# الوحدة الخامسة

# الشروط العلمية للاختبار الجيد

# الشروط العلمية للاختبارات والمقاييس

الاختبار عبارة عن موقف «مقنن» يستدعى سلوكا (استجابة) محددا من المختبر يعبر عن شيء معين يراد قياسه، وهو أيضا «تمرين» أو «سؤال» يهدف إلى التعرف على «قدرة» أو «سمة» أو «خصوصية» لدى المختبر. وتبعا لتعريف بارو Barrow ومك جى McGee فإن الاختبار هو **«مجموعة من الأسئلة أو المشكلات أو التمرينات تعطى بهدف التأكد من معرفة الشخص أو قدراته أو استعداداته أو كفاءته»**.

من ثم ليس صحيحا أى «موقف» أو «سؤال» أو «تمرين» يصلح كاختبار، فالأمر يتطلب شروطا ومواصفات محددة ينبغى توافرها حتى يحق لنا القول أن هذا «اختبار» أو «مقياس». فمثلا على مستوى اختبارات الأداء erformance Tests فى التربية البدنية والرياضة لايصح القول أن أى «تمرين» يصلح «كاختبار»، بينما العكس صحيح تمامًا، فالاختبار عبارة عن «تمرين مقنن» والتقنين هنا يعنى شروطا ومواصفات معينة هى موضوع هذا الجزء من الكتاب.

وتوافر ظروف وأبعاد التقنين لا تعنى بالضرورة أن جميع الاختبارات أو «التمرينات المقننة» على نفس الدرجة من الجودة، فلكل اختبار وزن وقدر يختلف عن غيره، فهناك اختبار جيد وآخر متوسط من حيث الصلاحية والثقل العلمى، كما أن الأمر نسبى؛ بمعنى أن هناك اختبارات تعد صالحة وجيدة لمجتمع معين، فى حين أن الأمر ليس كذلك لو استخدم هذا الاختبار فى مجتمع آخر له مواصفات مختلفة **«فالاختبارات التى يتم بناؤها وتقنينها على عينات تمثل مجتمع المستفيدين تعد أصلح من غيرها التى تم بناؤها وتقنينها على عينات تمثل مجتمعا آخر مهما بلغت درجة التشابه من المجتمعين»**[1].

ما سبق لايعنى عدم وجود استفادة من الاختبارات الأجنبية، هذا ما لم نقصده على الإطلاق.. فمجال الاستفادة موجود من حيث المقارنة، أى مقارنة مستويات أبنائنا بالنسبة لأقرانهم فى الخارج، ويفضل فى ذلك الاختبارات الدولية، كما يمكن استخدام الاختبارات الأجنبية فى تقويم أبنائنا شريطة أن يتم التأكد من صلاحيتها ومناسبتها لمجتمعنا وفى ضوء معايير Norms جديدة توضع خصيصا لأبنائنا.. وإن ما نقصده أن الاختبار المحلى الذى بنى وقنن على أبنائنا أنسب من مثيله الأجنبى على جميع المستويات وفى جميع المستويات وفى جميع الظروف شريطة جودته بالطبع.

وهناك بعد آخر يجب عدم إهماله، فالاختبارات المتداولة بالفعل فى المجال ليست هى الأصلح فى جميع الأحوال والظروف، كما أن المعايير والمستويات تحتاج إلى تعديل مستمر بما يتناسب والتحسن الطارئ على قدرات الأفراد على هذه الاختبارات، من ثم فإعادة التقنين والتأكد من الصلاحية أمر وارد أيضا على ما هو متداول حاليا من اختبارات.. ويعبر عن ذلك باروBarrow ومك جى McGee بقولهما: **«ربما لا تكون بعض الاختبارات والمقاييس المتداولة الاستخدام هى أحسن ما يمكن استخدامه، فقد تكون مستنفدة للوقت والجهد فى إجرائها، أو غير محددة فى قياسها لعناصر معينة، وقد تكون وضعت لقياس**

**الأداء المتوسط فقط؛ لذلك يجب الاهتمام بإعادة تقويم مجموعات الاختبارات المستخدمة وتكوين مجموعات جديدة فى ميادين الأداء الحركى»**، وربما يكون هذا ما دعا جلاسو Glassow إلى إجراء دراسته المستفيضة التى تضمنت إعادة تقويم اختبار يونج وموسير Young & Moser لكرة السلة بهدف تقديم صورة أفضل لهذا الاختبار، وربما يكون هذا هو السبب أيضا الذى حفز جاكسون Jackson إلى إجراء دراسته لنقد عوامل القوة Muscular Strength التى توصل إليها فليشمان Fleishman باستخدام أسلوب التحليل العاملى Factor Analysis(*)، ناهيك عن هذه الأمثلة فهى عديدة وكثيرة.

مما سبق يتضح أن هناك صفات ومواصفات وشروطا يجب مراعاتها فى الاختبارات المستخدمة، وفيما يلى نركز على الشروط العلمية للاختبارات الجيدة.

### ١ـ الصدق Validity:

يعتبر «الصدق» أهم شروط الاختبار الجيد، فالاختبار الصادق هو الذى ينجح فى قياس ما وضع من أجله، فمثلا إذا كنا بصدد قياس اللياقة البدنية Physical Fitness فإن صدق الاختبار المستخدم لتحقيق هذا الغرض يعنى أنه يقيس اللياقة البدنية فعلا وليس شيئا آخر.

والدارس للتاريخ الحديث لتطور حركة القياس فى المجال الرياضى يعلم جيدا أن هناك العديد من الاختبارات كانت تستخدم فى الماضى لقياس قدرات بدنية معينة، فى حين أثبتت الدراسات الحديثة أنها تقيس قدرات مخالفة تمامًا.

ويشير تيلر Tyler أن الصدق يعتبر أهم اعتبار يجب توافره فى الاختبار، وحوله يتردد هذا السؤال:

**ماذا يقيس هذا الاختبار؟:**

فقياس قدرات الإنسان ـ أطفالا أو بالغين، أنفسنا أو آخرين ـ يعتبر غير ذى جدوى إذا لم نجد جوابا شافيا لهذا السؤال، وربما يكون الأمر أخطر من مجرد عدم الجدوى؛ حيث يمكن أن يؤدى التسليم الخاطئ بالمعنى الذى تعنيه درجة الشخص على اختبار ما إلى اتخاذ قرارات تؤدى إلى سوء التكيف وارتكاب أخطاء فادحة.. وخاصة إذا استخدمت نتائج الاختبار فى الاختيار والتصنيف والتوجيه والتنبؤ.. وهى مجالات تعد من أهم فوائد استخدام الاختبارات. فأى فرد يستطيع كتابة عدة أسئلة، ولكن سرعان ما يكتشف ـ وربما بعد فوات الأوان ـ أن أسئلته كانت تقيس شيئا آخر غير الذى كان يقصده.

ويشير رايتسون Wrightstone وجاستمان Justman وروبينز Robbins إلى أن صدق الاختبار يعنى **«المدى الذى يحقق به الاختبار أو أى متغير آخر الغرض الذى وضع من أجله»**.

ويشير أيضا بارو Barrow ومك جى McGee إلى أن الصدق يعنى **«المدى الذى يؤدى فيه الاختبار الغرض الذى وضع من أجله، حيث يختلف الصدق وفقا للأغراض التى يراد قياسها، والاختبار**

**الذى يجرى لإثباتها»**. ويرى فان دالين Van Dalen أن وسيلة القياس تكون صادقة إذا كانت تقيس ما تدعى قياسه، وأشار أيضا إلى أن الصدق يعتبر من المقومات ذات الأهمية القصوى فى الاختبار.

ويرى فؤاد أبوحطب وسيد أحمد عثمان أن صدق الاختبار يعنى **«ما يقيسه الاختبار وإلى أى حد ينجح فى قياسه،وهو موضوع لايقتصر على عملية القياس، وإنما قد يمتد إلى المنهج التجريبى بصفة عامة والذى يتطلب اختبار الفروض العملية للتحقق من صحتها»**.

ويعرف كيركون Kurkon الصدق بأنه **«إلى أى مدى يؤدى الاختبار عمله كما يجب»**.

وتشير رمزية الغريب إلى أن صدق الاختبار فى قياس ما وضع من أجله يكون بالنسبة لناحيتين هما:

١ ـ قياس السمة المراد دراستها أو الوظيفة التى يقيسها.

٢ ـ طبيعة العينة أو المجتمع المراد دراسة السمة كعينة مميزة لأفراده.

وعن كيورتن Cureton يتضح أن للصدق مظهرين هما:

ـ المظهر الأول وهو الثبات Reliability.. فالاختبار الصادق يكون أيضا ثابتا فى معظم الأحيان، إلا إذا تدخلت عوامل تحول دون ذلك.

ـ والمظهر الثانى هو ما أسماه بالتعلق Relevence.. ويقصد به مدى اقتراب درجات الاختبار من الدرجات الحقيقية الخاصة بالعينة كلها.

والصدق نسبى Relative.. بمعنى أن الاختبار يكون صادقا بالنسبة للمجتمع الذى قنن فيه، فاختبار جرى ١٥٠٠م قد يكون صادقا لقياس الجلد الدورى التنفسى Cardiovascular Endurance للمرحلة الثانوية، فى حين أن الأمر قد لا يكون على نفس الدرجة من الصدق إذا استخدم نفس الاختبار لقياس نفس القدرة للمرحلة الابتدائية. ولكن هذا لا يعنى بالضرورة أنه لا يوجد اختبارات صالحة لجميع الأعمار والمراحل، فهذا وارد أيضا مثل قياس درجة الحرارة بالترمومتر وضغط الدم بجهازه المعروف.

وأيضا الصدق نوعى Specific، أى أن الاختبار يكون صالحا لقياس ما وضع لقياسه دون غيره، وصلاحية أو صدق الاختبار تحدد عادة بمعامل صدقه Coefficient of Validity.

وتشير الملاحظات الإمبيرقية والتجارب الميدانية فى ضوء الاختبارات العملية أو اختبارات الأداء المتداولة فى التربية البدنية والرياضة إلى أن نقاء الاختبارات من حيث الصدق لا تصل إلى الواحد الصحيح (١٠٠٪)، وقد يرجع ذلك إلى وحدة النمو البيولوجى للإنسان فيما يتعلق بالقدرات البدنية والمهارات الحركية الممارسة، من ثم فوجود اختبار- مثلا ـ لقياس السرعة Speed خالصة يصعب توافره.. حيث يتطلب الأداء فى مثل هذه النوعية من القياس وجود قدرات أخرى كالمرونة Flexibility والجلد Endurance والقوة Strength.. إلخ. ولكن الأمر يتطلب أن يكون للسرعة الغلبة فى الاختبار؛ لذلك نجد دائما أن اختبارات القدرات البدنية لا تحقق صدقا نقيا ١٠٠٪، ويشير الخبراء إلى أن درجة الصدق المقبولة يجب ألا تقل عن ٨٩, ٠، وهذا يعنى أن الاختبار يقيس السرعة بما هو قدره ٨٩, ٠، فى حين يوجد ١١, ٠ تمثل متغيرات أخرى فى الاختبار. وهذا أمر شائع ومتداول بكثرة فى المجال.

وحيث إن الصدق يعبر عن التنبؤ، فإن قيمته تمثل أهمية كبيرة؛ لذلك فإنه كلما ارتفعت قيمة معامل الصدق زادت القدرة التنبؤية للاختبار، والجدول رقم (١) يوضح الكفاية التنبؤية تبعا لقيمة معامل الصدق:

جدول رقم (١)

الكفاية التنبؤية تبعا لقيمة معامل الصدق

| معامل الصدق | الزيادة فى النسبة المئوية للكفاية التنبؤية | احتمالات النجاح والفشل بين (١٠٠) يختارون بوساطة الاختبار | |
|---|---|---|---|
| | | احتمال النجاح | احتمال الفشل |
| صفر | صفر | ٥٠ | ٥٠ |
| ٠,١٠ | ٠,٥ | ٥٠,٢٥ | ٤٩,٧٥ |
| ٠,٢٠ | ٢,٠ | ٥١ | ٤٩ |
| ٠,٣٠ | ٥,٠ | ٥٢,٥ | ٤٧,٥ |
| ٠,٤٠ | ٨,٠ | ٥٤ | ٤٦ |
| ٠,٥٠ | ١٣,٠ | ٥٦,٥ | ٤٣,٥ |
| ٠,٦٠ | ٢٠,٠ | ٦٠ | ٤٠ |
| ٠,٧٠ | ٢٩,٠ | ٦٤,٥ | ٣٥,٥ |
| ٠,٨٠ | ٤٠,٠ | ٧٠ | ٣٠ |
| ٠,٩٠ | ٥٦,٠ | ٧٨ | ٢٢ |
| ٠,٩٥ | ٦٩,٠ | ٨٤,٥ | ١٥,٥ |
| ٠,٩٨ | ٨٠,٠ | ٩٠ | ١٠ |

وعن تير Tyler يتضح أن بيان مدى صدق الاختبارات يتحدد من خلال مضامين ثلاثة يجب على من يستخدم الاختبارات أن يراعيها وهى:

**أولا:** الدراسة الشاملة لجميع البيانات المتاحة عن الاختبارات المستخدمة، وخاصة إذا كان الهدف من استخدام هذه الاختبارات هو اتخاذ قرارات تخص الأفراد أو الجماعات؛ وذلك قبل أى محاولة لتفسير الدرجات المشتقة من الاختبارات.

**ثانيا:** استخدام الاختبارات فى المواقف المعنية التى تقيسها فقط، وبخاصة إذا كانت النية مبيتة على استخدام نتائج الاختبارات فى التنبؤ أو الاختيار.

**ثالثا:** إذا كنا نرغب فى استخدام الاختبارات فى مواقف عملية تتضمن أفرادا، أواستخدامها فى بحوث بحتة تهدف إلى زيادة معرفتنا النظرية عن الفروق الفردية.. يجب أن نتذكر دائما أفكارنا عن

ماهية هذه السمات وكذلك عما تقيسه الاختبارات، وأن هذا يجب أن يتغير كلما حصلنا على بيانات جديدة.

أما عن معامل الصدق Coefficient of Validity فهو معامل الارتباط بين درجات الأفراد فى الاختبار ودرجاتهم فى المحك (أيا ما كان نوعه مادامت قيما رقمية)، أى أن معامل الصدق هو أحد تطبيقات معامل الارتباط.. ورغم كون معامل الارتباط يتمثل فى قيمة نظرية من (ـ ١) إلى (+ ١) وهذا هو الحال فى معامل الصدق، إلا أنه لا يحدث أن ترتبط الدرجة فى الاختبار مع المحك ارتباطا سالبا (على أى وجه)، أى أن القيم الحقيقية والواقعية لمعامل الصدق تنحصر بين (صفر) إلى ( + ١).

**وفيما يلى نستعرض بعض أنواع الصدق:**

- **الصدق الظاهرى** Face Validity:

يعتبر هذا النوع من الصدق أقل أنواعه أهمية، أو بمعنى آخر فإنه يعتبر أضعفها وأقلها استخداما فى المجال، حيث يعتمد ـ فى معظم الأحوال ـ على منطقية محتويات الاختبار ومدى ارتباطها بالظاهرة المقاسة.

فالصدق الظاهرى يمثل الشكل العام للاختبار، أو مظهره الخارجى من حيث مفرداته ومدى وضوح هذه المفردات وموضوعيتها ووضوح تعليماتها.

وقد يطلق على هذا النوع من الصدق اسم «صدق السطح».. ، أى كيف يبدو الاختبار مناسبا وملائما للفرد (أو الأفراد) الذى نقيسه، ومدى ارتباط فقرات الاختبار بالمتغير المقاس.. وهذا النوع يتطلب الانتقال عبر المراحل التالية:

ـ البحث عما «يبدو» أن الاختبار يقيسه.

ـ الفحص المبدئى لمحتويات الاختبار.

ـ النظر إلى فقرات الاختبار، ومعرفة ماذا «يبدو» أنها تقيسه، ثم مطابقة ذلك بالوظائف المراد قياسها.. فإذا اقترب الاثنان كان الاختبار صادقا صدقا سطحيا.

وحساب هذا النوع من الصدق يتطلب التحليل المبدئى لفقرات الاختبار لمعرفة ما إذا كانت تتعلق بالجانب المقاس، وهذا أمر يرجع إلى ذاتية الباحث وتقديره.. ومن هنا تأتى المحاذير.

- **الصدق المنطقى** Logical Validity:

قد يطلق عليه صدق المضمون أو المحتوى Content Validity أو الصدق بالتعريف Validity by definition حيث يعنى مدى جودة تمثيل محتوى الاختبار لفئة من المواقف أو الموضوعات التى يقيسها. فوفقا لهذا النوع من الصدق يعتبر الاختبار صادقا إذا مثلت تقسيماته وتفرعاته تمثيلا سليما، ويتم ذلك أثناء تصميم الاختبار فى ضوء الأبعاد التالية:

ـ تحديد السمة أو الظاهرة أو الخصوصية قيد البحث تحديدا منطقيا (بالتحليل الشامل).

ـ التعرف على أبعاد السمة أو الظاهرة أو الخصوصية المقاسة، وأهمية كل جزء فيها والوزن النسبى لكل جزء أو بُعد من هذه الأجزاء أو الأبعاد، و ذلك بالنسبة للاختبار ككل.

ـ وضع مفردات الاختبار بما يتفق مع الأبعاد أو الأجزاء التى استقر عليها الرأى فى ضوء المرحلتين السابقتين.

فمثلا لو كنا بصدد وضع اختبار لقياس القوة العضلية Muscular Strength فإن الأمر يتطلب دراسة هذه القدرة والتعرف فى ضوء المنطق المستمد من الإطار المرجعى على القدرات الفرعية التى تتضمنها هذه القدرة. وقد يشير هذا الإجراء إلى أن الأشكال البارزة أو القدرات الفرعية للقوة العضلية هى:

١ ـ القوة الحركية Dynamic Strength

٢ ـ القوة الثابتة Static Strength

٣ ـ القوة المتفجرة Explosive Strength

وهنا يلزم ترشيح الاختبارات لتغطية هذه الأبعاد فى ضوء الوزن النسبى لأهميتها (قد تتفق جميع الأبعاد فى وزنها وقيمتها).. وعن طريق منطقية وتطابق الاختبارات مع هذه الأبعاد يكون الاختبار صادقا، حيث يعتمد الأساس النظرى لهذا الأسلوب من الصدق على أن صدق المفردات دليل على صدق الاختبار كله.

ويقاس هذا النوع من الصدق عن طريق التحليل المنطقى لمحتويات الاختبار ومدى مطابقتها مع محتويات الجانب المقاس.

### • الصدق التنبؤى Predective Validity:

يعتمد الصدق التنبؤى على مسلمة محددة هى أن السلوك الإنسانى له صفة الثبات النسبى فى المواقف المستقبلية.. فمع افتراض ثبات هذه المسلمة يمكن التنبؤ بما ستكون عليه استجابة المختبر فى المستقبل للمواقف المحددة قيد البحث، فالمواقف المستقبلية تمثل «المحك» فى هذا التصميم.

فالمدى الذى يسير فيه الاختبار إلى ما يتوقع من نجاح فى المستقبل، وذلك فى الميدان الذى استعمل فيه الاختبار للتنبؤ.. هذا المدى يحمل فى طياته قيمة الصدق التنبؤى للاختبار.

**ويجب مراعاة ثلاثة أبعاد أساسية فى هذا النوع من الصدق هى:**

(أ) حساب القيمة التنبؤية للاختبار.

(ب) الاعتماد على فكرة أن السلوك له صفة الثبات النسبى فى المواقف المستقبلية، وهنا يجب توافر مبررات لذلك.

(جـ) التنبؤ يحتاج إلى فترة بين تطبيق الاختبار ثم جمع البيانات عن المحك فى فترة تالية للاختبار بما يعرف بالطريقة التتبعية Follow - up method، أى تتبع الأداء الواقعى للخاصية أو الظاهرة قيد البحث.

والأمر فى الصدق التنبؤى يختلف عن مثيله فيما يعرف بالصدق التلازمى Concwrrent Validity رغم أن فكرة استخدام «المحك» فيهما تكاد تكون متفقة من حيث الشكل العام، غير أن الاختلاف يتضح فى الوقت الذى يقاس فيه المحك لمقارنته مع الاختبار، فهنا ـ الصدق التنبؤى ـ يكون القياس فى فترة لاحقة عن طريق تتبع أداء المختبر على الظاهرة قيد البحث، أو عن طريق نجاحه المتسقبلى فى أمور تعتمد فى المقام الأول على الظاهرة قيد البحث، أما فى الصدق التلازمى فجمع درجات المحك يسبق أو يتزامن مع جمع درجات الاختبار.

ما سبق يمثل الصدق التنبؤى ويدل على مدى مسايرة التنبؤات التى نصل إليها من الاختبار مع دليل (محك) نجمعه فى وقت تال على إجراء الاختبار.. وهذا يتفق تماما مع ما قام به نوكسKnox عندما أثبت صدق اختباره الخاص بقياس القدرة فى كرة السلة، حيث استخدم اختباره هذا فى اختيار عشرة لاعبين كرة سلة من طلاب المدارس الثانوية (تم اختيارهم بناء على نتائج الاختبار) وقام بتتبعهم فوجد أنهم نجحوا فى تكوين فريق يتنافس على دورى جامعة ولاية أوريجون لكرة السلة(*).

أما عن الطرق الإحصائية المتداولة لتحديد القيم الرقمية لهذا النوع من الصدق فهى:

١ ـ طريقة النسب المئوية Percentage method

٢ ـ طريقة المتوسطات Average method

٣ ـ طريقة الارتباط Correlation method

- **الصدق التلازمى** Concurrent Validity:

يمثل الصدق التلازمى العلاقة بين الاختبار ومحك موضوعى تجمع البيانات عليه وقت أو قبل إجراء الاختبار. وهذا إجراء يسمح بالتعرف على مدى ارتباط الدرجة على الاختبار بمحكات الأداء الراهنة أو مركز الفرد حاليا.. وهذا إجراء يختلف عما يتحدث فى الصدق التنبؤى الذى يتضمن ارتباط الاختبار بمحكات تجمع فى فترة لاحقة لوقت تطبيق الاختبار.

ومن أفضل الصور لاستخدام هذا النوع من الصدق فى مجال التربية البدنية ما أشار إليه كل من جنسن Jensen وهرست Hirst فى تحديد صدق اختبار الوثب العمودى Vertical Jump عن طريق إيجاد معامل الارتباط بين درجات الأفراد على هذا الاختبار ودرجاتهم على أربع مسابقات للقدرة فى ألعاب القوى (بلغ معامل الصدق ٠,٧٨) على أساس أن المتغير الحاسم فى هذه النوعية من الأداء الحركى يتطلب القدرة Power أو القوة المتفجرة Explosive Strength.

- **الصدق التجريبى:**

يعتبر الصدق التجريبى من أفضل أنواع الصدق وأكثرها شيوعا، حيث يعتمد على إيجاد معامل الارتباط بين الاختبار الجديد واختبار آخر سبق إثبات صدقه فى قياس الظاهرة قيد البحث. فمثلا إذا كنا

بصدد قياس اللياقة البدنية Physical Fitness عن طريق اختبار جديد بنى وقنن لهذا الغرض فمن الممكن إيجاد صدق هذا الاختبار عن طريق إيجاد معامل الارتباط بينه وبين اختبار آخر ثبت صدقه لقياس اللياقة البدنية.

ونظرا لأن هذا النوع من الصدق يمثل إجراء صالحا للتنبؤ بدرجات المحك[1] أو الميزان من درجات الاختبار، وذلك باستخدام قانون الانحدار (انحدار درجات المحك على درجات الاختبار) وحيث إن هذا النوع يعتمد فى المقام الأول على صدق المحك أو الميزان نفسه فإن الأمر يتطلب بذل جهد صادق فى اختيار المحك بحيث يكون على درجة عالية من الصدق.

ومن عيوب استخدام اختبار آخر كمحك فى إيجاد الصدق بهذا الأسلوب ما يلى:

١ ـ حداثة محال الظاهرة قيد البحث وقلة الاختبارات (أو محكات أخرى) السابقة التى يمكن استخدامها كمحك أو ميزان.

٢ ـ الشك فى صلاحية الاختبارات المتداولة فى المجال لقياس الظاهرة قيد البحث كمحك للاختبار الجديد.

ويجب ألا يغيب عن خاطرنا أن الصدق التجريبى (أو الصدق العملى أو صدق الوقائع Emperical Validity ـ كما يطلق عليه البعض) لا يقتصر فى أساليبه على استخدام الاختبارات فقط كمحكات، ولكن الأمر يمتد إلى استخدام المواقف العملية والنجاح الميدانى وأساليب أخرى كمحكات للظواهر قيد الدراسة، ومن ثم فإن ما يذكر عن الصدق التلازمى أو الصدق التجريبى يعتبر ضمنيا تفرعات منطقية لهذا النوع من الصدق. ويشير فان دالين Van Dalen إلى أن الصدق التجريبى يمكن تحقيقه عن طريق:

(١) صدق التجانس الداخلى (قدرة الاختبار على التمييز).

(٢) طريقة المحك الخارجى.

**• الصدق العاملى** Factorial Validity:

يعتبر هذا النوع من الصدق من أفضل الأنواع المتداولة، حيث يعتمد على أسلوب إحصائى متقدم هو التحليل العاملى Factor Analysis، فالتشبعات البارزة للاختبارات على عواملها المقبولة تمثل القيم المحددة لصدق هذه الاختبارات، وهذا يمثل الصدق العاملى الذى هو فى الواقع معامل الارتباط بين الاختبار وبين ما هو شائع أو مشترك أو عام فى مجموعة الاختبارات الخاضعة للتحليل. وسواء كان هدف التحليل هو الحصول على عامل أو عوامل طائفية أو عوامل أولية فإن التشبعات المشاهدة للاختبار على العامل تعتبر دليلا على صدقه وتحديد قيمة هذا الصدق[2].

---

(١) المحك CRITERION هو معيار نحكم به على اختبار أو نقيمه، وقد يكون مجموعة من الدرجات أو المقاييس أو التقديرات أو الإنتاج.. إلخ، صمم الاختبار للتنبؤ بها، أو الارتباط معها، كمقياس لصدقها.. وهو أيضا ـ المحك ـ مجموعة من المفاهيم أو الأفكار المستخدمة فى الحكم على محتوى الاختبار عن طريق تقدير مضمونه أو صدقه المنطقى.

وفى هذا الخصوص يقول تيلر Tyler: إن هناك سؤالا هاما متعلقا بالثبات هو:

**ما مدى الدقة أو مدى الاطراد الذى يقيس به الاختبار ما يقيسه؟**

ويشير رايستون Wrightstone وجاستمان Justman وروبينز Robbine إلى أن ثبات الاختبار يعنى **«درجة ثبات ما يقيسه الاختبار»**، كما يعرف بارو Barrow ومك جى McGee الثبات بكونه **«هو مدى اتساق الاختبار فى قياس ما يقيسه»** كما تشير رمزية الغريب إلى أن ثبات الاختبار يعنى أن يعطى الاختبار نفس النتائج إذا ما أعيد على نفس الأفراد فى نفس الظروف.

ويعنى الثبات أيضا ثبات درجات المفحوصين على الاختبار إذا تكرر قياسه أكثر من مرة، شريطة ألا تتدخل عوامل أخرى (كالنمو مثلا) تغير من حالة المفحوص فى الشىء الذى يقيسه الاختبار. ويقول فان دالين Van Dalen: إن الاختبار يعتبر ثابتا إذا كان يعطى نفس النتائج باستمرار، إذا ما تكرر تطبيقه على نفس المفحوصين وتحت نفس الشروط.

والشكل العلمى للثبات يتعلق بنسبة تباين الخطأ Errer Variance والتباين الحقيقى من التباين الكلى للاختبار، وهذا يعنى فى مفهومه العام مدى اعتماد الفروق الفردية فى درجات الاختبار على أخطاء الصدفة Chance error والاختلافات الحقيقية فى الصفة الخاضعة للقياس.

وقد نكون أقرب إلى الحقيقة عنها من الخطأ إذا قلنا أنه لا يوجد اختبار يخلو من خطأ الصدفة، سواء كان السبب فى ذلك الفاحص أو المفحوص أو الاختبار نفسه أو الظروف المحيطة بعملية القياس، وهذا يعنى أن هناك علاقة عكسية بين قيمة ثبات الاختبار وخطأ الصدفة.

فأخطاء القياس العرضية، أو تباين الخطأ، أو خطأ الصدفة.. تعنى عدم تطابق نتائج تطبيق نفس الاختبار على نفس الأفراد إذا ما كرر القياس، وهذه الأخطاء قد تكون قليلة، بما لا يؤثر على قيمة الثبات، وقد تكون كبيرة بحيث تؤثر تأثيرا واضحا على ثبات الاختبار.

ويجب ملاحظة أن تباين الخطأ يتعلق بالظروف الاختبارية التى لا تصل من قريب أو بعيد بغرض الاختبار، وحسم هذه الظروف يقلل من نسبة تباين الخطأ، وبالتالى ترفع من ثبات الاختبار واستقراره.. ويجب أن ننوه هنا إلى أن بعض ما يعتبر تباين الخطأ قد يكون تباينا حقيقيا (فى الصفة المقاسة نفسها) فى موقف آخر، فإذا كان اتجاهنا مثلا يتجه نحو قياس تأثر القوة العضلية بظروف تباين الحالة النفسية يصبح الأمر مرتبطا بهدف الاختبار نفسه، وهذا جزء من التباين الحقيقى وليس من تباين الخطأ.

والثبات يعنى الاتساق Consistat، بمعنى أن درجات الفرد لا تتغير جوهريا بتكرار إجراء الاختبار عليه، أو أن موقف الفرد النسبى لايتغير فى مجموعه.

كما يعنى الثبات أيضا الاستقرار Stability، وهذا يعنى أنه إذا تكررت عملية قياس الفرد الواحد بنفس الاختبار فإن درجاته على الاختبار فى المرات المتكررة للقياس تظهر شيئا من الاستقرار.

كما يعنى الثبات الموضوعية Objectivity، وهذا يعنى أن الفرد يحصل على نفس الدرجة على الاختبار لو اختلف المحكمون.

### • الصدق الذاتى:

ويطلق عليه أيضا مـؤشر الثبات Index of Reliability، وهو صدق الدرجـات التجريبية بالـنسبة للدرجات الحقيقية التى خلصت من شوائب أخطاء الصدفة، ومن ثم فإن الدرجات الحقيقية True Scores هى الميزان أو المحك الذى ينسب إليه صدق الاختبار.

وحيث إن ثبات الاختبار يعتمد على ارتباط الدرجات الحقيقة للاختبار بنفسها إذا أعيد الاختبار على نفس المجموعـة التى أجرى عليها فى أول الأمر. لهـذا كانت الصلة وثيقة بين الثـبات والصدق الذاتى، شريطة أن يحسب الثبات بأسلوب الاختـبار ـ إعادة الاختبار Test - retest، ويحسب عن طريق المعادلة التالية:

الصدق الذاتى= $\sqrt{\text{الثبات}}$

ويجب ملاحظة أن النتـيجة المسـتخلصة من هذه المعـادلة تمثل الحد الأقصى المتـوقع للصدق وليس القيمة الحقيقية لصدق الاختبار.

### • صدق المفهوم Construct Validity:

هو الارتباط بين الجوانب التى يقـيسها الاختـبار وبين مفهوم هذه الجوانب، أى عند اسـتخدام هذا النوع من الصدق يلزم (بطريقة أو بأخرى) تحديد ما نقصد بمصطلح يصف ما يقيسه الاختبار.

فمثلا إذا كنا بصدد قياس القوة الثابتة باخـتبار الشد على الديناموميتر، ونحن ننظر إلى هذه القدرة على أنها بذل القوة القصوى فى شكل عمل ثابت ضد مقاومة ثابتة (الـديناموميتر) فإن هذا الاختبار يعتبر صادقا فيما وضع من أجله.

ويعـيب هذا النوع من الصدق أنه يعـتمـد على مـفهـوم الباحث للظاهرة أو السـمة أو القـدرة أو الخصوصية قيد البحث، فإن لم يحالفه الحظ فى هذا المفهوم أصبح الإجراء كله خاطئا. وهذا أمر اعتبارى ويصعب الاعتماد عليه.

ويقـاس هذا النوع من الصدق على أسـاس تفسـير نتـائج الاخـتبـار عن طريق النظرية التى وضع المقياس بناء عليها.

### • الصدق عن طريق الاتساق داخل وحدات الاختيار:

يمكن استخدام هذا الأسلوب فى حساب الصدق، شـريطة أن لا يعتمد عليه فقط، بمعنى أنه يلزم أن تسبقه طرق أخرى أصـيلة للتأكد من الصدق. أما فى حالة استخدام هـذا الأسلوب منفردا فهو يقيس الثبات وليس الصدق. . ، راجع هذا الأسلوب فى الثبات (تالى).

### ٢ـ الثبات Reliability:

الثبات يمثل العـامل الثانى فى الأهمية بعد الصدق فى عمليـة بناء وتقنين الاختبارات، وهو ـ أى الثبات ـ يعنى أن يكون الاختبار على درجة عالية من الدقـة والإتقان والاتـساق والموضوعية والاطراد فيما وضع لقياسه.

وبالطبع فإن الثبات نسبى Relative حيث يختلف باختلاف الاختبارات والأفراد والسمة المقاسة.

وعن أسباب أخطاء الصدفة فإنها عديدة ومتنوعة، منها الانفعال، وحالة الجو، وسوء الحظ، والتخمين أو التوقع.. وعن ثورنديك Thorndike يمكن حصر هذه الأسباب فيما يلى:

**١ ـ سمات الفرد العامة الدائمة:**

(أ) مستوى قدرة الفرد فى واحدة أو أكثر من السمات العامة، والتى تعمل فى عدد من الاختبارات.

(ب) مهاراته العامة وطريقته فى حل الاختبارات.

(جـ) قدرته العامة على فهم التعليمات.

**٢ ـ سمات الفرد الخاصة الدائمة:**

(أ) مستوى قدرة الفرد فى السمات المطلوبة لأداء هذا الاختبار المعين دون غيره.

(ب) المعلومات والمهارات النوعية الخاصة بنوع معين من مفردات الاختبار.

١ ـ أنواع معينة من مفردات الاختبار.

٢ ـ أثر عوامل الصدفة المتعلقة بمدى معرفة الممتحن لنوع معين من المفردات.

**٣ ـ سمات الفرد العامة المؤقتة:**

(أ) الصحة.

(ب) التعب.

(جـ) الدافعية.

(د) التوتر الانفعالى.

(هـ) منطق الاختبار نفسه.

(و) فهم طريقة أداء أسئلة الاختبار.

(ز) الظروف الخارجية المحيطة بالمختبر ـ مثل الهدوء، و درجة الحرارة، ونوع الإضاءة، ومدى استراحته على مقعده.. إلخ.

**٤ ـ سمات الفرد الخاصة المؤقتة:**

(أ) العوامل النوعية الخاصة بالاختبار ككل.

١ ـ فهم الأعمال الخاصة المطلوبة من الاختبار.

٢ ـ الحيل النوعية الخاصة بمعالجة أحد مقدرات الاختبار أو مادته.

٣ ـ التأهب العقلى الوقتى لحل اختبار معين.

(ب) العوامل النوعية الخاصة ببعض بنود الاختبار:

١ ـ تذبذب الذاكرة.

٢ ـ التذبذب الذى لا يمكن التنبؤ به والمتعلق بتركيز الانتباه والدقة.. إلخ.

٣ ـ عوامل الصدفة المنتظمة التى تؤثر فى إجراء الاختبار:

(أ) ظروف إجراء الاختبار الخاصة بالزمن، ووضوح التعليمات.. إلخ.

(ب) العوامل المؤثرة فى ترتيب السمات أو الأداء.

**٥ ـ التباين الذى لم يبين سابقا (الحظ):**

عامل الحظ أو الصدفة التى أدت إلى اختيار سؤال أو إجابة معينة، أو الخاصة بتخمين إجابة صحيحة معينة.

ويشير رايتسون Wrightstone وجاستمان Justman وروبينز Robbins إلى أن معامل الثبات Coefficint of Reliability يعنى **«معامل الارتباط بين صورتين متكافئتين لاختبار واحد، أو بين درجات اختبار يطبق مرتين على نفس التلاميذ بينهما فترة من الزمن (لا تغير ظروف المفحوص) أو معامل الارتباط بين نصفى اختبارما»**. وهذا التعريف يتضمن تحديدا واضحا لأنواع الثبات.

ويشير بارو Barrow ومك جى McGee إلى نفس المفهوم السابق تقريبا فيما يتعلق بتعريف (وأنواع) الثبات، وهو: معامل الثبات يعنى معامل الارتباط Cefficient of correlation بين شكلين للاختبار Two forms، ويسمى هذا النوع بالصورة المتبادلة alternate - form reliability أو الصور المتوازية للثبات Parallelform reliability.

كما يعنى أيضا معامل الارتباط بين الدرجات عند إعادة تطبيق نفس الاختبار، ويسمى هذا النوع الثبات بالاختبار ـ إعادة الاختبار Test - retest reliability.

كما يعنى أيضا معامل الارتباط بين نصفى الاختبار، ويسمى هذا النوع بالثبات بالتجزئة النصفية Split - half reliability.

كما يمكن التوصل إلى معامل الثبات عن طريق استخدام معادلات كودر ـ ريتشاردسون Koder - Richardson formalas.

**أما عن أنواع الثبات فيشير فان دالين Van Dalen أنها:**

١ ـ الثبات بطريقة إعادة الاختبار: يعطى الاختبار لنفس المفحوصين مرتين ثم يحسب معامل الارتباط بين التطبيقين.

٢ـ الثبات بطريقة الصور المتكافئة: يعد صورتين متكافئتين ويطبقان على المفحوصين، ثم يحسب معامل الارتباط بين نتائج الصورتين.

٣ ـ الثبات بطريقة التجزئة النصفية: يطبق الاختبار مرة واحدة فقط، ولكن تقسم بنوده عشوائيا إلى نصفين، ويحسب الارتباط بين درجات النصفين.

وفيما يلى نستعرض الطرق المختلفة الشائعة لإيجاد معامل الثبات ولكن قبل أن نقوم بذلك يجب أن ننوه إلى أنه مادام جميع طرق الثبات تقريبا تعتمد على معامل الارتباط فإنه من المستحيل أن نحصل على اطراد تام بين مجموعتين من الدرجات ؛لأن عوامل الصدفة موجودة دائما، أى أن معامل الارتباط لن يكون دائمًا واحد صحيح.

## • الثبات عن طريق إعادة الاختبار:

يعتبر أسلوب الثبات عن طريق الاختبار ـ إعادة الاختبار ـ Test - retest reliability أكثر طرق إيجاد عامل الثبات صلاحية بالنسبة لاختبارات الأداء فى التربية البدنية والرياضة، وأكثرها شيوعا واستخداما أيضا. حيث تعتمد على فكرة إعادة تطبيق الاختبار على نفس الأفراد وفى نفس الظروف مرة أخرى بعد انقضاء فترة زمنية، إذ يعبر معامل الارتباط بين التطبيقين (الأول والثانى) عن معامل الثبات أو معامل الاستقرار Coefficient of stability كما يطلق عليه البعض.

وقد تكون مشكلة هذا النوع من الثبات هى الفترة الزمنية الواقعة بين التطبيقين، حيث يتفق جميع الخبراء على أنه لا يجب أن يتخللها تعلم أو نمو فى قدرات المفحوص فيما يتعلق بالظاهرة الخاضعة للقياس. ،فمن المعروف أن درجة الاستقرار الوقتى Temporal Stability ترتبط بطول الفترة الزمنية بين التطبيقين، وكثير من الاختبارات أظهر ثباتا عاليا عندما أعيد تطبيقها بعد فترة زمنية قصيرة من التطبيق الأول، فى حين انخفض ثباتها بشدة عندما طالت هذه الفترة الزمنية.. والتفسير الإحصائى لهذه الظاهرة هى أنه كلما قلت الفترة الزمنية بين التطبيقين كلما عزى هذا إلى تفسير اختلافات تباين الخطأ، بينما فى حالة طول هذه الفترة فإن الأمر قد يرتبط بالتباين الحقيقى للاختبار.

والجانب الآخر من المشكلة هو أن المبالغة فى تقصير الفترة الزمنية بين التطبيقين قد يؤثر فى القيمة الحقيقية لثبات الاختبار، إذ يؤثر التطبيق الأول على نتائج التطبيق الثانى، وهذه ملاحظة تستحق التقدير خاصة فى الاختبارات المعرفية فى التربية البدنية والرياضة.

ونرى أن استخدام فترة أسبوع كفاصل بين التطبيقين فى حالة اختبارات الأداء Performance tests فى التربية البدنية تعتبر إجراء مناسبا للحصول على معامل للثبات بهذا الأسلوب، فى حين أن الأمر يختلف بالنسبة لاختبارات الورقة والقلم ـ المعرفية مثلا ـ حيث يتطلب الأمر فترة أطول تحدد فى ضوء ظروف الاختبار ومضمونه ونوعية المفحوصين.

## • الثبات عن طريق التجزئة النصفية:

يعتبر أسلوب التجزئة النصفية Split - half relibility من أكثر طرق الثبات استخداما فى اختبارات الورقة والقلم، ولكنها لا تصلح لاختبارات الأداء فى التربية البدنية والرياضة، وتتلخص إجراءات هذه الطريقة فى تقسيم الاختبار إلى نصفين يراعى فيهما تساوى عدد الوحدات وكذلك القيم الإحصائية لهما (المتوسط، الانحراف المعيارى، معامل الارتباط..) وهذا يستدعى ضرورة توافر شرط تجانس نصفى الاختبار. فمثلا إذا كان لدينا اختبار معرفى (يقيس الجانب المعرفى فى كرة السلة) مكونا من ٣٠ فقرة،

فمن الممكن (وهذا هو الشائع فى ظل شرط ترتيب الوحدات حسب مستوى الصعوبة) حساب معامل الثبات عن طريق تقسيم الاختبار إلى نصفين متكافئين، كان يتضمن القسم الأول الفقرات ذات الأرقام الفردية، ويتضمن القسم الثانى الفقرات ذات الأرقام الزوجية، وهذا يؤيد ما ذكره رايتسون Wrightstone وجاستمان Justman وروبنز Robbins أن الثبات بين نصفى الاختبار يتضمن الارتباط بين درجات نصف الاختبار مع درجات النصف الآخر، وعموما يتكون أحد نصفى الاختبار من الأسئلة ذوات الأرقام الفردية والنصف الآخر من الأسئلة ذوات الأرقام الزوجية.

وعلى الرغم من أن فان دالين Van Dalen قد أشار إلى إمكانية أن يكون التقسيم عشوائيا، إلا أننا نرى أن هذا قد يكون مناسبا عند تساوى جميع فقرات الاختبار من حيث مستوى الصعوبة[1]، ولكننا لا نؤيد هذا الإجراء إذا اختلفت وحدات الاختبار فى مستويات صعوبتها.

وعلى الرغم من شيوع هذا النوع لإيجاد معامل الثبات، إلا أن توفير تكافؤ مناسب لنصفى الاختبار قد يكون صعب التحقيق، حيث توجد معوقات كثيرة ولتحقيق ذلك نذكر منها على سبيل المثال:

- الفروق الواردة فى طبيعة الأسئلة أو الوحدات.
- الفروق الواردة فى مستوى صعوبة الوحدات أو الأسئلة.
- الآثار التراكمية للحماس فى الأداء أو العمل،وكذلك الممارسة والتعب والملل وغيرها من العوامل التى تختلف فى بداية الاختبار عنها فى نهايته.

لذلك يتطلب الأمر - لتلافى ما سبق من معوقات - للحصول على تقسيم موضوعى للاختبار تحديد مستوى صعوبة الوحدات وتوفير التكافؤ الإحصائى من حيث المتوسطات والانحرافات المعيارية ومعاملات الارتباطات والصدق.. إلخ.

ويشير الخبراء إلى أن هذا الأسلوب - التجزئة النصفية - فى مضمونه يعنى الثبات بالاتساق الداخلى Internal consistency وأن هذا النوع يعطينا معامل ثبات نصفى الاختبار فقط وليس الاختبار كله، وذلك بالمقارنة مع الأسلوب السابق عرضه وهو الاختبار - إعادة الاختبار وكذلك الأسلوب الذى سيلى عرضه وهو الصور المتكافئة.

### • الثبات عن طريق الصور المتكافئة:

يطلق على هذا النوع من الثبات أسماء عدة هى:

- الصور المتبادلة Alternale - form.
- الصور المتوازية Parallel - form.
- الصور المتكافئة Equivalence - form.

---

(١) اختبار متدرج الصعوبة Scaled test يعنى أن تكون أسئلة الاختبار مرتبة ترتيبا تصاعديا من حيث درجة الصعوبة.

وكلها تحمل نفس المعنى. . وتستخدم هذه الطريقة لإيجاد معامل الثبات عندما يتوافر صورتان (أو أكثر) متكافئتين من الاختبار، ومعنى التكافؤ هنا هو أن يكون قد تم بناء وتصميم كل صورة من الصورتين على حدة وبطريقة مستقلة بشرط توافر عدد من المواصفات المحددة وهى:

- شروط التكافؤ الإحصائى (المتوسط، الانحراف المعيارى، معامل الارتباط، معامل الصدق الداخلى. . .إلخ).
- تساوى عدد الأسئلة فى الصورتين.
- تماثل صياغة الأسئلة فى الصورتين.
- تماثل المحتوى فى الصورتين.
- تساوى مستوى الصعوبة فى الصورتين.
- تماثل متغيرات القياس فى الصورتين (التعليمات، الزمن. .إلخ).

فالصورة المتكافئة (البديلة) وفقا لما أشار إليه رايتسون وآخرون تعنى «**صورتان أو أكثر لاختبار ما، وهى متشابهة تماما من حيث دلالة القياس ودرجة الصعوبة لوحدات الاختبارات، وتعطى متوسط تقدير متشابها وتشتتا واحدا إذا ما طبقت كل منها على مجموعة واحدة**».

وترجع أهمية توافر شروط التكافؤ فى الصورتين إلى أن إيجاد الثبات بهذا النوع يعتمد على إيجاد معامل الارتباط بين الصورتين، وهذا يعبر عن معامل التكافؤ أيضا[1].

وكمثال لاستخدام هذه الطريقة فى إيجاد الصدق ما قام به حسن معوض ومحمد محمود عبدالدايم عندما قاما باقتباس وتعريب اختبار «كارلوس ل. وير» لقياس الاتجاهات نحو التربية الرياضية، عن طريق إيجاد معامل الارتباط بين صورتى الاختبار (الصورة الأولى تتضمن ٣٠سؤالا، والصورة الثانية تتضمن ٢٩ سؤالا) حيث بلغت قيمة معامل الثبات (معامل الارتباط بين الصورتين) ٧٢٣, .[2].

### • الثبات عن طريق الاتساق داخل وحدات الاختبار:

يختلف الاتساق داخل وحدات الاختبار Inter - item consistency عن الاتساق الداخلى Internal consistency السابق ذكره فى طريقة التجزئة النصفية. حيث إن الاتساق داخل وحدات الاختبار يعنى استقرار استجابات المفحوص على أسئلة الاختبار واحدا بعد الآخر، وهذا يشير إلى التجانس homogeneity الكلى للاختبار، أما التجانس الداخلى السابق الإشارة إليه فى طريقة التجزئة النصفية فإنه يعنى التجانس بين نصفى الاختبار فقط.

وعلى الرغم من أن البعض يرى أن الاتساق الداخلى (كما هو الحال فى التجزئة النصفية) والاتساق بين وحدات الاختبار (هذا النوع) يمثلان بعدين لنوع واحد يطلق عليه معامل الاتساق الداخلى، إلا أننا

---

(١) معامل التكافؤ يعنى تقدير معامل الثبات الناتج من تطبيق صورتين متكافئتين لاختبار ما.

نرى أن الاتساق بين وحدات الاختبار يعتبر أكثر دقة من الاتساق الداخلى حيث يتعامل مع كل مفردة من مفردات الاختبار، وهذا يقترب فى معناه العام من تحليل التباين Analysis of Variance ويعتبر أسلوب كيودر ـ ريتشاردسون Kuder - Richardson[1] من أشهر الأساليب المستخدمة فى هذا المضمار، ومن أنماط هذه الأساليب المعادلتان التاليتان:

$$\text{أولا: ر} = \left(\frac{\text{ن}}{\text{ن} - ١}\right)\left(\frac{\text{ع}^{٢} - \text{مجـ ص ح}}{\text{ع}^{٢}}\right)$$

حيث ر = معامل ثبات الاختبار.

، ن = عدد الأسئلة التى يتكون منها الاختبار.

، ع = الانحراف المعيارى لدرجات الاختبار.

، مجـ ص ح = يحسب بأعداد نسبة الأفراد الذين أصابوا فى إجاباتهم على كل سؤال، ونسبة الذين أخطأوا، ثم تضرب النسبتين فى بعضهما البعض ثم نجمع حواصل الضرب بالنسبة لجميع الأسئلة.

$$\text{ثانيا: ر ١١} = \frac{\text{ن ع}^{٢} - \text{م(ن ـ م)}}{(\text{ن} - ١)\,\text{ع}^{٢}}$$

حيث ر= معامل الثبات.

، ن = عدد أسئلة الاختبار.

، ع٢= تباين درجات الاختبار.

، م= متوسط درجات الاختبار.

### ٣ـ الموضوعية Objectivity:

موضوعية الاختبار تعنى عدم تأثر الاختبار بتغيير المحكمين، أو أن الاختبار يعطى نفس النتائج مهما كان القائم بالتحكيم.

ويشير فان دالين Van Dalen إلى أن الاختبار يعتبر موضوعيا إذا كان يعطى نفس الدرجة بغض النظر عمن يصححه.. وهذا يعنى استبعاد الحكم الذاتى للمحكم، أى أنه كلما زادت الذاتية قلت الموضوعية.

ويعرف بارو Barrow ومك جى McGee الموضوعية بكونها «**درجة الاتساق بين درجات أفراد مختلفين لنفس الاختبار، ويعبر عنه بمعامل الارتباط**».

كما يرى كل من رايتسون Wrightstone وجاستمان Justman وروبينز Robbins أن الاختبار الموضوعى يعنى اختبارا يستبعد منه الرأى الشخصى للمصحح.

---

(١) معادلة كودر ـ ريتشاردسون هى معادلة لتقدير ثبات الاختبار من معلومات عن فقرات الاختبار، أو معرفة الدرجة المتوسطة، والانحراف المعيارى، وعدد الفقرات فى الاختبار، وهى تدل على الاتساق الداخلى للاختبار.

ويشير الخبراء إلى أن الموضوعية تتضمن اتجاهين أساسيين هما:

١ ـ ثبات المصحح أو الحكم Reliability of scorer أو Inter - scorer reliability، وهو ثبات يعبر عن تباين الخطأ الذى يمكن إرجاعه إلى الفروق الفردية بين الممتحنين أوالمحكمين أو القضاة ويؤثر فى التباين الكلى للاختبار.

ويتم الحصول على هذا النوع من الثبات عن طريق معامل الارتباط بين محكمين أو أكثر يقومون بالتحكيم لنفس الأفراد ولنفس الاختبار فى نفس الوقت.

٢ ـ ثبات المفحوص نفسه، وهذا يمكن الحصول عليه من معامل الثبات عن طريق إعادة الاختبار مع ثبات المصححين فى مرتى التطبيق.

مما سبق يتضح أن ثبات المصحح وثبات المفحوص يكونان معا معامل الموضوعية Coefficient of Objectivity.

ومن الأمور الهامة فى الموضوعية وضوح الاختبارات من حيث التعليمات والصياغة والإجراءات والشروط.. إلخ، وكلما زادت شروط التقويم ـ وكان للاختبار قواعد ثابتة للتصحيح ـ ارتفع معامل الموضوعية، حيث إن زيادة الشروط يتيح أسسا موحدة للتقويم مما يؤدى إلى تقليل الاعتبارية Subjectivity فى التقويم.

كما يجب التأكد من فهم المفحوص لتعليمات الاختبار والدور المطلوب منه قبل بدء الاختبار، وفى هذا الصدد يصبح من الملح الإجابة على السؤالين التاليين:

١ ـ ماذا يقصد واضع السؤال أو مصمم الاختبار؟

وهذا التساؤل يزيل اللبس الممكن وقوع المفحوص فيه.

٢ ـ ما هى الإجابة التى ينتظرها؟

وذلك لتلافى حدوث خطأ فى فهم الواجبات المكلف بها المفحوص.

مما سبق يتضح أن الموضعية تتأثر بدرجة كبيرة عندما يكون التقويم على أساس ذاتى أو اعتبارى، فالاختبار الذاتى أو الاعتبارى Subjective test هو اختبار يصحح على أساس اتجاهات Attitudes وآراء Opinions وميول Interests ومعايير Norms وأفكار المصحح نفسه، وهذه هى المشكلة لأن الموضعية فى معناها الدقيق تعنى أن درجة المفحوص لا تتأثر بشخص الفاحص وإنما بناء على قدرة المفحوص على ما يقيسه الاختبار.

فى حين أن الاختبار الموضوعى Objective test فإنه اختبار يصحح بطريقة لا تتأثر بشخص من يقوم بتصحيحه، فالنتائج واحدة برغم اختلاف المصححين.

ويشير باومجرتنير Baumagrtnir وجاكسون Jackson إلى أن إيجاد معامل الموضوعية عن طريق إيجاد معامل الارتباط بين تسجيل محكمين للاختبار يعتبر إجراء سليما فى التربية البدنية، وأشار إلى أن هذا النوع من الموضوعية يطلق عليه اسم الثبات التخمينى Rater Reliability.

### ٤ـ المعايير والمستويات Norms and Standards:

الاختبارات الجيدة تتضمن معايير Norms أو مستويات Standards، حيث تمثل هذه المعايير أو المستويات القيم المعيارية الموازية للقيم الخام المستخلصة من الاختبارات.

ووجود المعايير يسمح للمختبر أن يتعرف على مركزه النسبى فى المجموعة، وهذا يعتبر إجراء هاما وضروريا لتحقيق شروط التقويم المثلى.. كما يجب ملاحظة أن المعايير ليست مستويات مثلى نسعى إليها، وإنما هى قيم تحدد مركز الفرد النسبى وتساهم أيضا فى وضع درجات كلية لوحدات مختلفة فى وحدات قياسها خاصة فى اختبارات الأداء البدنى.

**وفيما يلى نعرض طرق اشتقاق الدرجات المعيارية:**

#### ١ـ الدرجة الزادية Z. Score:

الدرجة الزادية تمثل انحراف الدرجات الخام Raw Scores عن متوسطها الحسابى، والانحراف عن المتوسط الذى يحدد بصفر يكون فى حدود ±٣ انحراف معيارى كحد أقصى.

بناء على ذلك فالدرجة الزادية تمثل درجة معيارية متوسطها صفر والحدود القصوى لانحرافها المعيارى ±٣.

كما أن الدرجة المعيارية «Z» تمتد عادة بين -٣ إلى +٣ انحراف معيارى، ومتوسطها الحسابى= صفراً، وانحرافها المعيارى يساوى واحد، كما أن الانحراف المعيارى الموجب يعنى زيادة الدرجة الخام عن المتوسط الحسابى فى حين الانحراف المعيارى السالب يعنى نقصان الدرجة الخام عن المتوسط الحسابى. كما أن انحراف الدرجات الخام على جانبى المتوسط تمثل قيما متساوية، فالقيمة (+١) تماثل (-١)... وهكذا.

وفيما يلى القانون المستخدم فى تحويل الدرجات الخام إلى درجات زادية معيارية:

$$\text{الدرجة ذ} = \frac{\text{س} - \bar{\text{س}}}{\text{ع}} \qquad Z = \frac{x - \bar{x}}{SD}$$

حيث ذ = الدرجة الزادية المعيارية.

س = الدرجة الخام.

$\bar{\text{س}}$ = المتوسط الحسابى للدرجات الخام.

ع = الانحراف المعيارى للدرجات الخام.

وفى حالة الاختبارات الزمنية تستخدم المعادلة كما يلى:

$$\text{الدرجة ذ} = \frac{\bar{\text{س}} - \text{س}}{\text{ع}} \qquad Z = \frac{\bar{x} - x}{SD}$$

وفيما يلى الخطوات الواجب استخدامها لاستخراج الدرجة الزادية المعيارية من الدرجات الخام المشتقة مباشرة من نتائج الاختبارات.

١ ـ يتم ترتيب الدرجات الخام تصاعديا أو تنازليا.

٢ ـ يحسب المتوسط الحسابى لدرجات الخام عن طريق جمع قيم المشاهدات على عددها من خلال القانون:

$$س = \frac{مجـ س}{ن}$$

حيث س = المتوسط الحسابى للقيم المشاهدة.

ن = عدد القيم المشاهدة.

٣ ـ يتم حساب الانحراف المعيارى للدرجات الخام.

٤ ـ تحديد أعلى القيم وأقلها، والأرقام الخام التى تقع بين أعلى القيم وأقلها.

٥ ـ تطبق معادلة الدرجة «ذ» على كل قيمة من القيم الواقعة ما بين أعلى قيمة وأقل قيمة.

تستخدم نفس الخطوات فى حالة الاختبارات الزمنية مع مراعاة استخدام المعادلة المخصصة للاختبارات الزمنية.

**هذا، وتتميز هذه الدرجة بما يلى:**

١ ـ سهولة الحساب والتفسير والفهم.

٢ ـ من أنسب طرق الدرجات المعيارية عندما يقترب توزيع القيم المشاهدة من التوزيع الاعتدالى (المنحنى الاعتدالى).

٣ ـ هى درجة متوسطها الحسابى دائما= صفر، وانحرافها المعيارى= ١.

٤ ـ تعتمد على انحراف الدرجة الخام عن المتوسط الحسابى للقيم المشاهدة فقط دون الحاجة لأى عمليات إحصائية أخرى.

### ٢. الدرجة التائية T. Score:

من أكثر الطرق المستخدمة فى بحوث التربية البدنية والرياضة وخاصة فى اختبارات الأداء البدنى، وهى تعالج عيوب الطريقة السابقة (Z.Score) المتمثلة فى وجود قيم سالبة وصغر الدرجات. وهى درجة متوسطها الحسابى ٥٠ وانحرافها المعيارى= ١٠. لذلك فإن نواتجها دائما تكون موجبة.

وفيما يلى القانون المستخدم لتحويل الدرجات الخام إلى درجات تائية:

$$\text{الدرجة التائية (ت)} = ٥٠ + \frac{١٠ (س - \bar{س})}{ع}$$

$$T = 50 + \frac{10 (x - \bar{x})}{SD}$$

حيث ت= الدرجة التائية المعيارية.

س= الدرجة الخام.

$\bar{س}$= متوسط الدرجات الخام.

ع= الانحراف المعيارى للدرجات الخام.

أما عن الخطوات المستخدمة لاستخراج الدرجة التائية (ت) فهى نفس الخطوات المستخدمة فى استخراج الدرجة الزادية ما عدا المعادلة المستخدمة..، حيث يستخدم هنا معادلة الدرجات التائية.

فى حالة الاختبارات الزمنية تستخدم المعادلة بالشكل التالى:

$$ت = ٥٠ + \frac{١٠ (\bar{س} - س)}{ع}$$

**وفيما يلى أهم مميزات هذه الدرجة:**

١ ـ جميع نواتجها موجبة (لا توجد درجات سالبة).

٢ ـ توفر إمكانية التخلص من كسور الدرجات (قاعدة جبر الكسر).

٣ ـ سهلة الفهم والاستخدام فى بحوث الأداء البدنى.

٤ ـ من أكثر الطرق مناسبة فى حالة عدم انتظام انتشار الدرجات الخام (عدم تحقق المنحنى الاعتدالى).

ويعيب هذه الطريق أن هناك إمكانية لحصول درجتين خام على درجة معيارية واحدة نتيجة لعمليات التقريب، وهى أكثر فى عملياتها الحسابية مقارنة مع الدرجة الزادية.

**٣ـ الدرجة المئينية** Percentile Score:

من أكثر الطرق استخداما فى اختبارات التربية البدنية والرياضة، وهى عبارة عن النسبة المئوية للقيم التى تقع تحت درجة خام معينة.

ويشير فرانك Frank إلى أن مصطلح المئينى Percentile هو عبارة عن القيمة التى تقع دونها نسبة معلومة من الدرجات أو التوزيع التكرارى Frequency Destribution للدرجات، ومثال على ذلك أن ٦٠٪ من الأفراد يقعون تحت القيمة المئينية ٦٠، ٧٥٪ من الأفراد يقعون تحت القيمة المئينية ٧٥.. وهكذا. ومن ثم فإن المئينيات تدل على الوضع النسبى للفرد مقارنة بمجموع أفراد مجموعته.

والرتبة المئينية عبارة عن الدرجة التى تحدد مكان اللاعب بالنسبة لتوزيع مجموعة من الدرجات الخام، وتستخدم لمقارنة أداء الفرد بأداء غيره من المجموعة التى ينتمى إليها لتحديد موقعه بالنسبة للمجموعة.

وفى هذه الطريقة يستخدم القانون التالى: