مجلة العلوم الاجتماعية والإنسانية، جامعة باتنة1 المجلد (23) – العدد (01) (جوان، 2022) - ص 631- 648
ISSN 1111-5149 /E. ISSN 2430-2588

تقييم إجراءات التحقق من صدق الاختبارات النفسية من خلال عينة من الدراسات العربية والأجنبية

Evaluate Procedures to Verify the Validity of Psychological Tests through a Sample from Arab and Foreign Studies

سكينة تشلابي *[1]، مخبر القياس والدراسات النفسية، (جامعة لونيسي علي)
es.tchalabi@univ- blida2.dz
عبد العزيز بوسالم[2]، مخبر القياس والدراسات النفسية، (جامعة لونيسي علي)،
boussalem.azzizz@yahoo.fr

تاريخ الاستلام	تاريخ القبول
2021-03-10	2022-01-17

ملخّص

يعتبر الصدق أهم دليل على صلاحية الاختبار النفسي، لذلك هدفت هذه الدراسة إلى محاولة وضع بعض المعايير التي يستند إليها عند إعادة التحقق من صدق الاختبارات النفسية، ومن ثم اعتماد تلك المعايير لتقييم إجراءات التحقق من صدق 91 اختبار نفسي في 53 دراسة عربية وأجنبية، وذلك باستخدام شبكة تقييم من صنع الباحثين، وبإتباع المنهج الوصفي التقويمي، وقد أظهرت النتائج أن صدق المحكمين هو الأكثر استخداما يليه صدق الاتساق الداخلي، وفي ما يخص كفاية هذه الأدلة فـ58 اختبار لم تكن طرق التحقق من صدقه كافية سواء في الدراسات العربية أو الأجنبية، إلا أن هذه الأخيرة كانت أكثر تقصيرا فـ78.2٪ منها لم تكن كافية في حين أن الدراسات العربية لم تكن كافية بنسبة 58.8٪، أما في ما يخص العينة التي استخدمت لإعادة التحقق فـ27 اختبار عينته لم تكن ممثلة و12 اختبار حجم العينة فيه غير كاف، إلا أن الباحثين في جميع الدراسات ما عادا دراسة واحدة، اختاروا اعتماد الاختبار والعمل به.

الكلمات المفتاحية: صدق؛ إعادة تحقق؛ معايير؛ صلاحية.

Abstract

Validity is the most important evidence of the validity of a psychological test, so this study aimed to try to set standards upon which to re-check the validity of psychological tests, and then adopt those standards to evaluate procedures for verifying the validity of 91 psychological tests in 53 Arab and foreign studies, using an evaluation network. Made by researchers, following the descriptive evaluation method, The results showed that the validity of the experts was the most commonly used validity, followed by the validity of the internal consistency. With regard to the adequacy of this evidence, 58 tests were not sufficient methods to verify its validity, whether in Arab or foreign studies, but the latter was more short-lived, 78.2% of them were not sufficient, while the Arab studies were not sufficient by 58.8%. As for the sample that was used for re-verification, 27 tests sampled were not representative and 12 tests of sample size were insufficient. However, researchers in all studies except for one study chose to rely on the test and work on.

Keywords: validity; reverification; criteria; validity.

*المؤلف المرسل

مقدّمة

إن استخدام أي اختبار نفسي، يفرض على صاحبه مجموعة من الشروط والمعايير الواجب التقيد بها قبل الاستخدام، سواء كان الغرض من ذلك الاختبار التشخيص أو التصنيف أو غيرها، فإنه لزاما عليه التحقق من صلاحية هذا الاختبار وتطابقه للغرض المراد منه، والعينة التي سيطبق عليها، وغيرها من الشروط، وهذا انطلاقا من أن الصدق خاصية نسبية، فكما ورد في (فرج، 2017 : 265) بأن الصدق ليس أمرا مطلقا حسب كل منThorndike وHagen بل يختلف من اختبار لآخر، بحيث لا نستطيع أن نقول إن الاختبار إما صادق أو غير صادق، بل نقول إنه صادق بقدر أو بدرجة ما، والصدق تقدير لما إذا كان الاختبار "يقيس ما نريد أن نقيسه به، وكل ما نريد أن نقيسه به، ولا شيء غير ما نريد أن نقيسه به"

ولهذا قد يضطر الباحث أو الفاحص بصفة عامة إلى إعادة التحقق من صدق وثبات الاختبار النفسي الذي يريد الاعتماد عليه، وبخاصة إذا كان هذا الأخير غير مكيف للبيئة الجديدة، أو تم تغيير مجتمع البحث عن المجتمع الأصلي للاختبار، وغيرها من الضوابط، وهذا ما أشار إليه أيضا كل من (هوييت و كرامر، 2016: 400) فالصدق والثبات حسبهما ليست خصائص ذاتية للاختبارات النفسية؛ لذلك فإنها ستختلف وفقًا للسياق والغرض من القياس والعينات المختلفة للمشاركين، لذلك نلاحظ أنه يمكننا استخدام الاختبارات التي تصمم لغرض البحث فقط، رغم انخفاض صدقها وثباتها نوعا ما، على عكس الاختبارات التي تهدف إلى تقييم الأفراد في المواقف السريرية أو التعليمية الضرورية يجب أن تتمتع بثبات وصدق أعلى؛ لأنها تستخدم لتقييم الأفراد. وهنا وجب التنبيه إلى أنه لإعادة التحقق من صلاحية الاختبارات النفسية، علينا الالتزام بإجراءات تقنية معينة ومحددة حسب نوع كل اختبار والغرض منه وغيرها، ولا يجب الاكتفاء بصدق وثبات الاختبار كما هو مقدم في دليله إلا في حالات معينة.

كذلك أشار (بوسالم، 2017) إلى أن معظم الاختبارات والمقاييس الشائعة في الدول العربية، يتم إعدادها في الغرب، وبخاصة في الولايات المتحدة؛ لذلك فهي ثقافيًا غير متحررة إلى حد ما، وبالتالي يصعب استخدامها في الواقع. والثقة بموضوعية نتائجها. ومع ذلك أثبت (نعموني و براق، 2015) أنه رغم ذلك تم نقل هذه الاختبارات من الدول الغربية بترجمتها إلى اللغة العربية في أحسن الأحوال واستخدامها. وهذا ما جعل أغلب الاختبارات النفسية التي نستخدمها غربية المنشأ محلية التطبيق. وهو ما يستدعي على الأقل إعادة التحقق من صدقها وثباتها.

فمن خلال الاطلاع على بعض الدراسات الأكاديمية التقييمية لصلاحية الاختبارات النفسية ، مثل دراسة (Irene, Luisa, Luca, & Alvisa, 2017) ودراسة (Polit D. , 2015) ودراسة (بوسالم، 2017) وكذا دراسة (بن صافي و طاجين، 2013)، تبين لنا أن العديد من الباحثين يكتفون بالخصائص السيكومترية الواردة للاختبارات في البيئات الأصلية رغم تغير

العينة والهدف من الاختبار وغيرها، في حين عمل البعض الآخر على استخراج دلالات صدق وثبات جديدة للاختبارات كإجراء إحصائي روتيني لإعادة التحقق من صلاحية الاختبارات النفسية، دون الأخذ بالاعتبار طبيعة السمة المقاسة وما يناسبها من دلالات الصدق، أو مدى توافق نوع الصدق المستخدم مع الغرض من الاختبار، وغيرها من التجاوزات والأخطاء 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 وتفسير النتائج.

ومن خلال كل ما سبق يمكن طرح التساؤلات التالية:

- ما هي الدوافع والمعايير التي يعتمد الباحثون عليها عند إعادة التحقق من صدق الاختبارات النفسية؟
- ما الأدلة الأكثر استخداما في إعادة التحقق من صدق الاختبارات النفسية في الدراسات الأكاديمية العربية والأجنبية؟
- هل تعدّ أدلة الصدق المستخرجة كافية في الدراسات العربية والأجنبية؟ وأيهما أكثر تقصيرا؟ وهل تم تبرير استخدام تلك الأدلة دون غيرها؟
- هل كان كل من حجم وتمثيل العينة، وأيضا قرارات الباحثين مناسبة؟

2-أهداف الدراسة

تهدف الدراسة الحالية إلى:

- تحديد دوافع إعادة التحقق من صلاحية الاختبارات النفسية.
- ضبط أهم المعايير الواجب أخذها بعين الاعتبار عند التحقق من صدق الاختبارات النفسية.
- مناقشة مصغرة لإشكالية الاتساق الداخلي بين الصدق والثبات.
- الكشف عن مدى التزام الباحثين في عينة من الدراسات العربية والأجنبية في احترام هذه المعايير.

3-أهمية الدراسة

تتمحور أهمية هذه الدراسة حول قلة الدراسات والمراجع العربية منها خاصة، التي تهدف إلى ضبط الأسباب والدوافع وراء إعادة التحقق من صلاحية الاختبارات، وكذا تحديد معايير واضحة للتحقق من الصدق مستندة إلى مراجع مهمة وموثوق بها في القياس النفسي.

4-مفهوم الصدق

إنّ التعريف القاموسي لمصطلح "الصدق " يعني:" ما إذا كان شيء ما متأسسا جيدا، أو منطقيا أو يمكن تبريره". (هوييت وكرامر، 2016، ص400)، وانطلاقا من هذا التعريف، وغير بعيد عن الفكرة نفسها عرف ميسك الصدق بأنه "حكم تقييمي متكامل للدرجة التي تدعم بها

الأدلة التجريبية والأسباب النظرية كفاية وملاءمة الاستدلالات والإجراءات القائمة على درجات الاختبار. فعلى هذا النحو، الصدق هو ملخص استقرائي لكل الأدلة الموجودة والعواقب المحتملة لتفسير الاختبار واستخدامه".(Messick S. , 1987, p. 1)

كما هو واضح، فإن الصدق لم يعد يتعلق بالدرجة الكمية، بل هو خاصية تعنى بتفسير وتبرير دقة وملاءمة النتائج أكثر مما هي مرتبطة بالاختبار في حد ذاته.

5-متى يجب إعادة التحقق من صدق الاختبار النفسي؟

عادة ما يلجأ معظم الباحثين والمتخصصين إلى استخدام الاختبارات النفسية الجاهزة بدل بنائها، وذلك طبعا توفيرا للجهد والوقت والمال. فيقوم الباحثون بالاستعانة بدل ذلك بالاختبارات النفسية الجاهزة. إلا أن الإشكال هنا هو استخدام هذه الاختبارات كما هي دون إعادة التحقق من صلاحيتها متجاهلين المتغيرات التي قد تؤثر فيها، وكأن مستخدم الاختبار جازم بصلاحيته لكل الأزمنة والعينات وغيرها، لذلك حاول الباحثون تلخيص أهم الأسباب التي تلزم مستخدم الاختبار على إعادة التحقق من صلاحيته.

5-1-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 وغيره، لذلك وجب على مستخدم الاختبار أن يتحقق من مطابقة أن بينات صدق الاختبار مناسبة للغرض الذي يريد الباحث استخدامه فيه، فحسب كل من (Crocker & Algina, 1986, p. 304) عند اختيار اختبار لغرض معين، على مستخدم الاختبار مسؤولية واضحة لضمان أن له أدلة صدق مناسبة للاستخدام المقصود في هذا الموقف. أي أن مستخدم الاختبار لا يكتفي بصدق الاختبار المقدم من طرف مطوره عندما يختلف الغرض أو الهدف منه، بل وجب عليه إعادة التحقق من صدقه واختيار الإجراءات المناسبة حسب الغرض الذي يريد الباحث استعماله فيه. وهذا ما أكده (هوييت و كرامر، 2016: 393) بقولهما إن" الصدق والثبات ليست خصائص ذاتية للمقاييس السيكولوجية، فهما سوف يتباينان وفقا لسياق القياس والغرض منه". وغير بعيد عن هذه الفكرة أضاف علام (2000) أن الاختبار أو الأداة يمكننا استخدامها لأغراض مختلفة ومتعددة حسب المواقف، لذلك وجب أن يتم التحقق من صدق الأداة لكل غرض منها. (علام، 2000، صفحة 190) وهذا بالضبط ما أوضحناه أيضا (Anastasi & Urbina, 1997) بأن الاختبار نفسه عندما يستخدم لأغراض مختلفة ينبغي أن يتم التحقق من صدقه بطرق مختلفة.

5-2-مجتمع وعينة الاختبار

عند تصميم أي اختبار نفسي يعمد مصممه إلى وصف العينة التي صنع 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، فالصدق حسب (Gay, Mills, & Airasian, 2012, p. 262) متعلق بالتفسير المطلوب والعينة المدروسة. وبتعبير آخر: لا يفترض أن نقول "الاختبار صادق بصفة مطلقة " وإنما ينبغي أن نقول: "هذا الاختبار صادق لهذا التفسير، وهذه العينة". وبالتالي بمجرد تغيير المجموعة، تتغير معه خصائص الاختبار، فلا يمكننا الجزم بصلاحيته إلا بالتحقق من ذلك، فحسب (Anastasi & Urbina, 1997, p. 189) ربما يقيس الاختبار نفسه أغراضا مختلفة عندما يُطبق على أفراد مختلفين. 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. (فرج، 2017: 265)

5-3-عامل الزمن

قد يختلف صدق الاختبار أيضاً بين فترة زمنية وأخرى، فنتيجة للتقدم في التعليم والتطور الاجتماعي والثروة التكنولوجية الاتصالية حسب (فرج، 2017) أصبحت البنود التي كانت صادقة في قياس الذكاء في عشرينيات وثلاثينات القرن الماضي لا تصلح الآن ولا يمكن استمرار وجودها في مقياس صدق للذكاء في القرن الواحد والعشرين. كما يتعين أن يضع الباحث في اعتباره عند تقدير صدق اختبار ما أن الأمر يتطلب مراجعة إجراءات الصدق بين فترة وأخرى؛ لإعادة تقدير تأثير المتغيرات الزمنية على المفاهيم التي يتضمنها الاختبار. فالاختبار الصادق في فترة معينة قد لا يكون صادقاً في فترة تالية.

5-4-عامل الثقافة

يبدو عامل الثقافة على درجة كبيرة من الأهمية عند نقل أو ترجمة اختبار من حضارة ذات خصائص معينة إلى حضارتنا المحلية التي تتسم بخصائص نوعية واضحة، تنعكس على القيم وأساليب السلوك السائدة ودلالتها الاجتماعية. (فرج، 2017، ص238) كذلك الأخذ بعين الاعتبار هذا 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 العربية وهو بالطبع أمر خاطئ.

6-معايير الصدق

رغم فصلنا لأنواع الصدق وكأن لكل نوع إجراءات ومعايير خاصة به، إلا أننا نؤكد أننا نتفق مع الكثير من علماء القياس في مجال علم النفس الذين دعوا إلى تجميع أنواع الصدق تحت عنوان صدق البناء منهم (Cronbach L. , 1951) (Anastasi A)،. 1988) (Messick S. , 1995)، فالفصل هنا كان بعد اتفاق العديد من المتخصصين الدوليين بنسبة (77 ٪ اتفاق) في دراسة (Mokkink, et al., 2010, p. 746) وتبرير ذلك أنه على مستوى التصميم والأساليب هذه الأشكال الثلاثة من الصدق تختلف.

ومن أجل تقييم جودة المقاييس، قدم الباحثون بعض المعايير المتعلقة بكل نوع على حدة في شكل فقرات منفصلة، وقد تم استنباط هذه المعايير المتفق عليها في المقالات والمراجع التالية، مع الإشارة إلى المراجع في حال اختلاف الآراء:

(Messick، 1994)(AERA, APA, & NCME, 1999) (Cronbach & Meehl, 1955) (Anastasi A. , 1988) (Reynolds & Livignston, 2012) (Anastasi & Urbina, 1997) (Polit & Beck, 2006) (Hambleton & Kanjee, 1993) (Lehmnn & Mehrens, 1973) (Wilson & Cleary, 1995) (Guyatt, Feeny, & Patrick, 1993) (Keszei, Novak, & Streiner, 2010) (Roach, 2006) (Kimberlin & Winterstein, 2008) (Fayers & Machin, 2007) (Gay, Mills, & Airasian, 2012) (Aaronson, et al., 2002) (Crocker & Algina, 1986) (Mokkink, et al., 2010) (Messick S. , 1995) (Cortina, 1993) (Outco, 2002) (علام، 2000)

1-6-معايير خاصة بصدق المحتوى

- يجب بداية تحديد الهدف من استخدام الاختبار، أي الغرض منه: التشخيص، والتمييز، والتقييم، أو التنبؤ وغيرها.

-المجموعة المستهدفة: أي المجتمع الذين تم تطوير الاختبار من أجله. هذا مهم للحكم على أهمية وشمولية الفقرات. وبالتالي إثبات صلاحية المقياس للعينة المراد دراستها.

-المفاهيم التي يهدف الاختبار إلى قياسها: للحكم على مدى ملاءمة الاختبار لغرض معين، من المهم أن يقدم المطورون والمستخدمون إطارًا واضحًا للمفهوم العام المراد قياسه. وعلى مستخدم الاختبار التحقق من مطابقة هذا المفهوم للسمة المراد دراستها.

-اختيار الفقرة وتقليل الفقرات: يجب تبرير طرق اختيار الفقرات أو تقليلها، وبخاصة عند الترجمة، وتنفيذ دراسة تجريبية لفحص قابلية القراءة والفهم والإبلاغ عنها. ويجب أن تعكس الفقرات الموجودة في الاختبار المجالات المهمة للمجتمع المستهدف قيد الدراسة. لذلك يجب إشراك العينة المستهدفة في أثناء اختيار الفقرة للإبقاء عليها أو إلغائها.

- تفسير الفقرات: عند ترجمة أو إعادة صياغة المفردات لسبب ما، يجب أن تكون الفقرات قصيرة وبسيطة وألا تحتوي على كلمات صعبة أو مصطلحات لغوية معقدة، وأيضا ألا تختلف لا في المعنى ولا السياق عن الفقرات الأصلية، كذلك يجب ألا تتكون البنود من سؤالين في الوقت نفسه .وأيضا يجب تحديد الفترة الزمنية التي تشير إليها الأسئلة بوضوح وتبريرها.

- بما أنه لا يوجد اختبار إحصائي لتقييم صحة المحتوى على وجه التحديد، فإن الباحثين يستخدمون منهجاً نوعياً، يعتمد على النقاط المذكورة سالفا، وعلى تقييم الخبراء؛ أي ما يطلق عليه صدق المحكمين. ويجب أن يكون الرقم القياسي للتطابق المقبول بين المحكمين 0.80 على الأقل، ويفضل أن يكون أعلى من 0.90.

-إن أفضل المجالات مناسبة لملائمة هذا النوع من الصدق هو مع الاختبارات التحصيلية، مع إمكانية استخدامه مع اختبارات مهنية معينة، في حين أن التحقق من الصدق بهذه الطريقة يكون غير مناسب فيما يخص اختبارات الاستعدادات واختبارات الشخصية حسب Anastasi)

(AERA, APA, & Urbina, 1997). على عكس ما أشارت إليه وثيقة اللجنة المشتركة (NCME, 1999)إمكانية التحقق من صدق المحتوى بالنسبة لاختبارات الاستعدادات واستبيانات الشخصية والميول، كذلك مقاييس التوافق والسلوك الاجتماعي، حيث يمكن الاعتماد على عينات من السلوك أو عينات من الملاحظات في مواقف مدروسة، ولكن لا ينبغي في هذه الحالات الاقتصار على صدق المحتوى، وإنما يجب الاستناد إلى أدلة من جوانب الصدق الأخرى.

-وجب بنا كذلك التنويه إلى عدم الخلط بين صدق المحتوى والصدق الظاهري، فالصدق الظاهري لا يعد صدقا بالمعنى التقني، فحسب (Anastasi & Urbina, 1997, p. 166) "فهو لا يشير إلى ما يقيسه الاختبار بالفعل، وإنما ما يبدو ظاهريا أنه يقيسه".

-كذلك هنالك نوع آخر وهو الصدق الذاتي الذي يستخرج من خلال حساب الجذر التربيعي للثبات، الأمر الذي يتعارض مع مبادئ القياس الأساسية التي تنص على أن كل اختبار صادق ثابت، وليس كل اختبار ثابت صادق.

2-6-معايير خاصة بصدق المحكي

-إن دليل الصدق المحكي مرتبط عادة بمقاييس الاستعداد والميول والشخصية، واختبارات الاستعدادات بصفة عامة تبدو أكثر تعلقا من غيرها بصدق التنبؤ، كذلك وجب الإشارة إلى أن هذا الأخير مهم للغاية للاختبارات التي تستخدم في تصنيف الأفراد وانتقائهم.

-الصدق المحكي هو العلاقة بين درجة اختبار معين وبعض المعايير الخارجية. ويجب أن يكون هذا المعيار مقياسا مقبولا على نطاق واسع، مع خصائص أداة التقييم نفسها، أي أداة أو معيار يعدّ "المعيار الذهبي".

-لا يمكن العثور على "المعيار الذهبي" بسهولة في جميع السمات. 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.

-لتفسير معامل الصدق ينبغي أن يأخذ الباحث بعين الاعتبار عدداً من الظروف المصاحبة، فالارتباط الذي يتم الحصول عليه ينبغي بالطبع أن يكون مرتفعاً بدرجة تجعله "دالاً إحصائياً". وعادة نعطي تقييمًا إيجابيًا لصدق الاختبار إذا كان الارتباط بالمعيار 0.70 على الأقل.

-يجب الإشارة دائما إلى المدة الزمنية التي استغرقها الباحث بين التطبيقين وتدوينها.

-إن عدم تجانس العينة يكون موائماً لقياس الصدق، فكلما اتسع نطاق الدرجات، زادت قيمة الارتباط.

-عند استخدام معامل الارتباط بيرسون، وجب التحقق من شروطه أولا من حيث حجم العينة الذي يجب ألا يقل عن 50 مفردة، وأيضا يشترط أن تكون العينة ممثلة للمجتمع عدا العلاقة الخطية وما ذكر سابقا.

3-6-معايير صدق التكوين الفرضي

لقد قام فريق COSMINفي دراسة (Mokkink, et al., 2010, p. 740) بالتمييز بين جوانب من صدق البناء، وهي الصدق الهيكلي، واختبار الفرضيات، والصدق عبر الثقافات بنسبة (76٪ اتفاق) بين الخبراء.

-ينبغي أن يقيِّم صدق التكوين الفرضي (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 الصلاحية عبر الثقافات فقط للأدوات المترجمة.

تذكر أنستازي وأوربينا (Anastasi & Urbina, 1997) أن الاتساق الداخلي أساسا يقيس التجانس، ولذلك فإن الإضافة التي توفرها بيانات الاتساق الداخلي لصدق الاختبار محدودة، ففي ضوء غياب بيانات من خارج الاختبار نفسه، فلن نعرف إلا القليل عما يقيسه الاختبار. وعلى هذا فمن الأفضل الاكتفاء باستخدام أساليب الاتساق الداخلي في مجالي تحليل البنود لأغراض التمييز والصعوبة، أو لحساب ثبات الاختبار. وعبرت كورتينا (Cortina, 1993, p. 101) أنه ليس من المناسب اعتبار الاتساق الداخلي مقياساً للصدق لأنه لا يخبرك ما إذا كانت الفقرات تقيس البناء الذي يراد قياسه. فإنه يخبرك فقط ما إذا كان الاختبار يقيس شيئا مشتركا، ولكن ما هو ذلك لا يزال غير معروف.

كما يطرح الاتساق الداخلي إشكالا عويصا، لطالما صادف الباحثين عند استخدامه، بحيث يجد الباحث نفسه في تناقض بين نتائج الاتساق الداخلي والتحليل المنطقي لبنود الاختبار، حيث إنه في الاتساق الداخلي عند دراسة الارتباط بين الفقرات، إذا ما تبين أن أحد البنود أو عدة بنود غير دالة، فإنه وجب حذفها انطلاقا من المنطق الإحصائي. في المقابل فإن المتفحص لتلك البنود قد يجدها تجسد وتعكس السمة المدروسة أكثر من البنود الأخرى التي قبلت إحصائيا ربما، وهنا يوقع الباحث في إشكال محرج؛ هل يحذف الفقرة وهي مهمة بالنسبة للسمة، أم يبقيها وهي غير دالة. ويرجع (تيغزة، 2017) السبب في هذا التناقض إلى تشويه الاستجابات.

-هناك عدة استراتيجيات لتأكيد صلاحية البناء من خلال اختبار الفرضية. من أهمها المقارنة الطرفية، وأيضا الصدق التقاربي والتباعدي وغيرها.

-إن معايير تفسير معاملات الارتباط المستخدمة في أدلة هذا النوع من الصدق، مثل الاتساق الداخلي أو التحليل العاملي وغيرها، هي نفسها المعايير التي ذكرناها سابقا عند تفسير معاملات ارتباط الصدق المحكي.

-أما الصلاحية عبر الثقافات فهي الأدلة التي تدعم الاستدلال على أن الاختبار الأصلي والاختبار المترجم متساويان من حيث التكييف الثقافي. فيجب النظر في ترجمة الاختبار. فمثلا عند ترجمة البنود هل ترجمها مترجمون مستقلون مرة أخرى؛ أي الترجمة والترجمة الخلفية، وإذا

كانت الترجمة قد نقحتها لجنة خبراء، وغيرها من الضوابط التي تدل على صلاحية المقياس عبر الثقافات.

7- استراتيجية البحث وعينة الدراسة والأساليب الإحصائية

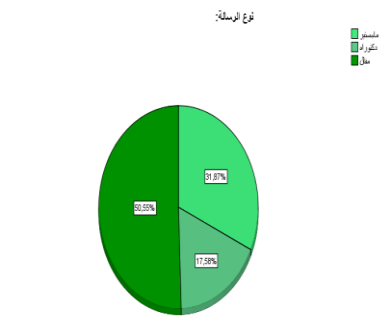
7-1-استراتيجية ومنهج البحث

أجري البحث عن عينة الدراسة التي تتمثل في الدراسات والمقالات سواء العربية أو الأجنبية من خلال قواعد البيانات الإلكترونية، فبالنسبة للدراسات العربية تم اعتماد قاعدة **المعرفة.** وأما في ما يخص الدراسات الأجنبية فقد تم الاعتماد على قاعدة Science Directوذلك باستخدام مزيج من عناوين الموضوعات التي تستخدم فيها عادة الاختبارات النفسية، ومن هذه العناوين مثلا: (الوسواس القهري، والاكتئاب، والاحتراق النفسي، والقلق، والذكاء.....)، وقد تم الاعتماد هنا على المنهج الوصفي التقويمي؛ وذلك لأنه ملائم لمتطلبات الدراسة وأهدافها، ويعتمد هذا المنهج في وصفه على التكرارات والنسب المئوية من أجل وصف وتقييم عينة من الدراسات العربية والأجنبية التي تم فيها إعادة التحقق من صدق الاختبارات النفسية قبل استخدامها.

7-2- عينة الدراسة

الجدول والشكل رقم (01): توزيع العينة حسب نوع الدراسة

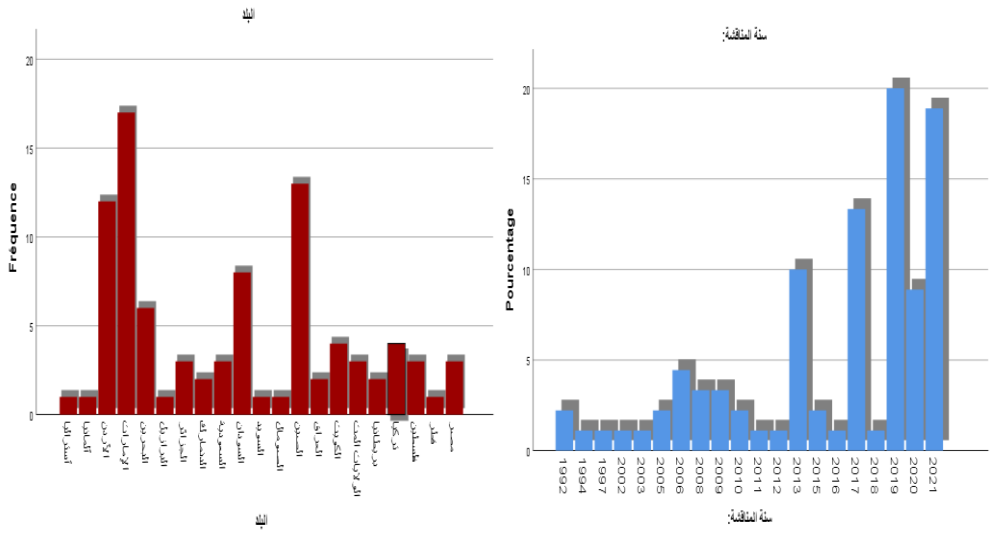
نوع الدراسة	تكرارات	النسب المئوية
ماجستير	29	%31.9
دكتوراه	16	%17.6
مقال	46	%50.5
الكلية	91	%100



المصدر: من إعداد الباحثين بالاعتماد على برنامج spss

تم فرز عناوين المقالات وملخصاتها والتحقق منها. ومن ثم اعتماد الدراسات التي تستوفي الشروط فقط، أما المواد التي لا تستوفي معايير القبول، فقد استبعدت استناداً إلى عناوينها أو مضمونها، وكانت معايير القبول في الدراسة هي الاعتماد على اختبار نفسي أولا، ثانيا أن يكون هذا الاختبار جاهزا، أي تم بناؤه في دراسة أخرى، ثالثا إعادة التحقق من صلاحيته، رابعا استبعاد الدراسات التي اعتمدت على بناء أو تكييف أو تقنين تلك الاختبارات. وقد تم في الأخير تقييم 91 اختبارا نفسيا أعيد التحقق من صلاحيته (للاطلاع على أسماء الاختبارات اضغط هنا). موزعا على 53 دراسة (للاطلاع على عناوين الدراسات وأسماء أصحابها اضغط هنا)، منها 12 أجنبية و41 عربية من مقالات ورسائل دكتوراه وماجستير موزعة كما في الجدول والشكل التاليين:

الشكل رقم (03): توزيع العينة حسب سنة المناقشة **الشكل رقم (04):** توزيع العينة حسب البلد



المصدر: من إعداد الباحثين بالاعتماد على برنامج spss

الجدول رقم (02): أنواع الصدق

أنواع الصدق المستخدمة في الدراسات

أنواع الصدق المستخدمة في الدراسات	التكرارات	النسبة المئوية
الصدق التمييزي	1	%1,1
الصدق الذاتي	1	%1,1
الصدق الظاهري	3	%3,3
الصدق الظاهري، الصدق الذاتي، صدق الاتساق الداخلي	1	%1,1
الصدق الظاهري، الصدق المحكي، الصدق التمييزي	1	%1,1
الصدق الظاهري، صدق الاتساق الداخلي	9	%9,9
الصدق الظاهري، صدق الاتساق الداخلي، الصدق التمييزي	1	%1,1
الصدق الظاهري، صدق البناء	3	%3,3
الصدق الظاهري، صدق المقارنة الطرفية، صدق الاتساق الداخلي	1	%1,1
الصدق الظاهري، صدق المقارنة الطرفية، صدق البناء، صدق الاتساق الداخلي	2	%2,2
الصدق المحكي	1	%1,1
الصدق المحكي، صدق الاتساق الداخلي	4	%4,4
صدق الاتساق الداخلي	10	%11,0
صدق الاتساق الداخلي، الصدق التمييزي، التحليل العاملي	1	%1,1
صدق البناء	1	%1,1
صدق البناء، الصدق المحكي	1	%1,1
صدق المحتوى	4	%4,4
صدق المحتوى، الصدق الظاهري، الصدق التمييزي	1	%1,1
صدق المحكمين	12	%13,2
صدق المحكمين، الصدق المحكي	1	%1,1
صدق المحكمين، صدق الاتساق الداخلي	5	%5,5
صدق المقارنة الطرفية، صدق البناء، الصدق المحكي	1	%1,1
لم يتم التحقق من الصدق	26	%28,6
الكلية	91	%100

كذلك عمد الباحثان إلى محاولة سحب العينة من عدة بلدان مختلفة، وفي سنوات متفاوتة كما هو واضح في الشكل أعلاه؛ كمحاولة للتوسع في تباين العينة بحيث تراوحت سنوات المناقشة للرسائل والنشر بالنسبة للمقالات بين 1992 و2021، أما البلدان فتوزعت على أكثر من 20 بلدا عربيا وأجنبيا.

7-3-الأساليب الإحصائية

من أجل الإجابة على تساؤلات الدراسة تم الاستعانة بالحزمة الإحصائية للعلوم الاجتماعية النسخة (SPSS V26) لاستخدام الإحصاء الوصفي من تكرارات، ونسب مئوية، وجداول متقاطعة.

8-أداة الدراسة

قام الباحثان باستخراج معلومات من الدراسات التي تستوفي معايير القبول، باستخدام شبكة تقييم منظمة ومسبقة التصميم. وشملت هذه الشبكة: بيانات عامة مثل اسم صاحب المقال أو الدراسة وسنة نشرها وعنوان الدراسة واسم الاختبار المستخدم، وكذلك البلد والمشرف بالنسبة لرسائل الماجيستر والدكتوراه، كذلك تضمنت الشبكة أسئلة حول أصل الاختبار: هل هو من البيئة نفسها أو بيئة مختلفة، وبيانات خاصة حول الصدق. (للاطلاع عليه اضغط هنا)

9-معايير تقييم جودة الصدق

استخدمت المعايير المفصلة أعلاه للحكم على جودة عملية إعادة التحقق من الصدق في عينة من الدراسات العربية والأجنبية، وقد استخرجت هذه المعايير من العديد من الدراسات المهمة في القياس النفسي مثل: (Messick S.، Validity، 1987) (Anastasi & Urbina, 1997) (Cronbach & Meehl, 1955) وغيرهم الذين تم ذكرهم بالتفصيل في عنصر المعايير أعلاه، وقد تم بناء أداة الدراسة والحكم على الدراسات والمقالات بناء على هذه المعايير.

10- عرض النتائج

بعد محاولة تنظيم وعرض المعايير، سنتطرق إلى عرض نتائج الدراسة حسب التساؤلات المطروحة سابقا:

- ماهي الأدلة الأكثر استخداما في إعادة التحقق من صدق الاختبارات النفسية في الدراسات؟

كما هو واضح من خلال الجدول (02) فإن أكثر أنواع الصدق استخداما في إعادة التحقق من صلاحية الاختبارات النفسية هو صدق المحكمين بـ 12 تكرارا، يليه صدق الاتساق الداخلي منفردا دون أي إجراء آخر بـ 10 تكرارات، ثم مصاحبا للصدق الظاهري بـ 9 تكرارات، ثم مصاحبا لصدق المحكمين هذه المرة بـ 5 تكرارات، لتبقى التكرارات الأخرى موزعة بصفة متقاربة على الأنواع الأخرى. ماعدا 26 دراسة لم تتحقق من الصدق، بل اكتفت بالإشارة إلى أن الاختبار صادق.

وللإشارة فالتسمية هنا كما تم ذكرها في الدراسات الأكاديمية. وهذا لا يعني أن الباحث حقا استخدم هذا النوع من الصدق، فصدق المحتوى غالبا يستخدم للدلالة على صدق المحكمين كما سنوضحه عند مناقشة النتائج.

- هل تم تبرير استخدام تلك الأدلة دون غيرها؟ وهل هي كافية؟

الجدول رقم (03): كفاية الطرق المستخدمة

الخيارات	التكرارات	الدراسات الأجنبية		الدراسات العربية	
كافية بتحفظ	8	1	4.35%	7	10.3%
لا	58	18	78.2%	40	58.8%
نعم	25	4	17.3%	21	30.8%
الكلية	91	23	100%	68	100%

الجدول رقم (04): تبرير استخدام الأدلة

الخيارات	التكرارات	النسبة المئوية
لا	85	93,4
نعم	6	6,6
الكلية	91	100,0

المصدر: من إعداد الباحثين بالاعتماد على برنامج spss

يتضح لنا من خلال الاطلاع على الجدول (03) المتعلق بكفاية الطرق المستخدمة، أنها لم تكن كذلك في 58 اختبارا، وهو ما بدا واضحا فقط من خلال عرض الأنواع الأكثر استخداما، بحيث أن صدق المحكمين وصدق الاتساق الداخلي والصدق الظاهري كان لهم النصيب الأكبر، وهو عكس ما نصت عليه المعايير. في حين أن 25 اختبارا كانت طرق التحقق من صدقه كافية، فقد تم استخدام العديد من الأدلة لإثبات صدق الاختبار الواحد. لتبقى في الأخير 8 اختبارات تم اعتبار طرق التحقق من صدقها كافية بتحفظ، وقد اعتبرت كافية لكثرة الأدلة المستخدمة في التحقق من الصدق، لكن تم التحفظ عليها لكثرة الأخطاء التي وقع فيها الباحثون. وكما هو واضح من الجدول، فإن الدراسات الأجنبية أكثر تقصيرا من العربية فيما يخص كفاية الأدلة؛ بحيث اعتمدت معظمها على الصدق الظاهري فقط مع ألفا كرونباخ كدليل على الثبات في أغلب الدراسات التي تم تحليلها. أما بالنسبة للجدول (04) فقد تجاهلت أغلبية الدراسات ذلك رغم أهمية التبرير في التفسير، وحتى في الثقة في النتائج فـ6 اختبارات فقط تم تبرير اختيارات الصدق فيها، لتبقى 85 طريقة لم يبرر اختيارها.

- هل كان كل من حجم وتمثيل العينة وقرارات الباحثين مناسبة؟

الجدول رقم (05): تمثيل العينة

الخيارات	التكرارات	النسبة المئوية
غير ذلك	41	45,1
غير ممثلة	27	29,7
لم يتم ذكرها	8	8,8
ممثلة	15	16,5
الكلية	91	100,0

الجدول رقم (06): حجم العينة

الخيارات	التكرارات	النسبة المئوية
غير ذلك	39	42,9
لا	12	13,2
لم يتم ذكرها	8	8,8
نعم	32	35,2
الكلية	91	100,0

المصدر: من إعداد الباحثين بالاعتماد على برنامج spss

من خلال الجدول أعلاه يتضح لنا أن 32 اختبارا تم دراسة صدقه في عينة مناسبة من حيث الحجم، وفي 15 اختبارا فقط كانت مناسبة من حيث التمثيل، ولم يتم ذكر تفاصيل العينة في 8 اختبارات، وبالتالي لا ندري إذا كانت مناسبة وممثلة أم لا، في حين أن 12 اختبارا تم التحقق من صدقه في عينة غير مناسبة من حيث الحجم، أي أقل من 50 فرا، و27 اختبارا لم تكن العينة فيه ممثلة ولا مناسبة. أما 41 و39 من خانة غير ذلك، فهي متعلقة بالدراسات التي لم تحتج إلى العينة في التحقق من الصدق مثل الصدق الظاهري، ورغم ذلك هنالك دراستان تم ذكر حجم العينة دون الحاجة إليه.

- أما فيما يخص قرارات الباحثين:

الجدول رقم (07): قرارات الباحثين

الخيارات	التكرارات	النسبة المئوية
رفض	1	1,1
قبول	90	98,9
الكلية	91	100,0

المصدر: من إعداد الباحثين بالاعتماد على برنامج spss

من خلال الجدول أعلاه يتضح لنا أن كل الباحثين اعتمدوا الاختبارات التي تحققوا من صدقها، باستثناء اختبار واحد فقط لم يعتمده الباحث (أنور أحمد عيسى 2013) وقرر إلغاءه، ورغم أن العديد من البنود كانت ضعيفة وأحيانا غير دالة في الاختبارات التي قبلت، إلا أن الباحثين تجاهلوا ذلك، وفي أحسن الأحوال كان الباحث يقوم بحذف تلك البنود، والإبقاء على المقياس دون إعادة التحقق من صلاحيته مرة ثانية، رغم تغير هيكل الاختبار.

11-مناقشة النتائج

لقد حاول الباحثان عرض بعض النتائج فقط، وباختصار شديد غير، أنَ هذا لا يمنع الدخول في بعض التفاصيل والأخطاء التي تم مصادفتها في أثناء تحليل تلك الدراسات، فكما سبق وأشرنا مثلا في تعليقنا على الجدول رقم (02)، فإن الكثير من الباحثين يطلقون أسماء على الصدق غير التي يستخدمونها، فمثلا استخدم صدق المحكمين للدلالة على الصدق الظاهري، ففي دراسة (عائشة الشامسي 2005) مثلا قد صرحت باستخدامها لصدق المحكمين، لكن عند التدقيق في الإجراءات، وجد الباحثان أنها اعتمدت الصدق الظاهري فقط، وهنالك العديد من الدراسات التي وقعت في الخطأ نفسه، كذلك نجد الكثير من الباحثين يستخدمون صدق المحكمين بوصفه صدق المحتوى مثل دراسة (ديانا ابزاخ 2019) و (محمد المحترم 2017)، وغير بعيد عن صدق المحكمين، لاحظ الباحثان أن العديد من الدراسات وقعت في عدة تجاوزات في أثناء استخدام هذا النوع من الصدق منها: عدم احتساب نسبة الاتفاق بين

المحكمين، والاختيار غير الموفق لهم، فنجد 16 اختيارا غير موفق مقابل 11 اختيارا موفقا (للاطلاع على النتائج التي لم تعرض في هذه الدراسة اضغط هنا)، ونقصد بغير موفق مثلا رغم ترجمة الاختبار، إلا أننا لا نجد ضمن المحكمين أي مختص لغوي، وهو ما لاحظناه في العديد من الدراسات على سبيل المثال لا الحصر دراسة (هنادي الجبالي 2019) و(مؤيد حميدي 2019) وقد اتفقت نتائج دراسة كل من (Wittkowski, Vatter, Muhinyi, Garrett, & Henderson, 2020) و (بوسالم، 2017) مع نتائج هذه الدراسة.

كذلك لاحظ الباحثان خلال عملية التحليل أن من أكثر الأخطاء شيوعا هي قبول الاختبارات رغم ضعف معاملات الارتباط، سواء في الاتساق الداخلي أو في الصدق المحكي أو غيرها من الإجراءات، التي تستلزم استخدام معامل الارتباط، فمثلا في دراسة (عبد الكريم بني صخر 2017) ودراسة (هاني الجراح 2018) لم تتجاوز معاملات الارتباط 0.40 ، بل إن بعض الدراسات لم تتجاوز معاملاتها 0.20 مثل دراسة (عبد الرحيم النوايسة ووجدان خليل الكركي 2020) ودراسة (محمد مبارك 2010) إلا أن الباحثين احتفظوا بتلك البنود الضعيفة والاختبارات، وفي أحسن الأحوال كان بعض الباحثين يقوم بحذف البنود غير الدالة فقط مثل دراسة (محمد الصبان وسماح عيد الحربي 2019)، والبعض الآخر قرر الاحتفاظ حتى ولو كانت غير دالة مثل دراسة (عائشة آل علي 2002)، في حين لم يتطرق بعض الباحثين إلى ذكر مستوى الدلالة، بل اكتفوا بذكر قيمة معامل الارتباط مثل دراسة (هادي الحويلة 2017) الذي قرر أيضا حذف البنود الضعيفة التي قدرت قيمتها بأقل من 0.25 والاحتفاظ بما هو أعلى من تلك القيمة، وهو ما اتفقت معه دراسة (Irene, Luisa, Luca, & Alvisa, 2017). وهذا ما يدفعنا للبحث في السبب وراء هذه القيم المنخفضة التي غالبا ما كان وراءها حجم العينة الذي لم يتجاوز في هذه الدراسة 25 فردا، وبالتالي فإن هذا يأخذنا إلى مسار آخر، وهو التذكير بوجوب احترام شروط معامل الارتباط بيرسون سواء في حجم العينة أو طريقة سحبها أو خطية العلاقة وغيرها من الشروط التي سبق ذكرها في المعايير أعلاه.

وفي السياق نفسه، لاحظ الباحثان أن العديد من الدراسات غيرت من هيكل الاختبار سواء بحذف البنود لتقليص الاختبار، أو تغيير خيارات الإجابة من ليكرت ثلاثي إلى خماسي، ومع وجود العديد من الأبعاد إلا أن الباحثين استخدموا الاتساق الداخلي مثل دراسة (عائشة آل علي 2002) ودراسة (عبد الرحيم النوايسة ووجدان الكركي 2020) وكان الأصح استخدام التحليل العاملي، وهو أيضا ما اتفقت معه نتائج دراسة (Polit D. , 2015) ودراسة (Irene, Luisa, Luca, & Alvisa, 2017).

وأيضا وجب التنويه إلى أمر مهم وهو اكتفاء العديد من الباحثين بذكر أن الاختبار صادق ظاهريا فقط. والأغرب من ذلك حسابه عن طريق ألفا كرونباخ مثل دراسة (الدود أحمد 2013) وهو الأمر نفسه الذي اعتمدته معظم الدراسات الأجنبية التي اكتفت بحساب

الاتساق الداخلي عن طريق ألفا كرونباخ لإثبات صلاحية المقياس بصفة عامة مثل دراسة (McAndrews, B & Eric A & Michael J Zvolensky 2021) وقد توصلت دراسة (Richardson, & Stopa, 2019) إلى النتائج نفسها واستهجنت قيام الباحثين بهذا.

وفي الأخير، فإن هذه التجاوزات وجب تداركها والتفكير فيها من قبل الباحثين؛ لأن هنالك تناقضا صريحا فيما يقومون به، بحيث يعمد الباحثون إلى إعادة التحقق من صلاحية الاختبارات النفسية، مفترضين أنها قد تكون غير صالحة سواء لهذه البيئة أو العينة أو غير ذلك، ولكن عند الوصول إلى النتائج تظهر صحة هذا الفرض، يرفض الباحثون التسليم بذلك، الأمر الذي يدفعنا إلى التساؤل عن الدافع وراء إعادة التحقق من صلاحية الاختبارات النفسية إذا لم نأخذ بنتائجها.

12-الخاتمة

إن إعادة التحقق من صدق الاختبارات النفسية ليس عملية تقنية محضة كما يتصورها الكثير من الباحثين، بل تستلزم الاختبارات قبل اعتمادها الكثير من التحليل والنقد، إضافة طبعا إلى الجانب السيكومتري منها، وهو أمر يغفله الكثير من الباحثين، وتوصلت هذه الدراسة إلى وجود العديد من التجاوزات عند إعادة التحقق من صدق الاختبارات النفسية، فمن هذا المنطلق نوصي الباحثين ومستخدمي الاختبارات بصفة عامة بالتريث قليلا قبل الثقة في نتائج الاختبارات واعتمادها، حتى بالنسبة للاختبارات العالمية، فقيمة الاختبار قد تضيع عند استخدامه أو تكييفه أو تقنينه بطريقة خاطئة، وقد حاولنا إلقاء نظرة خاطفة على المعايير الواجب مراعاتها عند التحقق من الصدق، إلا أنها لا تزال تحتاج إلى الكثير من التفاصيل التي لا يمكن تدوينها في مقال واحد وبالتالي ننصح المتخصصين في القياس بالتوسع في هذه النقطة، وإضافة ما يمكن إضافته وما غاب عن الباحثين، إضافة إلى أن صلاحية الاختبارات لا تعتمد على صدق الاختبار فقط، بل هنالك العديد من العناصر الأخرى التي وجب دراستها والتعمق فيها مثل الثبات والتحيز وغيرها. وفي الأخير لا يمكن الادعاء بأن عينة الدراسات المستخدمة لأغراض توضيحية في هذه الدراسة تمثل كل الدراسات الأكاديمية في هذا المجال.

قائمة المراجع

قائمة المراجع باللغة العربية

1 امحمد تيغزة. (2017). توجهات حديثة في تقدير صدق وثبات درجات أدوات القياس: تحليل نظري تقويمي تطبيقي. مجلة العلوم النفسية والتربوية (4(1))، 7-29.

2 دينيس هوييت، و دونكان كرامر. (2016). مقدمة لطرائق البحث في علم النفس (ط1). (صلاح الدين محمود علام، المترجمون) عمّان، الأردن: دار الفكر.

3 صفوت فرج. (2017). القياس النفسي (ط6). مصر: مكتبة الأنجلو المصرية.

4 صلاح الدين محمود علام. (2000). القياس والتقويم التربوي والنفسي ،أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.

5 عبد الرحمان بن صافي، و علي طاجين. (2013). تقييم طرق تقدير صدق أدوات القياس المستخدمة في البحوث النفسية والتربوية بجامعة الجزائر 2. سلوك مخبر تحليل المعطيات الكمية والكيفية (3).

6 عبد العزيز بوسالم. (2017). الصدق الثقافي للاختبارات النفسية المطبقة في الجزائر ومشكلة التكييف من أجل الصلاحية. مجلة العلوم النفسية والتربوية (4(2)).

7 مراد نعموني، و محمد براق. (2015). تطبيق الاختبارات النفسية في عملية الاختيار والتوظيف في الجزائر. المدرسة العليا للتجارة (27)، 64-83.

قائمة المراجع باللغة الأجنبية

8 Aaronson, N., Alonso, J., Burnam, A., Lohr, K., Patrick, D., Perrin, E., et al. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res , 11 (3), 193-205.

9 AERA, APA, & NCME. (1999). Standards For Education and Psychological Testing. Wachington: AERA.

10 Anastasi, A. (1988). Psychological testing. New York: Macmillan.

11 Anastasi, A., & Urbina, S. (1997). Psychological Testing. Upper Saddle River: Prentice Hall.

12 Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. J Apple Psychol , 78 (1993), 98-104.

13 Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Rinehart and Winston.

14 Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. Psychometrika , 16 (1951), 297-334.

15 Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. Psychological Bulletin , 52 (4), 281-302.

16 Fayers, P., & Machin, D. (2007). . Quality of life. Assessment, analysis, and interpretation. The assessment, analysis, and interpretation of patient-reported outcomes. chichester: John Wiley & Sons.

17 Gay, L., Mills, G., & Airasian, P. (2012). Educational research: Competencies for analysis and applications. Boston: Pearson.

18 Guyatt, G., Feeny, D., & Patrick, D. (1993). Measuring health related quality of life. Annals of Internal Medicine , 118 (8), 622-629.

19 Hambleton, R., & Kanjee, A. (1993, Apr). Enhancing the Validity of Cross-Cultural Studies: Improvementsin Instrument Translation Methods. National Center for Education Statistics .
20 Irene, M., Luisa, S., Luca, G., & Alvisa, P. (2017). Instruments evaluating the quality of the clinical learning environment in nursing education: A systematic review of psychometric properties. Inernational Journal of Nursing Studies , 68, 60-72.
21 Keszei, A., Novak, M., & Streiner, D. (2010). Introduction to health measurement scales. J Psychosom Res , 68 (4), 319-323.
22 Kimberlin, C., & Winterstein, A. (2008). Validity and reliability of measurement instruments used in research. Am Journal health syst pharm , 65 (23), 76-84.
23 Kirschner, B., & Guyatt, G. (1985). A methodological framework for assessing health indices. J Chronic Dis , 38, 27-36.
24 Lehmnn, I., & Mehrens, W. (1973). Measurement and Evaluation In Education and Psychology. Rinehart and winston.
25 McAndrews, Z., Richardson, J., & Stopa, L. (2019). Psychometric properties of acceptance measures: A systematic review. Journal of Contextual Behavioral Science , 12, 261-277.
26 Messick, S. (1987). Validity. New Jeresy: Educational Testing Service Princeton.
27 Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol , 50 (1995), 741-749.
28 Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., et al. (2010). 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), 737-745.
29 Outco, S. A. (2002). Assessing health status and quality-of-life instruments: Attributes and review. Quality of Life Research , 11, 193-205.
30 Polit, D. (2015). Assessing measurement in health: Beyond reliability and validity,International. Journal of Nursing Studies , 52, 1746-1753.
31 Polit, D., & Beck, C. (2006). The content validity index: are you know what's being reported? Critique and recommendations. Res Nurs Health , 29 (5), 489-497.
32 Reynolds, C., & Livignston, R. (2012). Mastering modern psychological testing: Theory and methods. Pearson Education.
33 Roach, K. (2006). Measurement of health outcomes: reliability, validity and responsiveness. Journal Prosthet Orthot , 18 (1S), 8-12.
34 Wilson, B., & Cleary, P. (1995). Linking clinical variables with health-related quality of life. JAMA , 273, 59-65.
35 Wittkowski, A., Vatter, S., Muhinyi, A., Garrett, C., & Henderson, M. (2020). 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, 101-906.