

المحاضرة الثانية عشر  
المرحلة الثانية  
مادة القياس والتقويم  
اعداد  
الاستاذ الدكتور  
علي سموم الفرطوسي  
2025/2024

## -الثبات:

عندما نستخدم اختبارًا من أجل الحصول على معلومات تساعدنا في اتخاذ قرار ما، فإننا نواجه مشكلة أساسية تتعلق باختيارنا للاختبار الذي يمكن أن يفيدنا في اتخاذ القرار المناسب والوصول من أجل ذلك إلى الحقيقة، وقد تكون هناك بدائل عديدة لاختبارات أو أدوات قياس أخرى يمكن استخدامها لأغراض اتخاذ القرار. ولكن، يخطر ببالنا سؤال جوهري:

- أي هذه الاختبارات أو أدوات القياس هو الأفضل في تقديم المعلومات المفيدة؟

- وما هي الأسس التي نستند إليها في اختيار أداة القياس؟

ولا شك أن أمورًا كثيرة يمكن أخذها بعين الاعتبار عند تقييم جودة أداة القياس، فالصدق والثبات والموضوعية والقابلية للاستعمال أمور يجب أن تؤخذ بعين الاعتبار خاصة وأن "الصدق يشير إلى الدرجة التي يمكن فيها لأداة القياس أن تقدم معلومات ذات صلة بالقرار الذي سيبنى عليها، بينما يشير الثبات إلى درجة الدقة أو الضبط والإحكام في عملية القياس بحيث تعطينا معاملات الثبات فكرة عن درجة الاتساق أو التوافق في نتيجة القياس عند تكراره. أما القابلية فإنها تعني عددًا من العوامل ترتبط بالاقتصاد في الكلفة والوقت والجهد والملاءمة والقابلية للتفسير. وكلها عوامل تؤخذ بالاعتبار فيما إذا كان استخدام اختبار ما أو أداة قياس ما على نطاق واسع أمرًا عمليًا". (ملحم 2005).

## الثبات Reliability:

الثبات هو صفة من الصفات التي يجب أن تتصف بها أداة القياس الجيدة، ويقصد بالثبات (ثبات القياس) أي كم تكون علامة اختبار ما متسقة وغير مختلفة من وقت لآخر. (دروزة 2001)

ولكي نوضح معنى الثبات بشكل مبسط دعونا نفترض الافتراض التالي:

قام أحد معلمي اللغة العربية بإجراء اختبار لطلبة الصف العاشر، وبعد أن قام بتصحيح الأوراق ورصد العلامات تخلص من الأوراق. ولكنه بعد ذلك فوجئ بالطلبة يطالبونه بمراجعة الاختبار! فماذا فعل؟ قرر أن يعيد الاختبار للطلبة. وبعد تصحيح الاختبار الثاني وضع جدولاً مقارنةً لعلامات الاختبارين الأول والثاني ورصد فيه العلامات ورتب الطلبة. فماذا كانت النتيجة؟

لمعرفة النتيجة علينا أن نطلع على الجدول الذي أعده المعلم وهو الجدول التالي:

علامات طلاب الصف العاشر لمادة اللغة العربية للاختبارين الأول والثاني:

### الأسماء ونتائج الاختبار

تطبيق الاختبار الثاني		تطبيق الاختبار الأول		الاسم
المرتبة	العلامة	المرتبة	العلامة	
1	95	1	93	سامي
2	89	2	85	سعيد
3	85	3	82	سعد

4	82	4	79	سالم
5	65	5	68	سليمان
6	64	6	63	سلام
7	60	7	58	إسلام
8	51	8	47	سلطان

فيذا قارنا بين درجتى الاختبار ماذا نلاحظ:

- علامات الطلبة جميعهم قد ارتفعت في المرة الثانية باستثناء علامة سليمان فقد انخفضت من 68 إلى 65.

- ارتفاع علامات الطلبة في المرة الثانية كان طفيفاً وهو يتراوح بين أربع علامات وعلامة واحدة.

- لم تتغير رتب الطلاب بل تميزت بالثبات في المرة الثانية.

- ارتفاع علامات الطلاب في المرة الثانية كان نتيجة أثر التدريب أو الخبرة التي مروا بها عند أدائهم للاختبار في المرة الأولى. أو نتيجة لمراجعتهم لمادة الاختبار بعد الاختبار الأول.

وقد عرف بعض الباحثين الثبات بطرق مختلفة منها:

- مدى الاتساق بين البيانات التي تجمع عن طريق إعادة تطبيق نفس المقاييس على نفس الأفراد أو الظواهر، وتحت نفس الظروف أو تحت ظروف متشابهة إلى أكبر قدر ممكن. (Gay, 1990)

- الثبات قد يعني الاستقرار Stability: بمعنى أنه لو كررت عمليات قياس الفرد الواحد لأظهرت شيئاً من الاستقرار. (ملحم 2005)

- والثبات قد يعني الموضوعية Objectivity: بمعنى أن الفرد يحصل على نفس الدرجة كائناً من كان الإحصائي الذي يطبق الاختبار. (ملحم 2005).

- ويعرف إحصائياً بأنه: نسبة التباين الحقيقي إلى التباين الكلي، أو مربع معامل الارتباط بين العلامات الحقيقية والعلامات الظاهرية، وبما أننا لا نعرف مطلقاً العلامات الحقيقية، فلا يمكن حساب الثبات بهذه الطريقة، وكل ما يتوفر لدينا هو العلامات الظاهرية، وبالتالي لا بد من الاستفادة منها بطريقة ما لتقدير الثبات أي الحصول على مؤشر إحصائي نحكم من خلاله على دقة القياس، ويسمى هذا المؤشر بمعامل الثبات ( Reliability Coefficient) (عودة 2002)

وقد وضعت د: دروزة بعض الملاحظات التي توضح مفهوم الثبات كما

يلي:

1- تشير كلمة الثبات إلى النتائج المتعلقة بوسيلة تقويمه وليس إلى الوسيلة التقويمية ذاتها. ووسيلة التقويم الواحدة تكون لها عدة درجات ثبات وذلك حسب المجموعة المتحننة والموقف الذي استعملت فيه، وهكذا يفضل استعمال كلمة ثبات نتائج أو علامات الاختبار على ثبات الاختبار ذاته.

2- علامات الاختبار ليست ثابتة بشكل مطلق، بل هي ثابتة (أو يمكن تعميمها) على فترات مختلفة من الزمن أو على عينات مختلفة من الأسئلة، أو على عدة مصححين أو مقدرين. ومن الممكن أن تكون علامات اختبار ما متسقة في ناحية معينة من هذه النواحي، وليست كذلك في ناحية أخرى، والنوع المناسب من الاتساق في ناحية معينة تمليه الحالة التي يجب أن تستعمل فيها النتائج. فلو أردنا معرفة ما سيكون عليه الأفراد في وقت ما في

المستقبل فإن ثبات الاختبار مهم جداً، في حين إذا أردنا قياس تغير القلق من لحظة لأخرى، فلا نحتاج إلى مقياس في درجة عالية من الثبات، وعندها قد نستعمل الاختبار على فترات متباعدة للحصول على المعلومات المرغوبة.

3- الثبات شرط ضروري ولكنه ليس كافياً للصدق. فالاختبار الذي يعطي نتائج غير ثابتة لا يمكن أن يعطي معلومات صادقة عن السلوك المراد قياسه، وفي المقابل فإن الاختبار ذا الدرجة العالية من الثبات قد لا يقيس الشيء الصحيح (الذي يفترض أن يقيسه)، وهكذا فإن الثبات المنخفض يؤدي إلى انخفاض درجة الصدق التي نحصل عليها، لكن الثبات العالي لا يضمن درجة من الصدق عالية.

4- الثبات ذو طبيعة إحصائية، إذ علينا أن نجري الاختبار على مجموعة من الناس مرة أو عدة مرات في ظروف متشابهة، ثم يحسب ثبات النتائج. وهذا الثبات يعبر عنه إما بواسطة التغيرات في رتبة الأشخاص المتينة بالنسبة للمجموعة، أو عن طريق التغير الذي نتوقعه في علامة فرد محدد. فالثبات من النوع الأول يسمى معامل ارتباط الثبات (Reliability Coefficient) والثبات من النوع الثاني يسمى الخط المعياري للمقياس (دروزة 2001).

## طرق حساب معامل الثبات:

اختلفت طرق حساب معامل الثبات للقياس وذلك بناء على نوع القياس والغرض منه ومدى الخطأ المعياري لكل قياس. وقد ظهرت عدة طرق لحساب معامل الثبات ومن هذه الطرق:

### 1- طريقة الإعادة (Test- Retest) Methode

تقدير الثبات بهذه الطريقة، يعطى الاختبار مرتين للمجموعة نفسها من الطلاب مع فترة زمنية معينة بين المرتين، ثم يحسب معامل الارتباط بين علامتي الاختبار للمجموعة في المرتين، ومعامل الارتباط هذا هو مقياس للثبات، إذ يبين لنا كم كانت النتائج مستقرة خلال فترة زمنية معينة، فإذا كانت النتائج ثابتة ومستقرة بشكل عالٍ فإن الطلاب الذين سجلوا علامات عالية في الإجراء الأول من الاختبار، يميلون إلى أن يسجلوا علامة عالية في الإجراء الثاني، أما بقية الطلاب فيبقون في مواقعهم النسبية في الإجراءين. (دروزة 2001).

أما عيوب هذه الطريقة فهي:

- أ- عند إعادة تطبيق الاختبار سيتذكر الطلبة إجابات بعض الأسئلة وهذا يزيد من ثبات النتائج.
- ب- إذا كانت الفترة قصيرة بين الاختبار وإعادة الاختبار، فإن الذاكرة تلعب دورها وهذا أيضا يرفع من معامل الثبات.

ج- يَألف التلاميذ الاختبار، وتصبح لديهم خبرة فيه وفكرة عنه، فترتفع علامتهم عند الاختبار فيقل معامل الثبات. (أبو لبدة، 1985).

وتعتمد هنا الفترة الزمنية المطلوبة بين الاختبارات على الاستعمال الذي نريده من هذه النتائج، التنبؤ بعلامات طالب في الجامعة بناء على علامات اختبار في الصف السابع. (دروزة 2001)

ويستخرج معامل الثبات في هذه الطريقة بحساب معامل الارتباط بين درجات التطبيق الأول والثاني لنفس الأفراد، وكلما اقترب المعامل من الواحد الصحيح دل ذلك على ارتفاع ثبات المقياس والعكس صحيح.

## 2- طريقة الصور المتكافئة Equal Forms Methode

تعتبر الصور المتكافئة للاختبار نماذج بنيت طبقاً لمواصفات واحدة، ولكنها تألفت من عينات مستقلة، من منطقة سلوك محددة. وعلى هذا فإن اختبارين متكافئين للقراءة، يجب أن يتضمنا أسئلة لها الصعوبة نفسها، ويسأل فيها النوع نفسه من الأسئلة. إذا كان لدينا صورتان من الاختبار، فيمكننا أن نطبق إحدى الصورتين ثم نتبعها بالأخرى.

وبحساب الارتباط بين الصورتين، نحصل على معامل مناسب للثبات.

وعلى الرغم من أن هذه الطريقة تقدم أساساً سليماً جداً لتقدير الدقة في اختبار نفسي أو تربوي، إلا أنها تثير عدداً من المشكلات العلمية، أهمها:

- أ- أنها تتطلب توافر صورتين متكافئتين تماماً للاختبار.
- ب- أنها تحتاج إلى توافر وقت يسمح باختبار كل فرد مرتين.

والطريقة التي تعتمد على استخراج الارتباط بين صورتين متكافئتين تطبقان في المادة بفواصل زمني يمتد إلى عدة أيام أو عدة أسابيع، تمثل الطريقة المفضلة في تقدير الثبات.

وتستعمل هذه الطريقة بعد مرور فترة من الزمن بين إجراء شكلي الاختبار. في هذه الظروف يعطينا معامل الثبات مقياسًا للاستقرار (Stability & Equivalence) ويعتبر هذا المقياس أفضل مقياس للثبات؛ لأنه يشمل:

- جميع مصادر التباين في علامات الاختبار.

- ثبات خاصية التلميذ المراد قياسها.

- يمثل عينة الاختبار كلها.

### 3- طريقة الثبات النصفى Split- Half Method

قد يكون من الصعب على الباحث أن يطبق اختبارين متكافئين على التلاميذ، أو قد يتعذر عليه فحص الطلبة مرتين في الاختبار نفسه. لذلك يتم اللجوء إلى تقسيم الاختبار إلى نصفين، يفترض أنهما متكافئان. ومن الممكن تجميع نصفى الاختبار على أساس تفحص دقيق للمحتوى والصعوبة لكل فقرة، وبذل جهد منظم لموازنة المحتوى ومستوى الصعوبة في النصفين. ولكن الطريقة الأبسط، والتي يكثر استخدامها، هي وضع الأسئلة ذات الأرقام الفردية في النصف الأول والزوجية في النصف الثانى، وحساب الارتباطات بين النصفين، هو معامل الثبات في هذا الاختبار، ويلاحظ أن التجزئة مرتبطة فقط بتصحيح الاختبار، أما تطبيقه فيتم مرة واحدة.

ويقدر معامل ثبات الاختبار بهذه الطريقة بتطبيق معادلة سبيرمان براون كما يلي:

$$\text{معامل ثبات الاختبار} = \frac{\text{معامل الثبات النصفى } 2 \times \text{كله (بعد التصحيح)} + 1}{\text{معامل الثبات الحالي} + 1}$$

$$0.75 = \frac{1.20}{1.60} \times \frac{2}{1 + 0.60}$$

وما يميز هذا المعامل:

يدلنا إذا كان عاليًا على تكافؤ نصفى الاختبار وبالتالي كفاية العينة من حيث تمثيلها للمحتوى.

لا يعطينا شيئاً عن التغيرات التي تطرأ على الفرد من وقت لآخر. (دروزة 2001)

#### 4- طريقة كودر - ريتشاردسون (Kuder-Richardson)

في عام 1937 نشر كل من كودر وريتشاردسون مقالة في إحدى المجلات العلمية تحت عنوان نظريات حساب معاملات الثبات للاختبارات، وفي هذه المقالة ذكرا عددا من الصيغ المستخدمة لحساب معاملات الثبات للاختبارات كان من أشهرها على الإطلاق الصيغتان المعروفتان 20، 21 وتعتمد المعادلة

20 على نسبة أولئك الذين ينجحون في كل فقرة من فقرات الاختبار وعلى الانحراف المعياري لمجموع العلامات، وحساب هذه المعادلة مرهق نوعاً ما، إلا إذا توفرت معلومات بخصوص نسبة الذين اجتازوا كل فقرة من الاختبار ولكن النتيجة تساوي معدل مجموع معاملات الارتباط الممكنة لاختبار التكافؤ النصفى، أما المعادلة 21 فهي أسهل ولو أنها غير دقيقة ويمكن تطبيقها على نتائج أي اختبار بعد أن يصحح ويعرف عدد الإجابات الصحيحة، ويعبر عن المعادلة كما يلي:

$$M (K-M-1)$$

$$\text{Reliability Estimate (Kr 21)} = K$$

$$K - 1 Ks2$$

$$\text{Reliability Estimate (Kr21)} = k$$

والمعادلة عدد البنود × تباين درجات الاختبار - المتوسط (عدد البنود - المتوسط) <sup>بالعربية:</sup>

$$\frac{\text{(عدد البنود - 1)} \times \text{تباين درجات الاختبار}}$$

حيث إن تباين درجات الاختبار =  $\frac{\text{(عدد الأفراد} \times \text{مجموع المربعات}}{100}$  - مربع الدرجات)  
المتوسط <sup>مجموع</sup> =  $\frac{\text{ن}}$

S = الانحراف المعياري لعلامات الامتحان.

$K =$  عدد فقرات الامتحان.

$M =$  المتوسط المعياري لعلامات الامتحان.

وهذه المعادلة تعطي تقديرًا للثبات قريبًا من معادلة (20) وفي أغلب الأحيان يكون تقدير الثبات أقل دقة، ولعل السهولة التي تطبق بها المعادلة هي الفائدة المرجوة منها. (دروزة 2001)

من مميزات هذه الطريقة:

- أن فقرات الاختبار يفترض أن تكون متجانسة.
- أن هذه الطريقة لا تصلح للاختبارات التي يلعب فيها الزمن دورا أساسيًا.

## 5- طريقة اتفاق المقيمين Rates Agreement Method

تعتبر طريقة اتفاق المقيمين من الطرق المعروفة أيضا في حساب معامل ثبات الاختبار، وفي هذه الطريقة يحسب معامل ثبات الاختبار عن طريق حساب معامل الارتباط بين تقييم المقيمين للمجموعة نفسها من الأفراد، وتسمى هذه الطريقة أيضا باسم ثبات المصححين، وقد يلجأ إلى مثل هذه الطريقة حين يصعب استخدام الطرق الأخرى في حساب معامل الثبات.

## 6- طريقة الخطأ المعياري Standard Error Measurement

تعتبر طريقة الخطأ المعياري في القياس من الطرق المعروفة في حساب معامل الثبات. وفي هذه الطريقة يطبق الاختبار أكثر من مرة على العينة، ويحسب الخطأ المعياري لدرجات العينة فكلما كان الخطأ المعياري كبيراً كان معامل الثبات متدنياً والعكس صحيح.

### العوامل المؤثرة في قياس الثبات:

1- عدد أسئلة- فقرات- الاختبار: حيث ترتفع القيمة العددية لمعامل الثبات تبعاً لزيادة عدد أسئلة الاختبار، بمعنى أن معامل ثبات الاختبار الطويل أكبر من معامل ثبات نفس الاختبار عندما ينقص عدد أسئلته إلى النصف أو الثلث أو أية نسبة أخرى.

2- زمن الاختبار: فثبات الاختبار يتأثر بالزمن المحدد له. وتشير الدراسات التي أجراها كل من لينوجست وكوك Lindquist & Cook إلى أن معامل الثبات يزداد تبعاً لزيادة الزمن المستغرق للإجابة عن فقرات الاختبار

حتى يصل إلى الحد المناسب للإجابة عن فقرات الاختبار فيصل الثبات إلى نهايته العظمى ثم يقل الثبات تبعاً لذلك كلما زاد الزمن عن ذلك الحد.

3- تباين قدرات المفحوصين: معامل ثبات درجات الاختبار لمجموعة متجانسة من التلاميذ ينقص في قيمته العددية عن معامل ثبات درجات نفس الاختبار على مجموعة أخرى أقل تجانساً من المجموعة الأولى.

4- مستوى قدرات المفحوصين: يختلف مستوى قدرات المفحوصين من شعبة إلى شعبة أخرى فقد تتضمن الشعبة الواحدة تلاميذ يتميزون بمستوى مرتفع من الذكاء ومستوى تحصيل عالٍ، بينما تتميز شعبة أخرى بمستوى منخفض من التحصيل وذكاء أقل. وبالتالي، فإن ما يناسب تلاميذ الشعبة الأولى من اختبارات لا يناسب تلاميذ الشعبة الأخرى، مما يجعل درجات التلاميذ تنخفض في الشعبة الأخرى.

5- التخمين: فزيادة التخمين تنقص ثبات أي اختبار؛ لأن الإجابة التي تعتمد على التخمين في المرة الأولى لإجراء الاختبار، لا تعتمد على نفس هذا التخمين في المرات القادمة لإجراء ذلك الاختبار. وبذلك تضعف الصلة بين مرتي التطبيق للاختبار فتتخفف بذلك القيمة العددية لمعامل الثبات، وتختلف الاختبارات في مدى تأثيرها بالتخمين تبعاً لنوعها. وتعتبر الاختبارات المتعلقة بنوع الاختبار من متعدد أكثر أنواع الاختبارات تأثراً بالتخمين. وعلى الفاحصين أن يدققوا في اختيار أسئلتهم ويصيغوها بعبارات تقلل من فرص التخمين بها ليصلوا بذلك إلى مستويات عالية من الثبات.

6- صياغة أسئلة الاختبار: ذلك أن الأسئلة الغامضة، الخادعة، العاطفية، الطويلة تقلل من ثبات الاختبار بينما الأسئلة الواضحة في صياغة فقراتها الموضوعية القصيرة تزيد من ثبات الاختبار. وعلى الباحث أن يصيغ أسئلته بعبارات واضحة تمامًا تحقق له الثبات الحقيقي المرجو.

7- مدى صعوبة الاختبار: إذا تألف الاختبار من أسئلة سهلة جدًا أو صعبة جدًا. فإن درجات التلاميذ عليه تكون متقاربة، وتقلل من الثبات، وإذا ما أراد الباحث أن يزيد من ثبات اختباريه فإن عليه أن يؤلف الاختبار من أسئلة تتراوح في مدى صعوبتها بين (0.25 - 0.75) وأفضل الأسئلة كان مستوى صعوبته يساوي (0.50).

8- حالة التلميذ: ويتأثر الثبات بحالة التلميذ التي يكون فيها التلميذ (علميًا وصحياً ونفسياً) ومستوى تدريبه على الموقف الاختباري عند أدائه الاختبار. وكلما كانت حالة التلميذ العلمية والصحية والنفسية عالية كلما زاد من معامل ثبات الاختبار.