مجلة العلوم الاجتماعية والإنسانية، جامعة باتنة1 المجلد(22) ـ العدد (01) (جوان، 2021م) -ص 243 – 258
ISSN 1111-5149 /E.ISSN 2430-2588

بعض الأخطاء الناجمة عن جمع أدلة الصدق والثبات في البحوث التربوية والنفسية

Some Errors Arising from Gathering Evidence of Validity and Reliability in Educational and Psychological Research

فاروق طباع [1] (جامعة سطيف2)، ftebbaa05@yahoo.fr *
يوسف خنيش [2]، (جامعة سطيف2)، khenniche_y@yahoo.fr

تاريخ الاستلام	تاريخ القبول
2020-10-03	2021-01-11

ملخّص

يقع الباحثون أحياناً في المجالات التربوية والنفسية في الأخطاء أثناء جمع الأدلة عن الصدق والثبات في مرحلة إعداد أدوات البحث، والتي يمكن أن تؤثّر على موثوقية النتائج لأنّ جودة النتائج تعتمد بالأساس على جودة أدوات البحث. فعلى الرّغم من التطوّرات التي حدثت في القياس إلاّ أنّ بعض ممارسات الباحثين لا تزال تقليدية في تقدير الصدق والثبات، ويلجؤون إلى استخدام طرق غير ملائمة، ويقدّمون أدلة غير كافية، وينتهكون افتراضات بعض الطرق، ويجمعون أدلة غير كافية. وفي هذا الإطار يهدف المقال إلى تقديم نظرة عن الأخطاء الناجمة عن جمع أدلة الصدق والثبات في البحوث التربوية والنفسية، وذلك باستعراض التطوّرات التي حدثت في أدبيات القياس الحديثة التي لا تعكس الممارسات السّائدة لدى الباحثين في هذا الشّأن.

كلمات مفتاحية:

أخطاء الممارسات، أدلة الصدق، أدلة الثبات، البحوث.

Abstract

In educational and psychological settings sometimes researchers commit errors while collecting evidence on validity and reliability in elaborating research tools. These errors can affect the dependability of the results as the quality of the results depends mainly on the quality of the research tools. Despite developments in measurement, some researchers' practices are still traditional in estimating validity and reliability, resorting to inappropriate methods, providing insufficient evidence, violating assumptions of some methods, and gathering insufficient evidence. In this context, this article aims to provide a view of these errors resulting from gathering evidence of validity and reliability in educational and psychological research, and reviews the developments that have occurred in modern measurement literature that do not reflect the prevailing practices among researchers.

Keywords:

Errors practices; Validity evidence; Reliability evidence; Research.

* المؤلف المرسل

مقدّمة

تُستخدم في البحوث التربوية والنفسية مجموعة من الأدوات لجمع البيانات (استبيانات، مقاييس، اختبارات، شبكات ملاحظة...وغيرها)، التي على أساسها يتم تقديم تقرير عن النتائج، وتخضع جودة البحوث إلى جودة الأدوات المستخدمة، وهذا ما يسمح في النّهاية بالحصول على نتائج قابلة للتعميم على سياقات وعيّنات وفترات أخرى، ومن أجل تحقيق جودة البحث يجب أن يتم جمع بيانات باستخدام أدوات تتمتع بصدق وثبات كافيين، وذلك على اعتبار أن جودة النتائج تخضع إلى الشروط السيكومترية للأدوات.

ويتم التحقّق من جودة أدوات البحث من خلال جمع أدلة ملائمة وكافية حول صدقها (أي صلاحيتها) وثباتها (أي موثوقيتها) باستخدام أساليب متعدّدة وردت في الكثير من أدبيات القياس، ويعدّ الصدق من أكثر الاعتبارات أهمية، وخاصية جوهرية، ومعيارا أساسيا في تقييم جودة أدوات البحوث النفسية والتربوية. لذلك يُعدّ الصدق أهم اعتبار في تطوير وتقييم الاختبارات (أو أي أداة أخرى) (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014, p. 11). ويتم قياس الصدق من خلال جمع أدلة نظرية وأخرى عملية كافية تسمح بتدعيم تفسيرات درجات أدوات البحث للاستخدامات المستهدفة لها.

ويُعدّ الثبات أيضاً من الاعتبارات المهمة التي تُبنى عليه أدوات البحث، وقد حظي باهتمام أكبر من طرف المؤلفين منذ فترة زمنية بعيدة ؛ نظراً لأهميته في الحصول على نتائج ذات موثوقية عالية من أدوات القياس. فالثبات يهتم باتّساق القياسات عبر مختلف الظروف التي تُطبق وتُصحّح وفقها أدوات القياس(Meyer, 2010).وتُستخدم في تقدير ثبات أدوات البحث أساليب متعدّدة، وتتضمن هذه الأساليب افتراضات وإجراءات معينة حسب كل طريقة أو مجموعة من الطرق، فمنها ما يندرج ضمن طرق النظرية الكلاسيكية، ومنها ما يندرج ضمن نظرية إمكانية التعميم، ومنها ما يدرج ضمن نظرية الاستجابة للمفردة(Crocker & Algina, 2006).

وفي إطار إعداد البحوث يتم بناء أدوات القياس في المجالات التربوية والنفسية باتّباع خطوات أولية ضرورية تتمحور حول؛ 1)- تحديد الأهداف الأولية التي تُستخدم فيها درجات الأداة، 2)- تحديد السلوكات التي تُمثّل البناء أو نطاقه بتحويله إلى مجموعة السلوكات أو العبارات باستخدام تحليل المحتوى، ومراجعة الأبحاث، والأحداث العرضية الحرجة، والملاحظات المباشرة، وأحكام الخبراء، والأهداف التربوية، 3)- إعداد مجموعة مواصفات للأداة تصف بدقة نسبة العبارات الممثّلة لكل نوع من السلوكات المحددة في الخطوة الثّانية، 4)- بناء ملف أولي للعبارات باختيار صيغة مناسبة، ومدى ملاءمتها لفئة المفحوصين، واختيار وتدريب مُعدّي العبارات، وكتابة العبارات، ومراقبة تقدّم مُعدّي العبارات ونوعيتها، 5)- مراجعة العبارات وتعديلها عند الضرورة (الاستعانة بالمختصين، ومراجعة العبارات من حيث الدقة والصياغة والقواعد والغموض والملاءمة، 6)- التجريب الأولي للعبارات على عينة من المفحوصين ومراجعتها إذا لزم الأمر قبل

طبعها نهائيا، وتحليلها إحصائياً من حيث الصّعوبة والتباين والتمييز، 7)- تطبيق العبارات على عينة كبيرة ممثّلة لمجتمع المفحوصين، 8)- تحديد الخصائص الإحصائية للدرجات، وعند الضرورة حذف العبارات التي لا تتفق مع المعايير المُحدّدة مسبقاً، 9)- تصميم وإجراء دراسات الصدق والثبات للصيغة النهائية للأداة للتحقّق من الخصائص السيكومترية الضرورية، 10)- إعداد دليل يفيد في التطبيقات والتصحيح وتفسير الدرجات (Crocker & Algina, 2006).

وتعدّ المرحلة التاسعة ذات أهمية كبيرة في تحديد مدى ملاءمة أداة البحث لجمع بيانات متّسقة وملائمة للاستخدام في تحقيق الأغراض التي أُعدّت من أجلها، فدراسة الصدق والثبات مرحلة حرجة يتخذ فيها الباحث قرارات حول صلاحية الأداة للاستخدام في جمع البيانات. ولكن في بعض الحالات لا تزال الممارسات البحثية في هذا المجال تعتمد على طرق تقليدية، ويقع الباحثون في مشكلات اختيار أساليب قياس غير ملائمة نظراً لعدم استيفاء شروط وافتراضات استخدامها، والاكتفاء ببعض الأدلة دون الأخرى (تيغزة، 2017).

كما لا يولي الباحثون الاهتمام بجمع أدلة كافية عن صلاحية أدواتهم، كما أن تكوينهم - خاصة في البيئة المحلية - لا يرقى إلى المستوى المطلوب في مجال القياس النفسي والتربوي نظراً لعدم اطلاعهم على التطورات الحاصلة في هذا المجال، مما جعل الأساليب المستخدمة - أغلبها تقليدية- غير قادرة كفايةً على جمع أدلة رصينة لتأييد صدق وثبات القياسات النفسية والتربوية.

وعلى هذا الأساس يهدف هذا المقال إلى تقديم بعض الأخطاء التي يقع فيها الباحثون في جمع أدلة عن صدق وثبات أدوات البحث التربوي والنفسي، وقبل ذلك تم تحديد مفهوم الصدق والثبات وإبراز بعض التطوّرات التي حصلت في أدبيات القياس والأساليب المستخدمة في جمع أدلة عن صدق وثبات أدوات القياس، وربطها بالممارسات الخاطئة للباحثين التي لا تعكس التطوّرات الحاصلة في هذا الشأن.

1. مفهوم الصدق

يعدّ الصدق خاصية جوهرية ومعيارا أساسيا في تقييم جودة أدوات البحوث التربوية والنفسية، ويشير ببساطة إلى قدرة أداة على قياس ما أُعدّت لقياسه. وقد تطوّر خلال العقود القليلة الماضية وأصبح لا يُنظر إليه بمدى ملاءمة الأداة للغرض الذي وُضعت من أجله، وإنّما يشير إلى حكم تقييمي شامل لمدى قدرة الأدلة الإمبريقية والأسس المنطقية النظرية على تدعيم كفاية وملاءمة التفسيرات والعمليات المعتمدة على درجات الاختبار أو أدوات تقييم أخرى (Messick, 1989).

كما تشير "وثيقة معايير العملية الاختبارية النفسية والتربوية Standards for" Educatioanal and Psychological Testing إلى أن الصدق هو مدى تدعيم الأدلة والنظرية تفسيرات درجات الاختبار التي تتطلّبها الاستخدامات المقترحة للاختبارات (AERA, APA, & NCME, 2014).حيث يسمح الصدق بالإجابة على سؤالين أساسيين، يتعلق الأول بـ "هل تعكس

درجات الاختبار التي تُزوّد بالمعلومات المطلوبة للاستخدامات المقترحة؟" و "هل الدرجات مفيدة في اتخاذ قرارات صائبة؟

وتنطوي حسب "وثيقة معايير العملية الاختبارية النفسية والتربوية" عملية التحقّق من الصدق أو تأييد الصدق validation على جمع الأدلة ذات الصلة بالسمة المُقاسة لتوفير أساس علمي سليم لتفسير الدرجات المقترحة، وتفسيرات درجات الاختبار للاستخدامات المقترحة هي التي يتم تقييمها، وليس الاختبار نفسه، فعندما يتم تفسير درجات الاختبار بأكثر من طريقة يجب التحقّق من صدق كل تفسير مقصود، ويجب أن تشير التأكيدات حول الصدق إلى تفسيرات معينة للاستخدامات المحدّدة للأداة، ومن الخطأ استخدام العبارة "صدق الاختبار" بشكل قاطع.

يختلف مصطلح الصدق عن مصطلح تأييد الصدق أو التحقّق من الصدق، فتأييد الصدق يتضمن العمليات أو الطرق المستخدمة لتدعيم الصدق وتفسير اختلاف الدرجات، ويجب أن تظهر أيضاً القيم الشّخصية والاجتماعية المرجوة أو غير المرجوة من تلك العملية (Hubley & Zumbo, 2013, p. 11). وبالتالي تأييد الصدق عملية مستمرة، والتي أشار إليها Messick (1989) بأنها "قضية لبناء دليل منطقي بالأساس لمتابعة الاستخدام الفعلي للاختبار، والبحث الفعلي عن زيادة فهم دلالة درجات الاختبار(p. 13) ".

وفي هذا الإطار يشير الصدق حسب Crocker & Algina (2006) إلى الاستدلالات المستمدة من درجات الاختبار لتحقيق غرض معين ضمن مجموعة شروط وُضعت مسبقاً، ويشير تأييد الصدق إلى العملية التي يتم من خلالها جمع الأدلة التجريبية لتدعيم استخدام درجات الاختبار للأغراض الموضوعة.

يبدو أن التعريفات السابقة للصدق تختلف في عدة جوانب عن التعريفات التقليدية التي ترى بأن الصدق هو قدرة الأداة على قياس ما صُمّمت لقياسه، وإذا كان الاختبار (أو أي أداة أخرى) لا يقيس ما عُني بقياسه فإنّ استخدامه يُصبح مضلّلاً(Urbina, 2004).ومن الجوانب التي أضافتها نظرية الصدق الحديثة إلى مفهوم الصدق - ظهرت ضمنيا- في ثلاثة جوانب مترابطة:

- صدق درجات الأداة تنتج من كل الأدلة المُجمّعة لتدعيم تفسيرها واستخداماتها، وبذلك يهتم الصدق دائما بالدرجة أو بالأحرى بالمقدار، وتأييد الصدق يشير إلى العملية التي تتم بواسطتها جمع أدلة الصدق، وتبدأ بتوضيح مُطوّر الاختبار للإطار المفاهيمي والأساس المنطقي للأداة.

- الصدق مفهوم نظري ودليل إمبريقي لتفسير درجات الأداة، فصدق الاستدلالات المحصلة على أساس درجات الاختبار لمختلف الأغراض المطلوبة يمكن تأكيدها أو دحضها، فقد أعلنت وثيقة معايير العملية الاختبارية التربوية والنفسية بأنّ تأييد الصدق مسؤولية مشتركة بين مُطوّر الاختبار الذي يقدم أدلة وأسسا منطقية للاستخدام المستهدف من الاختبار، ومُستخدم الاختبار الذي يُقيّم الأدلة الموجودة ضمن السياق الذي استخدم فيه الاختبار(AERA, APA, & NCME, 2014, p. 11).

-نتيجة للأغراض المتعددة التي يمكن أن تُطبّق فيها درجات أدوات البحث، فإنّ الأسس الاستدلالية لتفسير الدرجات يمكن أن تُستمد من طرق متنوعة، ويمكن الحصول على مساهمات عن أدلة صدق الدرجات من أي بحث منظم يُدعّم أو يضيف إلى دلالتها بغض النظر عن الذي يقوم أو متى يقوم به.

وقد توسّع مفهوم الصدق وفق النظرية الحديثة التي طوّرها الباحثون أمثال ميسيك Messick، وكاين Kane، وكرونباخ Cronbach، وخاصة أعمال الباحث ميسيك (1995 ; Messick,1989)الذي قدم إطاراً فكرياً لتأييد صدق التكوين الفرضي لاستخدام وتفسير درجات أدوات القياس، وقد أصبح الصدق مفهوماً موحداً ومتكاملاً يتضمن ستة جوانب أساسية، وهي أدلة المحتوى، وأدلة العمليات والتأصيل النظري، والأدلة البنائية، وأدلة إمكانية التعميم، والأدلة الخارجية، وأدلة العواقب. وقد اختلفت هذه النظرة عن النظرة التقليدية التي تُقسّم الصدق إلى ثلاثة أنواع؛ صدق المحتوى، وصدق المحك، وصدق التكوين الفرضي. وتتلخّص أدلة الصدق الحديثة فيما يلي:

- جانب المحتوى، ويتضمن أدلة وثيقة الصّلة بالمحتوى، ومدى تمثيله، ومعايير جودته الفنية.
- -جانب التأصيل النظري والعمليات الذي يشير إلى أساس المنطق النظري لاتّساق استجابات المفحوص، مُتضمّنة تمثيل العمليات المستخدمة في الاستجابة، إضافة إلى الأدلة الإمبريقية المرتبطة بانغماس المستجيبين فعلاً في العمليات في أثناء التقييم.
- الجانب البنائي الذي يُقيّم مدى دقة بنية التقدير في مجال التكوين الفرضي موضع الاهتمام .
- جانب إمكانية التعميم يفحص مدى تعميم خصوصيات الدرجات وتفسيراتها عبر مجموعات الأفراد، والمؤسسات، والمهام، مُتضمّنة تعميم صدق ارتباطات محك الأداة.
- الجانب الخارجي يشتمل على دلائل تقاربية وتمييزية من مقارنات متعددة السمات- متعددة الطرق، مثلها أدلة متعلقة بالمحك وفائدة التطبيق.
- جانب العواقب الذي يُقيّم مُتضمنات القيمة في تفسير الدرجات كأساس للنشاط، مثل العواقب المرجوة وغير المرجوة لاستخدام الأداة، وبصورة خاصة النظر إلى مصادر عدم الصدق المتعلقة بمسائل التحيّز، والإنصاف، والعدالة التوزيعية.

أكد (1995) Messick أن الصدق مفهوم موحد ومتكامل يندرج ضمن صدق التكوين الفرضي، ومع ذلك أشار إلى أن تأييد الصدق يجب أن يدمج ستة جوانب متكاملة، فالجوانب الستة للتكوين الفرضي تُطبَق على كل القياسات التربوية والنفسية، فهي تُقدم أسلوباً لتوجيه قضايا الصدق المتعدّدة بشكل متبادل، ويتطلب الإجابة عليها تقديم تبريرات لاستخدام وتفسير

الدرجات، والعلاقة بين الأدلة والاستدلالات التي أثارها يجب أن تُحدد بؤرة تأييد الصدق، بحيث تدمج حجج منطقية ونظرية مقنعة بأن الأدلة المحصل عليها تُدعم الاستدلالات.

وقد أشارت "معايير العملية الاختبارية التربوية والنفسية" (AERA, APA & NCME, 2014) إلى أنَ المغزى من براهين الصدق التي تدمج أدلة متعدّدة هو تقديم وصف متماسك إلى الدّرجة التي تُدعّم بها الأدلة المتاحة والنظرية التفسيرات المرجوة لدرجات الاختبار لاستخدامات معينة، وتشمل الأدلة المجمعة من دراسات جديدة إضافة إلى الأدلة المستفادة من البحوث المنشورة.

2. مفهوم الثبات

تطور مفهوم الثبات تطوراً سريعاً خلال العقود القليلة الماضية شأنه شأن الصدق، ولكنه حظي باهتمام أكبر من طرف المؤلفين باعتباره أحد المسائل المهّمة التي تُبنى عليه أدوات البحث. والثبات بمفهومه التقليدي يشير إلى اتّساق القياسات من خلال إعادة تطبيق أداة القياس على مجتمع من الأفراد، ولكن هذه النظرة تقليدية حول الثبات، لأن النظرة الحديثة تُشير إلى اتّساق القياسات عبر مختلف أبعاد الموقف، مثلا: عبارات، فترات، مقيّمين، صيغ...، بمعنى أن الثبات يضمن أن تكون درجات الأفراد المختبرين نفسها تحت مختلف الظروف أي قابلة للتكرار عبر المواقف والظروف التي تُطبّق فيها أدوات القياس.

وردت في أدبيات القياس طرق عديدة لتقدير ثبات أدوات القياس، وتضمنت هذه الطرق تقدير معاملات الثبات والأخطاء المعيارية للقياس التي تقوم على النظرية الكلاسيكية للاختبارات (التطبيق وإعادة التطبيق، تقديرات المحكمين، الصيغ المتكافئة، الاتساق الداخلي). وبناء عليه قدمت النظرية الكلاسيكية طرقاً مختلفة لتقدير ثبات القياسات النفسية والتربوية تعتمد على ثلاث طرق أساسية، حيث يعتمد بعضها على تطبيق اختبارين، وتعتمد طرق أخرى على تطبيق اختبار واحد، في حين تعتمد طرق أخرى على التباين المشترك بين البنود (Crocker & Algina, 2006).

أما طريقة الاستقرار والتكافؤ فهي تجمع بين طريقة الاختبار-إعادة الاختبار وطريقة الصيغ المتكافئة، وتُصمّم دراسة الثبات بتطوير صيغتين من الاختبار تُطبّقان خلال فترتين مُختلفتيْن، والارتباط بين الدرجات في الصيغتيْن المأخوذتيْن من أوقات مختلفة تُنتج معامل ثبات يُتوقّع أن يكون أقل من أو يساوي كلاً من معامل الاستقرار ومعامل التكافؤ.

تتضمن الطرق التي تعتمد على تطبيق اختبار واحد طرق التجزئة النصفية التي يُطبّق فيها مُطوّر الاختبار صيغة واحدة من الاختبار على عينة من الأفراد، وتجزئة البنود إلى اختبارين فرعيين يشتمل كل نصف على عدد البنود نفسها، وتُستخدم في تقدير ثبات من درجات نصفي الاختبار: طريقة سبيرمان- براون، وطريقة رولون، وطريقة قاتمان.

أما الطرق التي تعتمد على التباينات المشتركة بين البنود، فتتمثل في تطبيق الأداة على عينة واحدة خلال فترة واحدة، ومن أشهر الطرق التي تعتمد على التباين المشترك: طريقة

كيودر-ريتشاردسون، وطريقة ألفا، وطريقة هويت لتحليل التباين. ويُعدّ معامل ألفا من أكثر طرق تقدير الثبات استخداماً في تقارير البحوث النفسية والتربوية على الإطلاق (Cronbach & Shavelson, 2004 ; Laveault, 2012).

ويُستخدم معامل ألفا في حساب التجانس الداخلي للبنود ثنائية التصحيح أو متدرجة التصحيح مثل: اختبارات المقال أو مقاييس الاتجاهات، وتُستخدم صيغتا كيودر-ريتشاردسون رقم 20 في البنود ثنائية التصحيح، وصيغة كيودر-ريتشاردسون رقم 21 بافتراض تساوي صعوبة جميع البنود، أما طريقة هويت فتعتمد على استخدام أسلوب على تحليل التباين بمعالجة الأفراد والبنود على أنها مصادر للتباين.

وتعتمد طريقة الثبات بين المقدّرين على تطبيق مجموعة واحدة من البنود (مثلا: اختبار ذات مهام مفتوحة) ويتم جمع الملاحظات فيها من طرف اثنيْن أو أكثر من المقدّرين (الملاحظين أو المصحّحين)، ويهتم صانع القرار في هذه الحالة بتجانس الملاحظات أو البيانات عبر المقدّرين، وتُستخدم في تقدير الثبات وفق هذه الطريقة على معاملات الاتفاق، ومعامل كاندل.

تندرج الطرق التي طُوّرت لتقدير ثبات القياسات النفسية والتربوية ضمن النظرية الكلاسيكية للاختبارات، وتوجد طرق أخرى لتقدير الثبات على غرار معامل ألفا الطّبقي، ومعامل أوميغا، ومعامل الثبات المركب (تيغزة، 2017). وقد عالجت هذه الطرق أوجه القصور في معامل ألفا في حالة عدم استيفاء شروط تطبيقه، حيث يتطلب معامل ألفا أن تكون السّمة المُقاسة متجانسة الأبعاد أي أنها تتضمن بُعداً واحداً فقط، وضرورة استيفاء شرط أن يكون تباين الدّرجات الحقيقية للعبارات متساوياً ولا تختلف هذه الدرجات إلاّ بمقدار ثابت واحد (تيغزة، 2017).

3. بعض الأخطاء الناجمة عن جمع أدلة الصدق

من الممارسات الشّائعة لدى الباحثين في التربية وعلم النفس قياس الصدق الداخلي أو صدق الاتّساق الداخلي لأدوات البحث، وبعدها يُكرّرون العملية باستخدام معادلة الاتساق الداخلي لألفا اعتقاداً منهم بأنّ كلا منهما يقيس جانباً مستقلاً عن الآخر، أي أن الاتّساق الداخلي يسمح بتقدير الارتباط بين العبارات ببُعدها والارتباط بين البُعد والاختبار ككل، وهذا تكرار لقياس الثبات مرتيْن، مرة تحت مُسمّى الثبات عن طريق الاتّساق الداخلي باستخدام معادلة ألفا، ومرة تحت مُسمّى صدق الاتّساق الداخلي (تيغزة، 2017).

وهكذا يُوظّف الاتّساق الداخلي للاستدلال على توفر الصدق وعلى توفر الثبات على حدّ سواء، رغم أن نظرية الصدق الحديثة تعتبر الصدق الداخلي دليلاً من الأدلة الستة التي تُبرهن على توفر الصدق(Messick, 1995)، فلا يكتفي الفاحص للأدلة التي تعتمد على البنية الداخلية فقط للتحقّق من صدق أدوات القياس.

وهناك ممارسة أخرى شائعة بين الباحثين في استخدام طريقة الصدق الداخلي، فعندما يتم تقدير معاملات ارتباط البنود بالبُعد أو بالدّرجة الكلية للأداة، ويتم الحكم على صدق البنود على الدلالة الإحصائية باستبعاد كل عبارة غير دالة إحصائياً، ممّا يؤدي إلى بخس تمثيل التكوين

الفرضي الذي اعتبره Messick (1995) تهديداً للصدق البنائي، الذي يؤدي إلى عدم تغطية أبعاد السّمة المُقاسة. ويذكر تيغزة (2017) كذلك أن هذا الإجراء يضرّ بالصدق أكثر مما ينفعه؛ لأنّ الدلالة النظرية للعبارة لا توزاي بالضرورة الدلالة الإحصائية، وأن هذه الأخيرة تتأثر بعوامل أخرى لا علاقة لها بالثّانية، منها تجانس الإجابات أو تباينها، وحجم العينة، وقوّة الاختبار الإحصائي.

ازداد الاهتمام في السنوات الأخيرة بترجمة وتكييف أدوات القياس (اختبارات، مقاييس، استبيانات...) نظراً لحاجة الباحثين للحصول على نسخ لغوية وثقافية متعدّدة لاستخدامها في المجالات النفسية والتربوية لاتّخاذ قرارات فردية أو جماعية. إلاّ أنّ هذه العملية تتطلّب إجراءات لجمع أدلة عن الصدق والثبات وتقديم المعايير، وذلك لأنّ تطوير أداة نفسية أو تربوية لمجموعة ثقافية أخرى تتطلّب أكثر من التّرجمة الحرفية التي كانت في السابق، وفي الغالب ممارسة شائعة.

أشارت Merenda (2005) إلى أن مجال الاختبارات (أو أدوات أخرى) النفسية والتربوية يُعاني من ممارسات خاطئة ناجمة عن سوء تطبيق وتفسير الاختبارات بشكل منتظم على المفحوصين الذين تُسبّب لهم مُعيقات لغوية وعوامل ثقافية تحولا دون صدق الاختبار، كسوء استخدام الأدوات عبر الثّقافات، والتفسير الخاطئ للدّرجات المُحصّلة الناجمة عن تطبيق التقييمات غير الملائمة للثقافة المستهدفة. لذلك فإنّ ترجمة وتكييف الاختبارات التربوية والنفسية ممارسة يمكن أن تتأثر بمصادر تحيّز متعدّدة راجعة إلى تحيّز البناء الذي يتعلق بعدم تكافؤ التكوينات الفرضية بين المجموعات الثقافية، وتحيّز الطريقة الذي ينتج من مشكلات في تطبيق الأداة، وتحيّز العبارة الذي ينتج في العادة من عدم ملاءمة الترجمات كاختيار كلمات غير صحيحة(van de Vijver & Hambleton, 1996).

وفي هذا الإطار قدّمت الهيئة الدولية للاختبار International Test Commission دليلاً يتضمن إرشادات للترجمة والتكييف التي يجب مراعاتها لضمان التكافؤ بين الاختبار الأصلي والاختبار المُكيّف من خلال نشر النسخة الأولى (International Test Commission, 2005) والنسخة الثّانية(International Test Commission, 2017)من الدليل التي تُستخدم قبل وفي أثناء وبعد عملية الترجمة والتكييف مُتضمّنة عمليات التحكيم والعمليات الإمبريقية لجمع أدلة عن تكافؤ وثبات وصدق درجات الاختبار في لغات وثقافات متعدّدة، إضافةً إلى إجراءات تطبيق الاختبار، وتفسير درجاته، وتوثيق عملية تكييفه.

وبالنظر إلى الممارسات الشائعة فإنّ أغلب الباحثين في التربية وعلم النفس لا يتبعون الإجراءات الأساسية في جمع أدلة عن صدق أدوات القياس، فعلى سبيل المثال فإنّ معظمهم لا يطلبون "ترخيصاً من صاحب الاختبار (أو أي أداة أخرى) قبل القيام بأي تكييف، وهذه ممارسة منتشرة في البيئة المحلية، رغم أن دليل الهيئة الدولية للاختبار ينص على "ضرورة الحصول على ترخيص من حامل حقوق الملكية الفكرية للاختبار قبل القيام بأي تكييف" (International Test Commission, 2017). إضافة إلى عدم تقديم أدلة كافية لتدعيم معايير وثبات وصدق

النسخة المُكيّفة للاختبار في المجتمعات المقصودة باستخدام الإجراءات الإحصائية المتقدّمة في الصدق والثبات كالتحليل العاملي التوكيدي، وتحليلات إمكانية التعميم، ونماذج نظرية الاستجابة للمفردة.

ويُستخدم في العادة التحليل العاملي الاستكشافي للتحقق من صدق أدوات البحث، الذي يسعى إلى اختزال تعدّد المتغيرات أو المؤشّرات المُقاسة إلى عدد قليل من المتغيرات الكامنة التي تُلخّصها، والكشف عن البنية العاملية الكامنة أو مساحات الدلالة المشتركة التي تكمن وراء تعدّد المتغيرات المُقاسة (تيغزة، 2011). ويُستخدم التحليل العاملي الاستكشافي في البحوث التربوية والنفسية بالأساس من أجل الكشف عن البنية العاملية من خلال تحديد عدد العوامل التي تتضمنها أداة البحث، وطبيعتها، والعبارات التي تتشبّع على كل عامل من العوامل المستخلصة.

ولكن غالبا ما يستخدم الباحثون في تقدير صدق أدوات القياس تلقائياً طريقة المكوّنات الأساسية لاستخراج العوامل رغم توفر طرق أخرى أكثر صلاحيةً من الطريقة المألوفة، مع العلم أن هذه الطريقة لا تصلح للكشف عن البينة العاملية للسّمة، وإنّما تصلح لاختزال عدد العوامل إلى عدد أقل، وهذا يعود إلى نقص الثقافة الإحصائية، ونقص تكوين الباحثين، وتشجيع بعض الحزم الإحصائية لاستخدام هذه الطريقة من خلال تنصيبها بشكل تلقائي (تيغزة، 2017).

رغم أهمية استخدام التحليل العاملي الاستكشافي في تحديد البنية العاملية لأدوات البحث إلا أنها عملية غير كافية ومحدودة، حيث لا يمكن الاكتفاء بها من طرف الباحثين في تأييد صدق أدوات البحث، وذلك على اعتبار أن هذه الطريقة لا تسمح بالتأكد من مُطابقة النموذج المفترض للأداة مع البيانات التي يتم جمعها من العينة التي يفترض أن تُمثّل المجتمع الإحصائي. وهذه العملية تتم باستخدام التحليل العاملي التوكيدي الذي يُستخدم في التحقّق من ملاءمة نموذج القياس المُستخدم مع بيانات العينة (Byrne, 2010).

وتُقدّم طريقة التحليل العاملي التوكيدي التي تندرج ضمن طرق النمذجة بالمعادلات البنائية توفر مزايا في تقدير الصدق البنائي، حيث تقوم بالحكم على ملاءمة النموذج العاملي الذي يعكس بنية معينة للمفهوم على عديد الأدلة (محك الدقة، محك الاقتصاد، المحك النظري)، والمقارنة بين النماذج المتنافسة بمرونة كبيرة بناء على مؤشّرات المُطابقة، وعدم اشتراطها استقلالية أخطاء القياس بافتراض ارتباط بعض منها (تيغزة، 2017).

ورغم ما تقدمه طريقة التحليل العاملي التوكيدي من أدلة موثوقة حول الصدق البنائي إلاّ أن استخدامها خاصة في البيئة المحلية ـ أو حتى العربية لا زال ضئيلاً بالمقارنة مع المميّزات التي تتمتع بها، وتجدر الإشارة أن هذه الطريقة تسمح أيضاً بتقديم مؤشرات مهمّة عن الثبات (مثلا: طريقة أوميغا الموزونة).

4. بعض الأخطاء الناجمة عن جمع أدلة الثبات

استُخدمت طرق تقدير الثبات المُطوّرة في النظرية الكلاسيكية للاختبار بشكل واسع من طرف طلاب الدراسات العليا في إنجاز رسائلهم العلمية، ومن طرف الباحثين في إنجاز بحوثهم وتقديم تقارير حولها، لكن في كثير من الأحيان ما يسوء استخدامها أو تفسيرها بشكل صحيح. فعلى سبيل المثال يُعدّ معامل ألفا كرونباخ أكثر شهرةً واستخداماً في البحوث، فبعد ظهوره ذُكر في أكثر من 17600 منشور منذ بداية الخمسينات(Dunn, Baguley & Brunsden, 2013) وحتى أن كرونباخ في حدّ ذاته وزميله شافلسون سنة 2004 لم يتصوّرا أن يتوسّع وينتشر استخدام معامل ألفا بهذا الشّكل (Cronbach & Shavelson, 2004).

ورغم ذلك كثيراً ما يتم انتهاك افتراضات استخدامه، حيث إن الكثير من الباحثين يتهافتون لاستخدامه في بحوثهم، ولكن دون استيفاء افتراضاته الأساسية التي يقوم عليها. فمعامل ألفا يتطلب أن تكون العبارات متجانسة الأبعاد (أي تتضمن بُعداً أو عاملاً واحداً)، ويكون تباين الدرجات الحقيقية للعبارات متساوية، حيث لا تختلف إلاّ بمقدار ثابت (تيغزة، 2017؛Laveault, 2012).

ويطغى على تقديرات الثبات في البحوث النفسية والتربوية استخدام معامل ألفا، وطرق التجزئة النصفية (خاصة طريقة سبيرمان-براون، وقاتمان)، والصيغ المتكافئة، وهذا يعود بالأساس إلى سهولة استخدامها من جهة، وإمكانية تقديرها باستخدام البرامج الإحصائية على غرار برنامج SPSS من جهة أخرى. وتجدر الإشارة إلى وجود طرق أخرى لتقدير الثبات أكثر تلاؤماً مع طبيعة البيانات (عدم استيفاء شرط تجانس العبارات، وتعدد عوامل السّمة المُقاسة) مثل: معامل ألفا كرونباخ الطّبقي، ومعامل أوميغا، ومعاملات أخرى للثبات في حالة اتخاذ قرارات مطلقة، مثل: معامل لوفنجستون، ومعامل برينان وكاين(Meyer, 2010).

إضافة إلى عدم مراعاة الفترة الفاصلة بين الاختبارين في أثناء استخدام ثبات الاختبار-إعادة الاختبار خاصة إذا كانت السّمة المُقاسة متغيرة نسبياً، فاختبارات التحصيل مثلا تتطلب أن لا تكون الفترة الفاصلة طويلة حتى لا يحدث نضجاً لدى الأفراد المُختبرين، ولا تكون قصيرة فيحدث تذكّرا للإجابات المقدّمة في الفترة الأولى، في حين أن مقاييس الدافعية واستراتيجيات التنظيم الذاتي والاتجاهات على سبيل المثال سمات ثابتة نسبياً يمكن أن تكون الفترة الفاصلة بين التطبيقين أكبر.

وفي العادة يستخدم الباحثون طرق تقدير الثبات التي تعتمد على القرارات النسبية (مقارنة الفرد بين زملائه لتحديد مكانته)، مثل: معامل ألفا، معامل قاتمان، معامل كيودر-ريتشاردسون، في حين أن طرق تقدير الثبات التي تعتمد على القرارات المطلقة (تحديد أداء الفرد بالمقارنة بمحك أو هدف خارجي)، مثل معامل لوفنجستون، معامل كاين وبرينان أقل استخداماً، وقد يعود ذلك إلى عدم معرفة الباحثين بهذه المعاملات أو تركيزهم على المعاملات الأكثر استخداماً في البحوث الأخرى (مثل: معامل ألفا). إضافة إلى قلّة استخدام معاملات الثبات حتى في حالة القرارات النسبية (مثل: معامل ألفا الطّبقي، ومعامل أوميغا) التي لا تفترض تكافؤ العبارات، وتساوي تباين الدرجات الحقيقية للعبارات، وتجانس السّمة المُقاسة.

لا زالت عمليات التحقّق من ثبات أدوات البحث تتمحور بالأساس حول أساليب النظرية الكلاسيكية للاختبارات رغم ما تزخر به الأدبيات من أساليب وطرق إحصائية حديثة متقدّمة تساعد على جمع أدلة كافية وذات موثوقية حول الثبات. فالنظرية الكلاسيكية للاختبارات رغم بساطتها وتقديمها لأساليب متعدّدة لتقدير الثبات إلاّ أنّها محدودة من حيث قدرتها على التحكّم في مصادر الخطأ المتعدّدة، لأنّها تتحكّم في مصدر خطأ أحادي غير مميّز (Bertrand & Blais, 2004).حيث تسمح بتقدير مصدر الخطأ الراجع إلى معاينة المحتوى (طرق التجانس الداخلي)، أو معاينة صيغة الأداة (طريقة التكافؤ)، أو إلى معاينة فترة تطبيق الأداة (طريقة الاختبار-إعادة الاختبار)، أو إلى معاينة المقدّرين أو الملاحظين (طرق اتّساق التقديرات).

ومن التطوّرات التي حدثت في القياس ظهور نظرية إمكانية التعميم التي تُعدّ إطاراً فكرياً وإحصائياً لتقدير ثبات أو اتّساق القياسات السلوكية(Brennan, 2001 ; Shavelson & Webb, 1991). حيث تسمح نظرية إمكانية التعميم بالتحكّم في المصادر المتعدّدة لخطأ القياس التي تؤثر على اتساق درجات الأفراد سواء الرّاجعة إلى العبارات أو الملاحظين أو الفترات أو الصّيغ، وتفاعلاتها الممكنة فيما بينها أو بينها وبين الأفراد، فهي على عكس النظرية الكلاسيكية التي تسمح بالتحكّم بخطأ قياس أحادي غير مميّز راجع لأحد الأبعاد المذكورة.

يمكن أن توفّر نظرية إمكانية التعميم بالمقارنة مع النظرية الكلاسيكية للاختبارات مزايا مختلفة؛ كإمكانية إدماجها مصادر متعدّدة للخطأ وتقدير (ثبات الاختبار-إعادة الاختبار، والاتساق الداخلي، والصدق الظاهري، والثبات بين التقديرات)في الوقت نفسه،وتقدير ليس فقط تأثير أبعاد القياس على حدة ولكن تأثيرات التفاعل أيضاً، وتوسيع ثبات القياسات مع الاقتصاد من التكاليف والوقت، وتقديم معلومات عن ثبات التفسيرات النسبية وحتى التفسيرات المطلقة (Yin & Shavelson, 2008) إلاّ أن استخداماتها في تقدير ثبات القياسات التربوية والنفسية قليلة لا ترقى إلى مستوى المزايا التي تتمتّع بها.

فقد أشار(Briesch, Swaminathan, Welsh & Chafouleas, 2014) أنه على الرّغم من نقاط القوة التي تتمتّع بها نظرية إمكانية التعميم إلاّ أن جهود الباحثين في مجال التربية وعلم النفس كانت بطيئة في تبني واستخدام هذه النظرية، وقد يعود ذلك إلى عدم اكتمال فهم أسسها المفاهيمية أو خطواتها العملية التي ينطوي عليها تصميم وتنفيذ دراسات إمكانية التعميم أو إلى مزيج منهما. كما أنّ أساليب تقدير معلمات نظرية إمكانية التعميم (مكوّنات التباين، معاملات إمكانية التعميم) معقّدة جداً خاصة بالنسبة للباحثين الذين يمتلكون معارف محدودة في تحليل التباين، وخاصّة في التصميمات متعدّدة الأبعاد ذات عدّة شروط أو عدّة أبعاد للقياس.

وفي الغالب، يقدم الباحثون تقارير عن ثبات أدوات البحث التي يستخدمونها في جمع بياناتهم من خلال وصف معاملات الثبات المُحصّلة، وذلك دون مراعاة تقديم تقرير عن خطأ أو أخطاء القياس، ومتوسط الارتباط بين العبارات، لأنّ هذا يسمح للقارئ بتقدير حجم الاتّساق

الداخلي الناتج عن الارتباطات العالية بين العبارات وليس الناتج عن تضخّم قيمة معامل الثبات بسبب العدد الكبير من البنود مع وجود ارتباط ضئيل بينها(Laveault, 2012).

فالكثير من البحوث المنشورة تُقدّم بالأساس تقريراً عن معاملات الثبات التي يتم وفقها تحديد مدى اتّساق درجات الأداة، ولكن كثيراً ما يتم إغفال تقديم معلومات مهمّة عن الخطأ المعياري للقياس الذي يُعدّ أفضل جزء من المعلومات المطلوب تقديم تقرير حولالأداة وليس معامل الثبات، فالخطأ المعياري يبيّن الشك المتعلق بكل درجة، ويمكن فهمه بسهولة من طرف مفسّري الاختبار المحترفين، وحتى الأشخاص غير المتمرّسين في نظرية الإحصاء، والأشخاص العاديين الذين تُوصف لهم الدرجات المحصّلة من الاختبارات (Cronbach & Shavelson, 2004, p. 413).

خاتمة ونتائج الدراسة

أوضحنا في هذا المقال أن عملية فحص صدق وثبات أدوات البحوث التربوية والنفسية مرحلة مهمّة تحتاج من الباحث استخدام أساليب مناسبة، وكافية لجمع أدلة ذات مصداقية من أجل اتّخاذ قرارات ملائمة حول النتائج المحصلة، وأن التأكّد من صدق وثبات الأدوات يسمح بالحصول على نتائج قابلة للتعميم على سياقات مختلفة وعلى مجتمعات مختلفة من الأفراد. فمن خلال فحص الصدق يقوم الباحث بجمع أدلة نظرية وعملية كافية وملائمة لتدعيم تفسيرات درجات الأدوات التي تتطلبها الاستخدامات المفترضة، ومن خلال فحص الثبات يتأكّد الباحث من اتّساق الدّرجات المُحصّلة من الأدوات عبر مختلف الظروف من خلال قابلية تعميمها على نطاق أوسع من الظروف (عبارات، فترات، صيّغ، ملاحظين...).

وبالنظر إلى الممارسات الحالية للباحثين–خاصة في السياق المحلي والعربي- فإنّ جمع أدلة عن صدق وثبات أدوات البحث عرضة لجملة من الأخطاء؛ منها ما يتعلق بانتهاك افتراضات بعض الأساليب المستخدمة في الثبات والصدق، ومنها ما يتعلّق باستخدام أساليب تقليدية غير ملائمة؛ ومنها ما يتعلق بعدم اتباع إجراءات مناسبة في جمع الأدلة. وقد استعرضنا مجموعة من هذه الممارسات الخاطئة التي تمّ تلخيصها فيما يلي:

- -تكرار قياس الثبات مرتيْن من خلال تقدير الصدق الداخلي (ارتباط العبارات ببُعدها، والارتباط بين البُعد والأداة ككل) وتقدير الثبات باستخدام معامل ألفا، بالرغم أن طريقة صدق الاتّساق الداخلي تدل على الثبات والصدق على حد سواء، فهما وجهان لعملة واحدة.
- -استبعاد العبارات غير الدّالة إحصائياً في أثناء استخدام طريقة الصدق الداخلي (ارتباط العبارات بالبُعد أو بالدرجة الكلية) على اعتبار أنها غير صادقة، في حين أن هذا يؤدي إلى عدم تمثيل التكوين الفرضي المُقاس، مما يضرّ بالصدق أكثر ممّا يفيده.
- -بعض الباحثين لا يطلبون رخصة من مالك حقوق الملكية قبل القيام بأيّ تكييف، وهذه ممارسة منتشرة بين الباحثين رغم أهمية هذا الإجراء وضرورته للقيام بعملية الترجمة والتكييف الذي اعتبرته الهيئة الدولية لترجمة وتكييف الاختبارات شرطاً أولياً يجب مراعاته.

-عدم تقديم أدلة كافية لتدعيم معايير وصدق وثبات النّسخ المُكيّفة للاختبار في المجتمعات المقصودة باستخدام الإجراءات الإحصائية المتقدّمة كالتحليل العاملي التوكيدي، وتحليلات إمكانية التعميم، ونماذج نظرية الاستجابة للمفردة.

-شيوع استخدام طريقة المكوّنات الأساسية من طرف الباحثين للكشف عن البنية العاملية للسّمة المُقاسة رغم عدم صلاحيتها في ذلك، وإنّما تصلح فقط لاختزال العوامل إلى عدد أقل، لأنّ هذه الطريقة لا تُصفي العبارات من تباين الخطأ والتباين الخاص.

-قلّة استخدام التحليل العاملي التوكيدي –خاصة في البيئة المحلية- الذي يسمح بالتحقّق من مدى ملاءمة النموذج المفترض للأداة مع بيانات العينة التي توفّر مزايا متعدّدة في تقدير الصدق البنائي باستخدام أدلة مختلفة.

- انتهاك افتراض استخدام معامل ألفا الذي انتشر كثيراً بين الباحثين في تقدير الثبات، فافتراض أحادية البُعد (تجانس عبارات الأداة) والتكافؤ بالأساس "تو" (تساوي تباين الدّرجات الحقيقية للعبارات التي لا يجب أن تختلف إلاّ بمقدار ثابت) شروط لا يتم استفاؤها أو التحقّق من توفّرها لدى أغلب الباحثين.

- اكتفاء بعض الباحثين بتقديم معاملات الثبات لتحديد مدى اتّساق درجات الأداة، ويغفلون عن تقديم معلومات مهّمة عن الخطأ المعياري للقياس، وهذا مهم جداً ويعدّ أفضل جزء من المعلومات المطلوب تقديمها في وصف ثبات أدوات البحث.

- ارتكاز معظم الباحثين على استخدام أساليب النظرية الكلاسيكية للاختبارات في التحقّق من ثبات أدوات البحث (خاصة معامل ألفا) رغم محدوديتها في قدرتها على التحكّم في مصادر خطأ القياس المتعدّدة التي تؤثر على الثبات.

- قلّة استخدام نظرية إمكانية التعميم رغم المزايا المتعدّدة التي تتمتّع بها بالمقارنة مع النظرية الكلاسيكية، كقدرتها على التحكم في المصادر المتعدّدة لخطأ القياس فيالوقت نفسه، وقياسها تأثير أبعاد القياس وتفاعلاتها فيما بينها، وتقديمها معلومات عن ثبات التفسيرات النسبية والمطلقة، وتوسيعها لثبات القياسات مع الاقتصاد من الوقت والجهد والتكاليف.

- طرق تقدير الثبات المستخدمة من طرف معظم الباحثين تعتمد أغلبها على اتّخاذ قرارات نسبية (استخدام أداة القياس لمقارنة الأفراد فيما بينهم في السّمة المُقاسة) دون التركيز على أساليب تقدير الثبات المعتمدة على القرارات المطلقة (استخدام الأداة لمقارنة أداء الفرد بمحك أو بهدف معين) التي تتوفّر في أدبيات الباحث (مثل: معامل لوفنجستون، معامل برينان وكاين، معامل سوبكوفياك...).

بناءً على ما تقدّم يحتاج الباحثون إلى مراعاة بعض الممارسات في أثناء جمع أدلة عن صدق وثبات أدوات البحث، وهذا يتطلّب الاطلاع على مستجدات القياس التربوي والنفسي،وتصحيح بعض الممارسات الخاطئة التي تقلّل من جودة أدوات جمع البيانات التي تؤدّي في النهاية إلى نتائج مضلّلة. فاطلاع الباحث

واستخدامه للأساليب الحديثة للقياس، على غرار أساليب التحليل العاملي التوكيدي، ونظرية إمكانية التعميم، وأساليب نظرية الاستجابة للمفردة، وتوصيات الهيئة الدولية للاختبار يمكن أن يحسّن من جودة البحث التربوي والنفسي.

قائمة المصادر و المراجع

أولا: المراجع باللغة العربية

1. تيغزة، أمحمد بوزيان. (2011). *التحليل العامل الاستكشافي والتوكيدي: مفاهيمهما ومنهجيتهما بتوظيف حزمة SPSS وليزرل LISREL*. عمان: دار المسيرة.
2. تيغزة، أمحمد بوزيان. (2017). توجهات حديثة في تقدير صدق وثبات درجات أدوات القياس: تحليل نظري تقويمي وتطبيقي. *مجلة العلوم النفسية والتربوية*، 4(1)، 7-29.

ثانيا: المراجع باللغة الاجنبية

3. American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington: DC: Author.
4. Bertrand, R., & Blais, J. G. (2004). *Modèles de mesure: L'apport de la théorie de réponse aux items*. Canada : Presses de l'Université du Québec .
5. Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag .
6. Briesch, A. M., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design implementation, and interpretation. *Journal of School Psychology*, 52(1), 13–35.
7. Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routeldge.
8. Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.
9. Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient Alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391-418.
10. Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412.
11. Hubley, A. M., & Zumbo, B. D. (2013). Psychometric characteristics of assessment procedures: An overview. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology* (pp. 3-19). Washington, DC: American Psychological Association.
12. International Test Commission. (2005). *International guidelines on test adaptation*. Availible online at : www.intestcom.org

13. International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2ndEd.). Availible online at : www.InTestCom.org
14. Laveault, D. (2012). Soixante ans de bons et mauvais usages du alpha de Cronbach. *Mesure et évaluation en éducation*, 35(2), 1-7.
15. Laveault, D., & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. Bruxelles: De Boeck.
16. Merenda, P. F. (2005). Cross-cultural adaptation of educational and psychological testing. In R. K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 321-341). Mahwah, NJ: Lawrence Erlbaum Publishers.
17. Messick, S. (1989). Validity. In R. L. Linn (Ed.). *Educational measurement* (pp. 13-103). New York: Macmillan Publishing.
18. Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749 .
19. Meyer, J. P., (2010). *Reliability*. New York: Oxford University Press.
20. Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage Publications.
21. Urbina, S. (2004). *Essentials of psychological testing*. New Jersey: John Wiley & Sons.
22. van de Vivjer, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1(2), 89-99 .
23. Yin, Y., & Shavelson, R.J. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education*, 21(3), 273-291.