مجلة علم النفس التطبيقي
قسم علم النفس - كليـة الآداب، جامعة المنوفية

هل المقاييس النفسية كـ"فرش أسنان

ترجمة
أ.د./ عبد المنعم شحاته
قسم النفس، كلية الآداب جامعة المنوفية

المجلد (٢) العدد (٢) يناير ٢٠٢٤ م

هل المقاييس النفسية كـ"فرش أسنان"[1]

Psychological measures aren't toothbrushes.

By
Malte Elson, Ian Hussey, Taym Alsalti & Ruben C. Arslan

ترجمة

أ.د./ عبد المنعم شحاته

قسم النفس، كلية الآداب جامعة المنوفية

ملخص

معظم المقاييس النفسية تستخدم مرة أو اثنتين، وهذا مؤشر تكاثرها وتباينها بدرجة تهدد مصداقية البحث العلمي، ولتجنب هذا يجب أن تؤكد كتيبات توحيد اجراءات القياس أن المقاييس النفسية مقننة بما يساعد على تكرار استخدامها.

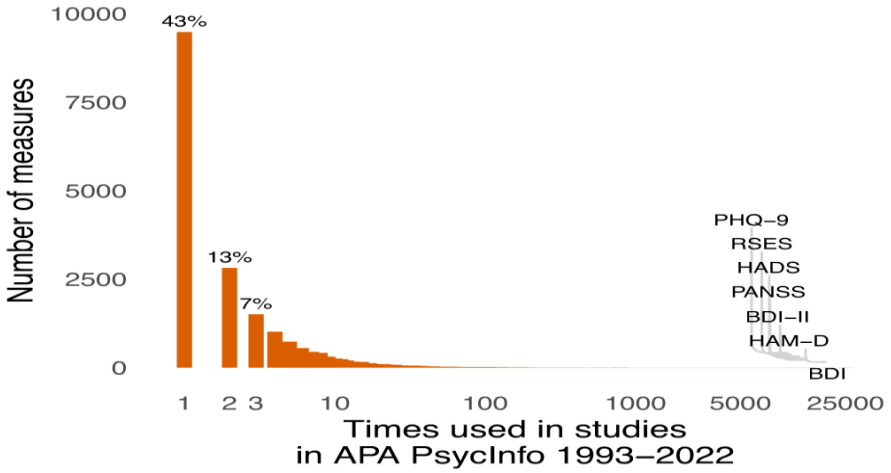
مقدمة

تعاني المفاهيم والمقاييس النفسية من مشكلة فرشاة الأسنان (١): بمعنى أن الأخصائي النفسي الذي يحترم ذاته لا يريد استخدام مقياس أعده غيره، مما سبب تكاثر المقاييس، فوفقا لبيانات قاعدة معلومات البحوث النفسية من ١٩٩٣ حتى ٢٠٢٢ فإن ٣٤٪ من مقاييسها استخدم مرة واحدة، و١٣٪ مرتين، و٧٪ ثلاث مرات كما يوضح شكل (١)، وأن المقاييس واسعة الاستخدام هي التي في المجال العيادي (الاكلينيكي)، حيث تجاوز مرات استخدام بعضها عشرة ألاف مرة.

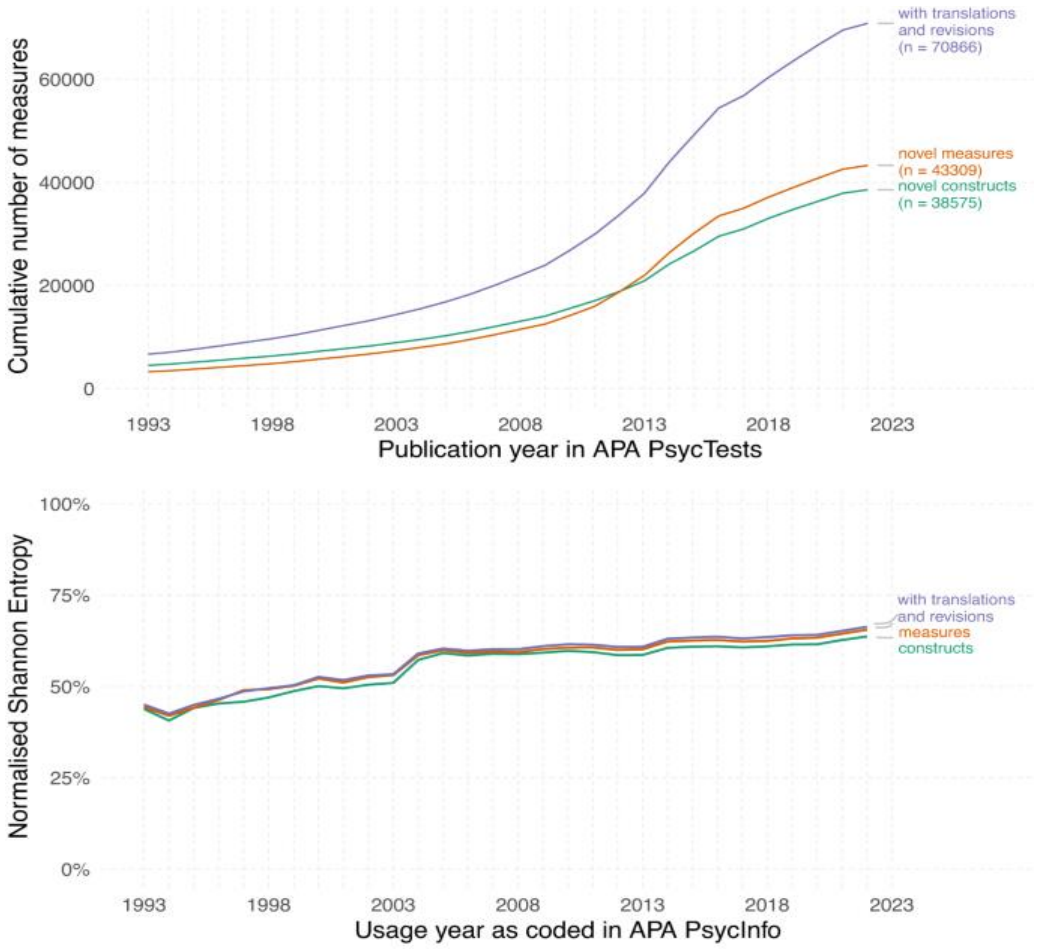
لا يميل الباحثون للاتفاق على معايير المقاييس بمرور الوقت، بل يتزايد التفتت حال كون المفاهيم جديدة (خط أخضر) والمقاييس جديدة (خط برتقالي) ومقاييس تمت مراجعتها (خط أرجواني)(١٠) كما يوضح شكل (٢). يحدث التكاثر ببساطة لأن الباحثين يروجون لعلاماتهم

[1] communications psychology (2023) 1:25 | https://doi.org/10.1038/s44271-023-00026-9 | www.nature.com/commspsychol

التجارية الخاصة لرفع سعرها، لذا من الصعب اكتشاف المقاييس القابلة لإعادة الاستخدام في الكتابات العلمية الأكاديمية بمستوييها الواسع والضيق، وقد يبدو للوهلة الأولى تكاثر القياس غير ملفت للانتباه أو حتى مفيدا، فعلى أية حال من الذي يعارض دراسة الظاهرة نفسها من خلال مناحي منهجية متعددة؟، وهنا، نرى أن التكاثر يشكل في الواقع عائقًا خطيرًا لتراكمية العلم.



شكل (١) معدل استخدام المقاييس النفسية في الثلاثين عاما الماضية



شكل (٢) تكاثر المقاييس وفقا لتوزيعها

كثرة الأسماء

تقوم بعض المقاييس في الواقع بقياس أشياء مختلفة، لكنها تشترك في تسميات مماثلة أو حتى الأسماء نفسها (على سبيل المثال يوجد بقاعدة اختبارات جمعية علم النفس الأمريكية ما لا يقل عن ١٩ اختبار مختلف اسمه "استخبار السلوك المخطط له"، ١٥ منها مقاييس للرضا عن العمل، و١١ تقيس كفاءة الذات). وهناك مقاييس أخرى تحدد نفس الشيء لكن تحت عنوان مختلف. تُعرف هذه المغالطات باسم "تغمة" Jingle و"خشخشة" Jangle، وهي

تهديدات شائعة موثقة جيدًا لصدق الأبحاث النفسية وقابليتها للاستعادة، على سبيل المثال تنطوي بحوث الانفعال (٢) على مغالطة اسمية، فلا يحدث أن اسم المقياس يخبرك عن محتوياته أو ما يقيسه (٣).

مرونة غير معلنة

حتى عندما يصرح المؤلفون باستخدام المقياس ذاته للمفهوم نفسه، فإن كل شيء ليس على ما يرام بعد، نتيجة مرونة القياس التي تم أو لم يتم الكشف عنها، فالشائع أن لتغييرات في المقياس عواقب نفسية معروفة أو غير معروفة (٤). من التغييرات التي تحدث: حذف البنود وإضافتها وتعديلها في مقاييس التقرير الذاتي، وتجميع الدرجات الإجمالية بطرق مختلفة في المهام المعملية، أو المحفزات المختلفة وفترات التجارب، كل هذا بينما لا يشير الباحثون إلى نفس المفهوم فحسب، بل في الواقع إلى الأداة نفسها بالاسم (٥). حتى عندما يتم الكشف عن جميع القرارات، سوف تكشف منهجية عرض الدراسات أن العديد منها استخدمت خوارزميات تجميع فريدة لتصحيح البنود على سبيل المثال، التي غالبا يكون لها عواقب نفسية غير معروفة.

التعميم عبر العينات

معايير الاختبار الراهنة ضرورية لمقارنات يعتد بها بين درجات الأفراد على الاختبار وبين قيم عينة ممثلة للمجتمع (٦)، فحتى المقاييس واسعة الاستخدام لم يتم معايرتها أبدًا في المجتمع الذي يتم استخدامها فيه، أو أن معايير الاختبار المتاحة قديمة جدًا. وبدون المعايير الحالية، لا يستطيع مستخدمو المقاييس الحكم على ما إذا كانت إجراءات اختيار العينات الخاصة بهم قد تؤدي لتحيز، مما يجعل من الصعب الحكم على قابلية التعميم، من المفترض أن تكون أحجام التأثير "الموحدة"، مثل الارتباطات، متحيزة عندما يكون كل ما هو متاح للتقنين هو الانحراف المعياري المحدود داخل العينة (٧). قد لا يتم تجميع أحجام التأثير هذه مباشرة عبر العينات دون تصحيحات معقدة، لكنه مع ذلك يتم بشكل روتيني في التحليلات اللاحقة meta–analyses.

أدبيات البحث التي تبدو متماسكة وصارمة لقارئ طارئ أو حتى قارئ متمرس قد يكون في الواقع أي شيء آخر غير ذلك، فالتوليف المعقول للأدلة والتحليلات اللاحقة سوف تصحح الفروق في اختيار العينة والتباين والثبات وأي مصدر أخر لعدم تجانس القياس، ومع ذلك، فإن القيام بذلك بشكل صحيح غالبًا ما يكون صعبًا لأن المعلومات المطلوبة مفقودة (٧)، وأساليب تصحيح التحيز تعتمد على افتراضات غير قابلة للاختبار في كثير من الأحيان.

هنا تصبح مهددات القياس هي:

أ) الافتقار لعايير امبريقية أو اجرائية قوية.

ب) غياب الشفافية عند كتابة التقارير.

ج) عدم وجود مرجعيات مشتركة (أي معايير الاختبار) في القياس.

وهي مهددات خطيرة لتراكم الأدلة ذات المعنى المطلوب لتركيب البحوث، على سبيل المثال، لتعزيز معامل ثبات مقياس ما في دراسة أولية يقوم الباحثون غالبا بحذف ما يعتبرونه بنود "ضعيفة الأداء" (٨)، لكن هذا الاسلوب في حد ذاته يفتقد الثبات وينتهي لتوصيات متناقضة بشأن ما تم حذفه من بنود (٩). وبدون الحصول أيضًا على معيار خارج العينة لهذا القياس المعدل، أو قابلية للاعادة توضح أن حذف البنود المختارة يحسّن القياس باستمرار، فإن هذا لا يحسن الاستخدام الحالي أو المستقبلي للقياس، وبوجه خاص عندما يكون حذف هذا البند ينقصه الوصف المفصل كما يحدث في كثير من الأحيان (٨)، فبدون هذا الوصف لا يمكن تحسين الاجراء. وبالتالي، عندما تكون مرونة القياس موجودة، قد لا تتم مقارنة الدراسات ذات المقاييس التي تستخدم الإسم نفسه بشكل مفيد، سواء تمت مقارنتها بعضها مع بعض أو في اطار تجميع أكبر للبحوث. علاوة على ذلك، لا يمكن تطبيق أدلة صدق جديدة لمقياس موجود إلا بشكل ضيق للغاية، أو مع قدر كبير من عدم اليقين عندما يكون من غير الواضح أي الدراسات تتطابق مع البروتوكول الذي تم التحقق من صدقه. وحتى القرارات الأكثر غرابة في الدراسة يمكن أن تظل دون أن يلاحظها أحد، وتظل تطبيقاتها غير معروفة، طالما لم يكن ممكنا إجراء مقارنة ذات معنى مع معيار ثبت صدقه، فلا يستطيع أحد أن يستنتج أن نتائج

الدراسة كنتاج رحلة صيد سمك، بمعنى أنها تؤدي إلى مؤلفات تبدو متجانسة ولكنها في الواقع نتاج عملية تراكم الصيد.

ارشادات تقنين البحوث في العلوم السلوكية

يجب أن يكون علم النفس جادًا في تقنين مقاييسه، وهذه الجدية لم تحدث بعد، والسؤال المهم هو من يفرضها لرفع القيمة العلمية للقياس وكبح النزعة التجارية لمعدي المقاييس؟، الاجابة أنه يوجد كثير من أصحاب المصلحة القادرون على تشكيل سلوك الباحثين من خلال سياسات ذات معنى، على رأسها سياسة النشر بدوريات علمية تشجع روح العلم، وتعزز مكانته من خلال رفع جودته.

نقترح في الجدول التالي ارشادات تقنين البحث في العلوم السلوكية Standardization of Behavior Research (SOBER)، والتي تتعلق بوجه خاص بمرونة القياس ومعاييره.

جدول (١) سياسات النشر بالدوريات العلمية النفسية

السياسة	يمتثل لها الباحث	ينفذها المحرر والمحكم
يبين عدم التكرار	عند اعداد مقياس جديد، وفر دليل أنه ليس تكرارا لمقياس أخر أو مزايدة على صدقه بعينة منفصلة، بواسطة تحليلات ارتباطات متغير كامن بمجموعة مقاييس كبيرة منتقاة.	التحقق من الصلاحية والمعايير في البيانات المستقلة، أو تفسير لماذا هذا ليس ضروريا. رفض الدراسات التي تستخدم مقاييس جديدة أو محدَّثة دون تقديم دليل صدق.
توضيح الالتزام بالبروتوكول	عند اتباع إجراء قياس نُشر في مكان آخر، اثبت التزاما بالبروتوكولات ذات الصلة من خلال مشاركة مواد الدراسة.	التحقق من ادعاءات الالتزام بالبروتوكول من خلال مقارنة المواد والاقتباسات، وتخصص المجلات موارد لهذه المهمة.
تبرير التعديلات	إثبات أن أي انحراف أو تعديل على مقياس موجود إما أنه ذو معنى (مثل لمعالجة عدم ثبات القياس بين العينات) أو غير ذي صلة،	تثبيط تعديل المقاييس الموجودة من خلال حذف البنود ضعيفة الأداء مثلا، إذا لم يستطع الباحث اظهار مدى تأثيرها

السياسة	يمتثل لها الباحث	ينفذها المحرر والمحكم
	بتقديم دليل صدقه في عينة منفصلة، وتوثيق متى حدث، وإذا أمكن تقييم قوة الاستنتاجات.	على القياس خارج العينة، تحفز المجلات البحث المنهجي تقديم أدلة صدق المقاييس شائعة الاستخدام بدلاً من الإجابة عن أسئلة البحث الموضوعية.
التسجيل المسبق	تحديد مبكر لتفاصيل الاجراءات والتصحيح، وتسجيل كل انحراف قد يضعف الاستنتاجات، وينبغي أن تسجل مسبقا القرارات التحليلية والإحصائية برمز بدلاً من وصف سردي.	مطالبة المؤلفين بتقديم مبررات اتخاذ قرار بقياس معين، ومقارنتها بما سبق تسجيله أو استخدامه من تعديلات له.
إعداد التقارير الشاملة	تقرير كافة البنود، المنبهات، التعليمات، المعالم الإجرائية أو غيرها من خصائص القياس المستخدمة في الدراسة أو طورت أثناء اجرائها.	التحقق من وجود بيانات ومواد شاملة تتجاوز ما ورد في المخطوطة.
تسهيل عملية توليف البحوث	الإبلاغ عن الانحرافات المعيارية والوسائل (بغض النظر عما إذا تمت مشاركة البيانات)، ولا تُبلغ بشكل حصري عن أحجام التأثير النسبية إلى التباين داخل العينة (ما يسمى بأحجام التأثير القياسية مثل ارتباطات (R2 ,Cohen's d	الإصرار على إحصاءات وصفية كاملة لجعل التحليل اللاحق الدقيق ممكنًا. أقترح توحيد أحجام التأثير باختبار المعايير بدلا من التباين داخل العينة.

بيانات هذا الجدول متاحة بالرابط https://doi.org/10.1038/s44271-023-00026-9.

المضي قدما

ندعو جمهور الباحثين في مجال علم النفس إلى:-

١) الاعتراف أن مرونة القياس مشكلة خطيرة تهدد مصداقية العلم مثلها مثل مسائل القرصنة المعترف بها على نطاق واسع.

٢) وضع واستخدام معايير يتم تحديثها بشكل مستمر، والتحقق من صلاحيتها للقياس بواسطة إرشادات إعداد تقارير قياسية يتم الحفاظ عليها وتحديثها بأدلة جديدة.

ولتحقيق هذه الغاية، فإننا نتحيل وجود مستودع[١] واحد لبروتوكولات القياس، تكون متاحة لجميع الباحثين، وهو ما يسد الفجوة التي تركها الناشرون الأكاديميون بعد تراجع مقاييس الورقة والقلم، مع تطبيقات أدوات الدراسة المستخدمة على نطاق واسع، ومعايير الاختبار، وقواعد التصحيح القياسية، والبيانات الوصفية القابلة للقراءة آليًا. من شأن هذا المستودع أن يسهل اكتشاف المقاييس، والتدقيق في قرارات الاعداد، والوعي بالعلم meta-science من خلال تقييم منهجي لممارسات البحث. الأمر الذي سيسمح بنماذج لغوية كبيرة لتقييم تداخل الدلالات وثباتها بشكل مستقل، وربما حتى تقدير الصدق التمييزي والتقاربي لمقاييس طورت حديثًا مقابل تلك الموجودة بالفعل في مستودع المقاييس.

من الضروري التدقيق في تفاصيل إجراءات بناء المقاييس السابقة لمعرفة كيفية تفسير نتائجها وزيادة امكانية استخدامها مستقبلا، تسمح شفافية تفاصيل مقاييسنا الدقيقة للأخرين أن يستخدمونها بأمانة، كما تسمح باختبار كفاءتها عبر الدراسات، لم يتم بعد توظيف جوانب الشفافية هذه وفوائدها العلمية في مجالنا.

إذا كنا نريد بناء قاعدة أدلة تراكمية في علم النفس نحتاج إلى تقنين مقاييسنا وبروتوكولاتها والتوقف عن اعادة المزج واعادة التدوير والبدء في اعادة استخدام المقاييس حتى لا تصبح كفرشاة أسنان.

[١] كما هو الحال عند نشر الكتب، حيث الايداع بدار الكتب المصرية (مكتبة "الكونجرس" كذلك)، ومثل بنك المعامل (للكليات العملية) الذي أسسه المجلس الأعلى للجامعات باقتراح من أ.د./ طريف شوقي حين كان نائبا لرئيس جامعة بني سويف للدراسات العليا والبحوث (المترجم).

المراجع

1. Mischel W. The Toothbrush Problem. APS Obs. 21. Accessed May 22, 2023.
https://www.psychologicalscience.org/observer/the-toothbrush-problem.
(2008).
2. Weidman, A. C., Steckler, C. M. & Tracy, J. L. The jingle and jangle of emotion
assessment: imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. Emot. Wash DC 17, 267–295 (2017).
3. Lilienfeld, S. O. & Strother, A. N. Psychological measurement and the replication crisis: four sacred cows. Can. Psychol. Psychol. Can. 61, 281–288 (2020).
4. John, L. K., Loewenstein, G. & Prelec, D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol. Sci. 23, 524–532 (2012).
5. Elson, M. Examining psychological science through systematic meta-method
analysis: a call for research. Adv. Methods Pract. Psychol. Sci. 2, 350–363 (2019).
6. Crocker L. M., Algina J. Introduction to Classical and Modern Test Theory. Cengage Learning (2008).
7. Wiernik, B. M. & Dahlke, J. A. Obtaining unbiased results in meta-analysis:
the importance of correcting for statistical artifacts. Adv. Methods Pract. Psychol. Sci. 3, 94–123 (2020).
8. Cortina, J. M. et al. From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. J. Appl. Psychol. 105, 1351–1381 (2020).
9. Kopalle, P. K. & Lehmann, D. R. Alpha inflation? The impact of eliminating
scale items on Cronbach's Alpha. Organ Behav. Hum. Decis. Process 70, 189–197 (1997).
10. Wilcox, A. R. Indices of qualitative variation and political measurement. West
Polit. Q 26, 325–343 (1973).