسة. أما في ما يخص نتائج بنود هذا الإجراء بالتفصيل فهي كما يلي: بداية البند الأول والذي فحواه: بالنسبة إلى النظرية التقليدية CTT: هل تم إجراء الاتساق

[299] CTT : Classical Test Theory النظرية الكلاسيكية للقياس
[**] Rasch Model نموذج راش
[***] IRT : Item Response Theory نظرية الاستجابة للفقرة

الداخلي، التحليل العاملي الاستكشافي أو التوكيدي؟ تحصل 25 اختبار على تقييم V، و70 اختبار على تقييم A ، ولم يتحصل أي اختبار على تقييم مشكوك فيه D. وفي الأخير تحصل اختبارين فقط على تقييم **I**، ومن أجل فهم هذه التقييمات وجب ذكر المعايير التالية: يتم الحصول على التقييم جيد جدا **V** إذا تم استخدام التحليل العاملي التوكيدي وذلك لأن الاختبار هو اختبار جاهز يتم إعادة التحقق من صلاحيته، كذلك يمكن تقييم التحليل العاملي الاستكشافي بجيد جدا أيضا في حالة تم تغيير بنية الاختبار أو تم ترجمة الاختبار مثلا. أما التقييم **A** فيصلح للاختبارات التي تم التحقق من صدق هيكلها باستخدام التحليل العاملي الاستكشافي للاختبار الجاهز أو تم استخدام صدق الاتساق الداخلي (الارتباط بين البنود والمحور وبين المحاور والدرجة الكلية) وهو ما أضافته الباحثة. التقييم **D** لا يمكن منحه في هذا الإجراء. في حين أن التقييم **I** يتم اعتماده إذا لم يتم إجراء أي مما سبق، أي يذكر الباحث أنه تحقق من أحادية البعد أو صدق الاتساق الداخلي لكنه لم يجري أي إجراء من الإجراءات الثلاثة السابقة.

وبخصوص البند الثاني بالنسبة للنظرية الحديثة IRT/Rasch: هل يتناسب النموذج المختار مع سؤال البحث؟ فكانت النتائج بأنه لم يتم استخدام النظرية الحديثة في أي دراسة من الدراسات 171. ورغم ذلك قررت الباحثة عرض المعايير الخاصة به باختصار يمكن منح التقييم **V** إذا كان النموذج المختار يناسب السؤال بشكل جيد. **A** يفترض أن النموذج المختار يناسب سؤال البحث ولكنه غير موصوف جيدا. **D** مشكوك فيه إذا كان يناسب بشكل جيد سؤال البحث. **I** النموذج المختار لا يتناسب مع سؤال البحث.

وبالنسبة للبند الثالث هل كان حجم العينة المشمول في التحليل كافيا؟ فقد تحصل 43 اختبار على تقييم V . وتحصل 11 اختبار على تقييم A. في حين تحصل 38 اختبار على تقييم D. أما الاختبارات 5 الباقية تحصلت على تقييم I. وهذا حسب المعايير التالية: للحصول على تقييم

Vبالنسبة للنظرية التقليدية: في التحليل العاملي الاستكشافي أو التوكيدي يجب أن تكون العينة 7 مرات ضعف عدد البنود وأكبر أو تساوي 100، وفي الاتساق الداخلي أن يكون حجم العينة أكبر من 50. وبالنسبة لنموذج راش يجب أن تكون أكبر أو تساوي 200، أما حسب نظرية الاستجابة للفقرة أكبر أو تساوي 1000.

أما للحصول على تقييم **A** بالنسبة للنظرية التقليدية: في التحليل العاملي الاستكشافي أو التوكيدي يجب أن تكون العينة على الأقل 5 مرات ضعف عدد البنود وأكبر أو تساوي 100، وفي الاتساق الداخلي أن يكون حجم العينة يساوي 50. وبالنسبة لنموذج راش يجب أن تكون بين 100–199 ، أما حسب نظرية الاستجابة للفقرة بين 500–999. أما تقييم **D** بالنسبة للنظرية التقليدية: في التحليل العاملي الاستكشافي أو التوكيدي يجب أن تكون العينة 5 مرات ضعف عدد البنود ولكن أقل من 100، وفي الاتساق الداخلي أن يكون حجم العينة أقل من 50. وبالنسبة لنموذج راش يجب أن تكون بين 50–99، أما حسب نظرية الاستجابة للفقرة تكون بين 250–499. وفي الأخير للحكم على البند بالتقييم **I** فبالنسبة للنظرية التقليدية في التحليل العاملي الاستكشافي أو التوكيدي أن تكون العينة أقل من 5 مرات ضعف عدد البنود، وفي الاتساق الداخلي أن يكون حجم العينة أقل من 20. وبالنسبة لنموذج راش يجب أن تكون أقل من 50، أما حسب نظرية الاستجابة للفقرة أقل من 250.

أما تقييمات البند الرابع الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ تم تقييم 22 اختبار بـV أي جيد جدا. و68 اختبار بـ D أي مشكوك فيه. و7 اختبارات بـ I أي غير مناسب. وكانت معايير هذا البند هي: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا المعيار في تقييم هذا البند. **D** توجد عيوب لكن ثانوية مثلا طريقة الدوران للعوامل غير موصوفة. **I** عيوب منهجية مهمة مثل طريقة الدوران غير ملائمة.

كما هو واضح من المعايير السابقة المتعلقة بالبند الأول فإن التقييم V يخص التحليل العاملي التوكيدي، أما التحليل العاملي الاستكشافي في حالة تم تغيير بنية الاختبار أو ترجمته ونقله إلى بيئة جديدة يأخذ التقييم V أما في غير تلك الحالات فإنه يأخذ التقييم A، وفي هذا الدراسة لم يأخذ ولا تقييم A وهو الأمر الصائب لأنه يصلح عند استكشاف بينة التحقق من اختبار جديد، أما في الدراسات التي تمثل عينة بحثنا والتي هي مرتبطة بإعادة التحقق من صلاحية الاختبارات فإن الأصح هو التحليل العاملي التوكيدي، وبالنسبة للتقييم V في هذا البند فهو موزع بين 17 تقييم للتحليل العاملي التوكيدي و8 تقييمات للتحليل العاملي الاستكشافي، أما التقييم A فكان كله من نصيب الدراسات التي تحققت من الصدق الهيكلي باستخدام الاتساق الداخلي (العلاقة بين البنود والمحور والعلاقة بين المحاور والدرجة الكلية) وقدر عدد هذه الدراسات بـ71 دراسة.

1-4-عرض نتائج السؤال الرابع: هل تم التحقق من ثبات الاتساق الداخلي للاختبار بجميع أشكاله باستخدام الإجراءات الصحيحة؟

قبل عرض نتائج هذا السؤال، تجدر الإشارة إلى أن مبادرة COSMIN لم تتطرق إلى التجزئة النصفية إطلاقا، ولأنها تم استخدامها في العديد من الدراسات التي تمثل عينة الدراسة الحالية، فقد أدرجتها الباحثة في البنود وفي المعايير، كما سيتم توضيحه في الجدول والمعايير الخاصة به تاليا:

الجدول 13: ثبات الاتساق الداخلي

ثبات الاتساق الداخلي	V	A	D	I	N
هل تم حساب إحصاء الاتساق الداخلي لكل مقياس أحادي البعد (فرعي) على حدة؟	119	/	11	22	19
هل تم التحقق من التجزئة النصفية باعتماد إجراءات سليمة؟	5	4	16	0	146
هل تم حساب ألفا كرونباخ أو أوميغا بالنسبة للنتائج المستمرة–متغير متصل–؟	114	/	25	1	31
بالنسبة إلى الدرجات ثنائية التفرع: هل تم حساب ألفا كرونباخ أو كيودر ريتشاردسون 20–KR؟	10	/	1	0	160
بالنسبة للدرجات المستندة إلى IRT: هل تم حساب الخطأ القياسي للثيتا (SE (θ)) أو معامل الموثوقية لقيمة السمات الكامنة المقدرة (مؤشر الفصل (الموضوع أو العنصر))؟	0	/	/	0	171
هل توجد أي عيوب أخرى مهمة؟	90	/	61	1	19

المصدر: من إعداد الباحثة

إن ثبات الاتساق الداخلي هو الطريقة الأكثر استخداما في التحقق من صلاحية الاختبارات النفسية في عينة الدراسة الحالية، بحيث استخدم في التحقق من صلاحية 152 اختبار. وسيتم هنا التفصيل في كل إجراء تم استخدامه وسيتم بداية عرض النتائج حسب ترتيب البنود بداية بالبند الأول: هل تم حساب إحصاء الاتساق الداخلي لكل مقياس أحادي البعد (فرعي) على حدة؟ وقد تقييم 119 اختبار بتقييم V أي جيد جدا، و11 اختبار تم تقييمه مشكوك فيه D. ليبقى 22 اختبار تحصل على تقييم **I** أي غير مناسب أو غير كاف، وقد تم اعتماد المعايير التالية في التقييم: يتم الحصول على التقييم جيد جدا **V** إذا تم حساب ثبات الاتساق الداخلي لمقياس أحادي البعد أو لكل مقياس فرعي على حدى. أما التقييم **A** فإنه غير معتمد في هذا البند. وفيما يخص التقييم **D** فإنه

يتم منحه إذا كان من غير الواضح هل تم استخدام ثبات الاتساق الداخلي مع مقياس أحادي البعد أو متعدد الأبعاد. في حين أن التقييم **I** يتم اعتماده إذا تم استخدام الاتساق الداخلي مع مقياس متعدد الأبعاد.

وفيما يخص البند الثاني: هل تم التحقق من التجزئة النصفية باعتماد إجراءات سليمة؟ تم تقييم 5 اختبارات بتقييم V . وأيضا 4 اختبارات بتقييم A. في حين تحصل 16 اختبار على تقييم D. ولم يتم منح أي اختبار تقييم I. وقد استندت هذه التقييمات إلى المعايير التالية: يتم منح التقييم **V** إذا تم تقسيم الاختبار بشكل جيد واعتماد معامل التصحيح المناسب. **A** يفترض أن الطريقة سليمة لكن لم يتم التفصيل فيها مثل ذكر أن العدد الكلي لبنود الاختبار فردي وتم استخدام معادلة جوتمان بدل سبيرمان براون التصحيحية. **D** مشكوك في أنه تم تقسيم الاختبار بشكل سليم. **I** تم تقسيم الاختبار بطريقة خاطئة، أو تم استخدام معادل تصحيح غير مناسب.

وبالنسبة للبند الثالث هل تم حساب ألفا أو أوميغا من كرونباخ بالنسبة للنتائج المستمرة ؟ فقد تحصل 114 اختبار على تقييم V. وتحصل 25 اختبار على تقييم D. في حين تحصل اختبار واحد على تقييم I. وذلك حسب المعايير الآتية: للحصول على تقييم **V** تم حساب ألفا كرونباخ أو بديله أوميغا في حالة لم تتحقق شروط ألفا كرونباخ. أما التقييم **A** فلم يتم اعتماده في هذا البند. وبالنسبة للتقييم **D** فإنه يمنح في حالة تم حساب الارتباطات الإجمالية بين البنود أي أنه لم يتم حساب ألفا لكل بعد أو استخدام أوميغا. وفي الأخير للحكم على البند بالتقييم **I** فإنه لم يتم حساب لا ألفا كرونباخ و لم يتم حساب الارتباطات الإجمالية بين البنود.

أما البند الرابع: بالنسبة إلى الدرجات ثنائية التفرع: هل تم حساب ألفا كرونباخ أو KR-20؟ تم تقييم 10 اختبار بجيد جدا V. واختبار واحد بمشكوك فيه D. بالاعتماد على هذه المعايير: للحصول على تقييم **V** تم حساب ألفا كرونباخ أو كيودر ريتشاردسون 20. أما التقييم **A** و**D** فلم

يتم اعتمادهم في هذا البند. ليبقى التقييم **I** الذي يمنح في حالة لم يتم حساب لا ألفا كرونباخ و لا كيودر ريتشاردسون ولم يتم حساب حتى الارتباطات الإجمالية بين البنود.

في حين أن البند الخامس بالنسبة للدرجات المستندة إلى IRT: هل تم حساب الخطأ القياسي للثيتا (SE (θ)) أو معامل الموثوقية لقيمة السمات الكامنة المقدرة (مؤشر الفصل (الموضوع أو العنصر))؟ فكانت النتائج بأنه لم يتم استخدام النظرية الحديثة كما ذكرنا سابقا في أي دراسة من الدراسات 171. أما بخصوص المعايير الخاصة به فهي منح التقييم **V** إذا تم حساب الخطأ القياسي لثيتا أو حساب معامل الموثوقية. ولا يتم منح التقييم **A** و**D** في هذا البند. في حين أن التقييم **I** يتم منحه إذا لم يتم حساب الخطأ القياسي لثيتا أو حساب معامل الموثوقية.

وفي الأخير فإن تقييمات البند السادس الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ تم منح 90 اختبار تقييم V أي جيد جدا. و61 اختبار تقييم D أي مشكوك فيه. واختبار واحد تقييم I أي غير مناسب. وكانت معايير هذا البند هي: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا التقييم في هذا البند. **D** توجد عيوب لكن ثانوية. **I** توجد عيوب منهجية مهمة.

إن ثبات الاتساق الداخلي هو الطريقة الأكثر استخداما في هذه الدراسة للتحقق من ثبات الاختبارات النفسية، أما الإجراء الأكثر استعمالا في هذه الطريقة من بين جميع إجراءات الاتساق الداخلي هو ألفا كرونباخ بحيث استخدم في التحقق من ثبات 149 اختبار، في حين أنه تم استخدام التجزئة النصفية في التحقق من ثبات 25 اختبار، وقد صاحب ألفا كرونباخ التجزئة النصفية في 19 اختبار، وقد صنف 16 منها بـ D أي مشكوك فيه، وذلك بسبب استخدام أساليب غير مناسبة مثل استخدام سبيرمان براون مع عدم توفر شرط التساوي بين نصفي الاختبار، أما كيودرريتشاردسون تم استخدامه في دراستين فقط، أما أوميغا ماكدونالدز فلم يتم استخدامه في

التحقق من ثبات أي اختبار، رغم أن 33 دراسة تم تقييمها ب I و D في البند الأول الذي يكشف عن استخدام الاتساق الداخلي مع اختبار أحادي البعد أو لا، إلا أن 26 دراسة منها التي كانت من المفروض أن تستخدم أوميغا ماكدونالدز بدل ألفا كرونباخ، أي 26 دراسة تحققت من ثبات الاختبار باستخدام ألفا كرونباخ لاختبار متعدد الأبعاد، و دراسة واحدة استخدمت كيودرريتشاردسون، في حين 6 دراسات تم التحقق من ثباتها باستخدام التجزئة النصفية للاختبار ككل رغم أنه متعدد الأبعاد.

1-5-عرض نتائج السؤال الخامس: هل تحققت الدراسات التي استخدمت ثبات الاتساق الداخلي من أحادية البعد قبل استخدامه؟

إن ألفا كرونباخ يشترط أحادية البعد قبل استخدامه، أي أنه لا يصلح مع الاختبارات المتعددة الأبعاد، ويتم اقتراح بديله أوميغا في تلك الحالة، أو ينصح على الأقل باستخدام ألفا لكل بعد على حدى، ولذلك تؤكد قائمة COSMIN على ضرورة الكشف عن بنية الاختبار وعدد العوامل والأبعاد التي يحتويها قبل استخدام ألفا كرونباخ أو ثبات الاتساق الداخلي بصفة عامة، ولذلك عمدت الباحثة إلى هذا الجدول المتقاطع من أجل الكشف عن مدى التزام الباحثين بهذا الشرط، وقد أتت النتائج كما هي موضحة في الجدول التالي:

الجدول 14: الجدول المتقاطع بين ثبات الاتساق الداخلي والكشف عن بنية الاختبار

		ثبات الاتساق الداخلي				
		مرتفع	متوسط	منخفض	منخفض جدا	الكلية
هل تم إجراء تحليل عامل استكشافي أو تأكيدي؟	N	42	10	14	1	67
	I	1	0	0	0	1
	A	36	14	8	3	61
	V	17	3	2	1	23
	الكلية	96	27	24	5	152

المصدر: من إعداد الباحثة

فضلت الباحثة عرض هذا الجدول المتقاطع بين البند الأول للمربع الثالث والذي يكشف عما إذا تم التحقق من أحادية البعد والمربع الرابع المتعلق بثبات الاتساق الداخلي، بدل عرض المربع الرابع للصدق الهيكلي مع المربع الثالث لثبات الاتساق الداخلي، وذلك لأن هذا الجدول يكشف بطريقة مفصلة أكثر عن الإجراءات المستخدمة في الكشف عن أحادية البعد، وبداية فقد استخدام ثبات الاتساق الداخلي والتحقق من بنية الاختبار باستخدام التحليل العاملي الاستكشافي والتوكيدي (انظر التعليق الموجود تحت الجدول 12) في 23 دراسة، وقد تحصل 17 اختبار منهم على تقييم مرتفع في ثبات الاتساق الداخلي، في حين تحصلت 3 اختبارات على تقييم متوسط، واختبارين على تقييم منخفض، أما تقييم منخفض جدا فقد كان من نصيب اختبار واحد فقط. وقد تم منح

التقييم منخفض ومنخفض جدا بسبب تدني قيمة ألفا كرونباخ غالبا. مثل الاختبار الذي تم منحه التقييم منخفض جدا في دراسة (Richard 2018) حيث قدرت قيمة ألفا كرونباخ بـ0.20.

أما صدق الاتساق الداخلي (العلاقة بين البنود والدرجة الكلية والعلاقة بين المحاور والدرجة الكلية) والذي هو في الحقيقة لا يظهر بينة الاختبار أو عدد العوامل، وإنما يكشف فقط عما إذا كانت البنود تقيس شيئا مشتركا، ورغم أن الباحثة لا تحبذ استخدامه كما سيتم التفصيل فيه في الجزء الخاص بمناقشة النتائج إلا أنها قامت بتقييمه، وقد التحقق من الاتساق الداخلي في 61 دراسة، 36 اختبار منها تم تقييمه في ثبات الاتساق الداخلي بمرتفع، و14 اختبار بمتوسط، في حين تم تقييم 8 اختبارات بتقييم منخفض، وأخيرا 3 اختبارات بتقييم منخفض جدا. وفي الغالب تم منح التقييم منخفض ومنخفض جدا لسببين هو انخفاض قيمة ألفا وأيضا لاستخدامها مع مقياس متعدد الأبعاد مثل دراسة (القصاص 2014) والذي استخدم ألفا للمقياس ككل رغم أنه يحتوي على أبعاد، وبالإضافة إلى ذلك قيمة ألفا كرونباخ كانت أصغر من 0.70 وهي القيمة التي حددتها قائمة معايير COSMIN.

وبالنسبة للدراسات التي ارتأت أنها استخدمت الصدق الهيكلي ولكنها لم تقم بأي إجراء من إجراءاته فقد وجدت دراسة واحدة فقط، والتي تم منحها التقييم I في البند الذي يكشف عن بنية العامل، ورغم ذلك كانت جودتها مرتفعة في ثبات الاتساق الداخلي، والسبب راجع لكون هذه الدراسة والتي هي دراسة (سلمان والعلي 2021) ذكرت في جزئية الصدق الهيكلي أنه تم تطبيق الاختبار والتحقق من وضوح فقراته، وكأن الإجراء هنا هو التحقق من صدق المحتوى عند العينة، الذي دفع بالباحثة إلى تقييمه بغير مناسب I، أما ألفا كرونباخ فقد تم حسابها لكل بعد على حدى، وقيمها كلها تجاوزت 0.70، ولذلك قيمت الجودة الكلية لها في هذا الاختبار بمرتفع.

1-6-عرض نتائج السؤال السادس: هل تم التحقق من صلاحية الاختبارات المنقولة ثقافيا باستخدام الخطوات المحددة؟

إنَ الصدق الثقافي كما ذكرنا سابقا في الفصل النظري، ليس مرتبط فقط بالاختبارات المترجمة بل هو مرتبط بخطر تحيز الاختبار، سواء لنقله من بيئة مختلفة، أو لتغيير مجتمع البحث أو غيرها من الأسباب، وعلى العموم فإن هذا الإجراء نادرا ما استخدم كما هو موضح في الجدول أسفله:

الجدول 15: صدق الاختبارات المنقولة ثقافيا

الصدق الثقافي	V	A	D	I	N
هل كانت العينات متشابهة في الخصائص الأخرى باستثناء المتغير المدروس؟	0	1	3	0	167
هل تم استخدام منهج مناسب لتحليل البيانات؟	0	1	2	1	167
هل كان حجم العينة المدرجة في التحليل كافيا؟	1	0	1	2	167
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	1	/	1	2	167

المصدر: من إعداد الباحثة

لقد تم التحقق من صدق الاختبارات المنقولة ثقافيا في 4 دراسات فقط، وسيتم عرض تقييمها حسب البنود كما يلي: البند الأول هل كانت العينات متشابهة في الخصائص الأخرى باستثناء المتغير المدروس؟ للأسف لم يتم منح أي اختبار تقييم V أي جيد جدا، في حين منح اختبار واحد فقط تقييم A. لتمنح الاختبارات الثلاثة الباقية تقييم مشكوك فيه D. وتم ذلك من خلال المعايير التالية: للحصول على التقييم جيد جدا **V** يجب أن تقدم الدراسة أدلة واضحة على أن العينات كانت متشابهة في الخصائص الأخرى باستثناء المتغير الذي تم تغييره. وللحصول على التقييم **A** فإن الدراسة تذكر أن العينات متشابهة لكن لم تقدم دليل واضح على ذلك. وفي ما يخص التقييم **D**

فإنه يتم منحه إذا كان من غير الواضح هل العينات كانت متشابهة في الخصائص الأخرى باستثناء المتغير الذي تم تغييره أم لا. وفي الأخير يتم منح التقييم **I** إذا كانت العينات غير متشابهة في الخصائص الأخرى باستثناء المتغير الذي تم تغييره.

وبالنسبة للبند الثاني هل تم استخدام منهج مناسب لتحليل البيانات؟ لم يتم تقييم أي اختبار من الاختبارات بتقييم جيد جدا في هذا البند V. في حين تم تقييم اختبار واحد فقط بتقييم A أي مناسب.أما التقييم مشكوك فيه فقد تم منحه لاختبارين أي التقييم D. وتم منح اختبار واحد تقييم I. وقد استندت هذه التقييمات إلى المعايير التالية: يتم منح التقييم **V** إذا تم استخدام الأساليب المناسبة مثل التحليل العاملي التوكيدي بالنسبة للنظرية التقليدية، أو استخدام نماذج الانحدار الترتيبي وقياس الثوابت، واستخدام تحليل الأداء التفاضلي للبند بالنسبة لنظرية الاستجابة للفقرة. ويتم تقييم الاختبار بـ **A** إذا افترض أن المنهج والأساليب كانت مناسبة لكن لم يتم وصفها بوضوح، أم التقييم **D** فيتم منحه إذا كان المنهج المستخدم غير واضح أو مشكوك في صحته. ليبقى التقييم **I** والذي يتم منحه إذا كان المنهج غير مناسب.

وبالنسبة للبند الثالث هل كان حجم العينة المدرجة في التحليل كافيا؟ فقد تحصل اختبار واحد على تقييم V. وتحصل اختبار واحد أيضا على تقييم D. في حين تحصل الاختبارين الباقيين على تقييم I. وذلك حسب هذه المعايير: V يتم منحه إذا كان حجم العينة يفوق حجم عدد البنود بـ7 أضعاف وأكبر أو يساوي 100 بالنسبة للتحليل العاملي، أما بالنسبة لتحليلات الانحدار والتحليلات القائمة على راش ونظرية الاستجابة للفقرة فإنها تتطلب حجم عينة 200 فرد لكل مجموعة. أما التقييم **A** فيتم منحه إذا كان حجم العينة 5 أضعاف عدد البنود وأكبر أو يساوي 100 أو من 5 إلى 7 أضعاف ولكن أصغر من 100 بالنسبة للتحليل العاملي، وبالنسبة لتحليلات الانحدار والتحليلات القائمة على راش ونظرية الاستجابة للفقرة فإنها تتطلب حجم عينة 150 فرد لكل

مجموعة. وبالنسبة للتقييم **D** فإنه يمنح إذا كان حجم العينة 5 ضعف عدد البنود وأكبر 100 بالنسبة للتحليل العاملي، أما بالنسبة لتحليلات الانحدار والتحليلات القائمة على راش ونظرية الاستجابة للفقرة فحجم العينة 100 فرد لكل مجموعة. وفي بالنسبة للتقييم **I** فإنه يمنح إذا كان حجم العينة أصغر من عدد البنود بأقل من 5 أضعاف وأقل من 100 بالنسبة للتحليل العاملي، أما بالنسبة لتحليلات الانحدار والتحليلات القائمة على راش ونظرية الاستجابة للفقرة ف حجم العينة في هذا التقييم أقل من 100 فرد لكل مجموعة.

أما البند الرابع والأخير في هذا المربع وهو: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ تم منح اختبار واحد تقييم V أي جيد جدا. واختبار واحد تقييم D أي مشكوك فيه. واختبارين تقييم I أي غير مناسب. وكانت معايير هذا البند هي: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا التقييم في هذا البند. **D** توجد عيوب لكن ثانوية. **I** توجد عيوب منهجية مهمة.

رغم أنه لم يتم التحقق من صدق الاختبار المنقول ثقافيا إلا في 4 دراسات، غير أن تلك الدراسات تحصلت على تقييمين جيدين جدا وتقييمين مناسبين، وعلى العموم فإن هذا النوع من الصدق ورغم أهميته إلا أنه لم يلقى نصيبه من الاستخدام أو حتى إلقاء الضوء عليه من طرف الباحثين، وربما هذا عائد في هذه الدراسة لكونها استبعدت الدراسات التي تقوم بالتكييف والتقنين.

1-7-عرض نتائج السؤال السابع: إلى أي مدى التزم الباحثون بالمعايير الموضوعة عند استخدامهم التطبيق وإعادة التطبيق وكذا الصور المتكافئة في التحقق من صلاحية الاختبار؟

للإجابة على هذا السؤال جمعت قائمة COSMIN بين طريقتين من طرق إعادة التحقق من ثبات الاختبار، والتي تشترك في نقطة رئيسية وهي وجود تطبيقين للاختبار، وتختلف في ماهية هذا الاختبار، بحيث تعتمد طريقة التطبيق وإعادة التطبيق على اختبار واحد يتم تطبيقه على مرتين، في حين تعتمد الصور المتكافئة على تطبيق اختبارين متكافئين في جميع الجوانب مرة واحدة (تطبيق الاختبار الأول وتطبيق الصورة المكافئة له وبالتالي أصبح لدينا تطبيقين). ولذلك فشروط تطبيق هاذين النوعين وحتى الأساليب الإحصائية المستخدمة هي نفسها، ماعدا إضافة شرط التكافؤ بالنسبة للنوع الثاني الخاص بالصور المتكافئة. ونتائج التزام الباحثين بهذه الشروط موضحة في الجدول:

الجدول 16: ثبات التطبيق وإعادة التطبيق

ثبات الإعادة	V	A	D	I	N
هل كانت العينة مستقرة في البناء المراد قياسه بين التطبيقين؟	8	9	33	0	121
هل كانت الفترة الزمنية مناسبة؟	40	/	10	0	121
هل كانت ظروف وصور الاختبار متماثلة في التطبيقين؟ على سبيل المثال: ظروف التطبيق، البيئة، التعليمات؟	2	7	41	0	121
بالنسبة للدرجات المستمرة: هل تم حساب معامل الارتباط داخل الصف ICC*؟	1	0	40	0	130
بالنسبة للدرجات ثنائية التفرع/ الاسمية/ الترتيبية: هل تم حساب كابا؟	0	/	/	8	163
بالنسبة للدرجات الترتيبية: هل تم حساب كابا المرجحة؟	0	/	1	0	170
بالنسبة للدرجات الترتيبية: هل تم وصف مخطط الترجيح؟ على سبيل المثال خطي تربيعي	0	1	/	/	170
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	35	/	15	0	121

المصدر: من إعداد الباحثة

لقد استخدم صدق التطبيق وإعادة التطبيق، والصور المتكافئة في 50 دراسة. وقبل عرض تكرارات كل إجراء سيتم بداية عرض النتائج حسب ترتيب البنود: البند الأول هل كانت العينة مستقرة في البناء المراد قياسه بين التطبيقين؟ تم منح 8 اختبارات تقييم V أي جيد جدا، و3 اختبارات تقييم مناسب A. ليأخذ التقييم D أعلى تكرار بـ33 اختبار تم تقييمه مشكوك فيه. وقد تم اعتماد المعايير التالية في التقييم: يتم الحصول على التقييم جيد جدا **V** إذا قدمت أدلة واضحة

* ICC Intarclass Correltion Coefficent معامل الارتباط داخل الصف

أن العينة كانت مستقرة. أما التقييم **A** فإنه يتم الحصول عليه إذا افترض أن العينة كانت مستقرة ولكن الأدلة لم تكن واضحة. وفي ما يخص التقييم **D** فإنه يتم منحه إذا كان من غير الواضح ما إذا كانت العينة مستقرة. في حين أن التقييم **I** يتم اعتماده إذا لم تكن العينة مستقرة. وبالتالي يتضح أن أغلبية الدراسات هنا أي 33 دراسة من أصل 50 دراسة لم يكن واضح هل كانت العينة مستقرة أو لا.

وفي ما يخص البند الثاني: هل كانت الفترة الزمنية مناسبة؟ تم تقييم 40 اختبار بتقييم V . أما 10 اختبارات الباقية فتحصلت على تقييم D. ولم يتم منح أي اختبار تقييم I. وقد استندت هذه التقييمات إلى المعايير التالية: يتم منح التقييم **V** إذا كان الفاصل الزمني مناسب، ويعتبر أسبوعين هو الفاصل الزمني الأكثر ملاءمة، ولكن تبقى الفترة الزمنية مرتبطة بالسمة المدروسة وعينة الدراسة. **A** لا يتم منح هذا التقييم في هذا البند. **D** مشكوك في ما إذا كان الفاصل الزمني مناسب أو لم يتم تحديد الفاصل الزمني. **I** الفاصل الزمني غير مناسب. كما لاحظنا من الجدول السابق أنه لم يتم منح الاختبار التقييم I، في أي دراسة أي أن الفاصل الزمني كان مناسب في أغلب الاختبارات، في حين أ، 10 اختبارات فقط كان تم تقييمها D وفي الغالب حصلت على هذا التقييم بسبب عدم ذكر الفترة بين التطبيقين.

وبالنسبة للبند الثالث هل كانت ظروف وصور الاختبار متماثلة في التطبيقين؟ على سبيل المثال: ظروف التطبيق، البيئة، التعليمات؟ تحصل اختبارين على تقييم جيد جدا V. وتحصلت 7 اختبارات على تقييم مناسب A . في حين تحصل 41 اختبار على تقييم D مشكوك فيه. وذلك حسب المعايير الآتية: للحصول على تقييم **V** ينبغي أن تكون الظروف متشابهة مع تقديم الأدلة الواضحة على ذلك أو في حالة وجود اختلاف ولكنه اختلاف مقصود كجزء من الدراسة. أما التقييم **A** فيشترط أيضا أن تكون ظروف الاختبار متشابهة. وبالنسبة للتقييم **D** فإنه يمنح في حالة لم يكن

واضحا إذا كانت ظروف الاختبار متشابهة أم لا. وفي الأخير للحكم على البند بالتقييم **I** فهذا يعني أن ظروف الاختبار لم تكن متشابهة. وعليه فكما هو واضح فإن أغلب الدراسات وبالضبط 41 دراسة لم يكن واضحا فيها إذ ما كانت الظروف متشابهة أم لا، وربما يرجع هذا الشح في التفاصيل إلى أن الباحثين يقومون بإعادة التحقق من صدق وثبات الاختبارات النفسية وليس بناءها وبالتالي ربما هذا دفع بالباحثين إلى التغاضي عن ذكر جميع هذه التفاصيل المهمة.

أما البند الرابع: بالنسبة للدرجات المستمرة: هل تم حساب معامل الارتباط داخل الصف ICC؟ تم تقييم اختبار واحد بجيد جدا V. و 40 اختبار الباقية بمشكوك فيه D. وهذا التقييم راجع للمعايير التالية: للحصول على تقييم **V** تم حساب ICC وصيغتها أو النموذج المتعلق بها موصوف. أما التقييم **A** فيتم منحه إذا تم حساب ICC لكن الصيغة أو النموذج غير موصوف، أو تم حساب بيرسون أو سبيرمان ولكن بشرط توفر الدليل على عدم وجود التغيير المنهجي و**D** تم حساب بيرسون أو سبيرمان بدون دليل على حدوث تغير منهجي أو مع حدوث تغير منهجي. ليبقى التقييم **I** الذي يمنح في حالة لم يتم حساب ICC ولم يتم حساب سبيرمان أو بيرسون. ولذلك نجد أن اختبار واحد فقط تحصل على تقييم جيد جدا لأن أغلبية الاختبارات تستخدم معامل الارتباط بيرسون في التطبيق وإعادة التطبيق، ولا تشير بأي طريقة إلى عدم وجود تغيير منهجي، وبالتالي صنف 40 اختبار في التقييم D لاعتمادهم على بيرسون بالنسبة للاختبارات ذات الدرجات المتصلة، أما 9 اختبارات الباقية فتوزعت على الدرجات الترتيبية والثنائية كما سيتضح في البندين التاليين.

في البند الخامس بالنسبة للدرجات ثنائية التفرع/ الاسمية/ الترتيبية: هل تم حساب كابا؟ فكانت النتائج بأنه 8 اختبارات تم تقييمها I. وذلك راجع لكون المعايير تمنح التقييم **V** في حالة تم حساب كابا كوهين. ولا يتم منح التقييم **A** و**D** في هذا البند. في حين أن التقييم **I** يتم منحه إذا لم

يتم حساب كابا كوهين. وبالتالي لم تستخدم أي دراسة معامل كابا كوهين رغم أنه الاختبارات 8 كانت ثنائية التفرع في خيارات الاستجابة.

وبالنسبة للبند السادس الذي كانت صياغته: بالنسبة للدرجات الترتيبية: هل تم حساب كابا المرجحة؟ فقد كانت النتائج بأنه تم تقييم اختبار واحد D. وفق المعايير التالية: يمنح التقييم **V** في حالة تم حساب كابا المرجحة. ولا يتم منح التقييم **A** و**I** في هذا البند. في حين أن التقييم **D** يتم منحه إذا تم حساب كابا الغير مرجح أو كابا غير موصوف. وبما أن الاختبار تم تقييمه D في هذا البند فيمكن للقارئ الاستنتاج أن الباحث لم يستخدم كابا المرجحة في التحقق من الارتباط بين التطبيقين، رغم أن درجات الاختبار ترتيبية.

وكتتمة للبند السادس جاءت الصيغة البند السابع كما يلي بالنسبة للدرجات الترتيبية: هل تم وصف مخطط الترجيح؟ على سبيل المثال خطي تربيعي؟ في هذا البند تم بطبيعة الحال تقييم اختبار واحد فقط وقد تم تقييمه بـ A. استنادا إلى هذا المعايير: يمنح التقييم **V** في حالة تم وصف مخطط الترجيح. في حين أن يمنح التقييم **A** إذا لم يتم وصف مخطط الترجيح. أما التقييمات **D** و**I** فلا تمنح في هذا البند. وبما أنه كما ذكرنا في البند السادس لم يتم استخدام كابا المرجحة فبطبيعة الحال لم يتم وصف مخطط الترجيح.

وأخيرا بالنسبة للبند الثامن: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ تم منح 35 اختبار تقييم V أي جيد جدا. ليبقى 15 اختبار من أصل 50 تم تقييمه D أي مشكوك فيه. وكانت معايير هذا البند هي: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا التقييم في هذا البند. **D** توجد عيوب لكن ثانوية. **I** توجد عيوب منهجية مهمة. غالبا ما كانت الباحثة تمنح التقييم D في هذا البند بسبب أن الارتباطات بين الاختبارين أقل من 0.70.

وكما ذكرنا في بداية هذا التعليق على هذا الجدول فإننا سنتطرق إلى جزئية التفصيل في التكرارات المتعلقة بكل إجراء على حدى، وعليه فقد تم استخدام التطبيق وإعادة التطبيق في التحقق من ثبات 49 اختبار، في حين تم استخدام الصور المتكافئة في التحقق من ثبات اختبار واحد فقط.

1-8- عرض نتائج السؤال الثامن: هل قام الباحثون بالكشف عن أخطاء القياس؟

رغم أنه لم يتم التحقق من أخطاء القياس في أي اختبار من اختبارات عينة الدراسة الحالية، إلا أن الباحثة قررت عرض الجدول الخاص بأخطاء القياس من أجل وصف وتوضيح المعايير التي وضعتها قائمة COSMIN عند التحقق والكشف عن أخطاء القياس.

الجدول 17: أخطاء القياس

أخطاء القياس	V	A	D	I	N
هل كانت العينة مستقرة في البناء المراد قياسه بين التطبيقين؟	0	0	0	0	171
هل كانت الفترة الزمنية بين التطبيقين مناسبة؟	0	/	0	0	171
هل كانت ظروف الاختبار متماثلة في التطبيقين؟	0	0	0	0	171
بالنسبة للمتغير المتصل: هل تم حساب الخطأ المعياري (SEM)* أو أصغر تغيير (SDC)** أو حدود الاتفاقية (LoA)***؟	0	0	/	0	171
بالنسبة للدرجات ثنائية التفرع / الاسمية / الترتيبية: هل تم حساب النسبة المئوية (الإيجابية والسلبية) للاتفاق؟	0	0	/	0	171
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	0	/	0	0	171

المصدر: من إعداد الباحثة

* SEM Standard Error of Measurement خطأ القياس المعياري
** SDC Smallest Detectable Change أصغر تغيير يمكن اكتشافه
*** LoA Limits of Agreement حدود الاتفاقية

بما أنه لم يتم التحقق من خطأ القياس في أي دراسة، فإن الباحثة انتقلت مباشرة إلى عرض معايير هذا المربع، والتي كما هو واضح من الجدول تشترك بنوده الثلاثة الأولى والبند السادس الأخير مع المربع السادس الذي يدرس التطبيق وإعادة التطبيق، لذلك فإن الباحثة لم تدرج معايير البند الأول والثاني والثالث والسادس هنا وللإطلاع عليها انظر التعليق أسفل (الجدول رقم 16). أما في ما يخص البند الرابع بالنسبة للمتغير المتصل: هل تم حساب الخطأ المعياري في القياس (SEM) أو أصغر تباين يمكن اكتشافه (SDC) أو حدود الاتفاقية (LoA)؟ فيتم الحصول على التقييم **V** إذا تم حساب (SEM) أو (SDC) أو (LoA). ويتم تقييم البند بـ **A** في حالة افتراض أنه تم حساب (LoA) من خلال البيانات المقدمة. أما التقييم **D** فلا يتم منحه في هذا البند. في حين أن التقييم **I** فيتم منحه في حالة تم حساب (SEM) على أساس ألفا كرونباخ أو على أساس الانحراف المعياري من عينة أخرى.

وتم تقييم البند الخامس والذي كانت صياغته كما يلي: بالنسبة للدرجات ثنائية التفرع / الاسمية / الترتيبية: هل تم حساب النسبة المئوية (الإيجابية والسلبية) للاتفاق؟ ومعاييره كالتالي: لمنح البند التقييم **V** يجب أن يتم حساب النسبة الايجابية والسلبية للاتفاق. أما التقييم **A** فيتم منحه في حالة تم حساب نسبة الاتفاق فقط. وبالنسبة للتقييم **D** فهو غير معتمد في هذا البند. ليبقى التقييم **I** فيتم منحه في حالة تم لم يتم حساب نسبة الاتفاق.

1-9-عرض نتائج السؤال التاسع: ما مدى استخدام الباحثين للصدق المحكي والتزامهم بخطوات التحقق منه؟

رغم أن العديد من الباحثين ذكروا في دراساتهم أنهم تحققوا من الصدق المحكي إلا أنه بالتدقيق في الإجراءات التي تم استخدامها والاختبار الذي تم اعتماده كمحك، نجد أن الباحثين

أجروا الصدق التقاربي وليس الصدق المحكي والذي يشترط وجود اختبار ذهبي، لذلك نجد أن هذه الطريقة بإجراءاتها السليمة قليلة الاستعمال كما سيتضح من خلال الجدول التالي:

الجدول 18: تقييم الصدق المحكي

الصدق المحكي	V	A	D	I	N
بالنسبة للدرجات المتصلة: هل تم حساب الارتباطات أو (AUC)* المنطقة الواقعة أسفل منحنى تشغيل المستقبل؟	2	/	/	0	169
بالنسبة للدرجات ثنائية التفرع: هل تم تحديد الحساسية والنوعية؟	0	/	/	0	171
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	1	/	1	0	169

المصدر: من إعداد الباحثة

كما هو واضح من خلال الجدول السابق دراستين فقط من أصل 171 دراسة استخدمت الصدق المحكي في التحقق من صلاحية الاختبار النفسي، وتم تقييم كل من الدراستين بـ V في ما يخص البند الأول: بالنسبة للدرجات المتصلة هل تم حساب الارتباطات أو (AUC) المنطقة الواقعة أسفل منحنى تشغيل المستقبل؟ استنادا إلى المعايير التالية: يتم تقييم هذا البند بـ**V** في حالة تم حساب معامل الارتباط بين اختبارين كلاهما متغيران متصلان، ويتم حساب (AUC) في حالة كانت درجات الاختبار متصلة ودرجات المحك ثنائية التفرع. أما التقييم **A** والتقييم **D** فلا يتم اعتمادهم في تقييم هذا البند، وفي الأخير يتم منح التقييم **I** إذا لم يتم حساب معامل الارتباط أو (AUC).

أما البند الثاني الذي يهتم بالدرجات ثنائية التفرع والذي كانت صياغته: هل تم تحديد الحساسية والنوعية؟ فلم حساب الحساسية والنوعية لأي من الدراستين، أما معايير هذا البند هي:

* AUC Area Under the Curve المنطقة الواقعة أسفل المنحنى

بـ**V** تم حساب الحساسية والنوعية. أما التقييم **A** والتقييم **D** فلا يتم اعتمادهم في تقييم هذا البند، وفي الأخير يتم منح التقييم **I** إذا لم يتم حساب معامل الارتباط أو (AUC).

وفي ما يخص البند الأخير هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ فاختبار واحد تم تقييمه V. والاختبار الثاني تم تقييمه D. وبالنسبة لمعاييره فهي مثل المعايير الموجودة لنفس البند في الجداول السابق وهي: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا التقييم في هذا البند. **D** توجد عيوب منهجية طفيفة. **I** توجد عيوب منهجية مهمة. ولقد تم منح التقييم D في هذا البند بسبب قيمة معامل الارتباط التي تقل عن 0.70. وهي القيمة التي حددتها قائمة COSMIN كما سيتضح في الجداول المتعلقة بالجودة الكلية لكل طريقة من طرق الصدق والثبات.

1-10-عرض نتائج السؤال العاشر: هل تم التحقق من صدق الفرضيات بالشكل المطلوب حسب معايير COSMIN؟

للإجابة على هذا السؤال قسمت قائمة COSMIN صدق الفرضيات إلى جزأين الأول متعلق بمقارنة الاختبار بأدوات أخرى، والثاني متعلق بالمقارنة بين المجموعات، ولذلك سيتم تقسيم هذا السؤال إلى سؤالين فرعيين.

1-10-1-السؤال الفرعي الأول: هل تم الالتزام بالإجراءات اللازمة للتحقق من الصدق التقاربي والتباعدي؟

في هذا الجدول يتم تقييم الاختبارات التي اعتمدت على التحقق من الصدق باستخدام اختبارات تقيس نفس السمة، أو سمة مشابهة، أو سمة مغايرة، وتختلف المعايير حسب السمة التي يقيسها الاختبار المستخدم للمقارنة، كما سيتضح من الجدول التالي:

الجدول 19: تقييم الصدق التقاربي والتباعدي

الصدق التقاربي والتباعدي	V	A	D	I	N
هل من الواضح ما تقيسه أداة (أدوات) المقارنة؟	16	/	/	0	155
هل كانت الخصائص السيكومترية الخاصة بأداة (أدوات) المقارنة كافية؟	5	1	8	2	155
هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟	12	4	0	0	155
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	8	/	8	0	155

المصدر: من إعداد الباحثة

لقد تم استخدام إجراء الصدق التقاربي في التحقق من صلاحية 16 اختبار، غير أن 5 دراسات فضلت إضافة إجراء الصدق التباعدي معه، أي من بين 16 دراسة قامت بالتحقق فقط من الصدق التقاربي 5 دراسات قامت بالتحقق من الصدق التقاربي والتباعدي معا مثل دراسة (Treaster 2021).

ولتقييم مدى التزام الباحثين بالإجراءات اللازمة سيتم عرض وتحليل كل بند على حدى، وبالنسبة للبند الأول والذي تمت صياغته كما يلي: هل من الواضح ما تقيسه أداة (أدوات) المقارنة؟ تم تقييم كل الاختبارات التي تم التحقق من صدقها بهذه الطريقة بتقييم جيد جدا، بحيث تم منح V للاختبارات 16، وذلك لأن المعايير الخاصة بهذا البند تقول بأنه يتم الحصول على التقييم جيد جدا **V** إذا كان البناء الذي تقيسه أداة المقارنة واضح وملائم. أما التقييم **A** و **D** فإنه غير معتمد في هذا البند. وفي ما يخص التقييم **I** فإنه يتم منحه إذا كان البناء الذي تقيسه أداة المقارنة غير واضح أو غير ملائم.

وغير بعيد عن نفس فكرة تقييم أدوات المقارنة، تمت صياغة البند الثاني كالتالي: هل كانت الخصائص السيكومترية الخاصة بأداة (أدوات) المقارنة كافية؟ وقد تم تقييم 5 اختبارات بتقييم V.

واختبار واحد بتقييم A. في حين تحصلت 8 اختبارات على تقييم D. في حين تحصل اختبارين على تقييم I. وقد استندت هذه التقييمات إلى المعايير التالية: يتم منح التقييم **V** إذا كانت الخصائص السيكومترية لأداة المقارنة كافية وجيدة في مجتمع مشابه لمجتمع الدراسة الحالي. **A** إذا كانت الخصائص السيكومترية لأداة المقارنة كافية ولكن ليس واضح إذا المجتمع مشابه لمجتمع الدراسة الحالي. **D** بعض المعلومات عن الخصائص السيكومترية في مجتمع ما. **I** لا توجد معلومات عن الخصائص السيكومترية لأداة المقارنة أو أنها ضعيفة وغير كافية.

وبالنسبة للبند الثالث هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟ فقد تحصل 12 اختبار على تقييم V جيد جدا. وتحصلت 4 اختبارات على تقييم A مناسب. وذلك على أساس هذه المعايير: للحصول على تقييم **V** فإن الأساليب الإحصائية المطبقة اعتبرت مناسبة، مثل استخدام معامل الارتباط بيرسون مع توزيع الدرجات أو متوسط الدرجات. أما التقييم **A** فيفترض فيه أن الأساليب الإحصائية المطبقة مناسبة. وبالنسبة للتقييم **D** فإنه يمنح في حالة اعتبار أن الأساليب الإحصائية المطبقة لا تعتبر الأمثل. وفي الأخير للحكم على البند بالتقييم **I** فإن الأساليب الإحصائية المطبقة اعتبرت ليست مناسبة.

وأخيرا تم تقييم البند الرابع الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ بـ8 اختبارات تقييم V أي جيد جدا. و8 اختبارات تقييم D أي مشكوك فيه. بالاعتماد على المعايير التالية: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا التقييم في هذا البند. **D** توجد عيوب لكن ثانوية. **I** توجد عيوب منهجية مهمة.

1-10-2-السؤال الفرعي الثاني: إلى أي درجة تم تتبع الخطوات المطلوبة من قبل الباحثين للتحقق من الصدق التمييزي بين المجموعات؟

اعتمدت الدراسات في هذا الإجراء على تقييم صدق المقارنة الطرفية غالبا، رغم أن الصدق التمييزي في الحقيقة أعم من ذلك فهو لا يقتصر إجراء المقارنة بين الفئة الدنيا والعليا، كما معروف وشائع بل يضم أي إجراء للمقارنة بين المجموعات مثل المقارنة بين عينة من العاديين والمرضى في سمة ما، كما سنوضحه في الجدول التالي:

الجدول 20: تقييم الصدق التمييزي بين المجموعات

الصدق التمييزي	V	A	D	I	N
هل تم تقديم وصف مناسب للخصائص الهامة للمجموعات الفرعية؟	5	8	7	/	151
هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟	7	12	1	0	151
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة ؟	11	/	6	3	151

المصدر: من إعداد الباحثة

رغم أن هذا الإجراء لا يقتصر فقط على المقارنة الطرفية، بل هو أشمل من ذلك، بحيث يشمل جميع الإجراءات التي تقارن بين المجموعات، إلا أن إجراء المقارنة الطرفية هو الأكثر شيوعا رغم أنه الأقل فعالية في الكشف عن صدق الاختبار، ومن أصل 20 دراسة استخدمت الصدق التمييزي بين المجموعات 18 دراسة اعتمدت على المقارنة الطرفية في حين دراسة واحدة اعتمدت على الفرق بين مجموعتين (عاديين ومرضى)، لتبقى دراسة واحدة ذكرت أنها قامت بدراسة الفرق بين مجموعتين لكن لم تذكر أي تفاصيل أخرى يمكن من خلالها تحديد ما هو الإجراء المستعمل بالضبط. ولتقييم الإجراءات المستخدمة في عينة الدراسة سيتم بداية بعرض نتائج البند

الأول: هل تم تقديم وصف مناسب للخصائص الهامة للمجموعات الفرعية؟ تم تقييم 5 اختبارات بتقييم V أي جيد جدا، و8 اختبارات بتقييم مناسب A. و7 اختبارات بتقييم مشكوك فيه D. وقد تم اعتماد المعايير التالية في التقييم: يتم الحصول على التقييم جيد جدا **V** إذا قدمت الدراسات وصف واضح للخصائص الهامة للمجموعات الفرعية. أما التقييم **A** فإنه يمنح في حالة اعتبر الوصف مقبول إلى حد ما لأغلب الخصائص الهامة للمجموعة الفرعية. وفي ما يخص التقييم **D** فإنه يتم منحه إذا اعتبر الوصف ضعيف. في حين أن التقييم **I** لم يتم اعتماده في هذا البند.

وفي ما يخص البند الثاني: هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟ تم تقييم 7 اختبارات بتقييم جيد جدا V . و12 اختبار بتقييم A. في حين تحصل اختبار واحد على تقييم D. ولم يتم منح أي اختبار تقييم I. وقد استندت هذه التقييمات إلى المعايير التالية: يتم منح التقييم **V** إذا كانت الأساليب الإحصائية المطبقة مناسبة، مثل استخدام اختبار ت لدراسة الفروق وغيرها لكن يجب تغاضي النظر عن الدلالة الإحصائية هنا لأن لا معنى لها. أما التقييم **A** فيفترض فيه أن الأساليب الإحصائية المطبقة مناسبة. وبالنسبة للتقييم **D** فإنه يمنح في حالة اعتبار أن الأساليب الإحصائية المطبقة لا تعتبر الأمثل. وفي الأخير للحكم على البند بالتقييم **I** فإن الأساليب الإحصائية المطبقة اعتبرت ليست مناسبة.

وفي الأخير فإن تقييمات البند الثالث هنا الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ فقد وزعت كما يلي: تم منح 11 اختبار تقييم V أي جيد جدا. و6 اختبارات تقييم D أي مشكوك فيه. و 3 اختبارات تقييم I أي غير مناسب. وكانت معايير هذا البند هي: **V** لا توجد أي عيوب مهمة. **A** لا يتم اعتماد هذا التقييم في هذا البند. **D** توجد عيوب لكن ثانوية. **I** توجد عيوب منهجية مهمة. وعادة ما يتم تقييم الاختبارات

في هذا البند بـ D أو I لأن الباحثين يهملون ذكر التفاصيل المهمة التي تساعد على التقييم، أو بسبب قيم معاملات الفروق أو الارتباط التي تكون منخفضة.

1-11-عرض نتائج السؤال الحادي عشر: هل تم تقييم الاستجابة للاختبارات النفسية بالطريقة الصحيحة حسب المعايير التي وضعتها COSMIN؟

للإجابة على هذا السؤال تم تقسيم الجزء المتعلق بالاستجابة إلى أربعة أقسام وتحوي على نفس الأسئلة أو البنود التي تم طرحها في الجزء المتعلق بالصدق المحكي وصدق الفرضيات بقسيمه (مقارنة الاختبار بأدوات أخرى، المقارنة بين المجموعات الفرعية)، بالإضافة إلى بنود تقيس اختبار الفرضيات حول الحجم المتوقع للتدخل، والفرق الوحيد هو أن الاستجابة-كما ذكرنا في هذا الفصل أعلاه-تدرس صحة التغيير، وبالتالي تستند الاستجابة دائما على فرضيات التوقع، مثلا اختبار الفرضيات حول الاختلافات المتوقعة في التغييرات بين المجموعات الفرعية، أو اختبار الفرضيات بين درجات التغيير في الاختبار قيد المراجعة والتغيير في درجات أدوات المقارنة. أي أن الباحث دائما عند تحققه من صلاحية الاستجابة يصيغ فرضيات مبنية على التوقع، وبالتالي عند تطبيق الاختبار قيد المراجعة إما يتحقق توقعه وبالتالي الاختبار قادر على كشف التغيير، أو لا يتحقق توقعه وبالتالي الاختبار غير قادر على كشف التغيير.

رغم أنه لم يتم تقييم الاستجابة لأي اختبار في أي دراسة من دراسات العينة 171 فقد قامت الباحثة بعرض الأسئلة الفرعية الأربعة التي يتم تقييم الاستجابة من خلالها، وذلك من أجل عرض المعايير الخاصة بها.

1-11-1-السؤال الفرعي الأول: هل تم الالتزام بالإجراءات اللازمة عند توقع نتائج مقارنة الاختبار قيد المراجعة مع المحك؟

كما سيتضح من الجدول فإن البنود هنا هي نفسها البنود التي تقيم الصدق المحكي، والفرق الوحيد هو في صياغة فرضيات، كما هو واضح من خلال الجدول التالي:

الجدول 21: تقييم استجابة المحك

استجابة المحك	V	A	D	I	N
بالنسبة للدرجات المستمرة: هل تم حساب الارتباطات أو المنطقة الواقعة أسفل منحنى تشغيل المستقبل؟	0	/	/	0	171
بالنسبة للدرجات ثنائية التفرع: هل تم تحديد الحساسية والنوعية؟	0	/	/	0	171
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	0	/	0	0	171

المصدر: من إعداد الباحثة

في ما يخص البند الأول: بالنسبة للدرجات المتصلة هل تم حساب الارتباطات أو (AUC) المنطقة الواقعة أسفل منحنى تشغيل المستقبل؟ استنادا إلى المعايير التالية: يتم تقييم هذا البند بـ**V** في حالة تم صياغة فرضية يتوقع فيها حجم الارتباط ومن ثم حساب معامل الارتباط بين اختبارين كلاهما متغيران متصلان، ويتم حساب (AUC) في حالة كانت درجات الاختبار متصلة ودرجات المحك ثنائية التفرع، مع فرضية التوقع دائما. أما التقييم **A** والتقييم **D** فلا يتم اعتمادهم في تقييم هذا البند، وفي الأخير يتم منح التقييم **I** إذا لم يتم حساب معامل الارتباط أو (AUC) أو لم يتم صياغة الفرضية بالشكل المطلوب.

أما البند الثاني الذي يهتم بالدرجات ثنائية التفرع والذي كانت صياغته: هل تم تحديد الحساسية والنوعية؟ فمعاييره هي: بـ**V** تم صياغة فرضية التوقع مع حساب الحساسية والنوعية. أما التقييم **A** والتقييم **D** فلا يتم اعتمادهم في تقييم هذا البند، وفي الأخير يتم منح التقييم **I** إذا لم يتم صياغة فرضية التوقع بالشكل المطلوب أو لم تم حساب معامل الارتباط أو (AUC).

وفي ما يخص البند الأخير هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ فمعياره هو نفس المعيار الموجود أسفل الجدول رقم (18).

1-11-2-السؤال الفرعي الثاني: هل تم الالتزام بالإجراءات اللازمة عند توقع نتائج مقارنة الاختبار قيد المراجعة مع اختبار آخر (يقيس نفس السمة أو سمة مختلفة عنه)؟

مثل ما ذكر في الجدول فإن بنود هذا الجدول هي نفس بنود الجدول رقم (19) إلا أن الباحثة قررت عرضها من أجل إدراج المعايير الخاصة بها.

الجدول 22: تقييم الاستجابة مع الاختبار التقاربي أو التباعدي

الاستجابة مع الاختبار التقاربي أو التباعدي	V	A	D	I	N
هل من الواضح ما تقيسه أداة (أدوات) المقارنة؟	0	/	/	0	171
هل كانت خصائص القياس الخاصة بأداة (أدوات) المقارنة كافية؟	0	0	0	0	171
هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟	0	0	0	0	171
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	0	/	0	0	171

المصدر: من إعداد الباحثة

قبل عرض المعايير وجب الإشارة إلى أنها نفس معايير الجدول رقم (19) فقط يختلف هنا شرط صياغة فرضية التوقع. أي أن الباحث في الاستجابة يتوقع مآل اختباره، وبالتالي يهدف إلى التحقق من صحة توقعه. بالنسبة لمعايير البند الأول والذي تمت صياغته كما يلي: هل من الواضح ما تقيسه أداة (أدوات) المقارنة؟ فهي كالتالي: يتم الحصول على التقييم جيد جدا **V** إذا كان البناء الذي تقيسه أداة المقارنة واضح وملائم لفرضيات البحث. أما التقييم **A** و **D** فإنه غير معتمد في هذا البند. وفي ما يخص التقييم **I** فإنه يتم منحه إذا كان البناء الذي تقيسه أداة المقارنة غير واضح أو غير ملائم للفرضيات.

أما البند الثاني: هل كانت الخصائص السيكومترية الخاصة بأداة (أدوات) المقارنة كافية؟ يستند إلى المعايير التالية: يتم منح التقييم **V** إذا كانت الخصائص السيكومترية لأداة المقارنة كافية وجيدة في مجتمع مشابه لمجتمع الدراسة الحالي. **A** إذا كانت الخصائص السيكومترية لأداة المقارنة كافية ولكن ليس واضح إذا المجتمع مشابه لمجتمع الدراسة الحالي. **D** بعض المعلومات عن الخصائص السيكومترية في مجتمع ما. **I** لا توجد معلومات عن الخصائص السيكومترية لأداة المقارنة أو أنها ضعيفة وغير كافية.

وبالنسبة للبند الثالث هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟ وتتمثل معاييره في ما يلي: للحصول على تقييم **V** فإن الأساليب الإحصائية المطبقة اعتبرت مناسبة للفرضيات المصاغة، مثل استخدام معامل الارتباط بيرسون مع توزيع الدرجات أو متوسط الدرجات. أما التقييم **A** فيفترض فيه أن الأساليب الإحصائية المطبقة مناسبة للفرضيات. وبالنسبة للتقييم **D** فإنه يمنح في حالة اعتبار أن الأساليب الإحصائية المطبقة ليست الأمثل للفرضيات. وفي الأخير للحكم على البند بالتقييم **I** فإن الأساليب الإحصائية المطبقة اعتبرت ليست مناسبة للفرضيات.

وأخيرا تم تقييم البند الرابع الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية ؟ فمعياره هو نفس المعيار الموجود أسفل الجدول رقم (19)

1-11-3-السؤال الفرعي الثالث: هل تم الالتزام بالإجراءات اللازمة في دراسة الاستجابة التمييزية؟

كما سبق وذكرت الباحثة فإن الجدول والمعايير هنا هي نفسها مع الجدول رقم(20).

الجدول 23: تقييم الاستجابة التمييزية

الاستجابة التمييزية	V	A	D	I	N
هل تم تقديم وصف مناسب للخصائص الهامة للمجموعات الفرعية؟	0	0	0	/	171
هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟	0	0	0	0	171
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	0	/	0	0	171

المصدر: من إعداد الباحثة

في ما يخص معايير البند الأول: هل تم تقديم وصف مناسب للخصائص الهامة للمجموعات الفرعية؟ هي نفسها معايير الجدول رقم (20)

وفي ما يخص البند الثاني: هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟ فإن معاييره هي نفسها بالنسبة للجدول رقم (20)، فقط تم إضافة فرضيات التوقع.

وفي الأخير فإن تقييمات البند الثالث هنا الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ معاييره أيضا هي نفسها معايير الجدول رقم (20).

1-11-4-السؤال الفرعي الرابع: هل تم اختبار فرضيات التدخل بشكل صحيح؟

في هذا الجدول سيتم عرض بنود جديدة خاصة فقط بمربع الاستجابة وهي كما يلي:

الجدول 24: استجابة التدخل

استجابة التدخل	V	A	D	I	N
هل تم تقديم وصف مناسب للتدخل؟	0	/	0	0	171
هل كان التصميم والأساليب الإحصائية مناسبة لاختبار الفرضيات؟	0	0	0	0	171
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	0	/	0	0	171

المصدر: من إعداد الباحثة

يتم تقييم البند الأول والذي محتواه: هل تم تقديم وصف مناسب للتدخل؟ حسب المعايير التالية: يمنح إذا تم وصف التدخل بالتفصيل وبوضوح. أما التقييم **A** فهو غير معتمد عليه في هذا البند. وبالنسبة للتقييم **D** فإنه يمنح في حالة اعتبار أن الوصف سيء. ويتم تقييم البند بـ **I** إذا لم يتم وصف التدخل إطلاقا.

وفي البند الثاني هل كان التصميم والأساليب الإحصائية مناسبة لاختبار الفرضيات؟ فإن معاييره هي **V** إذا كانت الأساليب الإحصائية المطبقة مناسبة، مثل استخدام متوسط درجة التغيير ومعامل نورمان للاستجابة. أما التقييم **A** فيفترض فيه أن الأساليب الإحصائية المطبقة مناسبة للفرضيات. وبالنسبة للتقييم **D** فإنه يمنح في حالة اعتبار أن الأساليب الإحصائية المطبقة ليست الأمثل للفرضيات. وفي الأخير للحكم على البند بالتقييم **I** فإن الأساليب الإحصائية المطبقة اعتبرت ليست مناسبة للفرضيات مثل الانحراف المعياري أو معامل MIC الذي يقيم الحد الأدنى من التغيير.

أما البند الثالث الذي كانت صياغته: هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟ فمعاييره أيضا هي نفسها معايير الجدول رقم (18).

1-12-عرض نتائج السؤال الثاني عشر: هل كانت الخصائص السيكومترية التي تم استخدامها في إعادة التحقق من صلاحية الاختبارات النفسية ذات جودة مقبولة؟

من أجل الإجابة على هذا السؤال قامت الباحثة بتجزئته إلى عدة أسئلة فرعية، بحيث يقيم كل سؤال فرعي جودة خاصية محددة حسب قائمة COSMIN، في الص المرتبطة بتقييم الجودة. (أنظر الملحق01)، وبالطبع تم تقييم جودة الإجراءات التي تم استخدامها في هذه الدراسة.

1-12-1-السؤال الفرعي الأول: ما مدى جودة إعادة التحقق من صدق المحتوى؟

سيتم في هذا الجدول عرض جودة إعادة التحقق من صدق المحتوى عند العينة والمحكمين مع بعض، حسب معايير الجودة التي اقترحتها قائمة COSMIN، وكانت النتائج كالتالي:

الجدول 25: جودة صدق المحتوى

جودة صدق المحتوى	التكرارات	النسبة المئوية
منخفض جدا	12	%7
منخفض	20	%11.7
متوسط	19	%11.1
مرتفع	11	%6.4
N	109	%63.7
الكلية	171	%100

المصدر: من إعداد الباحثة

من خلال الجدول السابق نستنتج أن العدد الإجمالي للدراسات التي تحققت من صدق المحتوى سواء بعرضه على المحكمين أو بعرضه على العينة قد بلغ 62 دراسة، وكما ذكرنا سابقا

57 منها تحققت من صدق المحتوى عند المحكمين فقط، و3 دراسات تحققت من صدق المحتوى عند العينة فقط، في حين أن دراستين فضلتا التحقق من صدق المحتوى عند المحكمين والعينة وهو ما تنصح به مبادرة COSMIN. وعلى العموم فقد تم تقييم 32 اختبار بتقييم سلبي، بحيث 20 اختبار تم تصنيفه بجودة منخفضة في هذا الإجراء و12 اختبار بتصنيف جودته بمنخفضة جدا، في حين أن 19 كانت جودته متوسطة، ليبقى في الأخير 11 اختبار صنف ذو جودة عالية في هذا الإجراء. لم يكن إجراء صدق المحتوى في المستوى المطلوب من حيث الجودة، رغم أن الباحثة حاولت تخفيف صرامة معاييره إلى حد ما وجعلها مرنة قدر الإمكان فمثلا يشترط في صدق المحتوى بجميع فروعه أن يتشارك باحثين في تحليل البيانات، وهو الأمر الذي يصعب توفره خاصة في رسائل الماجستير والدكتوراه، الأمر الذي دفع بالباحثة بالتعامل مع البند كأنه إجراء مفقود ومنحه التقييم N، أي أنه لم يطبق، بدل منحه D والذي يمنح في حالة مشاركة باحث واحد فقط في تحليل البيانات. كذلك لتقييم صدق المحتوى تم جمع جودة كل من صدق الملائمة الشمولية والفهم بالنسبة للعينة، الأمر الذي خفض من جودة صدق المحتوى في كثير من الأحيان لأن صدق الشمولية غالبا ما كان منخفض الجودة مقارنة بصدق الملائمة. وفي الأخير سيتم التفصيل أكثر في هذه النقطة وغيرها في الجانب المتعلق بمناقشة النتائج.

1-12-2-السؤال الفرعي الثاني: ما مدى جودة إعادة التحقق من الصدق الهيكلي في عينة الدراسة الحالية؟

يتم في هذا الجدول عرض الجودة الكلية للصدق الهيكلي في عينة الدراسة الحالية:

الجدول 26: جودة الصدق الهيكلي

جودة الصدق الهيكلي	التكرارات	النسبة المئوية
منخفض جدا	13	%7.6
منخفض	30	%17.5
متوسط	35	%20.5
مرتفع	19	%11.1
N	74	%43.3
الكلية	171	%100

المصدر: من إعداد الباحثة

كما يتضح من خلال الجدول أعلاه فإن النسبة الأكبر من الدراسات التي استخدمت الصدق الهيكلي، تم تصنيف جودة إعادة التحقق فيها، في خانة المتوسط بنسبة 20.5% أي بمقدار 35 دراسة من أصل 97 دراسة. ليحتل الرتبة بعدها تصنيف منخفض بمقدار 30 دراسة، ليأتي بعدهما تصنيف مرتفع الذي بـ19 دراسة، وأخيرا منخفض جدا بـ13 دراسة. وقد تم تصنيف الجودة بالاعتماد على عدة معايير أهمها الالتزام بالشروط والمعايير المذكورة سابقا في (الجدول رقم 12). بالإضافة إلى بعض المعايير الأخرى مثل أن يكون التحليل العاملي التوكيدي والاستكشافي أكبر من 0.95، و RMSEA* أصغر من 0.06 وSRMR** أصغر من 0.08. كذلك عدم انتهاك الاستقلال المحلي أي الارتباط المتبقي بين العناصر بعد ضبط العامل المهيمن يكون أصغر من 0.20. وبالنسبة للاتساق الداخلي أكبر من 0.70. ولهذا السبب غالبا انخفضت جودة الصدق

* Root Mean Square Error of Approximation خطأ جذر متوسط التربيع للتقريب

** SRMR Standardized Root Mean Residues متوسط الجذر المعياري للبقايا

الهيكلي لأن العديد من الدراسات لم تلتزم بهذه المعايير وتبقي على البنود الضعيفة جدا، مما يضعف من جودة الاختبار كما سيتم التفصيل فيه في مناقشة النتائج.

1-12-3-السؤال الفرعي الثالث: ما مدى جودة إعادة التحقق من ثبات الاتساق الداخلي في عينة الدراسة الحالية؟

سيتم في هذا الجدول الكشف عن جودة ثبات الاتساق الداخلي، مع ذكر المعايير المعتمدة في التصنيف:

الجدول 27: جودة ثبات الاتساق الداخلي

جودة ثبات الاتساق الداخلي	التكرارات	النسبة المئوية
منخفض جدا	5	%2.9
منخفض	24	%14
متوسط	27	%15.8
مرتفع	96	%56.1
N	19	%11.1
الكلية	171	%100

المصدر: من إعداد الباحثة

على عكس الجدول السابق فإن تصنيف الاتساق الداخلي أو بالضبط ألفا كرونباخ إن صح التعبير لأنه الأكثر استخداما في التحقق من ثبات الاتساق الداخلي في 149 اختبار من أصل 152، قد تم الحكم على جودته بأنه مرتفع في 96 اختبار، وهو أمر يدعوا إلى الاطمئنان في حالة التزام الباحثين بالتحقق من الصدق الهيكلي والتأكد بأن الاختبار أحادي البعد كما ذكرنا سابقا، ومن الجيد أيضا أن التصنيف الذي يلي التصنيف مرتفع من حيث أكبر عدد للدراسات هو التصنيف متوسط بـ27 اختبار، ليليه 24 اختبار تم تصنيف جودته بمنخفض، ليبقى في الأخير 5 اختبارات تم تصنيف جودتهم بمنخفض جدا. وقد صنفت الباحثة الاختبارات في هذا التصنيف انطلاقا من

المعايير التالية والتي وضعتها COSMIN: أولا وجود أدلة كافية (أقل درجة هي منخفض) في الصدق الهيكلي، ثانيا قيمة كل من ألفا والمعادلات التصحيحية وغيرها أكبر أو يساوي 0.70 لكل مقياس أحادي البعد، أو لكل مقياس فرعي. وهو الأمر الذي استوفته معظم الاختبارات والذي رفع من قيمة تصنيفها، رغم أن العديد من الدراسات لم تلتزم بأحادية البعد كما سيتم الاستفاضة فيه في الجزء المتعلق بمناقشة النتائج.

1-12-4-السؤال الفرعي الرابع: ما مدى جودة التحقق من الصدق الثقافي في عينة الدراسة الحالية؟

رغم أن عدد الدراسات التي تحققت من جودة الصدق الثقافي قليل مقارنة بالدراسات التي قامت بترجمة الاختبار ونقله من بيئة مختلفة، إلا أن جودة هذا المربع عموما كانت كما يلي:

الجدول 28: جودة الصدق الثقافي

جودة الصدق الثقافي	التكرارات	النسبة المئوية
منخفض جدا	2	%1.2
منخفض	1	%0.6
متوسط	1	%0.6
مرتفع	0	%0
N	167	%97.7
الكلية	171	%100

المصدر: من إعداد الباحثة

كما هو واضح من الجدول لم يتحصل ولا اختبار على تصنيف مرتفع، فقد توزعت الاختبارات الأربع كما يلي: اختبارين تم تصنيف جودتهم بمنخفض جدا، واختبار واحد تم تصنيفه بمنخفض، أما آخر اختبار فتم تصنيفه بمتوسط. ولقد حازت الاختبارات على هذا التصنيف، بسبب أن جودة الاختبار في هذا الإجراء مرتبطة بشكل رئيسي بالأساليب الإحصائية الواجب تطبيقها في حالة

تغيير البيئة الأصلية التي بني الاختبار من أجلها بأي طريقة كانت، وتتمثل هذه الأساليب في: التحليل العاملي الاستكشافي والتوكيدي بالنسبة للنظرية التقليدية، أو استخدام نماذج الانحدار الترتيبي وقياس الثوابت، واستخدام تحليل الأداء التفاضلي للبند بالنسبة لنظرية الاستجابة للفقرة، وللأسف لم يتم استخدام أي من هذه الأساليب في التحقق من الصدق الثقافي، ما عاد في دراسة (عيسى 2013) التي استخدمت الصدق العاملي الاستكشافي وقت منح الاختبار تصنيف متوسط لأن قيم التشبعات كانت جد منخفضة.

1-12-5-السؤال الفرعي الخامس: ما مدى جودة التحقق من ثبات التطبيق وإعادة التطبيق في عينة الدراسة الحالية؟

تراوحت جودة هذا الإجراء وتمايزت من اختبار لآخر كما هو وموضح في الجدول أسفله:

الجدول 29: جودة التطبيق وإعادة التطبيق

جودة إعادة التطبيق	التكرارات	النسبة المئوية
منخفض جدا	5	%2.9
منخفض	20	%11.7
متوسط	22	%12.9
مرتفع	3	%1.8
N	121	%70.8
الكلية	171	%100

المصدر: من إعداد الباحثة

لقد تم استخدام هذا الإجراء في 59 دراسة، 49 تم التحقق منها باستخدام التطبيق وإعادة التطبيق، وواحدة تم استخدام إجراء تطبيق صورتين متكافئتين، وصنفت جودة الاختبارات المستعملة في هذه الدراسات كما يلي: 22 اختبار كأعلى نسبة تم منحه تصنيف متوسط من حيث الجودة، يليها 20 اختبار اعتبرت جودته منخفضة، و5 اختبارات منخفضة جدا، في حين أن 3 اختبارات

فقط تم تصنفيها بمرتفع. وهذا راجع لكون معايير الجودة التي وضعتها قائمة COSMIN لم تحترم في أغلب الدراسات، مثل عدم استخدام معامل الارتباط داخل الصف ICC الذي أكدت عليه القائمة مرارا وتكرارا، كذلك معامل كابا بالنسبة للمتغيرات الثنائية والترتيبية، أيضا قيم كل منهما اشترطت القائمة أن تكون أكبر أو تساوي 0.70 حتى وإن استخدم معامل الارتباط بيرسون أو سبيرمان فإن هذا الشرط مازال قائما، ولكن في الحقيقة لم تتقيد معظم الدراسات بهذه الشروط مما جعل تقييم الجودة لهذا المربع يتراوح في غالبيته بين المتوسط والمنخفض في حين فقط 3 دراسات صنف مستوى جودتها بمرتفع. وسيتم التوسع في مناقشة هذه الفكرة أكثر عند مناقشة النتائج مع الاستدلال ببعض الدراسات.

1-12-6-السؤال الفرعي السادس: ما مدى جودة التحقق من الصدق المحكي في عينة الدراسة الحالية؟

لقد سبق وذكرت الباحثة عند عرضها للإجراءات المستخدمة في التحقق من صلاحية الاختبارات النفسية في التعليق تحت (الجدول رقم05)، وكذلك في الجزء المتعلق بالسؤال المطروح حول الالتزام بإجراءات الصدق المحكي، السبب وراء ندرة استخدام هذا النوع من الصدق في التحقق من صلاحية الاختبار، فكما هو موضح في الجدول التالي لم يتم استخدام هذا النوع إلا في دراستين وصنفت جودتهما كما يلي:

الجدول 30: جودة الصدق المحكي

جودة الصدق المحي	التكرارات	النسبة المئوية
منخفض جدا	0	%0
منخفض	0	%0
متوسط	1	%0.6
مرتفع	1	%0.6
N	169	%98.8
الكلية	171	%100

المصدر: من إعداد الباحثة

لقد تم تصنيف جودة الصدق المحكي في إعادة التحقق من صدق اختبار أنماط السيادة النصفية في دراسة (روبي 2016) بمتوسط بسبب أن معايير قائمة COSMIN تشترط أن العلاقة بين الاختبار و بين المحك والذي يعتبر في هذه الحالة هو النسخة المستحدثة لنفس الاختبار الذي يراد التحقق من صدقه، أن تكون أكبر من 0.70 أو تساويها ونفس الشرط في حالة تم استخدام المنطقة الواقعة أسفل المحنى، وهو الشرط الذي قصرت فيه هذه الدراسة بحيث أن العلاقة قدرت بـ0.43 في البعد الثالث و0.59 في البند الثاني أي أنها تجاوزت 0.70 في بعد واحد فقط، مما دفع بالباحثة بتقييم جودتها بمتوسط لأن باقي الشروط المهمة توفرت أهمها خاصية المعيار الذهبي وحجم العينة أكثر من 50. أما الاختبار الثاني فتم تقييم جودته بمرتفع لأنه استوفى جميع الشروط مثل الاختبار الذهبي وحساب المنطقة الواقعة أسفل المنحنى وحجم العينة قدر 147.

1-12-7-السؤال الفرعي السابع: ما مدى جودة التحقق من الصدق التقاربي والتباعدي لعينة الدراسة الحالي؟

يتضح من الجدول التالي أن جودة التحقق من الصدق التقاربي والتباعدي تراوحت بين المنخفضة والمرتفعة، كما سيتم التفصيل فيه في الجدول أسفله:

الجدول 31: جودة الصدق التقاربي والتباعدي

جودة الصدق التقاربي والتباعدي	التكرارات	النسبة المئوية
منخفض جدا	0	%0
منخفض	2	%1.2
متوسط	9	%5.8
مرتفع	5	%2.9
N	155	%90.6
الكلية	171	%100

المصدر: من إعداد الباحثة

من خلال الجدول يتضح لنا أن في تصنيفات الجودة لهذا الإجراء لم يتم تصنيف ولا اختبار بمنخفض جدا وهو أمر محمود بطبيعة الحال، وحتى الاختبارات التي قيمت بمنخفض فعددها قليل بحيث فقط اختبارين من أصل 16 اختبار تم تصنيفها في خانة منخفضة الجودة، أما الاختبارات المتوسطة الجودة من حيث الصدق التقاربي فقد قدر عددها بـ9 اختبارات وهي النسبة الأعلى في هذا الإجراء، لتبقى 5 اختبارات تم تصنيفها مرتفع. وهي تقييمات لا بأس بها في الحقيقة، بحيث أن 14 اختبار (9متوسط و5 مرتفع) تعتبر جودة التحقق من صدقه التقاربي مقبول في حين اختبارين فقط تعتبر جودتهم ضعيفة. وذلك وفق للمعايير التي أضافتها قائمة COSMIN في تقييم الجودة الكلية لهذا الإجراء ألا وهي: يجب أن تكون الارتباطات مع الأدوات التي تقيس بنى متشابهة أكبر أو يساوي 0.50. و يجب أن تكون الارتباطات مع الأدوات التي تقيس بنية لديها

علاقة مع البنية المراد دراستها ولكن ليست هي نفسها من 0.30 إلى 0.50. في حين أن الارتباطات مع الأدوات التي تقيس سمة مختلفة يجب أن تكون أصغر من 0.30. ويسمح بتجاوز هذه المعايير بـ0.10 كأقصى تقدير. وكما هو واضح من النتائج السابقة فقد تم الالتزام بهذه المعايير غالبا إلا في بعض الدراسات.

1-12-8-السؤال الفرعي الثامن: ما مدى جودة التحقق من صدق التمييزي في عينة الدراسة الحالية؟

رغم قلة عدد الدراسات التي استخدمت هذا الإجراء في التحقق من الصدق إلا أن جودة الاختبارات التي استخدم فيها تنوعت كما هو واضح في الجدول التالي:

الجدول 32: جودة الصدق التمييزي

جودة الصدق التمييزي	التكرارات	النسبة المئوية
منخفض جدا	1	%0.6
منخفض	4	%1.8
متوسط	8	%4.7
مرتفع	7	%4.1
N	152	%88.9
الكلية	171	%100

المصدر: من إعداد الباحثة

تقاربت الاختبارات من حيث العدد في التصنيفين متوسط ومرتفع، فقد تحصلت 8 اختبارات على تصنيف متوسط الجودة، في حين تحصلت 7 اختبارات على تصنيف مرتفع الجودة، أما منخفض الجودة فقد قيمت به 4 اختبارات، ليبقى اختبار واحد فقط تم تقييمه بمنخفض الجودة. رغم أنه تمت إضافة معيار واحد فقط بالإضافة إلى المعايير السابقة الذكر لهذا الإجراء وهو أن

التغييرات أو الفروق بين المجموعتين يجب أن تكون ذات معنى. إلا أن السبب في تصنيف الاختبار بمنخفض جدا في هذا الإجراء راجع إلى عدة أسباب منها عدم ذكر يحدد بالضبط الأسلوب الإحصائي المستخدم، كذلك من غير الواضح هل تم استخدام المقارنة الطرفية أو الفروق بين المجموعات وغيرها من العيوب التي جعلته في خانة المنخفض جدا.

2-مناقشة النتائج

إن مجال صلاحية الاختبارات النفسية يصعب فيه تحديد أهداف وتحقيقها وفقط، فكل هدف وكل تساؤل كان يأخذنا إلى تساؤل آخر، وهو ما يمكن ملاحظته من خلال عدد الأسئلة الفرعية التي تم إدراجها في هذه الدراسة فقد وصل عددها الكلي إلى 22 سؤال فرعي. ولقد عرضت الباحثة نتائج الدراسة من خلال احترام الترتيب الذي فرضته COSMIN في قائمتها لذلك نجد في عرض النتائج جدول يدرس الصدق ليله جدول آخر يدرس الثبات ثم نعود إلى الصدق مرة أخرى إلى آخره، الأمر الذي ربما يبدو للقارئ أنه خلط في بداية الأمر، لكن في الحقيقة هذا الترتيب مدروس ومقصود، وعلى الباحث الذي يقبل على التحقق أو إعادة التحقق من صدق وثبات الاختبارات النفسية الالتزام به أيضا، فأول خطوة يقوم بها الباحث هي التحقق من خطوات بناء الاختبار وهو يعتبر المربع الأول في قائمة COSMIN، تليه خطوة التحقق من صدق محتواه بجميع إجراءاته (الصدق الظاهري، صدق المحتوى للعينة وللمحكمين) والذي يمثل المربع الثاني في القائمة، لتأتي الخطوة الثالثة وهدفها التحقق من بينة الاختبار في العينة المقصودة وذلك من خلال الصدق الهيكلي في المربع الثالث، ليقوم بعدها الباحث بعد أن تحقق من أحادية البعد في الاختبار بالتحقق من ثبات اتساقه الداخلي والذي يعتبر المربع الرابع في القائمة، لتأتي بعدها خطوة التحقق من صدقه الثقافي إن لزم الأمر طبعا في المربع الخامس، أما المربع السادس فهو يقيم صدق الإعادة والصور المتكافئة، ليأتي بعد كل ذلك المربع السابع والذي يهدف إلى تقييم خطأ القياس، وفي نفس التسلسل المنطقي وبعد اعتماد الباحث على الخطوات التي يراها مناسبة وضرورية من الخطوات السابقة يصل إلى مرحلة تقييم بناء الاختبار انطلاقا من فرضياته الخاصة من خلال المربع

التاسع، ليبقى في الأخير المربع العاشر والذي يدرس الاستجابة ومآل الاختبار وبالتالي من الطبيعي أن يكون هو آخر إجراء وآخر مربع يمكن الاعتماد عليه.

بداية وقبل الانطلاق في مناقشة النتائج بالاعتماد على الترتيب السابق الذكر لقائمة COSMIN، فضلت الباحثة طرح سؤال يمدنا بنظرة عامة عن الإجراءات الأكثر استخداما في إعادة التحقق من صلاحية الاختبارات النفسية قبل تقييمها، من خلال طرحها السؤال التالي والذي يمثل **السؤال الأول في هذه الدراسة وهو: ما هي الطرق الأكثر استخداما من طرف الباحثين في إعادة تحققهم من صدق وثبات الاختبارات النفسية في عينة من الدراسات الأكاديمية العربية والأجنبية؟** وتوصلت النتائج إلى أن ثبات الاتساق الداخلي هو الطريقة الأكثر استخداما، وبالضبط يعتبر ألفا كرونباخ هو الإجراء الأكثر استخداما في هذا النوع من الثبات بحيث استخدم في التحقق من ثبات 149 اختبار. وهو ما اتفقت معه دراسة كل من (Polit, 2015) و (Tian, Cao, & Feng, 2019) وذلك راجع حسبهما لكون هذا الإجراء يعتبر سهلا مقارنة بباقي إجراءات كالتطبيق وإعادة التطبيق، أو بناء نسختين لنفس الاختبار وغيرها من الإجراءات. أما ثاني أكثر نوع استخدامًا فهو الصدق الهيكلي أو ما يعرف بصدق الاتساق الداخلي والذي استخدم في التحقق من صلاحية 97 اختبار، وهو ما توصلت إليه أيضا دراسة (بودلال و موراس، 2021) إلا أنها لم تتفق مع الدراسة الحالية من حيث المرتبة الأولى فقد توصلت نتائجها إلى أن صدق المحكمين هو الأكثر استخداما. في حين احتل في الدراسة الحالية المركز الثالث من حيث الاستخدام وهو ما يختلف أيضا مع دراسة كل من (بن صافي و طاجين، 2016) و(بوسالم، د س) في حين اتفق مع دراسة (Newton, Soleimani, Kirkland, & Gokiert, 2017). وترجع الباحثة ذلك لكونها لم تفصل بين الصدق والثبات في هذا الترتيب على غرار الدراسات الأجنبية، في حين أن الدراسات العربية التي سبق ذكرها قد فصلت بين الصدق والثبات وتم حساب تكرار كل نوع على حدى.

كذلك غير بعيد عن صدق المحكمين يجب الإشارة إلى أن صدق المحكمين هو إجراء فقط من إجراءات صدق المحتوى شأنه شأن الصدق الظاهري، وحتى وإن كان هذا الإجراء مهم إلا أنه يبقى غير كاف، بالإضافة إلى ذلك يقع الباحثون في العديد من التجاوزات الأخرى عند استخدام صدق المحتوى، وهو ما بدى جليا عند عرض نتائج **السؤال الثاني: هل صدق المحتوى الذي تحقق منه هؤلاء الباحثين تم بطرق سليمة موافقة لمعاييرCOSMIN؟** والذي بدوره تفرع إلى خمس أسئلة تقيس (الملائمة، الشمولية والفهم عند العينة، والملائمة والشمولية عند المحكمين) من خلال العديد من البنود، وكانت النتائج في مجملها متراوحة بين V، A، وD أي بين جيد جدا، مناسب، ومنخفض. وغالبا كانت الملائمة والشمولية والفهم عند العينة أحسن في تقييمها من الملائمة والشمولية عند المحكمين رغم قلة الدراسات التي اعتمدت العينة في الحكم على صدق المحتوى بحيث لم تتجاوز 5 دراسات، ويرجع انخفاض جودة المحتوى عند المحكمين لكون الشمولية لم يتم تقييمها بالشكل المناسب فـ41 اختبار من أصل 59 تم تقييمه بمشكوك فيه، كذلك الأمر الذي خفض من جودة صدق المحتوى عند المحكمين هو المنهج المستخدم سواء في الملائمة أو الشمولية بحيث تحصلت كلاهما على تقييم مشكوك فيه أو غير كاف بـ 32 و37 تكرار على الترتيب. وهو ما أكدته أيضا نتائج دراسة (Ashra, Barnes, Stupple, & Maratos, 2021) بحيث توصلت إلى نفس التقييم مشكوك فيه لهذا البند. لقد لاحظت الباحثة أن التقييم المرتفع في هذا المربع كان من نصيب البند الذي يقيم عدد المحكمين أو حتى المفحوصين بالنسبة للعينة، فقد تم تقييم هذا البند في صدق المحكمين بـV 42، وفي صدق المحتوى للعينة تم تقييم البند بـV أي جيد جدا للدراسات 5 كلها. وبالتالي يعتبر هذا أكثر معيار تم احترامه من قبل الباحثين، وهو ما يتنافى مع ما ذكره (السردي، 2012) في نتائج بحثه بحيث توصل إلى أن الاعتماد على عدد قليل من المحكمين يعتبر من أهم الأخطاء التي وقع فيها الباحثين في هذا النوع من الصدق.

أماَ في ما يخص نتائج **السؤال الثالث: ما نسبة الدراسات التي تحققت من الصدق الهيكلي للاختبار النفسي بالشكل المطلوب؟** كما تمت الإشارة إليه سابقا عند وصف أداة الدراسة الحالية، وأيضا أسفل الجدول رقم (11)، فإن قائمة COSMIN لم تدرج الاتساق الداخلي (الارتباط بين البنود والمحور أو بين المحاور والدرجة الكلية)، ربما ذلك راجع لكونه لا يكشف عن أحادية البعد أو عن عدد العوامل، أو ربما راجع لما ذكرته (أنستازي و أوربينا، 2015) بأنه "لا يمكن الخروج بدلالات حول صدق ما يقيسه الاختبار من خلال المعلومات المتعلقة بالارتباط بين البنود والدرجة الكلية". ورغم ذلك لم يكن بإمكان الباحثة تجاهله أو عدم تقييمه إطلاقا، لما له من استخدام واسع في الدراسات، فقد تم استخدامه في التحقق من صدق 70 اختبار، وكل هذه الدراسات عربية ما عدا 5 دراسات أجنبية (3 من الولايات المتحدة و1 من كندا والأخيرة من ألبانيا). ولذلك بعد تشاور الباحثة مع المشرف قررت إضافته وإدراجه في نفس المرتبة مع التحليل العاملي الاستكشافي. وقد اعتمدت الباحثة في وضع معاييره على مقالة (Terwee, et al., 2007). وكما يتضح من خلال البند الأول للجدول رقم (11) فإن الاتساق الداخلي هو الإجراء الأكثر استخداما في هذا النوع من الصدق بحيث تحصل 70 اختبار على تقييم A، في حين أن التحليل العاملي الاستكشافي تم استخدامه في 8 دراسات وتم تقييمه في جميع تلك الدراسات بـ V بدل A كما هو مذكور في المعايير سابقا، لكونه استخدم في حالة تغيير بنية الاختبار من إضافة وحذف بنود مثل دراسة (Yi lee 2012)، أو ترجمة الاختبار وتغيير البيئة الأصلية له مثل دراسة (عمرو يوسف 2017) وبالتالي تم تقييمه V، أما لو تم استخدامه من أجل إعادة التحقق فقط بدون أي تعديلات فإن تصنيفه يكون A. ليبقى 17 اختبار من أصل 25 تم تقييمه أيضا بـV اعتمد في التحقق من صدقه على التحليل العاملي التوكيدي، والذي يعتبر الخيار الأفضل للغرض المستخدم من أجله بحيث أن هدف الباحثين هو إعادة التحقق من صدق وثبات الاختبار وليس البناء. وبالتالي يعتبر التحليل

العاملي هو الخيار المفضل إلا في حالة تم تغيير بنية الاختبار وترجمته كم ذكرنا سابقا فإن التحليل العاملي الاستكشافي يصبح هو الخيار الأول مدعما بعدها بالتوكيدي، ليبقى دائما الاتساق الداخلي (العلاقة بين البنود والمحور أو بين المحاور والدرجة الكلية) هو الخيار الأخير والذي يستلزم معه استخدام صدق المحتوى كأحسن خيار، أو أي إجراء من الإجراءات الأخرى للتحقق من الصدق، لأن هذا النوع كما ذكرنا في الجزء النظري لا يكشف لنا عما يقيسه الاختبار إنما فقط يثبت أن الاختبار يقيس شيئا مشتركا. وتوصلت الباحثة أيضا في النتائج الخاصة بهذا المربع، أن بعض الباحثين ارتكبوا العديد من التجاوزات أهمها ضعف معاملات الارتباط والذي كان السبب وراء تقييم البند الخاص بالعيوب والذي جاء بصيغته التالية: هل كان هناك أي عيوب أخرى مهمة في طريقة الاستخدام أو الأساليب الإحصائية الأخرى؟ وقد تحصل 75 اختبار على تقييمات بين D و I أي مشكوك فيه وغير مناسب، وغالبا بسبب قيم معاملات الارتباط والتشبعات والتباينات وغيرها، وقد لاحظت الباحثة أن بعض الدراسات اعتمدت على معيار 0.10 كحد أدنى لمعامل الارتباط بمعنى أن البنود التي يكون ارتباطها بالبعد أقل من هذا المعيار هي التي تحذف في حين أن البند الذي تصل قيمته لـ0.11 هو بند صالح. ولم تصل الباحثة إلى أصل هذا المعيار المستخدم رغم البحث الموسع عن مصدره، لذلك تعتقد أنه ربما هو خطأ شائع، ومن بين الدراسات التي اعتمدت هذا المعيار دراسة علي (2012)، في حين نجد بعض الدراسات مثل دراسة (الدود 2013) و(بوخاري 2017) قد قررت الاحتفاظ بجميع البنود بغض النظر عن قيمتها ودلالتها، فقد وصلت قيم بعض البنود إلى 0.01 وبالطبع هي غير دالة، إلا أن الباحثين قررا الاحتفاظ بها دون أي تبرير كأهمية البند مثلا. وقد توصلت العديد من الدراسات السابقة إلى نفس النتائج منها دراسة (Irene, Luisa, Luca, & Alvisa, 2017).

وفي الأخير فإن الهدف الرئيسي من استخدام الصدق الهيكلي هو الكشف والتحقق من بنية الاختبار، ومن ثم التعامل معه على أساس النتائج، فمثلا يعتبر ألفا كرونباخ غير مناسب لمقياس متعدد الأبعاد وبالتالي، يتم تشجيع الباحثين الذين يقومون بتقييم ثبات الاتساق الداخلي للاختبارات على استخدام البدائل، مثل ألفا كرونباخ الطبقية كما اقترح (تيغزة، 2017) أو أوميغا ماكدونالدز كما اقترح كل (Peters, 2014, p. 61) و (Mokkink, et al., 2017, p. 1174). وللكشف عن مدى التزام الباحثين بهذا الأمر تم الإجابة عن السؤال **الخامس: هل تحققت الدراسات التي استخدمت ثبات الاتساق الداخلي من أحادية البعد قبل استخدامه؟** وهو الأمر الذي أكدت عليه مبادرة COSMIN مرارا في دراستها، وذلك لما يطرحه ألفا كرونباخ من نتائج مزيفة عند استخدامه مع اختبارات متعددة الأبعاد، فكما تم الإشارة إليه سابقا في الجزء النظري فإن ألفا كرونباخ يتأثر بعدد البنود، كذلك الهدف منه هو دراسة الاتساق الداخلي بين البنود وبالتالي استخدامه مع اختبارات متعددة الأبعاد يناقض الأساس الذي وضع من أجله، ولذلك اقترحت COSMIN التحقق من أحادية البعد أولا، ويتم ذلك من خلال التحليل العاملي الاستكشافي والتوكيدي، وكما كشفنا عنه في مناقشة في البند الأول من المربع السابق، فإن عدد الدراسات التي تحققت من الصدق الهيكلي باستخدام هذه الإجراءات هو 25 دراسة 8 منها استخدمت تحليل عامل استكشافي و17 تحليل عاملي توكيدي، وقد تقاطع هذا البند مع جدول ثبات الاتساق الداخلي في 23 دراسة، أي أن 23 اختبار تم التحقق من أحادية البعد فيه قبل استخدام ثبات الاتساق الداخلي، وقد تم تقييم 17 اختبار منها بمرتفع، في حين صنفت 3 اختبارات بمتوسط واختبارين بمنخفض وفي الأخير اختبار واحد بمنخفض جدا. وعليه فإن 23 اختبار فقط من أصل 152 اختبار تم التحقق من أحادية البعد فيه قبل استخدام ثبات الاتساق، وعلى الرغم من أن هذا عكس توصيات مبادرة COSMIN، إلا أن الباحثة ترى بأنه لم يشكل خطأ منهجيا في عينة الدراسة الحالية، بما أن

هذه العينة تتكون من الدراسات التي أعادت التحقق من صلاحية الاختبارات، وبالتالي في أغلب الأحيان اعتمدت تلك الدراسات على الأبعاد الموجودة في الدراسات السابقة أو من الدراسات الأصلية للاختبار، في حين يعتبر خطأ منهجيا إذ ما تعديل الاختبار والتغيير في بنيته، ثم اللجوء إلى الاتساق دون التحقق من الصدق الهيكلي، وهو ما تمت مناقشته عند محاولة الإجابة عن التساؤل **السؤال الرابع: هل تم التحقق من ثبات الاتساق الداخلي للاختبار بجميع أشكاله باستخدام الإجراءات الصحيحة؟** وكما ذكرنا سابقا فقد توصلت نتائج هذه الدراسة إلى أن معامل ألفا كرونباخ هو الإجراء الأكثر استخداما (انظر الجدول رقم 13)، خاصة في الدراسات الأجنبية التي نادرا ما استغنت عن ألفا كرونباخ، فمن أصل 84 دراسة أجنبية فقط 4 دراسات لم تستخدم ألفا كرونباخ، في حين أن 22 دراسة عربية من أصل 87 دراسة فضلت عدم استخدامه. وفي نفس المربع الخاص بثبات الاتساق الداخلي تم استخدام التجزئة النصفية في التحقق من ثبات 25 اختبار، وقد صاحبه ألفا كرونباخ في 19 اختبار، في حين أن كيودرريتشاردسون تم استخدامه في دراستين فقط، أما أوميغا ماكدونالدز فلم يتم استخدامه في التحقق من ثبات أي اختبار، رغم أن 33 دراسة تم تقييمها ب I و D في البند الأول الذي يكشف عن ما إذا تم استخدام الاتساق الداخلي مع اختبار أحادي البعد أو لا؟ (انظر الجدول رقم 14)، إلا أن 26 دراسة منها استخدمت ألفا كرونباخ مع اختبارات متعددة الأبعاد بدل أوميغا ماكدونالدز أو ألفا الطبقية كما ذكرنا سابقا. و7 دراسات الأخرى الباقية من 33 دراسة، فضلت 6 منها استعمال التجزئة النصفية مع اختبار متعدد الأبعاد، لتبقى آخر دراسة فضلت استخدام كيودرريتشارسون.

وقد لاحظت الباحثة عند تحليل البيانات إشكالا مهما وهو أن 44 دراسة اعتمدت على ألفا كرونباخ وفقط في التحقق صلاحية الاختبار، وكلها دراسات أجنبية، ولم تجد الباحثة إجابة لهذا الإشكال سوى بعض التخمينات أو التوقعات التي استنتجتها من خلال تبريرات بعض الباحثين

عند تحققهم من الصدق والثبات معا، مثل دراسة (Zeko 2019) التي ذكرت عند إعادة تحققها من الصدق، بأنها قامت بذلك لكون الاختبار طبق على مجتمع جديد غير الذي بني من أجله، كذلك برر (Strimaitis 2018) إعادة تحققه من الصدق بحذفه لبعض البنود، وفي نفس السياق أشار (Viola 2012) عند استخدامه للصدق العاملي بأنه يتناول صلاحية هذا المقياس لأنه لم يجد بحث سابق يعزز به صلاحية المقياس، وهو بالضبط ما لاحظته الباحثة عند الدراسات التي لم تقم بالتحقق من الصدق، فقد اكتفت فقط بألفا كرونباخ وقامت بدعم دراستها بعرض الصدق والثبات للمقياس الأصلي، وأحيانا دعمه ببعض الدراسات الأخرى التي قامت بالتحقق من صلاحية هذا الاختبار، ويأخذنا هذا التحليل إلى منحى جديد وجب مناقشته وهو إذا افترضنا أن الباحثين انطلقوا من فكرة أن الاختبار صادق وصالح وبالتالي ليسوا بحاجة إلى إعادة التحقق منه، فلماذا يعاد التحقق من الثبات؟ ولما تستخدم دائما ألفا كرونباخ بدل الإجراءات الأخرى؟ للأسف لم تجد الباحثة إجابة لهذا السؤال. فحتى لو انطلقنا من افتراض أن الباحثين قاموا بإعادة التحقق فقط كإجراء روتيني أو منهجي، وبأنهم جازمين بصلاحية المقياس لاستخدامهم الحالي، فإننا نجد من يناقض هذا الافتراض فمثلا دراسة (Hancerogla 2015) قامت بإعادة التحقق من صلاحية الاختبار باستخدام ألفا كرونباخ وفقط، رغم أنها غيرت مجتمع الدراسة بحيث أن الاختبار صمم للبالغين إلا أنها استخدمته للأطفال بعد تعديل صياغة البنود طبعا. وأيضا دراسة كل من (Zhanget al 2021) و(Petal et al 2012) الذين قاموا بترجمة الاختبار والاكتفاء أيضا بألفا كرونباخ كدليل على صلاحيته. وفي الأخير فإن الباحثة تستنكر هذا الأمر، ويوافقها في ذلك (McAndrews, Richardson, & Stopa, 2019) الذين توصلت دراستهم إلى أن ألفا كرونباخ هي الأكثر استخداما وأن جودتها مقبولة إلا أنهم استهجنوا إجراءاها منفردة. فكما هو متعارف علي الصدق يوجب الثبات وليس العكس، وحتى

هذا لا يعني الاكتفاء بالصدق فكلاهما ضروري، فالصدق يدل على أن الاختبار يقيس ما وضع لقياسه في حين أن الثبات يدل على دقة الاختبار ونسبة الثقة في نتائجه، وبالتالي وجب التحقق منها معا. بالإضافة إلى التحقق من عدم تحيز البنود ومنه الاختبار، خاصة للاختبارات المنقولة ثقافيا والذي التمسنا غيابه في عينة الدراسة الحالية وذلك في محاولتنا الإجابة على **السؤال السادس: هل تم التحقق من صلاحية الاختبارات المنقولة ثقافيا باستخدام الخطوات اللازمة؟** فكما هو موضح في الجدول رقم (15) فقد تم التحقق من صدق الاختبارات المنقولة ثقافيا في 4 دراسات فقط، رغم أن عدد الدراسات التي نقلت الاختبار ثقافيا–اعتمادا على تعريف COSMIN– أكثر من هذا العدد، فالدراسات التي قامت بترجمة الاختبارات فقط قد قدرت بـ13 دراسة، بالإضافة إلى بعض الدراسات التي غيرت مجتمع الدراسة بأكمله مثل دراسة (Hancerogla 2015)، والدراسات التي أضافت العديد من البنود مثل دراسة (نشوف في جدول المواصفات)، إلا أن معظم هذه الدراسات كما ذكرنا سابقا اكتفت فقط بألفا كرونباخ، وقد بررت الباحثة ندرة استخدام هذا النوع من الصدق في البداية لكونها استبعدت الاختبارات التي يتم بناؤها وتكييفها من عينة الدراسة، إلا أنها تداركت أن الأمر لا علاقة له بذلك من خلال الاطلاع على العديد من الدراسات السابقة والتي أغلبها توصلت إلى نفس النتائج مع أنها لم تستبعد الاختبارات التي يتم تكييفها وبنائها، ومن هذه الدراسات (Cassidy, Bradley, Bowen, Wigham, & Rodgers, 2018)

(Saini, Hoffmann, Pantelis, Everall, & Bousman, 2019). في حين أن دراسة (Stallwood, et al., 2021) توصلت إلى أنه في 65 دراسة لـ6 اختبارات نفسية تقيس الاكتئاب، لم يتم تقييم الصدق الثقافي في أي دراسة منها. وهو أيضا ما أيدته دراسة (Wittkowski, Vatter, Muhinyi, Garrett, & Henderson, 2020) والتي كشفت أن العديد من الدراسات تهدف إلى تكييف مقياس معين مع مجموعات عرقية ولغوية مختلفة وبعضها

تضمن الترجمة الخلفية والإجراءات الضرورية الأخرى، إلا أنه لم يقم أي منها بتقييم الصدق الثقافي من خلال مقارنة مجموعات متعددة حسب تحليل العوامل، أو اختبار أداء العنصر التفاضلي.

أما في ما يخص نتائج السؤال السابع: **إلى أي مدى التزم الباحثون بالمعايير الموضوعة عند استخدامهم التطبيق وإعادة التطبيق وكذا الصور المتكافئة في التحقق من صلاحية الاختبار؟** فقد أظهرت نتائج المربع السادس بأنه تم استخدام صدق التطبيق وإعادة التطبيق في 49 دراسة في حين استخدمت الصور المتكافئة في دراسة واحدة فقط، وقد أرجعت الباحثة ذلك لكون الباحثين أصلا ليسوا هم من قاموا ببناء الاختبار المستخدم في الدراسة فكيف سيتم بناء نسخة مكافئة من أجل التحقق من صدق الاختبار، لذلك استخدم هذا النوع في دراسة واحدة فقط وهي دراسة (حنا 2019) والذي استخدم صورتين متكافئتين لنفس الاختبار، أي أن الاختبار أصلا يحتوي على صورتين متكافئتين أ و ب وقد استخدم الباحث في دراسته صورة واحدة أما الصورة الثانية فقد قام باستخدامها فقط من أجل التحقق من الصدق. وكما هو واضح من الجدول رقم (16) فإن الفترة الزمنية كانت مناسبة بحيث تم تقييمها في 40 دراسة بجيد جدا، على عكس استقرار العينة والظروف التي لم تكن هناك معلومات كافية حولها، وتعتقد الباحثة أن هذا الشح في التفاصيل راجع إلى أن الباحثين يقومون بإعادة التحقق من صدق وثبات الاختبارات النفسية، وليس بناءها وبالتالي ربما هذا دفع بالباحثين إلى التغاضي عن ذكر جميع هذه التفاصيل المهمة. مما جعل التصنيف في الغالب D أي مشكوك فيه، وقد اتفقت دراسة (Wittkowski, Vatter, Muhinyi, Garrett, & Henderson, 2020) مع نتائج هذه الدراسة في جزئية معامل الارتباط داخل الصف الذي لم يستخدم إلا مرة واحدة في هذه الدراسة مما منح هذا البند التقييم D في 40

دراسة من أصل 41 دراسة كانت نتائجها ذات متغيرات متصلة، ولم يتم تقديم أي دلائل على غياب الخطأ المنهجي لاستخدام معامل الارتباط بيرسون.

وغير بعيد عن معامل الارتباط، والذي يعتبر معيار مهم في الإجابة على **السؤال التاسع: ما مدى استخدام الباحثين للصدق المحكي والتزامهم بخطوات التحقق منه؟** وكما هو موضح في الجدول رقم (18) فإن دراستين فقط من استخدمت هذا النوع من الصدق في التحقق من صلاحية الاختبارات النفسية، وقد تقييم كلا الاختبارين بـV في البند الأول، في حين قيم اختبار واحد بـV والآخر بـD في البند الثالث. ولم يتم تقييم البند الثاني لكون ولا اختبار منهما ثاني التفرع. ولقد لاحظت الباحثة أن العديد من الدراسات مثل (Georgievaa, Tomasa, & Navarro–Pérez, 2021) و(Zeilinger, et al., 2022) توصلت إلى نفس النتائج من حيث ندرة استخدام هذا النوع من الصدق، وذلك طبعا لصعوبة الحصول على الاختبار الذي يعتبر معيارا ذهبيا، فالمعيار الذهبي حسب COSMIN، إما يكون عبارة عن نسخة مختصرة للاختبار الأصلي، أو نسخة مستحدثة، أو اختبار جد معروف بصلاحيته عبر العديد من الدراسات.

عند عرض نتائج **السؤال العاشر: هل تم التحقق من صدق الفرضيات بالشكل المطلوب حسب معايير COSMIN؟** تم تقسيم هذا السؤال إلى قسمين، يدرس القسم الأول كل من الصدق التقاربي والتباعدي، أما القسم الثاني فيدرس الصدق التمييزي، وبالنظر إلى الجدول رقم (19) يتضح لنا أن 16 دراسة قامت بمقارنة اختباراتها باختبارات أخرى تقيس نفس السمة أو سمة مغايرة عنها، وقد تم تقييمها جميعا بـV في البند الأول الذي يكشف عن سمة الاختبارات المقارنة، في حين أن 8 اختبارات تم تقييمها بـD أي مشكوك فيه و2 بـI أي غير مناسب في البند الثاني الذي يتسآل عن الخصائص السيكومترية للاختبارات المقارنة، وهو الذي صادفناه في العديد من المربعات بحيث دائما ما تكون هنالك معلومات لم يتم الإفصاح عنها، في حين أن البند الثالث الذي يقيم

الطرق الإحصائية المستخدمة فأغلب الدراسات كان تقييمها جيد جدا، غير أن تلك الأساليب الإحصائية رغم أن الاختيار كان سليما، إلا أنها حازت 8 تقييمات D بسبب أنها لم تحترم القيم المنصوص عليها حسب قائمة معايير COSMIN. أما في القسم المتعلق بالصدق التمييزي رغم أن هذا الإجراء لا يقتصر فقط على المقارنة الطرفية، إلا أنه الأكثر استخداما، فمن أصل 20 دراسة استخدمت الصدق التمييزي بين المجموعات؛ 18 دراسة منها اعتمدت على المقارنة الطرفية (الجدول رقم 20). أما في ما يخص تقييم 20 دراسة، فلم يتم تقديم وصف واضح للمجموعات المستخدمة في 7 دراسات، ودراسة واحدة فقط لم تكن الطريقة الإحصائية فيها مناسبة وهي دراسة (الفرطوسي 2013) والتي اعتمدت اختبار ت بين عينتين مترابطتين بدل عينتين مستقلتين، وبالنسبة للبند الثالث الذي يقيم العيوب الأخرى فإن 9 دراسات قُيمت بين مشكوك فيه وغير مناسب. وغالبا بسبب قيم الأساليب الإحصائية المنخفضة.

لقد تجاوزت الباحثة مناقشة **نتائج السؤال الثامن: هل قام الباحثون بالكشف عن أخطاء القياس؟ والسؤال الحادي عشر: هل تم تقييم الاستجابة أو المآل للاختبارات النفسية بالطريقة الصحيحة حسب المعايير التي وضعتها COSMIN؟** وذلك لأنه كما هو واضح من خلال الاطلاع على نتائج الجدول رقم (17) والجداول (19) (20) (21) فإن جميع الدراسات المدرجة في العينة لم تتحقق من خطأ القياس ولا الاستجابة. وهو الأمر الذي اتفقت معه عدة دراسات منها: (Haile, et al., 2022) ودراسة (Irene, Luisa, Luca, & Alvisa, 2017).

وفي الأخير وبعد عرض ومناقشة نتائج كل مربع على حدى والتفصيل في تقييم إجراءاته سيتم ختام هذه الدراسة بمناقشة تصنيف الجودة الكلية، من خلال الإجابة عن السؤال **الثاني عشر: هل كانت الخصائص السيكومترية التي تم استخدامها في إعادة التحقق من صلاحية الاختبارات النفسية ذات جودة مقبولة؟** وكما اتضح من خلال عرض نتائج هذا السؤال في الجداول التالية:

(25–26–27–28–29–30–31–32) فقد تراوحت جودة الدراسات المشمولة من مرتفع إلى منخفض جدا. تم تصنيف إحدى عشر دراسة بجودة مرتفع من حيث صدق المحتوى؛ بينما تم تصنيف 19 دراسة من أصل 62 على أنها متوسطة، وتصنيف جودة 20 اختبار منخفض، أما منخفض جدا فقد صنفت جودة 12 دراسة بها، وكان السبب الرئيسي للتصنيف منخفض ومنخفض جدا في تلك الدراسات، هو وجود مشاكل منهجية في الخطوات المستخدمة والتي كانت في الغالب غير واضحة مثلا ليس من الواضح ما إذا تم تدريب من قام بالمقابلات، أو إذا تم استخدام دليل المقابلة، أم تم تسجيل الاجتماعات والمقابلات، وما إلى ذلك.

وفي ما يخص جودة المربع الثالث، توصلت النتائج إلى 19 دراسة من الدراسات 97 التي قيمت الصدق الهيكلي جودتها مرتفعة، و35 دراسة كانت ذات جودة مناسبة أي متوسطة، و30 دراسة أخرى منخفضة، و13 دراسة جودتها منخفضة جدا. وترجع أسباب انخفاض جودة 43 دراسة غالبا إلى انخفاض كل من قيم الأساليب الإحصائية المستخدمة، وحجم العينة المشمول في الدراسات.

كانت أغلبية الدراسات التي قيمت ثبات الاتساق الداخلي ذات جودة مرتفعة بـ 96 دراسة، أما متوسطة الجودة فقد كان عدد الدراسات 27، في حين أن 24 دراسة ذات جودة منخفضة، و5 دراسات فقط ذات جودة منخفضة جدا. وذلك لكون ألفا كرونباخ لا تحتاج إلى الكثير من الشروط والمعايير مثل باقي الإجراءات، كذلك كانت أغلب القيم فوق 0.70، مما سمح للدراسات أن تتحصل على جود كافية، ما عاد 29 دراسة انخفضت جودتها غالبا بسبب استخدام ألفا كرونباخ مع اختبارات متعددة الأبعاد.

لم يكن لأي من الدراسات الأربع التي قيمت الصدق الثقافي تصنيفا مرتفعا، فقد تم تصنيف جودة دراسة واحدة متوسطة والأخرى منخفضة أما الدراستين الباقيتين فتم تصنيفهما بمنخفض جدا،

والسبب راجع في ذلك إلى حجم العينة المشمول والأساليب الإحصائية المستخدمة، فمثلا لم يتم استخدام تحليل الأداء التفاضلي في أي دراسة.

جودة الاختبارات في المربع السادس لم تكن في المستوى المطلوب، فقط 3 اختبارات من أصل 50 اختبار تحصلت على تصنيف مرتفع في التطبيق وإعادة التطبيق، في حين أن 22 اختبار تم تصنيفه متوسط، و20 اختبار جودته منخفضة، وفي الأخير 5 اختبارات ذات جودة منخفضة للغاية، والسبب في ذلك هو أن معظم المعلومات كانت غير واضحة فمثلا لم يكن من الواضح هل العينة كانت مستقرة، ولم يكن من الواضح هل الظروف كانت متشابهة بين التطبيقين، كذلك من أسباب تدني جودة هذا المربع هو الأسلوب الإحصائي المستخدم فدائما ما استخدم معامل الارتباط بيرسون بدل معامل الارتباط داخل الصف إلا في دراسة واحدة.

أما الصدق المحكي وهو المربع الثامن، وذلك لكون المربع السابع يخص خطأ القياس الذي لم تتحقق منه أي دراسة، في حين أن الصدق المحكي أصلا لم يستخدم إلا في دراستين، كانت جودة إحداهما مرتفعة والأخرى متوسطة، والسبب في كون تدني درجة جودة الدراسة الثانية إلى الرتبة متوسط هو أن قيم معاملات الارتباط بين أبعاد الاختبارين كانت أقل من 0.70، مما خفض من جودة هذه الدراسة.

وأخير المربع التاسع والذي يقيم صدق الفرضيات والذي قسمته الباحثة حسب القائمة إلى قسمين: قسم يعتمد على المقارنة بين الاختبارات (الصدق التقاربي والتباعدي)، والآخر يعتمد على المقارنة بين المجموعات (الصدق التمييزي)، وكانت جودة فحص الصدق التقاربي والتباعدي مرتفعة في 5 دراسات ومتوسطة في 9 دراسات وأخيرا منخفضة في دراستين، ولم يتم تصنيف أي دراسة بمنخفض جدا. وكان السبب الرئيسي في التصنيف المنخفض لتلك الدراستين هو عدم توفر المعلومات الكافية عن صلاحية المقياس المحك، وأيضا انخفاض قيمة معاملات الارتباط بين

الاختبارين. أما جودة الصدق التمييزي فقد تراوحت بين المنخفض جدا والمرتفع، فتحصلت 7 اختبارات على جودة مرتفعة، و8 اختبارات على جودة متوسطة، في الوقت الذي كانت فقط 4 اختبارات ذات جودة منخفضة واختبار واحد ذو جودة منخفضة جدا. وكالعادة فإن أهم الأسباب التي تؤدي انخفاض جودة الاختبار هي قيم معاملات الارتباط التي تكون جد منخفضة، وأيضا عدم وضوح المعلومات المتعلقة بالمجموعات. وبالنسبة للمربع العاشر الذي يدرس الاستجابة فكما ذكرنا سابقا لم تقم أي دراسة بالتحقق من الاستجابة وذلك راجع لكون هذا النوع من الخصائص يحتاج إلى دراسات طولية.

3-الاستنتاج العام:

بعد مراجعة وتحليل عينة البحث المتمثلة 171 اختبار نفسي، من أجل الوصول إلى هدف الدراسة الذي يتمحور حول تقييم طرق إعادة التحقق من صدق وثبات الاختبارات النفسية، تم التوصل إلى العديد من الاستنتاجات التالية:

لم تستخدم أي دراسة من الدراسات 171، النظريات الحديثة في القياس مثل نموذج راش ونظرية الاستجابة للفقرة. رغم أن عينة الدراسة قد حُصرت بين 2012-2022، وما لا يقل عن 80 اختبار منها تم التحقق منه في الخمس سنوات الأخيرة فقط. إلا أن الباحثين فضلوا استخدام النظرية التقليدية في القياس.

لا يعتبر الصدق الذاتي من أنواع الصدق بأي شكل من الأشكال، لعدة أسباب أهمها أن هذا الصدق يناقض مبدأ القياس الذي ينص على أن الصدق يوجب الثبات والعكس غير صحيح، كذلك قيمة جذر الثبات دائما أكبر منه، وبالتالي تصبح دائما قيمة الصدق الذاتي أكبر من ثباته وهو أمر غير مبرر منطقيا وإحصائيا.

الاتساق الداخلي هو الإجراء الأكثر استخداما سواء كثبات أو كصدق. فقد تم استخدام ثبات الاتساق الداخلي في 152 اختبار، والصدق الهيكلي في 97 اختبار. وهو الأمر الذي تم الإشارة إليه في الجانب النظري، فلا يمكن الاعتماد فقط على الاتساق الداخلي للتحقق من صدق وثبات الاختبارات النفسية، رغم أهمية هذان الإجراءان لكن لابد من مصاحبتهما بإجراءات أخرى.

الصدق المحكي هو الصدق الأقل استخداما بسبب صعوبة الحصول على الاختبار الذهبي. والاختبار الذهبي كما سبق وذكرنا هو النسخة المستحدثة للاختبار الحالي، النسخة الأصلية للاختبار في حالة استخدام النسخة المختصرة، والاختبارات التي توثقت صلاحيتها عبر العديد من الدراسات.

ضرورة استخدام الصدق الثقافي بإجراءاته الصحيحة مثل دراسة الأداء التفاضلي للبند، للاختبارات التي تم تغيير بيئتها. فلا يمكن تغيير مجتمع الدراسة مثلا وعدم التحقق من صلاحية البنود، لأنه في هذه الحالة يصبح الصدق والثبات بإجراءاتهما المعهودة غير كافية، لذلك تم إضافة الصدق الثقافي والذي يدرس عدم تحيز البنود وملاءمتها للاستخدام الحالي.

غالبية القيم الناتجة عن مختلف الأساليب الإحصائية كانت ضعيفة، خاصة عند التحقق من صدق الاتساق الداخلي، في حين ثبات الاتساق الداخلي وبالضبط ألفا كرونباخ كانت قيمه مرتفعة غالبا، وذلك استنادا إلى معايير قائمة COSMIN.

الاتساق الداخلي كصدق (العلاقة بين البنود والبعد، والعلاقة بين الأبعاد والدرجة الكلية) لا يكشف عما يقيسه الاختبار، بل فقط يكشف أنه يقيس شيئا مشتركا لذلك لا يمكن الاكتفاء به بل وجب دعمه بأدلة أخرى.

ألفا كرونباخ لا يصلح إلى للاختبارات أحادية البعد أو للمقاييس الفرعية كل على حدى، فهو لا يصلح مع لمقياس متعدد الأبعاد، لذلك وجب في هذه الحالة الاعتماد على بديلها أوميغا ماكدونالدز أو ألفا الطبقية.

تراوحت جودة جميع الخصائص السيكومترية عموما بين منخفضة ومرتفعة. فلم تعتبر ولا خاصية ذات جودة منخفضة جدا، لأنه حتى وإن كانت بعض الدراسات أخطأت في تقديرها، كانت الأخرى قد قامت باستخدامها وفق ما هو مطلوب، ويعتبر صدق المحتوى هو الإجراء الأكثر انخفاضا من حيث الجودة في حين تعتبر ألفا كرونباخ هي الأحسن جودة.

لم تعتبر الباحثة الصدق الظاهري الذي يقوم فيه الباحث بقراءة الاختبار فقط صدقا، ولم تقم بتقييمه في هذه الدراسة، وتم تقييمه فقط في حالة صاحبته باقي إجراءات التحقق من صدق المحتوى.

في جميع الدراسات 171 المستخدمة، لم يتم حذف ولا اختبار لعدم صلاحيته ما عدا اختبار واحد فقط، رغم أن العديد من الاختبارات كانت ذات جودة جد منخفضة.

يعتبر إصدار معايير COSMIN الأخير لعام 2018 هو الأكثر صرامة وتعقيدًا في التطبيق من باقي الإصدارات، بسبب أن العملية متعددة الخطوات. وتحتاج إلى العديد من الإجراءات من أجل الوصول إلى الجداول السابقة. إلا أن الباحثة تنصح بشدة باستخدامها ليس فقط من أجل الدراسات التقييميه بل أيضا من أجل تقييم جودة أي اختبار قبل استخدامه.

الخاتمة

الخاتمة

بناءً على ما تطرقنا إليه في دراستنا الحالية سواء في جانبها النظري أو من خلال نتائجها، فإنه يتضح لنا أن الطرق التي يتم إعادة التحقق بها من صلاحية الاختبارات النفسية قبل استخدامها، ليس بالجودة المطلوبة، فقد اقترف الباحثون عدة أخطاء منهجية وإحصائية، منها ما تم التركيز عليها وتقييمها في هذه الدراسة، ومنها ما سيتم اقتراحه على باحثين آخرين لإلقاء الضوء عليها وتقييمها كما يجب، بحيث فضلت الباحثة التقيد بأهداف البحث والمتمثلة في الكشف عن طرق التحقق من صدق وثبات الاختبارات النفسية، وتقييم تلك الطرق بالاعتماد على قائمة المعايير COSMIN، وتسليط الضوء على الأساليب الإحصائية التي تم انتقادها واقتراح البديل لها، وكل هذا من أجل ضمان جودة الاختبارات النفسية ومن ثم الوثوق في نتائجها وإمكانية تعميم تلك النتائج، خاصة وأن الاختبارات النفسية تعتبر من أهم وسائل القياس في علم النفس. لذلك لا يجب التعامل معها على أنها أدوات صالحة لكل الظروف والأزمنة، بل تحتاج هذه الأخيرة دائما إلى إعادة التحقق من صدقها وثباتها وصلاحيتها للاستخدام المرجو، أو التحقق من جودة خصائصها وسلامة بناءها قبل أي استخدام.

الاقتراحات

في ضوء نتائج الدراسة الحالية تقترح الباحثة ما يلي:

- إجراء دراسات مماثلة بالاعتماد على قائمة المعايير COSMIN في تقييم الاختبارات النفسية التي تم بناءها أو تكييفها.
- القيام بمراجعات منهجية شاملة للاختبارات النفسية المستخدمة في مراكز الرعاية الصحية بأنواعها مثل المستشفيات والعيادات النفسية من أجل تقييم جودتها. اعتمادا على معايير دولية.
- إجراء دراسات خاصة بالانتهاكات التي يرتكبها الباحثون عند نقل الاختبارات إلى بيئة جديدة.
- إجراء مزيد من الدراسات على مدى مناسبة دلالات الصدق والثبات للسمات التي تقيسها الاختبارات النفسية المستخدمة في البحوث والدراسات.
- إجراء المزيد من الدراسات حول التجاوزات التي يقع في الباحثون عند ترجمة الاختبارات النفسية.
- تخصيص مخابر ومراكز خاصة، لبناء الاختبارات النفسية وتكييفها. وإعادة التحقق من صلاحيتها.
- تكوين الطلبة، الأساتذة والباحثين في مجل الإحصاء والذي يعتبر أساس الخصائص السيكومترية.
- عقد ورشات ودورات من أجل تصحيح الأخطاء التي أصبحت شائعة ومقبولة عند إعادة التحقق من صلاحية الاختبارات النفسية.

- على الهيئات العلمية العليا اشتراط ضرورة مرور أي اختبار نفسي على المخابر المتخصصة في القياس من أجل التحقق من جودته قبل أي استخدام.
- وضع ضوابط رادعة من أجل منع استخدام أي اختبار دون عرض موافقة صاحبه.
- اشتراط تقديم موافقة صاحب الاختبار على أي تعديل يقوم به الباحث.
- اعتماد قائمة المعايير COSMIN أو القوائم الدولية الأخرى المعترف بها، في جميع المخابر من أجل تقييم جودة أي اختبار والتحقق من صلاحيته.

قائمة المراجع

المراجع

المراجع العربية:

1. إبراهيم أنيس، عبد الحليم منتصر، عطية الصوالحي، و محمد خلف الله: المعجم الوسيط (ط4). القاهرة: مجمع اللغة العربية–مكتبة الشروق الدولية، 2004.

2. ابراهيم محمد محاسنة: *القياس النفسي في ظل النظرية التقليدية و النظرية الحديثة*. عمان: دار جرير للنشر والتوزيع، 2013.

3. إحسان الأغا: *البحث العلمي"منهجه–عناصره–أدواته"*. غزة: مطبعة مقداد، 1997.

4. أحلام معمري: أسس وضع اختبارات تعليم اللغة العربية للناطقين بغيرها. *مجلة الباحث في العلوم الانسانية والاجتماعية*. (35). 2018، ص 595–606.

5. أحمد محمد سيد: طيف الاختبارات النفسية وإشكاليات القياس في الوطن العربي. *مجلة دراسات نفسية وتربوية . 1* (1). 2008، ص 81–131.

6. أحمد محمد عبد الخالق: *استخبارات الشخصية* (ط3). القاهرة: دار المعرفة، 2000.

7. أزهار محمد غليون: تقويم الاختبارات النهائية للدراسات العليا بكلية التربية صنعاء في ضوء معايير الاختبار الجيد. *المجلة العلمية لكلية التربية ، 1* (8). 2009، ص 176–200.

8. إسماعيل عبد الفتاح الكافي: *اختبارات الذكاء والشخصية*. مركز الاسكندرية للكتاب، 2001.

9. امحمد بوزيان تيغزة: نظرية الصدق الحديثة ومتضمناتها التطويرية لواقع القياس. *كلية التربية قسم علم النفس* : *جامعة الملك سعود*،2008.

10. امحمد تيغزة: توجهات حديثة في تقدير صدق وثبات درجات أدوات القياس: تحليل نظري تقويمي تطبيقي. مجلة العلوم النفسية والتربوية. 4(1). 2017 ، ص7-29.

11. آن انستازي، و سوزان أوربينا: *القياس النفسي*. ترجمه: صلاح الدين محمود علام. دار الفكر العربي ناشرون وموزعون. عمان، الأردن. (ط1)، 2015.

12. ايمان عبد الكريم ذيب: أثر طول الاختبار على تحديد درجة القطع لاختبار تحصيلي محكي المرجع في مادة علم نفس الشخصية. *مجلة مداد الآداب . 1* (6).2020، ص 330-408.

13. بشير معمرية: *أساسيات القياس النفسي وتصميم أدواته*. جامعة الحاج لخضر، باتنة: دار الخلدونية للنشر والتوزيع، 2012.

14. بشير معمرية: *القياس النفسي وتصميم أدواته للطلاب والباحثين في علم النفس والتربية*. الجزائر: دار الخلدونية للنشر والتوزيع، 2007.

15. بلال بن عرعور، و ليندة عمارة: مواصفات الباحث والبحث العلمي ملتقى وطني حول آداب وآخلاقيات مهنة التدريس الجامعي والبحث العلمي. *مجلة الباحث للعلوم الرياصية والاجتماعية . 2* (1). 2019، ص 373-387.

16. جلال الدين بوعطيط، و هناء بوحارة.: الاختبارات النفسية ودورها في تفعيل العمليات المهنية بمؤسسات العمل الواقع والأهمية. *المجلة العربية نفسانيات* (50-51). 2016، ص 37-47.

17. حيدر إبراهيم ظاظا: درجة توافق دلالات صدق وثبات الاختبارات المقننة المستخدمة في رسائل الماجستير المقدمة في كليات التربية في الجامعات الأردنية مع دلالات صورها الأصلية. *دراسات العلوم التربوية ، 38* (2). 2011، ص2400-2416.

18. خالد إبراهيم حسن الكردي: *مناهج البحث في علم النفس*. الخرطوم: جامعة النيلين، 2004.

19. خالد العتيبي: مستويات أداء الطلبة الثانوية العامة في المملكة العربية السعودية في اختبار القدرات العامة. *مجلة العلوم التربوية والنفسية. 19* (1).2018، ص 613-638.

20. عبد القادر رحيم: اتجاهات معاصرة في كيفية استخدام أدوات القياس النفسي. *مجلة أبحاث البصرة (العلوم الانسانية). 01* (22). 2008، ص 195-205.

21. دنيا عدائكة، و يمينة بن موسى: معايير الجودة في بناء الاختبارات التحصيلية الجيّدة. *مجلة السراج في التربية وقضايا المجتمع. 1* (1).2017، ص 216-227.

22. دينيس هوبيت، و دونكان كرامر: *مقدمة لطرائق البحث في علم النفس* (ط1). ترجمه: صلاح الدين محمود علام. دار الفكر. عمّان، الأردن. (ط1)، 2016.

23. رونالد ك هامبلتون، بيتر ف ميريندا، و تشارلز د سبيلبيرغر: *تكييف الاختبارات التربوية والنفسية للتقييم عبر الثقافات*.ترجمه: هالة برمدا. مكتبة العبيكان. الرياض. (ط1)، 2006.

24. رونالد ليفينجستون، و سيسيل رينولدز: تأليف *إتقان القياس النفسي الحديث النظريات والطرق*. ترجمه: صلاح الدين محمود علام.دار الفكر ناشرون وموزعون. عمان، الأردن، 2013.

25. ريم نصر قصاب: تقنين اختبار أوتيس– لينون للقدرة المدرسية على عينة من طلاب التعليم الأساسي في مدينة حماة. رسالة دكتوراه غير منشورة. كلية التربية: جامعة دمشق، 2014.

26. زياد بركات: *القياس والتقويم النفسي والتربوي بين النظرية والتطبيق*. جامعة القدس: عمادة البحث العلمي للنشر والتوزيع، 2008.

27. زينة مهلل، و عبد الرحمان عباس: النقل الثقافي للاختبارات النفسية ذات المنشأ الغربي في البيئة الجزائرية. *المجلة الجزائرية للطفولة والتربية* . *4* (1). 2016، ص 231–260.

28. سامي محمد ملحم: *القياس والتقويم في التربية وعلم النفس* (ط 1). عمان: دار المسيرة للنشر والتوزيع والطباعة، 2000.

29. سامي محمد ملحم: *القياس والتقويم في التربية وعلم النفس* (ط3). عمان: دار المسيرة للنشر والتوزيع، 2005.

30. سليمان عبد الواحد يوسف ابراهيم: *علم النفس التعليمي (نماذج التعلم وتطبيقاته في حجرة الدراسة)* (ط1). عمان: دار أسامة للنشر والتوزيع، 2013.

31. سليمة سلات: اشكالية استخدام الاختبارات والمقاييس والاستبيانات في مجال التوجيه المدرسي. *مجلة الروائز . 1* (1).2017، ص 24-28.

32. سوسن شاكر الجلبي: *أساسيات بناء الاختبارات والمقاييس النفسية والتربوية* (ط 1). دمشق: مؤسسة علاء الدين للطباعة والتوزيع، 2005.

33. سوسن شاكر مجيد: *اسس بناء الاختبارات والمقاييس النفسية والتربوية* (ط 3). جامعة بغداد: مركز ديبونو لتعليم التفكير، 2014.

34. سيسيل رينولدز، و رونالد ليفينجستون. (2013). تأليف *إتقان القياس النفسي الحديث النظريات والطرق. ترجمه:* صلاح الدين محمود علام. دار الفكر ناشرون وموزعون. عمان، الأردن، 2013.

35. صباح حسين العجيلي: *اساسيات في القياس والتقويم* (ط1). عمان، الاردن: دار وائل للنشر والتوزيع، 2017.

36. صفوت فرج: *القياس النفسي* (ط6). مصر: مكتبة الأنجلو المصرية، 2017.

37. صلاح الدين محمود علام: *الاختبارات والمقاييس التربوية والنفسية* (ط1). الأردن: دار الفكر، 2006.

38. صلاح الدين محمود علام: *الإختبارات والمقاييس التربوية والنفسية* (ط6). عمان، الأردن: دار الفكر ناشرون وموزعون، 2018.

39. صلاح الدين محمود علام: *القياس والتقويم التربوي والنفسي ،أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة: دار الفكر العربي، 2000.

40. صونيا عيواج، صفاء عثمان، و جمال بلبكاي: البحث العلمي- أسسه وأدواته: الاختبارات والمقاييس النفسية نموذجا. *مجلة الباحث للعلوم الرياضية والاجتماعية . 3* (1). 2020، ص 316-327.

41. عبد الرحمان بن صافي: دراسة تقييمية لطرق تقدير صدق وثبات أداوت جمع المعطيات. أطروحة دكتوراه في علم النفس وعلوم التربية. كلية العلوم الانسانية والاجتماعية: جامعة مستغانم رسالة غير منشورة، 2017.

42. عبد الرحمان بن صافي، و علي طاجين: تقييم طرق تقدير صدق أدوات القياس المستخدمة في البحوث النفسية والتربوية بجامعة الجزائر 2. *سلوك مخبر تحليل المعطيات الكمية والكيفية* (3)، 2016.

43. عبد الرحمان عباس، و زينة مهلل: النقل الثقافي للاختبارات النفسية ذات المنشأ الغربي في البيئة الجزائرية. *المجلة الجزائرية للطفولة والتربية*. *4* (1). 2016، ص 231-260.

44. عبد الرحمن بن معتوق بن عبد الرحمن زمزمي: تقنين اختبار تورانس للتفكير الابتكاري الشكل (ب) على الطلاب الصم وضعاف السمع في المرحلة المتوسطة بمنطقة مكة المكرمة. رسالة دكتوراه غير منشورة. كلية التربية: جامعة أم القرى، 2008.

45. عبد العزيز بوسالم. (بلا تاريخ). منهجية مقترحة لإعادة التحقق من صدق وثبات الاختبارات النفسية. من portal.arid.my

46. عبد العزيز بوسالم: *القياس في علم النفس وعلوم التربية الأسس النظرية والمبادئ التطبيقية*. الجزائر: دار قرطبة للنشر والتوزيع، 2014.

47. عبد العزيز بوسالم، و أحمد كريش: الأداء التفاضلي للبند في الاختبارات النفسية، مصادره وتفسيره. *مجلة العلوم الاجتماعية*. 2016، ص 127-142.

48. عبد الناصر الجراح، و ميساء أبو أحمد: أثر طريقة تقديم المعلومات ونوعها وفترات الاحتفاظ في القدرة على التعرف والاستدعاء لدى طلبة جامعة اليرموك. *المجلة الأردنية في العلوم التربوية*. *11* (1). 2015، ص 45-60.

49. عز الدين بوداري، و الصالح بوعزة. (2019). أهمية الاختبارات والمقاييس النفسية في البحوث النفسية. *مجلة الوقاية و الأرغنوميا، جامعة الجزائر 2 . 7* (1). 2019، ص 70-86.

50. علي حسين هاشم الزاملي: *بناء وتقنين المقاييس النفسية*. جامعة القادسية: دار الكتب في بغداد، 2017.

51. علي صلاح عبد المحسن حسن: الأخطاء الشائعة في بناء الاختبارات الموضوعية وتأثيرها علي الخصائص السيكومترية للاختبارات التحصيلية لدى طلاب كلية التربية بجامعة أسيوط. *المجلة العلمية لكلية التربية*. *34* (2). 2018، ص 322-402.

52. عماد حنون الكحلوت: فاعلية برنامج تدريبي لتنمية مهارات إعداد الاختبارات النفسية وتطبيقها وتفسير نتائجها لدى الطلبة المرشدين النفسيين. *مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية*. *4* (14).2016، ص 133-172.

53. فارس إساعدي، و ايمان عزي: مواصفات الاختبار التحصيلي الجيّد. *مجلة الدراسات والبحوث الاجتماعية* . *3* (2).2015، ص 189-200.

54. فتيحة بودلال، و محبوبة موراس: دراسة تقييميه لقياس الخصائص السيكومترية (الصدق والثبات) في البحوث العلمية. *مجلة ضياء للبحوث النفسية والتربوية*. *1* (2).2015، ص 111-131.

55. كمال عبد المجيد حزيمة: الكشف عن فاعلية أداتي لكشف الموهوبين من وجهة نظر مدريسهم. *مجلة كلية التربية*. *1* (27). 2016، ص 175-194.

56. ل جاي، ج ميلز، و ب ايرسيان: *البحث التربوي كفايات للتحليل والتطبيق*. ترجمه: صلاح الدين محمود علام. دار الفكر ناشرون وموزعون. الأدرن، 2012

57. ليندة كروكر، و جايمس ألجينا: تأليف *مدخل إلى نظرية القياس التقليدية والمعاصرة*. ترجمه: هند عبد المجيد الحموري، و زينات يوسف دعنا. دار الفكر ناشرون وموزعون.عمان ، الاردن، 2017.

58. ليونا تايلر: *الاختبارات والمقاييس* (ط 3). بيروت: دار الشروق للنشر والتوزيع، 1988.

59. محمد ابن منظور: *لسان العرب* (ط6). بيروت: دار صادر، 1994.

60. محمد أحمد الخطيب، و أحمد حامد الخطيب: *الاختبارات والمقاييس النفسية*. عمان: دار ومكتبة الحامد للنشر والتوزيع، 2010.

61. محمد بلعالية: الاختبارات التحصيلية كأسلوب فعال للتقويم التربوي. *مجلة الخلدونية للعلوم الانسانية والاجتماعية . 13* (1). 2021، ص 53-61.

62. محمد خماد: واقع استخدام الاختبارات النفسية في عملية الارشاد. *مجلة الدراسات والبحوث الاجتماعية . 2* (3). 2014، ص 186-201.

63. محمد عبد السلام أحمد: *القياس النفسي والتربوي* (ط1). القاهرة: ملتزمة الطبع والنشر مكتبة النهضة المصرية، 1960.

64. محمد عرفات جخراب، ربيعة جعفور، و دنيا عدائكة: كفايات إعداد الاختبارات التحصيلية الجيدة لدى أساتذة التعليم الثانوي. *مجلة العلوم الانسانية . 26* (4).2015، ص 515-540.

65. مروان العطية: *المعجم الجامع* (ط1). الأردن: دار الغيداء للنشر والتوزيع، 2018.

66. مصطفى محمود الامام، أنور حسين عبد الرحمان، و صباح حسين العجيلي: *التقويم والقياس*. الأردن: دار الأيام، 2016.

67. موسى النبهان: *أساسيات القياس في العلوم السلوكية* (الإصدار ط2). عمان، الأردن: دار الشروق للنشر والتوزيع، 2013.

68. نادية مصطفى العساف، و ختام محمد الوزان: أسس تصميم اختبارات اللغة العربية للناطقين بغيرها. *مجلة العلوم الانسانية والاجتماعية . 41* (1). 2014، ص 173-183.

69. نايف مخائيل امطانيوس: *بناء الاختبارات والمقاييس النفسية والتربوية وتقنيتها* (ط1). عمان: دار الاعصار العلمي للنشر والتوزيع، 2016.

70. همام عبد الله السردي: تقويم أدوات البحث المستخدمة في البحوث التربوية لدى طلبة الدراسات العليا في الجامعة الإسلامية بغزة. رسالة ماجستير غير منشورة . كلية التربية: الجامعة الإسلامية ، 2012.

71. ياسر خلف الشجيري، و عبد الكريم حيدر الزهيري: *اتجاهات حديثة في القياس والتقويم النفسي والتربوي* (ط1). الأردن: مكتبة المجتمع العربي ودار الإعصار العلمي، 2022.

72. زايدي لمين و ليلى سهل: الاختبارات ودورها في تحقيق مبدا الجودة التعليمية. 22(49). 2020، ص 269-296.

المراجع الأجنبية:

1. Adkins, j.: An Examination of Bias in Oral Reading Fluency: Differential Effects across Race, Gender, and Socioeconomic Status. Arizona State University Unpublished doctoral dissertion, 2013.

2. Adoga, R.: psychological test in counselling. *prestige journal of conselling psychology* .2020, pp. 232-248.

3. AERA: Standards for Reporting on Empirical Social Science Research in AERA Publications:American Educational Research Association. *Educational Researcher . 35* (6). 2006, pp. 33-40.

4. Alrfooh, A. E: type of avheivement tests which are preferred by outstanding students at al-hussein bin talal university. *Journal of Education and practice .13* (3). 2012, pp. 149-156.

5. Ananthi, B: Non- Referenced test, criterion- Referenced test and self- Referenced test Frameworks. *journal of information and computation of sciences*. *10* (5). 2020, pp. 605-618.

6. Anugerah, R.: Evaluation of question items in the National Comprehensive Arabic Language Test at the level of the Islamic secondary school for the academic year 2014-2015. *Jurnal pendidikan Bahasa Arab dan kebahasaar aban .5* (2). 2017, pp. 150-175.

7. AREA, A. E., APA, A. P., & NCME, N. C: *Standards for Educational and Psychological Testing.* Washington: American Educational Research Association, 2014.

8. Ary, D., Jacobs, L., Sorensen, C., & Walker, D: *Introduction to research in education* (9ed). Belmont: Wadsworth, 2014.

9. Ashra, H., Barnes, C., Stupple, E., & Maratos, F.: A Systematic Review of Self-Report Measures of Negative Self-Referential Emotions Developed for Non-Clinical Child and Adolescent Samples. *Clinical Child and Family Psychology Review , 24*. 2021, pp 224-243.

10. Axelsson, T: intelligence testing: ethnicity, and construction of the deviant child: foucauld and special education in sweden. *Nordic journal of social research*. *7* (1). 2016, pp. 22-35.

11. Beichi, A., Tchalabi, S., & Alsalhi, N: Internal Consistency Between Validity and Reliability in. *Journal of Statistics Applications & Probability*. *11* (1). 2022, pp 207-216.

12. Bringle, R., Mindy, P., & Hudson, M: *The Measure of Service learning* (Ed 2). Washington: American Psychological Association, 2004.

13. Camili, G., & Shepard, L: Methods of identifying biased test item. *Sage Publications*, 1994.

14. Campbell, D. T., & Fiske, D. W: Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin .56* (2). 1959, pp 81-105.

15. Campbell, L. F., Stanford, L. D., Alfonso, V. C., Cuttler, M. J., Demers, S. T., Hass, G. A., et al: APA Guidelines for psychological Assessment and Evaluation. *American Psychological Association*. 2020, pp. 1-27.

16. Cassidy, S., Bradley, L., Bowen, E., Wigham, S., & Rodgers, J: Measurement properties of tools used to assess suicidality in autistic and general population adults: A systematic review. *Clinical Psychology Review . 62*. 2018, pp56-70.

17. Chen, X., & Gabrenya, W. K: In search of cross-cultural competence: A comprehensive review of five measurement instruments. *International Journal of Intercultural Relations . 82* .2021, pp33-55.

18. Cleary, T., & Hilton, T: An investigation of item bias. *Educational and Psychological Measurement .28*.1968, pp 61-75.

19. Coaley, K: *An introduction to psychological Assessment and psychometrics.* London: Sage Publication, 2010.

20. Cohen, R. J., & Swerdik, M: *Psychological testing and Assessment* (09 ed.). united stastes of America: MC Graw- Hill Education, 2018.

21. Cohen, R., & Swerdlik, M: *Psychological testing and assessment:An introduction to tests and measurement* (7ed ed.). new york: Mc Gruw-Hill, 2009.

22. Cortina, J: What is coefficient alpha? An examination of theory and applications. *J Apple Psychol . 78* . 1993, pp98-104.

23. Costello, A., & Osbourne, J: *Best practices in exploratory factor analysis: Four four recommendations for getting the most from your analysis.* Retrieved November 11, 2020, from http://pareonline.net/pdf/v10n7.pdf, 2015.

24. Cronbach, L: Coefficient alpha and the internal structure of tests. *Psychometrika , 16* (4). 1951, pp297-334.

25. Cronbach, L. J: TESTN RELIABILITY: ITS MEANING AND DETERMINQTION. *PSYCHOMETRIKA . 12* (1). 1974, pp 1-16.

26. Cronbach, L: Test Validation. *Educational Measurement* . 1971, pp 443-507.

27. Cronbach, L., & Meehl, P: Construct validity in psychological tests. *Psychological Bulletin . 52* (4). 1955, pp 281-302.

28. cropley, A: Defining and Measuring creativity: Are creativity tests worth using? Journal of Educational Measurement.23 (2). 2000, pp. 72-79.

29. de Vet, H., Terwee, C., Mokkink, L., & Knol, D: Measurement *in Medicine: a practical guide, 2011.*

30. de Vet, H., Terwee, C., Mokkink, L., & Knol, D: Measurement in Medicine: a practical guide. *Cambridge University Press* , 2011.

31. Dobson, F., Choi, Y., Hall, M., & Hinman, R: Clinimetric properties of observer- assessed impairment tests used to evaluate hip and groin impairments: a systematic review. *Arthritis Care Res (Hoboken) . 64* (10). 2012, pp 1565-1575.

32. Domino, E., & Domino, L: *Psychological testing an introduction* (2ed). cambridge: cambridge university press, 2006.

33. Drost, E: Validity and Reliability in social science Research. *Research Educational Resarch and perspectives .38*. 2011, (1), p. 110.

34. Ebel, R. L: Obtaining and Reporting Evidence on Content Validity. *Educational and Psychological Measurement*. *16* (3).1956, pp 269-282.

35. Ebel, R: Must all tests be valid? *American Psychologist*. *16* (10). 1961. 640-647.

36. El-Den, S., Chen, T. F., Gan, Y. L., Wong, E., & O'Reilly, C. L: The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. *Journal of Affective Disorders*. *225*. 2018, pp 503-522.

37. Engellant, A., Holland, D., & Piper, T: Assessing convergent and discriminant validity of the motivation construct for the technology integration education (TIE) model. *16* (1). 2016, 3-16.

38. Furr, R., & Bacharach, V: *Psychometrics: An Introduction* (2e ed.). London: Sage, 2014.

39. Georgievaa, S., Tomasa, J., & Navarro-Pérez, J: Systematic review and critical appraisal of Childhood Trauma Questionnaire — Short Form (CTQ-SF). *Child Abuse & Neglect. 120*. 2021, pp 105-223.

40. Gierl, M: Using a multidimensionality-based DIF analysis to identify and interpret constructs that elicit group differences. *Educational Measurement Issues and Practice*. *24* (1).2005, pp. 3-14.

41. Gomez-Benito, J., Sireci, S., Padilla, J., Hidalgo, I., & Benitez, I: Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema*. *30* (1). 2018, 104-109.

42. Gregory, R: *Psychological testing histiory principles and applications* (7ed). united states of America: Peason Education, 2014.

43. Gronlund, N. E. (1981). *Measurement and evaluation in teaching* (4th ed.). Macmillan Pub.

44. Haile, T., Habatmu, K., Derese, A., Gouse, H., Lawrie, S., Cella, M., et al: Assessing cognition in people with severe mental disorders in low- and middle-income countries: a systematic

review of assessment measures. *Soc Psychiatry Psychiatr Epidemiology . 57*. 2022, pp 435-460.

45. hogan, T: *psychological testing a practical introduction* (3ed). Uited state of America: Wiley, 2013.

46. Holland, P., & Thayer, D: *Differential item performance and the Mantel-Haenszel procedure.* New Jersey: Educational Testing Service, 1986.

47. Holland, P., & Wainer, H: Differential item functioning, Lawrence Erlbaum. *Associates Inc*, 1993.

48. International Test Commission: *The ITC Guidelines For Translating and Adapting Test* (2ed). (N. E. Boudouda, A. A. Bensouilah, M. E. Seridi, M. Elafri, S. Ben Youssef, & K. Gana, Trans.), 2017.

49. Irene, M., Luisa, S., Luca, G., & Alvisa, P: Instruments evaluating the quality of the clinical learning environment in nursing education: A systematic review of psychometric properties. *Inernational Journal of Nursing Studies , 68*. 2017, pp 60-72.

50. Keszei, A., Novak, M., & Streiner, D: Introduction to health measurement scales. *J Psychosom Res*. *68* (4). 2010, p 319-323.

51. Klein, S. P., & Kosecoff, J. P: *Determining how well a test measures your objectives (CSE Report 94).* Los Angeles, University of California Los Angeles: Center for the Study of Evaluation, 1975.

52. Kubisyzn, T., & Borich: *Education testing and measurement.* United state of America: Wiley, 2013.

53. Lehmnn, I., & Mehrens, W: *Measurement and Evaluation In Education and Psychology.* New York: Rinehart and winston, 1973.

54. Lemke, E., & Wiersa, w: *Principales of psychologycal Mesurement.* chicago, 1976.

55. Mawardin, M. R: Educational Calendar: Specification of a good test. *Jurnal Pendidikan Bahasa dan sastra Arab*. *2* (2). 2020, pp. 161-175.

56. Mayer, J: *Reliability.* London: Oxford university press, 2010.

57. McAndrews, Z., Richardson, J., & Stopa, L: Psychometric properties of acceptance measures: A systematic review. *Journal of Contextual Behavioral Science . 12*. 2019, pp 261-277.

58. McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., & Mulrow, C: *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*. Retrieved avril 17, 2020, from PRIMSA: http://www.prisma-statement.org/, 2020.

59. Messick, S: *Validity.* New Jeresy: Educational Testing Service Princeton, 1987.

60. Messick, S: Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist . 50*. 1995, pp 741-749.

61. Messick, S: *Validity of psychological assessment: validation of inferences from persons responses and performances as scientific inquiry into score meaning.* New jersey: Educational Testing Service, 1964.

62. Miller, D., Linn, R., & Gronlund, N: *Measurement and Assessment in teaching* (10 ed ed.). United states of America: Person Education, 2014.

63. Mohajan, H. K: Two criteria for good Measurements in Reserch: validity and Reliability. *Annals of spiru Haret university .17* (3). 2017, pp. 58-82.

64. Mokkink, L., de Vet, H., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., et al: COSMIN risk of bias checklist for assessing the methodological quality of studies on the measurement properties of Patient-Reported. *epub ahead of print.* 2017, pp. 1171-1179.

65. Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., et al: The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement istrument: an international Delphi study. *Qual Life Res.19* (4). 2010, pp 539-549.

66. Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., et al.: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported

outcomes. *Journal of Clinical Epidemiology*. *63* (7). 2010, pp 737-745.

67. Mokkink, L., Vet, H., Prinsen, C., Patrick, D., Alonso, J., & Bouter, L: COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res , 27* (5), 2018.

68. Newton, A., Soleimani, A., Kirkland, S., & Gokiert, R: A Systematic Review of Instruments to Identify Mental Health and Substance UseProblems Among Children in the Emergency Department. *Academic Emergency Medicine*. *24* (5). 2017, pp 552-568.

69. Nictko, A., & Brook hart, S: *Educational assessment of students englercood chiffs.* NJ Mersill, 2018.

70. Nunnally, J., & Bernstein, I: *Psychometric theory* (3 ed). New York: MC Graw-Hill, 1994.

71. Ong, C., Krafft, J., Levin, M., & Twohig, M: A systematic review and psychometric evaluation of self-report measures for hoarding disorder. *Journal of Affective Disorders. 290*. 2021, pp 136-148.

72. Passmore, J: *psychometrics in coaching using psychological and psychometric tools for development.* London and Philadelphia: kogan page limited, 2008.

73. Peters, G: The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist .16* (2). 2014, pp 56-69.

74. Pham, J: A student's perspective on assessment and the affective domain. *Am J Pharm Educ . 79* (8), 2015.

75. Polit, D: Assessing measurement in health: Beyond reliability and validity,International. *Journal of Nursing Studies .52*. 2015, pp 1746-1753.

76. Polit, D., & Yang, F: *Measurement and the measurement of change: a primer for the health professions.* Philadelphia: Wolters Kluwer, 2016.

77. Popham, W: The Mismeasurement of Educational Quality. *School Administrator .57* (11). 2000, pp 12-15.

78. Price, P., Jhangiani, R., & Chiang, C: *Research Methods in Psychology.* Retrieved November 11, 2020, from https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/, 2016.

79. Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., et al. (2018). COSMIN guideline for systematic reviews of patient-reported outcome maesures. *Quality of Life Research . 27* (5). 2018, pp 1147-1157.

80. Reitsma, J., Rutjes, A., Whiting, P., Vlassov, W., Leeflang, M., & Deeks, J: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Retrieved from The Cochrane Collaboration: http://srdta.cochrane.org/, 2009.

81. Rolin, J. L: *Savoir, Comprendre, Aprendre Leçons de Psychométrie* (éd 03). Suisse: Université Savoie Mont Blanc, 2018.

82. Ronna C, T., & Laurie, C: Indexes of Item-Objective Congruence for Multidimensional Items. *International Journal of Testing*. 2003, pp 163-171.

83. Saether, R., Helbostad, J., Riphagen, I., & Vik, T: Clinical tools to assess balance in children and adults with cerebral palsy: a systematic review. *Dev Med Child Neurol. 55* (11). 2013, pp 988-999.

84. Saini, S., Hoffmann, C., Pantelis, C., Everall, I., & Bousman, C: Systematic review and critical appraisal of child abuse measurement instruments. *Psychiatry Research . 272*. 2019, pp 106-113.

85. Salmond, S: Evaluating the reliability and validity of measurement instruments. *Orthop Nurs .27*. 2008, pp 19-30.

86. Santee, J. A., Marszalek, J. M., & Hardinger, K. L: A critique of validity analysis from instruments assessing the affective domain. *Currents in Pharmacy Teaching and Learning.11* (2), 2019.

87. Sartori, R: Face validity in personality tests: Psychometric instruments and projective techniques in comparison. *Quality & Quantity .44* (4). 2010, pp 749-759.

88. Setiwati, F. A: Aptitude test's presdictive Ability for Academic success in psychology student. *Psychological Reaserch Student . 3* (1). 2020, pp. 1-12.

89. shafizan, s: Item analysis of student comprehensive test for research in teaching beginner string ensemble usingmodel based teaching among music students in public universities. *1* (12). 2013, pp. 1-14.

90. Souza, A., Alexandre, N. M., & Guirardello: Edinêis de BritoPropriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde .26* (3). 2017, pp 649-659.

91. Stallwood, E., Monsour, A., Rodrigues, C., Monga, S., Terwee, C., Offringa, M., et al: Systematic Review: The Measurement Properties of the Children's Depression Rating Scale–Revised in Adolescents With Major Depressive Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry .60* (1). 2021, pp 119-133.

92. Streiner, D., Norman, G., & Cairney, J: *Health measurement scales A practical guide to their development and use* (5ed). New York: Oxford University Press, 2015.

93. Terwee, C. B., Bot, S. D., de Boer, M. R., Van der windt, D. A., Knol, D. L., Dekker, J., et al: Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. *60* (1). 2007, pp 34-42.

94. Terwee, C. B., Schellingerhout, J. M., Verhagen, A. P., Koes, B. W., & de Vet, H. C: Methodological Quality of Studies on the Measurement Properties of Neck Pain and Disability Questionnaires: A Systematic Review. *Journal of Manipulative and Physiological*. *34* (4). 2011, pp 261-272.

95. Terwee, C., Mokkink, L., Knol, D., Ostelo, R., Bouter, L., & de Vet, H: Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. *21* (4). 2012, pp 651-657.

96. Terwee, C., Prinsen, C., Chiarotto, A., Westerman, M., Patrick, D., Alonso, J., et al: COSMIN methodology for evaluating

the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research .27* (5). 2018, 1159-1170.

97. Tian, L., Cao, X., & Feng, X: Evaluation of psychometric properties of needs assessment tools in cancer patients: A systematic literature review. *PloS One*. *14* (1). 2019, pp 210-240.

98. Urbina, S: *Essentials of psychological testing.* canada: John Wiley & Sons,Inc, 2004.

99. Urbina, S: *Essentials of psychological testing* (2 ed ed.). United states of America: Wiley, 2014.

100. Vrijman, C., Linthorst Homan, M., Limpens, J., Van der veen, W., Wolkestorfer, A., Terwee, C., et al: Measurement properties of outcome measures for vitiligo. A systematic review. *Arch Dermatol*. *148* (11). 2012, pp 1302-1309.

101. Wittkowski, A., Vatter, S., Muhinyi, A., Garrett, C., & Henderson, M: Measuring bonding or attachment in the parent-infant-relationship: A systematic review of parent-report assessmentmeasures, their psychometric properties and clinical utility. *Clinical Psychology Review .82*. 2020, pp 101-906.

102. Zeilinger, E., Novakovic, I., Komenda, S., Franken, F., Sobisch, M., Mayer, A., et al: Informant-based assessment instruments for dementia in people with intellectual disability: A systematic review and standardised evaluation. *Research in Developmental Disabilities . 121*. 2022, pp 104148.

الملاحق

الملحق 01: شبكة COSMIN للدراسة الحالية

فقط تلك الأجزاء من الصندوق التي يجب إكمالها والتي تتوفر معلومات عنها

التقييمات: V= جيد جدا A = مناسب D = مشكوك فيه I = غير كاف; N= لا ينطبق

	a2. الاتساق الداخلي (هل البنود ملائمة؟)
1	هل تم استخدام طريقة مناسبة لسؤال العينة عما إذا كان كل بند ملائما مع حالتهم؟
2	هل تم اختبار كل بند في عدد مناسب من العينة؟
3	هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة م العينة؟
4	هل كانت الاجتماعات مع العينة أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟
5	هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟
6	هل تم استخدام منهج مناسب لتحليل البيانات؟
7	هل شارك باحثان على الأقل في التحليل؟

	b2- الاتساق الضائع (هل البنود ضائعة؟)
8	هل تم استخدام طريقة مناسبة لتقييم شمولية الاختبار؟
9	هل تم اختبار كل بند في عدد مناسب من العينة؟
10	هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة م العينة؟
11	هل كانت الاجتماعات أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟
12	هل تم تسجيل الاجتماعات أو المقابلات ونسخها حرفيًا؟
13	هل تم استخدام منهج مناسب لتحليل البيانات؟
14	هل شارك باحثان على الأقل في التحليل؟
	c2. سؤال العينة عن فهم الاختبار
15	هل تم استخدام طريقة نوعية مناسبة لتقييم مدى فهم تعليمات اختبار والعناصر وخيارات الاستجابة وفترة الاسترجاع؟
16	هل تم اختبار كل عنصر في عدد مناسب من العينة؟
17	هل تم استخدام أشخاص ذوي خبرة كافية في المقابلة م العينة؟
19	هل كانت الاجتماعات أو المقابلات مبنية على موضوع مدروس أو دليل مقابلة؟
20	هل تم استخدام منهج مناسب لتحليل البيانات؟
21	هل شارك باحثان على الأقل في التحليل؟

	d2. الاتساق الداخلي (هل تعكس البنود بشكل صحيح البناء؟)
22	هل تم استخدام طريقة مناسبة لسؤال المحكمين عما إذا كان كل عنصر مناسبًا لبناء الاهتمام؟
23	هل تم تضمين المحكمين من جميع التخصصات ذات الصلة بالسمة المدروسة؟
24	هل تم اختبار كل بند في عدد مناسب من المحكمين؟
25	هل تم استخدام منهج مناسب لتحليل البيانات؟
26	هل شارك باحثان على الأقل في التحليل؟

	2e. لوحة تحكم خاصة بالاختبار
27	هل تم استخدام طريقة مناسبة لتقييم شمولية الاختبار؟
28	هل تم إدراج محكمين من جميع التخصصات ذات الصلة؟
29	هل تم اختبار كل بند في عدد مناسب من المحكمين؟
30	هل تم استخدام منهج مناسب لتحليل البيانات؟
31	هل شارك باحثان على الأقل في التحليل؟

	3- لوحة التحكم الكمي
1	يحتوي على أبعاد؟
2	بالنسبة إلى النظرية التقليدية CTT: هل تم إجراء تحليل عامل استكشافي أو تأكيدي؟
3	بالنسبة للنظرية الحديثة IRT / RAsch: هل يتناسب النموذج المختار مع سؤال البحث؟
4	هل كان حجم العينة المشمول في التحليل كافياً؟
5	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟
	4- لوحة تحكم الاتساق الداخلي
1	هل تم حساب إحصاء الاتساق الداخلي لكل مقياس أحادي البعد (فرعي) على حدة؟
2	هل تم التحقق من التجزئة النصفية باعتماد إجراءات سليمة
3	هل تم حساب ألفا أو أوميغا من كرونباخ بالنسبة للنتائج المستمرة (متغير متصل)
4	بالنسبة إلى الدرجات ثنائية التفرع: هل تم حساب ألفا كرونباخ أو KR-20؟
5	بالنسبة للدرجات المستندة إلى IRT: هل تم حساب الخطأ القياسي للثيتا (SE (θ)) أو معامل الموثوقية لقيمة السمات الكامنة المقدرة (مؤشر الفصل (الموضوع أو العنصر))؟
6	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟

	5- لوحة التحكم النوعي
1	هل كانت العينات متشابهة بالنسبة في الخصائص الأخرى باستثناء المتغير المدروس؟
2	هل تم استخدام منهج مناسب لتحليل البيانات؟
3	هل كان حجم العينة المدرجة في التحليل كافياً؟
4	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟
	6- لوحة تحكم خاصة بالأبعاد المتعددة
1	هل كانت العينة مستقرة في البناء المراد قياسه بين التطبيقين؟
2	هل كانت الفترة الزمنية مناسبة؟
3	هل كانت ظروف وصور الاختبار مماثلة في التطبيقين؟ على سبيل المثال ظروف التطبيق، البيئة ، التعليمات
4	بالنسبة للدرجات المستمرة: هل تم حساب معامل الارتباط داخل الصف (ICC)؟
5	بالنسبة إلى الدرجات ثنائية التفرع / الاسمية / الترتيبية: هل تم حساب كابا؟
6	بالنسبة إلى الدرجات الترتيبية: هل تم حساب كابا المرجحة؟
7	بالنسبة إلى الدرجات الترتيبية: هل تم وصف مخطط الترجيح؟ على سبيل المثال خطي تربيعي
8	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟

	7- أمراض الكبد
1	هل كانت العينة مستقرة في البناء المراد قياسه بين التطبيقين؟
2	هل كانت الفترة الزمنية مناسبة؟
3	هل كانت ظروف وصور الاختبار مماثلة في التطبيقين؟ على سبيل المثال ظروف التطبيق، البيئة ، التعليمات
4	بالنسبة إلى الدرجات المستمرة: هل تم حساب الخطأ القياسي في القياس (SEM) ، أو أصغر تغيير يمكن اكتشافه (SDC) أو حدود الاتفاقية (LoA)؟
5	بالنسبة إلى الدرجات ثنائية التفرع / الاسمية / الترتيبية: هل تم حساب النسبة المئوية (الإيجابية والسلبية) للاتفاق؟
6	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟

	8- أمراض الغدد الصماء
1	بالنسبة للدرجات المتصلة: هل تم حساب الارتباطات أو المنطقة الواقعة أسفل منحنى تشغيل المستقبل؟
2	بالنسبة للدرجات ثنائية التفرع: هل تم تحديد الحساسية والنوعية؟
3	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟

	9- أمراض القلب
	9 أ. مقارنة مع أدوات قياس النتائج الأخرى (صلاحية متقاربة)
1	هل من الواضح ما تقيسه أداة (أدوات) المقارنة؟
2	هل كانت الخصائص السيكومترية الخاصة بأداة (أدوات) المقارنة كافية؟
3	هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟
4	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟
	9 ب. مقارنة بين المجموعات الفرعية (التمييزية أو صحة المجموعات المعروفة)
1	هل تم تقديم وصف مناسب للخصائص الهامة للمجموعات الفرعية؟
2	هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟
3	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟

	IN- الجهاز التنفسي
	10 أ. نهج المعيار (أي المقارنة بالمعيار الذهبي)
1	بالنسبة إلى الدرجات المستمرة: هل تم حساب الارتباطات بين درجات التغيير ، أو المنطقة الواقعة تحت منحنى مشغل المستقبل (ROC)؟
2	بالنسبة للمقاييس ثنائية التفرع: هل تم تحديد الحساسية والخصوصية (متغيرة مقابل عدم تغييرها)؟
3	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟
	10 ب. نهج البناء (أي اختبار الفرضيات ؛ المقارنة مع أدوات قياس النتائج الأخرى)
4	هل من الواضح ما تقيسه أداة (أدوات) المقارنة؟
5	هل كانت خصائص القياس الخاصة بأداة (أدوات) المقارنة كافية؟
6	هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟
7	هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟

10 ج. نهج البناء: (أي اختبار الفرضيات: المقارنة بين المجموعات الفرعية)	
هل تم تقديم وصف مناسب للخصائص الهامة للمجموعات الفرعية؟	8
هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟	9
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	10
IN د. نهج البناء: (أي اختبار الفرضيات: قبل وبعد التدخل)	
هل تم تقديم وصف مناسب للتدخل؟	11
هل كانت الطريقة الإحصائية مناسبة للفرضيات المراد اختبارها؟	12
هل هناك عيوب مهمة أخرى في طريقة الاستخدام أو الأساليب الإحصائية المستعملة؟	13

الجدول 2. جودة خصائص القياس

الاستجابة					صدق الفرضيات		الصدق المحكي	أخطاء القياس	اعادة التطبيق والصور المتكافئة	الصدق الثقافي	ثبات الاتساق الداخلي	صدق الهيكلي	صدق المحتوى					الاختبارات
													سؤال المختصين		سؤال العينة			
المقارنة القبلية البعدية	المقارنة بين المجموعات	المقارنة مع أدوات اخرى	صدق المعيار الذهبي		المقارنة بين المجموعات	المقارنة مع أدوات أخرى							الشمولية	ملاءمة	الفهم	الشمولية	ملاءمة	

الجدول 3. جودة الدليل لخصائص القياس للاختبار

	اختبار		مقياس اختبار	
	تقييم شامل	**جودة الأدلة**	**تقييم شامل**	**جودة الأدلة**
	+ / − / ?	**مرتفع ، متوسط ، منخفض ، منخفض جدا**	**+ / − / ?**	**مرتفع ، متوسط ، منخفض ، منخفض جدا**
صلاحية المحتوى				
الملاءمة				
الشمولية				
الفهم				
صدق البناء				
ثبات الاتساق الداخلي				
الصلاحية عبر الثقافات				
اعادة التطبيق والصور المتكافئة				
خطأ في القياس				
معيار صحة				
صدق الفرضيات				
إستجابة				

الملحق 02: الشبكة COSMIN الأصلية

COSMIN box 1. Standards for evaluating the quality of PROM development

Ratings: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

1a. PROM design

General design requirements	
1	Is a clear description provided of the construct to be measured?
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?
3	Is a clear description provided of the target population for which the PROM was developed?
4	Is a clear description provided of the context of use (i.e. discriminative, evaluative purpose, and/or predictive)
5	Was the PROM development study performed in a sample representing the target population for which the PROM was developed?
Concept elicitation (relevance and comprehensiveness)	
6	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?
7	Were skilled group moderators/ interviewers used?
8	Were the group meetings or interviews based on an appropriate topic or interview guide?
9	Were the group meetings or interviews recorded and transcribed verbatim?
10	Was an appropriate approach used to analyse the data?
11	Was at least part of the data coded independently?
12	Was data collection continued until saturation was reached?
13	For quantitative studies: was the sample size appropriate?
	SUBTOTAL QUALITY CONCEPT ELICITATION STUDY *Lowest score of items 6-13*

	TOTAL QUALITY OF THE PROM DESIGN *Lowest score of items 1-13*

1b. Cognitive interview study or other pilot test

14	Was a cognitive interview study or other pilot test performed? *If NO skip items 15–35*

	General design requirements
15	Was the cognitive interview study or other pilot test performed in a sample representing the target population?
	Comprehensibility
16	Were patients asked about the comprehensibility of the PROM? *If NO or not clear, skip items 17–25*
17	Were all items tested in their final form?
18	Was an appropriate qualitative method used to assess the comprehensibility of the PROM instructions, items, response options, and recall period?
19	Was each item tested in an appropriate number of patients?
20	Were skilled interviewers used?
21	Were the interviews based on an appropriate interview guide?
22	Were the interviews recorded and transcribed verbatim?
23	Was an appropriate approach used to analyse the data?
24	Were at least two researchers involved in the analysis?
25	Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM?
	SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY *Lowest score of items 15–25*

	Comprehensiveness
26	Were patients asked about the comprehensiveness of the PROM? *If NO or not clear, skip items 27–35*
27	Was the final set of items tested?
28	Was an appropriate method used for assessing the comprehensiveness of the PROM?

29	Was each item tested in an appropriate number of patients?
30	Were skilled interviewers used?
31	Were the interviews based on an appropriate interview guide?
32	Were the interviews recorded and transcribed verbatim?
33	Was an appropriate approach used to analyse the data?
34	Were at least two researchers involved in the analysis?
35	Were problems regarding the comprehensiveness of the PROM appropriately addressed by adapting the PROM?
	SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY *Lowest score of items 15, 26-35*

	TOTAL QUALITY OF THE PILOT STUDY *Lowest score of items 14-35*

	TOTAL QUALITY OF THE PROM DEVELOPMENT STUDY *Lowest score of items 1-35*

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

	2a. Asking patient about relevance
1	Was an appropriate method used to ask patients whether each item is relevant for their experience with the condition?
2	Was each item tested in an appropriate number of patients?
3	Were skilled group moderators/interviewers used?
4	Were the group meetings or interviews based on an appropriate topic or interview guide?
5	Were the group meetings or interviews recorded and transcribed verbatim?
6	Was an appropriate approach used to analyse the data?
7	Were at least two researchers involved in the analysis?
	SUBTOTAL QUALITY OF RELEVANCE STUDY *Lowest score of items 1-7*
	2b. Asking patients about comprehensiveness
8	Was an appropriate method used for assessing the comprehensiveness of the PROM?
9	Was each item tested in an appropriate number of patients?
10	Were skilled group moderators/interviewers used?
11	Were the group meetings or interviews based on an appropriate topic or interview guide?

12 Were the group meetings or interviews recorded and transcribed verbatim?
13 Was an appropriate approach used to analyse the data?
14 Were at least two researchers involved in the analysis?

SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY *Lowest score of items 8–14*

2c. Asking patients about comprehensibility

15 Was an appropriate qualitative method used for assessing the comprehensibility of the PROM instructions, items, response options, and recall period?
16 Was each item tested in an appropriate number of patients?
17 Were skilled group moderators/interviewers used?
18 Were the group meetings or interviews based on an appropriate topic or interview guide?
19 Were the group meetings or interviews recorded and transcribed verbatim?
20 Was an appropriate approach used to analyse the data?
21 Were at least two researchers involved in the analysis?

SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY *Lowest score of items 15–21*

2d. Asking professionals about relevance

22 Was an appropriate method used to ask professionals whether each item is relevant for the construct of interest?
23 Were professionals from all relevant disciplines included?
24 Was each item tested in an appropriate number of professionals?
25 Was an appropriate approach used to analyse the data?
26 Were at least two researchers involved in the analysis?

SUBTOTAL QUALITY OF RELEVANCE STUDY *Lowest score of items 22–26*

2e. Asking professionals about comprehensiveness

27 Was an appropriate method used for assessing the comprehensiveness of the PROM?
28 Were professionals from all relevant disciplines included?
29 Was each item tested in an appropriate number of professionals?
30 Was an appropriate approach used to analyse the data?
31 Were at least two researchers involved in the analysis?

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

Article reference:

	3. Structural validity
	unidimensionality or structural validity?
1	For CTT: Was exploratory or confirmatory factor analysis performed?
2	For IRT/Rasch: does the chosen model fit to the research question?
3	Was the sample size included in the analysis adequate?
4	Were there any other important flaws?
	TOTAL *Lowest score of items 1-4*
	4. Internal consistency
1	Was an internal consistency statistic calculated for each unidimensional (sub)scale separately?
2	For continuous scores: Was Cronbach's alpha or omega calculated?
3	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?
4	For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?
5	Were there any other important flaws?
	TOTAL *Lowest score of items 1-5*

	5. Cross-cultural validity|measurement invariance
1	Were the samples similar for relevant characteristics except for the group variable?
2	Was an adequate approach used to analyse the data?
3	Was the sample size included in the analysis adequate?
4	Were there any other important flaws?
	TOTAL *Lowest score of items 1-4*

	6. Reliability
1	Were patients stable in the interim period on the construct to be measured?
2	Was the time interval appropriate?

3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions
4	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?
5	For dichotomous/nominal/ordinal scores: Was kappa calculated?
6	For ordinal scores: Was a weighted kappa calculated?
7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic
8	Were there any other important flaws?
	TOTAL *Lowest score of items 1–8*

7. Measurement error

1	Were patients stable in the interim period on the construct to be measured?
2	Was the time interval appropriate?
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions
4	For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?
5	For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?
6	Were there any other important flaws?
	TOTAL *Lowest score of items 1–6*

8. Criterion validity

1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?
2	For dichotomous scores: Were sensitivity and specificity determined?
3	Were there any other important flaws?
	TOTAL *Lowest score of items 1–3*

9. Hypotheses testing for construct validity

9a. Comparison with other outcome measurement instruments

(convergent validity)

1 Is it clear what the comparator instrument(s) measure(s)?

2 Were the measurement properties of the comparator instrument(s) adequate?

3 Was the statistical method appropriate for the hypotheses to be tested?

4 Were there any other important flaws?

TOTAL *Lowest score of items 1–4*

9b. Comparison between subgroups (discriminative or known-groups validity)

5 Was an adequate description provided of important characteristics of the subgroups?

6 Was the statistical method appropriate for the hypotheses to be tested?

7 Were there any other important flaws?

TOTAL *Lowest score of items 5–7*

10. Responsiveness

10a. Criterion approach (i.e. comparison to a gold standard)

1 For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?

2 For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?

3 Were there any other important flaws?

TOTAL *Lowest score of items 1–3*

10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)

4 Is it clear what the comparator instrument(s) measure(s)?

5 Were the measurement properties of the comparator instrument(s) adequate?

6 Was the statistical method appropriate for the hypotheses to be tested?

7 Were there any other important flaws?

TOTAL *Lowest score of items 4–7*

10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)

8 Was an adequate description provided of important characteristics of the subgroups?

9 Was the statistical method appropriate for the hypotheses to be tested?

10 Were there any other important flaws?

TOTAL *Lowest score of items 8-10*

10d. Construct approach: (i.e. hypotheses testing: before and after intervention)

11 Was an adequate description provided of the intervention given?

12 Was the statistical method appropriate for the hypotheses to be tested?

13 Were there any other important flaws?

PROM **Study 1**

	RATING	**RATING**	**RATING**
	+ / − / ?	+ / − / ?	+ / − / ?
	rater 1	**rater 2**	**Consensus**
Structural validity			
Internal consistency			
Cross-cultural validity			
Measurement invariance			
Reliability			
Measurement error			
Criterion validity			
Construct validity			
Responsiveness			

الملحق 03: جداول ملخصات المعايير

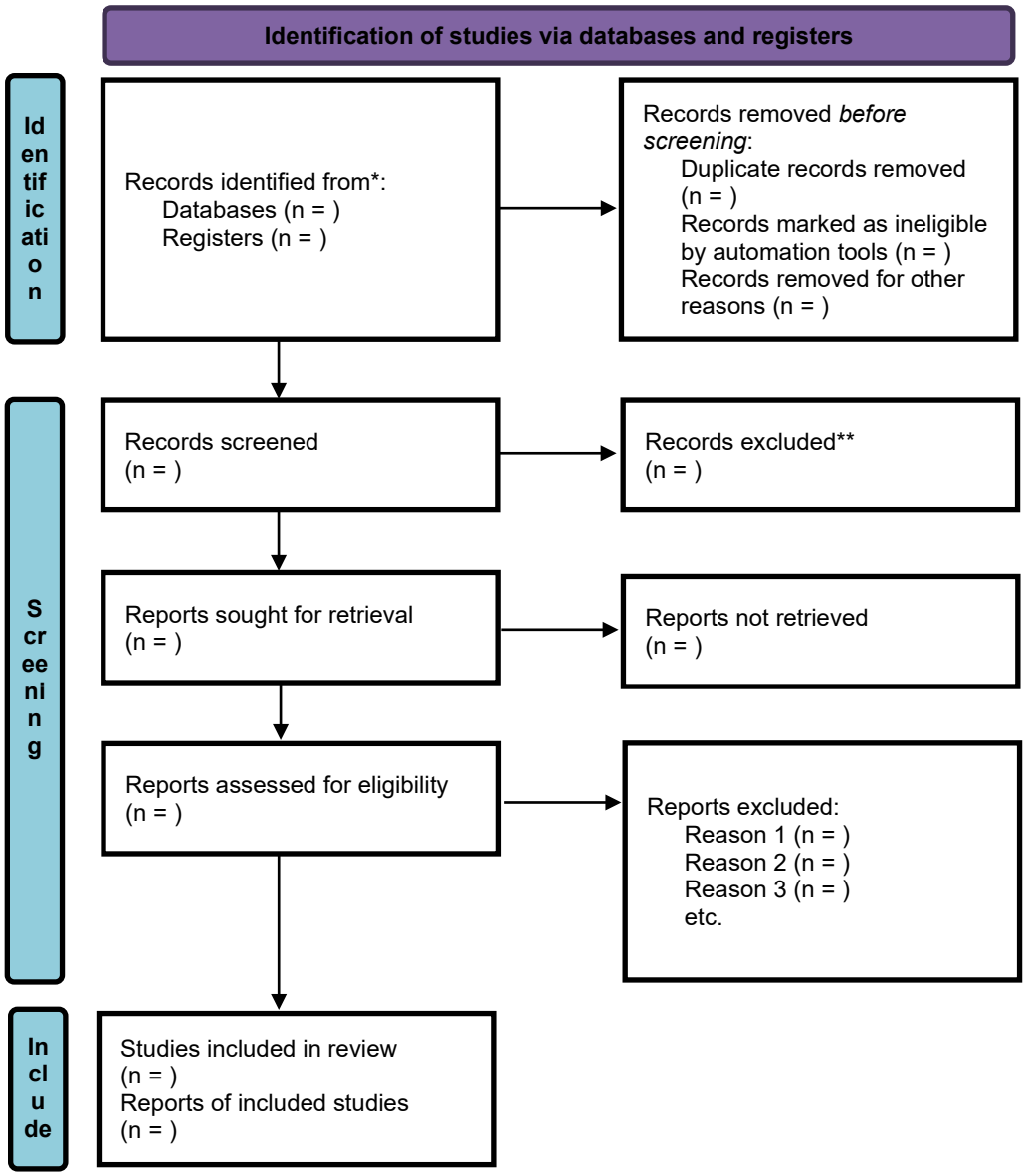
Measurement property	Rating[1]	Criteria
Structural validity	+	CTT: CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08[2] IRT/Rasch: No violation of unidimensionality[3]: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 AND no violation of local independence: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 AND no violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND

		adequate model fit: IRT: $\chi^2 > 0.01$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > -2 and <2
	?	CTT: Not all information for '+' reported IRT/Rasch: Model fit not reported
	–	Criteria for '+' not met
Internal consistency	+	At least low evidence[4] for sufficient structural validity[5] AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale[6]
	?	Criteria for "At least low evidence[4] for sufficient structural validity[5]" not met
	–	At least low evidence[4] for sufficient structural validity[5] AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale[6]

Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	–	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA < MIC[5]
	?	MIC not defined SDC or LoA > MIC[5]
	–	
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis[7]
	?	No hypothesis defined (by the review team)
	–	The result is not in accordance with the hypothesis[7]
Cross-cultural validity\measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
	?	No multiple group factor analysis OR DIF analysis performed
	–	Important differences between group factors OR DIF was found
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70

	?	Not all information for '+' reported
	–	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	+	The result is in accordance with the hypothesis[7] OR AUC ≥ 0.70
	?	No hypothesis defined (by the review team)
	–	The result is not in accordance with the hypothesis[7] OR AUC < 0.70

الملحق 04: منهجية PRISMA الأصلية



الملحق 05: جدول مواصفات عينة الدراسة

اسم الاختبار	سنة إعداد الاختبار	سنة تطبيق الاختبار	مجتمع الدراسة الأصلي	مجتمع الدراسة الحالي	نوع الدراسة أو الغرض من الاختبار	زمن تطبيق الاختبار	عدد الأبعاد أو المحاور أي المقاييس الفرعية	خيارات الاستجابة	عدد البنود في الدراسة الحالية	عدد البنود في الدراسة الأصلية	اللغة الأصلية للاختبار	النسخة المستخدمة: المترجمة/ الأصلية/ المستحدثة/ المكيفة	تم التعديل	إعادة التحقق بعد التعديل
مقياس أساليب مواجهة الضغط	1985/ 2007	2012	والدا الأطفال المعاقين	والدا الأطفال التوحديين	تشخيصية	/	5	5	47	42	الانجليزية	المترجمة	نعم	نعم
مقياس التوافق النفسي والاجتماعي	1939/ 1998	2017	أطفال كالفورنيا/ أطفال مصر	أطفال الجالية العراقية في الدنمارك	علائقية	/	2	2	44	50	الانجليزية	المترجمة	نعم	لا
مقياس التوافق النفسي	1934/ 1960/ 1996	2012	طلبة الثانوية والجامعة الأمريكية	طلبة الجالية السودانية في قطر	علائقية	/	4	2	58	160	الانجليزية	المكيفة	نعم	لا
مقياس السلوك العدواني	2003	2017	المراهقين	طلبة الثانوية	علائقية	/	3	5	40	40	العربية	الأصلية	نعم	لا

مقياس السلوك العدواني	2009	2017	/	SOSأطفال قرية	تشخيصية	/	3	5	46	48	العربية	الأصلية	نعم	نعم
اختبار عوامل الشخصية	/1957 1971	2017	طلبة الإعدادية والثانوية	المراهقين	علائقية	45-60د	16	3 ماعدا في عامل الذكاء 2	187	187	الانجليزية	المترجمة	لا	لا
مقياس الكفاية الذاتية المدركة	2008	2013	/	الأحداث	تشخيصية	/	4	5	33	33	العربية	الأصلية	لا	لا
مقياس الغضب	/1995 2002	2013	طلبة جامعيين	الأحداث	تشخيصية	/	/	2	30	30	الانجليزية	المكيفة	لا	لا
مقياس الوعي الذاتي	2011	2015	/	طلبة جامعيين	علائقية	/	2	5	26	26	العربية	الأصلية	لا	لا
مقياس الحاجات الارشادية	2006	2013	المراهقين	طالبات الإعدادية	علائقية	/	/	2	27	27	العربية	الأصلية	لا	لا
قائمة بيك للاكتئاب	/1979 1998	2015	/	طلبة الجامعة	علائقية	/	/	4	21	21	الانجليزية	المترجمة	لا	لا
اختبار واطسن- جليسر للتفكير الناقد	1994	2013	/	طالبات الإعدادي	تشخيصية	30د	5	2	40	40	الانجليزية	المترجمة	نعم	نعم

مقياس الاسلوب المعرفي	2003	2014	طلبة الإعدادية	طلبة الإعدادية	علائقية	15د	/	2	37	37	العربية	الأصلية	نعم	لا
اختبار آرمسترونغ للذكاء اللغوي	1984	2017	تلاميذ الإعدادية	/	تشخيصية علاجية	/	/	3	14	14	الانجليزية	المترجمة	نعم	نعم
اختبار الانتباه	1970/ 2008	2012	تلاميذ المرحلة الإعدادية	/	تشخيصية علاجية	25د	/	2	25	25	الانجليزية	المترجمة	لا	لا
بطارية كاتل	1963/ 1997	2019	طلاب المرحلة الثانوية	/	تشخيصية	45-50د	2	2	92	92	الانجليزية	المترجمة	لا	لا
استبيان نصوع التخيل البصري	1973	2012	تلاميذ المرحلة الإعدادية	/	تشخيصية	/	/	5	16	16	الانجليزية	المترجمة	نعم	نعم
مقياس الحاجات النفسية	2014	2015	الطلبة الموهوبين في المرحلة الثانوية	طلبة التعليم ما بعد أساسي	علائقية	/	/	3	47	47	العربية	الأصلية	لا	لا
مقياس مفهوم الذات	1964/ 1982	2015	طلبة الجامعة	/	علائقية	/	6	5	80	80	الانجليزية	المترجمة	نعم	نعم
مقياس الكفاية الاجتماعية	1995/ 2012	2014	المعلمات المطلقات والمرملات	/	علائقية	/	3	5	42	42	الانجليزية	المترجمة	نعم	نعم
مقياس أنماط	1962/	2015	المتفوقين في المرحلة	/	علائقية	/	4	2	28	28	الانجليزية	المترجمة	نعم	نعم

الشخصية الثمانية	2013			الثانوية									
اختبار الدافع للإنجاز للأطفال الراشدين	1970/ 1991	2015	الأطفال والراشدين	معلمي صعوبات التعلم	تشخيصية	/	/	5	28	28	الانجليزية	المترجمة	لا	لا
اختبار ستروب لتسمية الألوان	1935/ 2003	2012	/	الأطفال التوحديين والأطفال المعاقين ذهنيا	مقارنة	25ث لكل بطاقة	3	2	300	300	الانجليزية	المستحدثة	لا	لا
اختبار وكسلر للذاكرة	1967/ 2008	2013	/	طلاب الكليات	علائقية	/	7	/	/	/	الانجليزية	المكيفة	نعم	لا
مقياس مفهوم الذات	1992	2012	/	الوالدي	علائقية	/	4	3	50	44	العربية	الأصلية	نعم	لا
مقياس أساليب معاملة الأمهات	1991	2012	الأباء والأمهات	الأمهات	علائقية	/	5	3	146	75	العربية	الأصلية	نعم	نعم
مقياس لبعض أشكال العنف الأسري	2010	2012	طلاب المدارس الثانوية	طلاب المرحلة الثانوية	علائقية	/	4	5	41	58	العربية	الأصلية	نعم	لا
اختبار أساليب المعاملة الوالدية	1991	2013	طلاب المدارس الثانوية	تلاميذ مرحلة الأساس	علائقية	/	5	3	54	38	العربية	الأصلية	نعم	نعم
مقياس الرضا عن	1982	2013	/	المسنين	علائقية	/	/	3	13	9	العربية	الأصلية	نعم	لا

الحياة عند المسنين														
مقياس الرضا عن الحياة	2001	2015	/	المتزوجين	تشخيصية	/	5	4	40	40	العربية	الأصلية	لا	لا
مقياس أنماط حل المشكلات	2002/ 2013	2015	/	زوجات مرضى الفصام العقلي	علاجية	/	2	5	24	25	الانجليزية	المترجمة	نعم	نعم
مقياس بيك للاكتئاب	1988/ 1996	2015	/	النساء	علاجية	/	/	4	21	21	الانجليزية	المترجمة	لا	لا
مقياس بيك للاكتئاب	/	2014	/	جرحى الحروق	علائقية	/	/	4	21	21	الانجليزية	المترجمة	لا	لا
مقياس هوية الذات	2012	2015	/	طلبة الثانوية	علائقية	/	/	5	42	42	العربية	الأصلية	لا	لا
مقياس المناخ الأسري	2013	2016	طلبة الصف التاسع	النساء السجينات	علائقية	/	8	5	98	100	العربية	الأصلية	نعم	لا
مقياس غزة للخبرات الصادمة	2014	2016	/	الأطفال الأيتام وأمهاتهم	علائقية	/	/	2	29	29	العربية	الأصلية	لا	لا
مقياس بيك للاكتئاب	1988/ 2014	2015	/	أمهات المصابين بتشوه خلقي	علاجية	/	/	4	21	21	الانجليزية	المترجمة	لا	لا
مقياس الأفكار	1987	2014	/	أفراد الشرطة	علائقية	/	13	5	52	52	العربية	الأصلية	لا	لا

اللاعقلانية													
مقياس الدعم الاجتماعي المدرك	2015	2016	الناجيين من السكتة الدماغية في نيجيريا	المطلقات	علائقية	/	3	5	12	14	الانجليزية	الأصلية	نعم	نعم
مقياس الهوية النفسية	1989/ 1994	2015	المتزوجين والعازبين	طلبة الجامعة	علائقية	/	2	6	64	55	الانجليزية	المترجمة	نعم	نعم
اختبار كالفورنيا لمهارات التفكير الناقد	2003	2015	/	طلبة الجامعة	علائقية	/	5	2	34	34	الانجليزية	المترجمة	لا	لا
مقياس السلوك الصحي	1999	2016	/	طلبة الصف السابع والثامن والتاسع	علائقية	/	3	4	21	21	العربية	الأصلية	لا	لا
اختبار تورانس للتفكير الابداعي	1992/ 2016	2018	/	/	علائقية	10د لكل بعد	3	بدائل	7	7	الانجليزية	المترجمة	لا	لا
مقياس العجز المتعلم	2006	2014	/	طلاب المرحلة المتوسطة	علائقية	/	3	5	30	29	العربية	الأصلية	نعم	نعم
مقياس الذكاء الانفعالي	2003	2015	طلبة	طلبة الجامعة	علائقية	/	5	5	55	55	العربية	الأصلية	لا	لا
مقياس قلق العولمة	2014	2017	/	طلبة الجامعة	علائقية	/	/	5	30	30	العربية	الأصلية	لا	لا

مقياس النضج(النمو) الانفعالي	1973	2015	13 سنة فما فوق	طلبة الجامعة	علائقية	60-90د	6	ترتيب الفقرات	/	/	الانجليزية	المترجمة	لا	لا
مقياس التكيف النفسي والاجتماعي	1988	2015	/	طلبة الجامعة	علائقية	/	4	3	76	76	العربية	الأصلية	لا	لا
مقياس العوامل الخمسة الكبرى للشخصية	1992/ 2002	2016	/	الطلبة العمانيين بالجامعة الأردنية	تشخيصية	/	5	5	60	59	الانجليزية	المترجمة	نعم	لا
مقياس الاكتئاز القهري	2004	2017	/	طلبة الثانوية	علاجية	/	/	5	23	23	الانجليزية	الأصلية	نعم	نعم
مقياس الاكتئاب للأطفال والمراهقين	2011	2012	/	المراهقين	علائقية	/	8	3	40	40	العربية	الأصلية	لا	لا
مقياس ماسلاش للاحتراق النفسي	1981/؟	2017	/	معلمي الأقسام النهائية	تشخيصية	/	3	7	22	22	الانجليزية	المترجمة	لا	لا
قائمة إكسفورد للسعادة	1989/ 2010	2014	/	طلبة الجامعة	علائقية	/	/	4	29	29	الانجليزية	المترجمة	لا	لا
اختبار آيزنك للشخصية	1975/ 1991	2016	للراشدين والأطفال	الجانحين بالشرق الجزائري	تشخيصية	/	4	2	91	91	الانجليزية	المترجمة	لا	لا

اختبار أنماط السيادة النصفية للمخ	/1984 1988	2016	طلبة الدراسات العليا في البيئة الأمريكية	المراهقين	علائقية	/	3	3 بدائل	40	40	الانجليزية	المترجمة	لا	لا
مقياس الأفكار اللاعقلانية	1985	2017	/	المدمنين على المخدرات	علاجية	/	3	2	52	52	الانجليزية	المترجمة	لا	لا
مقياس نوعية العلاقة الزوجية	1991	2014	/	الأزواج	علائقية	/	3	4	25	25	الانجليزية	الأصلية	نعم	نعم
مقياس أبعاد السلفة الوالدية	2000	2014	المراهقين	طلاب الجامعة	علائقية	/	6	3	56	56	العربية	الأصلية	نعم	نعم
استبيان نمط الشخصية	1988	2012	/	الأطباء اليمنيين	علائقية	/	/	5	30	30	الانجليزية	المترجمة	نعم	نعم
مقياس أساليب مواجهة الضغوط	2006	2014	/	الأساتذة الجامعيين	علائقية	/	15	3	60	60	الانجليزية	المترجمة	لا	لا
مقياس حل المشكلات	1997	2016	/	تلاميذ الثانوية	تشخيصية	/	5	4	40	40	العربية	الأصلية	لا	لا
مقياس التكيف النفسي والاجتماعي	2010	2017	/	ممارسي وغير ممارسي السباحة	مقارنة	/	2	5	46	46	العربية	الأصلية	لا	لا

مقياس الهوية النفسية	2011	2020	/	أمهات طالبات العاشر أساسي	مقارنة	/	8	3	64	40	العربية	الأصلية	نعم	نعم
مقياس استخدام مواقع التواصل الاجتماعي	2017	2019	/	طلاب	علائقية	/	/	3	18	18	العربية	الأصلية	نعم	لا
مقياس اتخاذ القرار	1979/ 2011	2020	/	طلبة جامعة مؤتة	علائقية	/	/	5	33	27	العربية	المستحدثة	نعم	نعم
مقياس الاحتراق النفسي لمديري المدارس	1995	2019	مديري المدارس	مديري مدارس	علائقية	/	3	5	23	23	الانجليزية	المترجمة	نعم	نعم
قائمة تشخيص الاكتئاب	2004	2019	/	مرضى السكري	علائقية	/	2	5	31	31	الانجليزية	الأصلية	لا	لا
مقياس الاحتراق النفسي	1977/ 1998	2019	/	حكام كرة القدم	تشخيصية	/	/	7	15	15	الانجليزية	المترجمة	لا	لا
مقياس الاعتقاد بالكفاءة الذاتية	2007	2017	/	الإعاقة البصرية	علائقية	/	4	؟	32	32	العربية	الأصلية	نعم	لا
مقياس معنى الحياة	2007	2019	/	طلبة الجامعة	علائقية	/	4	4	20	20	العربية	الأصلية	لا	لا
مقياس الذكاء	2016	2018	/	طلبة الجامعة	علائقية	/	/	5	42	42	العربية	الأصلية	لا	لا

الأخلاقي													
مقياس الاتجاه	؟	2016	/	طلبة الجامعة	علائقية	/	/	5	26	26	الانجليزية	المترجمة	نعم	نعم
مقياس الشعور بالوحدة النفسية	/1982 1988	2017	/	طالبات الجامعة	علائقية	/	/	4	20	20	الانجليزية	المترجمة	لا	لا
مقياس الوساوس القهرية	/1977 1992	2021	/	طلاب الجامعة	تشخيصية	/	/	2	32	32	الانجليزية	المترجمة	لا	لا
مقياس التشاؤم الدفاعي	2008	2021	/	طلبة الجامعة	علائقية	/	/	5	24	24	العربية	الأصلية	لا	لا
مقياس النموذج العوامل الستة	/2006 2013	2021	/	طلبة الجامعة	تشخيصية	25د	6	5	42	42	الانجليزية	المترجمة	لا	لا
مقياس الأمن العاطفي	2020	2021	/	طلبة الجامعة	علائقية	/	/	5	32	32	العربية	الأصلية	لا	لا
اختبار تورانس للتفكير الإبداعي	1992	2018	/	تلامذة الروضة	علائقية	10-15د	3	/	/	/	الانجليزية	الأصلية	لا	لا
مقياس مهارات التفكير ما وراء المعرفي	2012	2017	/	طلبة الجامعة	علائقية	/	3	3	30	30	العربية	الأصلية	لا	لا

مقياس المرونة النفسية	2003/ 2010	2016	/	طلبة الجامعة	علائقية	/	4	5	25	25	الانجليزية	المترجمة	لا	لا
مقياس الاغتراب النفسي	2012	2016	/	طلبة الجامعة	علائقية	/	7	5	50	50	العربية	الأصلية	لا	لا
مقياس حاجات الأطفال العمانيين	2006	2016	/	الأطفال العمانيين	تشخيصية	/	6	5	69	89	العربية	الأصلية	نعم	نعم
مقياس المسايرة الاجتماعية	2011	2016	/	الطلبة المراهقين	علائقية	/	/	3	16	12	العربية	الأصلية	نعم	نعم
مقياس هولاند للذكاء المهني	2015/؟	2019	/	طلبة الصف العاشر	علائقية	/	6	2	120	120	الانجليزية	المترجمة	لا	لا
مقياس الذكاء العاطفي	2010	2019	/	المرشدين النفسيين	علائقية	/	5	5	55	55	العربية	الأصلية	لا	لا
اختبار التحمل النفسي	2003	2022	/	لاعبي كرة القدم	علائقية	/	/	4	30	30	العربية	الأصلية	لا	لا
مقياس الاتصال الزواجي	2014	2021	/	الأزواج	علائقية	/	3	3	81	81	العربية	الأصلية	لا	لا
Insomnia	2016/ 2017	2019	/	موظفات في المستشفى	تشخيصية	/	2	3	8	8	الانجليزية	المترجمة	لا	لا

Mental Health Screening Tool for Depressive Disorders	2018/ 2019	2021	/	مواطنون كوريون	تشخيصية	/	/	5	12	12	الكورية	المستحدثة	لا	لا
Five-Factor Inventory	1992	2020	/	معلمي اللغة الانجليزية	تشخيصية	/	5	؟	60	12	الانجليزية	الأصلية	نعم	لا
Peer-rating of Victimisation	1985	2014	ضحايا التنمر	ضحايا التنمر	تشخيصية	/	3	5	39	39	الانجليزية	الأصلية	لا	لا
Parent Achievement Orientation	1994	2016	11-17 سنة	تلاميذ المرحلة الثانوية	علائقية	15د	3	5	24	8	الانجليزية	الأصلية	نعم	نعم
Social Situation Questionnaire	1982	2016	/	طلبة	علاجية	/	/	7	41	16	الانجليزية	الأصلية	نعم	نعم
Brief Cope	1997	2016	/	أطفال التوحد	علائقية	/	14	4	28	28	الانجليزية	الأصلية	نعم	نعم
PTSD Symptoms and severity	1993	2016	/	البالغين	علائقية	/	3	5	17	17	الانجليزية	الأصلية	لا	لا

Locus of control	1966	2017	/	طلاب الجامعة تخصص علم النفس	علائقية	30د أو أقل	/	2	29	29	الانجليزية	الأصلية	لا	لا
Dyadic Adjustment Scale	/1976 2005	2019	/	الآباء	تشخيصية	/	/	6	4	4	الانجليزية	الأصلية	لا	لا
Depressive Symptoms	1977	2014	الكبار من ذوي الاحتياجات الخاصة والبالغين والطلاب	المعرضين للانتحار	تشخيصية	/	/	3	20	20	الانجليزية	الأصلية	لا	لا
Police Stress	1994	2015	/	موظفي الشرطة	علائقية	/	2	5	86	126	الانجليزية	الأصلية	نعم	لا
Mindful Attention Awareness scale for children	2004	2015	/	الأطفال	علائقية	/	/	6	15	15	الانجليزية	الأصلية	نعم	لا
Initial Client Impression	2004	2012	/	طلبة دراسات عليا في علم النفس	تشخيصية	/	4	7	13	17	الانجليزية	الأصلية	نعم	نعم

														Inventory
لا	لا	الأصلية	الانجليزية	10	10	5	2	/	علائقية	المصابين بالسرطان والألم العضلي	/	2021	2013	Life Orientation scale
لا	لا	الأصلية	الانجليزية	52	52	2	3	/	علائقية	طلاب الجامعة	/	2017	2001	Self Harm Behaviors Questionnaire
لا	لا	الأصلية	الانجليزية	؟	؟	3	3	/	علائقية	أمهات الأطفال في الحضانة	/	2020	2008	The Social Improvement System Rating Scales
لا	لا	الأصلية	الانجليزية	20	20	7	4	/	علائقية	طلاب الجامعات من الأقلية العرقية	/	2012	2003	Cultural Intelligence Scale
لا	لا	الأصلية	الانجليزية	5	5	4	/	/	علائقية	الشباب الذين عاشوا في دار رعاية	/	2012	2008	Self Image and Social engagement
لا	لا	الأصلية	الانجليزية	16	16	6	3	20د	علائقية	البالغين البيض	/	2018	2004	Psychosocial costs of racism

													to whites scale
نعم	نعم	المترجمة	الانجليزية	10	10	4	/	/	علائقية	البالغين الألبانيين	/	2019	1995	Self efficacy
نعم	نعم	الأصلية	الانجليزية	15	30	بدائل	/	15-25د	تشخيصية	طلاب علم النفس	/	2018	2008	Numeracy
لا	لا	الأصلية	الانجليزية	45	45	5	3	/	مقارنة	طلبة جامعيين من ذوي الاحتياجات الخاصة والعاديين	/	2017	1996	Outcome Questionnaire
لا	لا	الأصلية	الانجليزية	36	36	5	4	/	تشخيصية	تلاميذ	/	2016	2011	Devereux Student Strengths Assessment
لا	لا	الأصلية	الانجليزية	30	30	6	5	/	علائقية	مشاركين عبر الانترنيت	/	2020	2009	Moral Foundation Questionnaire
لا	لا	الأصلية	الانجليزية	10	10	7	2	/	تشخيصية	البالغين	/	2021	2011	Interpersonal Needs Questionnaire
لا	لا	الأصلية	الانجليزية	60	60	5	2	/	علائقية	تلاميذ المتوسط	/	2013	1994	Positive and Negative Affect

														Schedule Form
لا	لا	الأصلية	الانجليزية	16	16	3	2	8-16د	علائقية	دكاترة متدربين في علم النفس	/	2012	2010	Cannabs Use Problems Identification Test
لا	لا	الأصلية	الانجليزية	16	16	5	2	/	علائقية	طلبة علم النفس	/	2016	1975	Evaluation Apprehension
لا	لا	الأصلية	الانجليزية	35	35	5	/	20د	علائقية	ضباط الشرطة	/	2017	2014	Police Daily Hassles and Uplifts Scale
لا	لا	الأصلية	الفرنسية	45	45	5	/	10د	تشخيصية	النساء المغتصبات في الصغر	/	2018	2013	Détresse psychologique
لا	لا	الأصلية	الفرنسية	23	23	5	2	/	تشخيصية	التلاميذ	/	2019	1985	Anxiété et dépression
لا	لا	المترجمة	الانجليزية	11	11	3	3	40د	علائقية	المدربين	/	2014	/2004 2014	Le Coach-Athlete Relationship Questionnaire

l 'Échelle des expériences sexuelles d'Arizona	/2000 2011	2014	/	البالغين	تشخيصية	/	/	6	6	6	الانجليزية	المترجمة	نعم	لا
Échelle des comportementint erpersonnels du thérapeute	2000/؟	2016	/	المعالجين النفسيين	علائقية	/	3	7	35	12	الانجليزية	المترجمة	نعم	لا
L’anxiété état trait a été mesurée avec l’Inventaire d’Anxiété état–trait	1993/؟	2017	/	النساء المعنفات	تشخيصية	10د	2	4	20	20	الانجليزية	المترجمة	نعم	لا
Values In Action Inventory Strengths	2004/؟	2013	/	البالغين ذوي الاضطرابات العقلية	علاجية	/	24	5	240	240	الانجليزية	المترجمة	نعم	لا
Marlow Crowne Social Desirability	/1960 1982	2018	/	مرضى السرطان	تشخيصية	/	/	2	13	13	الانجليزية	المترجمة	نعم	لا

													Scale
لا	لا	الأصلية	الانجليزية	14	14	5	/	/	علاجية	مرضى القلق النفسي	/	2020	1993	Quality of life.
لا	لا	المترجمة	الانجليزية	12	12	5	/	/	علائقية	الأمهات	/	2018	/2007 2012	Intolerance of uncertainty
نعم	نعم	المترجمة	الانجليزية	10	10	4	2	/	علائقية	والدي الأطفال المصابين بالعمى	/	2013	/1965 2000	The Rosenberg self-esteem
لا	لا	المترجمة	الانجليزية	39	39	5	2	/	علائقية	أسر أطفال ما قبل التمدرس	/	2021	/1995 2002	The authoritative and authoritarian parenting styles
لا	لا	الأصلية	الانجليزية	10	10	5	/	/	علائقية	المراهقين	المراهقين والبالغين	2012	2009	The Identity Distress Survey
لا	لا	المترجمة	الانجليزية	13	13	3	/	/	علائقية	المراهقين مرضى السكري	/	2021	/1972 2007	Symptoms of depression and anxiety
لا	نعم	المترجمة	الانجليزية	27	27	5	4	/	علائقية	الممرضات	/	2016	/1983	Interpersonal

Reactivity Index	2010		/										
Self-esteem	/1965 1982	2020	/	طلبة الجامعة	علائقية	/	/	5	10	10	الانجليزية	المترجمة	لا	لا
Paternal postpartum depression	/1987 2003	2021	/	اضطراب ما بعد الولادة	علائقية	/	/	4	10	10	الانجليزية	المترجمة	لا	لا
Psychological flexibility	/2011 2017	2018	/	الرياضيين	علائقية	/	/	7	7	7	الانجليزية	المترجمة	لا	لا
Active concealment	/1987 2014	2017	/	ذوي الاضطرابات النفسية	علائقية	10د	/	5	19	19	الانجليزية	المستحدثة	نعم	نعم
General self-efficacy	/1992 2013	2014	/	الطلبة الجامعيين	علائقية	/	/	4	10	10	الانجليزية	المترجمة	نعم	لا
Math self-efficacy	2003	2018	/	تلاميذ	علائقية	/	/	4	5	5	الانجليزية	المترجمة	لا	لا
Academic stress	2005	2017	/	طلاب جامعيين دوليين	تشخيصية	/	/	5	53	53	الانجليزية	المترجمة	لا	لا
The explicit achievement	2012	2020	/	رياضيين	علائقية	/	/	6	6	6	الانجليزية	المترجمة	لا	لا

														motive
لا	نعم	الأصلية	الانجليزية	20	20	5	5	/	علائقية	التلاميذ	/	2021	2007	Teacher motivation
لا	لا	المترجمة	الانجليزية	3	3	5	/	/	علائقية	الطلبة	/	2020	/1995 2020	Implicit beliefs of intelligence
نعم	نعم	المترجمة	الانجليزية	14	14	5	4	/	علائقية	الطلبة	/	2017	/1993 2000	Achievement and social goals
لا	لا	المترجمة	الانجليزية	19	19	5	3	/	علائقية	التلاميذ	/	2021	/1990 2005	Perceived teacher support
لا	لا	المترجمة	الانجليزية	20	20	5	/	/	علائقية	الموظفين في مستشفيات وهب الأعضاء	/	2020	/2000 2013	Emotional intelligence scale
لا	لا	المترجمة	الانجليزية	26	26	4	6	/	علائقية	المراهقين	/	2021	/1996 2018	Sleep problems
لا	لا	المترجمة	الانجليزية	5	5	5	/	/	علائقية	الطلاب الجامعيين	/	2021	/1982 2021	Student evaluations of teaching quality in terms of

													learning
نعم	نعم	الأصلية	الانجليزية	9	9	7	/	/	علائقية	ممرضة جراحية	/	2020	1996	the emotional exhaustion subscale of the Maslach Burnout
نعم	نعم	الأصلية	الألمانية	3	3	؟	/	/	تشخيصية	العلماء الأكاديمين	/	2020	1994	Career turnover intentions
لا	لا	الأصلية	الانجليزية	8	8	7	/	/	تقييميه	مهني في رعاية الأطفال	/	2020	2008	Vicarious trauma
لا	لا	المترجمة	الانجليزية	28	28	5	4	/	علائقية	إعلامي	/	2021	/1983 2005	Empathy
لا	لا	الأصلية	الانجليزية	10	10	10	/	/	تشخيصية	المصابين بقلق البالغين	/	2016	2008	The fear of positive evaluation scale
لا	لا	المترجمة	الانجليزية	6	6	4	/	/	علائقية	السيدات	/	2013	/1980 2013	Dietary restraint status

Positive coping	/1998 2020	2021	/	المقيمين في الحجر	علائقية	/	فيه أبعاد	4	12	12	الانجليزية	المترجمة	نعم	نعم
Dark personality traits	2010	2020	/	طلبة الجامعة	علائقية	/	3	7	12	12	الانجليزية	الأصلية	لا	لا
PTSD checklist	2013	2021	/	طلبة علم النفس	علائقية	/	4	5	20	20	الانجليزية	الأصلية	لا	لا
Obsesive compulsive	/2002 2007	2020	/	مرضى الوسواس القهري	علائقية	/	6	؟	18	18	الانجليزية	المترجمة	لا	لا
Child Anxiety sensitivity index	1991	2021	/	مرضى الاكتئاب والوسواس	علائقية	/	4	3	18	18	الانجليزية	الأصلية	لا	لا
Employee engagement	2003	2018	/	الموظفين	علائقية	/	3	5	17	17	الانجليزية	الأصلية	لا	لا
Ethnic and national identification	2012	2018	/	الأتراك والمغاربة المقيمون بهولندا	تشخيصية	/	2	5	8	8	الانجليزية	الأصلية	لا	لا
Gender related discrimination	2015	2021	/	الأفراد الذين يعانون الاكتئاب والرفض	تشخيصية	/	/	2	17	17	الانجليزية	الأصلية	لا	لا
Shyness scale	/2002	2020	/	طلبة الجامعة	علائقية	/	4	5	17	17	الانجليزية	المترجمة	لا	لا

												2009	
لا	لا	المترجمة	الانجليزية	28	28	؟	3	/	علائقية	طلبة	/	2021	/1994 2012	Childhood trauma questionnaire
لا	لا	المستحدثة	الصينية	20	20	4	2	/	علائقية	الأمهات	/	2013	1999	Trait coping style
لا	لا	الأصلية	الانجليزية	49	49	4	/	/	علائقية	التلاميذ	/	2016	2006	The autism social skills profile
لا	نعم	المترجمة	الانجليزية	34	39	؟	8	/	علائقية	الأطفال المصابين بالتوحد	/	2012	1990	Sibling relationship Questionnaire
لا	لا	الأصلية	الانجليزية	50	50	4	10	/	علائقية	أسر أطفال التوحد	/	2012	1973	Brief symptom rating
لا	لا	الأصلية	الانجليزية	14	14	؟	4	/	علائقية	مدراء تنفيذيين	/	2019	2011	Perceived flexibility
لا	لا	المترجمة	الانجليزية	6	6	5	/	/	علائقية	الممرضات	/	2017	/2002	Psychological

distress	2008												
Rumination	2017/ 2017	2021	/	تلاميذ المتوسط	علائقية	/	/	4	10	10	الانجليزية	المترجمة	لا	لا
Leader inclusiveness	2010	2012	/	الموظفين	علائقية	/	/	5	9	9	الانجليزية	الأصلية	لا	لا

تعتبر كل نسخة من الاختبار بمثابة اختبار منفصل.

عنوان الدراسة	اللغة	البلد	الفئة العمرية	حجم العينة	الخصائص السيكومترية المستخدمة	عدد المحكمون	من هم المحكمون	نوع الدراسة	اسم صاحب الاختبار الحالي	اسم صاحب الاختبار المترجم أو المكيف	اسم صاحب الاختبار الأصلي
الضغوط الأسرية وأساليب مواجهتها لدى والدي الأطفال التوحديين في دولة الكويت	العربية	الكويت	أباء أطفال التوحد الذين سنهم بين 6-12	53	صدق المحكمين/ اتساق داخلي ألفا كرونباخ/ تجزئة	11	أعضاء هيئة التدريس بجامعة الخليج	ماجستير	فاطمة عبد الله عيد الشمري	الشراح	Vitaliano et al

						نصفية				
L.Thorpe W.Clark E.Tiger	العادلي	حنان ماجد خضير	ماجستير	أعضاء هيئة التدريس من عدة جامعات	9	الصدق الظاهري/الصدق البنائي(القدرة التمييزية للفقرات/علاقة الفقرة بالمجموع) التجزئة النصفية /ألفا كرونباخ	150	11-15 سنة	الدنمارك	العربية	التنشئة الأسرية وعلاقتها بالتوافق النفسي والاجتماعي لأطفال الجالية العراقية في الدنمارك
Humm. M Bel	عمر بن علي	سلمى عبد العظيم أحمد	ماجستير	لا توجد تفاصيل	7	الصدق الظاهري/ صدق البناء(الاتساق الداخلي) ألفا كرونباخ/ الصدق الذاتي	40	14-18 سنة	قطر	العربية	العنف الأسري وعلاقته بالتوافق النفسي لدى المراهقين
باظة	/	محمد حسين أحمد ناصر	ماجستير	أعضاء الهيئة الإدارية والتدريسية	11	الصدق الظاهري/صدق البناء(الاتساق الداخلي) ألفا كرونباخ	377	/	فلسطين	العربية	العنف المدرسي وعلاقته بالمهارات الاجتماعية لدى طلبة المرحلة الثانوية في المدارس الحكومية في مدينتي رام الله والبيرة
عياش	/	سمر عيسى إبراهيم صباح	ماجستير	ذكر فقط أن المحكمين ينتمون إلى 4 جامعات	لا توجد تفاصيل	الصدق الظاهري/ الاتساق الداخلي ألفا كرونباخ	52	حتى 15 سنة	فلسطين	العربية	أثر برنامج إرشادي يستند إلى اللعب والفن في خفض السلوك العدواني لدى أطفال قرية الأطفال SOS في محتفظة بيت لحم

				مختلفة							
عوامل الشخصية وأعراض ما بعد الصدمة المتعلقة بجرائم الاغتصاب لدى المراهقين	العربية	الكويت	/	30	صدق البناء (الاتساق الداخلي) ألفا كرونباخ/ التجزئة النصفية	/	/	ماجستير	أنور عيدان الخليفة	عطية هنا وآخرون	Cattel.R وزملاؤه
أثر برنامج إرشاد جمعي يستند إلى العلاج السلوكي المعرفي في تحسين الكفاية الذاتية المدركة وخفض سلوك الغضب والاكتئاب لدى الأحداث الجانحين	العربية	الكويت	16-19 عام	ص60/ ث30	صدق المحتوى/صدق البناء (الاتساق الداخلي) ألفا كرونباخ/ التجزئة النصفية	10	أساتذة مختصين في الإرشاد في الكويت والجامعات الأردنية	دكتوراه	حمود أحمد الرشيدي	/	أبو رمان
أثر برنامج إرشاد جمعي يستند إلى العلاج السلوكي المعرفي في تحسين الكفاية الذاتية المدركة وخفض سلوك الغضب والاكتئاب لدى الأحداث الجانحين	العربية	الكويت	16-19 عام	ص60/ ث30	صدق المحتوى/ صدق البناء (الاتساق الداخلي) التطبيق وإعادة التطبيق/ ألفا كرونباخ	10	أساتذة مختصين في الإرشاد في الكويت	دكتوراه	حمود أحمد الرشيدي	عواد	Spielbrgere
الوعي الذاتي وعلاقته بالإقناع الاجتماعي لدى طلبة الجامعة	العربية	العراق	/	ص /400	الصدق الظاهري/ صدق البناء (الاتساق الداخلي	12	أساتذة جامعات	ماجستير	تقى بدري عزيز	/	القرة غولي

				مختلفة في العراق		بين الفقرات والمحور وبين الفقرات والدرجة الكلية/الصدق التمييزي) التطبيق وإعادة التطبيق	ث50			
قياس العنف الأسري وعلاقته بالحاجات لدى طالبات المرحلة الإعدادية	العربية	العراق	/	أساتذة جامعات العراق	12	الصدق الظاهري/ صدق البناء (الصدق التمييزي)/ التطبيق وإعادة التطبيق	ص 400/ ث100	ماجستير	زهراء هادي الفرطوسي	/	امزيان زبيدة
تناقض إدراك الذات والاكتئاب وعلاقتهما بالعوامل الخمسة الكبرى الشخصية لدى طلبة الجامعة	العربية	العراق	/	أساتذة في علم النفس والقياس والتقويم	12	الصدق الظاهري/ صدق الفقرات التطبيق وإعادة التطبيق	ث50	دكتوراه	عمار عبد الجبار قدوري السلماني	عبد الستار ابراهيم	Beck
تأثر النظام الذكي لمعالجة المعرفة (R.I.S.K) في تنمية التفكير الناقد والإبداعي لدى طالبات الخامس الإعدادي	العربية	العراق	/	القياس والتقويم وعلم النفس	14	صدق الترجمة/ الصدق الظاهري/ الصدق العاملي التجزئة النصفية/ كيودرتشاردسون20	ص 400/ ث100	ماجستير	هبة مجيد عيسى	ترجمة الباحثة نفسها	Waston-Glaser
الأسلوب المعرفي الاندفاع-التروي وعلاقته بالقدرة على اتخاذ القرار لدى طلبة مرحلة الدراسة الإعدادية	العربية	العراق	/	أساتذة في علم النفس وعلوم التربية	13	الصدق الظاهري الاختبار وإعادة الاختبار	ص28/ ث40	ماجستير	إنصاف المالكي	/	عياش

أثر تعقد المادة اللفظية وتعليمات توجيه حركة العين على الفهم القرائي وحل المشكلات لدى تلاميذ المرحلة الإعدادية	العربية	مصر	11 سنة فما فوق	40	صدق المحكمين/ الصدق العاملي ألفا كرونباخ/ التجزئة النصفية/ الاتساق الداخلي	12	أساتذة علم النفس والمناهج وطرق التدريس واللغة العربية	دكتوراه	عمرو محمد إبراهيم يوسف	ترجمة الباحث نفسه	Armstrong
أثر برنامج قائم على حي الاستطلاع في تنمية بعض العمليات المعرفية ومهارات التفكير لدى تلاميذ المرحلة الإعدادية	العربية	مصر	/	30	صدق الاتساق الداخلي التطبيق وإعادة التطبيق	/	/	ماجستير	عاصم عبد المجيد كامل أحمد	خيرى عجاج	Berlyne
أثر برنامج تدريبي في إعادة تنظيم وإثراء البنية المعرفية في تحسين التحصيل الدراسي والاحتفاظ لدى طلاب المرحلة الثانوية	العربية	مصر	8-13 سنة والراشدين	40ص/ 80ث	صدق المحكمين/ صدق المحتوى/ الصدق الذاتي التجزئة النصفية/ ثبات الصور المتكافئة/ كيودرريتشاردسون/ التطبيق وإعادة التطبيق	12	اساتذة جامعيين ومعلمين	دكتوراه	وحيد ناجي فهيم حنا	ميخائيل	Cattell
أثر تعقد المادة اللفظية وتعليمات توليد الصور الذهنية على الاستدعاء لدى تلاميذ المرحلة الإعدادية	العربية	مصر	/	50	صدق المحكمين/ الصدق التلازمي/ الصدق العاملي ألفا كرونباخ/ التجزئة النصفية/ الاتساق الداخلي	10	متخصصين في علم النفس	ماجستير	عمرو محمد يوسف	ترجمة الباحث نفسه	Marks

مدى توفر الحاجت النفسية وعلاقتها بالصلابة النفسية لدة عينة من الطلبة الموهوبين في منطقة تبوك	العربية	السعودية	/	40	الصدق الظاهري/ صدق البناء (الاتساق الداخلي) التطبيق وإعادة التطبيق/ ألفا كرونباخ	6	أساتذة في الجامعة الأردنية والسعودية	ماجستير	سارة سابيتان الشامان	/	الحجرية
العلاقة بين عدد ساعات استخدام شبكات التواصل الاجتماعي والاتزان الانفعالي ومفهوم الذات لدى طلبة جامعة تبوك	العربية	السعودية	/	ص30/ ث34	الصدق الظاهري/ صدق البناء (الاتساق الداخلي) التطبيق وإعادة التطبيق/ ألفا كرونباخ	13	أساتذة في الجامعة الأردنية والسعودية	ماجستير	أحمد صالح رجاء البلوي	الداود	Piers-Harris
درجة انتشار التشوهات المعرفية لدى المعلمات المطلقات والمرملات وعلاقتها بالكفاية الاجتماعية لديهن في منطقة تبوك	العربية	السعودية	/	30	الصدق الظاهري/ صدق البناء (الاتساق الداخلي) التطبيق وإعادة التطبيق/ ألفا كرونباخ	10	أساتذة في الجامعة الأردنية والسعودية	ماجستير	أسماء عبد الرحمان عودة الجوفي	العتيبي	Walker-Mcconnell
أنماط الشخصية الثمانية استنادا ليونج وعلاقتها بجودة الحياة لدى طلبة المرحلة الثانوية المتفوقين تحصيلا والعاديين في المملكة العربية السعودية	العربية	السعودية	14-19سنة	45	الصدق الظاهري/ صدق البناء (الاتساق الداخلي) التطبيق وإعادة التطبيق/ ألفا كرونباخ	10	أساتذة في الجامعة الأردنية والسعودية	ماجستير	هاني بن عايد الثبيتي	الخميس	Myers-briggs
اتجاهات معلمي صعوبات التعلم نحو برنامج صعوبات التعلم ومستوى دافعية الانجاز لديهم في محافظة يتبع بالمملكة العربية	العربية	السعودية	/	30	الصدق الظاهري/ صدق البناء (الاتساق الداخلي) التطبيق وإعادة التطبيق/	11	متخصصين في المجال	ماجستير	فهد إسماعيل البلوي	موسي	Hermans

السعودية						ألفا كرونباخ				
الفروق في أداء الوظائف التنفيذية بين الأطفال التوحديين ذوي الأداء الوظيفي المرتفع والأطفال المعاقين ذهنيا بدرجة بسيطة في المملكة العربية السعودية	العربية	السعودية	/	ث10/ 20ص	تطبيق وإعادة التطبيق الصدق التمييزي	/	/	ماجستير	رحاب حمد حمدي الصاعدي	ترجمة الباحث نفسه	Stroop/ Golden Freshwater Golden
التذكر والتفكير الإبتكاري وعلاقتهما ببعض سمات الشخصية (دراسة على طلاب الكليات الأكاديمية والعسكرية)	العربية	السودان	/	32	الصدق الظاهري/صدق الاتساق الداخلي ألفا كرونباخ/ سبيرمان براون	5	أساتذة في قسم علم النفس	دكتوراه	عثمان موسى حريكة إبراهيم	محمد حسين	Wechseler
اتجاهات الوالدين نحو تعليم البنات وعلاقتها بمفهوم الذات لدى الأسر الصومالية المقيمة بالسودان	العربية	السودان	/	50	صدق المحكمين/ الاتساق الداخلي ألفا كرونباخ/ الصدق	3	أساتذة علم النفس	ماجستير	عبد الله معلم علي	/	محمد صوالحة
مشكلات الكلام لدى طفل الروضة وعلاقتها بأساليب معاملة الأمهات والوضع الاقتصادي الاجتماعي للأسرة	العربية	السودان	/	60	الصدق الظاهري/ الصدق العاملي/الاتساق الداخلي ألفا كرونباخ/ سبيرمان براون	8	أساتذة علم النفس	دكتوراه	غادة محمد أحمد حمدي	/	أنور رياض وعبد العزيز المغيضب

العنف الأسري وعلاقته بالتوافق النفسي لدى المراهقين (دراسة وصفية على الطلبة السودانيين من عمر 14–18 سنة)	العربية	السودان	/	40	الصدق الظاهري/صدق البناء (الاتساق الداخلي) الثبات/الصدق الذاتي	7	أساتذة علم النفس	ماجستير	سلمى عبد العظيم حسب الرسول أحمد	/	يامن سهيل
صعوبات التعلم النوعية وعلاقتها بأساليب المعاملة الوالدية وسمات الشخصية لدى تلاميذ مرحلة الأساس	العربية	السودان	/	40	الصدق الظاهري/صدق الاتساق الداخلي ألفا كرونباخ/ سبيرمان براون	لا توجد تفاصيل	أساتذة علم النفس	دكتوراه	الدود يوسف الدود أحمد	/	عبد الرحيم والمغيضب
علاقة مرحلة الشيخوخة بالروح المعنوية والرضا عن الحياة –دراسة ميدانية بمؤسسة الوفاء لإعادة تأهيل العجزة والمسنين–	العربية	السودان	/	39	الصدق الظاهري/ الاتساق الداخلي ألفا كرونباخ	7	أساتذة علم النفس	دكتوراه	رابح محمد أمنيسي ديبك	/	محمد العزبي
البروفيل النفسي لعينة من المتزوجين مضطربي الحياة الجنسية	العربية	فلسطين	/	50	صدق المحكمين/ صدق الاتساق الداخلي طريقة الإعادة/ ألفا كرونباخ	10	/	ماجستير	أحمد فهمي خضر	/	عبد المقصود

فاعلية العلاج المعرفي السلوكي بأسلوب حل المشكلات في التخفيف من الضغوط النفسية لدى زوجات مرضى الفصام العقلي	العربية	فلسطين	/	31	صدق المحكمين/صدق الاتساق الداخلي ألفا كرونباخ/التجزئة النصفية	15	أعضاء الهيئة التدريسية والبحث العلمي	ماجستير	أسماء عبد القادر غراب	مصطفى المصري	D'zurilla T Nezu A MaydeuOlivares A
فاعلية برنامج علاجي باستخدام السيكودراما في خفض أعراض اضطراب الاكتئاب لدى النساء	العربية	فلسطين	18-60 سنة	30	صدق الاتساق الداخلي/ الصدق التمييزي ألفا كرونباخ/ التجزئة النصفية	/	/	ماجستير	إيناس حسين أحمد	أحمد عبد الخالق	Beck
صورة الجسم والاكتئاب في علاقتهما ببعض المتغيرات لدى جرحى الحروق في قطاع غزة	العربية	فلسطين	/	30	صدق الاتساق الداخلي/ الصدق التمييزي ألفا كرونباخ/التجزئة النصفية	/	/	ماجستير	مسرة محمود يوسف الشاعر	/	Beck
التغير القيمي وعلاقته بهوية الذات والاغتراب النفسي لدى طلبة الثانوية العامة في قطاع غزة	العربية	فلسطين	14-18 سنة	50	الصدق الظاهري/صدق الاتساق الداخلي ألفا كرونباخ/ التجزئة النصفية	1	المشرف	ماجستير	سناء عادل ابراهيم كباجة	/	الوحيدي لبنى

المناخ الأسري وعلاقته ببعض متغيرات الشخصية لدى النساء السجينات في قطاع غزة	العربية	فلسطين	/	28	صدق الاتساق الداخلي ألفا كرونباخ/ التجزئة النصفية	/	/	ماجستير	رانيا محمد جعفر البيك	/	سفيان أبو نجيلة
الخبرات الصادمة لدى أمهات الأيتام وعلاقتها بالتوافق النفسي والاجتماعي لدى الأم والطفل	العربية	فلسطين	/	30	صدق الاتساق الداخلي ألفا كرونباخ/ التجزئة النصفية	/	/	ماجستير	منيرة محمود جودة	/	قوتة
فاعلية برنامج علاجي معرفي سلوكي مقترح لتخفيف الاكتئاب لدى أمهات الأطفال المصابين بتشوهات خلقية	العربية	فلسطين	/	30	صدق الاتساق الداخلي/ الصدق التمييزي ألفا كرونباخ/ التجزئة النصفية	/	/	ماجستير	أسماء محمود سعود	/	Beck
الأفكار اللاعقلانية وانفعال الغضب لدى أفراد الشرطة في ضوء بعض التغييرات	العربية	فلسطين	/	50	صدق الاتساق الداخلي/ الصدق التمييزي ألفا كرونباخ/ التجزئة النصفية	/	/	ماجستير	موسى زهير حسن القصاص	/	الريحاني سليمان
الدعم الاجتماعي المدرك وعلاقته بالأمن النفسي لدى المطلقات	العربية	الأردن	/	30	صدق المحكمين/صدق البناء (الاتساق الداخلي) الاختبار وإعادة	9	أساتذة علم النفس بالجامعات	ماجستير	تهاني رزق علي بدير	ترجمة الباحث نفسه	Mohammad Alsadat Yim

						الاختبار/ألفا كرونباخ			الأردنية			Chinna
الهوية النفسية وعلاقتها بالتفكير اللاعقلاني لدى طلبة جامعة مؤتة	العربية	الأردن	14-56سنة	53	صدق المحكمين/ صدق الاتساق الداخلي الاختبار وإعادة الاختبار/ ألفا كرونباخ	10	أساتذة علم النفس بالجامعات الأردنية	ماجستير	رقية مثقال الصرايرة	الربابعة	Adams Huh Bennion
مستوى التفكير الناقد لدى طلبة الجامعة مؤتة وعلاقته بالدافع المعرفي	العربية	الأردن	/	53	صدق المحكمين/ صدق الاتساق الداخلي الاختبار وإعادة الاختبار/ ألفا كرونباخ	10	أساتذة علم النفس بالجامعات الأردنية	ماجستير	سرى راضي المحادين	العبدلات	/
السلوك الصحي لدى طلبة المراهقة المبكرة وعلاقته بالشعور بالأمن النفسي وتنظيم الوقت	العربية	الأردن	/	40	الصدق الظاهري/ صدق الاتساق الداخلي/ الصدق التمييزي الاختبار وإعادة الاختبار/ألفا كرونباخ	10	متخصصين بعلم النفس والقياس والإرشاد التربوي	ماجستير	ميسون أمين الثبيتات	/	شكري
أثر استخدام اللوح التفاعلي في التفكير الإبداعي في مادة الرياضيات لدى طالبات	العربية	الأردن	/	17	صدق الاتساق الداخلي الاختبار وإعادة الاختبار/	/	/	ماجستير	دينا عزمي عبد الله أبو نحلة	سيف الدين	Torrance

الصف الثامن في الأردن					معامل ثبات المصحيحين						
علاقة العجز المتعلم بالعنف المدرسي لدى الطلاب: دراسة ميدانية على طلاب المرحلة المتوسطة	العربية	الأردن	/	75ص 57ث	صدق المحكمين/ صدق الاتساق الداخلي الاختبار وإعادة الاختبار/ ألفا كرونباخ	11	أساتذة علم النفس بالجامعات الأردنية	ماجستير	فايز الجهني	/	الزواهرة
الذكاء الوجداني وعلاقته بالسعادة والثقة بالنفس لدى طلبة جامعة مؤتة	العربية	الأردن	/	50	صدق المحكمين/ صدق الاتساق الداخلي ثبات الإعادة/ ألفا كرونباخ	8	أساتذة علم النفس في جامعة مؤتة	ماجستير	حسن عطا الكساسبة	/	النبهان وكمالي
قلق العولمة وعلاقته بالمسؤولية الاجتماعية لدى طلبة جامعة مؤتة	العربية	الأردن	/	43	صدق المحكمين/صدق الاتساق الداخلي ثبات الإعادة/ ألفا كرونباخ	11	أساتذة الجامعات الأردنية	ماجستير	فداء ضامن العبيسات	/	علي
النضج الأخلاقي وعلاقته بالتكيف النفسي والاجتماعي لدى طلبة جامعة مؤتة	العربية	الأردن	/	84ص 46ث	صدق المقارنة الطرفية ثبات الإعادة	/	/	ماجستير	محمد جبريل المرابحة	؟	Rest

النضج الأخلاقي وعلاقته بالتكيف النفسي والاجتماعي لدى طلبة جامعة مؤتة	العربية	الأردن	/	63 ص وث ماعدا التجزئة 46	صدق المحكمين/ صدق الاتساق الداخلي ألفا كرونباخ/ التجزئة النصفية/ ثبات الإعادة	11	أساتذة من جامعة مؤتة وجامعة الخليج العربي	ماجستير	محمد جبريل المرابحة	/	الديب
مستوى انتشار السلوك الانسحابي لدى الطلبة العمانيين الدارسين بالجامعات الأردنية وعلاقته بالعوامل الخمس الكبرى	العربية	الأردن	/	35	صدق الاتساق الداخلي الاختبار وإعادة الاختبار/ ألفا كرونباخ	/	/	ماجستير	عبير مسلم خميس النصيري	الأنصاري	Costa & McCrae
فاعلية برنامج إرشادي في وقف الأفكار والأفعال القهرية لدى الطلبة الذين يعانون من الاكتناز القهري	العربية	الأردن	/	42	الصدق الظاهري/ صدق الاتساق الداخلي ألفا كرونباخ/ الاختبار وإعادة الاختبار	10	أساتذة الإرشاد النفسي	دكتوراه	نهى عوض محمد النوايسة	ترجمة الباحثة نفسها	Forst et al
الأفكار اللاعقلانية لدى المراهقين وعلاقتها بالاكتئاب	العربية	الجزائر	/	45	صدق المحك/ الصدق الذاتي التجزئة النصفية	/	/	ماجستير	محمد روبي	/	أحمد عبد الخالق
مستويات الاحتراق النفسي لدى معلمي الأقسام النهائية وعلاقتها ببعض المتغيرات	العربية	الجزائر	/	100	الصدق الظاهري/ الاتساق الداخلي	6	أساتذة علم النفس في جامعة تلمسان	دكتوراه	زكية بن عامر	؟	Maslach & Jackson

				وسعيدة		التطبيق وإعادة التطبيق/ ألفا كرونباخ/ التجزئة النصفية				
Argyle & Martin & Crossland	نشوى دردير	بلال نكديل	ماجستير	/	/	صدق المقارنة الطرفية ألفا كرونباخ	40	/	الجزائر	العربية	الأفكار اللاعقلانية وعلاقتها بالسعادة لدى طلبة الجامعة
Eysanck H	أحمد عبد الخالق	إبراهيم موسى	دكتوراه	/	/	صدق المقارنة الطرفية ألفا كرونباخ	26	/	الجزائر	العربية	الأفكار اللاعقلانية والاستعدادات الذهانية لدى عينة من الجانحين بالشرق الجزائري
Torrance	علي محمد هاشم	محمد روبي	دكتوراه	/	/	الصدق التطابقي التطبيق وإعادة التطبيق	74	16-18سنة	الجزائر	العربية	أنماط السيادة النصفية للمخ لدى المراهقين وأثرها في تفعيل العلاقة بين الأفكار اللاعقلانية والصحة النفسية
الرحاني	/	سهام بوخاري	دكتوراه	/	/	صدق الاتساق الداخلي التجزئة النصفية	40	16-27سنة	الجزائر	العربية	مدى فعالية العلاج العقلاني الانفعالي للتخفيض من الأفكار اللاعقلانية والتخفيف من قلق الانزعاج لدى المدمنين على المخدرات
Pierce et al	ترجمة الباحث نفسه	علي خزف الله	دكتوراه	/	/	صدق الاتساق الداخلي/ الصدق التمييزي ألفا كرونباخ	194	/	الجزائر	العربية	نوعية العلاقة الزوجية وعلاقتها بمهارات الذكاء العاطفي

التغيرات الاجتماعية والاقتصادية وانعكاساتها على السلطة الوالدية كما يدركها الأبناء في الأسرة الجزائرية	العربية	الجزائر	/	50	الصدق الظاهري/ الصدق التمييزي/ الصدق الذاتي إعادة الاختبار/ ألفا كرونباخ	11	أستاذ جامعي	دكتوراه	حمودة سليمة	/	محمد بيومي خليل
نمط الشخصية (أ) و(ب) وعلاقتها بالضغوط واستراتيجيات مواجهتها لدى الأطباء اليمنيين	العربية	الجزائر	/	100	صدق المحكمين/ صدق الاتساق الداخلي/ الصدق التمييزي ألفا كرونباخ/ التجزئة النصفية	7	أساتذة جامعة الجزائر وجامعة صنعاء	دكتوراه	ناصر الدين زبدى	ترجمة الباحث نفسه	ناصر الدين زبدى
إدارة الضغوط النفسية وعلاقتها بسمتي الشخصية المناعية (الصلابة النفسية والتوكيدية) في ضوء الذكاء الانفعالي	العربية	الجزائر	/	50	صدق الاتساق الداخلي/ الصدق التمييزي ثبات التجزئة النصفية/ ألفا كرونباخ	/	/	دكتوراه	مريامة حنصالي	زيزي ابراهيم	Carver & Scheier
مهارات حل المشكلات لدى التلاميذ	العربية	الجزائر	/	75	الاتساق الداخلي ألفا كرونباخ	/	/	مقال	خليدة مهرية	/	نزيه حمدي
دراسة مقارنة لجودة الحياة وعلاقتها بالتكيف النفسي والاجتماعي بين ممارسي وغير	العربية	الأردن	/	61	صدق المحتوى	4	يحملون درجة الدكتوراه	مقال	محمد حسن أبو الطيب	/	عبير عبد الرزاق عبده

ممارسي السباحة كبار السن						التجزئة النصفية					
مستوى الهوية النفسية والرضا الحياتي لدى طالبات الصف العاشر الأساسي من وجهة نظرهن وأمهاتهن: دراسة مقارنة	العربية	الأردن	/	30	10	الصدق الظاهري/ صدق الاتساق الداخلي إعادة الاختبار/ ألفا كرونباخ	أعضاء هيئة التدريس في الجامعات الأردنية	مقال	فاطمة عبد الرحيم النوايسة ووجدان خليل الكركي	/	العويلي
إدمان الطلاب على استخدام مواقع التواصل الاجتماعي وعلاقته بالأمن النفسي والتورط في الجرائم السيبرانية	العربية	الأردن	/	50	؟	صدق المحكمين/ صدق الاتساق الداخلي ألفا كرونباخ	؟	مقال	عبير محمد الصبان وسماح عيد الحربي	/	خليوي
عادات العقل وعلاقتها بالقدرة على اتخاذ القرار والاتزان الانفعالي لدى طلبة جامعة مؤتة 2016	العربية	الأردن	/	50	11	الصدق الظاهري/ صدق الاتساق الداخلي إعادة الاختبار/ ألفا كرونباخ	أساتذة جامعة مؤتة	مقال	وجدان خليل الكركي وشذى سلامة العواودة	Al Hawari	عبدون
العلاقة بين درجة الاحتراق لدى مديري المدارس الحكومية في الأردن وأدائهم الوظيفي من وجهة نظرهم	العربية	الأردن	/	15	10	صدق المحكمين/ الاتساق الداخلي إعادة الاختبار	أساتذة الجامعة الأردنية	مقال	صالح عبابنة	ترجمة الباحث نفسه	Freidman
اضطراب صورة الجسم وعلاقته بالشعور	العربية	السعودية	/	100	10	صدق المحكمين/ الصدق التلازمي/التجزئة النصفية/	أساتذة علم	مقال	لولوة صالح الرشيد	؟	Zimmerman et al

بالاكتئاب لدى عينة من مرضى السكري						ألفا كرونباخ		النفس				
مستويات الاحتراق النفسي لدى حكام كرة القدم في المملكة العربية السعودية	العربية	الإمارات	/	10		صدق المحتوى/ الاتساق الداخلي إعادة الاختبار	؟	أعضاء هيئة التدريس	مقال	عبد العزيز عبد الكريم المصطفى	علاوي	Weinberg & Richardson
الرضا عن الحياة وعلاقته بكل من الاعتقاد في الكفاءة الذاتية والاكتئاب والوحدة النفسية لدى عينة من ذوي الإعاقة البصرية	العربية	الكويت	/	25		صدق التقاربي والتباعدي/الاتساق الداخلي/ ألفا كرونباخ/ القسمة النصفية	/	/	مقال	أمثال هادي الحويلة	/	شويخ
معنى الحياة وعلاقته بكل من الرضا عن الحياة وبعض أبعاد التفكير الايجابي لدى طلبة الجامعة بدولة الكويت	العربية	الكويت	/	50		صدق التكوين الفرضي/صدق المحك التلازمي الاتساق الداخلي ألفا كرونباخ	/	/	مقال	سعاد البشر وحسن الحميدي	/	أبو غزالة
الذكاء الأخلاقي وعلاقته بكل من تقدير الذات ودافعية الانجاز لدى طلبة جامعة العين	العربية	الإمارات	/	40		الصدق الظاهري/ صدق الاتساق الداخلي ألفا كرونباخ/ ثبات الإعادة	7	من ذوي الخبرة	مقال	هاني يوسف الجراح	/	الربعي
الاتجاه نحو العلم وعلاقته بالدافع المعرفي	العربية	الإمارات	/	200		صدق الترجمة/ الصدق	؟	؟	مقال	بشرى أحمد جاسم	ترجمة	؟

لدى طلبة جامعة الشارقة						الظاهري/ صدق البناء ألفا كرونباخ				الباحث نفسه
الأفكار اللاعقلانية وعلاقتها بالشعور بالوحدة النفسية لدى عينة من طالبات الجامعة في ضوء التخصص والتحصيل الدراسي	العربية	السعودية	/	38	/	صدق الاتساق الداخلي ألفا كرونباخ/ التجزئة النصفية/ ثبات الإعادة	/	/	مقال	رجاء مريم ومنيرة الشمسان	الشناوي	Russel
فعالية نموذج لبعض المتغيرات المرتبطة باضطراب تفكك الأنا لدى طلاب الجامعة	العربية	مصر	/	258	/	الاتساق الداخلي ألفا كرونباخ/ التجزئة النصفية	/	/	مقال	علي شعيب وأيريني غبريال	عبد الخالق أحمد	Hodgson
الخوف من الحسد وعلاقته بالتشاؤم الدفاعي لدى طلبة الجامعة	العربية	العراق	/	54	/	صدق المحكمين/ صدق المقارنة الطرفية/ الاتساق الداخلي ألفا كرونباخ	/	/	مقال	ياسمين جرجيس يونس	/	آل غزال
نموذج العوامل الستة للسعادة لدى طلبة الجامعة	العربية	العراق	/	42	؟	الصدق الظاهري/ صدق البناء ألفا كرونباخ	؟	؟	مقال	سارة سلمان وماجدة العلي	الجمال	Rosemry A Abbott al
الأمن العاطفي وعلاقته بالاستقرار النفسي	العربية	سوريا	/	32	13	الصدق الظاهري/ الصدق	أساتذة الجامعة		مقال	سنان الأسدي	/	الجبوري

لدى طلبة الجامعة					التمييزي/ صدق الاتساق الداخلي التطبيق وإعادة التطبيق		العراقية				
أثر مستوى جودة الروضة التعليمي في تنمية التفكير الإبداعي لدى تلامذة الصف الأول الأساسي في مدارس التعليم الأهلي في أمانة العاصمة صنعاء	العربية	اليمن	/	30	التطبيق وإعادة التطبيق وثبات المصححين	/	/	مقال	هناء الفلفلي عزة غانم إلهام المرتضى	؟	Torrance
أثر مهارات التفكير ما وراء المعرفي على الدافعية للانجاز لدى طلاب قسم علم النفس بكلية الآداب	العربية	ليبيا	/	40	صدق المحكمين/ صدق المقارنة الطرفية/الصدق الذاتي/ألفا كرونباخ/ التجزئة النصفية	؟	؟	مقال	عبد الحكيم سعود	/	منى ابراهيم
المرونة النفسية كمتغير وسيط بين أحداث الحياة الضاغطة والصحة النفسية لدى عينة من طلبة جامعة دمشق	العربية	سوريا	/	70	صدق المحتوى/ صدق البناء ثبات الإعادة/ ألفا كرونباخ	6	أعضاء هيئة التدريس علم النفس	مقال	أحمد الزعبي	العاسمي	Connor and Davison
الاغتراب النفسي وعلاقته بالتوافق النفسي لدى عينة من طلبة جامعة دمشق القاطنين بالمدينة الجامعية	العربية	سوريا	/	53	صدق المحتوى / الاتساق الداخلي ثبات الإعادة/ تجزئة	؟	؟	مقال	غسان الرحيلي	/	نعيسة

						النصفية/ ألفا كرونباخ					
حاجات الأطفال وفق النوع الاجتماعي والصف الدراسي من وجهة نظرهم لدى طلبة الصف السادس والثامن والعاشر بسلطنة عمان	العربية	عمان	6-12 سنة	؟	صدق المحكمين ألفا كرونباخ	؟	مختصين في محال علم النفس واللغة العربية	مقال	سالم الكحالي	/	الزبيدي وسليمان
أثر قلق الانفصال عن جماعة الرفاق في مستوى المسايرة الاجتماعية لدى عينة من الطلبة المراهقين في قضاء حيفا	العربية	فلسطين	/	41	صدق المحتوى/ صدق الاتساق الداخلي ثبات الإعادة/ ألفا كرونباخ	8	أساتذة جامعة اليرموك	مقال	منار مصطفى وعنات أبو عصبة	/	أبو نمر
أنماط الشخصية المهنية لدى طلبة الصف العاشر الأساسي في محافظة معان في ضوء نظرية هولاند وعلاقتها ببعض المتغيرات	العربية	الأردن	/	35	صدق الاتساق الداخلي ثبات الإعادة/ ألفا كرونباخ	/	/	مقال	عمر محاسنة وعوده مراد	عطار	؟
الذكاء العاطفي وعلاقته بجودة الحياة لدى المرشدين النفسيين في لواء قصبة إربد	العربية	الأردن	/	20	صدق الاتساق الداخلي ألفا كرونباخ	/	/	مقال	سحر السعدي		الأسطل
التردد النفسي والتحمل النفسي وعلاقتهما ببعض المهارات الأساسية بكرة القدم التباعد	العربية	الجزائر	/	4	الصدق الظاهري التطبيق وإعادة التطبيق	؟	خبراء في المجال الرياضي	مقال	لؤي ناجي حمد	/	بهاء مرقس

إيناس لطفي	/	سهام بوخاري	مقال	/	/	الاتساق الداخلي ألفا كرونباخ	35	/	الجزائر	العربية	سوء الاتصال الزواجي وعلاقته بتطور بعض أعراض اضطراب الوسواس القهري عند الأزواج
Chiu et al	Han et al	Chao Wang, Junming Daib, Jue Li	مقال	/	/	ألفا كرونباخ	68	22-51 سنة	الصين	الانجليزية	Mediating effects of hair cortisol on the mutual association of job burnout and insomnia: A retrospective exploratory study
Choi et al	Yoon et al	Yeoul Han, Jihee Jang, Eunsil Cho, Kee-Hong Choi	مقال	/	/	ألفا كرونباخ	1000	8-13 سنة	كوريا	الانجليزية	Investigating how individual differences influence responses to the COVID-19 crisis: The role of maladaptive and five-factor personality traits
Costa & McCrae	/	Shima Ahmadi-Azad, Hassan Asadollahfam , Masoud Zoghi	مقال	/	/	ألفا كرونباخ	107	25-40 سنة	إيران	الانجليزية	Effects of teacher's personality traits on EFL learners' foreign language enjoyment
Masterp, Morrison	/	Freddie Anibal	ماجستير	/	/	الصدق التقاربي	644	/	الولايات	الانجليزية	The Utiliye of the Olweus Bully Victim Questionnaire In Identifying

Pelleyrini		Pastrana Rivera				الاختبار وإعادة الاختبار/ ألفا كرونباخ			المتحدة		Stably peer victimized Children
Paulson	/	Paul Deschamps	دكتوراه	/	/	التحليل العاملي ألفا كرونباخ	247	13-17 سنة	الولايات المتحدة	الانجليزية	The of Individual Character Strengths In Addescent Academic Achievement
Furnham & Bochner	/	Clint E Thomson	ماجستير	/	/	الصدق التقاربي ألفا كرونباخ	22	/	بريطانيا	الانجليزية	Helping The Transition Through Building New Friendships A Psychological Perspective On Supporting International Students Acculturation and Social Integration
Carver	/	Grace Yoder	ماجستير	/	/	التحليل العاملي ألفا كرونباخ	300	/	الولايات المتحدة	الانجليزية	Coping life satisfaction and empowerment in caregivers of children with an autism spectrum disorder
Weather et al	/	Ilanas Pinsky	ماجستير	/	/	ألفا كرونباخ	248	أكبر من 18 سنة	الولايات المتحدة	الانجليزية	Attachment Quality and Sexual Satisfaction and Sexual Functioning in Romantic Relationship for combat Veterans

Rotter	/	Lauren Johnson	ماجستير	/	/	الاتساق الداخلي التطبيق وإعادة التطبيق	79	/	الولايات المتحدة	الانجليزية	The Relationship between Mock Jurors Locus of control Belief in a just world and NGRI verdict
Spanier's	Sabouri net	Marie deschenes	دكتوراه	/	/	ألفا كرونباخ	113	/	كندا	الانجليزية الفرنسية	Le satisfaction conjugale et le stress parental comme prédateurs de l'alliance dans la triade familiale : la prise de perspective comme modérateur
Radloff	/	Catherine Rowe	ماجستير	/	/	ألفا كرونباخ	71	18-64 سنة	الولايات المتحدة	الانجليزية	Health Related quality of life and suicidal behaviors in primary care patient: conditional indirect effects via interpersonal needs and depressive symptoms
Hart,wearing & Heady		Fiona Johnston	ماجستير	/	/	ألفا كرونباخ	114	19-64 سنة	هولندا	الانجليزية	Polis Stress General Well being and job satisfaction; The Moderating effects of social
Benn	/	Learna Hancerogla	ماجستير	/	/	ألفا كرونباخ	108	7-13 سنة	بريطانيا	الانجليزية	Investigating the effects of mindfulness on children's escecuaive function emotional

											regulation stress and academic performance compared to control condition
Gushue	/	Yi Jung Lee	دكتوراه	طلاب دراسات عليا في علم النفس و2 أخصائيين نفسانيين	12	صدق المحتوى/ التحليل العاملي الاستكشافي الصدق التقاربي ألفا كرونباخ	56	/	كندا	الانجليزية	Exploring the Impact of Asian stereotype endorsement multicultural counseling competence and motivation to respond without prejudice on white therapists clinical judgment
Carver		Morgan Kate Treaster	دكتوراه	/	/	الصدق التقاربي والتباعدي ألفا كرونباخ	508	/	الولايات المتحدة	الانجليزية	Optimist and physical Health Related Quality of live in chronic Illness: Mediating effects of control beliefs and health behaviors
Gutierrez, Osman, Barrios & Kopper	/	Sherry Woods	ماجستير	/	/	ألفا كرونباخ	329	/	الولايات المتحدة	الانجليزية	The relationship between no suicidal self injury frequency and suicidal behaviors, Depression and Anxiety: A curvilinear analysis

Gresham & Elliott	/	Amy Halling	دكتوراه	/	/	ألفا كرونباخ	25	35 سنة	الولايات المتحدة	الانجليزية	Psychosocial Effects of Shared Book Reading
Ang & Earley	/	Teresa Ann Smith	دكتوراه	/	/	ألفا كرونباخ	135	20-40 سنة	الولايات المتحدة	الانجليزية	A study of ethnic minority college students: Relationship Among the big five personality traits cultural intelligence and psychological well being
Roa & Irvine	/	Wendy Viola	ماجستير	/	/	التحليل العاملي ألفا كرونباخ	569	/	الولايات المتحدة	الانجليزية	Adolescent males similarity emotional safety and change in strengths based programming
Spanierman & Heppner	/	Whitney Black	دكتوراه	/	/	ألفا كرونباخ	240	/	الولايات المتحدة	الانجليزية	An Examination of Relation among fear guilt self compassion and multicultural attitudes in white adult
Schwarzer & Jerusalem	ترجمة الباحثة نفسه	Elda Zeko	دكتوراه	عينة من الألبانيين	20	صدق المحتوى/ الصدق التقاربي/ صدق الاتساق الداخلي ألفا كرونباخ	153	20-64 سنة	ألبانيا	الانجليزية	A Cross Sectional study of education acculturation and self efficacy among recent Albania

Galli et al	/	Jacob L Strimaitis	ماجستير	/	/	الصدق التقاربي/ الصدق التنبؤي ألفا كرونباخ	110	/	الولايات المتحدة	الانجليزية	Conditional reflection and the religion reflection scale: how familiarity with different domains influences cognitive reflection
La mbert & Burlingame	/	Annette Tardif	دكتوراه	/	/	الصدق التلازمي ألفا كرونباخ/ التطبيق وإعادة التطبيق	157	/	الولايات المتحدة	الانجليزية	Reported psychological distress and willingness to utilize mental health services for disabled and non disabled university student
Le Buffe N & Shapiro	/	Amy Compton	دكتوراه	/	/	ألفا كرونباخ	2604	/	الولايات المتحدة	الانجليزية	Examining the longitudinal stability of mental health in early elementary school students
Graham et al	/	Hannah Reas	دكتوراه	/	/	ألفا كرونباخ	418	/	مكسيك	الانجليزية	The missing moral dimension: perception of transgressions and the moderating role of moral foundations on psychological distress
Bryan	/	Kelley C Berto	دكتوراه	/	/	الاتساق داخلي ألفا كرونباخ	712	18 فما فوق	الولايات المتحدة	الانجليزية	Interpersonal Needs and Suicide Risk: Examining Indirect Effects of Internal Hostility and Feeling

											Forgiven
Watson & Clark	/	Rachel Kim	دكتوراه	/	/	ألفا كرونباخ	263	/	كولومبيا	الانجليزية	Differential Impact of Racial Microaggressions on Asian Americans: Relationship to Perpetrator & Power Status
Bashford et al	/	Alexandra G Startyner	دكتوراه	/	/	ألفا كرونباخ	123	22-40 سنة	الولايات المتحدة	الانجليزية	The Influence of Doctoral Psychology Trainees'Personal Cannabis Use, Perceptions of Cannabis'Risks, and Attitudes Toward Substance Use
Daly & Miller	/	Jessica Ann Minney	دكتوراه	/	/	الاتساق الداخلي ألفا كرونباخ	273	18 فما فوق	الولايات المتحدة	الانجليزية	I solution: the optimum environment for creativity? The relationship between the experience of ostracism and creativity
Brough	/	Nikki Geeson	ماجستير	/	/	التحليل العاملي ألفا كرونباخ	674	21-65 سنة	نيوزلندا	الانجليزية	Determining the impact of trauma and daily organizational hassles on psychological distress and burnout in New Zealand police

											officers; and the moderating role of social support
Lambert	/	Maryline Germain Bédard	دكتوراه	/	/	ألفا كرونباخ	81	/	كندا	الفرنسية	Attachement, qualité et stabilité conjugale au sein d'une population clinique victime d'abus sexuel en enfance
Fortin Coutu-Wakulezyk	/	Jonathan Girard	دكتوراه	/	/	ألفا كرونباخ	910	12-17 سنة	كندا	الفرنسية	Effet de la Pression Parental À la Réussite Scolaire, de L'importance Parental, de L'anxiété et de la Dépression Sur La Consommation De Psychotropes Chez L'adoliscent selon lage et le sexe
Jowett & Ntounamis	MarcAndré	Cindy Fillion	دكتوراه	/	/	ألفا كرونباخ	200	/	كندا	الفرنسية	Le phénomène de la passion sportive chez les entraineurs de football: Une étude des liens entre la passion le leadership et la qualité de la relation
McGahuey et al	Bourassa et	Noémie Bigras	ماجستير	/	/	ألفا كرونباخ	304	18 سنة فما فوق	كندا	الفرنسية	L'agression Sexuelle en enfance et L'ajustement psychosexuel :

	Brassard										Le rôle des capacités du soi
Otis et Pelletier	؟	Héléne Larouche	دكتوراه	/	/	الاتساق الداخلي	22	23-66 سنة	كندا	الفرنسية	L'efficacité de processus thérapeutiques brefs et ses lien avec le soutien aux besoins de client de la part du thérapeute et l'alliance thérapeutique
Spielberger	؟	Hinda HAMDI	دكتوراه	/	/	التحليل العاملي الاستكشافي التطبيق وإعادة التطبيق	78	18-66 سنة	فرنسا	الفرنسية	Les stratégies de coping en contexte de violences conjugales Étude de leurs effets sur la souffrance psychique et la qualité de vie des femmes victimes de violences conjugales
Peterson et Seligman	؟	Martine Regourd	دكتوراه	/	/	ألفا كرونباخ/ التطبيق وإعادة التطبيق	24	18-60 سنة	فرنسا	الفرنسية	Psychologie positive et EMDR : Question et perspectives
Crowne & Marlowe	Berton volla & Berger	Louis Richard	دكتوراه	/	/	ألفا كرونباخ	61	14-20 سنة	كندا	الفرنسية	L'impact d'une expédition thérapeutique sur la motivation et le temps consacré A' l'activité physique chez des adolescentes

	on										atteints d'un cancer
Endicott et al	/	Keith Bredemeier, Shari Lieblich	مقال	/	/	الصدق المحكي ألفا كرونباخ	147	/	الولايات المتحدة	الانجليزية	Pretreatment levels of rumination predict cognitive-behavioral therapy
Carleton et al	Takebayashi et al	Taishi Kawamoto & Kaichiro Furutani	مقال	/	/	الصدق التمييزي ألفا كرونباخ	500	/	اليابان	الانجليزية	The mediating role of intolerance of uncertainty on the relationships between perfectionism dimensions and psychological adjustment/maladjustment among mothers
Rosenberg	Banos & Guillen	Juan Carmona, Remedios Liria, David Gongora,Maria Teresa Daza, Manuel Alcoba	مقال	/	/	ألفا كرونباخ/ التطبيق وإعادة التطبيق	61	/	إسبانيا	الانجليزية	Anxiety, psychological well-being and self-esteem in Spanishfamilies with blind children. A change in psychologicaladjustment?
Robinson, Mandleco,	Chen &	Yuhan Luo, Fumei Chen,	مقال	/	/	التحليل العاملي التوكيدي	690	31-82 سنة	الصين	الانجليزية	Profiles of maternal and paternal parenting styles in Chinese

Olsen, & Hart	Luster	Xinghui Zhang, Yu Qin, Zhang Yanfan,g Li Qing Zhou, Yun Wang				ألفا كرونباخ					families: Relations to preschoolers' psychological adjustment
Berman et al	/	Rachel Wiley, Steven Berman	مقال	/	/	ألفا كرونباخ	60	12-19 سنة	الولايات المتحدة	الانجليزية	The relationships among caregiver and adolescent identity status, identity distress and psychological adjustment
Beck & Beck	Raitasalo	Alho Iina, Joro Mirka, Juntunen Laura, Muotka Joona, LappalRaimo	مقال	/	/	ألفا كرونباخ	65	12-16 سنة	فنلندا	الانجليزية	Adolescents with poorly controlled type 1 diabetes: Psychological flexibility is associated with the glycemic control, quality of life and depressive symptoms
Davis	Limpo et al	Joana Duarte, Jose Pinto-Gouvei, Barbara Cruz	مقال	/	/	ألفا كرونباخ	280	/	البرتغال	الانجليزية	Relationships between nurses' empathy, self-compassion and dimensions of professional quality of life: A cross-sectional study

Rosenberg	Yamamoto, Matsui, & Yamanari	Tetsuya Kawamoto	مقال	/	/	الصدق التمييزي ألفا كرونباخ	1042	24-18 سنة	اليابان	الانجليزية	The moderating role of attachment style on the relationship between selfconcept clarity and self-esteem
Cox et al	Garcia Esteve et al	Juanita Nieto, Anna Gimene,Antoni Tormo, Francesc Mussons	مقال	/	/	الحساسية والنوعية ألفا كرونباخ	84	/	إسبانيا	الانجليزية	Paternal symptoms of anxiety and depression in the first month after childbirth: A comparison between fathers of full term and preterm infants
Bond et al	Chan, Chi, Lin, and Ye	Wen Hsin Changa, Chia-Huei Wub, Che-Chun Kuoc, Lung Hung Chend	مقال	/	/	الصدق التمييزي ألفا كرونباخ	76	/	الطيوان	الانجليزية	The role of athletic identity in the development of athlete burnout: The moderating role of psychological flexibility
Link	Ilic et al	Diane Quinn, Bradley Weisz,	مقال	/	/	التحليل العاملي ألفا كرونباخ	288	76-19سنة	الولايات المتحدة	الانجليزية	Impact of active concealment of stigmatized identities on physical

		ElizabetLawner									and psychological quality of life
Jerusalem & Schwarzer	Leme et al	Jose Oliveir, Marina Coimbra, Anne Fontaine	مقال	/	/	التحليل العاملي التوكيدي ألفا كرونباخ	620	18-30 سنة	البرتغال	الانجليزية	Family support in the transition to adulthood in Portugal e Its effects on identity capital development, uncertainty management and psychological
NCES	/	Grigg,Perera,McIlveena,Svetleff	مقال	/	/	ألفا كرونباخ	400	/	أستراليا	الانجليزية	Relations among math self efficacy, interest, intentions, and achievement: A social cognitive perspective
Gadzella &Masten	/	Mehmet Karaman, JoshuaWatson	مقال	/	/	الصدق التباعدي ألفا كرونباخ	307	/	الولايات المتحدة	الانجليزية	Examining associations among achievementmotivation, locus of control,academic stress, and life satisfaction: A comparison of U.S. and international undergraduate students

Schonbrodt & Gerstenberg	/	Philipp Kopp, Veit Senne, Hugo Kehr, Peter Gropel	مقال	/	/	ألفا كرونباخ	255	18-60 سنة	ألمانيا	الانجليزية	Achievement motive, autonomous motivation, and attendance at fitness center: A longitudinal prospective study
Taylor & Ntoumanis, 2007	/	Inok Ahn, Ming Ming Chiu, Helen Patrick	مقال	/	/	صدق تقاربي ألفا كرونباخ	697	/	الصين	الانجليزية	Connecting teacher and student motivation: Student-perceived teacher need-supportive practices and student need satisfaction
Dweck et al	Gleru, Loyen, & Rikers	Michiel Boncquet, Bart Soenens,Karine Verschueren, Jeroen Lavrijsen, Nele Flamant, Maaren Vansteenkiste	مقال	/	/	ألفا كرونباخ	2564	/	الولايات المتحدة	الانجليزية	Killing two birds with one stone: The role of motivational resources in predicting changes in achievement and school well-being beyond intelligence
Wentzel's	Midgley et al	Teresa Gonçal es, Markku Niemivirta, Marina	مقال	/	/	التحليل العاملي التوكيدي ألفا كرونباخ	386	/	البرتغال	الانجليزية	Identification of students' multiple achievement and social goal profiles

		S.Lemos									and analysis of their stability and adaptability
Babad	Ouyang	Xiao-xian Li, Shao-Ying Gong, Hong-po Zhang, Quan-lei Yu, Zhi-jin Zhou	مقال	/	/	التحليل العاملي التوكيدي ألفا كرونباخ	512	/	الصين	الانجليزية	Perceived teacher support and creative self-efficacy: The mediating roles of autonomous motivation and achievement emotions in Chinese junior high school students
Deniz ME, Ozer E, Isık E	Petrides KV, Furnham A	Handan Alan, Feride Eskin Bacaksiz, Arzu Kader Harmancı Seren, and Hasan Anıl Kurt	مقال	/	/	ألفا كرونباخ	104	/	تركيا	الانجليزية	Evaluating the Relationship Between Burnout Levels and Compassion Fatigue, Emotional Intelligence, and Communication Skills of Organ Transplant Coordinators
Bruni et al	Zhang et al	Xiaoting Liu, Lijin Zhang, Guoqiang Wu, Rui Yang, Yuan Liang	مقال	/	/	ألفا كرونباخ	1226	11-15 سنة	الصين	الانجليزية	The longitudinal relationship between sleep problems and school burnout in adolescents: A cross-lagged panel analysis

Marsh	Daumiller, Grassinger, Engelschalk, & Dresel	Daumiller, Raven Rinas, Julia Hein, Stefan Janke, Oliver Dickh¨auser, Markus Dresel	مقال	/	/	التحليل العاملي التطبيق وإعادة التطبيق	307	/	ألمانيا	الانجليزية	Shifting from face-to-face to online teaching during COVID-19: The role of university faculty achievement goals for attitudes towards this sudden change, and their relevance for burnout/engagement and student evaluations of teaching quality
Maslach et al	/	Farinaz Havaei , Oscar Astivia, Maura MacPhee	مقال	/	/	التحليل العاملي ألفا كرونباخ	537	/	كندا	الانجليزية	The impact of workplace violence on medical-surgical nurses' health outcome: A moderated mediation model of work environment conditions and burnout using secondary data
Baillod and Semmer	/	Luisa Barthauer, Philipp Kauche, Daniel Spurk, Simone Kauffeld	مقال	/	/	ألفا كرونباخ	385	/	ألمانيا	الانجليزية	Burnout and career (un)sustainability: Looking into the Blackbox of burnout triggered career turnover intentions

Vrklevski & Franklin	/	Hazen, Carlson, Bowers, Fessinger, Mossman, Bahm, Hauptman, Brank, Gilkerson	مقال	/	/	ألفا كرونباخ/ التطبيق وإعادة التطبيق	63	/	الولايات المتحدة	الانجليزية	Evaluating the Facilitating Attuned Interactions(FAN) approach: Vicarious Trauma, professional Burnout, and reflective practice
Davis	Siu & Shek	Mingxiao Liu, Ning Wang, Pengcheng Wang, Haomeng Wu, Xianger Ding, Fengqing Zhao	مقال	/	/	التحليل العاملي ألفا كرونباخ	119	/	الصين	الانجليزية	Negative Emotions and Job Burnout in News Media Workers: A Moderated Mediation Model of Rumination and Empathy
Weeks, Heimberg, &Rodebaugh	/	Taylor Dryma, Shani Gardne, Justin W.Weeks, Richard G.Heimberg	مقال	/	/	ألفا كرونباخ	129	/	الولايات المتحدة	الانجليزية	Social anxiety disorder and quality of life: How fears of negative andpositive evaluation relate to specific domains of life satisfaction

Herman and Polivy's	Boyce	Jessi ca A.Boyc, RoelineG.Kuij, DavidH.Gleaves	مقال	/	/	ألفا كرونباخ	107	/	نيوزلندا	الانجليزية	Positive fantasies or negative contrasts: The effect of media body ideals on restrained eaters' mood, weight satisfaction, and food intake
Xie	Li et al	Angyang Li, Shuo Wang, Minmin Cai, Ruiqi Sun, Xiangping Liu	مقال	/	/	التحليل العاملي التوكيدي ألفا كرونباخ	337	/	الصين	الانجليزية	Self-compassion and life-satisfaction among Chinese self-quarantined residents during COVID-19 pandemic: A moderated mediation model of positive coping and gender
Jonason & Webster	؟	Heyun Zhang, Huanhuan Zhao	مقال	/	/	صدق البناء ألفا كرونباخ	675	17-22 سنة	الصين	الانجليزية	Dark personality traits and cyber aggression in adolescents: A moderated mediation analysis of belief in virtuous humanity and self-control
Weathers et al	/	AndreaM.Despotes, Benjamin C. Darnell, Sarah R. Lee, David P.	مقال	/	/	ألفا كرونباخ	402	18-34 سنة	الولايات المتحدة	الانجليزية	Past adversity and obsessive-compulsive symptoms: The mediating roles of posttraumatic cognitions and posttraumatic

		Valentiner									stress symptoms
Foa et al	Gonner et al	Miegel, Moritz, Wagener, Cludiu, Jelinek	مقال	/	/	ألفا كرونباخ	105	18-70 سنة	ألمانيا	الانجليزية	Self-esteem mediates the relationship between perfectionism and obsessive compulsive symptoms
Silverman et al	/	Shaw, Halliday, Tonarely, Ehrenreich-May	مقال	/	/	ألفا كرونباخ	277	8-17 سنة	الولايات المتحدة	الانجليزية	Relationship of affect intolerance with anxiety, depressive, and obsessive-compulsive symptoms in youth
Schaufeli and Bakker	/	Lalatendu Jena, Sajeet Pradha, Nrusingh Panigrahy	مقال	/	/	التحليل العاملي التوكيدي	394	29-35 سنة	الهند	الانجليزية	Pursuit of organisational trust: Role of employee engagement, psychological well-being and transformational leadership
Verkuyten & Martinovic	/	Zhang,Verkuyten, Weesie	مقال	/	/	التحليل العاملي التوكيدي ألفا كرونباخ	1939	/	هولندا	الانجليزية	Dual identity and psychological adjustment: A study among immigrant-origin members
Testa et al	ترجمة الباحث	Jiwen Zhang, Herman Hayming	مقال	/	/	ألفا كرونباخ	361	/	الصين	الانجليزية	The buffer of resilience in the relations of gender-related

	نفسه	Lo, Alma Maylan Au									discrimination, rejection, and victimization with depression among Chinese transgender and gender non-conforming individuals
Henderson and Zimbardo	Wang	Fengqiang Gao, Yue Sun, Yuci Zhou, Mohan Sang, Jinzhe Zhao, Lei Han	مقال	/	/	التحليل العاملي التوكيدي ألفا كرونباخ	539	/	الصين	الانجليزية	Shyness and depression: The mediating roles of interpersonal competence, dormitory belonging and inferiority
Bernstein et al.	Sar, Ozturk, and Ikikardes	Gizem Akcan, Erdinç ¨Oztürk, Barıs¸han Erdo˘gan	مقال	/	/	ألفا كرونباخ	735	/	تركيا	الانجليزية	The investigation of the mediating role of coping trategies on the relationship between childhood traumas, depression and alcohol use disorder in university students
Jiang	Jin et al	Wei Zhang, Ting-ting Yan, Ya-song Du, Xiao-hong Liu	مقال	/	/	ألفا كرونباخ	102	/	الصين	الانجليزية	Relationship between coping, rumination and posttraumatic growth in mothers of children with autism spectrum disorders

Bellini	/	Danielle Ung, Nicole McBrid, Amanda Collier, Robert Selles, Brent Small, Vicky Phares, Eric Storch	مقال	/	/	ألفا كرونباخ	81	9-17 سنة	الولايات المتحدة	الانجليزية	The relationship between peer victimization and the psychological characteristics of youth with autism spectrum disorder
Buhrmester & Furman	ترجمة الباحث نفسه	Michael A.Petal, Richard P.Hastings, Susie Nash, Louise M. Hall, Helen Joannidi,Alan Dowey	مقال	/	/	ألفا كرونباخ	166	5-17 سنة	بريطانيا	الانجليزية	Psychological adjustment and sibling relationships in siblings of children with Autism Spectrum Disorders: Environmental stressors and the Broad Autism Phenotype
Derogatis, Lipman, & Covi	/	Susan Gau , Miao Chou, Huey Chiang, Ju Lee,Ching Wong, Wen Chou,Yu Wu	مقال	/	/	ألفا كرونباخ/ التطبيق وإعادة التطبيق	151	/	الطيوان	الانجليزية	Parental adjustment, marital relationship, and family function in families of children with autism

Hoge		Lalatendu Jena, Pratishtha Bhattacharyya,Sajeet Pradhan	مقال	/	/	ألفا كرونباخ	288	/	الهند	الانجليزية	Am I empowered through meaningful work? The moderating role of perceived flexibility in connecting
Kessler et al	Furukawa et al	Keiko Kunie, Norito Kawakami, Akihito Shimazub, Yuki Yonekura, Yuki Miyamoto	مقال	/	/	ألفا كرونباخ	789	/	اليابان	الانجليزية	The relationship between work engagement and psychological distress of hospital nurses and the perceived communication behaviors of their nurse managers: A cross-sectional survey
Lei et al	Lian, Sun, Niu, & Zhou	Shuai Lian, Xiao Sun, Geng Niu, Xiu Yang, Zong Zhou, Chen Yang	مقال	/	/	ألفا كرونباخ	754	/	الصين	الانجليزية	Mobile phone addiction and psychological distress among Chinese adolescents: The mediating role of rumination and moderating role of the capacity to be alone
Carmeli, Reiter-Palmon, and Ziv	/	Reuven Hirak, Ann Chunyan	مقال	/	/	ألفا كرونباخ	279	/	الولايات المتحدة	الانجليزية	Linking leader inclusiveness to work unit performance

الملحق 06: نموذج شامل لخصائص القياس

