

مقدمة عامة إلى ثبات وصدق الاختبارات النفسية  
**An Introduction to the Reliability and Validity**

مقرر: علم النفس العيادي (471 نفس)

د. احمد الشايع

## مقدمة عامة إلى ثبات وصدق الإختبارات النفسية

### An Introduction to the Reliability and Validity

مقرر: علم النفس العيادي 471 نفس  
د. احمد الشايع

#### 1. اعتبارات مشتركة بين المفهومين

- كلاهما من المواصفات الأساسية للإختبارات النفسية.
- لا يغني أحدهما عن الآخر.
- الثبات شرط ضروري للصدق إلا إنه غير كافٍ لتحقيقه.
- الإختبار الثابت ليس صادق بالضرورة.

#### 2. الثبات Reliability:

##### 1.2. تعريفه و معناه:

يشير الثبات حسب المفهوم الحديث إلى مدى إتساق الدرجات على المقياس، وتجانسها وعدم إضطرابها؛ وعكسه هو عدم الإتساق أي عدم وجود نظام يحكم الدرجات التي يحصل عليها الأفراد على المقياس. وفي المفهوم التقليدي له يشير الثبات إلى عدم إختلاف الدرجات التي يحصل عليها الأفراد على المقياس إذا أُعيد تطبيقه عليهم مرةً أخرى.

وحسب النظرية الكلاسيكية للقياس فإن الدرجة على أي إختبار نفسي هي حصيلة نوعين من التباين، هما تباين الصواب True variance وتباين الخطأ Error variance (الدرجة = جزء صواب + جزء خطأ). وكلما إرتفع تباين الصواب وقل تباين الخطأ، إرتفعت بالتالي قيمة الثبات المستخرجة، والعكس صحيح. ويختص تباين الصواب بالمقدار الحقيقي للسمة المقاسة، بينما يشير تباين الخطأ إلى جملة من العوامل الأخرى الغير معروفة إجمالاً، والتي تؤثر في الدرجة على الإختبار دون أن تكون خاصة بالسمة المقاسة.

## 2.2. أنواع الثبات:

نظراً لأن الثبات يختص بتقدير قيمة التباين الحقيقي (الصواب) عن طريق تحديد مقدار تباين الخطأ و عزله، فتنوع طرق التحقق من الثبات بتنوع مصادر تباين الخطأ المُحتملة في المقاييس على إختلاف أنواعها. فكل نوع من أنواع الثبات يختص بتقدير نوع معين من مصادر الخطأ. ومن أنواع الثبات مايلي:

### 1.2.2. ثبات الإتساق عبر الزمن Temporal stability (ثبات الإعادة Test Retest):

ويختص هذا النوع بتقدير قيمة تباين الخطأ الناتج عن إختلاف الفترة الزمنية. ويُستخرج عن طريق تطبيق الإختبار على مجموعة من الأفراد وإعادة تطبيقه على نفس المجموعة بعد مرور فترة زمنية معينة، و احتساب قيمة معامل الارتباط بين الدرجات على المقياس في التطبيق الأول و الثاني. فقيمة معامل الارتباط المُستخرجه هي قيمة الثبات.

### 2.2.2. ثبات الإتساق الداخلي Internal Consistency :

ويختص بتقدير قيمة تباين الخطأ الناتج عن عدم تناسق فقرات الإختبار مع بعضها البعض، أي مدى التجانس بينها. فكلما أرتفع تجانس الفقرات زادت قيمة الثبات المُحتسب بهذه الطريقة. ومن الطرق المشهورة لتقدير هذا النوع من الثبات: مُعامل ألفا كرونباخ، ومعادلة جتمان، ومعادلة كودر ريتشاردسون ( 20 و 21)، وطريقة التجزئة النصفية Split half كمعادلة سبيرمان براون.

### 3.2.2. ثبات الصور المتكافئة Parallel-Forms Reliability

ويتم تقدير هذا النوع من الثبات عن طريق بناء نسخة مكافئة تماماً للإختبار، ومن ثم تطبيق كلا النسختين على نفس المجموعة من الأفراد وإحتساب قيمة معامل الارتباط بينهما.

### 4.2.2. ثبات المصححين (أو ثبات المُلاحظين) Inter-Rater or Inter-Observer Reliability

ويهتم بتقدير قيمة تباين الخطأ الناتج عن إختلاف المصححين (أو المُلاحظين) في حال استخدام الإختبارات التي تعتمد في إستخراج درجاتها على تقدير الفاحص. ومن أمثلة هذه الإختبارات: بعض أجزاء إختبارات

التفكير الإبتكاري، بعض أجزاء إختبارات الذكاء، وطرق الملاحظة المقتنة في أساليب القياس السلوكي، وأخيراً المقابلات التشخيصية شبه المقتنة. وفي هذا النوع يتم تصحيح الإختبار بواسطة عدة فاحصين، ومن ثم يتم إحتساب قيمة الإتفاق بينهم، وتكون قيمة المعامل المستخرج هي قيمة الثبات، وتشير إلى تقدير قيمة الخطأ الناتج عن إختلاف المصححين أو الملاحظين.

### 3.2. مالحد الأقصى لمعامل الثبات؟ وكم قيمة معامل الثبات المطلوبة؟

يتدرج معامل الثبات على إختلاف الطرق المستخدمة من صفر كحد أدنى إلى واحد صحيح. ولا يمكن لقيمه أن تأخذ الإشارة السلبية إطلاقاً. فمن الممكن لقيمة الثبات أن تكون مثلاً: 0.45 أو 0.78 أو 0.52 أو حتى 0.99، ولكنها لا يمكن أن تكون (-0.34) أو (0.78 -) مثلاً.

أما عن القيمة المطلوبة للثبات، فكلما أرتفعت القيمة كلما كان هذا أحسن بالطبع (لماذا؟). إلا أن هذا لايتوفر دائماً في الواقع، وعليه يمكن للباحث أن يحدد القيمة المطلوبة بناء على الغرض من تطبيق الإختبار. فإن كنا نهدف إلي إستخدام الإختبار لأغراض بحثية بحتة بإستخدام مجموعات بحثية (يستثنى من هذا التطبيق لأغراض البحث ولكن بإستخدام حالات فردية Single case reserach)، فيمكننا القبول بقيمة ثبات لا تقل عن 0.40 . ولكن للأغراض الإرشادية و العيادية التي تستخدم الإختبارات لأغراض جمع المعلومات وتقييم الأفراد فمن المفضل ألا تقل قيمة الثبات عن 0.80 كحد أدنى. أخيراً فيجب القول أنه لا توجد قاعدة متفق عليها بين الباحثين في هذا الخصوص، وإنما هناك معايير تحكمية Arbitrary أقترحها بعض المتخصصين، ومنهم كرونباخ .

كما يجب الإنتباه أن الغرض من تطبيق الإختبار يُحدد أيضاً نوع الثبات المطلوب و الأكثر أهمية من غيره. فعندما نستخدم إختبار ما لأغراض متابعة التقدم العلاجي عند مريض ما ( مراقبة مستوى تحسن المريض مع تقدم العلاج) فهذا يتضمن بالضرورة إعادة تطبيق نفس الإختبار على هذا المريض عدة مرات على فترات مختلفة. وهنا في هذه الحالة يكون الثبات بالإعادة (الإستقرار عبر الزمن) هو أكثر أنواع الثبات حاجة، مع أهمية بقية الأنواع. وفي مثال آخر لو كنا نستخدم مقابلة تشخيصية بغرض تصنيف شكاوى المريض (التشخيص)، فيكون ثبات الملاحظين هو الأكثر إلحاحاً و أهمية، مع أهمية بقية الأنواع، وهكذا.

### 3. الصدق Validity :

#### 1.3. تعريفه و معناه:

يشير المفهوم التقليدي للصدق إلى أن الإختبار الصادق هو الذي يقيس كل ما أُعد لقياسه، ولا يقيس غير ما أُعد لقياسه. إلا أن المفهوم الحديث يربط بين صدق الإختبار و تحقيقه لأغراض معينة فقط. إذن، وحسب المفهوم الحديث، فالصدق ليس صفة ملازمة للإختبار دائماً، وإنما هو محصور بإستخدامات معينة له. فقد نقول أن إختبار ( أ ) صادق في التنبؤ بالتحصيل الدراسي أو في تحديد المصابين بإضطرابات شخصية أو في التنبؤ بالأفراد الذين سيستفيدون من العلاج النفسي أو في المساعدة برفع دقة التشخيص بإضطرابات المزاج أو في تحديد شدة الأعراض لإضطراب نفسي ما، وهكذا. فإطلاق صفة الصدق على إختبار نفسي ما دون تحديد الإستخدامات التي هو صادقٌ فيها يعتبر ممارسة غير دقيقة.

فمن غير الدقيق أن نقول أن إختبار وكسلر لذكاء الأطفال صادق، ونتوقف. بل يجب أن نحدد الإستخدامات و الأغراض البحثية أو التطبيقية التي يختص بها هذا الصدق الذي نتحدث عنه. كأن نقول أنه صادق في التعرف على الأطفال المصابين بالتخلف العقلي، أو أنه صادق في التنبؤ بالنجاح المدرسي، أو أنه صادق في تحديد الأطفال المعرضين أكثر من غيرهم لخطر الإصابة المستقبلية بإضطراب ما.

وبناء على هذا المفهوم للصدق، يمكن تقرير أمرين:

الأول: أن الصدق ليس صفة ملازمة للإختبار تدور معه وإنما حل وكيفما أُستُخدم.

الثاني: أن الطرق المختلفة للصدق ماهي إلا مصادر لجمع الأدلة و الشواهد Evidence التي تدعم صدق الإختبار في تلبية الغرض و الإستخدام المرجو منه، أيأ كان هذا الغرض (كيف نحكم على صدق إختبارات المركز الوطني للقياس والتقويم التي يلزم الطلاب والطالبات بإجرائها قبل دخول الجامعة؟).

### 2.3. أنواع الصدق (طرق التحقق من الصدق):

#### 1.2.3. وحدة مفهوم الصدق:

تذهب الإتجاهات الحديثة في الصدق على أنه مفهوم موحد United Construct وكي. و ما الأنواع المختلفة له إلا أنواع مختلفة من الطرق التي تسهم نتائجها في النهاية في دعم أو رفض هذا المفهوم الكلي، و أن كل نوع من هذه الأنواع يختص بمصدر محدد من المعلومات التي نجمعها بغرض تقييم صدق إختبار ما في تحقيقه لغرض أو أغراض محددة. إذن، يمكن لنا تصور الصدق كمفهوم على أنه أناء كبير يتدفق إليه الماء من عدة جداول، وهذه الجداول تمثل الطرق المختلفة لجمع المعلومات. ومن متضمنات هذه النظرة القول بأن مختلف طرق التحقق من الصدق تساند بعضها البعض، وأن أيّاً من الأنواع لا يغني عن الآخر بأي حالٍ من الأحوال.

#### 2.2.3. صدق المحتوى Content validity :

ويختص بجمع الأدلة الخاصة بحسن تمثيل مكونات الإختبار للنطاق الذي تمثله السمة المقاسة. فإذا كنت تقيس سمة كالذكاء مثلاً، فيختص هذا النوع من الصدق بجمع الأدلة الخاصة بما اذا كان هذا الإختبار يقيس كل العناصر والمكونات التي يتضمنها الذكاء حسب النظرية التي تعتمدها أو التعريف الذي صمم الإختبار على أساسه. مثال آخر، إذا كنا نريد بحث صدق المحتوى لمقياس إكتتاب صمم لقياس الإكتتاب كما هو موجود في الدليل التشخيصي DSM، هنا كل ماعلينا عمله هو المقارنه بين محتوى الإختبار و بين المحتوى الموجود في الدليل، وجمع كل ماله علاقه بالتحقق من ذلك بشكل علمي.

وتُعد جداول المواصفات Table of Specifications الخاصة بالإختبارات التحصيلية من أوضح الأمثلة على صدق المحتوى. كما أن خطوات صدق المحكمين حيث يطلب فيها من مجموعة من الخبراء الحكم على فقرات الإختبار من حيث مدى إنتمائها أو عدم إنتمائها لنطاق السمة المقاسة (هل تقيسها أو لا ) مثلاً آخر على إجراءات جمع الأدلة الخاصة بصدق المحتوى.

### 3.2.3. الصدق المتعلق بالمحكات Criteria related validity :

وتنقسم الأدلة التي يتضمنها هذا النوع إلى قسمين: الصدق التلازمي Concurrent validity و الصدق التنبؤي Predictive validity . وتتميز الأدلة التي يختص بها هذا النوع بأنها تعتمد على محكات خارجية مستقلة - نسبياً على الأقل - عن الإختبار محل الدراسة. وهذه المحكات إما أن تكون حاضرة الآن فنكون نتحدث عن النوع الأول (الصدق التلازمي)، أو تكون مستقبلية فنكون عندئذ نتحدث عن النوع الثاني (الصدق التنبؤي).

ومن الأمثلة على النوع الأول، أن نحسب قيمة معامل الارتباط بين إختبار أعدّ بالأساس للمساعدة في تشخيص التوحد وبين تشخيص المختصين للتوحد، أو بينه وبين درجات إختبار آخر ثبت صدقه لهذا الغرض، فهنا نجمع أدله إرتباطية بهدف التحقق من صدق الإختبار في تحقيق هذا الإستخدام (تشخيص التوحد). ومن ذلك أيضاً إحتساب العلاقة بين درجات الطالبات على إختبار مقرر مقاييس الذكاء والشخصية الفصلي وبين درجاتهن على إختبار آخر يقيس نفس المحتوى.

أما في النوع الثاني فنحن نبغي التحقق من صدق الإختبار في التنبؤ بأداء المفحوص في المستقبل، وهذا أمر ملح في العمل الإرشادي أو العيادي. فقد نريد معرفة من من المرضى سيستفيد أكثر من جلسات الإرشاد؟ ومن من الأطفال ستظهر لديه أعراض اضطراب تشتت الإنتباه وفرط الحركة مستقبلاً كي أستهدفهم وعائلاتهم ببرامج الوقاية الأولية مثلاً.

ومن الأمثلة الحقيقية على هذا النوع من الصدق، قيام باحثين ببناء إختبار نفسي يهدف لتحديد خطر الإصابة المستقبلية بالاضطراب الوجداني ثنائي القطب. فقد قام هؤلاء بتطبيق إختبارهم الجديد على مجموعة واسعة من الأفراد، وبعد ذلك تتبعوهم لقراءة عشر سنوات، حيث كانوا يقابلونهم على فترات مختلفة ويقومون بإجراء فحوصات نفسية عليهم كل ستة أشهر، وبإنتهاء فترة المتابعة ( 10 سنوات) وجدوا أن مانسبته 75 % من الأفراد الذين نالوا درجات عالية على المقياس (30 فأعلى من 45) أصيبوا بالاضطراب، الأمر الذي يشكل دليلاً يدعم صدق الإختبار في تحديد مستوى الخطورة لهذا الإضطراب.

### 4.2.3. صدق المفهوم Construct validity:

ويسمى بالعربية بأسماء مختلفة منها: صدق البناء، صدق التكوين، صدق التكوين الفرضي. ويُنظر لهذا النوع على أنه المظلة الكبرى لجميع أشكال و أنواع الأدلة والمعلومات الخاصة بالصدق. و أنه ربما ينضوي تحته كل من صدق المحتوى وصدق المحكات، وربما حتى الأنواع المختلفة للثبات أيضاً. إلا أن هذه النظرة لا تتجاوز الجانب التنظيري الأكاديمي فحسب. أما في الممارسة، فمع الإقرار بأهميته و مركزيته، إلا أنه غالباً ما يُقدّم كمظلة لأنواع محددة من المعلومات التي تُجمع للتحقق من الصدق، ومُختلفة عن تلك التي يتضمنها صدق المحتوى أو صدق التعلق بمحكات.

بإختصار، يتضمن صدق المفهوم أمرين: الأول هو تحديد توقعات وفروض نظرية عن المفهوم الذي يقيسه الإختبار، والثاني التحقق من هذه الفروض علمياً. إذن فمن الضروري أن يتوفر لنا إطار نظري يحدد عناصر المفهوم وطبيعته وعلاقاته مع غيره من المفاهيم الأخرى و ماله علاقة بالتغيرات التي قد تطرأ عليه حالياً أو مستقبلياً، أو أي إفتراضات خاصة بالمفهوم المُقاس كما تحددها النظرية أو النظريات ذات العلاقة.

فلو أخذنا مفهوم القلق كحالة كما تُقدمه نظرية سبيلبرج مثلاً، فكي أفحص صدق المفهوم لإختبار يقيس هذا المفهوم حسب هذه النظرية، من الممكن أن أضع الفروض التالية أولاً:

أ. القلق كحالة يرتبط إيجابياً بالقلق كسمة.

ب. القلق كحالة موجود عند النساء أكثر من الرجال.

ج. يزداد القلق كحالة بشكل حاد قبيل المواقف التي تتضمن تهديداً للفرد.

د. يتذبذب مستوى القلق كحالة خلال اليوم الواحد بشكل أكثر حدة من تذبذب القلق كسمة.

وهذه الإفتراضات تم إشتقاقها من النظرية التي قدّمت هذا المفهوم أساساً. ويجب إختبار مدى إنطباقها على الإختبار الذي نريد التحقق من صدقه.

ومن الممكن أن يدخل ضمن إجراءات التحقق من صدق المفهوم العديد من الطرق الفرعية (أنواع مختلفة من المعلومات)، ومنها مثلاً لا حصراً:

▪ -التحليل العاملي.



- الفروق بين المراحل العمرية.
- أسلوب نمذجة العلاقات البنائية **Structural Equations Modelling** .
- الفروق بين الصفوف الدراسية / الفروق بين الجنسين.
- الصدق التقاربي **Convergent** والصدق التباعدي **Divergent** .
- مصفوفة الوسائل المختلفة والسمات المختلفة.