الثبات أم الصدق أولاً ؟

Reliability or Validity first ?

أ.د عزت عبد الحميد محمد حسن
أستاذ علم النفس التربوي
كلية التربية جامعة الزقازيق

ملخص:

يهدف هذا البحث إلى إلقاء الضوء على أهم مفهومين من مفاهيم القياس وهما: الثبات والصدق، وتحديد أيهما يجب أن يأتي أولاً أو يتم حسابه قبل حساب الآخر، بالإضافة إلى التعرف على الطرق الشائعة لحساب كل منهما، وكيفية تحسين كل من الثبات والصدق.

وقد تطرق البحث إلى عدة أمور تتعلق بالثبات والصدق وأجاب على أسئلة كثيرة منها: ما أفضل مستوى للثبات؟ كيف يمكن جعل الاختبار أكثر ثباتاً؟ وكيف يمكن جعل الاختبار أكثر صدقاً؟ ما مكونات أو أدلة الصدق؟ ومن خلال دراسة العلاقة بين الثبات والصدق تم التوصل إلى مجموعة من الاستنتاجات من أهما:

- أنه ينبغي حساب ثبات الاختبار أولاً ويأتي بعد ذلك حساب صدق الاختبار. نظراً لأنه ليس بالضرورة أن يكون كل ثابت صادق، ويعتبر حساب الثبات قبل الصدق بمثابة الفلتر الأول لتنقية أداة القياس من البنود أو العبارات غير الثابتة، ويأتي بعد ذلك الصدق الذي يعتبر بمثابة الفلتر الثاني لتنقية أداة القياس من البنود أو العبارات غير الصادقة.
- نظراً لأن الثبات شرط ضروري للصدق لكنه غير كاف، لذا فمن المنطقي أن نحسب الثبات قبل حساب الصدق؛ حتى نستوفي هذا الشرط أولاً قبل حساب الصدق.
- يجب استخدام أكثر من طريقة لحساب ثبات أدوات القياس وعدم الاكتفاء بطريقة واحدة؛ لأن طرق حساب الثبات تختلف في الكشف عن مصدر الخطأ في الاختبار. كما يجب حساب الصدق بطرق تغطي جميع مكونات الصدق كلما كان ذلك ممكناً.
- أن يكون الباحثين على وعي تام بأن استخدام أداة قياس غير صادقة يعني لا قياس.

الثبات أم الصدق أولاً ؟

Reliability or Validity first ?

أ.د عزت عبد الحميد محمد حسن

أستاذ علم النفس التربوي

كلية التربية جامعة الزقازيق

مقدمة:

يُعد الاهتمام بالخصائص السيكومترية لأدوات القياس ليس حديثًا، حيث استقطبت تلك الخصائص السيكومترية انتباه كثير من علماء القياس النفسي على مر العصور الماضية، لما لتلك الخصائص من أهمية بالغة، وخاصة إذا كانت تتعلق باختبارات أو أدوات قياس يتم في ضوء نتائجها اتخاذ قرارات أو عمل استدلالات أو تدخلات علاجية بشأن الأفراد أو الجماعات الذين تطبق عليهم تلك الاختبارات. ومن أهم موضوعات أو قضايا القياس موضوعين هما: **الثبات والصدق.**

فلكي نحصل على قياسات سليمة ودقيقة ينبغي أن تكون أدوات القياس على قدر كبير من الثبات والصدق وأن تكون خالية من التحيز والتشويه.

وقد شعر المؤلف الحالي بمشكلة هذا البحث أثناء حضوره المناقشات العلمية لرسائل الماجستير والدكتوراه سواء داخل مصر أو خارجها، حيث وَجَدَ أن بعض المناقشين أو المحكمين يخلطون خلطًا شديدًا بين المفهومين، ويتباينون في ترتيب كل منهما، حيث يصر كثير من المناقشين على وجوب حساب صدق أداة القياس أولاً وبعد ذلك يتم حساب ثباتها ولم يقدموا أي تبريرات لإصرارهم على ذلك، والقليل منهم من يعرف أنه يجب حساب ثبات أداة القياس أولاً وبعد ذلك يتم حساب صدقها، وهناك فريق ثالث لا يعطي لهذا الموضوع أي اهتمام، كما أنه على المستوى العالمي نجد كثير من الباحثين يتباينون أيضًا في ترتيب هذين المفهومين في عناوين الكتب والأبحاث الخاصة بهم. ولذا يهدف هذا البحث إلى القاء الضوء على هذا الجانب المهم من جوانب القياس النفسي والتربوي.

هدف البحث:

يهدف هذا البحث إلى إلقاء الضوء على أهم مفهومين من مفاهيم القياس وهما: الثبات

والصدق، وتحديد أيهما يجب أن يأتي أولاً أو يتم حسابه قبل حساب الآخر، بالإضافة إلى التعرف على الطرق الشائعة لحساب كل منهما، وكيفية تحسين كل من الثبات والصدق.

أولاً: الثبات: Reliability

يشير ثبات الاختبار إلى اتساقه (اتساق القياس)، فالاختبار الذي يتمتع بثبات جيد يعني أن مستخدم الاختبار سوف يحصل على نفس نتائج الاختبار عند تطبيقه عدة مرات، طالما أنه لا توجد عوامل خارجية أو أي عوامل دخيلة أخرى قد أثرت على نتائج الاختبار (Hilsenroth et al., 2004: 7).

أي أن الثبات يتعلق بما إذا كانت النتيجة قابلة للتكرار (Golafshani, 2003: 599). فالثبات هو اتساق القياس مع مرور الوقت أو استقرار القياس خلال الظروف المختلفة، والأسلوب الأكثر استخداما لتقدير الثبات هو قياس الارتباط (معامل الارتباط) وغالبًا ما يطلق عليه معامل الثبات. ومعامل الثبات هو العلاقة المتبادلة بين اثنين أو أكثر من المتغيرات (الاختبارات، والبنود، أو المقيِّمين) التي تقيس نفس الشيء (Drost, 2011: 108).

ويعرف الثبات إحصائيًا على أنه نسبة تباين الدرجة الحقيقة إلى تباين الدَّرجة المشاهدة (المقاسة). ففي العديد من طرق حساب الثبات يتم حساب معامل الثبات عن طريق حساب معامل ارتباط بيرسون بطريقة مباشرة (بين درجات الاختبار في التطبيقين، بين درجات الصورة (أ) والصورة (ب)، بين الدرجات التي تم الحصول عليها في الموقف ١ والموقف٢) ويُفسر معامل الارتباط الناتج أو معامل الثبات على أنه نسبة التباين في المقياس التي تكون ثابتة (Sechrest, 1984: 45).

وعمليًا لا يوجد اتساق أو ثبات تام، وقد وضعت نظريات ثبات الاختبار لتقدير آثار التضارب وعدم الاتساق على دقة القياس، ونقطة الانطلاق الأساسية لجميع نظريات ثبات الاختبار هو فكرة أن درجات الاختبار تعكس تأثير نوعين من العوامل:

١. العوامل التي تسهم في الاتساق: وتتمثل في استقرار خصائص الأفراد أو السمة المراد قياسها.
٢. العوامل التي تساهم في عدم الاتساق: سمات الأفراد أو الحالة التي يمكن أن تؤثر على نتائج الاختبارات ولكن ليس لديها ما تفعله مع السمة التي يجري قياسها، وتتمثل هذه العوامل في:

- الخصائص المؤقتة العامة للفرد: مثل الصحة، والتعب، والدافعية، والتوتر العاطفي.
- الخصائص المؤقتة والمحددة للفرد: فهم المهمة المحددة للاختبار، والحيل أو تقنيات التعامل مع المواد الخاصة بالاختبار، وتقلبات الذاكرة، والانتباه أو الدقة.
- مظاهر موقف الاختبار: التحرر من المشتتات، ووضوح التعليمات، وتفاعل شخصية أو الجنس، أو عرق مقدم الاختبار.
- عوامل الصدفة: الحظ في اختيار الإجابات عن طريق التخمين، والمشتتات المؤقتة (http://en.wikipedia.org/wiki/Reliability_psychometrics).

وأكثر الأشكال الشائعة للثبات هو: ثبات إعادة الاختبار test-retest reliability، والاتساق الداخلي internal consistency (Hilsenroth et al., 2004: 7):

حيث الثبات عن طريق إعادة الاختبار هو مقياس لاتساق الاختبار خلال فترة من الزمن، حيث يفترض ثبات إعادة الاختبار أن البناء (أو السمة) الذي يتم قياسه هو ثابت نسبيا مع مرور الوقت، مثل: نسبة الذكاء أو سمات الشخصية. فإذا كانت السمة موضع القياس من المحتمل أن تتغير بمرور الوقت (على سبيل المثال: القلق كحالة)، عندئذ يختار صانعي الاختبار عموما فاصل زمني أقصر (أسبوع واحد مثلاً). ويسجل ويفسر ثبات إعادة الاختبار على أنها معاملات الارتباط، ويعتبر ثبات إعادة الاختبار على أنه ممتاز إذا كان يساوي (٠.٩٠ فأكثر)، كما يعتبر ثبات إعادة الاختبار على أنه جيد إذا كان يساوي (٠.٨٠ فأكثر). أما إذا كانت السمة المقاسة ثابتة نسبيا لكن معامل ثبات إعادة الاختبار لتلك السمة حوالي (٠.٥٠) فإن هذا قد يعني أن هذا القياس غير ثابت أي أن الاختبار غير ثابت، وقد يرجع ذلك وجود عبارات ركيكة (مثل: النفي المزدوج الذي قد يصعب على معظم المستجيبين)، أو إلى بعض المتغيرات الخارجية أو المتغيرات الدخيلة التي تدخلت على السمة أثناء الفاصل الزمني لإعادة الاختبار. وتوجد مشكلة واحدة أخيرة لتفسير ثبات إعادة الاختبار وهي أنها قد يكون الثبات مرتفع ارتفاعًا زائفًا بسبب آثار الممارسة أو آثار الذاكرة، فقد يكون أداء المستجيب أفضل في التطبيق الثاني للاختبار وذلك لأن السمة موضع القياس تتحسن من خلال الممارسة. كما أن بعض الأفراد المستجيبين قد تكون استجاباتهم على الاختبار مماثلة لاستجاباتهم التي قدموها على نفس الاختبار في وقت سابق بسبب تذكرهم لكثير من تلك الإجابات. وأحد الحلول الممكنة لهذه المشكلة هو استخدام صور بديلة أو مكافئة للاختبار. وإذا كان مستخدم الاختبار مهتما بتغير السمة مع مرور الوقت ويشعر بالقلق إزاء آثار الممارسة أو الذاكرة فإن الصور البديلة أو المكافئة للاختبار يجب أن يضعها في الحسبان.

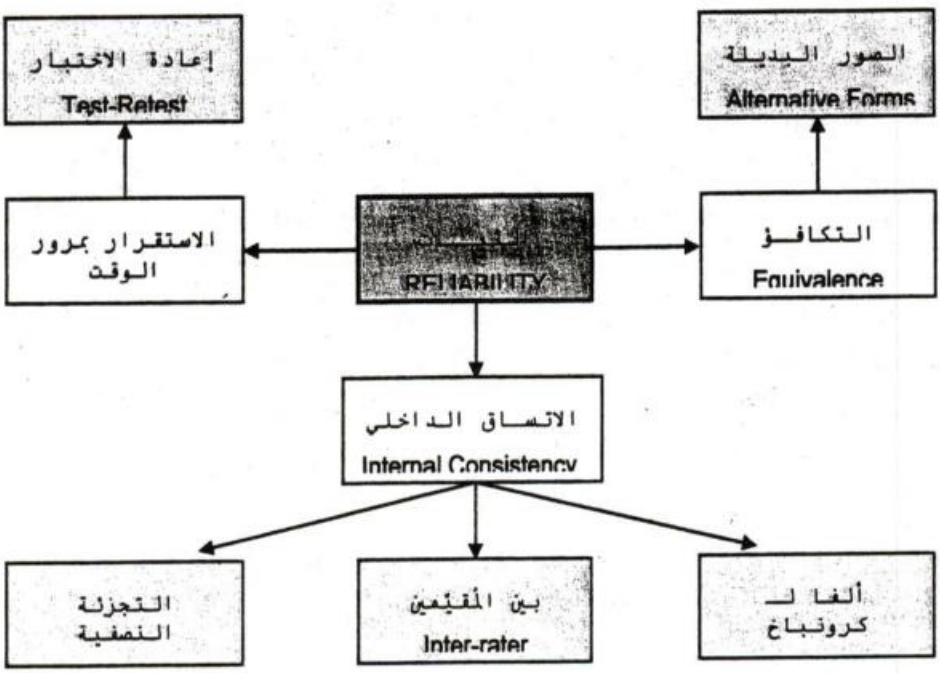
والاتساق الداخلي internal consistency هو مقياس لمدى ارتباط مفردات (أو عبارات) الاختبار بعضها ببعض. وأكثر أساليب ثبات الاتساق الداخلي شيوعا هو معامل ألفا لـ كرونباخ (١٩٥١)، وأحد الطرق البديهية لتفسير معامل ألفا كرونباخ هو أن ننظر إليه على أنه متوسط ارتباطات كل بند من بنود الاختبار مع البنود الأخرى على الاختبار. ويتم تفسير معامل ألفا مثل كثير من معاملات الارتباط (أي أنه يتراوح من صفر إلى ١)، وتشير قيمة ألفا الأكبر من أو تساوي (٠.٨٠) إلى ارتفاع الاتساق الداخلي للمقياس.

ويوضح الشكل رقم (١) الطرق الشائعة لحساب الثبات (Drost, 2011: 109)

وفيما يلي شرح مبسط لبقية الطرق الموضحة بالشكل السابق Phelan & Julie Wren, (2006):

ثبات الصور المتوازية Parallel forms reliability:

يقيس الثبات الذي يتم الحصول عليه عن طريق تطبيق صورتين متكافئتين لأداة القياس على عينة الدراسة في نفس الوقت، ويتم حساب معامل الارتباط بين درجات الأفراد في الصورتين لتقييم اتساق consistency درجات أداة القياس عبر الصور المتكافئة أو البديلة لأداة القياس.



شكل (١)

الطرق الشائعة لحساب ثبات أدوات القياس

ثبات المُقيّمين Inter-rater reliability

يقيس درجة اتفاق المُقدرين أو المحكمين في قرارات تقييمهم أو ملاحظتهم لنفس السلوك أو الأداء. ويُعد ثبات المقيمين أو المقدرين من أنواع الثبات المهمة والمفيدة؛ لأن المشاهدات الانسانية

قد لا تفسر بالضرورة الاستجابات بنفس الطريقة، فقد يختلف المقيمون حول مدى استجابات معينة أو المواد التي تظهر المعرفة للبناء أو المهارة التي يجرى تقييمها. ويستخدم هذا النوع من الثبات عندما يقوم مجموعة مُحكمين أو مُقدرين مختلفين بتقييم درجة بلوغ ملفات الانجاز معايير محددة. ويفيد هذا النوع من الثبات عندما تتأثر الأحكام بالعوامل الذاتية، حيث يكون أكثر احتمالا عند تقييم عمل فني بدلا من مسائل الرياضيات.

ثبات التجزئة النصفية Split-half reliability

وهو نوع أخر من أنواع ثبات الاتساق الداخلي، ويتم فيه تجزئة بنود الاختبار إلى نصفين، بإحدى طرق التجزئة ثم حساب معامل الارتباط بين نصفي الاختبار لتحديد معامل ثبات نصف الاختبار، ثم يصحح هذا المعامل بإحدى معادلات حساب ثبات التجزئة النصفية مثل: سبيرمان /براون أو جوتمان.

ما الهدف من تقدير الثبات ؟

الهدف من تقدير الثبات هو تحديد كمية التباين في درجات الاختبار التي ترجع إلى أخطاء في القياس وكمية التباين التي ترجع الى التغير في الدرجات الحقيقة. والدرجة الحقيقة هي السمة القابلة للتكرار للمفهوم الذي يتم قياسه، أي التي يمكن أن تتكرر عبر مناسبات مختلفة للقياس في حالة عدم وجود خطأ (http://en.wikipedia.org/wiki/Reliability_(psychometrics)

الثبات في ضوء نظرية القياس الكلاسيكية ونظرية الاستجابة للمفردة

الهدف من الثبات في ضوء نظرية القياس الكلاسيكية Classical test theory هو تقدير أخطاء القياس واقتراح سبل تحسين الاختبارات بحيث يتم الحد من تلك الأخطاء. والافتراض الرئيسي لنظرية الثبات هو أن أخطاء القياس تكون أساسًا بشكل عشوائي، وهذا لا يعني أن الأخطاء تنشأ من العمليات العشوائية. وخطأ القياس لأي فرد ليس حدثًا عشوائيًا تمامًا، ولكن يُفترض أن أسباب خطأ القياس لدى عدد كبير من الأفراد تكون متنوعة بحيث تكون أخطاء القياس بمثابة المتغيرات العشوائية. فإذا كان للأخطاء الخصائص الأساسية للمتغيرات العشوائية، فمن المعقول أن نفترض أن الأخطاء من المحتمل أن تكون إيجابية أو سلبية وأنها لا ترتبط مع الدرجات الحقيقية أو مع الأخطاء على الاختبارات الأخرى، وتفترض الآتي (http://en.wikipedia.org/wiki/Reliability_psychometrics):

١. متوسط خطأ القياس = صفرًا

٢. الدرجات الحقيقية والأخطاء غير مرتبطة.

٣. الأخطاء على المقاييس المختلفة غير مرتبطة.

وتظهر نظرية الثبات أن تباين الدرجات التي يتم الحصول عليها عبارة عن مجموع تباين الدرجات الحقيقية بالإضافة إلى التباين في أخطاء القياس، أي: $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$ وتشير هذه المعادلة إلى أن درجات الاختبار تختلف نتيجة لعاملين: التباين في الدرجات الحقيقية، والتباين بسبب أخطاء القياس. **ويوفر معامل الثبات** ($\rho_{xx'}$) مؤشرًا للتأثير النسبي لكل من الدرجات الحقيقية ودرجات الخطأ على درجات الاختبار التي يتم الحصول عليها (الدرجة المشاهدة)، وبوجه عام يُعرَّف معامل الثبات على أنه نسبة تباين الدرجة الحقيقية إلى التباين الكلي لدرجات الاختبار. وهذا يكافئ، واحد مطروحًا منه نسبة تباين الدرجة الخطأ إلى تباين الدرجة المشاهدة:

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

وللأسف لا توجد وسيلة مباشرة لملاحظة أو حساب الدرجة الحقيقية، ولذا يتم استخدام مجموعة متنوعة من الأساليب لتقدير ثبات الاختبار. وتختلف طرق حساب الثبات (سواء عن طريق إعادة الاختبار أو الصور المتكافئة، أو الاتساق الداخلي) في الكشف عن مصدر الخطأ في الاختبار.

كما أنه طبقًا لنظرية القياس الكلاسيكية يمكن التعبير عن الثبات كما يلي:

الدرجة المُقاسة (المشاهدة) = الدرجة الحقيقية ± درجة الخطأ

والدرجة الحقيقية لا يمكن معرفتها لأنه لا يوجد مقياس دقيق تمامًا. كما أن درجة الخطأ لا يمكن معرفتها أيضًا ولكن يمكن ضبط كمية الخطأ العشوائي والنظامي (DeVon et al., 2007: 160).

وكان من المعروف أن المنظرين لنظرية القياس الكلاسيكية أن دقة القياس ليست موحدة في جميع أنحاء نطاق القياس، فالاختبارات تميل إلى التمييز بشكل أفضل للمتقدمين للاختبار بمستويات معتدلة للسمة المُقاسة، والتمييز بشكل أسوأ بين المتقدمين للاختبار ذوي المستويات العالية والمنخفضة للسمة المُقاسة. وقد وسعت **نظرية الاستجابة للمفردة** Item response theory مفهوم الثبات من مؤشر واحد إلى وظيفة تسمي دالة المعلومات information function، وهذه الدالة عبارة عن معكوس الخطأ المعياري للدرجة المشاهدة المشروطة عند أي درجة معطاه للاختبار. **والهدف من تقدير الثبات** هو تحديد كمية التباين في درجات الاختبار التي

ترجع إلى أخطاء القياس وكمية التباين التي ترجع الى التغير في الدرجات الحقيقة. والثبات هو خاصية لدرجات المقياس بدلا من المقياس نفسه، كما أن الثبات الذي يتم حسابه من عينة أولى قد يختلف عن الذي يتم حسابه من عينة ثانية (بغض النظر عما هو متوقع بسبب تباينات المعاينة) إذا تم اشتقاق العينة الثانية من مجتمع مختلف لأن التغير الحقيقي مختلف في المجتمع الثاني.

أخطاء القياس Errors of Measurement

يوجد نوعين من أخطاء القياس هما: خطأ ثابت، وخطأ عشوائي. والخطأ العشوائي هو الذي يلعب دورًا مهما في نظرية القياس، ولكن أيضًا من المهم فهم الأخطاء الثابتة، فالخطأ الثابت هو الذي يكون متكرر بنفس القيمة في درجات جميع الأفراد. فإذا كان هناك نفس الخطأ بدرجات جميع أفراد العينة وتم اكتشاف هذا الخطأ فمن السهل التعامل مع هذا الخطأ الثابت وتصحيح درجات العينة وذلك إما بإضافة أو طرح هذا الخطأ الثابت من جميع الدرجات. إلاَّ أن المشكلة تصبح معقدة إذا كان الخطأ الثابت دالة للدرجات الأصلية (مثلاً: يكون أكبر في حالة الدرجات المرتفعة). أما الأخطاء العشوائية فهي تلك الأخطاء التي لا تكون مرتبطة من فرد لآخر، أو من موقف لآخر، أو من وقت لآخر، أو من اختبار لاختبار آخر. ومقدار الخطأ العشوائي (يتضمن التباين غير المُفسر) يمكن تقديره واستخدامه في تحديد حدود المدى الذي يمكن أن تقع فيه الدرجات الحقيقية، عن طريق حساب ما يسمى بالخطأ المعياري للقياس standard errors of measurement الذي يُحسب من المعادلة التالية:

$$SEM = \sigma\sqrt{1 - r_{xx}}$$ حيث:

(SEM) الخطأ المعياري للقياس، (σ) الانحراف المعياري لدرجات الاختبار، (r_{xx}) معامل ثبات الاختبار أو أداة القياس. وقيمة الخطأ المعياري للقياس الناتجة من المعادلة السابقة تحدد حدود المدى الذي يمكن أن تقع فيه الدرجات الحقيقية باحتمال يساوي (٠.٦٧)، أما المدى الذي يمكن أن تقع فيه الدرجات الحقيقية باحتمال يساوي (٠.٥٠) فيتم حسابه عن طريق أخذ (٠.٦٧٤٥) من الخطأ المعياري للقياس واستخدامه في حساب حدود ذلك المدى.

فمثلاً: إذا كان الانحراف المعياري لدرجات نسبة الذكاء = ١٥ ومعامل ثبات اختبار نسبة الذكاء = ٠.٨٥ فإن: $SEM = \sigma\sqrt{1 - r_{xx}} = 15\sqrt{1 - 0.85} = 5.8 \cong 6$

ومن هنا: إذا حصل أحد الأفراد على نسبة ذكاء تساوي ١١٥، لحساب المدى الذي يمكن أن تقع فيه الدرجة الحقيقية لهذا الفرد باحتمال يساوي (٠.٦٧) = ١١٥ ± ٦

أي أن الدرجة الحقيقة لهذا الفرد تقع في المدى الذي يمتد من ١٠٩ إلى ١٢١ باحتمال يساوي (٠.٦٧).

أما المدى الذي تقع فيه الدرجة الحقيقية لهذا الفرد باحتمال يساوي (٠.٥٠) = ١١٥ ± ٦ (٠.٦٧٤٥) = ١١٥ ± ٤

أي أن الدرجة الحقيقة لهذا الفرد تقع في المدى الذي يمتد من ١١١ إلى ١١٩ باحتمال يساوي (٠.٥٠) مع ملاحظة أن الخطأ المعياري للقياس يعتمد على معامل الثبات المستخدم في المعادلة لحسابه، ولذلك يتم تفسير الخطأ المعياري للقياس في ضوء مفهوم معامل الثبات المستخدم في حسابه (Sechrest, 1984: 31-32).

ويشير (Drost, 2011: 113) إلى أنه توجد مصادر عديدة للخطأ في الاختبار منها: التخمين على الاختبار، وتصحيح أوراق الإجابة بشكل غير صحيح، وتخطي سؤال عن غير قصد، وإساءة تفسير تعليمات الاختبار، ورغم ذلك فكلما ارتفع عدد بنود الاختبار انخفضت كمية الخطأ في الاختبار ككل.

ما أفضل مستوى للثبات ؟

يتم عرض ثبات الاختبار في شكل معامل الارتباط الذي تمتد قيمته من ٠ إلى ١ ويتم تفسير معاملات الثبات بناءً على قيمتها أو مقدارها، وبصفة عامة كلما ارتفع معامل الارتباط كان الاختبار أكثر اتساقا، ومعاملات الثبات الأقل من ٠.٥٠ تعتبر منخفضة، أما معاملات الثبات التي تمتد من ٠.٥٠ إلى ٠.٦٩ تعتبر متوسطة، ومعاملات الثبات التي تمتد من ٠.٧٠ إلى ٠.٧٩ مرتفعة نسبيًا، أما التي تمتد من ٠.٨٠ إلى ٠.٩٩ فتعتبر مرتفعة. ورغم ذلك، فإن معايير قوة العلاقة تختلف تبعا لموضوعية المفهوم موضع القياس (McGoeya et al., 2010: 106).

ويعتمد أفضل مستوى للثبات على الكيفية التي يتم بها استخدام المقياس. حيث يشير Nunnally (1978) إلى أن معامل الثبات الذي يساوي (٠.٧٠) فأكثر يكون كافيًا في المراحل الأولى من البحث على الاختبارات التنبؤية أو المقاييس المفترضة للبناء (أو للظاهرة موضع القياس)، ويؤكد أنه خلال هذه المرحلة محاولة زيادة معامل الثبات عن (٠.٨٠) غالبًا ما يكون مضيعة للوقت والأموال، وذلك لأن الارتباطات في هذه الحالة تكون ضعيفة بسبب خطأ القياس، وأنه للحصول على ثبات أعلى من (٠.٩٠) مثلا، يتطلب جهودًا مضنية في المُعايرة standardization وربما إضافة بعض البنود. أما في المواقف التطبيقية حيث يتم اتخاذ القرارات المهمة في ضوء درجات الاختبار، أوصى أن معامل الثبات الذي يساوي (٠.٩٠) على

الأقل أمر مرغوب فيه، وذلك لأن قدرًا كبيرًا من القرارات يعتمد على النتيجة الدقيقة التي أدلى بها الفرد على الاختبار. فمثلاً الأطفال ذوي نسبة الذكاء المنخفض دون ٧٠ يتم وضعهم في فصول خاصة، وفي هذه الحالة يوجد فرقًا كبيرًا بين ما إذا كان الطفل لديه معدل ذكاء يساوي ٦٥ أو ٧٥ في اختبار معين، فالطفل الذي لديه معدل ذكاء يساوي ٦٥ يتم أخذ القرار بوضعه في فصول خاصة، أما الذي حصل على معدل ذكاء ٧٥ فيتم وضعه في فصول العاديين.

وفي المقاييس الجديدة يكون معامل ألفا الذي يساوي (٠.٧٠) مقبولاً، ولكن تختلف آراء المؤلفين المتعلقة بالقيمة المثالية لمعامل ألفا كمقياس للاتساق الداخلي للاختبار، ويوصي العديد من الكتاب أن القيمة المثالية لمعامل ألفا ينبغي ألاَّ تقل عن (٠.٩٠)، وأن القيمة المثالية تساوي (٠.٩٥). ولكن يعتبر بعض المؤلفين أن قيم ألفا الأكبر من (٠.٩٠) تشير إلى وجود تكرار بأداة القياس وأنه ينبغي تقصير طولها وذلك بحذف بعض عباراتها (DeVon et al., 2007: 160).

ونظرًا لأن قيمة معامل الثبات تمتد من (صفر) إلى (١)، فإن القيمة (صفر) تشير إلى عدم الثبات بينما تشير القيمة (١) إلى الثبات التام. ونظرًا لأن معظم الاختبارات تتضمن بعض الأخطاء فإن معامل الثبات لن يصل إلى القيمة (١)، وعمومًا إذا كان معامل الثبات أعلى من (٠.٨٠) فإنه يمكن القول بأن الثبات جيد جدًا، أما إذا كان معامل الثبات أقل من (٠.٥٠) فإنه لن ننظر إلى الاختبار على أنه اختبار ثابت

http://fcit.usf.edu/assessment/basic/basicc.html.

كيف يمكن جعل الاختبار أكثر ثباتًا؟

يمكن تحسين ثبات الاختبار أو أداة القياس من خلال عدة أمور منها:

- كتابة البنود بشكل واضح، وكتابة تعليمات الاختبار بصورة يسهل فهمها.
- تدريب المُقيِّمون raters على قواعد واضحة لتقدير الدرجات.
- الطريقة الرئيسية لجعل الاختبارات هو أكثر ثباتًا جعلها أطول عن طريق إضافة المزيد من البنود، إلاَّ أن الاختبار الأطول أكثر احتمالاً لجعل المستجيبين يشعرون بالملل والتعب، الأمر الذي يؤدي خفض الاتساق في الاستجابة الدقيقة (Drost, 2011: 113).
- ليس عن طريق زيادة طولة وذلك بإضافة عبارات أخرى ولكن عن طريق نقص طول الاختبار وذلك بحذف العبارات ذات الارتباطات الضعيفة مع العبارات الأخرى، أي حذف العبارات السيئة التي ليس لها تأثير كبير على صدق الاختبار.

- عن طريق زيادة عدم تجانس العينة sample heterogeneity.
- تحسين ظروف تطبيق الاختبار (Sechrest, 1984: 49).
- وفي ضوء نظرية الاستجابة للمفردة يمكن تحسين الثبات من خلال طريقة القياس النفسي التي تسمى تحليل البند item analysis والتي تعتبر من الطرق الأكثر فعالية لزيادة الثبات، وتتكون هذه الطريقة من مؤشرين هما: صعوبات المفردة item difficulties وتمييز المفردة item discrimination ، حيث يتضمن مؤشر التمييز حساب الارتباطات بين البنود ومجموع درجات بنود الاختبار بأكمله، ويتم استبدال البنود الصعبة جدًا أو السهلة جدًا، وتلك البنود التي يقترب معامل تمييزها من الصفر أو ذات معامل التمييز السلبي وذلك ببنود أفضل الأمر الذي يؤدي إلى زيادة ثبات المقياس أو الاختبار.

http://en.wikipedia.org/wiki/Reliability_(psychometrics)

ثانيًا: الصـــدق: Validity

يتعلق الصدق بدقة أداة القياس وما إذا كانت فعلاً تقيس ما أعدت لقياسه (Golafshani, 2003: 599). أي أن الصدق يشير إلى مدى قياس الاختبار أو المقياس بدقة البناء الذي وُضع لقياسه، ويحدد الصدق مغزى وفائدة الاستدلالات المحددة التي يمكن الحصول عليها من درجات الاختبار. فالصدق يتعلق بما إذا كان الاختبار فعلا يقيس السمة أو السلوك الذي وُضع لقياسه، ومسألة صدق الاختبار من المسائل المهمة إلا أنها معقدة، فصدق أي اختبار نفسي لا يمكن أن يكون مطلقًا فهو نسبي بسبب عدم وجود معيار ذهبي للصدق في العلوم النفسية (Hilsenroth et al., 2004: 7).

فالصدق يتعلق بالمغزى من مكونات البحث، فعندما يقوم الباحثون بقياس السلوكيات فإنهم يكونوا مهتمين بما إذا كانوا يقيسون ما يهدفون قياسه، وقد يسألون عدة أسئلة منها: هل اختبار الذكاء يقيس الذكاء فعلاً؟ هل اختبار الدافعية للتعلم يتنبأ فعلاً بالتحصيل الدراسي؟ هذه الأسئلة وغيرها تتعلق بصدق أدوات القياس، على الرغم من أنه لا يمكن الإجابة عنها باليقين التام (Drost, 2011: 114).

مما سبق يتضح أنه يقصد بصدق أداة القياس أن تقيس البناء (أو المفهوم أو السمة) التي أُعدت لقياسه، وألاّ تقيس غيره أو تقيس بناء أو مفهوم أو سمة بجانب ما أُعدت لقياسه.

مكونات (أدلة) الصدق

المتأمل للتراث السيكومتري يلاحظ أن كثير من الكتابات تعرض مفاهيم الصدق على أنها أنواعًا للصدق، وفي الحقيقة أن **الصدق مفهوم واحد** ليس له أنواع بل – إن جاز التعبير – له عدة مكونات أو جوانب تتكامل مع بعضها البعض لتوفر أدلة متكاملة عن هذا المفهوم الوحيد وهو الصدق.

حيث يشير (Sechrest, 1984: 35) إلى أنه تم عرض عدة مفاهيم للصدق تحت ما يسمى بأنواع الصدق type of validity كما لو كانت مفاهيم منفصلة، ولكن حديثًا ينظر إلى هذا المفهوم على أنه مفهوم واحد single concept وأن هذه الأنواع تعتبر كمكونات للصدق وتقدم أنواعًا من الأدلة التي يمكن أن يستشهد بها لصالح الصدق.

وقد تباينت الكتابات السيكومترية في تصنيف ما يسمي أنواع أو مكونات أو أدلة أو جوانب أو طرق وصف الصدق، **ومن أفضل هذه التصنيفات** التصنيف الذي يتكون من أربع مكونات ينبغي أن يأخذها الباحثون في حسبانهم وهي: صدق الاستنتاج الإحصائي، والصدق الداخلي، والصدق الخارجي، وصدق البناء، وهذا التصنيف قدمه (Drost, 2011: 114-117) وهو باختصار كما يلي:

١) صدق الاستنتاج الإحصائي statistical conclusion validity

يختص صدق الاستنتاج الإحصائي بالعلاقة التي يتم اختبارها، حيث يشير صدق الاستنتاج الإحصائي إلى الاستنتاجات حول ما إذا كان من المعقول افتراض أن التغاير covariation يعطي مستوى ألفا المحدد التباينات التي تم الحصول عليها. وتوجد بعض التهديدات الرئيسية لصدق الاستنتاج الإحصائي مثل: انخفاض القوة الإحصائية، وانتهاك الافتراضات، وثبات المقاييس، وثبات المعالجة، وعدم التجانس العشوائي للمستجيبين.

٢) الصدق الداخلي internal validity

يتعلق الصدق الداخلي بصدق البحث نفسه، ويجيب عن أسئلة مثل: هل هناك علاقة سببية؟ هل هناك عوامل خارجية في البحث؟ فمثلاً: إذا أراد مدير لشركة بحث رضا الموظفين عن القيادة واستجاب ٥٠٪ فقط من الموظفين وجميعهم يحبون رئيسهم، فهل هذا المدير اختار عينة تمثل جميع الموظفين أم أنها عينة متحيزة؟ ومثال آخر: يتمثل في قيام أحد الباحثين بقياس الرضا الوظيفي قبل عيد العمال بعد تلقى الجميع منحة مالية، وأظهرت النتائج أن جميع الموظفين كانوا

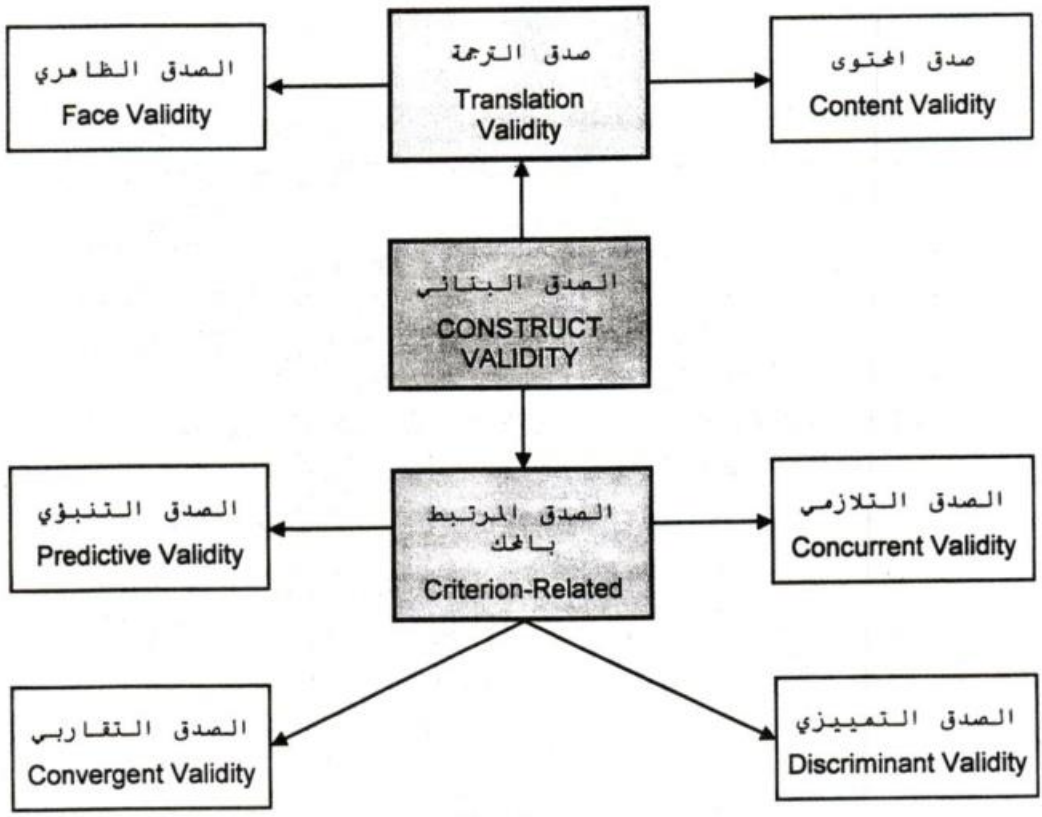
سعداء، فهل تشير النتائج إلى وجود رضا وظيفي حقيقي أم تظهر النتائج وجود تحيز؟ ومن هنا توجد العديد من المخاطر والتهديدات التي تهدد الصدق الداخلي لتصميم البحث منها: وقت وظروف تطبيق أدوات البحث، والنضج، وأداة القياس، وطريقة اختيار العينة، والتنافس، والإحباط.

٣) الصدق الخارجي external validity

يتعلق الصدق الخارجي بإمكانية تعميم النتائج على الأفراد الآخرين والمواقف والأوقات، فإذا كانت هناك علاقة سببية من البناء (أ) إلى البناء (ب)، يهتم الصدق الخارجي بأسئلة مثل: كيف يمكن تعميم هذه العلاقة عبر الأفراد أو المواقف أو الأوقات؟ وإلى أي مدي يستطيع الفرد تعميم ذلك؟ وهل العلاقة السببية التي لوحظت في مصنع معين يمكن الحصول عليها في مؤسسة عامة، أو في مؤسسة بيروقراطية، أو في قاعدة عسكرية؟.

٤) صدق البناء construct validity

يشير صدق البناء إلى كيفية ترجمة أو تحويل المفهوم أو الفكرة أو السلوك (أي البناء موضع الاهتمام) إلى واقع إجرائي أو عملي. فإذا كانت العلاقة سببية، ما سلوكيات أو بناءات السبب والنتيجة المشاركة في هذه العلاقة؟ وللتحقق من صدق البناء يتطلب التحقق من ستة مظاهر للصدق هي: الصدق الظاهري face validity ، وصدق المحتوى، الصدق التلازمي والتنبؤي، والصدق التقاربي والتمييزي وتنقسم هذه المظاهر الستة إلى فئتين هما: الصدق المرتبط بالمحك، وصدق الترجمة (الذي يركز على ما إذا كانت أداة القياس تعكس بدقة المعنى الحقيق للبناء وأنه تم ترجمة هذا البناء إجرائيًا بدقة). والشكل رقم (٢) يوضح مظاهر الصدق البنائي:



شكل (٢)

مظاهر الصدق البنائي

ومما هو جدير بالذكر أن مكونات أو مظاهر الصدق البنائي الموضحة بالشكل السابق يعتبرها بعض الباحثين أو المؤلفين على أنها أنواع منفصلة لمفهوم الصدق بوجه عام وليست مكونات لصدق البناء، ولكن المتأمل لتلك المكونات الموضحة بالشكل السابق يتضح له أنها مكونات لصدق البناء ويتفق مع المقصود بصدق البناء كما أشار إليه الباحثين.

فصدق البناء هو مدى قياس الاختبار للبناء أو السمة موضع الاهتمام بدقة وشمول، ويتعلق بفهم أبعاد وسمات البناء كما يقاس باختبار أو أداة قياس معينة، وهو أكثر من مجرد خطوة واحدة أو تقرير إحصائي بل هو عملية جمع المعلومات وبناء الأدلة التي تدعم دقة أداة القياس في وصف البناء الأكاديمي أو السلوكي أو المعرفي أو العاطفي، أو الاجتماعي (McGoeya et al, 2010: 110).

ويتعلق صدق البناء باستكشاف المعنى العلمي للبناء، وتتم هذه العملية من خلال ثلاثة عناصر رئيسية هي: (١) تحديد مجال المتغيرات الظاهرة المتعلقة بالبناء. (٢) من البحوث الإمبريقية والتحليلات الإحصائية، يتم تحديد إلى أي مدى تميل المتغيرات الظاهرة إلى قياس نفس الشيء، أو عدة أشياء مختلفة ويتم ذلك عادة عن طريق تطبيقات التحليل العاملي. (٣) إجراء الدراسات اللاحقة للفروق الفردية والتجارب المضبوطة لتحديد أي المقاييس المفترضة للبناء التي تعطي نتائج يمكن التنبؤ بها من الافتراضات النظرية المقبولة المتعلقة بالبناء (Nunnally, 1978: 98).

كما يشير صدق البناء إلى مدى استحواذ الاختبار على البناء النظري أو السمة المحددة، وتداخله مع بعض الجوانب الأخرى للصدق. وهذا يتطلب أن يرتكز الاختبار على إطار مفاهيمي أو نظرية تحدد بوضوح معنى البناء، وتفرده، وعلاقته بمتغيرات أخرى تقيس مجالات مماثلة. وفي القياس النفسي عادة يتم تقييم صدق البناء من خلال إعطاء مقاييس أخرى للسمة جنبًا إلى جنب مع المقياس الجديد المقترح لقياس تلك السمة ثم اختبار افتراض وجود علاقات بين المقياس الجديد وتلك المقاييس. فمثلاً: للتحقق من صدق البناء لمقياس جديد يقيس الإضطراب، يفترض أنه لكي يتمتع هذه المقياس بصدق البناء فأنه يرتبط إيجابيًا بمقياس صعوبات ضبط الانتباه ويرتبط سلبيًا بالإيثار (لأن الأدبيات تشير إلى ذلك)، كما يجب أن يظهر هذا المقياس الجديد ارتباطات ضعيفة مع البناءات الأخرى التي لا ترتبط نظريًا بالبناء موضع الاهتمام أو القياس. ومن المثير للاهتمام، أنه لا توجد طريقة واحدة لتحديد صدق البناء، فعادة ما يتم الجمع بين العديد من الأساليب والطرق المختلفة لتقديم صورة شاملة لصدق بناء المقياس. وبالإضافة إلى جانب طرق دراسة العلاقات الموضحة سابقًا، توجد طريقة تستخدم في كثير من الأحيان وهي طريقة التحليل العاملي، وفيها يُعطى الاختبار الجديد إلى مجموعة كبيرة من الأفراد (**وعدد من المشاركين يجب أن يكون على الأقل ١٠ أضعاف عدد البنود المدرجة في المقياس أو الاختبار**) ويتم إجراء التحليل العاملي للتعرف على البناءات أو الأبعاد التي تكمن وراء هذا المقياس أو الاختبار. وتوجد طريقة أخرى للتعرف على صدق البناء هو صدق التمييز discriminant validity ، وفيه يجب أن يكون المرتفعين على المقياس الجديد مرتفعين أيضًا على المقاييس التي تقيس ظواهر مشابهة للظاهرة التي يقيسها المقياس الجديد (Hilsenroth et al. 2004: 9).

وفيما يلي توضيح لمكونات الصدق البنائي (Hilsenroth et al. 2004: 8-9)، (Drost, 2011: 114-117):

١. الصدق الظاهري: Face validity

وهو أبسط أنواع الصدق، ويشير الصدق الظاهري إلى المدى الذي يكشف به البند غرض أو معنى الاختبار. على سبيل المثال، في البند التالي: (في الآونة الأخيرة لقد فكرت في قتل نفسي) له صدق ظاهري واضح كبند لقياس التفكير في الانتحار، مثل هذه البنود أو العبارات ذات الجوانب السلبية واضحة الصدق الظاهري أكثر عرضة للتلاعب من قبل المستجيبين، وذلك إما لإنكار أو إخفاء المشاكل أو الجوانب السلبية أو إلى التمارض أو المبالغة في حجم المشاكل. وبعض أدوات القياس النفسي مثل الاختبارات تفتقر إلى الصدق الظاهري إلاَّ أنها تظل تمتلك صدقًا عامًا، فالاختبارات أو البنود التي تقيس ما وُضعت لقياسه لكنها تفتقر إلى الصدق الظاهري يصعب على المستجيبين التلاعب في الاستجابة عليها، مثل العبارة (اعتقد في هذه اللحظة مجيء المسيح) التي ظهرت في الطبعة الأولى لقائمة مينسوتا للشخصية متعددة الأوجه MMPI (Hathaway & McKinley, 1943) التي تشبعت بمقياس الاكتئاب، لأن معظم عينة تقنين هذه القائمة كانت من المسيحيين الذين لا يعانون من الاكتئاب، حيث إن المسيحيين الذين يعانون من الاكتئاب يعتقدون أن المسيح لن يعود. ومن الواضح أن هذا البند ليس جيدًا للعديد من الأفراد من كل الأديان والمعتقدات الأخرى، وهكذا رغم افتقار هذا البند إلى الصدق الظاهري إلاَّ أنه قد يكون له بعض المزايا الجذابة، حيث إن البنود التي تتمتع بقدر من الصدق الظاهري قد تمثل على المدى الطويل اختبارًا أفضل.

٢. صدق المحتوى أو المضمون: content validity

يشير صدق محتوى الاختبار إلى كفاية أخذ عينات من محتوى البناء أو السمة التي يتم قياسها، ويجيب صدق المحتوى على أسئلة مثل: هل جميع جوانب هذا المفهوم أو السمة ممثلة ببنود مدرجة في الإختبار؟ نستخدم مثالاً على ذلك: اضطراب السلوك في الطفولة، فإذا كشفت أدبيات البحث عن جانبين رئيسيين لاضطراب السلوك في الطفولة هما: الانحراف والعدوان، فإن البنود المدرجة في الاختبار يجب قياس هذين الجانبين بنسبة متساوية نسبيا، ويعتمد بعض صانعي الاختبارات أيضا على الخبراء في هذا المجال، وهو ما يطلق عليهم المحكمين، حيث يقوم واضعي الاختبار باستنباط وسيلة لتلخيص أراء الخبراء المتعلقة بطبيعة السمة أو الظاهرة موضع القياس، ومن ثم يتم صياغة بنود الاختبار التي تعكس آراء الخبراء المتعلقة بالسمة، وينبغي أن تكون البنود التي تقيس السمة موضع الاهتمام بنسب متساوية في ضوء ما تكشف عنه أدبيات البحث أو ما يراه الخبراء حول تلك السمة. إلا أنه توجد بعض الحالات قد يصبح فيها عدد البنود

غير متوازنة، عندما تستخدم البرامج الإحصائية لحساب معامل ألفا كرونباخ ويقوم مُعد الاختبار بحذف كل البنود غير الثابتة عندئذ يحدث عدم التوازن في محتوى الاختبار، وهنا يضحي صانع الاختبار بصدق المحتوى في ضوء مذبحة معامل ألفا، وفي هذه الحالة، ربما يلجأ صانع الاختبار إلى إعادة صياغة البنود لجعلها أكثر ثباتًا.

٣. صدق المحك: Criterion validity

يشير صدق المحك (ويسمى أيضًا الصدق التنبؤي predictive أو التلازمي concurrent) مقارنة درجات الأفراد على الاختبار بدرجاتهم على محك خارجي، والمقياس الآخر (المحك الخارجي) يجب أن يكون مرتبط نظريًا مع البناء أو المفهوم موضع القياس، وأن يتم تقييم العلاقة بينهما عن طريق معامل الارتباط البسيط. ويُقسّم بعض علماء القياس النفسي صلاحية المحك إلى صدق تنبؤي وصدق تلازمي، ففي الصدق التنبؤي، يُعطى الاختبار الجديد لمجموعة من المشاركين الذين يتم تتبعهم خلال فترة زمنية معينة لمعرفة قدرة الاختبار الجديد على التنبؤ ببعض المتغيرات المهمة في وقت لاحق لدى هؤلاء المشاركين. فمثلاً: نجاح الطالب في الكلية يمكن التنبؤ به من خلال متوسط درجاته في الثانوية العامة، وللتأكد من وجود صدق تنبؤي للمقياس يجب أن يكون هناك علاقة كبيرة (معامل الارتباط لا يقل عن ٠.٥٠) بين المقياس الجديد والمعدلات اللاحقة للطلاب في الكلية. وهكذا، ففي الصدق التنبؤي يتم تطبيق الاختبار أو المقياس أولاً، ثم تؤخذ القياسات الأخرى لاحقًا، وبعد ذلك يتم حساب ارتباطات تلك القياسات بدرجات التطبيق الأصلي للاختبار أو المقياس.

وفي الصدق التلازمي، وهو الأكثر شيوعًا يتم تطبيق الاختبار أو المقياس المقترح على مجموعة من المشاركين وفي نفس الوقت يتم تطبيق أحد المقاييس ذات الصلة بالمقياس المقترح، وبعد ذلك يتم حساب معامل الارتباط بين درجات الأفراد المشاركين في المقياسين أو الاختبارين. إلا أن المشكلة هنا تكمن في أنه ليس من السهل الحصول على مقياس صادق يقيس الظاهرة أو السمة موضع القياس، ولذا يتم الاعتماد على مقاييس ذات صلة بالسمة موضع القياس.

ويتم التحقق من الصدق التقاربي من خلال حساب الارتباط بين مقاييس أو معالجات مختلفة لنفس البناء ويفترض أن تكون الارتباطات مرتفعة، أما الصدق التمييزي فيقاس عن طريق تحديد قدرة أداة القياس على التمييز بين المرتفعين والمنخفضين في المحك أي يكونوا متمايزين على أداة القياس، فمثلاً: عند حساب الصدق التمييزي لمقياس إضطراب السلوك يسجل الطلاب العدوانيين درجات أعلى من ضحاياهم على هذا المقياس، وكذلك يسجل الطلاب مرتفعي الذكاء

درجات تحصيل أعلى من منخفضي الذكاء عند حساب الصدق التمييزي لاختبار تحصيلي معين.

بعض طرق تحسين الصدق

١) التأكد على أن غايات وأهداف أداة القياس محددة بوضوح ومُعرَّفة إجرائيًا، وأن يتم تدوين توقعات عينة البحث.

٢) مطابقة مقياس التقييم أو التقدير بأهداف وغايات أداة القياس، ومراجعة أداة القياس من الخبراء والمتخصصين في المجال للحصول على ردود فعل طرف خارجي تجاه أداة القياس.

٣) عرض أداة القياس على العينة الاستطلاعية لتقييم صياغتها وتحديد جوانب الغموض وأي صعوبات أخرى.

٤) مقارنة أداة القياس بغيرها من أدوات القياس الأخرى أو البيانات التي قد تكون متاحة إذا كان ذلك ممكنا (Phelan & Julie Wren, 2006).

ثالثًا: العلاقة بين الثبات والصدق Relationship between reliability and validity

الثبات والصدق مفهومان غير متعارضين أو متنافيين، بل إنهما مرتبطان ارتباطًا وثيقًا. ولكي يكون الاختبار صادقًا يجب أن يكون ثابتًا أولاً؛ ومع ذلك، يمكن أن يكون الاختبار ثابتًا لكنه غير صادق. وبعبارة أخرى، الثبات شرط ضروري ولكنه ليس كافيًا للصدق reliability is necessary but not sufficient for validity. فمثلاً: قد يكون لدى الفرد مقياس للتوافق الوظيفي يتمتع بثبات عال جدًا عن طريق إعادة الاختبار، حيث تظل درجات الفرد هي نفسها على مدى فترة زمنية لمدة شهرين؛ ومع ذلك، فإن هذا المقياس قد لا يقيس التوافق الوظيفي بدقة، ولكنه يقيس الرضا الوظيفي، أي أنه غير صادق في قياس ما وُضع لقياسه. ومن هنا إذا كان المقياس لا يقيس ما يهدف إلى قياسه، حتى عندما تكون درجات الأفراد عليه ثابتة مع مرور الوقت، فإن هذه الدرجات تكون غير صادقة كوسيلة لاتخاذ قرار أو وضع خطة عمل بشأن المستجيب على هذا المقياس (McGoeya et al., 2010: 110).

وعلى الرغم من أن مفهومي الثبات والصدق يتم استخدامها جنبا إلى جنب وأحيانا بالتبادل، فإن كل مهما يعني أشياء مختلفة تماما. **فالقياس يمكن أن يكون ثابتًا ولكن ليس بالضرورة أن يكون صادقًا. ومع ذلك، ينبغي بالضرورة أولاً أن يكون القياس ثابتًا قبل أن يكون صادقًا؛ فالثبات في حد ذاته ليس هو الشرط الوحيد للصدق.** وكل من الثبات والصدق ضروري لتقديم قياسات دقيقة في الدراسات والبحوث. فالثبات هو درجة الحصول على نفس النتيجة عند تكرار التجربة،

أي أن الثبات يعني الاستنساخ reproducibility ، فالنتائج قد لا تكون صحيحة ولكنها يمكن أن تكون مستنسخة. أما الصدق فيشير إلى مدى صحة البيانات، ومدى كون نتائج أداة القياس حقيقة ذات معنى، أو أن نتائج أداة القياس يمكنها التنبؤ بالأداء في نوع آخر من الاختبارات. وتنشأ مشكلة الصدق بسبب أن القياس في العلوم الاجتماعية غير مباشر باستثناء حالات قليلة جدًا، وفي ظل هذه الظروف يكون الباحثون غير متأكدين تماما من أنهم يقيسون بدقة ما يزعمون قياسه. ومعظم الباحثين يعتقدون أن الاختبار الأعلى ثباتًا هو الأفضل، لكن الثبات هو أكثر من مجرد مفاضلته عن الصدق، فيمكن الحصول على مستويات أعلى من الثبات عن طريق زيادة عدد بنود أداة القياس، ولكن يتطلب الاختبار في هذه الحالة مزيدًا من وقت المستجيب (Moskowitz et al., 2005:161).

فالصدق والثبات مصطلحان أو مفهومان غالبا يتم استخدامهما بالتبادل عندما لا يكون الحديث يتعلق بأمور إحصائية، ولكن عند القراءة الإحصائية لاستخدام هذين المصطلحين، فإنهما يشيران إلى خصائص إحصائية أو طرق تجريبية مختلفة

http://web.cortland.edu/andersmd/STATS/ valid.html

ففي القياس النفسي يستخدم الثبات لوصف الاتساق العام للمقياس، ويُقال أن المقياس عال الثبات إذا أعطى نتائج مماثلة في ظروف متسقة، فمثلاً قياسات طول ووزن الأفراد غالبا ما تكون أكثر ثباتًا، والثبات لا يعني بالضرورة الصدق، فالمقياس الثابت الذي يقيس شيء ما باتساق ليس بالضرورة أن يقيس ما يُراد قياسه. فمثلاً يوجد العديد من الاختبارات الثابتة التي تقيس قدرات محددة، إلاّ أن جميعها ليس صالحًا للتنبؤ بالأداء الوظيفي. وانخفاض الثبات يؤثر على الصدق الكلي للاختبار، فالاختبار غير الثابت لا يمكن أن يكون صادقًا سواء لقياس سمات الفرد أو التنبؤ بدرجاته على المحك. في حين أن الاختبار الثابت يمكن أن يزودنا بمعلومات مفيدة عن الصدق (Reliability_(psychometrics)http://en.wikipedia.org/wiki/).

ويؤكد كثير من علماء القياس النفسي على أن: الثبات يعتبر مُكوّن ضروري للصدق إلا أنه ليس كافيًا (DeVon et al., 2007: 160). ولهذا يتطلب الصدق أن يكون الاختبار ثابتًا، ولكن يمكن أن تكون الأداة ثابتة إلاّ أنها تكون غير صادقة (Kimberlin & Winterstein, 2008: 2278).

كما يشير (McGoeya et al., 2010: 111) إلى أن الثبات شرط ضروري للصدق لكنه غير كاف Reliability is a necessary, but not sufficient condition of

validity. فلكي تكون أداة القياس صادقة يجب أولاً أن تكون ثابتة، ولكن ليس كل أداة ثابتة تكون صادقة. وبالتأكيد، فإن قضيتي الثبات والصدق تشكلان حجر الزاوية في الخصائص السيكومترية لأدوات القياس.

ويؤكد (Hilsenroth et al., 2004: 7-8) على أن **ثبات الاختبار يجب التحقق منه قبل التحقق من صدقه**، فالاختبار أو المقياس يمكن أن يكون ثابتًا (أي مستقر وقابل لتكرار الحصول على نفس الدرجات) إلاّ أنه قد يكون غير صادق، ولكن الاختبار لا يمكن أن يكون صادقًا دون أن يكون ثابتًا في المقام الأول.

وقد يشعر الباحث المبتدئ بشيء من الغموض فيما يتعلق بفكرة أن المقياس الثابت ليس بالضرورة أن يكون مقياسًا صادقًا. فقد أوضح بولين Bollen (١٩٩٠) أن الثبات هو ذلك الجزء من القياس الذي يكون خاليًا من الخطأ العشوائي البحت، وأنه لا يوجد في وصف الثبات ما يتطلب أن يكون هذا القياس صادقًا. فمن الممكن أن يكون المقياس ثابت جدًا إلاّ أنه غير صادق. وهكذا، فالثبات هو شرط ضروري ولكنه غير كاف للصدق (in: Drost, 2011: 107).

ونظرًا لأن الصدق هو الدرجة التي تقيس بها الأداة ما تفترض قياسه، فإذا كانت الأداة ليست ثابتة مع مرور الوقت، فإنها لا يمكن أن تكون صادقة، لأن النتائج يمكن أن تختلف تبعا لوقت تطبيق أداة القياس، ويمكن أن تكون أداة القياس ليست ثابته وليست صادقة، أو تكون ثابتة وليست صادقة، أو تكون ثابتة وصادقة. ومع ذلك، يجب أن تكون أداة القياس ثابتة من أجل أن تكون صادقة

http://web.unthsc.edu/info/200160/center_for_learning_and_developm
ent/2256/assessment-reliability_and_validity .

أما العلاقة الإحصائية بين الثبات والصدق فقد عبر عنها (Sechrest, 1984: 47-48) بالمعادلة التالية: $\frac{r_{xy}}{\sqrt{r_{xx}}} = \text{constant}$ *أي أن النسبة بين معامل الصدق (r_{xy}) إلى الجذر التربيعي لمعامل الثبات ($\sqrt{r_{xx}}$) نسبة ثابتة ولا تعتمد على طول الاختبار. وأقصى قيمة لثابت المعادلة السابقة هو الواحد الصحيح، وذلك عندما يساوي معامل الصدق الجذر التربيعي لمعامل الثبات. ونظرًا لأن معامل الثبات يمثل نسبة التباين الثابتة في أداة القياس، أي التباين الذي يمكن تفسيره، ومن هنا فالحد الأعلى لمعامل الصدق يساوي الجذر التربيعي لمعامل الثبات. فالاختبار الذي معامل ثباته يساوي (٠.٥٠) يكون له معامل صدق يساوي (٠.٧١) تقريبًا، ومن هنا فالمقياس*

الذي يتمتع بصدق جيد يكون ثابتًا.

والشكل التالي يوضح العلاقة بين الثبات والصدق (Shuttleworth, 2013):

شكل (٣) العلاقة بين الثبات والصدق

الخلاصة والاستنتاجات:

مما سبق يتضح ما يلي:

١) أن التراث السيكومتري أكد على أن الثبات شرط ضروري للصدق لكنه غير كاف، وذلك لأن كل صادق ثابت، ولكن العكس ليس بالضرورة أن يكون صحيحًا، فليس بالضرورة أن يكون كل ثابت صادق، فقد يكون الاختبار ثابتًا ولكنه لا يقيس السمة أو الصفة التي وُضع لقياسها. فمثلاً إذا كان ميزان معين يقدر وزن الفرد بزيادة ٣ كجم كل مرة، فإن هذا الميزان ثابت لأنه يقرأ نفس الوزن باتساق كل مرة، إلا أنه غير صادق لأنه يضيف ٣ كجم إلى الوزن الحقيقي الفرد.

٢) أنه ينبغي حساب ثبات الاختبار أولاً ويأتي بعد ذلك حساب صدق الاختبار، لأنه إذا تم حساب الصدق أولاً فلا داعي لحساب الثبات لأن كل صادق ثابت، ونظرًا لأنه ليس بالضرورة أن يكون كل ثابت صادق، فإذا تم حساب الثبات قبل الصدق تعتبر هذه العملية بمثابة **الفلتر الأول** لتنقية أداة القياس من البنود أو العبارات غير الثابتة، ويأتي بعد ذلك الصدق الذي يعتبر بمثابة **الفلتر الثاني** لتنقية أداة القياس من البنود أو العبارات غير الصادقة.

٣) نظرًا لأن الثبات شرط ضروري للصدق لكنه غير كاف، لذا فمن المنطقي أن نحسب الثبات قبل حساب الصدق؛ حتى نستوفي هذا الشرط أولاً قبل حساب الصدق.

٤) ينبغي حساب الثبات بأكثر من طريقة وعدم الاكتفاء بطريقة واحدة؛ لأن طرق حساب الثبات تختلف في الكشف عن مصدر الخطأ في الاختبار، ولكل طريقة مميزاتها وعيوبها. كما ينبغي حساب الصدق بطرق تغطي جميع مكونات الصدق كلما كان ذلك ممكنًا.

٥) أنه نظرًا لأهمية ثبات أداة القياس في الحصول على بيانات ثابتة ومتسقة، ونظرًا لأنه كلما انخفضت أخطاء القياس ارتفع الثبات، وأنه يمكن تحسين ثبات الاختبار أو أداة القياس من خلال عدة أمور تم توضيحها في هذا البحث، لذا ينبغي على الباحثين بوجه عام والمهتمين بالقياس النفسي بوجه خاص أن يكونوا على وعي بطرق تحسين الثبات، وأن يكونوا أكثر حرصًا على الاستفادة من هذه الطرق في تحسين ثبات أدوات القياس التي يستخدمونها في بحوثهم.

٦) وأخيرًا، نظرًا لأهمية صدق أدوات القياس في الحصول على قياسات دقيقة وصادقة للبناء أو المفهوم أو الظاهرة موضع القياس، ونظرًا لأنه يمكن تحسين صدق أداة القياس من خلال عدة عوامل، لذا يجب على الباحثين التأكد من صدق أدوات القياس التي يستخدمونها في بحوثهم، وأن يكونوا على وعي تام بأن استخدام أداة قياس غير صادقة يعني لا قياس.

References

1. DeVon, H. A.; Block, M. E.; Moyle-Wright, P.; Ernst, D. M.; Hayden, S J.; Lazzara, D. J.; Savoy, S. M.; Kostas-Polston, E. (2007). A Psychometric Toolbox for Testing Validity and Reliability. *Journal of Nursing Scholarship, 39* (2), 155-164.
2. Drost, E. A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives, 38*(1).105-123.
3. Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report. 8* (4), 597-607.
4. *Hammersley, M. (1987). Some notes on the terms validity and reliability.* British Educational Research Journal. 13*(1), 73-78.*
5. Hilsenroth, M. J.; Segal. D. L. & Hersen M. (2004). *Comprehensive handbook of psychological assessment.* New Jersey, Hoboken: John Wiley & Sons, Inc. .
6. Kimberlin, C. L. & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *Am J Health-Syst Pharm. 65*, 2276-2284.
7. McGoeya, K. E.; Cowanb, R. J.; Rumrillb, P. P. & LaVogue, C. (2010). Understanding the psychometric properties of reliability and validity in assessment. *Work, 36*, 105–111.
8. Moskowitz. H. R.; Porretta, S. & Silcher, M. (2005). *Concept research in food product design and development.* Iowa: Blackwell Publishing.
9. Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

10. Phelan, C. & Julie Wren, J. (2006). Exploring Reliability in academic assessment. (Available online: www.uni.edu/chfasoa /reliabilityandvalidity .htm).
11. *Sechrest, L. (1984). Reliability and validity.* Research Methods in Clinical Psychology, *24-53.*
12. Shuttleworth, M. (2013). Definition of Reliability. http://explorable.com/definition-of-reliability (18/11/2013).

Electronic websites

13. http://en.wikipedia.org/wiki/Reliability_(psychometrics) (1/12/2013).
14. http://web.cortland.edu/andersmd/STATS/valid.html (18/11/2013).
15. http://web.unthsc.edu/info/200160/center_for_learning_and_development/2256/assessment-reliability_and_validity. (20/11/2013.)
16. http://fcit.usf.edu/assessment/basic/basicc.html (25/11/2013).
17. http://writing.colostate.edu/guides/page.cfm?pageid=1388 (30/11/2013).