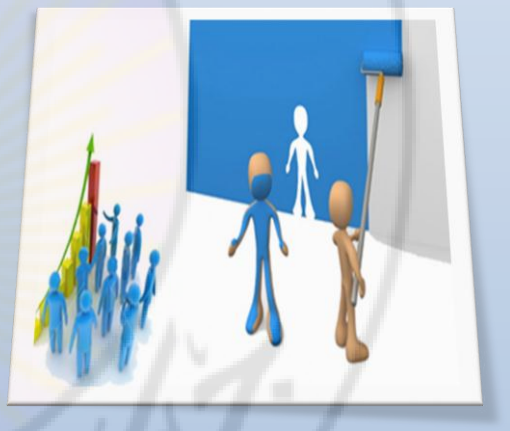
جامعة دمشق كلية العلوم الصحية

أساسيات في القياس والتقويم

د.شذى الميداني



Damascus University

الفصل الأول

مفاهيم أساسية في القياس والتقويم

مقدمة:

إن ما يميز الموضوعية التي يتصف بها أي علم من العلوم هو قدرته على تطبيق منهج القياس وبالتالي قدرته على التنبؤ ومن بعد ذلك التحكم، و يعدّ الإنسان هو المحور الأبرز والأهم لعملية القياس وذلك باعتباره كائن حي يدرك ويتعلّم ويفكر ويتذكر ويتأثرويؤثرفي المجتمع الذي يعيش فيه، ويؤثر في البيئة المحيطة به ويتأثر بها، ويتّخذ منها أداة لتفكيره وقياسه، لذا فإنّنا في أمس الحاجة إلى القياس.

أولاً: القياس التربوي:

١. تعريف القياس:

يشير القياس بشكل عام إلى الجوانب الكمية التي تصف خاصّة أو سمة معيّنة، مثل: ارتفاع سائل، أو حجم كرة، أو درجة حرارة، أو التحصيل الدراسي لطالب، كما يشير إلى عملية جمع المعلومات، وتنظيمها بطريقة مرتبة. وبذلك يتضمن مفهوم القياس من هذا المنظور كلاً من عمليتي جمع المعلومات وتنظيمها، بالإضافة إلى نتيجة هذه العملية.

ويشير مصطلح القياس إلى عملية التقدير الموضوعي الكمي وإعطاء قيم عددية للسمة أو الخاصية المقيسة بالاستناد إلى قواعد محددة تتم من خلالها المواءمة بين السمة أو الخاصية المقيسة والرقم الدال على وجودها. ويستخدم مصطلح التقويم في مجال التربية ليشير إلى العملية التي تسمح باتخاذ حكم حول قيمة الظاهرة التربوية موضع الدراسة.

كما يعرف القياس بأنه عملية وصف المعلومات (وصفاً كمياً)، أو بمعنى آخر استخدام الأرقام في وصف وتبويب وتنظيم المعلومات أو البيانات بطريقة سهلة موضوعية يمكن من فهمها ومن ثم تفسيرها دون صعوبة. ويمكن القول أيضاً أن القياس – كما يعرفه كامبل- إنما هو عملية تحويل الأحداث الوصفية إلى أرقام بناء على قواعد وقوانين معينة، وذلك يعني أن القياس هو عبارة عن عملية تحويل

الوصف الكيفي للظواهر إلى ما هو أسهل من حيث التعامل وأكثر طاعة وقابلية إلى التحويل من حالة إلى أخرى ألا وهو الأرقام.

فمثلاً عند قياس ذكاء الأفراد يصبح تحديد كمية ما يمتلكه كل فرد منهم من هذه السمة أمراً موضوعياً بحيث تكون عملية القياس في هذه الحالة قد اجتازت الخطوة الأولى إلى الخطوة التالية مباشرة وهي التقييم، وبالتالي أصبحت عملية القياس النفسي هي عملية قياس الفروق بين الأفراد في سمة ما أكثر منها كونها عملية قياس كمية ما يملكه كل فرد من هذه السمة، أو تلك التي يشتركون فيها، والهدف منها تحديد الوضع النسبي لكل فرد منهم على هذه السمة.

غير أنّه يمكن تعريف القياس بصورة أكثر دقّة بأنّه: تعيين فئة من الأرقام أو الرموز، بحيث تقابل فئة من الخصائص أو الأحداث، طبقاً لقواعد محدّدة تحديداً جيداً.

ويشتمل هذا التعريف على ثلاثة عناصر هي: فئات الأرقام أو الرموز، وفئات الخصائص أو الأحداث، وقواعد تعيين الأرقام أو الرموز للخصائص أو الأحداث.

ويمكن إعادة صياغة التعريف السابق لكي يناسب محتوى المجال التربوي والنفسي بحيث يصبح كالآتي:

قياس مقدار أو كمية ما يملك الشيء المقّاس من الخصائص المقاسة، وهذا يعني أنَّ القياس التربوي والنفسي يُعنّى باطلاق حكم على خصائص أو سمات الأفراد، حيث إنّنا لا نستطيع قياس الأفراد في ذاتهم، وإنّما نقوم بقياس خصائصهم أو سماتهم، مثال ذلك طول متعلم: نقيس طول الفرد من حيث الخاصية وهي هنا الطول.

٢. أهداف القياس:

للقياس أهداف عديدة ومتنوعة ومن أهمها:

أ- **تحديد الصفات الشخصية والنفسية والعقلية للإنسان وتصنيفها**، بهدف تعرّف مختلف جوانبها وتحديد المتغيرات المتعلّقة بها وذلك من أجل الوصول إلى القوانين التي تحكم السلوك والقدرات العقلية للأفراد، وبالتالي تحدد سلوكهم المجتمعي بمظاهره النفسية والتربوية والاجتماعية.

ب- **الحصول على معلومات تفيد المجتمع بكافة مستوياته (العام والخاص والفردي)**، فالمسؤول في المجال النفسي والمجال الاجتماعي والمجال الطبي وغيرها من المجالات يتوجب عليه بحكم عمله الوفاء بمطالب معينة؛ فهو مطالب مثلاً بتوجيه الفردين وفقاً لقدراتهم وميولهم، ومطالب في الوقت نفسه بتشخيص الحالات غير السويّة لتتلقى ما تتطلّبه من العلاج أو الرعاية.

ت- **الاختيار والتصنيف:** ويقصد به تحديد مستويات الأشخاص في بعض السمات المعيّنة وتصنيفهم وفقاً للمجال المناسب لكل منهم سواء تعلّق ذلك بالنواحي العملية أم التعليمية.

ث- **فحص الأهلية:** ويقصد به تحديد ما إذا كانت تتوافر لدى الفرد القدرة والأهلية والشروط اللازمة لتولي مهمة معينة أو الانخراط في عمل معين مثل الطب والتعليم وغيرهما.

١. عناصر القياس:

تتطلب عملية القياس استخدام أدوات خاصة تتصدى لسمات محددة في الأشياء والظواهر، وتسعى إلى التعبير عنها بلغة الكم، وللقياس ثلاثة عناصر أساسية وهي:

- **موضوع القياس.**
- **العدد.**
- **المقياس.**

- **موضوع القياس:** القياس لا يجري في الفراغ بل يتناول موضوعاً محدّداً يقرر الإنسان قياسه، والقياس لا يتصدى لأشياء أو أشخاص أو وقائع بل يتصدى لسمات وخصائص معينة فيها، فنحن لا نقيس الطاولة ذاتها بل نقيس طولها أو عرضها أو ارتفاعها، ولا نقيس الشخص نفسه بل نقيس طوله أو ذكاءه أو نضجه الانفعالي أو غير ذلك.

ومن المتعارف عليه أنَّ السمات أو الخصائص التي تشكل موضوعات القياس في التربية وعلم النفس مثل التعاونية والقلق والتكيف والتحصيل والذكاء ....الخ، تنطوي على الكثير من التعقد والتداخل، وأنَّ التحديد الواضح والدقيق للسمة موضوع القياس يمثل إحدى المهمات الصعبة التي يواجهها المختص في القياس النفسي والتربوي.

وتتحدد طريقة القياس بطبيعة الموضوع الذي يتم قياسه، فمن الموضوعات ما يمكن قياسه مباشرة (أبعاد منضّدة، قطعة من الأرض، وزن طفل....الخ)، ومن الموضوعات ما يستحيل قياسها بطريقة مباشرة ونلجأ إلى قياسها بطريقة غير مباشرة مثل (الذكاء، التحصيل الدراسي) عن طريق قياس أثارها.

فبعض السمات (القابلة للعزل) يمكن قياسها بسهولة في حين أنَّ بعضها الآخر كالصفات المزاجية وسمات الشخصية(غير المعزولة) من الصعب قياسها بسبب تشابكها وتداخلها مع غيرها من السمات. وفي الغالب يكون القياس المباشر أسهل وأكثر دقة من القياس غير المباشر، كما أنَّ التحديد الدقيق للسمة المراد قياسها يسهل إجراءات عملية القياس ويؤدي إلى نتائج دقيقة وصادقة.

- **العدد:** العنصر الثاني في عملية القياس هو العدد، فالقياس يعتمد على التعامل مع السمة موضوع القياس بلغة العدد، والعدد هو الرمز أو التعبير الكمي الدال على كميّة هذه السمة. ويحقق القياس ميزة هامة من خلال استخدامه العدد حيث أن عملية التقدير الكمي أو العددي توفر قدراً من الدقة لا يمكن توافرها في التقدير الكيفي. ويمكن من خلال ميزات العدد وقابليته للمقارنة، الوصول إلى استنواتج وعلاقات حول الموضوع المقيس يستحيل الوصول إليها بطرائق التقدير الأخرى غير الكمية. ثم إنّ العدد هو لغة الإحصاء ويستحيل دون الإحصاء تفسير نتائج القياس وإعطاؤها دلالتها.

ويحتوي العدد على أكثر من معنى، فقد يستخدم العدد بوصفه عدداً ومرة ثانية بوصفه رقماً (رقم منزل /المنزل السادس/، أو قاعة محاضرات /القاعة الثانية/) ومرة ثالثة بوصفه رتبة أو مرتبة (رتبة تلميذ بين مجموعة من زملائه في امتحان معين /الأول- الثاني- الثالث.../)، كما يمكن استخدامه كرموز للدلالة على أشياء وبدائل عن الأسماء دون أن يكون لها أية دلالة كمية وهنا يكون استخدامها بغرض وصفي، فمثلاً يمكن ترميز متغير الجنس وفق الآتي: (١ تدل على الذكور، ٢ تدل على الإناث).

لذا لا بدَّ من مراعاة قواعد المقابلة بين الأعداد والأشياء للعامل في القياس النفسي والتربوي، ولا بد من إدراك دلالة الأعداد التي يتعامل معها.

- **المقياس:** لكي تتم عملية القياس لا بدَّ من أن نستخدم مقياس أو أداة قياس محددة، بحيث تتوافق مع طبيعة السمة أو الخاصية المقيسة، وفي العصور الحديثة تنوعت أدوات القياس التي يستخدمها الانسان، وتفاوتت في مستوى دقتها ليصل بعضها إلى مستوى عال من الدقة حيث تمَّ تعييرها وفقاً لمواصفات محددة، حيث أن هناك مقاييس تتناول موضوعات مادية وتقيسها بصورة مباشرة كمقاييس الأبعاد والأوزان، وبعضها الآخر يتناول الموضوعات بطريقة غيرمباشرة مثل ميزان الحرارة الذي يقيس الحرارة من خلال الأثر الذي تتركه في عمود الزئبق، وإن معظم المقاييس التي تستخدم في مجال التربية وعلم النفس تتصدى لموضوعاتها بطريقة غير مباشرة.

وفي الواقع أنَّ المنطلق الأساسي المعتمد في بناء المقاييس النفسية والتربوية هو أنَّ عينة من استجابات الفرد يمكن أن تمثل السمة أو الخاصية التي يراد قياسها لدى الفرد بمجموعها الكلي. وانطلاقاً من هذه الناحية يتم وضع المقياس بحيث يشتمل عينة من المثيرات (الأسئلة) التي يفترض أن تؤدي إلى استثارة عينة من الاستجابات المعبرة عن السمة أو الخاصية موضوع القياس، ولكي يكون المقياس صادقاً لا بدَّ أن يشتمل على عينة من المثيرات تمثل سائر المثيرات (المجتمع الأصلي للمثيرات)، كما لا بدَّ أن يستدعي عينة من الاستجابات تمثل كافة الاستجابات (المجتمع الأصلي للاستجابات) التي تمثل السلوك أو السمة المراد قياسها. ويكون الحكم على أداء الفرد من خلال ما يظهر من استجاباته أو إجاباته عن بنود المقياس (عينة الاستجابات) وبالتالي يمكن تعميمه بحيث ينطبق على طائفة الاستجابات كلها التي

أخذت منها تلك العينة وافترض أنها تمثلها. ويكون التعميم من هذه الجهة من طبيعة المقياس التربوي وسمة من السمات الهامة المميزة له.

ويضاف عنصراً رابعاً في العلوم الإنسانية "**التعميم**"، ويقوم على تعميم النتيجة على العناصر المتماثلة، بينما في العلوم الطبيعية فالتعميم يحدث دون السعي لوجوده.

٣. مستويات القياس:

تتفاوت مستويات القياس وفقاً لنوع المتغير وطبيعته والهدف من عملية القياس، فلكي تتم عملية القياس بالدقة المطلوبة يجب مراعاة مستوى قياس المتغير، وتستند هذه المستويات إلى فرضيات رياضية ومنطقية، وتتخذ تنظيماً هرمياً يحقق فيه المستوى الأعلى فرضيات جميع المستويات الأدنى بالإضافةً إلى خصائص تميز المستوى الأعلى، وهذه المستويات:

الشكل (١) المستويات الهرمية للقياس

- **المستوى الاسمي Nominal Scale:** وهو أدنى مستويات القياس ويناسب جميع المتغيرات الكيفية أو النوعية التي تتطلب تصنيف الأفراد إلى مجموعات منفصلة للتمييز فيما بينها في سمة معينة، ويكون الهدف من عملية القياس في هذه الحالة هو التصنيف Classification الذي يراعي الفروق النوعية بين الأفراد. والأعداد المستخدمة في هذا المستوى من القياس تعد بمثابة رموز بسيطة تستخدم كأسماء لفئات أو مجموعات منفصلة ومتمايزة. ومن أمثلة متغيرات هذا المستوى: النوع ("١" ذكور- "٢" إناث)، والجنسية ("١" عربي- "٢" أجنبي)، والحالة الاجتماعية ("١" متزوج- "٢" عازب- "٣" مطلق- "٤" أرمل)، والانتماء إلى مؤسسات معينة، وهكذا.

وهذه الأعداد لا نستطيع إجراء عمليات حسابية عليها بحيث تكون ذات معنى ويمكن إجراء الجمع فقط للنوع الواحد، كما تُستخدم التكرارات والنسب المئوية في هذا المستوى.

- **المستوى الرتبيOrdinal Scale:** وفي هذا المستوى يمكن ترتيب الأفراد أو الأشياء وفقاً لخاصيّة أو سمة معينة. فمثلاً يمكن أن تطلب من الأفراد ترتيب تفضيلاتهم المهنية، إذ ربما يعطى فرد مهنة مهندس معماري ١، ويعطى مهنة محاسب في بنك الرقم ٢، وفني سيارات الرقم٣، وهكذا، وهذا يعني أن هذا الفرد قد أجرى نوع من القياس الرتبي لميوله المهنية وفقاً لمحكّات معيّنة استند إليها في هذا الشأن.

ونظراً لسهولة هذا النوع من القياس، فإنَّه يستخدم بكثرة في القياس التربوي والنفسي وبخاصة في قياس السمات الوجدانية التي تتعلق بالاتجاهات والتفضيلات والآراء والتقديرات. ولكن يجب ملاحظة أنَّه ليس من الضروري معرفة مقدار أو درجات السمة المقاسة لكي نتمكن من إجراء هذا الترتيب، بل يكفي مجرد معرفة ترتيب كل منها بالنسبة للآخر. كما أن القياس الرتبي لا يفترض أن الفروق بين الرتب تتساوى مع الفروق بين درجات السمة المقاسة.

وهذه الأعداد التي تمثل الرتب لا يمكن إجراء عمليات حسابية عليها كما في القياس الاسمي، وإنَّما نكتفي بالقول بأنَّ سامر أذكى من نبيل، أو تحصيل سامر أعلى من تحصيل نبيل، أو أنَّ محمود لديه اتجاه نحو المدرسة أكثر إيجابية من رامي، وهكذا، إضافةً لاستخدام النسب المئوية والتكرارات.

- **المستوى الفئوي Interval Scale:** وهذا المستوى من القياس يتعلّق بتحديد كمية الفرق بين شيئين، فعلى سبيل المثال يمكن حساب الفرق بين درجة متعلم حصل على (٩٣) درجة ومتعلم آخر حصل على (٨١) درجة في مادة العلوم.

وهنا يجب ملاحظة أن الصفر في هذا المستوى ليس له معنى حقيقي، أي أنّ الفرد الذي حصل على درجة (٠) في العلوم لا يعني عدم وجود أي معلومات لديه عن العلوم.

وهذا المستوى من القياس يستخدم كثيراً في القياس التربوي والنفسي، فنحن لا نقيس ذكاء الفرد أو سمات شخصيته أو ميوله قياساً مطلقاً، وإنما نقيس الفروق الحقيقية بين ذكاء شخصين مثلاً، أو نوازن بين ذكائه وذكاء زملائه الذين تم تطبيق نفس الاختبارعليهم، وتمثل الفروق المتساوية على متصل تدريج السمة (الذكاء) فروقاً متساوية في الذكاء بالفعل، ولكن لا يوجد صفر مطلق للذكاء أو لأي سمة أخرى من سمات شخصية الإنسان، فالسمات لا تنعدم لدى الأفراد.

- **المستوى النسبي Ratio Scale**: يتميز هذا المستوى من القياس بالخصائص التي تتوافر في المستوى الفئوي بالإضافة إلى وجود صفر مطلق على ميزان القياس يقابل بالفعل انعدام الخاصة أو السمة المقاسة، ومثال ذلك الأطوال والكتل وغير ذلك. ونظراً لأنه يمكن إجراء جميع العمليات الحسابية الأساسية في هذا المستوى فإنه يعد أعلى المستويات السابقة.

ويستخدم هذا النوع من القياس عادة في العلوم الطبيعية، ويندر استخدامه في العلوم السلوكية للأسباب السابقة، فمعظم الاختبارات والمقاييس التربوية والنفسية تؤدي عادة إلى قياسات فئوية، ولكن يجب مراعاة العمليات والإجراءات التي تمت الإشارة إليها فيما سبق. والجدول الآتي يوضح أمثلة عن مستويات القياس والعمليات الرياضية المستخدمة فيها:

جدول (١) مستويات القياس والعمليات الرياضية المستخدمة فيها

مستوى القياس	أمثلة	العمليات الرياضية
الاسمي	الجنس (ذكر، أنثى) الحالة الاجتماعية (متزوج، عازب، مطلق، أرمل)	العد، النسب المئوية
الرتبي	المؤهل العلمي، الحجم، الرأي.	العد، النسب المئوية، >، <
الفئوي	درجة الحرارة، درجة الاختبار، الذكاء	العد، النسب المئوية، >، <، الجمع، الطرح.
النسبي	الوزن، الطول، الدخل، السعر، الوقت	العد، النسب المئوية، >، <، الجمع، الطرح، الضرب، القسمة.

٤. أخطاء القياس:

على الرغم من المكانة التي يحتلها القياس في عصرنا وما حققته حركة القياس من نجاح وتقدم هائل في كافة المجالات ، وتضم عملية القياس العديد من احتمالات الخطأ نظراً لما يترتب عليها من خطوات، وإجراءات معقدة، وما تتطلبه من الشروط العديدة، وما تتوصل إليه من نتائج يترتب على الاختصاصي في القياس تفسيرها واستخراج دلالاتها، ومن هذه الأخطاء:

❖ **الخطأ في الأداة أو الوسيلة المستخدمة في القياس:**

في بعض الأحيان قد يستخدم في قياس شيء ما أداة قياس غير دقيقة مما ينطوي عليه وقوع خطأ في قياسه، ويظهر هذا الخطأ فيما لو أعيد قياس نفس الشيء بأداة قياس دقيقة.

❖ **الخطأ الناجم عن عدم ثبات الخاصية المقاسة:**

من الملاحظ أن صفة الثبات لا تكون مطلقة في الكثير من الظواهر التي نقوم بقياسها.

❖ **الخطأ الناجم عن الذي يقوم بعملية القياس (خطأ الإنسان):**

حالات الشعور بالتعب، الهدوء أو الاحساس بالأمن عنه وفي حالة الانفعال أو الخوف أو التحيز.

من الملاحظ أن الدقة في القياس تختلف من شخص لآخر في وحدة القياس المستخدمة، مما يشير إلى أن الخطأ الذي وقع فيه بعضهم يعود إلى عدم دقتهم في القياس، من الواضح أن هذا الخطأ من فعل الإنسان نفسه.

❖ **الخطأ الناجم عن النقص في الخبرة والتدريب في مجال القياس:**

إن إعداد المقاييس يتطلب تدريباً خاصاً وخبرة كافية وإلماماً كبيراً للذي يقوم بإعداد المقياس والذي يستخدم المقياس أيضاً، ويؤدي نقص الإلمام والتدريب إلى حدوث خلل فيه مما قد يجعله قاصراً عن تحقيق الهدف المرجو منه، كما قد يكون المقياس صحيح إلا أنّ سوء الاستخدام هو السبب.

١. **الفرق بين القياس والتقويم والتقييم:**

هناك خلط فيما يتعلّق باستخدام هذين المصطلحين ففي بعض الأحيان يستخدم التقويم كمرادف لمصطلح القياس فالقياس ما هو إلا جزء يقدم المعلومات التي يعتمد عليها التقويم.

القياس سابق للتقويم وأساس له ولا يجيب إلا على سؤال واحد هو (كم) وينتهي بإعطاء هذه الدرجة أما التقويم فإنّه يذهب إلى أبعد من ذلك فهو يسعى للإجابة على هذا السؤال (كم؟) وما قيمة هذه الدرجة إذا ما قورنت بالأهداف لإصدار حكم على الشيء أو على الشخص في ضوء نتائج القياس وفي ضوء الأهداف المحدّدة، أي أنّ هناك علاقة بين التقويم من ناحية والقياس من ناحية أخرى، حيث يتّضح كما سبق أنَّ التقويم أعم وأشمل منهما، ويمكن إجمال أبرز الفروق بين القياس والتقويم في النقاط الآتية:

١- يهتم القياس بوصف السلوك، أما التقويم فيحكم على قيمته، وعليه فالقياس يتضمن اهتماماً بالوسائل بغض النظر عن قيمة ما يوصف، أما التقويم فيتضمن اهتماماً بالمعايير، ومدى صلاحيتها، ووسائل تطبيقها، وتقدير أثرها.

٢- يقتصر القياس على التقدير (الوصف) الكمي للسلوك، ما يجعله يعتمد الأرقام في إعطاء النتيجة النهائية للموضوع المقاس أو المراد قياسه، أما التقويم فيشمل التقدير الكمي والنوعي (الكيفي) للسلوك، كما يشمل حكماً يتعلق بقيمة هذا السلوك وعليه فالتقويم أكثر شمولاً من القياس، والقياس يمثل إحدى الأدوات أو الوسائل المستخدمة فيه.

٣- يقدم القياس بعض المعلومات عن الموضوع المقاس، أما التقويم فيعد عملية تشخيصية علاجية في آن واحد.

٤- يعتمد القياس على الدقة الرقمية فقط، أما التقويم فيعتمد على عدد من المبادئ والأسس؛ ما يجعله أعم وأشمل كما ذكر سابقاً، ومن أبرز المبادئ والأسس التي يعتمد عليها التقويم: الشمول، والتشخيص، والعلاج ومراعاة الفروق الفردية، والتنوع في الوسائل المستخدمة.

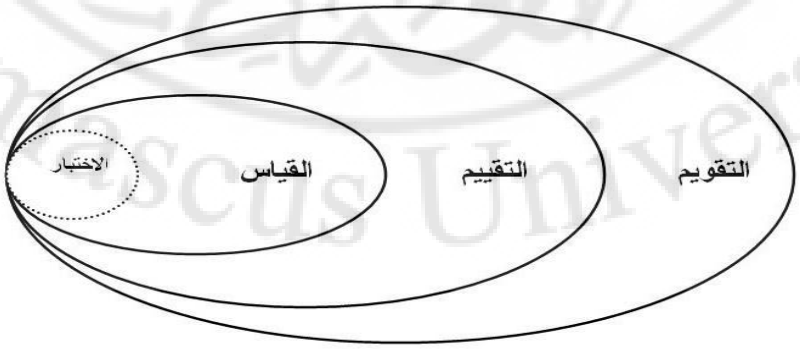
٥- يقتصر القياس على إعطاء وصف للموضوع المراد قياسه دون الاهتمام بالربط بين جوانبه، أما التقويم فيقوم على مقارنة الشخص مع نفسه ومع الآخرين.

٦- القياس أكثر موضوعية من التقويم، ولكنه أقل منه قيمة من الناحية التربوية؛ نظراً لأنّ معرفة النتائج بدقة وموضوعية من غير تقدير لقيمتها لا يعني شيئاً، أما إذا فسّرت تلك النتائج وقدرت قيمتها في ضوء معايير محددة، واتخذت نتائج هذا التقويم كأساس لمساعدة الفردين على النمو، فإنّها تصبح ذات فائدة كبيرة، وهذا ما تهدف له عملية التقويم التربوي.

٧- أما التقييم فله معنى أكثر محدودية من التقويم، ولكنه أكثر اتساعاً من القياس، فالتقييم هو عملية جمع البيانات وتشكيلها في صورة قابلة للتفسير، ومن ثم إصدار أحكام استناداً إلى هذا التقييم. فالتقييم يتطلب العناية بتحليل المعلومات المستمّدة من مختلف الأساليب وأدوات القياس، ويسهم في صنع قرارات ملائمة، بالاعتماد على الطرائق الأفضل للتحليل والتفسير اعتماداً على الهدف أو الغرض من التقييم.

٨- أي أنّ المعلّم يجري تقييمات لجمع معلومات عن تحصيل كل طالب من أجل اتخاذ قرارات تربوية مهمة، وبالطبع يختلف نوع المعلومات التي يسعى إليها المعلّمون لتوجيه العملية التدريسية، وتنظيم تعلم طلابهم بشكل مستمر عن المعلومات المطلوبة لتقويم المدرسة أو البرنامج التعليمي. فالتقويمات تخدم أفراداً مختلفين، وسياقات مختلفة، مما يُبرز الحاجة لمعلومات مختلفة

٩- وعلى الرغم من أنّ مفهوم التقييم من المنظور التاريخي شاع استخدامه في التعليم العلاجي والفئات الخاصة، إلا أنّه أصبح يستخدم بكثرة في وقتنا الحاضر في الصفوف المدرسية العادية بعد ما انتشر استخدام ما أصبح يطلق عليه التقييم البديل أو الواقعي Assessment Authentic في كثير من الأنظمة التعليمية في الدول المتطوّرة. والشكل الآتي مثالاً توضيحياً للعلاقة بين التقويم والتقييم والقياس والاختبار:



تمثيل تخطيطي بصري للعلاقة الجزئية بين مصطلحات القياس و التقييم و التقويم

شكل (٢) العلاقة بين مصطلحات القياس والتقييم والتقويم

٢. تصنيف أدوات التقويم وشروط صلاحيتها:

تعددت أدوات القياس والتقويم نتيجة لتعدد الأهداف والخصائص والسمات التي تقيسها، مما جعل المهتمون بها يشعرون بوجود خصائص مشتركة من ناحية معينة، وبالتالي تصنيفها وفق هذا التشابه، ومن أهم هذه التصنيفات:

تصنيف حسب طبيعة الأداء:

ويعود هذا التصنيف إلى كرونباخ، حيث صنفها في فئتين هما:

أقصى الأداء: وهي الأدوات التي يتم فيها إثارة دافعية المتعلم لتقديم أفضل ما عنده من إجابة، والحصول على أعلى علامة ممكنة، مثل اختبارات التحصيل، الاختبارات التي تقيس القدرات العقلية والمعرفية والمهارية.

الأداء العادي أو الطبيعي: وهي الأدوات التي تعكس سلوك المتعلم في الأوضاع العادية أو الطبيعية دون أي محاولة خارجية لتوجيه هذا السلوك، فمثلاً إذا كانت الأداة تقيس قلق المتعلم في الامتحان فإن القرارات التي يتم اتخاذها في ضوء النتائج تعتمد على صدق الإجابات لذلك تعتبر مقاييس الاتجاهات والميول والتكيف الاجتماعي من نوع الأداء العادي.

تصنيف حسب التصحيح ووضع الدرجة:

يعود هذا التصنيف إلى جلاسر الذي صنفها في فئتين هما:

معيارية المرجع: حيث يقارن أداء المتعلم على الاختبار بأداء مجموعته المعيارية، فقد تكون المجموعة من طلاب صفه، أو من هم في نفس المستوى الأكاديمي أو العمري محلياً أو عالمياً، كأن تفسر علامة المتعلم في الرياضيات مثلاً على أنه أعلى تحصيلاً من ٨٠% من طلاب صفه في مادة الرياضيات، وقد تفسر العلامة أو يتم الحكم على أداء المتعلم في امتحان معين من خلال موقع علامته بالنسبة للمتوسط الحسابي لعلامات الصف، فالمتوسط الحسابي هو المعيار.

محكية المرجع: حيث يقارن أداء المتعلم بمستوى أداء معين يتم تحديده بصرف النظر عن أداء المجموعة كأن يجيب المتعلم عن ٨٠% من أسئلة الاختبار على الأقل، ويستخدم هذا النوع من التفسير لنتائج القياس في مرحلة التعليم الأساسي ورياض الأطفال، مع ملاحظة تزايد الاهتمام به في جميع المراحل بسبب الاهتمام بالتعليم للوصول إلى حد الكفاية إذ يتوقع من المتعلم الوصول إلى حد معين كأن يحقق ٩٠% من الأهداف، أو يجيب عن ٨٥% من الأسئلة.

تصنيف حسب شكل (نوع) الفقرة:

هناك عدة أشكال للفقرات:

المطابقة (المزاوجة): تكون الفقرة على شكل جدول بعمود في أحدهما عمود المقدمات والثاني عمود الإجابات ويطلب من المتعلم أن يقوم باختيار الإجابة المناسبة للمقدمة بأسلوب يقترحه المعلم ويحدده في التعليمات بحيث تراعى سهولة التصحيح وأعمال المتعلمين.

الاختيار من بديلين: مثل (صواب، خطأ)، (نعم، لا)، (موافق، معارض)، (مرضي، غير مرضي).

الاختيار من متعدد: حيث يقدم للمتعلم ثلاثة بدائل أو أكثر وعليه أن يختار الإجابة الصحيحة وقد يختلف عدد البدائل باختلاف عمر المتعلم ومستواه الأكاديمي، تصنيف وطبيعة المادة أو نوع السمة المقاسة.

التكميل: حيث يطلب من المتعلم إكمال الجملة بوضع كلمة أو شبه جملة في الفراغ، أوتكميل شكل أو عدة كلمات.

الإجابة القصيرة: هي الأسئلة التي يطلب فيها من المتعلم أن يعطي الإجابة بكلمة أو جملة أو بضع جمل مثل تعريف المصطلحات، معاني المفردات، تسمية مواقع، وتأتي الإجابة القصيرة كحلقة وصل بين التكميل والأسئلة الإنشائية ذات الإجابة المحددة.

الإنشائية (المقالية): تختلف في درجة الإجابة وتتضمن الأسئلة الإنشائية المحددة المشار إليها أي أن حرية المتعلم في الإجابة مقيدة بطبيعة السؤال نفسه، وبالتالي يسهل تحديد زمن الإجابة وحجم الفراغ المتروك، كما تتضمن الأسئلة الإنشائية المفتوحة وهي الأسئلة التي لا يستطيع المتعلم معرفة متى يتوقف عن الإجابة عليها.

تصنيف حسب طريقة الإجابة:

يعد هذا التصنيف مماثلاً لتصنيف حسب شكل الفقرة وتقسم إلى فئتين:

الفئة التي يقوم بها المتعلم بانتقاء الإجابة: وهي الاختبارات التي يكون شكل فقراتها من نوع المطابقة، الصواب والخطأ، الاختيار من متعدد.

الفئة يقوم فيها المتعلم بصياغة الإجابة: تتضمن الفقرات ذات الإجابة المحددة مثل التكميل، الإجابة القصيرة، المسائل، الفقرات ذات الإجابة المفتوحة.

تصنيف حسب موضوعية التصحيح (حسب درجة تأثر العلامة بذاتية المصحح):

قد تتدخل ذاتية المصحح بالتأثر على درجة الاختبارات في حال التصحيح اليدوي من قبل المعلم فقد يكون التصحيح ذاتي.

أو يكون التصحيح موضوعي من خلال التصحيح الآلي أو بمفتاح مثقب كما يجري في أسئلة الاختيار من متعدد.

تصنيف حسب الجهة التي تعد الاختبار:

اختبارات من إعداد المعلم: وهي اختبارات غير رسمية يقوم المعلم بإعدادها بنفسه.

اختبارات مقننة (رسمية): حيث يعدها فريق من المختصين في مراكز القياس والاختبارات.

تصنيف حسب فترة الإجابة (هبكنزوآننس):

اختبارات قصيرة: قد تجري في نهاية الدرس أو الحصة أو الوحدة الدرسية.

امتحانات: وتجري في نهاية الفصل الدراسي أو السنة الدراسية.

تصنيف حسب سرعة الإجابة:

اختبارات السرعة: وتكون فيها سرعة الإجابة العامل الحاسم في تحديد أداء المتعلم، إذ يجب أن يتم إنهاء الإجابة على جميع فقرات الاختبار خلال مدة زمنية محددة.

اختبارات القوة: ويعطى المتعلم فيها زمناً مفتوحاً للإجابة بحيث يكون المتعلم قادراً على محاولة الإجابة عن كل سؤال، إلا أن صعوبة الأسئلة وقدرة المتعلم على الإجابة عليها هي التي تحدد أدائه بمعنى أن المتعلم لا يحصل على علامة كاملة بسبب صعوبة الأسئلة وليس بسبب كثرتها.

تصنيف حسب درجة المثير والاستجابة:

اختبارات إسقاطية: يكون فيها المثير والإجابة غير محددين كالاختبارات النفسية التشخيصية مثل اختبار رورشاخ (بقعة الحبر).

اختبارات محددة البناء: ويكون فيها المثير واضحاً، والمطلوب في السؤال أيضاً محدداً وهناك مفتاح إجابة محدد كاختبارات التحصيل والاستعداد أو اختبارات القدرات العقلية، والاختبارات التحصيلية هي من هذا النوع.

تصنيف حسب عدد الأفراد الذين يطبق عليهم الاختبار:

الاختبارات المتعلمية: وتطبق على فرد واحد في المرة الواحدة مثل الاختبارات الشفوية أو بعض اختبارات الذكاء مثل اختبار ستانفورد بينيه.

الاختبارات الجمعية: يمكن تطبيقها على عدد كبير من المتعلمين في الوقت نفسه.

تم عرض طرائق متعددة لتصنيف أدوات القياس والتقويم، ولابدّ من الإشارة إلى أن الأداة الواحدة قد تتبع لأكثر من تصنيف. فاختبار التحصيل يمكن أن يكون اختبار قوة، ويعمل كعينة، ويفسر تفسيراً محكي المرجع، ومن نوع أقصى الأداء.

تصنيف حسب سرعة الإجابة:

أ- **اختبارات السرعة:** وتكون فيها سرعة الإجابة العامل الحاسم في تحديد أداء الفرد، إذ يجب أن يتم إنهاء الإجابة على جميع فقرات الاختبار خلال مدة زمنية محددة.

ب- **اختبارات القوة:** ويعطى المتعلّم فيها زمناً مفتوحاً للإجابة بحيث يكون المتعلّم قادراً على محاولة الإجابة عن كل سؤال، إلا أن صعوبة الأسئلة وقدرة الفرد على الإجابة عليها هي التي تحدد أداءه، بمعنى أن المتعلّم لا يحصل على علامة كاملة بسبب صعوبة الأسئلة وليس بسبب كثرتها.

تصنيف حسب درجة المثير والاستجابة:

أ- **اختبارات إسقاطية:** يكون فيها المثير والإجابة غير محددين كالاختبارات النفسية التشخيصية مثل اختبار رورشاخ (بقعة الحبر).

ب- **اختبارات محددة البناء:** ويكون فيها المثير واضحاً، والمطلوب في السؤال أيضاً محدداً وهناك مفتاح إجابة محدد كاختبارات التحصيل والاستعداد أو اختبارات القرات الفعلية، والاختبارات التحصيلية هي من هذا النوع.

تصنيف حسب الكيفية التي يظهر فيها الأداء (أسلوب تقديم الإجابة):

أ- **اختبارات لفظية:** سواء كانت تحريرية (ورقة _ قلم) أو شفوية.

ب- **اختبارات الأداء المبرهن عملياً:** تتم في المختبرات والمشاغل العلمية.

تصنيف حسب درجة تحديد المجال:

أ- **الاختبار كعينة:** وهو الاختبار الذي اختيرت فقراته من جميع الفقرات المحتملة، التي تغطي المجال، ويحدث هذا عندما يكون المجال محدداً مثل اختبارات التحصيل المدرسية.

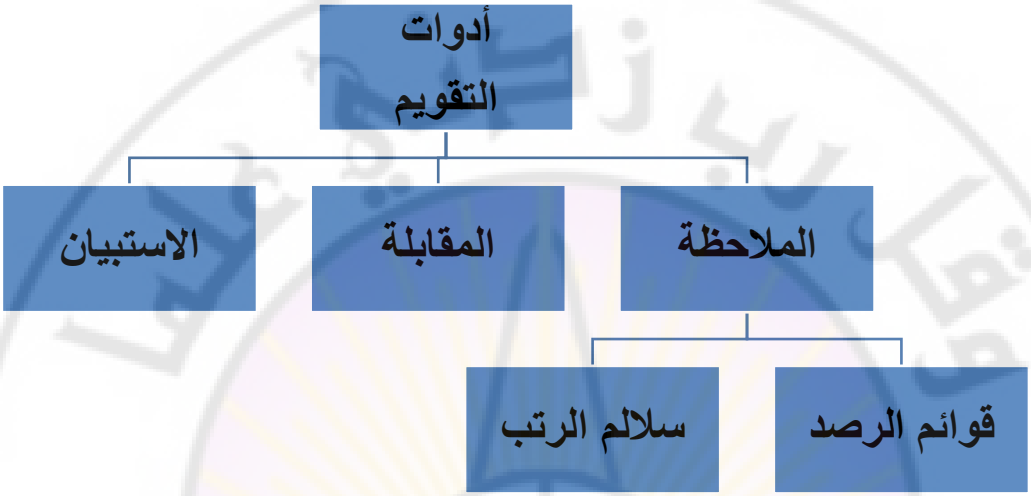
ب- **الاختبار كمؤشر:** هو الاختبار الذي اختيرت فقراته بشكل لا نستطيع القول بأنها تمثل المجال لأننا لا نستطيع أن نحدد جميع الفقرات المحتملة، ويحدث هذا عندما يكون المجال مفتوحاً مثل اختبار الذكاء. فالذكاء من السمات التي يصعب تحديد مجالها، كما لا يوجد اتفاق حول تعريفه مفهوم الذكاء، لذا يوجد هناك اختبارات مختلفة في القدرات لكنها تسمى اختبارات ذكاء. بمعنى أن المجال مفتوح لإضافة فقرات تقيس قدرات جديدة إلى أي اختبار في الذكاء.

تم عرض طرائق متعدّدة لتصنيف أدوات القياس والتقويم، ولابدّ من الإشارة إلى أن الأداة الواحدة قد تتبع لأكثر من تصنيف، فاختبارات التحصيل يمكن أن يكون اختبار قوة، ويعمل كعينة، ويفسر تفسيراً محكي المرجع، ومن نوع أقصى الأداء.

الفصل الثاني

أدوات التقويم

(الملاحظة – المقابلة – الاستبيان)



أولاً: الملاحظة

تعريف وفوائد الملاحظة:

التعريف:

يعتمد التقويم بالملاحظة على جمع المعلومات عن سلوك الفرد و وصفه وصفاً لفظياً ، وهو من أنواع التقويم النوعي Qualitative ، تدون فيه سلوكات الفرد من قبل المدرب أو مشرف التدريب ، أو الأقران أو صاحب القرار. إن هذا النوع من التقويم يتطلب تكرار الملاحظة فترة زمنية محددة ، وتنويع مصادر المعلومات ، للمساعدة في التعرّف على اهتمامات ، وميول واتجاهات المتدرّبين ، وتفاعلهم الاجتماعي مع زملائهم .

تُعطى الملاحظة دلائل مباشرة عن تعلم المتدرّبين ، وتشمل ما يعملون وما يستطيعون عمله وما لا يستطيعون عمله ، حيث توفر هذه المعلومات الفرصة للمدرب لوضع خطة لاستثمار قدرات المتدرّبين والبدء بتعزيز نقاط القوة لديهم .

التعريف الإجرائي

عملية يتوجه فيها الملاحظ بحواسه المختلفة نحو المفحوص ، بقصد مراقبته في موقف نشط ، وذلك من أجل الحصول على معلومات تفيد في الحكم عليه ، وفي تقويم مهاراته وقيمه وسلوكه وأخلاقياته وطريقة تفكيره

الفوائد:

- لا تهدد الأفراد ، وتزود الفاحصين بمعلومات تعجز وسائل التقويم الأخرى عن تزويدهم بها .

- تعطي الملاحظة اليومية للفاحص صورة جيدة عن تطورهم .

- وسيلة فعّالة وخاصة في المنهج التدريبي متعدّد مصادر التعلم ، حيث تعطي المتدرّبين مجالاً لاختيار المواقف التعليمية التي تناسب قدراتهم وميولهم ورغباتهم مما يتيح مجالاً واسعاً للملاحظة .

- تعطي فرصة للفاحص لتهيئة الجو لإيجاد مواقف تعلم يمكن عن طريقها ملاحظة بعض الاتجاهات والمهارات لدى المفحوصين .

- اكتشاف المشكلات حال ظهورها والعمل على حلها قدر الإمكان .

- تقديم تغذية راجعة فورية للمفحوصين .

- دراسة شخصية المفحوص.

أنواع الملاحظة وخصائصها وخطوات تصميمها:

الأنواع:

يمكن تقسيم الملاحظة إلى أنواع عديدة من أهمها :

١- **الملاحظة التلقائية :** عبارة عن صور مبسطة من المشاهدة والاستماع بحيث يقوم الملاحظ فيها بملاحظة السلوكات كما تحدث تلقائياً في الموقف الحقيقة .

٢- **الملاحظة المنظمة :** وهي الملاحظة المخطط لها مسبقاً والمضبوطة ضبطاً دقيقاً ، ويحدد فيها ظروف الملاحظة كالزمان والمكان والمعايير الخاصة للملاحظة .

الخصائص:

١- الحصول على معلومات عن بعض نتاجات التعلم التي لا يمكن توفيرها بواسطة طرائق التقويم الأخرى .

٢- توفير معلومات كمية ونوعية عن نتاجات التعلم مما يوفر درجة عالية من الثقة عند اتخاذ القرارات التدريبية ، وتوفر كذلك نوعاً من الشمولية في تقويم النتاجات التعليمية .

٣- تتمتع بمرونة عالية بحيث يمكن تكييفها أو تصميمها بما يتناسب مع النتاجات التعلمية المختلفة ومع المراحل العمرية المختلفة .

٤- توفير معلومات عن قدرات الفرد في مواقف حقيقية ، توفر فرصة للتنبؤ بتقدم الفرد ونجاحه في مهنته في المستقبل .

خطوات التصميم:

١- تحديد الغرض من الملاحظة .

٢- تحديد نتاجات التعلم المراد ملاحظتها .

٣- تحديد الممارسات والمهمات المطلوبة ومؤشرات الأداء .

٤- ترتيب الممارسات والمهمات ومؤشرات الأداء في جدول حسب تسلسل منطقي .

٥- تصميم أداة تسجيل لهذه الممارسات والمؤشرات مثل (سلم تقدير ، قائمة رصد ، ..)

أدوات الملاحظة:

١-قوائم الرصد

٢-سلالم التقدير

١) قوائم الرصد:

هي قوائم تشتمل على المكونات أو العناصر التي يتم تقديرها في عملية أو نتاج معين، ويقوم الفاحص بملاحظة كل من هذه العناصر أثناء أداء المفحوص للعملية ،أو في النتاج النهائي لتحديد ما إذا كانت العملية المعينة أو النتاج تحقق محكات الأداء كل على حده، ويضع المعلم علامة(صح)أمام المحك الذي تحقق، وذلك دلالة على انه تم ملاحظة هذا المحك ، و أنه متوافر بدرجة مرضية. وقد تتطلب قوائم الرصد الإجابة بنعم أو لا على كل عنصر من عناصرها.

والدرجة التي تقدر لكل محك هي عدد العلامات التي وضعت أمامه.

خطوات بناء قوائم الرصد:

- تحديد نواتج التعلم أو مؤشرات الأداء في المجال العام المراد قياسه.
- تحدد عينة من المواقف التي تتمثل فيها هذه النواتج.
- تحدد عينة من السلوك في هذه المواقف.
- تحدد أمام كل سلوك مقاييس تقدير (نعم / لا).
- يؤشر المقوم على مستوى أداء المتعلم طبقاً لما يلاحظه من مستوى جودة.
- تجمع الدرجات المقابلة لكل خطوة، حيث يقابل المستوى نعم الدرجة (١) ويقابل المستوى لا الدرجة (٠)، وتحتسب درجة المفحوص وفق هذا التدريج.

وينبغي تحديد درجة النجاح على القائمة كما يتفق عليها الخبراء وذلك قبل إجرائها، ولا يقتصر استخدام قائمة الملاحظة على تقويم أداء المفحوص للمهارة بل تستخدم أيضاً في تعلمها والتدريب عليها.

مثال لقائمة رصد:

فيما يأتي مثال لقائمة رصد، يمكن أن يستخدمها الفاحص لملاحظة مشاركة المفحوص في المناقشة.

جدول (١) قائمة رصد مشكلة التأتأة

العمليات	نعم	لا
١- يشعر بصعوبة في التعبير عن أفكاره		
٢- يعبر بالإشارة بدلاً من التعبير بالكلام		
٣- يشعر بصعوبة في التحدث مع زملائه		
٤- يشعر بالقلق والتوتر عند التحدث		
٥- يتحدث ببطء شديد		
٦- يكرر المقطع الأول من الكلمة		
٧- يكرر الكلمة الأولى من الجملة		

٢)سلالم الرتب :

تعد هذه السلالم من الطرائــق التحليلية في تقدير مكونات الأداء كل على حدة، بحيث لا يؤثر تقدير أي مكون من مكونات الأداء على تقدير بقية المكونات، ويتم تقدير كل مكونة على سلم تقدير يشتمل على نقاط متدرجة وليست ثنائية كما في قوائم الشطب.

سلم التقدير: أداة بسيطة تظهر فيما إذا كانت مهارات المتعلم متدنية أو مرتفعــة، حيث تخضع كل فقرة لتدريج من عدة فئات أو مستويات، حيث يمثل أحد طــرفيه انعدام الصفة أو وجود الصفة التي نقدرها بشكل ضئيل ويمثل الطــرف الآخر تمام أو كمال وجودها، وما بين الطرفين يمثل درجات متفاوتة من وجودها.

خطوات إعداد سلم التقدير:

١- تجزئة المهارة أو المهمة إلى مجموعة من المهام الأصغر، أو إلى مجموعة من السلوكات المكونة للمهارة المطلوبة.

٢- ترتيب السلوكات المكونة للمهارة المقاسة حسب تسلسل حدوثها أو بحسب تنفيذها من قبل المتعلمين.

٣- اختيار التدريج المناسب على سلم التقدير لتقدير مدى إنجاز المهارة، وذلك وفقاًلطبيعة المهارة ومجموعة السلوكات المتضمنة فيها والتي سيؤديها المتعلمون.

توجد أنواع متعددة من سلالم التقدير لعل أكثرها شيوعاً ما يلي:

سلالم التقدير الرقمية: يستخدم هذا النوع من السلالم في التقدير العددي أو الرقمي لمكونات عمليات أو نتاجات يقوم بها الطالب، وتصاغ هذه المكونات في عبارات بسيطة، بحيث يمكن ملاحظة السلوك الذي تشير إليه، ويلي كلا منها قيم تتراوح بين ٥ ، ١ مثلاً.

مثال١: سلم تقدير رقمي لمشكلة المصاداة وترديد الكلام لدى الطفل التوحدي

المهارة	بدرجة كبيرة جداً (٤)	بدرجة كبيرة (٣)	بدرجة متوسطة (٢)	بدرجة ضعيفة (١)
يكرر السؤال الموجه له.				
يردد الجمل البسيطة مع عكس الضمائر.				
يصدر أصوات عديمة المعنى.				
يفقد اللغة التي اكتسبها.				
يقلد الأصوات.				
يردد الكلمات دون فهم لمعناها.				
يصعب عليه استخدامن كلمات جديدة.				

سلالم التقدير البيانية Graphic Ratings:

تستخدم أوصاف معينة بدلا من الأرقام في تقدير مستويات جودة العمليات والنتاجات .

مثال: ضع علامة (×) على الخط عند النقطة التي تبين درجة تكرار بعض العبارات المتعلقة بمهارة اللغة الاستقبالية للعمر من سنتين إلى سنتين ونصف.

- يفهم كافة تراكيب الجملة.

_____

دائماً غالباً أحياناً نادراً

- يستجيب للأوامر المتضمنة على – تح – فوق - أسفل .

_____

دائماً غالباً أحياناً نادراً

- يستجيب للأوامر باستعمال أمرين بسيطين.

_____

دائماً غالباً أحياناً نادراً

- يظهر الاهتمام في شرح لماذا وكيف؟

دائماً غالباً أحياناً نادراً

سلالم التقدير الوصفية/ اللفظية:

يتم فيه صياغة أوصاف تعكس مستويات مختلفة من الأداء الفعلي الذي يقوم به الطالب ويقوم المدرس باختيار الوصف الذي ينطبق على الطالب، و يعد هذا النوع من أهم السلالم وأكثرها استعمالا ودقة في تقويم الأداء.

خطوات تصميم سلالم التقدير اللفظية:

١- وصف وبناء تصور للعمل الجيد.

٢- تحديد المعايير التي تمثل خصائص العمل الجيد.

٣- وصف مستويات الأداء المطلوب تقويمها.

٤- ناقش المعايير والمستويات مع المختصين وعدلها إن تطلب الأمر ذلك.

٥- صحح القائمة النهائية بالمعايير والمستويات.

مثال١: ميزان وصفي/ لفظي لبعض مكونات مهارة الاستماع للصف الثاني الإبتدائي

الأداء الدال على المهارة	ممتاز(٥)	جيد جدا(٤)	جيد(٣)	مقبول(٢)	ضعيف(١)
يفهم النص المسموع فهماً مجملاً دقيقاً	يحد الفكرة الأساسية من النص بشكل دقيق	يذكر معنى قريبا جدا للفكرة الأساسية للنص	يذكر معنى مقبولاً للفكرة الأساسية للنص	يذكر معنى يدور حول الفكرة الأساسية	لا يستطيع فهم النص وتحديد الفكرة الأساسية
يراعي مبادئ الاستماع	يحسن الانتباه والإنصات وعدم المقاطعة، ولكنه يقاطع المتحدث عند حدوث خطأ لا يمكن السكوت عنه	يحسن الانتباه والإنصات ولا يقاطع المتحدث أبداً	يحسن الانتباه والإنصات، إلا أنه غالباً يقاطع المتحدث.	يعرف مبادئ الاستماع، لكنه لايحسن الانتباه دائماً يقاطع المتحدث	لا يراعي مبادئ الاستماع

ثانياً: المقابلة :

وهي لقاء يتم بين الباحث والمجيب وعلى أساس ذلك يحدد هدف واضح للبحث ويعد الباحث استمارة خاصة لجمع المعلومات من العينات وهذه الطريقة تصلح لكافة المستويات التعليمية والثقافية.

إن المقابلة تستخدم للتعرف على الحقائق والتأكد من المعلومات بشكل دقيق من قبل الباحث مباشرة، والمقابلة هي مهمة لكشف المواقف الاجتماعية والتوصل إلى الاتجاهات والقيم الإنسانية والمعلومات التي تعطي حلا للمشكلة والهدف

من المقابلة اختبار الفروض وترجمة البحث والكشف عن استجابات المجيبين وذلك لان المقابلة محادثة هادفة مع تهيئة السؤال لمستوى المجيب مع ترتيب الأسئلة بشكل مناسب منسجم مع الحديث.

أنواع المقابلات:

١- المقابلة الفردية والجماعية:المقابلة الفردية هي التي تعطي للفرد الحرية بالإدلاء بآرائه أما المقابلات الجماعية فإنها تعطي للأفراد الفرصة للإدلاء بآرائهم بحرية تامة.

٢- المقابلة المقيدة:وتجري هذه المقابلة تحت نظام خاص مقيد بالأسئلة التي حددت مع الالتزام بالأجوبة المحددة وغالبا ما تهيئ استمارة خاصة بها.

٣-المقابلة الغير مقيدة:وهي المقابلة التي لا توضع لها قيود وتكون مفتوحة وغير رسمية وتعدل الأسئلة حسب ظروف المقابلة وهي مرنة.

٤-مقابلة التعمق:وهي أن يسمح الباحث للمجيب التحدث بحرية كاملة والباحث يستمع ويثني على المجيب ويدفعه للاستمرار بالحديث ليصل الباحث إلى كل المعلومات المطلوبة.

٥-المقابلة المركزة: وهي التركيز على خبر معلوم محسوس ويسعى الباحث لمعرفة الآثار المترتبة عليه.

تصنيف المقابلات:

ا-**المسحية:** وتستخدم لمسح آراء الرأي العام أو مسح الاتجاهات والميول ولتحديد آراء الأفراد بالضبط.

ب-**تشخيصية:** تشخص حالة ويبدأ البحث عن هذه الحالة.

ج-**العلاجية:** وهذه الحالة تستعمل لفهم المجيب ومن ثم وضع العلاج له.

د- **القابلة التوجيهية:** وهي إعطاء الأفراد إرشادا أو توجيها معينا من اجل حل للمشاكل المتعلقة بهم.

هـ-**المقابلة الاختبارية والقياسية:** وهي اختبار أو اقتباس حالة المجيب النفسية وفحص قدراته وكفاءته وتخصص استمارة لذلك.

الفصل الثالث

بناء الاستبانة وتوفير مستلزمات صلاحها

مقدّمة:

الاستبانة هي إحدى الأدوات الهامة في عملية البحث والقياس النفسي والتربوي يستخدمها المربّون على نطاق واسع بهدف جمع المعلومات والبيانات بطريقة منظمّة ودراسة المواقف والميول والآراء ومظاهر التكيّف الشخصي والاجتماعي. ولقد عرفت الاستبانة كطريقة في طرح الأسئلة منذ زمن بعيد، ويعود الفضل إلى علماء الاجتماع في القرن العشرين من أمثال هايمان وستوفر في الإلحاح على أهميتها وضرورة إخضاعها للتحليل المنطقي والإحصائي وتقنينها واستخدامها كأداة علمية لجمع المعلومات. وتفيد الاستبانة في دراسة مختلف جوانب العملية التربوية، وتحتل بالتالي مكانة خاصة في الدراسات والبحوث التربوية بأنواعها وخاصة منها البحث الوصفي.

وقد أشار لازارسفيلد وسيبر إلى أن ثلث البحوث التربوية المنشورة عام 1963 في اربعين مجلّة تربوية اعتمدت على الاستبانة كأداة بحث (LAZARSFELD AND SIEBER,1964). وذكر فان دالن أن "الاستبانة في بعض الدراسات، أو في مراحل معيّنة منها، هي الأداة العملية الوحيدة لاستخلاص المعلومات والبيانات المطلوبة لتأكيد الفرضية المطروحة أو رفضها." (VANDALEN, 1973, P.324).

وسوف نسعى في هذا الفصل إلى إلقاء الضوء على الاستبانة وموقعها في البحث وفي عملية القياس والتقويم التربوي وفاعليتها كأداة بحث وقياس، كما سنسعى إلى معالجة طائفة واسعة من المسائل المتّصلة بشكلها ونوعها وعملية تصميمها وتوفير مستلزمات صلاحيتها والتثبّت من هذه الصلاحية بما يمكن أن يحقّق بعض الفائدة لمن يعمل على إعداد استبانة تتحقّق فيها الشروط

شكل الاستبانة ونوعها:

تستلزم عملية تصميم الاستبانة وإعداد الأسئلة اللازمة لها أن يتّخذ الباحث قراراً بصدد شكلها (الشكل المغلق أو المفتوح أو المغلق – المفتوح)، وأن يقرّر ما إذا كان سيطبّقها على المبحوثين (المستجيبين) مباشرةً أو أنه سيرسلها لهم بالبريد (أو سينشرها على صفحات الصحف والمجلاّت أو على شاشة التلفزيون أو عن طريق الإذاعة أو الهاتف كما يحدث في حالات معيّنة). ولكل من هذه الأشكال والأنواع مزاياه وعيوبه. ويتحدّد اختيار الباحث لشكل الاستبانة ونوعها في ضوء مجموعة من المتغيّرات منها طبيعة الدراسة ومستلزماتها وأهدافها والأغراض الخاصة بالاستبانة، وإمكانات الباحث الواقعية، ونوع العيّنة المختارة وحجمها وتوزّعها الجغرافي... الخ.

تستخدم الاستبانة (غير البريدية) والتي يتولّى الباحث أو أحد مساعديه تطبيقها على المبحوثين في الحالات التي يمكن فيها جمع المبحوثين في مكان واحد وتوزيع الاستمارات عليهم كما هو الحال بالنسبة للطلبة في المدارس. ويشغل هذا النوع حيّزاً هاماً في البحوث التربوية ويتمتّع بمزايا هامة منها أنه يؤمّن الاتصال الحي والمباشر بين الباحث والمبحوثين مما يمكّن الباحث من شرح أهداف الدراسة ومغزاها وفائدتها وتوضيح النقاط الخافية عليهم، كما يمكّنه من حفز المبحوثين وتحميسهم واستثارة الدافعية لديهم للإجابة بعناية وأمانة، هذا بالإضافة إلى أنه يتيح تفحّص استمارات الاستبانة للتأكّد من تعبئتها تماماً من قبل المبحوثين كافة. وتتميّز الاستبانة غير البريدية بأنها قليلة التكاليف وأن نسبة الرّدود عليها تكون عالية جداً وأن وجود الباحث بنفسه (أو من ينوب عنه)

وتأكيده سرية البيانات يخفّف من قلق المبحوثين وشكوكهم، ممّا يؤدّي إلى رفع مستوى الصدق في استجابات المبحوثين (صدق الاستجابة) ويرفع بالتالي من مستوى صدق الاستبانة ككل. وتتطلّب هذه الطريقة تواجد شخصين أو أكثر للإشراف على التطبيق (بحسب حجم المجموعة ومستوى ثقافتها) وتقديم العون اللازم بطريقة غير موحية، وتفحص الاستمارات للتأكّد من تعبئتها تماماً (Oppenheim, 1966).

وتتفوّق الاستبانة البريدية على غير البريدية في إمكان تطبيقها على عيّنة كبيرة من الأفراد ومنتشرة في رقعة جغرافية واسعة وعلى مسافات متباعدة. وهي اقتصادية نسبياً وتنحصر تكاليفها في الطباعة والإرسال البريدي ويمكن تعميمها على إعداد كبيرة مع زيادة متواضعة في التكاليف. ثم إن عملية المعالجة والتحليل لنتائج هذه الاستبانة يسيرة وسهلة بالمقارنة مع طريقة المقابلة ولا تحتاج إلى طاقم مدرّب من الأشخاص العاملين في الميدان (Oppenheim, 1966). ولكن يعاب عليها أنها لا يمكن أن تطبّق على اشخاص محدودي الذكاء أو من درجة ثقافة متدنية، كما يعاب عليها أنها تؤدي إلى اعتماد ردود متواضعة وارتفاع نسبة الفاقد. وقد ظهر أنه في حال تطبيق هذه الاستبانة على عيّنة من الأفراد غير المهتمّين بموضوعها مباشرة فإن نسبة الرّدود تتراوح بين 40 – 60٪ وأن هذه النسبة نادراً ما تتجاوز 80٪ حتى إذا كان للمبحوثين مصلحة أو اهتمام مباشر بها (Cohen, 1980, p.88). وقد أجريت دراسات عديدة للكشف عن الفروق العقلية والشخصية القائمة بين المستجيبين وغير المستجيبين وأظهرت هذه الدراسات عدم وجود فروق بين المستجيبين وغير المستجيبين في أبعاد الشخصية الهامة. ويميل غير المستجيبين إلى أن يحقّقوا نجاحاً أكاديمياً أقل من المستجيبين، ويؤدّي انخفاض نسبة الرّدود إلى انحياز العيّنة نظراً لأن حالات عدم الاستجابة لا تخضع للتعيين العشوائي ولها محدّداتها الخاصة التي قد تختلف من مسح لآخر، ويصعب التغلّب على معضلة الفاقد وانحياز العيّنة دون معرفة طبيعة الانحياز. ويرى يورج أن الطريقة الأمثل

لمواجهة هذه المعضلة هي اختيار عدد من الحالات بطريقة عشوائية من فئة غير المستجيبين وإجراء مقابلات معهم. فإذا كانت هذه الفئة منتشرة على رقعة جغرافية فيمكن إجراء المقابلة مع الأقرب منهم إلى مكان وجود الباحث. وبعد الحصول على البيانات المطلوبة من هؤلاء لابد من مقارنة استجاباتهم لكل بند من بنود الاستبانة مع استجابات أولئك الذين ملأوها وأعادوها «فإذا لم تظهر فروق أو ظهرت فروق طفيفة يمكن الافتراض أن عيّنة المستجيبين ليست عيّنة منحازة، وإذا ظهرت فروق كبيرة فلا بد من الإشارة إليها ومناقشة دلالتها عند تقرير النتائج. ومن المفيد بهدف خفض نسبة الفاقد أن ترافق الاستمارة البريدية رسالة موجزة تشرح أغراض الدراسة وأهميتها وتحتوي على أسباب ومسوغات مقنعة لتعبئة الاستمارة وإعادتها. كما أن من المفيد اللجوء إلى رسائل التذكير والمكالمات الهاتفية والبرقيات (Cohen,1980, Borg, 1979). ولا يوجد اتفاق بين الباحثين حول الحد الأدنى المقبول لنسبة الردود والحد الأقصى المسموح به لنسبة الفاقد. ومنهم من يرى أن نسبة الفاقد لا يجوز أن تتعدّى 50٪ (النير، بلا تاريخ، ص135). والمهم في كل الأحوال هو أن تنخفض نسبة الفاقد إلى أدنى درجة كي لا تؤدي إلى إحداث نقص كبير في عدد أفراد العيّنة الكلية، كما أن من المهم أن تتوزّع هذه النسبة توزّعاً عشوائياً كي لا تؤثّر سلباً في صدق العيّنة وتمثيلها. فإذا جاء هذا التوزيع منحازاً ترتّب على الباحث إجراء بعض التصحيحات اللازمة. وعني عن البيان أن على الباحث أن يتوقّع مسبقاً امتناع عدد من المبحوثين عن الإجابة وان يحسب حساباً لذلك عند تحديد عدد أفراد العيّنة الذين سترسل لهم الاستبانة.

وتجدر الإشارة إلى أنه ظهر حديثاً في بعض البلدان المتقدّمة ما يعرف باسم "الاستبانة الهاتفية" وتعتمد هذه الطريقة في إجراء الاستبانة وتطبيقها على المبحوثين على أن يقوم الباحث أو أحد مساعديه بالاتصال هاتفياً بأفراد العيّنة وقراءة الأسئلة عليهم وتسجيل إجاباتهم عبر الهاتف. ومن مزايا الاستبانة الهاتفية أنها تؤمّن نوعاً من الاتصال المباشر بين الباحث والمبحوثين ممّا يعطي

المجال للباحث لشرح اهداف الدراسة واستثارة الدافعية لديهم للإجابة، إلاّ أن هذا النوع لا يصلح إلاّ في البلاد التي ينتشر فيها الهاتف انتشاراً واسعاً، كما لا يصلح إلاّ في حالة الاستبانة القصيرة حيث لا يرغب أكثر الناس في الإجابة عن أسئلة كثيرة عبر الهاتف. ومن الصعب تطبيق هذه الاستبانة على عيّنة كبيرة ومنتشرة انتشاراً واسعاً، ويمكن تطبيقها في حالات خاصة وعلى عيّنة صغيرة نسبياً، كما يمكن الإفادة منها في الدراسة الاستطلاعية.

إن اختيار الشكل الملائم للاستبانة (الشكل المغلق أو المفتوح أو الشكل المغلق – المفتوح) يخضع لمجموعة من المحدّدات أبرزها طبيعة الموقف الذي ستطبّق فيه الاستبانة وخصوصيته بالإضافة للأغراض الخاصة المرسومة للاستبانة. ومن الصعب الحكم على أحد الشكلين بأنّه مناسب أو غير مناسب في كل الحالات فلكل منهما مزاياه التي تتأكّد في مواقف وحالات معيّنة ولكل منهما عيوبه التي تبرز في مواقف وحالات أخرى. وعموماً يتفوّق الشكل المفتوح على الشكل المغلق في مجال الكشف عن دوافع المبحوثين ومواقفهم وفي إتاحة الفرصة لهم لاستخدام تعابيرهم الخاصة في الإجابة وتقديم استجابات "عفوية" لا تتقيّد بالبدائل التي يتضمّنها الشكل المغلق. فإذا كان من الصعب على الباحث أن يتوقّع إجابات المبحوثين، وإذا رغب في أن يتحدّث المستجيب بحرية بحيث يتمكّن فيما بعد من تعرّف آرائه الخاصة والأسباب الكامنة وراءها، وتحليل أقواله للحصول على تفصيلات مفيدة فما عليه إلاّ أن يستخدم الشكل المفتوح (Shofield, 1972). (Van Dalen, 1973) إلاّ أن الشكل المفتوح يعاني من عيوب كبيرة منها أن المبحوثين (المستجيبين) يفتقرون إلى الإشارات أو التلميحات التي يمكن أن توجّه تفكيرهم وقد لا يفهمون المقصود من السؤال (الأسئلة) فيحذفون عن غير قصد معلومات هامة وخاصة إذا كانوا من مستويات تعليمية متدنّية. كما أن عملية تفريغ وجدولة الإجابات عن هذا الشكل والتي غالباً ما تأتي متنوّعة تنوّعاً واسعاً، ليست بالعملية السّهلة وتستلزم وقتاً وجهداً كبيرين.

ولقد أشار أوبنهايم بهذا الصدد إلى أن «... من السهل طرح الأسئلة المفتوحة ولكن من الصعب الإجابة عنها، والأصعب هو تحليلها».

(Oppenheim, 1966, p.34)

ينطوي الشكل المغلق أو المقيّد على احتمالات أو بدائل محدّدة يضعها الباحث للكشف عن النواحي التي يرغب في دراستها لدى المبحوثين و«توجيه» تفكيرهم نحو تلك النواحي دون غيرها. ويتجنّب الباحث باستخدامه هذا الشكل احتمال عدم فهم الأسئلة أو سوء فهمها والاستطراد المحتمل في الإجابة عن الأسئلة المفتوحة، ومن هذه الزاوية يتمتّع بميزة هامة وهي أنه يضم أسئلة وبدائل محدّدة تقنّن طريقة الإجابة كما تقنّن طريقة تحليل الإجابات. ويفترض أن تؤدّي هذه الطريقة في طرح الأسئلة إلى «إنقاص عدد الذين يمتنعون عن الإجابة لعدم فهم السؤال وإنقاص عدد من يسيء فهم الأسئلة أو المقصود منها فتكون إجاباتهم غير دقيقة ولا معبّرة عن آرائهم» (Shofield, 1972, p.187) ويسهّل الشكل المغلق عملية التفريغ والجدولة والتحليل ويحقّق بذلك إلى درجة معيّنة شرطاً هاماً من شروط صلاحية الأداة وهو الاقتصاد في الوقت والجهد. والواقع أن هذا الشكل هو الشكل الأكثر انتشاراً بين الباحثين على الرغم من بعض الانتقادات الموجّهة له والتي تتركّز في أنه قد يضم أسئلة «مشحونة» وموحية ويمكن أن يؤدّي إلى الخسارة في عفوية المستجيب وتعبيريته، وأنه غالباً ما يعجز عن الكشف عن دوافع المستجيب والأسباب الكامنة وراء اختياره إجابة أو إجابات معيّنة دون غيرها، وقد يضطر المستجيب إلى إعطاء إجابة لا تعبّر عن أفكاره بدقّة طالما أنه مقيّد بمجموعة البدائل الجاهزة أو «أن يعطي إجابة تنسجم مع رغبات الباحث وتؤيّدها» (Van Dalen, 1973, p.326). ومن هنا لجأ الباحثون إلى استخدام الشكل المغلق – المفتوح ورأى بعضهم أن تترك صفحة بيضاء خاصة بالأسئلة المفتوحة. كما رأى بعضهم استخدام الأسئلة المفتوحة أو بعضها في مرحلة العمل الاستطلاعي بهدف تحويلها فيما بعد إلى أسئلة مغلقة. وتحقّق الأسئلة المفتوحة في هذه الحالة الأخيرة فائدة كبيرة فهي تفسح المجال لإسهام

أفراد العيّنة الاستطلاعية في اقتراح البدائل، وفي هذا إثراء للاستبانة من حيث أن المبحوثين أنفسهم يسهمون في عملية بنائها وتطويرها (برمانكوفا، 1979؛ ماتيوشكين، 1981).

تحديد الأغراض الخاصة للاستبانة:

لقد تمّت الإشارة إلى أن عملية تحديد نوع الاستبانة المراد تصميمها وشكلها تخضع لمجموعة من المتغيّرات أبرزها الأغراض الخاصة بهذه الأداة. ومن المفيد الإشارة ها هنا إلى أن تحديد الأغراض الخاصة بهذه الأداة لابد أن يسبق الخطوة العملية الأولى في تصميمها وهي إعداد الأسئلة في صورتها الأولية، ويسهم تحديد مشكلة البحث وفرضياته وأهدافه إسهاماً كبيراً في تحديد الأغراض الخاصة بالاستبانة والمجالات التي يسعى الباحث إلى تغطيتها من خلال الأسئلة التي سيعمل على إعدادها. ويرى بورج أن من الصعب على الباحث أن يوضّح الأسباب الكامنة وراء اختياره وطرحه أسئلة معيّنة دون غيرها وكيف سيقوم بتحليل استجابات المبحوثين دون أن يملك فكرة واضحة ومحدّدة عن مشكلة البحث وأهدافه. كما أن «من الصعب دون تحديد أغراض الاستبانة مسبقاً اتخاذ قرارات صائبة حول اصطفاء العيّنة وتصميم الاستبانة وطرائق تحليل البيانات» (Borg,1979, p.293). وتبعاً لذلك يترتّب على الباحث أن يعمل على تحديد مشكلة البحث وبيان عناصرها الرئيسية كما يترتّب عليه أن يعمل على تحديد المجالات الرئيسة التي ستدور حولها الأسئلة من خلال الفروض التي تطرحها الدراسة والمتغيّرات التي تتّجه تلك الفروض إلى كشف الصلة بينها. ومن المفيد بعد تحديد المجالات الرئيسة للاستبانة (وتفريعها إلى مجالات فرعية إذا تطلّب الأمر) وضع قائمة بالنقاط التي يتضمّنها كل مجال وتحديد عدد الأسئلة اللازمة لكل نقطة من تلك النقاط. وليس من الضروري أن يكون عدد الأسئلة متساوياً في المجالات أو النقاط كافة.

ولابد من أن يتناسب عدد الأسئلة اللازمة لكل مجال ولكل نقطة من النقاط التي يتضمّنها مع أهمية هذا المجال وأهمية النقاط التي يتضمّنها والوزن «النسبي» لكل منها (حسن، 1982، ص 350).

الصورة الأولية للاستبانة:

يتطلّب تصميم الاستبانة بصورتها الأولية إعداد مجموعة كبيرة من الأسئلة حول كل مجال من مجالات الاستبانة والنقاط التي يضمّها. ويفيد الإكثار في عدد الأسئلة عن الحد المطلوب في المفاضلة بينها واختيار الأصلح منها لأغراض الدراسة و«فربلتها» في ضوء نتائج التطبيق التجريبي. وليس ثمّة اتفاق بين الباحثين حول النسبة المئوية للأسئلة الزائدة عن العدد المطلوب، ويترتّب على واضع الاستبانة نفسه أن يقدّر هذه النسبة. والواقع أن عملية إعداد الصورة الأولية للاستبانة بعناصرها الكثيرة المتعدّدة ليست بالأمر السّهل وتتطلّب مهارات خاصة لدى مصمّم الاستبانة وقد تبرز الكثير من شخصيته وكفاءته العلمية. ولا تكفي المهارات التعبيرية والقدرة على الصياغة اللغوية السليمة وحدها لطرح الأسئلة ولابد من توافر مهارات وشروط أخرى.

أسأل؟ وكيف؟ ولماذا؟ وكيف سأقوم بتحليل الاستجابات؟ وقد يظنّ البعض أن عملية طرح الأسئلة هي عملية سهلة ولكنّها ليست كذلك في الواقع فما نسأله هو واضح لنا لمعرفتنا بالجواب وقد لا يكون واضحاً بالنسبة للمبحوث (Oppenheim,1966, p.39). ويشير فان دالن في هذا الصدد إلى أن «صياغة الأسئلة بهدف الحصول على إجابات صادقة هي فن قائم بذاته».

(Van Dalen, 1973, p.327)

ومن الباحثين من يرى ضرورة مراعاة التسلسل الزمني في صياغة الأسئلة (من الماضي إلى الحاضر فالمستقبل) والانتقال من المعلوم إلى المجهول ومن المحسوس والبسيط إلى المجرّد والأكثر تعقيداً (Shofild, 1972, p.189).

ومنهم من يرى أن توضّح الأسئلة وفق الترتيب المنطقي مع مراعاة الترتيب «السيكولوجي» وبحيث تسبق الأسئلة السّهلة والمشوّقة والمحايدة الأسئلة الأكثر صعوبة والأقل تشويقاً والمتّصلة بالنواحي الشخصية.

(Vam Dalen, 1973, p.328)

كما أن من الباحثين من يعتقد بوجود صعوبة في تقديم مبادئ وأسس ثابتة لترتيب الأسئلة ويرى أن يتم اختيار الترتيب الملائم للأسئلة «في ضوء مشكلات المسح الخاصة ونتائج العمل الاستطلاعي».

(Oppenheim, 1966, p.40)

وعموماً لابد ان تتوافر في الأسئلة صفات الوضوح والتحديد وأن يراعى في ترتيبها التسلسل المنطقي دون إهمال الترتيب «السيكولوجي» ومتطلّباته وان تتناسب مع سن المستجيب ودرجة تعليمه ومستوى فهمه، وتكون شاملة بحيث تغطّي تفاصيل البيانات المطلوبة. ويجب أن ينظر إلى السؤال على أنه مثير له استجابة ثابتة نسبياً وإلا فلا معنى لطرحه لأن الاستجابة (الاستجابات) سوف تتذبذب وتضعف مستوى الثبات (الوثوقية). كما يجب أن تصاغ الأسئلة بحيث «تكون الاستجابة على السؤال على أساس استجابتهم للضغوط الاجتماعية في الموقف، ويشار إلى هذه الإجابة عادة على أنها تعميمات جامدة» (جابر،1987، ص252). وعلى هذا النحو لابد أن تأخذ الأسئلة صيغاً واضحة المعنى والمبنى وان يتجنّب الباحث السؤال الذي يضم أكثر من فكرة أو نقطة واحدة (السؤال المركّب) فمثل هذا السؤال يمكن تجزئته بحيث يصبح سؤالين فرعيين أو أكثر. كما لابد من تجنّب الأسئلة التي قد تنطوي على أكثر من معنى واحد وتحتمل تفسيرات عديدة، والأسئلة غير المحايدة أو الموجبة أي التي توحي بجواب واحد مقبول، من مثل: «معظم الناس ينظّفون أسنانهم يومياً. ألا تفعل انت ذلك؟ أو «معظم الناس هذه الأيام يعتقدون ... ألا تعتقد انت بذلك ايضاً؟ «ومن الأفضل تجنّب الأسئلة التي تأخذ صيغة النفي بالإضافة إلى الأسئلة

التي تحتاج الإجابة عنها إلى ثقافة رفيعة، والأسئلة المثيرة والمزعجة من مثل: «هل قرأت صحفاً أو مجلاّت من أي نوع كان في أثناء حياتك المهنية كمعلّم أو هل شاركت في دورة تدريبية من أي نوع كان في أثناء حياتك المهنية كمعلّم؟».

ومن المناسب في حالات أن تحتل أسئلة البيانات الشخصية من مثل الأسئلة التي تتعلّق بالعمر والجنس والوضع العائلي... الخ مكانها في مواضع متأخّرة من الاستبانة لكي يستجيب لها المبحوث بصورة إيجابية وبعد أن يكون قد اقتنع بأهمية البحث. ولهذه الأسئلة أهمية خاصة في تصنيف العيّنة حسب المتغيّرات المطروحة للدراسة. (Oppenheim, 1966,p.58)

ومن الأمور التي يجب أن يأخذها الباحث بالحسبان ألاّ تكون الأسئلة طويلة وعبئاً على المبحوث من ناحية الوقت. فالزمن الطويل قد يؤدي إلى السأم والتعب ويؤثّر في جدية المستجيب ويضعف اهتمامه. ولا يوجد قاعدة عامة بصدد الطول المناسب للاستبانة فثمّة فروق كبيرة بين الناس في سرعة الفهم وتحمّل التعب وسرعة الاستجابة. وعموماً فإن احتمال إجابة أفراد العيّنة عن الاستبانة يزداد كلّما قصرت، كما أن الاستبانة التي يحتاج الشخص المتوسّط لتعبئتها إلى أكثر من ساعة تعدّ استبانة طويلة (الفير، بلا تاريخ، سيفيكينا، 1986). ومن هذه الزاوية لابد أن يؤخذ الوقت بالحسبان ومن خلال مراعاة الطبيعة الخاصة للعيّنة سواء من ناحية العمر أو المستوى التعليمي أو غير ذلك على أن تكون الاستبانة من الطول بحيث تغطّي جميع البيانات الضرورية للدراسة في الوقت نفسه. ومن الأمور الهامة في هذا السياق استثارة الدافعية لدى المبحوثين للإجابة بعناية وصدق عن كافة الأسئلة. وثمّة عوامل عديدة تؤدّي إلى رفع مستوى الدافعية والاهتمام لدى المبحوثين منها تفهّم أغراض الدراسة وقيمتها والاقتناع بأهميّتها وجدواها، والصياغة الواضحة والمحدّدة للأسئلة وتعليمات الإجابة، والابتعاد ما أمكن عن الأسئلة الشخصية المحرجة وتجنّب الأسئلة الاستفزازية والمثيرة للقلق (كالأسئلة التي تتعلّق بالدين أو الطائفة أو الدّخل غير المشروع... الخ.).

وكثيراً ما تنص التعليمات على أنّ الإجابة لن يطلع عليها أحد بل هي لأغراض البحث فقط. ويطلب عادة عدم كتابة الاسم إلاّ في حالات خاصة وعندما يرغب الباحث في إعادة تطبيق الاستبانة لمعرفة التغيّر (والثبات) في نظرات المبحوثين ومواقفهم (Shofield, 1972). ومن المفيد بصورة عامة الحفاظ على سرية الاسم بهدف الحصول على استجابات صريحة وكاشفة، وفي الحالات التي تتطلّب ذكر الاسم يمكن إعطاء المستجيبين أرقاماً بدلاً من مطالبتهم بذكر أسمائهم. كما أن من المفيد بهدف ترغيب المبحوثين أن تكون الاستبانة منظّمة وحسنة التبويب وجذّابة من ناحية الشكل والإخراج.

تعليمات الاستبانة:

للتعليمات المرافقة للاستبانة أهمية خاصة من حيث أنها ترشد المبحوث وتبيّن له ما يجب عليه أن يفعله وتؤثّر بالتالي تأثيراً كبيراً في صدق الاستجابة. ونظراً لعدم وجود احتكاك شخصي مباشر بين الباحث والمبحوث، وخاصة في الاستبانة البريدية، يترتّب على الباحث أن يولي عناية فائقة بصياغة التعليمات، وأن يتنبّأ باحتياجات المبحوث ويقدّم له التوجيهات اللازمة لفهم طريقة الإجابة. ويجب أن تتوافر في التعليمات شروط البساطة والوضوح والتحديد وأن تكون مباشرة ووافية بالغرض دونما تطويل زائد أو إيجاز غامض. ويفضّل أن تكتب بخط عريض أو يوضع تحتها خط. وتخضع التعليمات للتعديل والإضافة (والحذف أحياناً) في مراحل العمل جميعاً كما تخضع الأسئلة ذاتها. ويتم من خلال التجربة الاستطلاعية بوجه خاص واللقاء المباشر أو «المقابلات» التي يجريها الباحث مع أفراد العيّنة الاستطلاعية تعرّف الصعوبات «الواقعية» التي يواجهها المبحوثون في تعبئة استمارة الاستبانة، وفي ضوء ذلك يتم إجراء التعديلات اللازمة في التعليمات وصياغتها. ولا شك أن لعدم وضوح التعليمات وسوء فهم الأسئلة الناجم عنه أثر سلبي مباشر في صلاحية الاستبانة من حيث أنه يلحق ضرراً كبيراً بصدقها وموثوقيتها (ثباتها). وقد يركّز الباحث اهتمامه بالأسئلة ولا يولي اهتماماً كافياً بتعليمات الإجابة عنها مما يسيء إلى الأسئلة ذاتها وقد

يؤدّي إلى سوء فهمها ويدعو بالتالي إلى التشكيك بصحّة الإجابات. ومن هنا لابد من التنبّه لأهمية التعليمات وشمولها وأهمية صوغها في عبارات واضحة ومحدّدة وتعديلها وتنقيحها في المراحل اللاحقة حسب اللزوم.

التحليل المنطقي:

بعد إعداد الأسئلة وتعليمات الإجابة عنها في صورتها الأولية لابد من مراجعتها بصورة متأنّية وإخضاعها للتحليل المنطقي بهدف التحقّق من ملاءمتها للأغراض الخاصة بالاستبانة و«تمثيلها» للمجالات التي تسعى إلى تغطيتها، ومن وضوحها وتسلسلها المنطقي. ومن الجدير بالإشارة هنا أن التحليل العقلي المنطقي لهذه الأداة ليس عملاً متمّماً لها أو لاحقاً بها بل هو عمل مستمر و«مواكب» لعملية تصميمها، ويسبق ظهورها في صورتها الأولية ولكن تكون له في هذه المرحلة بالذات مكانة خاصة. ومن الضروري بهدف المزيد من التحليل عرض الأسئلة المقترحة مع تعليماتها على مجموعة من المختصّين والخبراء أو المحكمين) ممّن لهم صلة بموضوع الدراسة ودراية بتطوير الاستبانة. ويسهم إطلاع المختصّين والخبراء على الأغراض الخاصة المرسومة للاستبانة والمجالات التي تسعى إلى تغطيتها والنقاط الخاصة بكل مجال في التحقّق من تمثيل عناصر الاستبانة للبيانات المطلوبة وملاءمة الأسئلة المطروحة لكل نقطة في مجال ممّا يسهم في تأسيس صدق المحتوى أو الصدق المنطقي للاستبانة . والملاحظات التي يقدّمها أولئك المختصّون والنقاط المشتركة بينهم لابد أن تؤخذ بالحسبان في مراحل العمل اللاحقة وقد يكون لها دور كبير في تطوير الاستبانة ورفع مستوى صلاحيتها إذا أحسن الباحث الإفادة منها. وعموماً فإن للتحليل المنطقي الذي يقوم به الباحث بنفسه كما يقوم به مستعيناً بآراء المختصّين أهمية فائقة من حيث أنه يشكّل الحجر الأساسي في عملية تصميم الاستبانة ولابد أن يواكب هذه العملية حتى نهايتها. وقليلاً ما تولي «الأدبيات» والدراسات الخاصة بتصميم الاستبانة الأهمية اللازمة لهذا النوع من التحليل وغالباً ما تعطي الأولوية للعمل التجريبي والتحليل الإحصائي. والواقع أن من الأهمية بمكان أن يكون للتحليل

المنطقي مكانه الطبيعي وان تعطى له الأولوية طالما أن هذا التحليل هو الأساس الذي ترتكز عليه عملية التصميم برمّتها وأنه يضمن صدق المحتوى أو الصدق المنطقي للاستبانة ويفيد في دراسة جوانب هامة كالترتيب المنطقي (أو السيكولوجي) للأسئلة وصياغتها، هذا بالإضافة إلى أنه يسبق العمل التجريبي والتحليل الإحصائي ويحدّد مجال كل منهما واتجاهه وإجراءاته ويكون له دوره في هذه المرحلة ذاتها من مراحل العمل. وعلى هذا فإن تجاهل مصمّم الأداة لهذا النوع من التحلل، أو إجراءه بصورة شكلية أو «روتينية» من خلال تقديم الأسئلة في صورتها الأولية «على عجل» إلى بعض المختصّين دون تقديم بيانات تفصيلية وواضحة حول الأغراض الخاصة المحدّدة لها ومجالاتها الرئيسة والفرعية والأسئلة التي تغطي كلاً من هذه المجالات قد يلحق ضرراً كبيراً بالاستبانة وصدقها، وقد لا يمثّل أكثر من إجراء شكلي أو «استعراضي» يحقّق خدمة واحدة وهي تغطية عيوب الاستبانة أو نقاط ضعفها بالألقاب العلمية أو الأسماء البارزة لأولئك المختصّين.

الدراسة الاستطلاعية:

بعد أن ينتهي الباحث من إعداد الأسئلة بصورتها الأولية وإخضاعها للتحليل المنطقي من قبل الباحث نفسه ومجموعة من المحكمين وتعديلها وتنقيحها في ضوء هذا التحليل ينتقل إلى الخطوة التالية في تصميم الاستبانة وهي التجريب الأولي للاستبانة أو ما يعرف بالدراسة الاستطلاعية. ومن نافلة القول أن التحليل المنطقي ونصائح الخبراء ليست بديلاً عن الدراسة الاستطلاعية المنظمّة فهذه الدراسة تأتي مساندة لهذا التحليل ومتمّمة له. ويترتّب على الباحث هنا أن يختار عيّنة من الأشخاص من المجتمع نفسه الذي ستسحب منه عيّنة الدراسة فيما بعد. وقد لا يتجاوز عدد أفراد العيّنة الاستطلاعية 20 أو 30 شخصاً في حالات، وقد يكون من الضروري تجاوز هذا العدد بكثير في حالات أخرى. وتحقّق الدراسة الاستطلاعية فوائد عديدة للباحث حيث يقوم في ضوئها بإعادة صياغة الأسئلة التي يتبيّن أنها تعاني من الغموض،

كما يقوم بإلغاء الأسئلة غير اللازمة، وإضافة أسئلة جديدة أظهرت هذه الدراسة ضرورة تضمينها استمارة الاستبانة. ومن فوائد الدراسة الاستطلاعية أيضاً أنها تمكّن الباحث من «تحويل الأسئلة المفتوحة إلى أسئلة مغلقة بعد أن يكون قد اكتشف من خلال الإجابات المجالات التي يمسّها البحث ونوع الاستجابات اللازمة» (حسن، 1982، ص353). كما أن العمل الاستطلاعي يتيح دراسة كفاية هذه الأداة بجوانبها، بما في ذلك دراسة فعالية طرائق التفريغ والتحليل الكمّي» (Borg, 1979, p.352). ويفيد العمل الاستطلاعي في تحديد الزمن اللازم للإجابة عن الاستبانة وتحديد طولها، كما يفيد في تحديد الصعوبات المتّصلة بصياغة التعليمات وما إذا كانت محدّدة وواضحة وفي مستوى الفهم العام للمبحوثين. هذا بالإضافة إلى أنه يتيح الوقوف عند الأثر الذي يحدثه ترتيب الأسئلة بهدف إعادة تنظيم الاستبانة وإرجاء بعض الأسئلة التي تبدو أنها محرجة إلى أماكن أخرى من الاستمارة ممّا يسهم بالنتيجة في تخفيض نسبة الفاقد (من غير المستجيبين) (Oppenheim, 1966). وعلى الباحث هنا أن يقوم بتسجيل الملاحظات أو التعليقات التي تكون قد أثيرت حول الأسئلة وان يعمل على دراستها بصورة متأنّية مع الوقوف بصورة خاصة عند الأسئلة التي يمتنع الكثيرون عن الإجابة عنها لسبب أو لآخر والأسئلة التي تكثر الإجابات المحايدة عنها (من نوع لا ادري) أو الأسئلة التي قد تؤدّي إلى إجابة واحدة لدى المبحوثين جميعاً. وعموماً تسهم الدراسة الاستطلاعية والتحليل الشامل للأسئلة في حولها في تقديم بيانات مفيدة حول مؤشّرات صلاحية الاستبانة بصورة أولية (من حيث الصدق والثبات والكفاية)، وعلى الباحث أن يعمل على توظيف نتائج هذه الدراسة وإجراء التعديلات والتصحيحات اللازمة بما يخدم صلاحيتها ويسهم في رفع مستوى هذه الصلاحية.

التجريب التمهيدي والنهائي:

إن التجريب الأولي (الاستطلاعي) للاستبانة لا يكفي في معظم الأحيان، ويترتّب على الباحث أن يتابع عمله التجريبي بإجراء تجربة أخرى أو سلسلة من التجارب قبل الوصول إلى المرحلة الأخيرة وهي: مرحلة إعداد الاستبانة وإخراجها بصورتها النهائية. فبعد أن تعاد صياغة السؤال أو الأسئلة في ضوء التجربة الاستطلاعية لابد من إخضاعه للتجريب ثانية طالما أن هذه الصياغة الجديدة قد تؤدّي إلى تحيّزات أو صعوبات جديدة. وعندما يحوّل السؤال المفتوح إلى سؤال مغلق لابد من تجريبه في شكله الجديد. ويرى أوبنهايم أن «الأسئلة التي يتم اقتباسها أو استعارتها من استبانات جاهزة، معروفة بصلاحيتها، ولابد أن تخضع بدورها للتجريب لضمان الحصول على نتائج مرضية من العيّنة التي اختارها الباحث (Oppenheim, 1966,p.26). كما يقترح أوبنهايم تجريب الاستبانة على مراحل بحيث يتم تجريب كل فئة من فئات الأسئلة تغطّي مجالاً واحداً بعينه على حدة. وهذا سيجزّئ العمل التجريبي إلى عدد من العمليات الصغيرة، ويعطي الخبرة في التواصل مع المبحوثين، وفي شرح الغرض من الدراسة على نحو أفضل. وبالطبع فإن التجريب المرحلي لا يلغي ضرورة تجريب الاستبانة ككل فيما بعد فلكل منهما فوائده. وفي كل الأحوال لابد أن يولي الباحث اهتماماً خاصاً بالعمل التجريبي والتحليل الإحصائي للنتائج مع مراعاة شروط ومتطلّبات هذا العمل بمنتهى الدقّة، كما لابد أن ينظر إلى التجريب على أنه المجال الطبيعي «لغربلة الاستبانة وتنقيتها واستكمال أو تأمين مستلزمات صلاحيتها. وتخضع الحاجة لإعادة التجربة لمجموعة من المتغيّرات منها الأغراض الخاصة بالدراسة والنتائج التي أسفرت عنها التجربة الأولى حول الاستبانة وصلاحيتها، والشروط الواقعية المحيطة بالعيّنة، والإمكانات المتاحة للباحث. وعموماً يندر أن تفي التجربة الاستطلاعية وحدها بأغراض التجريب وفوائده وغالباً ما تتطلّب عملية تصميم الاستبانة والوصول بها إلى صورتها النهائية إجراء تجربة لاحقة أو سلسلة من التجارب (التمهيدية والنهائية) بهدف متابعة

عملية «تنقية» الاستبانة ورفع مستوى صلاحيتها. والأمر الذي يجب أن يؤخذ بالحسبان هو أن التجريب ليس غاية بذاته والإكثار منه أو تنويعه ليس دليلاً بحد ذاته على رفع مستوى جودة هذه الأداة. وعلى هذا فإن التجريب لا يجوز أن يكون إجراء «روتينياً» أو شكلياً يقوم به الباحث بهدف الظهور أو «الاستعراض». وتتحدّد قيمة العمل التجريبي برمّته وبكافة أشكاله وبما يتطلّبه من إجراءات المعالجة والتحليل الإحصائي، بما يقدّمه هذا العمل من بيانات «يوظّفها» الباحث فعلاً في صالح عملية التحقّق من صلاحية الاستبانة وفي رفع مستوى هذه الصلاحية إلى الحدود القصوى.

الشكل النهائي للاستبانة:

بانتهاء عملية التجريب والتحليل الإحصائي وبعد أن يكون الباحث قد انتهى من عملية التعديل والتنقيح و«التهذيب» للأسئلة وتعليماتها وتمكّن من اختيار أفضلها وأكثرها تحقيقاً للأغراض المرسومة تبدأ المرحلة الأخيرة في تصميم الاستبانة وهي تنسيق الاستمارة الخاصة بها وإعدادها في شكلها النهائي. وتتطلّب هذه المرحلة من جملة ما تتطلّبه طباعة الاستمارة على وجه واحد فقط لتكون واضحة وسهلة القراءة، وتخصيص المكان الكافي أمام كل سؤال أو تحته للإجابة عنه وإعطاء الأسئلة أرقاماً متسلسلة وتقسيمها إلى مجموعات توضع لها عناوين واضحة. كما تتطلّب وضع علامات مميّزة على الاستمارات لتسهيل التعرّف على كل فئة منها إذا كان الباحث يسعى إلى المقارنة بين إجابات الفئات المختلفة من المبحوثين، وإرسال ظرف معنْوَن عليه طابع بريدي إذا كانت الاستبانة بريدية . وعموماً فإنّ لشكل الاستمارة وإخراجها أثراً لا يستهان به في تعاون المبحوثين ويتوقّع أن ترتفع نسبة الرّدود بصورة ملحوظة مع زيادة الاهتمام بالشكل والإخراج.

صدق الاستبانة:

لقد تمّت الإشارة في السابق إلى أن عملية تطوير الاستبانة والوصول بها إلى شكلها النهائي تتطلّب التحقّق من صلاحيتها من خلال اختبار صدقها وموثوقيتها (ثباتها) وكفايتها. والواقع أن عملية تصميم الاستبانة لابد أن تواكبها عملية التحقّق من صلاحيتها إذا أريد لها أن تكون أداة بحث علمية ودقيقة وأن تتبعها أيضاً. ويتم التحقّق من الصدق المنطقي للاستبانة (صدق المحتوى) من خلال دراسة عناصر الاستبانة ويضمن هذا الشكل من أشكال الصدق شمول الاستبانة «للعناصر أو أنواع السلوك أو البيانات أو العوامل التي يجري قياسها دون زيادة أو نقصان» (حمدان، 1989، ص145). وقد أشار ايبانز إلى أن هذا الشكل من أشكال الصدق يتطلّب مراعاة أربع نقاط حيوية وهي: 1 – ان يكون الغرض من الاستبانة واضحاً تماماً في ذهن واضعها وأن يعمل على تعريف المبحوثين بهذا الغرض. 2 – أن يقرّر في ضوء هذا الغرض البيانات المطلوبة بمنتهى الدقّة. 3 – أن يحلّل هذه البيانات إلى عناصرها وأجزائها ويتأكّد من أن الأسئلة الموضوعة تغطي هذه العناصر والأجزاء و«تتّسق» معها. 4 – أن يضمن أن تكون الأسئلة مصاغة بحيث تحفّز على إعطاء الجواب المطلوب (Evans, 1968, p.41 – 45). وعموماً تفيد عملية اختبار صدق الاستبانة بهذه الطريقة في التثبّت مما إذا كانت الأسئلة تسأل فعلاً ما يراد منها أن تسأله، وهي تسبق المرحلة التجريبية وتستمر معها. وتنطوي هذه الطريقة على أهمية قصوى من حيث أنها تكشف عن ملاءمة محتوى الاستبانة وما تضمّه من البنود للموضوع المدروس بمجالاته الرئيسة والفرعية ودرجة تمثيلها له وللمتغيّرات التي يسعى الباحث إلى دراستها وكشف الصّلة بينها. ويؤدّي إهمال هذا الشكل من أشكال الصدق أو عدم الاهتمام به بصورة كافية إلى التشكيك كلّياً أو جزئياً بالاستبانة وصلاحيتها وقيمتها كأداة بحث بغض النظر عن الجهود التي قد يبذلها الباحث في مجال العمل التجريبي والتحليل الإحصائي. ومن الطرائق المتّبعة في اختبار صدق الاستبانة اللجوء إلى ما يسمّى بالصدق المحكي حيث تتم

مقارنة نتائج الاستبانة مع نتائج استبانة أخرى تؤخذ كمحك وتقيس الشيء أو (المتغيّر) نفسه الذي تقيسه الاستبانة. ويتميّز الصدق المحكي عن الصدق المنطقي أو الصدق عن طريق المحكمين في أنه يعبّر عن صفة الصدق بلغة الكم من خلال معامل الصدق (وهو معامل ارتباط بطبيعة الحال)، ويعاب عليه إمكان التشكيك في صدق الأداة ذاتها المأخوذة كمحكّ. ويرى أوبنهايم أنه لضمان الصدق في أسئلة الحقائق يمكن اللجوء إلى الصدق المحكي عن طريق ما يسمّى بالمراجعة المتقاطعة أو المتقابلة (Across – checks) حيث يكون هناك مصدر ثان مستقل للمعلومات وتتم المقابلة بين المعلومات المتحصّلة من الاستبانة وتلك المستمّدة من ذلك المصدر. ويمكن استخدام الوثائق والسجلاّت المختلفة كسجلاّت الخدمة العسكرية والدوائر الحكومية الأخرى، كما يمكن الاستعانة بالوالدين والمعلّم كمصدر للمعلومات، أي كمحك للصدق. ومن الباحثين من يقترح إجراء مقابلة مع عيّنة من المستجيبين للتحقّق من الصّدق ورفع مستواه في استخدام الاستبانة البريدية خاصة. وقد أشار أوبنهايم إلى أنه يمكن رفع مستوى الصدق بإجراء مقابلة شخصية مباشرة مع المستجيب وطرح مجموعة من الأسئلة عليه بسرعة قصوى ممّا يحفّزه على إعطاء معلومات حقيقية (Oppenheim,1966). وعموماً تعد مسألة الصدق هي المسألة المركزية ضمن طائفة واسعة من المسائل التي تثيرها عملية تطوير الاستبانة. ومن المفيد استخدام أكثر من طريقة واحدة في اختبار الصدق واللجوء إلى الصدق التجريبي بالإضافة إلى الصدق المنطقي.

الموثوقية والكفاية:

لا شك أن عملية التحقّق من صدق الاستبانة، على أهميتها، ليست دليلاً كافياً على صلاحية هذه الأداة، ولابد من التثبّت من موثوقيتها (ثباتها) بالإضافة إلى كفايتها لاستكمال عملية التحقّق من صلاحيتها. ويشير الثبات إلى اتساق البيانات والحصول على الإجابات نفسها ثانية في حال إعادة تطبيق هذه الأداة على حين أن الصدق يعبّر عمّا إذا كانت البنود (الأسئلة) تقيس فعلاً ما

وضعت لقياسه . ومن المعلوم أن أداة القياس قد تكون عالية في ثباتها وضعيفة في صدقها، فالأسئلة التي يجاب عنها بعبارات مقبولة اجتماعياً قد تحقق ثباتاً عالياً مع أنها ليست صادقة. إلاّ أن الصدق لا يمكن أن يصل إلى درجة معيّنة إذا لم تكن الأداة ثابتة بدرجة ما. وبعبارة أخرى لكي تكون الأداة على درجة عالية من الصدق لابد ان تكون عالية في ثباتها دون أن يدل ارتفاع مستوى الثبات بحد ذاته على صدق هذه الأداة. ومن هنا فإن دراسة الثبات (والتغيّر) في استجابات المبحوثين لا تنفصل عن دراسة مسألة الصدق، والواقع أن صدق الاستجابة (الاستجابات) التي يقدّمها المبحوثون تؤثر تأثيراً بالغاً ومباشراً في ثبات هذه الأداة كما تؤثر بطبيعة الحال في صدقها، وبالتالي فإن دراسة الثبات والصدق تتطلّب مراجعة تلك الاستجابات. ومن المفيد الإشارة هنا إلى أن استجابة المبحوث قد لا تكون حقيقية أو «منزّهة» أو صادقة في كل الحالات فقد تتأثر بالتفكير الرغبي وبالميل إلى إرضاء الباحث والرغبة في أن يكون شخصاً حسن الصّفات بنظر نفسه ونظر الآخرين، كما قد تتأثر بخداع الذاكرة (سيفكينا، 1986). وثمّة أسباب عديدة تؤدّي إلى إضعاف صدق الاستجابة وتؤدّي بالتالي إلى إضعاف ثبات الاستبانة وصدقها منها أن المبحوث قد لا يعرف الإجابة عن السؤال فيلجأ إلى التخمين، أو أنه قد لا يفكّر في الأسئلة ولا يتأمّلها جيداً. هذا بالإضافة إلى أن المبحوث قد لا يفهم التعليمات جيداً، وقد يخشى قول الصدق أو يشعر أن السؤال شخصي جداً في طبيعته (جابر، 1987) . ويتم التثبّت من صدق الاستجابة (الاستجابات) بعدّة طرق منها المراجعة الداخلية للاستجابات (Internal Checks). ومن اشكال هذه المراجعة تقديم بنود وهمية أو زائفة كتقديم اسم أو برنامج إذاعي غير موجود، فإذا أقرّ المستجيب مثل هذه البنود فإن هذا يدل على لجوئه إلى التخمين ولا مبالاته. ومن أشكالها ايضاً إعادة طرح السؤال في موضع آخر من الاستبانة بعد تعديل صياغته او وضعه في سياق آخر. فلو طرح السؤال التالي: (هل تذهب إلى المتحف؟) فيمكن دون إعادة السؤال نفسه، والتي غالباً ما تزعج المبحوث، وضع هذا السؤال في سياق آخر ليصبح بالشكل التالي: (هل تزور الأماكن التالية: السينما، حديقة الحيوان، المتحف، المعرض الفني؟). وفي بعض الحالات تتم المراجعة

الداخلية بطريقة منطقية فمثلاً إذا ادّعى شخص ما انه خاض الحرب فلابد من التأكّد انه كان في عمر معيّن وقت الحرب (Oppenheim, 1966). وغني عن البيان أن الطرائق المتّبعة في حساب الثّبات من مثل استعمال نماذج بديلة وإعادة إجراء الأداة ثم الربط بين النتائج يمكن استخدامها في التحقّق من ثبات الاستبانة. ومن الأجدى للباحث ألاّ يقتصر على طريقة واحدة للتحقّق من الثبات فلكل من الطرائق المتّبعة في حساب الثّبات مزاياها الخاصة. وتدعم كل منها الأخرى وتكملها ممّا يسهم بالنتيجة في تأكيد الثّبات وتعزيزه وتقريره بأدلّة واضحة.

وبالإضافة إلى الثّبات والصدق لابد أن تحقّق الاستبانة شرط الكفاية كغيرها من الأدوات وان يأخذ الباحث بالحسبان مسألة الاقتصاد في الوقت والجهد والنفقات ويعمل على إلقاصها إلى الحدود الدنيا شريطة الايلحق ضرراً بصدق الاستبانة وموثوقيتها. وعموماً تعدّ الاستبانة البريدية والاستبانة في شكلها المغلق أكثر كفاية من الاستبانة غير البريدية والأسئلة المفتوحة. ومن الضروري أن يتم اختيار الباحث لنوع الاستبانة وشكلها من منطلق مراعاة الشروط الثلاثة لصلاحيتها جميعاً مع تحقيق نوع من «التوازن» بين تلك الشروط ودون أن يكون الاهتمام بأحدها على حساب الأخر.

عيوب الاستبانة ومحاذيرها:

تفيد مراعاة الشروط الضرورية لتصميم الاستبانة والقيام بجملة الإجراءات اللازمة للتثبّت من صلاحيتها في تحاشي الكثير من مصادر الخطأ ولكنها لا تنفي كلية احتمالات الخطأ الذي يمكن أن يلحق بهذه الأداة الهامة من أدوات البحث.

وثمّة مصادر عديدة للخطأ تنعكس بصورة مباشرة أو غير مباشرة على الاستبانة ونتائجها وقد تؤدّي إلى إفسادها والتشكيك بها كأداة إذا لم يتنبّه

الباحث إليها ويعمل على تداركها أو التخفيف منها إلى الحدود القصوى. فإهمال المستجيبين أو سوء فهمهم لبنود الاستبانة، وارتفاع نسبة المستجيبين بما يتجاوز حدود التوقع، بالإضافة إلى أخطاء العيّنة والأخطاء التي يمكن أن تنجم عن عملية التفريغ والتحليل الإحصائي واستخلاص النتائج وغيرها من المؤثّرات يمكن أن تعمل منفردة أو مجتمعة وتقلّل من قيمة الإجراءات والخطوات المتّبعة في تصميم هذه الأداة والتحقّق من صلاحيتها، كما قد تؤدّي إلى التشكيك بمصداقيتها وقيمتها وتضعف وزنها كأداة بحث. ومن هذه الزاوية لابد أن تؤخذ مصادر الخطأ المشار إليها بالإضافة إلى أخطاء التصميم والانخداع ببساطة هذه الأداة وسهولتها الظاهرية على أنها من أهم الأسباب الكامنة وراء التشكيك بفاعلية هذه الأداة والإشارة إلى محدوديتها.

وتغطّي التحفّظات والانتقادات الشديدة الموجّهة لهذه الأداة جوانب عديدة تتّصل بالطبيعة الخاصة لها وصلاحيتها، والشروط الواقعية المحيطة بإجرائها والإفادة من نتائجها. ومن الباحثين من يرى أنه من الصعب الحصول على استجابات صادقة نظراً لأن استجابات المبحوثين تخضع لمؤثّرات سلبية عديدة وشديدة الفعالية. ويؤدّي ضعف مستوى الصدق في استجابات المبحوثين إلى عدم اتساق البيانات والتعارض بين الأقوال والأفعال وقد تكون بعض الاستجابات من نوع التعميمات الجامدة والتي قد تنطوي على شيء من الحذر والتحفّظ في قول الحقيقة و«مراعاة» السلطة. وقد أشار شوفيلد في هذا الصدد إلى أن «الكثير من الإجابات يحتمل أن تكون نمطية أو أن تحمل طابعاً استعراضياً خادعاً حتى عندما يبذل المبحوثون قصارى جهدهم للإجابة عن الأسئلة بعناية ويفكّرون فيها ملياً». (Shofield, 1972, p.193). وتبعاً لذلك يصعب الوصول إلى أكثر من تعميمات أولية و«غير نهائية» من بيانات الاستبانة، كما يصعب القول أنها تفضّل طريقة المقابلة في صدقها وثباتها مع الأخذ بالحسبان أن طريقة المقابلة تعاني من تدنٍّ واضح في مستوى الصدق والثبات.

(Shofield, 1972 نيرفر، 1983)

بالإضافة إلى ذلك يصعب عن طريق الاستبانة الكشف عن دوافع المبحوث ومشاعره وحالته النفسية. وليس من النادر «أن يحار الباحث عندما يجد أن المبحوثين يستخلصون معاني مختلفة كثيرة من الأسئلة التي اعتقد أنها واضحة ومحدّدة بصورة كافية» (Van Dalen, 1973, p.327).

كما أنه ليس من النادر في حال استخدام الأسئلة المغلقة ألاّ يتمكّن الباحث من تقديم عدد من الإجابات البديلة تتيح للمبحوث أن يعبّر بدقّة وبصورة ملائمة بسبب من صعوبة التنبؤ باستجابات المبحوث (المبحوثين) المحتملة. ويؤدّي هذا الأمر إلى تقيّد المبحوث بطائفة محدّدة من الإجابات وقد يدفعه إلى إحداها. وقد تتلوّن الأسئلة بالتحيّزات الشخصية للباحث وتكون مشحونة باتجاه واحد، وقد تطرح في أوقات وظروف غير مواتية. ومن الصّعب على الباحث أن يضمن تعاون المبحوث لعدم وجود احتكاك شخصي مباشر بينهما ونظراً لأن المبحوث يتعامل مع كرّاسة صمّاء من الأسئلة وقليلاً ما يتعامل مع الباحث نفسه على حين انه يمكن أن يتم هذا التعاون والتجاوب في حالة الملاحظة أو المقابلة الشخصية المباشرة (جابر، 1987، يرماكوفا، 1979). ومن الانتقادات الموجّهة للاستبانة أيضاً أنها تصلح للدراسات التي تهتم بجمع بيانات عن المظاهر الخارجية للسلوك ولا تصلح للدراسات التي يحتاج فيها الباحث إلى بيانات تتّسم بالعمق والفهم لطبيعة الظاهرة (أو الظاهرات) المدروسة. (ماتيوشكين، 1981، كابانوفا، 1982) . ويستتبع هذا أن تقتصر الاستبانة على الموضوعات التي تكوّن عنها المبحوث نظرة محدّدة وواضحة ويستطيع أن يعبّر عنها ببساطة (جابر، 1987). وغني عن البيان ان الاستبانة لا تصلح لصغار السن والأميين وأنها تتطلّب أكثر من مجرّد القراءة والكتابة.

الفصل الرابع

خطوات بناء الاختبارات والمقاييس وتقنينها

أولاً: خطوات بناء المقاييس :

لابد من أن تخضع عملية تصميم المقاييس لخطة عمل محكمة، تساعد في التحديد المسبق للمراحل والخطوات والإجراءات التنفيذية اللازمة. وتعكس هذه الخطوات شخصية مصمم المقياس وإبداعه الخاص. ويعتمد تحديد هذه الخطوات على أسس مشتقة من نظريات القياس المعاصرة وممارساتها المتقدمة. ومن الأسس الهامة التي يجب مراعاتها عند تصميم وبناء المقاييس ما يلي:

١-تحديد الفكرة العامة للمقياس والهدف الموجه له:

لكي تتم عملية تحديد الفكرة العامة للمقياس أو الأهداف الرئيسة له، لابد أولاً من من الاطلاع على الاختبارات المتوفرة والتأكد من مدى تلبيتها للغرض الذي وضعت لأجله. حيث تنطوي عملية تحديد وبلورة الهدف العام للمقياس على وضع تصور أولي تتم من خلاله بناء الخطوط العريضة لخطة بناء المقياس وتحديد الفرص والتحديات التي يمكن أن تواجه عملية بناء وتطوير المقياس وإخراجه إلى حيز الوجود.

تختلف المفاهيم العلمية في طبيعتها، حيث يجب تحديد مفهوم علمي محدد وواضح يمكن اختباره بواسطة المقياس ، فقد يكون الهدف من المقياس مثلاً تحديد مستويات الأفراد وفقاً لما يمتلكونه من خاصية أو استعداد معين، وقد يكون الهدف من مقياس آخر التمييز بين الأفراد وفقاً لترتيبهم على الخاصية، أو يكون الهدف تحديد الصعوبات أو الاضطرابات التي يعاني منها الفرد، وبالتالي يصبح الهدف من المقياس تحليل هذه الصعوبات بشكل تصنيفي مفصل.

فقد يهدف المقياس لتحديد أي من المتعلمين قد وصل أداؤه إلى المستوى المقبول، أو حقق الحد الأدنى المقبول، ويمكن أن يكون المقياس في مجال القدرات المعرفية حيث يصبح الهدف منه هو التعرف على أي من الأفراد يمكنه الإجابة عن بنود معينة، أو حل مشكلات ذات مستويات مختلفة من الصعوبة، ومن ناحية أخرى قد يستهدف المقياس تشخيص مشكلات محددة حيث يكون المطلوب تحديد شدة الصعوبات بشكل دقيق.

كما يجب تحديد نوع المقياس هل هو لفظي أم أدائي ، أو لفظي وأدائي معاً، ويجب تحديد المجتمع الأصلي للأفراد الذين سيطبق عليهم المقياس وتحديد الفئة العمرية وغير ذلك، كما يجب تحديد طريقة تطبيق المقياس هل هو فردي أم جماعي، وغير ذلك من الأمور التي ترتبط مباشرة بالهدف العام للمقياس وتؤدي في نهاية المطاف إلى تحقيق الغرض منه.

٢-تحديد الأهداف الفرعية الخاصة للمقياس:

يتضمن تحديد الأهداف الفرعية للمقياس فكرة ترجمة الأهداف العامة إلى أهداف خاصة ومحددة، كما يعد بمثابة ترجمة للفكرة الموجهة للمقياس أو الهدف العام منه إلى أهداف فرعية واضحة ومحددة، كما يمثل نقطة البداية الفعلية في بناء الاختبار ويؤثر في الخطوات اللاحقة جميعها.

ويتضمن ترجمة الأهداف العامة إلى أهداف خاصة تحديد بعض المفاهيم العلمية كالذكاء أوبعض سمات الشخصية كالانبساط أو الانطواء مثلاً، ونستدل على هذه الخصائص من خلال سلوكيات محددة، سواء كانت على شكل أفكار معبر عنها، أو سلوك ملاحظ، أو حلول لمشكلات يقدمها الفرد وفق خطوات متتابعة، أو استجابة لمنبهات مقننة، أو بعض الخصال الشخصية تتسم بالاستقرار نسبياً، ثم بعد ذلك يتم ترجمة هذه المفاهيم إلى وحدات صغيرة معيارية قابلة للقياس تمثل بنود المقياس، معنى ذلك أن المطلوب هو تحويل الأهداف إلى سلسلة من الخطوات والأعمال لنتمكن من كتابة البنود المناسبة التي يمكن أن تحقق تقديراً صادقاً للجوانب التي نقوم بقياسها، ويتم ذلك عن طريق تحديد عينة مقننة من هذا السلوك الذي يمثل السمة المقاسة بشكل جيد.

فإذا كان الغرض من المقياس الكشف عن الفروق الفردية بين الأفراد كان من الضروري استخدام الطرائق والأساليب الإحصائية الدقيقة التي تحقق هذا الهدف وتتيح التميز بينهم بالاعتماد على مستويات القدرة لديهم. وإذا كان الغرض من المقياس الكشف عن الصعوبات التي يعاني منها الأفراد وتشخيص جوانب القوة والضعف في أدائهم، كان لا بد أن يحتوي المقياس على عدد كبير من البنود لتغطية المجال المراد تشخيصه وأن تساعد هذه البنود على تحديد الصعوبة التي يعاني منها المفحوص وتحديدها، بالإضافة إلى تحديد نقاط القوة التي يمكن أن تفيد في مواجهة تلك الصعوبات. أما إذا كان المقياس يهدف إلى جمع معلومات وبيانات سريعة أو مسح سريع للسمة المقيسة بهدف التصفية أو الغربلة لأعداد كبيرة من المفحوصين، كان من الممكن التساهل قليلاً في بعض شروط الإجراء والاستعانة بأشخاص غير مدربين على التطبيق، هذا بالإضافة إلى إمكانية الاعتماد على معايير بسيطة وسهلة لتفسير نتائج المقياس.

٣-تحديد المجال الخاص الذي يقيسه المقياس وعينة السلوك المحددة له:

من الضرورة بمكان العمل على تحديد المجال الخاص لكل سمة يقيسها المقياس، فإذا كان المقياس موجهاً لقياس التطور اللغوي مثلاً أو العيوب النطقية أو نغمات الصوت مثلاً، فإنه من الضروري تحديد ما يتضمنه من عناصر أو مكونات رئيسة وفرعية، مع تحديد الأوزان النسبية لكل منها كي يتم تخصيص عدد البنود لكل عنصر استناداً إلى وزنه النسبي.

ولابد أيضاً من تحديد المجتمع للأفراد الذين سيطبق المقياس عليهم ، عند تحديد المجال الخاص الذي يقيسه المقياس، وذلك بهدف تجريبه وتعييره على عينات عشوائية من المجتمع الأصلي الذي تم تصميم المقياس من أجله. فإذا تم تصميم

المقياس لقياس ذكاء المرشحين لدخول الجامعة مثلاً، فإنه سيتم سحب العينة من الطلاب الذين تقدموا بطلبات التحاق للجامعة، وإذا صمم المقياس لقياس ذكاء المكفوفين أو ذوي الإعاقة مثلاً، سحبت العينات من تلك الفئة وهكذا.

٤-تحديد زمن المقياس وطوله:

لابد أن يقوم مصمم المقياس من تحديد الزمن الذي سيعطى للمفحوصين للإجابة عن بنود المقياس وعدد هذه البنود، ويمكن تحديد زم المقياس وطوله انطلاقاً من الأغراض التي يسعى لتحقيقها. فإذا كان المقياس يستخدم لأغراض الفرز والتصنيف وسيطبق على أعداد كبيرة من المفحوصين يجب أن يكون عدد البنود قليل في هذه الحالة، وأما إذا كان الغرض من المقياس هو تصنيف المفحوصين اعتماداً على مستويات القدرة لديهم، أو تصنيفهم في فئات مختلفة لاتخاذ قرار مهني او دراسي محدد، كان لابد من زيادة عدد البنود ورفع قدرتها التمييزية، وعندما يصبح الغرض من المقياس تشخيصي يصبح من الضروري جداً زيادة عدد البنود إلى الحدود القصوى المتاحة، وتغطية كافة جوانب المجال الذي يتم قياسه بهدف تعّف جوانب القوة وجوانب الضعف في أداء المفحوص أكثر من التركيز على مستوى صعوبة البنود. ويشار إلى أن زمن الاختبار يتحدد بنوع وشكل البنود، فإذا كان المقياس يتطلب انشاء استجابة من المفحوص سواء كانت كتابية أو لفظية أو أدائية، فإنه يختلف في عدد البنود عن المقياس الذي يتطلب اختيار اجابة مثل اسئلة الاختيار من متعدد .

٥-بناء بنود المقياس بصورتها الأولية:

تعد هذه الخطوة هي الخطوة الأهم والخطوة الحاسمة في عملية تصميم المقياس، وتعتمد هذه الخطوة على شخصية الفاحص مصمم المقياس بالدرجة الأولى، وتظهر براعته وقدرته الابداعية الخاصة على تطوير وتأليف البنود الاختبارية. ولا تقل هذه العملية أهمية عن عملية التحليل الكمي والكيفي لتلك البنود والتي تتطلب أن تسير وفق قواعد منهجية وأسس نظرية متفق عليها. ومن الضرورة الإشارة إلى أن عدد البنود في هذه المرحلة يجب أن لا يزيد زيادة كبيرة على العدد الكلي للبنود التي سيتضمنها المقياس بصورته النهائية. فإذا كان العدد المقرر لبنود المقياس هو مئة بند مثلاً كان من اللازم إعداد ما لا يقل عن ١٥٠ بند أو أكثر وذلك لإتاحة المجال أمام مصمم المقياس لغربلتها واختيار الأكثر صلاحية منها في المراحل اللاحقة.

٦-وضع تعليمات تطبيق المقياس:

يجب أن يتضمن المقياس تعليمات واضحة تحدد مهمة المفحوص بدقة عند قيامه بالإجابة على بنود المقياس، مما يمنع احتمال سوء الفهم من قبل المفحوص، ولاشك أنه من الضروري ذكر الوقت المعطى للإجابة وطريقة الإجابة عن البنود اللفظية وغير اللفظية، هذا بالإضافة إلى ما إذا كان سيسمح له بالتخمين أم لا عند الإجابة على البنود التي تفسح المجال

للتخمين. ومن الجدير بالذكر أنه يمكن تعديل التعليمات في ضوء نتائج التحليل الكيفي ونتائج التجربة الاستطلاعية والمبدئية التي يخضع لها الاختبار أثناء عملية التحليل الكمي.

٧-تحديد خطة تصحيح المقياس وتفسير النتائج:

من المهم تحديد طريقة تصحيح المقياس وتوزيع الدرجات على البنود الاختبارية المختلفة انطلاقاً من الوزن النسبي لكل منها، وتقرير ما إذا كان سيتم استخدام معادلة التصحيح من أثر التخمين في البنود الموضوعية. هذا بالإضافة إلى ضرورة توضيح طريقة تسجيل الإجابة هل سيتم على ورقة المقياس أم على ورقة أخرى منفصلة.

وفيما يتعلق بمسألة التخمين يميل معظم العاملين في القياس النفسي إلى ضرورة استخدام معادلة التصحيح من أثر التخمين ، فإذا كان المفحوص سيختار الإجابة الصحيحة من بين أربعة بدائل كان احتمال نجاحه عن طريق التخمين يساوي ٢٥% ، وإذا كانت بدائل الإجابة ثلاثة كان احتمال نجاحه عن طريق التخمين ٣٣%، وإذا كانت بدائل الإجابة اثنين كان احتمال نجاحه بالتخمين ٥٠%، وهكذا. وتتلخص الطريقة المتبعة في التصحيح من أثر التخمين في استخدام المعادلة التالية:

$$ع = ص - \frac{خ}{(ن-1)}$$

حيث يشير الرمز (ع) إلى العلامة المصححة.

ويشير الرمز (ص) إلى عدد البنود التي كانت الإجابة عنها صحيحة.

ويشير الرمز (خ) إلى عدد البنود التي كانت الإجابة عنها خاطئة.

ويشير الرمز (ن) إلى عدد البدائل في البند الواحد.

وتفترض المعادلة السابقة أن العلامة التي حصل عليها المفحوص هي أعلى من علامته الحقيقية نظراً للجوئه إلى التخمين وتعاقبه تبعاً لذلك، ولا تعير هذه المعادلة أي اهتمام للبنود المتروكة التي لم يجب عنها المفحوص اطلاقاً، ذلك أن هذه البنود لا تدخل ضمن عدد الإجابات الخاطئة.

٨-التحليل الكيفي للبنود:

يترافق هذا الإجراء مع عملية إعداد البنود وبنائها، كا يشكل خطوة لاحقة ومتممة لها. ويفيد في التأكد من أن البنود المعدّة تمثل بصدق السمة أو القدرة المقيسة والمعبرة عنها، كما يفيد في التأكد من أن هذه البنود تغطي الجوانب المختلفة للسمة أو القدرة المقيسة وتعكسها بحجمها الحقيقي. ويتناول التحليل الكيفي للبنود شكل البنود ومضمونها وملاءمتها للأغراض الخاصة، والتعليمات المرافقة مما يفيد في تأسيس صدق المحتوى أو الصدق المنطقي للاختبار، ومن الجدير بالذكر أنه لا توجد طريقة إحصائية تفيد في التوصل للصدق المنطقي للمقياس حيث يتم الاعتماد في هذه الخطوة على عملية التحليل

الكيفي فقط، وبالانتهاء من هذه الخطوة يكون المقياس بتعليماته وبنوده المختلفة قد أصبح جاهزاً للتطبيق التجريبي ثم التحليل الكمي.

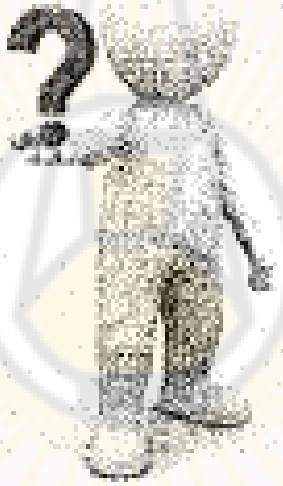
٩-التحليل الكمي للبنود:

يتم هذا النوع من التحليل بعد إجراء التجربة الاستطلاعية أو التجربة الأولية على عينات من الأفراد تشتق من المجتمع الصلي للأفراد الذين تم تصميم المقياس لأجلهم، ومن خلال هذا التطبيق يتم التعرف إلى مدى ملاءمة البنود للمفحوصين والتأكد من وضوح التعليمات وملاءمة الوقت المخصص للتطبيق، حيث يتم بعد ذلك إجراء بعض التعديلات الأولية اللازمة. واعتماداً على نتائج هذا التحليل يمكن تعرّف نقاط الضعف والقوة في المقياس والكشف عن مستوى السهولة والصعوبة والتمييز لكل مفردة من مفرداته والتحقق نهائياً من وضوح التعليمات وملاءمة البنود للمجال الذي وضعت لقياسه. وبعد ذلك تتم عملية تصفية البنود واختيار الأفضل لإعداد الصور النهائية للمقياس.

وتشتمل عملية التحليل الكمي للبنود مجموعة من الإجراءات من أهمها تحديد مستوى الصعوبة والسهولة لكل بند من خلال استخراج معامل السهولة والصعوبة الخاص بهذا البند ، بالإضافة إلى تحديد القدرة التمييزية للبنود أو ما يسمى صدق البنود، واستخراج مؤشرات الصدق والثبات.

الفصل الخامس

الصدق



الصدق

يعدُ مفهوما الصدق والثبات من أهم المفاهيم التي ترتكز عليها نظرية القياس النفسي إن لم تكن أهمها على الإطلاق. ومما يعزّز أهمية هذين المفهومين أنهما يمثّلان الشرطين الأساسيين من شروط صلاح الاختبار النفسي أيّاً كان نوعه. ومن المعلوم أن الاختبار النفسي يوضع أساساً لتلبية غرض أو مجموعة معينة من الأغراض، وليكون الأساس في اتخاذ عدد من القرارات المهمة. ولكي يؤدي الاختبار دوره على النحو الأمثل، ويلبي الغرض أو الأغراض المرسومة له، ويتيح، بالتالي اتخاذ قرارات صائبة لابد أن يقدّم معلومات واقعية وصحيحة حول السمة أو الخاصية التي توضع موضع القياس، أي يكون "صادقاً"، وأن يكون حساساً للفروق الدقيقة بين الأفراد في هذه السمة أو الخاصية، أي يكون "ثابتاً". وتتركز أهداف الفصل الحالي والفصل الذي يليه في إلقاء الضوء على هذين المفهومين ودراسة الطرائق التي يمكن اتباعها لتوفير قدر عال من الصدق والثبات في الاختبارات النفسية، والكشف عن الأهمية البالغة والمكانة التي يحتلها كل منهما في تلك الاختبارات.

معنى الصدق:

تجمع أدبيات القياس النفسي على أن مصطلح الصدق يشير أساساً إلى ما إذا كان الاختبار يقيس فعلاً ما أعدّ لقياسه، أو ما أردنا نحن أن نقيسه به. وقد أشار ن.ي.جرونلند في معرض حديثه عن الصدق إلى أن أقرب العبارات إلى هذا المفهوم وأكثرها دلالة عليه تلك العبارة التي يدلي بها الشاهد في المحكمة عادة التي تقول: "إنني أتعهّد أن أقول الحق، كل الحق، ولا شيء غير الحق". ومن التعريفات المهمة للصدق تعريف ليندكويست القائل: "إنه الدرجة التي يقيس بها الاختبار ما نريد قياسه"، وتعريف إدجرتون القائل: "إن الصدق يشير إلى المدى الذي تكون فيه أداة القياس مفيدة لهدف معين".

(نقلاً عن فرج، 1980، ص276 – 277)

ووفقاً لهذين التعريفين ولتعريفات أخرى عديدة لا يتسع المقام للوقوف عندها لابد من النظر إلى الصدق على أنه صفة نسبية لأداة القياس وليست مطلقة. وبدلاً من القول: إن الاختبار صادق أو غير صادق، من الأصوب أن نقول: إنه صادق بقدر ما. وهذا يعني بعبارة أخرى أن الصدق هو تعبير عن درجة، وأنه ليس ثمة اختبار صادق بصورة كلية أو مطلقة.

من جهة ثانية يرتبط الصدق مباشرة بالغرض أو الاستعمال الخاص الذي صمّم الاختبار من أجله، فالاختبار الذي أعدّ بغرض التنبؤ بأداء المفحوصين في فترة مقبلة قد لا يصلح لغرض التشخيص والكشف عن نقاط الضعف والقوة في أداء المفحوصين، أو لا يصلح لهذا الغرض بدرجة كافية. كما يرتبط الصدق مباشرة بالفئة أو الجماعة التي صمّم لها، فالاختبار الذي أعدّ لفئة أو فئات معينة من المعوقين قد لا يصلح للأسوياء، كما أن الاختبار الذي صمّم ليطبق على فئات معينة من الأسوياء، قد لا يصلح للمعوقين أو لجماعات أخرى من الأسوياء. ويتمثّل جوهر مفهوم الصدق في أن التعبيرات والصيغ العامة والمجردة قلّما تجدي في التعبير عن هذا المفهوم والكشف عن ماهيته، وبدلاً من استخدام تلك الصيغ والعبارات المجردة يستحسن ربط الصدق مباشرة بالغرض أو الاستعمال الخاص الذي صمّم الاختبار من أجله، ومراعاة ظروف الزمان والمكان وخصائص الجماعة التي سيطبق عليها.

وتجدر الإشارة إلى أن الأسماء التي تعطى للاختبارات النفسية لا تعبّر بالضرورة عما تقيسه هذه الاختبارات بالفعل. ومن هذا المنظور لا يصحّ أن يؤخذ الاسم الذي يحمله الاختبار على أنه تعبير أو انعكاس لما يقيسه هذا الاختبار، أو أنه مؤشر لصدقه. فالأسماء التي تحملها معظم الاختبارات النفسية تتّسم بالكثير من العمومية والغموض لدرجة يصعب معها تعرّف مجال السلوك الذي تغطيه هذه الاختبارات بدقة. وهذا يعني ببساطة أن السمة أو الخاصية التي يقيسها الاختبار فعلاً يتعذّر تحديدها من مجرد النظر في الاسم أو العنوان الذي يحمله هذا الاختبار، ولا بد من إخضاع هذا الاختبار لإجراءات معينة، قد يعتمد

الكثير منها على التجريب والإحصاء، لتحديد هذه السمة أو الخاصية المقيسة، والتثبّت من صدق الاختبار كما سنرى.

أشكال الصدق:

على الرغم من أن الباحثين استخدموا مصطلحات عديدة في التعبير عن الأنواع المختلفة من الصدق أو أشكال الصدق فإن التصنيف المعتمد لأنواع الصدق هو التصنيف الذي وضعته الجمعية الأمريكية لعلم النفس عام 1974 والذي جعل هذه الأنواع ضمن ثلاث فئات رئيسة، وهي: صدق المحتوى أو الصدق التمثيلي، والصدق المحكّي، والصدق البنيوي أو الافتراضي. وقد يكون من المفيد الإشارة، قبل الحديث عن هذه الأنواع، إلى أن بعض أنواع الصدق أكثر أهمية في أنواع معينة من الأدوات منها في أدوات أخرى. فصدق المحتوى مثلاً يؤدي دوراً بالغ الأهمية في اختبارات التحصيل على حين أن الصدق المحكّي يؤدي مثل هذا الدور في اختبارات القدرة والاستعداد. وبالطبع يمكن تصميم أداة القياس لتلبية أكثر من غرض واحد، وفي هذه الحالة لابد من الاستعانة بأكثر من نوع واحد من أنواع الصدق (Ahmann and Glock,1975). وهذا يعني أنه في الكثير من المواقف الاختبارية من الممكن والمفيد الجمع بين الأنواع المختلفة من الصدق.

- صدق المحتوى:

يرتكز صدق المحتوى أو الصدق التمثيلي كما يسمّى عادة على الفحص المنظم لمحتوى الاختبار لتحديد ما إذا كان عينة ممثلة لمحتوى الموضوع ومجال السلوك الذي يراد قياسه. ومن المعلوم أن الاختبار، أي اختبار، يتّجه أساساً إلى قياس عينة من السلوك، وأن مجموعة البنود التي يضمّها يفترض أنها عينة من المجتمع الأصلي " الأكبر" من البنود التي كان يمكن أن يضمّها. والواقع أن المشكلة الكبرى التي تواجه واضع الاختبار عادة هي ما إذا كانت مجموعة البنود التي أعدّها هي عينة ممثلة فعلاً لكل البنود المحتملة أو " للمجتمع الأصلي"

للبنود التي تنطوي عليها السمة أو الخاصية موضع القياس. ومع أنه لا تتوفر لدى واضعي الاختبارات عادة قائمة محدودة يمكن أن يشتقوا منها عينة من البنود فإن مطلب المعاينة يمثّل المطلب الأول والأهم في عملية تطوير الاختبار وبنائه. ويستحيل دون تحقيق هذا المطلب ضمان أن يقيس الاختبار ما أعدّ لقياسه وتمثيل السمة أو ميدان السلوك المقيس بصورة أمينة وصحيحة. ومن الواضح أنه بقدر ما تكون عينة البنود ممثلة للمجتمع الأصلي الأكبر من البنود بقدر ما يرتفع صدق المحتوى، وبقدر ما يضعف تمثيل العينة لهذا المجتمع يضعف صدق المحتوى. فإذا أمكن تحديد المحتوى الدراسي أو المجتمع الأصلي للبنود بـ 300 كلمة في الهجاء مثلاً تعلّمها الأطفال خلال الفصل واحتوى الاختبار على 60 كلمة منها، أي بلغت نسبة تمثيل العينة للمجتمع الأصلي 20٪ أمكن القول: إن حجم العينة مناسب، وإن الاختبار صادق إلى درجة كبيرة في محتواه. بيد أن المحتوى الذي يشكّل المجتمع الأصلي للبنود يصعب بل يستحيل تحديده وحصره كمياً على النحو السابق في معظم الأحيان. ومع ذلك فإن هذا النوع من أنواع الصدق يتطلب تحديداً ووصفاً دقيقاً للمحتوى أو المجال المقيس وتحليله إلى عناصره ومكوناته حتى يتمّ التأكد من أن البنود الموضوعة تغطّي الجوانب الأساسية لهذا المحتوى وبالنسب الصحيحة، ودون أن يطغى أي من هذه الجوانب على غيره أو يعطى له وزن أكبر من الوزن الذي يستحقه. ومن الواضح أن الاختبار بوصفه عينة مشتقة من مجتمع أصلي واسع ينطوي على قدر ما من خطأ العينة، وكلما ازداد عدد البنود التي يضمّها وكبرت العينة تضاءلت احتمالات الخطأ دون أن تنعدم، وأصبح الاختبار أكثر صدقاً وتمثيلاً للمجتمع الأصلي الافتراضي من البنود.

ليس ثمة صيغة كمية يمكن استعمالها للتحقق من صدق المحتوى أو تعبير عددي يمكن استعماله للدلالة عليه. فهذا النوع من أنواع الصدق يعتمد على التحليل المنطقي والمقارنة أو المطابقة بين محتوى الاختبار ومحتوى الموضوع المقيس، ولذلك يشار إليه في كثير من الأحيان بالصدق المنطقي. ثم إن هذا

النوع من الصدق لا يختص بنوع معين من أنواع المقاييس على الرغم من المكانة الخاصة التي يحتلها في مقاييس التحصيل. فالمقاييس التربوية والنفسية جميعها يجب أن تكون صادقة في محتواها بدرجة مناسبة، وأن يعكس أو يمثل كل منها السمة أو الخاصية موضع القياس على النحو الملائم. فإذا كنا بصدد إعداد مقياس للاتجاهات مثلاً فلابد أن نعمل على تغطية سائر الجوانب والمظاهر التي هي مدار البحث وبما ينسق مع الأهداف المرسومة، وكذلك الأمر فيما يتصل بمقياس الميول وغيره. بيد أن صدق المحتوى لا يكفي بحد ذاته لضمان الصدق في المقاييس النفسية، وقد يكون مضللاً إذا اقتصرت عليه هذه المقاييس نظراً لأنها "لا تتشابه تشابهاً" داخلياً كبيراً مع ميدان السلوك الذي تعدّ عينة منه ولا تعتمد على مقرر دراسي أو مجموعة موحدة من الخبرات السابقة يمكن منها صياغة مضمون الاختبار (أبو حطب، 1973، ص89). وعموماً فإن الاختبارات التي تستخدم لأغراض التنبؤ تتطلب الاعتماد على معايير أو محكات خارجية كما سنرى، ويصعب القول: إن هذه الاختبارات صادقة من مجرد فحص محتواها والتحقق من ملاءمتها للأهداف التي وضعت من أجلها. وقد يتعذّر تحديد الوظائف التي تقيسها تلك الاختبارات فعلاً دون الاعتماد على محكات عملية أو تجريبية تستخدم في التثبّت من صحة الفروض التي اعتمدت أساساً في بنائها. وتقول أنا أنستازي في هذا الصدد "مع أن صدق المحتوى يجب أن يدخل في المراحل الأولية لبناء أي اختبار فإن تأسيس الصدق (أو التصديق) Validation النهائي لاختبارات الشخصية والاستعداد يتطلب التحقق التجريبي بإجراءات أخرى... ومن المستحيل تحديد الوظائف النفسية التي يقيسها الاختبار من تفحّص بنوده. فالاختبار الذي يقيس الاستدلال الحسابي لدى طلاب المرحلة الثانوية قد يقيس الفروق الفردية في سرعة العمليات العددية لدى طلاب الجامعة" (Anastasi,1982,p.136).

- **الصدق الظاهري أو السطحي:**

من المهم التمييز بين صدق المحتوى أو الصدق المنطقي للاختبار والصدق الظاهري أو السطحي. فهذا الأخير ليس صدقاً بالمعنى الدقيق للكلمة ولا يشير إلى ما يقيسه الاختبار فعلاً بل إلى ما يبدو ظاهرياً أنه يقيسه سواء في نظر المفحوصين أنفسهم أم في نظر غيرهم من الأشخاص غير المختصين. ومما لاشك فيه أن الاختبار لابد أن يبدو صادقاً حتى يقبل المفحوصون عليه برغبة وتستثار دافعيتهم لأدائه بالحد الأقصى الذي تسمح به قدراتهم. وقد يؤدي غياب صفة الصدق الظاهري للاختبار حتى لو كان صدقه الحقيقي عالياً إلى موقف سلبي من جانب المفحوصين، واستيائهم مما ينعكس سلباً على أدائهم في الموقف الاختباري، كما قد يؤدي إلى التشكيك بالاختبار وقيمته ودوره في اتخاذ قرارات مهمة من جانب العديد من الأشخاص الذين قد يحكمون عليه من خلال مظهره وما يبدو لهم أنه يقيسه دونما تحليل دقيق لمحتواه وما قصده واضعه أن يقيسه أو ما يقيسه فعلاً. ومن المعلوم أن مقياس فكسلر – بلفيو ظهر أساساً لتغطية الحاجة إلى أداة ملائمة لقياس ذكاء الراشدين نظراً لأن المقاييس الشائعة في ذلك الحين بما فيها مقياس ستانفورد – بينيه أعدت للأطفال وتلاميذ المدارس، واستخدمت مواد مشوّقة للأطفال والتلاميذ بصورة خاصة مما جعلها تفتقر إلى الصدق الظاهري عند تطبيقها على الراشدين.

تؤكد آنا أنستازي أنه لا بد من توفير صفة الصدق الظاهري للاختبار حتى يكون أكثر فاعلية في المواقف العملية، ولضمان تعاون المفحوصين في الموقف الاختباري. ومن الواضح أن هذا التعاون والتجاوب يمثّل شرطاً ضرورياً لإظهار أقصى ما لديهم من قدرة، ويعدّ بالتالي، شرطاً ضرورياً لتوفير الصدق الحقيقي للاختبار. ويمكن أن يتحسن الصدق الظاهري بنظر أنستازي، بمجرد إعادة صياغة الأسئلة بحيث تبدو أكثر ارتباطاً بموقف القياس الفعلي. فمثلاً إذا أعدّ اختبار الاستدلال الحسابي للتطبيق على مجموعة من المشتغلين بالأعمال الميكانيكية فإن البنود يمكن صياغتها بحيث تبدو أكثر ارتباطاً بالعمل الميكانيكي ودون أن

يحدث أي تغيير في الوظيفة التي أعدت لقياسها (Anastasi,1982,p.136). وقد يكون من المفيد في الحالات التي يتعذّر فيها توفير الصدق الظاهري للاختبار تهيئة المفحوصين مسبقاً وإعلامهم بأن الاختبار ربما سيبدو ضعيف الارتباط بما درسوه لأن غرضه هو الكشف عن قدرتهم على تطبيق ما درسوه مثلاً في مواقف وأوضاع جديدة. وهنا قد تضعف الآثار السلبية لغياب هذا النوع من الصدق، ولا يكون الاختبار بمثابة "صدمة" أو تجربة مرة للمفحوصين. ومهما يكن من أمر فإن الصدق الظاهري للاختبار لابد أن يكون في خدمة الصدق الحقيقي لهذا الاختبار وليس بديلاً عنه، ولابد من تغليب الصدق الحقيقي حين يتعذّر الجمع بينهما.

- **الصدق المحكّي:**

يرتكز الصدق المحكّي على دراسة الارتباط بين درجات المقياس ودرجات مقياس آخر يعتمد كمحك. ويمكن تعريف هذا النوع من أنواع الصدق على أنه الدرجة التي يترابط عندها الأداء على مقياس مع الأداء على مقياس آخر عدّ محكاً للمقياس الأول وأساساً في الحكم على صلاحيته. ويتّضح من هذا التعريف أن الصدق المحكّي يتطلب إجراءات تجريبية إحصائية ويقوم على حساب معامل الارتباط بين الدرجات التي يحصل عليها المفحوصون في المقياس الذي يراد التثبّت من صدقه، والدرجات التي يحصل عليها أولئك المفحوصون أنفسهم في المقياس الثاني الذي عُدّ بمثابة محك. وكثيراً ما وصف هذا النوع من الصدق بالصدق التجريبي والصدق الإحصائي نظراً لاعتماده على التجريب ولغة الإحصاء، وبهدف تمييزه عن الأنواع الأخرى من الصدق التي تعتمد على المقارنة والتحليل المنطقي أو تعطيهما الأولوية.

يميّز بعض الباحثين بين نوعين لهذا الصدق وهما: الصدق التنبؤي Predictive Validity، والصدق التلازمي أو المصاحب Concurrent Validity، ويميل الكثير من الباحثين في الوقت الحاضر إلى دمج هذين النوعين في زمرة واحدة أو نوع واحد وهو الصدق المحكّي. ويرى ن. جرونلند أن هذا الدمج مناسب

"نظراً لأن طريقة التثبّت من الصدق والتعبير عنه واحدة في الحالتين" (Gronlund,1971,p.83). ومن الواضح أن كلاً من الصدق التنبؤي والصدق التلازمي يعتمد على استخدام محكّ، ويتطلب حساب معامل الارتباط بين الدرجات المتحصّلة على المقياس الموضوع والدرجات المتحصّلة على مقياس المحكّ. والفرق بينهما يرجع إلى الوقت الذي تجمع فيه بيانات المحكّ، كما يرجع إلى غرض القياس. ففي حين يتركز الاهتمام عند استخدام إجراءات الصدق التنبؤي بالكشف عما إذا كانت درجات المقياس تنبئ عن أداء المفحوصين أو إنجازهم في المستقبل مما يستدعي المقارنة بين درجات المقياس الأول (المتنبئ) ودرجاتهم على مقياس آخر للأداء اللاحق (المحكّ)، يتجه الاهتمام في حالة الصدق التلازمي أو المصاحب إلى تقدير الأداء الراهن للمفحوصين، ويتطلب هنا الصدق المقارنة بين درجاتهم على المقياس ودرجاتهم على المحكّ التي يحصلون عليها في الوقت نفسه تقريباً. وهذا يعني، بعبارة أخرى، أن الصدق التنبؤي يمتد بإطاره الزمني إلى المستقبل ويسعى إلى التعبير كمياً عن درجة الترابط بين نتائج الاختبار والأداء اللاحق للمفحوصين والذي يتوقع أن يترابط مع السمة أو الخاصية التي يقيسها الاختبار، على حين أن الصدق التلازمي يرجع في إطاره الزمني إلى الحاضر، ويطلعنا على درجة التوافق أو الترابط بين نتائج الاختبار ونتائج اختبار آخر يفترض أنه يقيس السمة ذاتها.

يفيد الصدق التنبؤي في الكشف عن فاعلية الاختبار في التنبؤ بنتائج معينة في المستقبل، ويلائم بخاصة اختبارات القدرات والاستعدادات، وقد يستعمل في بعض الاختبارات التحصيلية التي يكون غرضها الانتقاء والتصنيف. ويستعمل الصدق التلازمي بخاصة في الاختبارات التي يكون غرضها تشخيص الوضع الراهن بدلاً من التنبؤ بالأداء اللاحق أو بنتائج معينة في المستقبل، ويصلح بصورة خاصة لاختبارات الشخصية، وقد يستعمل في اختبارات القدرات والاستعدادات. والمسألة المركزية في كلا النوعين من الصدق هي توفير المحكّ الملائم، فمقياس المحكّ يمثّل العنصر الحاسم في عملية التحقق من الصدق المحكّي بنوعيه. وهذا

المقياس لابد أن يتصف بصفات معينة ويكون صالحاً بحد ذاته كي يصلح أساساً في الحكم على صلاحية غيره. وقبل الوقوف عند الصفات المرغوب بها في مقياس المحكّ قد يكون من المفيد التعرّف على أنواع المحكّات المستخدمة في المقاييس النفسية والتربوية.

أنواع المحكّات:

1. محك التحصيل الدراسي:

وهذا المحكّ هو أحد أهم المحكّات المستخدمة في التحقق من صدق اختبارات القدرة العقلية العامة (اختبارات الذكاء). وكثيراً ما توصف هذه المقاييس بأنها مقاييس الاستعداد المدرسي حيث تتمّ مقارنة الدرجات المتحصّلة عليها بالدرجات والتقديرات المدرسية ودرجات الاختبارات التحصيلية المختلفة.

غير أن محك التحصيل الدراسي لا يقتصر على اختبارات الذكاء العام، وقد يستخدم في حالات معينة في التحقق من صدق اختبارات القدرات الخاصة والمتعدّدة، كما يستخدم، ولكن بدرجة أقل، في التحقق من صدق اختبارات الشخصية أو جوانب معينة منها. ومن نافلة القول: إن الكثير من اختبارات التحصيل وخصوصاً تلك التي يعدّها المعلم كثيراً ما تلجأ إلى هذا المحكّ ذاته في تأسيس صدقها وخصوصاً حين تتوفر اختبارات تحصيلية مقننة تتمتّع بمواصفات فنية جيدة ويمكن اعتمادها بالتالي محكات للصدق.

2. محك الأداء في برنامج تعليمي أو تدريبي متخصص:

ويستخدم هذا المحكّ في التحقق من صدق اختبارات القدرات والاستعدادات الخاصة، ويرتكز على أداء الأفراد خلال مرحلة التعليم أو التدريب المتخصص. فمثلاً الأداء في مدارس الفن والموسيقى يستخدم في تقدير صدق اختبارات الاستعداد الفني أو الموسيقي. والعديد من اختبارات الاستعدادات المهنية

يتمّ تقدير صدقها باستخدام محكات التحصيل في كليات الحقوق، والطب، وطب الأسنان، والهندسة، وغيرها. وكذلك اختبار انتقاء الطيارين فإن محكه هو أداء الطيارين في مواقف التدريب الفعلية (Anastasi,1982, p. 140).

3. محك الأداء في العمل نفسه:

وهو أفضل من محك سجلات التدريب السابق نظراً لأنه يرتكز على سجلات تتبّعية للإنجاز الفعلي في العمل. ومن أمثلته سجل الإنتاج الفعلي للفرد على شكل عدد القطع المصنعة، أو عدد الأخطاء المرتكبة في سجلات محاسب. غير أن هناك الكثير من الأعمال مثل طبيب، ومعلم، وسكرتيرة، التي لا يمكن فيها الحصول على سجل موضوعي للإنجاز أو الإنتاج ولذلك كثيراً ما يتمّ اللجوء إلى تقديرات الرؤساء التي قد تتأثر بعوامل ذاتية وقد تنحاز سلباً أو إيجاباً ولا تعطي صورة صادقة عن إنجاز الفرد (ثورندايك وهيجن، 1989، ص60).

4. طريقة الفرق المتقابلة:

وتنطوي على محك أكثر تعقيداً وأقل تحديداً من المحكات السابقة. وتعتمد على المقارنة كأن نعمد إلى مقارنة درجات الطلاب في معهد الموسيقى بدرجات مجموعة غير منتقاة من طلبة الجامعة، وهذه الطريقة شائعة في اختبارات الشخصية أو اختبارات الاتجاهات والميول. ففي اختبارات الشخصية كثيراً ما تتمّ المقابلة بين العصابيين والأسوياء، وفي مقاييس الميول قد نلجأ إلى الجماعات المهنية المختلفة، وفي مقاييس الاتجاهات قد تستخدم الجماعات السياسية أو الدينية المتعارضة أو أي جماعات خاصة أخرى يمكن أن تعطي وجهات نظر متباينة أو متمايزة بصدد المسائل الخاضعة للدراسة.

5. طريقة الفروق الطرفية:

وتقترب هذه الطريقة من الطريقة السابقة من حيث أنها تقوم على مفهوم قدرة الاختبار على التمييز بين طرفي القدرة التي يقيسها، وتعتمد على المقارنة وحساب دلالة الفروق. غير أن المقارنة هنا تخص الفئتين العليا والدنيا من أفراد مجموعة واحدة ولا تخصّ مجموعتين من الأفراد. ويمكن إجراء هذه المقارنة بإحدى الطريقتين التاليتين:

1. مقارنة الفئات المتطرفة في الاختبار والمحكّ الخارجي: وبحسب هذه الطريقة تتمّ مقارنة الربع (أو الثلث) الأعلى في درجات الاختبار بالربع (أو الثلث) الأعلى في درجات المحكّ الخارجي، كما تتمّ مقارنة الربع (أو الثلث) الأدنى في درجات الاختبار بالربع (أو الثلث) الأدنى في درجات المحكّ الخارجي. وتجري هذه المقارنة عن طريق حساب الدلالة الإحصائية للفروق بين المتوسطات. فإذا لم تظهر فروق دالة إحصائياً عند مقارنة الربع الأعلى في درجات الاختبار بالربع الأعلى في درجات المحكّ، وإذا لم تظهر مثل هذه الفروق عند مقارنة الربع الأدنى في درجات الاختبار بالربع الأدنى في درجات المحكّ أمكن القول: إن الاختبار صادق بدلالة هذا المحكّ.
2. مقارنة الفئات المتطرفة في الاختبار نفسه: وتعتمد هذه الطريقة على المقارنة بين الفئات المتطرفة في الاختبار ذاته كأن يؤخذ الربع (أو الثلث) الأعلى من الدرجات المتحصّلة في هذا الاختبار (والذي يمثّل الفئة العليا) ويقارن بالربع (أو الثلث) الأدنى للدرجات فيه (والذي يمثّل الفئة الدنيا)، ثم تحسب الدلالة الإحصائية للفرق بين متوسطي الفئتين فإذا ظهرت هذه الدلالة عدّ الاختبار صادقاً (بدلالة الفرق بين الفئتين العليا والدنيا). وكثيراً ما تستخدم هذه الطريقة عند دراسة القدرة التمييزية للاختبار في المراحل الأولى لإعداده وبنائه.

وبالإضافة إلى المحكّات السابقة كثيراً ما يؤخذ الترابط بين درجات الاختبار التجريبي (الجديد) ودرجات اختبار آخر من نوعه أو قريب منه على أنه دليل للصدق. فمثلاً مقياس ستانفورد – بينيه كثيراً ما عُدّ محكاً للاختبارات الجماعية للذكاء. وترى أنستازي أن هذه الطريقة تفيد بخاصة حين يكون الاختبار التجريبي الجديد بديلاً مبسطاً أو مختصراً للاختبار الأصلي حيث لابد من اللجوء إلى هذه الطريقة عندئذ وعُدّ الاختبار الأصلي محكاً للاختبار الجديد (Anastasi,1982,p.142).

صفات المحكّ:

ليس من السهل إيجاد محك دقيق وملائم: فمقاييس المحكّ جميعها جزئية ونسبية ولكل منها مشكلاته الخاصة وحدوده. والمحكّ النهائي، كما يقول ثورانديك: "لابد أن يتناول تقويماً لنجاح الإنسان في مهنته في الحياة اليومية. غير أن هذا المحكّ النهائي غير متيسر، ولابد للباحث أن يكتفي ببدائل هي في الغالب جزئية وغير كافية" (ثورانديك وهيجن، 1989، ص60). ومن هنا لابد من العمل على اختيار البديل الأفضل والأكثر ملاءمة من بين البدائل المتاحة.

ويرى الكثير من الباحثين أنه لابد من توفر الصفات التالية في مقياس المحكّ وهي:

1. الصلة الوثيقة بالموضوع (Relevance):

ويعتبر المحكّ وثيق الصلة بالموضوع بقدر ما يكون الأداء على مقياس المحكّ مناظراً أو ممثلاً للنجاح في العمل. غير أن تقدير هذه الصلة يخضع للنظرة الشخصية الذاتية وقد يختلف من مقدّر لآخر، وليس هناك أي دليل تجريبي يمكن أن يدلنا على كم هذه الصلة. ولا بد من الاعتماد على أحكام المختصين لتقدير الدرجة التي يكون عندها مقياس المحكّ، وهو مقياس جزئي

ونسبي بطبيعة الحال، وثيق الصلة بالمحكّ النهائي للنجاح في العمل والدراسة (ثورانديك وهيجن، 1989، Mehrens,1973).

2. الثبات:

فمقياس المحكّ يجب أن ينطوي على قدر عالٍ من الثبات وبحيث نحصل عند تكراراه على نتائج متقاربة جداً. ومن الواضح انه إذا تذبذبت العلامات أو التقديرات على مقياس المحكّ من وقت لآخر افتقد إلى الاتساق والدقة. ويتعذّر في هذه الحالة الوثوق به واستعماله كمحك.

3. الخلو من الانحياز:

وتعني هذه الصفة ان مقياس المحكّ يجب أن يتيح فرصاً متساوية للأفراد لإظهار اقصى ما لديهم من اداء والحصول على العلامات أو التقديرات التي يستحقونها. والأمثلة على عوامل التحيز كثيرة. فالرؤساء قد يتفاوتون في تقديراتهم لمرؤوسيهم تفاوتاً كبيراً في بعض الأحيان، وإذا عرف المعلم أن طالباً ما حصل على علامة منخفضة في اختبار القدرة أو الاستعداد فإن هذه المعرفة قد تؤثر في الدرجة التي سيعطيها للطالب. وقد يميل هذا المعلم إلى إعطاء علامات عالية لمن حصل على علامات عالية فقط في اختبار القدرة. وهذا ما يرفع الترابط بين درجات الاختبار والدرجات على مقياس المحكّ بصورة مصطنعة (Anastasi,1982). ومن الواضح انه كلما أمكن التغلب على عوامل الانحياز تضاءلت أخطاء التقدير وازدادت قيمة المقياس المعتمد كمحك.

لقد ظهر مصطلح خاص للتعبير عن احتمالات الخطأ (أو التحيزات) التي يمكن أن تظهر عند استخدام مقياس المحكّ، ويعرف هذا المصطلح بـ "تلوّث المحكّ" Criterion Contamination. ويعود السبب في إطلاق هذه التسمية أو المصطلح إلى ان التقديرات التي يعطيها مقياس المحكّ قد تتأثر بقوة " او تتلوّث" بفعل المعرفة السابقة للمقدّر بدرجة المفحوص في الاختبار. وهذا ما يظهر واضحاً

حين يكون المدرس أو الرئيس في العمل مثلاً على علم مسبق بأن علامة فلان متدنية للغاية، أو مرتفعة للغاية، فيتأثر بها بشكل أو بآخر، ويعطي تقديرات موازية لها (أي تقديرات متدنية إذا كانت العلامة متدنية، ومرتفعة إذا كانت العلامة مرتفعة). وتبعاً لذلك فإن الأشخاص الذين ستعتمد تقديراتهم أساساً في الحكم على المفحوصين (أي تستخدم كمحك) لا يجوز إطلاعهم بأي حال من الأحوال على درجات المفحوصين في الاختبار موضع الدراسة. وهذا يعني أن درجات الاختبار التي ستستخدم في "اختبار الاختبار" على حدّ تعبير أنستازي وأوربينا، يجب المحافظة على سرّيتها وعدم تسريبها. هذا مع الإشارة إلى أنه ليس من السهل دائماً إقناع الأشخاص الذين تعتمد تقديراتهم كمحك (كالمدرسين والرؤساء في العمل، والضباط وغيرهم) بأن الحذر في هذه الأمور هو شيء ضروري، وقد يفشل هؤلاء الأشخاص تماماً في إدراك حقيقة أن درجات المفحوصين يجب أن تبقى جانباً حتى "تنضج" البيانات المحكّية، ويصبح بالإمكان القيام فعلاً بحساب الصدق المحكّي للاختبار (Anasstasi&Urbina,1997,p.120)

4. جدوى المحكّ أو مدى تيسره (Availability):

فعند اختيار مقياس المحكّ لا بدّ من مراعاة أمور واعتبارات عملية من مثل الكلفة والزمن، ولا شك أن الاعتبارات العملية على أهميتها لا يجوز أن تعطى الأولوية أو تكون على حساب الصفات والشروط الأخرى السابقة.

جداول التوقع:

تستخدم جداول التوقع في التحقق من الصدق التنبؤي للمقياس. فبدلاً من محاولة الحصول على تنبؤات دقيقة حول الأداء المقبل للأفراد عن طريق معامل الارتباط يمكن الاستعانة بجداول التوقع والتنبؤ من خلال زمر أو فئات واسعة يعبر عنها برتب أو تقديرات من مثل "متوسط" أو " تحت المتوسط" أو "فوق

المتوسط". وتقدّم هذه الطريقة فائدة قصوى بخاصة للمعلم نظراً لبساطتها وسهولة استخدامها. ويمكن إيضاحها من خلال الجدول التالي:

الجدول رقم (6): جدول توقع لدرجات اختبار الاستعداد المدرسي ودرجات اختبار تحصيلي في الدراسات الاجتماعية:

درجات اختبار الاستعداد المدرسي	درجات اختبار الدراسات الاجتماعية			
	تحت المتوسط (دون 55)	متوسط (55 – 75)	فوق المتوسط (فوق 75)	المجموع
فوق المتوسط (فوق 115)	0	2	3	5
متوسط (95 – 115)	2	7	2	11
تحت المتوسط (دون 95)	3	1	0	4
المجموع	5	10	5	20

واستناداً إلى هذا الجدول يمكن القول: إنه لا أحد من التلاميذ ممن درجاتهم في اختبار الاستعداد فوق المتوسط (أكثر من 115) حصل على تقدير تحت المتوسط (أقل من 55 درجة) في اختبار التحصيل. كذلك يمكن القول: إنه لا أحد من التلاميذ ممن كانت درجته في اختبار الاستعداد تحت المتوسط (أقل من 95) حصل على علامة في اختبار التحصيل تزيد على المتوسط (أكثر من 75). وعدد التلاميذ الذين حصلوا على 95 أو أكثر في اختبار الاستعداد وحصلوا على درجات في اختبار التحصيل أقل من 55 كان 2 من أصل 16. ويصبح من السهل تفسير هذا الجدول إذا استعملت النسب المئوية. كما يظهر في الجدول التالي. (نقلاً عن: Ahmann and Glock, 1975, p.233)

الجدول رقم (7): جدول توقع يرتكز على النسب المئوية:

الاستعداد المدرسي		درجة اختبار الدراسات الاجتماعية			
الفئة	درجات الاختبار	عدد الحالات	تحت المتوسط ٪	متوسط ٪	فوق المتوسط ٪
فوق المتوسط	فوق 115	5	0	40	60
متوسط	95 – 115	11	18	64	18
تحت المتوسط	تحت 95	4	75	25	0

ومن قراءة هذا الجدول يمكن القول مثلاً: إنه إذا حصل التلميذ على تقدير متوسط في اختبار الاستعداد المدرسي فإن احتمال حصوله على تقدير تحت المتوسط في اختبار التحصيل هو 18٪ (أي 2 من أصل 11)، واحتمال حصوله على تقدير متوسط في اختبار التحصيل هو 64٪ (أي 7 من أصل 11)، واحتمال حصوله على تقدير فوق المتوسط في اختبار التحصيل هو 18٪ (أي 2 من أصل 11)، كما يمكن القول أيضاً إنه من فئة فوق المتوسط في اختبار الاستعداد المدرسي (وعدد أفرادها 5) حصل 40٪ على تقدير متوسط في اختبار التحصيل و60٪ على تقدير فوق المتوسط فيه0 ومن فئة المتوسط في اختبار الاستعداد المدرسي (وعدد افرادها 11) حصل 18٪ على تقدير تحت المتوسط في اختبار التحصيل و64٪ على تقدير متوسط و18٪ على تقدير فوق المتوسط، ومن فئة تحت المتوسط في اختبار الاستعداد المدرسي (وعدد أفرادها 4) حصل 75٪ على تقدير تحت المتوسط في اختبار التحصيل و25٪ على تقدير متوسط في هذا الاختبار.

ولابد من الإشارة إلى أن جداول التوقع تفيد بخاصة في الاختبارات التي تستخدم لأغراض الانتقاء حيث يمكن استناداً إليها تحديد العلامة الدنيا أو الحد الأدنى للقبول. فإذا وجدنا مثلاً أن من نجح في تخصص دراسي أو برنامج تدريبي معين كان قد حصل على 75٪ فما فوق في امتحان القبول نثبّت هذه العلامة كحد أدنى للقبول. ويمكن القول في هذه الحالة إن امتحان القبول يتصف بالصدق التنبؤي إذ يمكننا من التنبؤ بنجاح من يحصل على 75٪ فما فوق ورسوب من يحصل على أدنى من ذلك.

ملاحظات مهمة حول دراسة الصدق المحكّي للاختبار:

تثير دراسة الصدق المحكّي للاختبار عدداً من الأسئلة والاستفسارات التي يتعين على العامل في القياس أخذها بالحسبان. وهذا ما يتّضح من خلال الملاحظات التالية:

1. **تجانس مجموعة المفحوصين:**

بما أن معامل الصدق المحكّي هو معامل ارتباط بين درجات الاختبار موضع الدراسة ودرجات المحكّ فإن كل ما يؤثّر في قيم معامل الارتباط يؤثّر في قيم معامل الصدق المحكّي. ويستتبع ذلك أنه كلما كانت مجموعة المفحوصين متجانسة انخفضت قيمة معامل الصدق المحكّي، وكلما كانت متباينة وازداد تباين الدرجات التي يعطيها الاختبار ارتفعت قيمة معامل الصدق المحكّي. وتبرز مسألة تجانس مجموعة المفحوصين في الاختبارات التنبؤية التي تعتمد نتائجها أساساً في انتقاء الأفراد. ففي هذه الاختبارات يستخرج معامل الصدق التنبؤي عادةً من النتائج المتحصّلة من تطبيق الاختبار التنبؤي على مجموعة الأفراد الذين تمّ انتقاؤهم بالفعل. وهذه المجموعة "المنتقاة" تكون بطبيعة الحال أكثر تجانساً من المجموعة الأصلية الأولى التي اختيرت منها المجموعة المنتقاة (أي

المجموع الكلي من الأفراد الذين تقدّموا للالتحاق بدراسة أو عمل معين قبل أن تتمّ عملية الانتقاء).

إن دراسة الصدق التنبؤي للاختبار الذي أعدّ بغرض الانتقاء تعتمد على الأفراد المقبولين عادة. ويستخرج معامل الصدق التنبؤي من خلال حساب الترابط بين درجات أولئك الأفراد المقبولين تحديداً في الاختبار التنبؤي ودرجاتهم في الاختبار المحكّي (أو المحك)، في حين ان الأفراد الذين طبق عليهم الاختبار ولم يقبلوا تستبعد درجاتهم تماماً. وهذا يعني ببساطة أن معامل الصدق التنبؤي يستخرج عادة من أداء مجموعة منتقاة. وهذا ما يؤدي بطبيعة الحال إلى ضيق مدى الدرجات التي يعطيها هذا الاختبار وضعف تباينها نظراً لأن هذه المجموعة تعدّ أكثر تجانساً من مجموعة الأفراد الذين سيطبق عليهم الاختبار في المستقبل بغرض انتقاء أفضلهم.

وفي ضوء ما سبق فإنه سيكون من المفيد عند تقويم الصدق التنبؤي للاختبار استناداً إلى معامل صدقه عدم التسرع في الحكم بعدم صلاحية هذا الاختبار، أو التشكيك بفاعليته في التنبؤ بالنجاح في الدراسة أو الأداء في العمل. فقد يعود انخفاض معامل الصدق التنبؤي إلى تجانس "أو ضعف تباين" أفراد المجموعة الذين طبق عليهم هذا الاختبار واستخرج من أدائهم بالذات هذا المعامل. وقد يعود إلى عامل آخر أو أكثر من عوامل أخرى عديدة يمكن أن تؤثر في الصدق التنبؤي للاختبار وتهبط بمعامل الصدق المستخرج.

2. ثبات درجات المحكّ:

فالمحكّات التي تستخدم في دراسة صدق الاختبار قد لا تكون على درجة عالية من الثبات أو الاتساق بحد ذاتها وخصوصاً حين تعتمد على التقديرات الذاتية ونتائج الملاحظة، كتقديرات الرؤساء في العمل، أو ملاحظات المعلم اليومية لأداء التلميذ، أو درجات الامتحانات المدرسية، أو غيرها. وقد يكون من

المناسب في الحالات التي يصعب، أو يتعذّر فيها إيجاد محكات على درجة عالية من الوثوقية والثبات استخدام أكثر من محكّ واحد، مع التنبّه إلى أن معاملات الصدق المستخرجة بدلالة هذه المحكّات لا تحمل المعنى ذاته الذي تحمله معاملات الصدق المستخرجة بدلالة محكات أخرى على درجة عالية من الثبات، كالاختبارات المقننة مثلاً.

3. الفترة الزمنية الفاصلة بين تطبيق الاختبار وتطبيق المحكّ:

فكلما ازدادت الفترة الزمنية الفاصلة بين تطبيق الاختبار الذي يدرس صدقه وتطبيق الاختبار الثاني الذي اعتمد محكاً للاختبار الأول ازداد احتمال ظهور عوامل دخيلة يمكن أن تؤثّر في الأداء الاختباري للمفحوصين، وتؤدي إلى خفض معامل الصدق المستخرج. وبالمقابل فإنه كلما تضاءلت الفترة الزمنية الفاصلة تضاءل الأثر المحتمل للعوامل الدخيلة التي يمكن أن تهبط بمعامل الصدق المستخرج. ولعل هذا الأمر بالذات هو من بين العوامل والأسباب التي تجعل معاملات الصدق التنبؤي أدنى عموماً من معاملات الصدق التلازمي، كما تجعل معاملات الصدق التنبؤي المستخرجة بعد انقضاء فترة زمنية طويلة (3سنوات مثلاً أو أكثر) أدنى عموماً من معاملات الصدق التنبؤي المستخرجة بعد مرور فترة زمنية قصيرة نسبياً (3 أو 6 أشهر مثلاً).

4. عدد المفحوصين:

ولعدد المفحوصين أثره الواضح في الاختبارات التنبؤية خاصة. فقد تطول الفترة الزمنية الفاصلة بين تطبيق الاختبار التنبؤي والمحكّ، ويؤدي ذلك إلى انقطاع بعض المفحوصين عن الدراسة، أو تسرّبهم، أو انتقالهم، أو تعرّضهم لظروف طارئة أخرى. ويجعل من الصعب، بالتالي، وربما من المستحيل، إخضاعهم للاختبار المحكّي. والواقع أنه ليس من النادر أن يجد الفاحص نفسه أمام عدد ضئيل من الأفراد الذين طبق عليهم الاختبار التنبؤي. لهذا السبب فقد يكون من

الضروري عند دراسة الصدق التنبؤي للاختبار التنبّه مسبقاً لهذا الأمر وسحب عينة من الأفراد أكبر من العدد اللازم (علام ، 2000).

تصحيح معامل الصدق:

سبقت الإشارة إلى أن معامل صدق الاختبار يتأثر مباشرة بثبات درجات المحكّ المعتمد في استخراج هذا المعامل، وأنه كلما كان المحكّ المعتمد على درجة عالية من الموثوقية والثبات أمكن الوثوق بدرجة أكبر بمعامل الصدق المستخرج بدلالة هذا المحكّ.

غير أن معامل صدق الاختبار لا يتأثر فقط بثبات درجات المحكّ المعتمد بل يتأثر أيضاً بثبات درجات الاختبار ذاته. فإذا لم يكن الاختبار موضع الدراسة على درجة عالية من الثبات والاتساق بضعف معامل صدقه الذاتي، والذي يمثّل الجذر التربيعي لمعامل الثبات، وهذا ما يؤثر سلباً في معامل صدقه المحسوب بدلالة المحكّ المعتمد. لهذا السبب يصبح من الضروري إعادة تقدير معامل الصدق بين الاختبار والمحكّ في ضوء معامل الثبات الخاص بالاختبار إضافة إلى معامل ثبات المحكّ.

ولإجراء تصحيح شامل لمعامل صدق الاختبار في ضوء ثبات كل من الاختبار والمحكّ تستخدم المعادلة التالية:

$$\text{رح أ} = \frac{\text{ل س ص}}{\sqrt{\text{ل س}^2 \times \text{ل ص}^2}}$$

حيث يشير الرمز رح أ إلى الارتباط بين المكونات الحقيقية أو التباين الحقيقي في درجات المتغيرين س، ص (الاختبار والمحكّ).

$ر_{س ص}$ إلى معامل الارتباط المحسوب بين الاختبار والمحكّ.

$ر_{س}$، $ر_{ص ص}$ إلى معاملات ثبات كل من س وص (أي الاختبار والمحكّ).

غير أن الاختبار، أي اختبار، لابد أن ينطوي على درجة ما من الخطأ، ولا يوجد اختبار ثابت تماماً. ومن هذه الزاوية فقد نكون بحاجة لحساب صدق الاختبار في ضوء وجود هذا القدر من الخطأ والاعتراف به دون اللجوء إلى تصحيح أخطاء الثبات. وفي هذه الحالة يمكننا استخدام صيغة مختصرة للمعادلة السابقة يستبعد منها معامل الثبات الخاص بالاختبار وتبقي على معامل ثبات المحكّ، وذلك على النحو التالي:

$$ر_{ح ا} = \frac{ر_{س ص}}{\sqrt{ر_{ص ص}}}$$

حيث يشير الرمز $ر_{ح ا}$ إلى الارتباط بين التباين الحقيقي في المحكّ (ح) والاختبار (ا).

ويشير الرمز $ر_{س ص}$ إلى معامل الارتباط المحسوب بين الاختبار والمحكّ.

ويشير الرمز $ر_{ص ص}$ إلى معامل ثبات المحكّ.

ولابد أن نلاحظ عند استخدام المعادلة الخاصة بتصحيح معامل الصدق أن هذه المعادلة لا تصحح أخطاء العينة، ولا أخطاء القياس التي قد يكون لها دورها في التأثير في معامل الصدق الذي نحصل عليه بحساب الارتباط بين الاختبار والمحكّ. (Guilford,1954,p.400، نقلاً عن: فرج، 1980، ص327)

- **الصدق البنيوي (الافتراضي):**

ويشير هذا النوع من الصدق، والذي يطلق عليه أيضاً صدق التكوين الفرضي، إلى الدرجة التي يقيس معها الاختبار السمة أو الخاصية التي يفترض أن يقيسها، وينطلق من أن الدرجات على الاختبار يجب أن تتنوع أو تتباين كما تتنبأ النظرية الخاصة بالسمة المقيسة أو المفهوم أو التكوين المفترض. ومن أمثلة هذه السمات أو التكوينات الفرضية (Constructs) الذكاء، والقدرة الرياضية أو العددية، والقدرة على الاستدلال، والفهم القرائي، والتفكير النقدي، والعصابية، والقلق... إلخ. والصدق البنيوي هو مفهوم شامل يتضمّن الأنواع الأخرى من الصدق ويعدّ بالتالي شرطاً مهماً للاختبارات النفسية والتربوية بأنواعها. وبدلاً من سؤال: " هل يقيس الاختبار ما يدعي واضعه أنه يقيسه" يصبح السؤال المركزي في هذا النوع من الصدق: "ما الذي يقيسه هذا الاختبار بالضبط؟" أو "ما معنى العلامات المتحصّلة على الاختبار؟" أو "هل تشير هذه العلامات إلى سمة بارزة أو بناء معين؟". ومن المعلوم أن أدوات القياس لا تمدّنا دائماً بقياسات "نقية" للسمات التي وضعت لقياسها. ومن هذه الزاوية يؤدي هذا النوع من الصدق دوراً مهماً من حيث أنه يفيد في تحديد طبيعة وقوة العوامل التي تؤثّر فعلاً في الأداء الاختباري. فعند تفحص هذا النوع من الصدق "لا ينحصر اهتمامنا بالسمة النفسية التي صمّم الاختبار لقياسها، وقد نهتم بأي عامل آخر يمكن أن يؤثر في الدرجات، فمثلاً قد يدعي واضع الاختبار أن اختباره يقيس الاستدلال الحسابي ولكن نحن قد نتساءل عن احتمال تأثر الدرجات بالقدرة القرائية أو السرعة أو عوامل أخرى مشابهة" (Gronlund,1971,p.91).

يتطلب الصدق البنيوي اللجوء إلى الاستنتاجات المنطقية بالإضافة إلى الوسائل التجريبية والإحصائية. والواقع أن الصدق البنيوي يتضمّن أنواع الصدق كافة كما أسلفنا. والتقنيات والإجراءات الخاصة بتأسيس صدق المحتوى والصدق المحكّي يمكن أن تندرج ضمن تقنيات وإجراءات الصدق البنيوي. فمقارنة الأداء الاختباري للفرق أو المجموعات المتقابلة كالعصابيين والأسوياء هو طريقة

من طرائق التحقق من الصدق البنيوي، وهو بطبيعة الحال طريقة للتحقق من الصدق المحكّي. ودراسة الترابط بين اختبارات الاستعداد الميكانيكي والأداء في أنواع عديدة من العمل تسهم في فهمنا للسمة التي يقيسها الاختبار مما يشير إلى دورها في الصدق البنيوي، بالإضافة إلى كونها طريقة في الصدق المحكّي. ويمكن تحديد الخطوات الواجب اتباعها في عملية التحقق من هذا النوع من الصدق على النحو التالي:

1. الدراسة المنطقية للأداة وتحديد السمات النفسية المفترضة أو البنى التي يمكن أن ينطوي عليها الأداء على تلك الأداة.
2. الدراسة المنطقية للسمات أو البنى التي تم تحديدها وتكوين فرضيات قابلة للاختبار حول مدى صلتها بالأداء على تلك الأداة انطلاقاً من نظرية تطرح تلك السمة (أو السمات).
3. إجراء بحوث للتحقق من كل فرضية بالوسائل التجريبية.

فإذا أردنا على سبيل المثال التحقق مما إذا كان اختبار ما يقيس الذكاء فيمكن أن نفترض أن الدرجات على هذا الاختبار تزيد بزيادة العمر، فإذا لم تزد يمكن أن نستنتج أحد أمرين: إما أن يكون الاختبار مقياساً غير صادق للذكاء، أو أن يكون هناك خطأ ما في النظرية التي كانت بمثابة المنطلق في تصميم هذا الاختبار.

يؤكد أريكسون أن طول عملية التحقق من الصدق البنيوي وتعقدها يجب ألا يكونا مسوّغاً لتجنبها. فليس بمستغرب، على سبيل المثال، أن نجد بين الطلاب من يظهر مهارة أو موهبة خاصة في الأوضاع الطبيعية ويعجز عن إظهار الحد الأدنى من تلك المهارة أو الموهبة في الموقف الاختباري. وفي هذه الحالة يمكن الافتراض بأن الاختبار يقيس القلق (بسبب حدود الوقت المفروضة أو إجراءات الموقف الاختباري الضاغطة). وليس بمستغرب أيضاً أن نجد بين الطلاب من يظهر أداءً عالياً في المختبر والمخزن وصالة الرسم ويظهر مع ذلك أداء ضعيفاً

على اختبارات الورقة والقلم المصمّمة لقياس التحصيل في تلك المجالات. ويمكن الافتراض في هذه الحالة أن الاختبار يقيس المهارة اللفظية أكثر مما يقيس الأداء في تلك المجالات (Erickson,1979,p.p.30 – 31). وهذا كله يظهر أهمية هذا النوع من الصدق من حيث أنه يتجه إلى الكشف عما يقيسه الاختبار فعلاً ولا ينحصر بالتأكد مما أراد واضعه أن يقيسه. والواقع أن الصدق البنيوي بوصفه مفهوماً معقداً وشاملاً لم يفهم دائماً بوضوح، وقد نظر إليه بعضهم وكأنه صدق المحتوى معبراً عنه من خلال السمة أو الخاصية النفسية المفترضة ولذا اعتمدوا على التقديرات الذاتية المحضة في التثبّت منه. ويؤكد أنستازي أن الصدق البنيوي ينطوي على قدر كبير من الأهمية، وأن هذا الصدق إذا طبق بطريقة متراخية فقد يفتح الطريق لنوع من التأكيدات الذاتية غير المحققة عن صدق الاختبار، وكل من يظن أن الصدق البنيوي يتمّ حين يفسر الاختبار كمقياس لخاصية لا يمكن تحديدها إجرائياً يرتكب خطأ فاحشاً برأي أنستازي "ولكي نكتشف ما يقيسه الاختبار فعلاً لابد أن نلجأ إلى التحقق التجريبي للعلاقات بين درجات الاختبار وبيانات أخرى خارجية".

(Anastasi,1982,p.155)

طرائق التحقق من الصدق البنيوي:

ليس هناك طريقة واحدة للتحقق من الصدق البنيوي، ولابد من جمع الأدلة من مصادر مختلفة. ويمكن استعمال الطرائق الخاصة بصدق المحتوى والصدق المحكّي كدليل جزئي لدعم الصدق البنيوي ولكن هذه الطرائق لا تكفي بحد ذاتها. فالصدق البنيوي يعتمد على الاستنتاجات المنطقية المشتقة من أنواع مختلفة من البيانات ومن الطرائق المتبعة في التحقق من الصدق البنيوي:

1. الطرائق التي تعتمد على التحليل المنطقي:

وترتكز هذه الطرائق على الفحص الدقيق لبنود الاختبار والأداء الذي تتطلبه ومقابلة نتائج هذا الفحص بالنظرية أو الفرضية المعتمدة في بناء الاختبار والتفسيرات التي قد تنبع عنها.

ويرى «كرونباخ أن هذا التحليل المنطقي يعدّ من أهم مصادر التوصل إلى فروض بديلة فيما يتعلق بالأداء في الاختبارات والمقاييس، فالمحكّم الذي لديه خبرة سابقة كبيرة بالأخطاء التي شابت الاختبارات السابقة يمكنه أن يكتشف جوانب الضعف في أداة القياس الجديدة. ولكن مع ذلك فإننا لا نستطيع باستخدام هذا التحليل المنطقي للمحتوى أن نفحص صدق أداة القياس، وإنما نستطيع أن نقدم فروضاً بديلة يمكن التحقق من صحتها بالأساليب الإمبريقية أو التجريبية.

من جهة أخرى، فإن من المفيد تحليل العمليات العقلية التي يستخدمها الأفراد في أدائهم أو في التوصل إلى إجاباتهم عن بنود الاختبار، لأن هذا التحليل يضيف أدلة جديدة تزيد من فهمنا للتكوينات الفرضية التي تقيسها هذه الاختبارات (علام ، 2000، ص227). ومن هذه الزاوية فإن تحليل بنود اختبار في العلوم مثلاً قد يشير إلى أن درجاته تتأثر بالمعرفة والفهم والقدرة الكمية. وتطبيق اختبار الاستدلال الحسابي على التلاميذ وجعلهم يفكرون "بصوت مرتفع" قد يشير إلى أن البنود تتطلب عملية الاستدلال المقصودة وتعدّ بالتالي مؤشراً للصدق، أو تتطلب فقط المحاولة والخطأ وتفتقر بالتالي إلى الصدق (Gronlund,1971 ,p.90).

2. طريقة الفرق المتقابلة:

إذا كان لدينا اختبار يفترض أنه يقيس الفروق الفردية في "الميل إلى المقامرة" فيمكن أن نؤلف مجموعة من المفحوصين المعروف عنهم أنهم يقامرون في كل مناسبة تقريباً ومجموعة أخرى من غير المقامرين. فإذا كان الاختبار حساساً للفروق فإن الأداء المتوسط لكل من هاتين المجموعتين على الاختبار سيختلف اختلافاً ذا دلالة. وهكذا إذا اتسقت النتائج مع التوقعات يكون لدينا الدليل بأن الاختبار يقيس ما افترض أن يقيسه وهو الميل إلى المقامرة، وإذا لم تتّسق النتائج مع التوقعات فيمكن أن نستنتج أن الاختبار غير صادق أو أن مفهومنا حول سمة " الرغبة في المقامرة" هو مفهوم مغلوط. ولا يكون لدينا بذلك أي أساس للادعاء بأن الاختبار حساس للفروق الفردية في السمة مدار البحث (Marthuza,1977,p.152). وبهذه الطريقة يمكن أن نتوقع أن درجات الاختبار التحصيلي مثلاً ستميّز بين جماعة متدربة وجماعة أخرى غير متدربة، كما يمكن أن نتوقع أن الطلاب الذين يحصلون على درجات عالية على مقياس القابلية الاجتماعية لديهم من الأصدقاء أكثر من الطلاب الذين يحصلون على درجات منخفضة على هذا المقياس.

3. الترابط مع اختبارات أخرى:

حيث يتوقع أن تترابط درجات الاختبار التجريبي ترابطاً موجباً مع درجات اختبار آخر يفترض أنه يقيس السمة نفسها، وأن تترابط بدرجة أقل مع درجات اختبارات تقيس سمات أخرى. فاختبار الاستعداد المدرسي مثلاً يتوقع أن يترابط ترابطاً عالياً مع غيره من اختبارات الاستعداد المدرسي، وبدرجة مقبولة مع اختبارات التحصيل، وبدرجة أقل مع اختبار الاستعداد الموسيقي.

لقد أشار كامبل (Campbell, 1960) في معرض حديثه عن صدق التكوين الفرضي إلى أن دراسة هذا النوع من الصدق يجب ألا تقتصر على مجرد

إظهار أن الاختبار يترابط ترابطاً عالياً مع متغيرات أخرى يُتوقّع (أو يجب) أن يترابط معها. ولابد أن تعمل هذه الدراسة، بالإضافة إلى ما سبق، على إظهار أن الاختبار لا يترابط ترابطاً دالاً مع متغيرات يتوقع، أو يجب، أن يختلف عنها. وقد وصفت العملية الأولى من هاتين العمليتين بـ "الصدق التقاربي للاختبار" Convergent Validation في حين أن العملية الثانية أطلق عليها "الصدق التمييزي للاختبار" Discriminant Validation. وعلى سبيل المثال إذا أعطى اختبار الاستدلال الكمي ترابطاً عالياً مع الدرجات التحصيلية في الرياضيات لدى مجموعة من الأفراد كان هذا دليلاً على صدقه التقاربي. وأما الصدق التمييزي لهذا الاختبار نفسه فلا بد من إقامة الدليل عليه من خلال إظهار ترابط منخفض وغير دال لهذا الاختبار مع درجات اختبار آخر مختلف كاختبار الفهم القرائي مثلاً، وذلك لأن متغير القدرة القرائية لا علاقة له بما يقيسه اختبار الاستدلال الكمي.

لقد اقترح كامبل Campbell وفيسك Fiske (1959) اسلوباً لدراسة الصدق التقاربي والتمييزي للاختبار اعتمد على ما يسمى "مصفوفة السمات المتعدّدة والطرق المتعدّدة Multitrait Multimethod Matrix. وقد كان الهدف من وراء إعداد هذه المصفوفة هو الكشف عن درجة ترابط الاختبار باختبارات مشابهة للتحقق من الصدق التقاربي للاختبار، وكذلك ترابطه باختبارات مختلفة للتحقق من الصدق التمييزي أو التمايزي للاختبار.

إن مصفوفة السمات المتعدّدة والطرق المتعدّدة هي أشبه ما تكون بالتصميم التجريبي المنظم الذي يوفّر أدلة عن كل من هذين النوعين من الصدق ويتطلب إعداد هذه المصفوفة قياس اثنتين أو أكثر من السمات بطريقتين أو أكثر ولإلقاء الضوء على هذه المصفوفة وطريقة إعدادها نعرض المثال الذي يسوقه كامبل وفيسك، والذي يظهر جميع الترابطات المحتملة بين الدرجات المتحصّلة حين تقاس ثلاث سمات، كل منها على حده، بثلاث طرق. وهذه السمات الثلاث المفترضة في المثال هي:

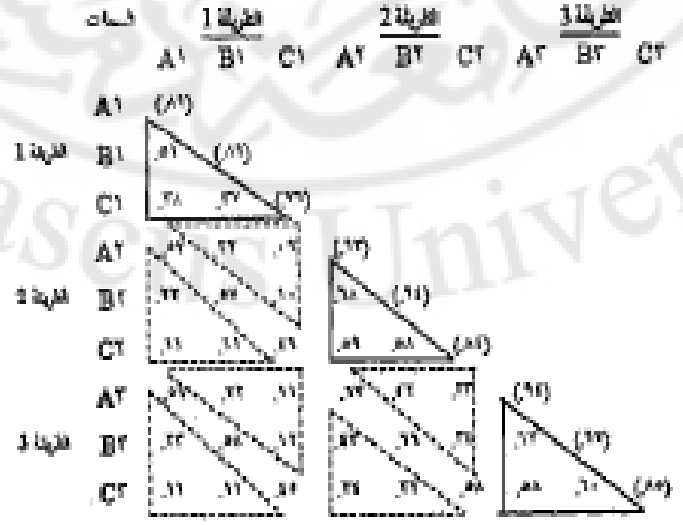
1. السيطرة.
2. الاجتماعية (أو حب الاختلاط بالآخرين) Sociability.
3. الدافعية للتحصيل.

وأما الطرق الثلاث التي يفترضها المثال فهي:

1. قائمة تعتمد على التقرير الذاتي.
2. أداة إسقاطية.
3. تقديرات الأقران.

وطبقاً لهذا المثال، والذي يعرضه الجدول رقم (7)، تشير الأحرف A، B، C إلى السمات الثلاث المفترضة في حين تشير الأرقام السفلية إلى الطرق الثلاث. وعلى هذا يشير A1 إلى درجات السيطرة في قائمة التقرير الذاتي، وA2 إلى درجات السيطرة في الأداة الإسقاطية، وA3 إلى درجات السيطرة كما تحدّدها تقديرات الأقران. كما يشير B1 إلى درجات الاجتماعية في قائمة التقرير الذاتي، وB2 إلى درجات الاجتماعية في الأداة الإسقاطية وهكذا....

الجدول رقم (8): مصفوفة السمات المتعدّدة والطرق المتعدّدة لمجموعة من البيانات الافتراضية (نقلاً عن: Anastasi &Urbina ,1997,p.131)



تتضمّن المجموعة الأولى من الترابطات المفترضة التي يعرضها الجدول السابق وهي الأرقام الواقعة بمحاذاة القطر الرئيسي للمصفوفة (بين الأقواس) – تتضمّن قيم معاملات ارتباط كل سمة بنفسها بعد أن أخضعت للقياس بإحدى الطرق (أو الأدوات) الثلاث (أي ارتباط A1 مع A1 وB1 مع B1 وC1 مع C1 وهكذا). وهذا يعني بعبارة أخرى أن هذه المجموعة من الترابطات والمؤلفة من تسعة ترابطات تشير إلى ارتباط كل اختبار بنفسه فيما يخصّ كل سمة، أي انها تشير إلى معامل ثبات كل من الاختبارات التسعة (ومن الطبيعي أن تكون قيم هذه الترابطات مرتفعة لكونها تعبر عن ارتباط كل اختبار بنفسه، وتعكس مدى اتساقه مع نفسه).

أما المجموعة الثانية من الترابطات التي يعرضها الجدول السابق فتمثّل قيم معاملات الصدق وهي الأرقام التسعة الواقعة بمحاذاة الخطوط القطرية القصيرة والملونة باللون الأسود (أي الأرقام: 0.57، 0.57، 0.46، 0.67، 0.66، 0.58، 0.56، 0.58، 0.45). وتشير معاملات الصدق هذه إلى ترابطات الدرجات المتحصّلة في السمة ذاتها بطرائق (أو أدوات) مختلفة، وهي في مثالنا هذا تشير إلى ترابط قائمة التقرير الذاتي مع الأداة الإسقاطية، وترابط قائمة التقرير الذاتي مع تقديرات الأقران، وترابط الأداة الإسقاطية مع تقديرات الأقران في كل سمة من السمات الثلاث (أي 3 ترابطات لكل سمة و9 ترابطات للسمات الثلاث). ومن الواضح أن معاملات الصدق المستخرجة هنا توفّر الدليل عن الصدق التقاربي للاختبار، وقد كانت قيم هذه المعاملات مرتفعة نسبياً ولكن دون أن تصل إلى قيم المجموعة السابقة من الترابطات التي تشير إلى معاملات الثبات وتعبر عن ارتباط كل اختبار بنفسه كما سبقت الإشارة.

وبالإضافة إلى الترابطات السابقة يتضمّن الجدول السابق الترابطات بين السمات المختلفة وقد قيست بالطريقة نفسها (وهي القيم المدوّنة داخل المثلثات التي تحدّها خطوط متصلة)، كما يتضمّن الترابطات بين السمات المختلفة وقد قيست بطرق مختلفة (وهي القيم المدونة داخل المثلثات التي تحدها خطوط

تتضمّن المجموعة الأولى من الترابطات المفترضة التي يعرضها الجدول السابق وهي الأرقام الواقعة بمحاذاة القطر الرئيسي للمصفوفة (بين الأقواس) – تتضمّن قيم معاملات ارتباط كل سمة بنفسها بعد أن أخضعت للقياس بإحدى الطرق (او الأدوات) الثلاث (أي ارتباط A1 مع A1 وB1 مع B1 وC1 مع C1 وهكذا). وهذا يعني بعبارة أخرى أن هذه المجموعة من الترابطات والمؤلفة من تسعة ترابطات تشير إلى ارتباط كل اختبار بنفسه فيما يخصّ كل سمة، أي انها تشير إلى معامل ثبات كل من الاختبارات التسعة (ومن الطبيعي أن تكون قيم هذه الترابطات مرتفعة لكونها تعبر عن ارتباط كل اختبار بنفسه، وتعكس مدى اتساقه مع نفسه).

أما المجموعة الثانية من الترابطات التي يعرضها الجدول السابق فتمثّل قيم معاملات الصدق وهي الأرقام التسعة الواقعة بمحاذاة الخطوط القطرية القصيرة والملونة باللون الأسود (أي الأرقام: 0.57، 0.57، 0.46، 0.67، 0.66، 0.58، 0.56، 0.58، 0.45). وتشير معاملات الصدق هذه إلى ترابطات الدرجات المتحصّلة في السمة ذاتها بطرائق (او ادوات) مختلفة، وهي في مثالنا هذا تشير إلى ترابط قائمة التقرير الذاتي مع الأداة الإسقاطية، وترابط قائمة التقرير الذاتي مع تقديرات الأقران، وترابط الأداة الإسقاطية مع تقديرات الأقران في كل سمة من السمات الثلاث (أي 3 ترابطات لكل سمة و9 ترابطات للسمات الثلاث). ومن الواضح أن معاملات الصدق المستخرجة هنا توفّر الدليل عن الصدق التقاربي للاختبار، وقد كانت قيم هذه المعاملات مرتفعة نسبياً ولكن دون أن تصل إلى قيم المجموعة السابقة من الترابطات التي تشير إلى معاملات الثبات وتعبر عن ارتباط كل اختبار بنفسه كما سبقت الإشارة.

وبالإضافة إلى الترابطات السابقة يتضمّن الجدول السابق الترابطات بين السمات المختلفة وقد قيست بالطريقة نفسها (وهي القيم المدوّنة داخل المثلثات التي تحدّها خطوط متصلة)، كما يتضمّن الترابطات بين السمات المختلفة وقد قيست بطرق مختلفة (وهي القيم المدونة داخل المثلثات التي تحدها خطوط

متقطعة). ولتوفير صفة الصدق البنيوي (أو صدق التكوين الفرضي) لابد أن تكون معاملات الصدق المحسوبة أعلى بوضوح من الترابطات بين السمات المختلفة المقيسة بطرق مختلفة، كما يتوقع أن تكون أعلى (ولكن بدرجة اقل) من الترابطات بين السمات المختلفة المقيسة بالطريقة نفسها. فمثلاً الترابط بين درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات السيطرة المتحصّلة من الأداة الإسقاطية (وهو ما يشير إلى الصدق التقاربي) يجب أن يكون أعلى من ترابط درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات الاجتماعية المتحصّلة من الأداة الإسقاطية. في الوقت نفسه يتوقع أن يكون هذا الترابط (أي الترابط بين درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات السيطرة المتحصّلة من الأداة الإسقاطية) أعلى (ولو قليلاً) من ترابط درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات الاجتماعية المتحصّلة أيضاً من قائمة التقرير الذاتي.

ويتبين مما سبق أن دراسة الصدق التقاربي والتمييزي وفقاً للمنهج الذي اختطه كامبل وفيسك تعتمد على دراسة الارتباط بين الطرق (أو الأدوات) المختلفة التي تقيس السمة نفسها (وهذا ما يعبر عنه بالصدق التقاربي) ثم مقارنة هذا الارتباط بالارتباط بين السمات المختلفة التي تقاس بالطريقة نفسها، وكذلك الارتباط بين السمات المختلفة التي تقاس بطرق مختلفة (وهذا ما يعبر عنه بالصدق التمييزي). وهذا يعني أن دراسة الصدق التقاربي لا تنفصل عن دراسة الصدق التمييزي، بل تصبّان في مجرى واحد، وتمثّلان جانبين لعملية واحدة متكاملة هدفها الأول والأخير هو التثبّت من صدق الاختبار باعتماد الأسلوب "التقاربي" من جهة والأسلوب التمييزي أو التباعدي من جهة أخرى.

4. التحليل العاملي:

ويهدف إلى تحديد العوامل أو السمات المشتركة بين مجموعة من الاختبارات وذلك بحساب معاملات الارتباط بين كل اختبار والاختبارات الأخرى فإذا ظهرت تجمعات معينة بينها فهذا يدل على وجود سمات أو عوامل مشتركة بينها، وهذا يعني أن السمة أو العامل المشترك نستدل عليه من وجود ترابط عالٍ بين عدد من الاختبارات. فمثلاً إذا كانت الاختبارات التي تترابط مع بعضها ترابطاً عالياً هي من نوع اختبارات المفردات وتكملة الجمل والمتشابهات والأضداد، وكانت معاملات ارتباط هذه الاختبارات ببقية الاختبارات منخفضة فيمكن أن نستنتج وجود عامل الفهم اللغوي. ومن الواضح أن التحليل العاملي يؤدي دوراً بالغ الأهمية في تقدير الصدق البنيوي من خلال الكشف عن السمة التي يقيسها الاختبار بأسلوب كمي إحصائي.

يتيح التحليل العاملي إنقاص عدد المتغيرات التي يمكن أن يوصف من خلالها أداء الفرد. فبدلاً من أن يوصف أداء الفرد من خلال 20 اختباراً أخضع لها هذا الفرد مثلاً، يمكن أن يوصف هذا الأداء من خلال خمسة أو ستة عوامل أمكن استخراجها من خلال التحليل العاملي لهذه الاختبارات وتبيّن أنها المسؤولة عن الترابطات الداخلية بينها. وعلى هذا يمكن التعبير عن أداء الفرد من خلال درجاته في خمسة أو ستة عوامل بدلاً من درجاته في الاختبارات العشرين. وكما سبقت الإشارة فإن الهدف الرئيس للتحليل العاملي هو تبسيط عملية وصف السلوك عن طريق خفض عدد المتغيرات التي تقيسها الاختبارات إلى عدد ضئيل من العوامل (أو السمات) المشتركة.

هذا ويمكن استخدام العوامل التي يتمّ استخراجها بالتحليل العاملي في وصف البنية العاملية للاختبار. وعلى هذا الأساس يمكن وصف كل اختبار بالاعتماد على العوامل الرئيسة المحددّة لدرجاته، إضافة إلى تشبّعه بكل من هذه العوامل وارتباطه بها وهذا ما يطلق عليه الصدق العاملي Factorial validity

للاختبار. وهكذا فإذا ترابط عامل الفهم اللفظي بمقدار 0.66 مع اختبار المفردات فإن الصدق العاملي لهذا الاختبار بوصفه مقياساً لسمة الفهم اللفظي هو 0.66.

وتجدر الإشارة هنا إلى أن الصدق العاملي يشير إلى ترابط الاختبار مع كل ما هو مشترك بين مجموعة من الاختبارات، أو غيرها من الأدوات "اللااختبارية". وهذا يعني أن التقديرات الرتبية والمقاييس المحكّية الأخرى يمكن استخدامها، إضافة لاختبارات أخرى، في استطلاع البنية العاملية لاختبار معين وتحديد السمات المشتركة التي يقيسها.

(Anastasi&Urbina ,1997,p.128)

تكامل الطرائق المتبعة في دراسة الصدق:

على الرغم من أن الأشكال أو الطرائق المتبعة في دراسة الصدق تصنّف عادة إلى ثلاثة أشكال رئيسة "متمايزة"، وهي: صدق المحتوى، والصدق المحكّي، وصدق التكوين الفرضي، وأن لكل منها أساليبه وإجراءاته الخاصة، فإن هذه الطرائق أو الأشكال لا تمثل في حقيقة الأمر أنواعاً متباينة من الصدق. وقد يكون من الخطأ النظر إليها بصورة منفصلة، أو الاعتقاد بأن أياً منها يغني عن الآخر أو يحل محله، والأصح أن ننظر إليها على أنها جوانب أو مظاهر متعدّدة لعملية واحدة متكاملة وهي عملية التحقق من الصدق بالمعنى الشامل. هذا مع الأخذ بالحسبان أنه ليس من الضروري دائما استخدام هذه الطرائق جميعها عند دراسة الصدق لاختبار محدد، وأن لكل من هذه الطرائق والأشكال استعمالاته الخاصة به، وذلك انطلاقاً من أن صدق أدوات القياس ليس مفهوماً عاماً، وإنما هو مفهوم نسبي يرتبط بمواقف، أو استخدامات معينة لمجموعات معينة من الأفراد.

إن اختيار شكل أو أكثر من أشكال الصدق عند دراسة الصدق لاختبار محدد يتوقف على الغرض أو الاستعمال الخاص الذي صمّم من أجله هذا

الاختبار. ومن هذا المنظور لابد من اختيار شكل او اكثر من أشكال الصدق يكون او تكون اكثر ملاءمة للغرض الذي صمّم الاختبار من اجله. في الوقت نفسه، لا يصح بأي حال من الأحوال، الاقتصار على شكل واحد من أشكال الصدق حين يكون هناك أكثر من غرض واحد للاختبار. فالاختبار الذي صمّم لتلبية أغراض عديدة ومختلفة لابد من دراسة صدقه بطرائق عديدة ومختلفة. وحين يستخدم الاختبار التحصيلي في التنبؤ بالأداء اللاحق للفرد في المرحلة الدراسية الأعلى مثلاً يصبح من المحتم دراسة صدقه باعتماد محك معين للأداء اللاحق، وعدم الاقتصار على دراسة صدق محتوى هذا الاختبار. وتظهر في الجدول التالي الطرائق (أو الأشكال) التي يمكن استخدامها عند دراسة الصدق لاختبار في الحساب أعطي كمثال، وارتباط كل منها بالغرض أو الاستعمال الخاص به (الجدول رقم (9)).

الجدول رقم (9) دراسة الصدق لاختبار في الحساب (صمّم لتلبية أغراض متعدّدة)

الغرض من القياس	السؤال التوضيحي	دليل الصدق
قياس التحصيل في الحساب "الابتدائي"	كم تعلّم التلميذ "س" من مادة الحساب؟	وصف المحتوى
اختبار للاستعداد (يهدف التنبؤ بالأداء اللاحق في الرياضيات)	إلى أي درجة يمكن أن ينجح التلميذ في المستقبل في دراسة الرياضيات؟	محك للتنبؤ
أداة لتشخيص صعوبات التعلّم	هل يشير أداء التلميذ إلى صعوبات محددة؟	محك للتنبؤ (يستخدم في الظرف الراهن)
قياس الاستدلال الكمي	كيف ترتبط درجة التلميذ بمؤشرات أخرى للقدرة الاستدلالية؟	تحديد السمة المفترضة

(نقلاً عن: Anastasi&Urbina ,1997,p.136)

لقد سبقت الإشارة إلى أن الصدق البنيوي أو الافتراضي هو مفهوم عام شامل يتضمّن أنواع الصدق كافة، وأن الأساليب والإجراءات والتقنيات الخاصة بتأسيس صدق المحتوى وكذلك الصدق المحكّي يمكن أن تندرج ضمن أساليب الصدق البنيوي وتقنياته وإجراءاته. والواقع أن البيانات والأدلة التي تجمع للتحقق من الصدق البنيوي للاختبار تتضمّن عادة من جملة ما تتضمّن أدلة عن صدق محتواه وصدقه المحكّي، مما يشير إلى أن الصدق البنيوي يعدّ الإطار الشامل الذي تلتقي ضمنه كما تتكامل من خلاله أشكال الصدق كافة. ولعل في هذه النظرة إلى الصدق البنيوي إشارة واضحة إلى تكامل الطرائق المتبعة في دراسة الصدق، وتأكيداً لدورها مجتمعة في هذه الدراسة، دون أن ينفي هذا حقيقة ارتباطها بأغراض محددة، وإمكانية الاقتصار على إحداها أو بعضها في الحالات التي يمكن أن تلبي الغرض المرسوم للاختبار.

العوامل المؤثرة في الصدق:

ثمة عوامل عديدة يمكن أن تؤثر في صدق الاختبار وتضعفه. وقد صنف ن. جرونلند هذه العوامل ضمن أربع فئات رئيسة وهي التالية:

أ. عوامل في الاختبار نفسه:

ومن أهمها:

أ. التعليمات غير الواضحة: فالتعليمات التي لا تشير بوضوح إلى ما يجب أن يفعله التلميذ، وكيف يجيب، وما إذا كان يسمح له بالتخمين، وكيف يسجل إجاباته، تضعف الصدق.

ب. المفردات والتراكيب الصعبة: فهي قد تجعل الاختبار يقيس الفهم القرائي أو مظاهر من الذكاء وليس مظاهر السلوك التي أعدّ لقياسها.

ج. البنود الصعبة جداً أو السهلة جداً: إذ أن البنود الصعبة جداً وكذلك السهلة جداً لا تميّز بين الأقوياء والضعاف تحصيلاً كما سنرى، وبالتالي فهي تضعف الصدق.

د. البنود الموحية بالإجابة: فمثل هذه البنود قد تقيس القدرة على اكتشاف الإجابات الصحيحة اعتماداً على الإيحاءات أو التلميحات التي تنطوي عليها البنود ذاتها ولا تكشف بالتالي الفروق في السمة أو الخاصية موضع القياس.

ه. عدد البنود: فالاختبار هو عينة من بنود أو أسئلة كثيرة يمكن أن تطرح وإذا كان قصيراً وتضمّن عدداً ضئيلاً من البنود يضعف تمثيله وبالتالي يضعف صدقه.

و. الغموض: فالغموض يؤدي إلى سوء التفسير وقد يضلّل المفحوصين الأقوياء أكثر من الضعاف ويجعل الاختبار يميّز بالاتجاه المعاكس.

ز. ترتيب البنود بصورة غير ملائمة: يرتّب الاختبار عادة وفق مبدأ التدرج في الصعوبة فإذا وضعت البنود الصعبة في البداية يقضي الطلاب الكثير من الوقت للإجابة عنها، وقد لا يتسع لهم الوقت للإجابة عن البنود السهلة. ثم إن هذا الترتيب قد يضعف دافعيتهم مما ينعكس سلباً على أدائهم.

ح. تخصيص مكان واحد للإجابة الصحيحة: فإذا أخذ الجواب الصحيح الرقم (3) أو الرمز (ج) مثلاً يسهل اكتشافه مما يؤدي إلى إضعاف الصدق بطبيعة الحال.

2. العوامل المتصلة بشروط الإجراء والتصحيح ومنها:

أ. الشروط البيئية المحيطة: فالحرارة والبرودة الزائدة والضوضاء وغيرها، هي من العوامل التي يمكن أن تؤثر سلباً في الأداء الاختباري.

ب. الوقت المخصص للإجابة: فإذا لم يكن كافياً أو كان طويلاً أكثر من اللزوم يتأثر الأداء الاختباري بالضغط على المفحوصين في الحالة الأولى وإرباكهم، وإتاحة الفرصة للغش والفوضى في الحالة الثانية.

ج. أخطاء التصحيح: وقد تنتج هذه الأخطاء عن عملية التقدير ذاتها كما في الأسئلة المقالية التي تعاني من عيوب التقدير الذاتي وضعف مستوى الثبات، وقد تنتج عن عملية جمع العلامات أو نقلها وما شابه ذلك.

د. الغش: فإتاحة الفرصة للغش أو تقديم المساعدة للمفحوصين يحولان دون الكشف عن الفروق الحقيقية بينهم، ويضعفان الصدق.

3. العوامل المتصلة باستجابات المفحوصين:

قد يتعرض المفحوصون لأزمات انفعالية حادة تؤثر في أدائهم الاختباري. ومن المعلوم أن بعض المفحوصين يعانون بشدة مما يسمى بـ " قلق الاختبار"، كما أن بعضهم قد تضعف دافعيتهم ولا يبذلون جهودهم على النحو المطلوب، وهذا ما يظهر بخاصة في الاختبارات التي تجرى لأغراض البحث التي قد لا يعيرها المفحوصون أو بعضهم الاهتمام الكافي نظراً لأنها قد لا تعني شيئاً بالنسبة لهم أو لا تقدّم أية فائدة من وجهة نظرهم.

بيد أن العامل الأقل وضوحاً بين العوامل المتصلة باستجابات المفحوصين والذي ينطوي مع ذلك على قدر كبير من الأهمية هو ما يسمى نمط الاستجابة (Response Set) والذي يظهر في ميل ثابت نسبياً لاتباع نموذج معين في الإجابة على البنود. فمثلاً بعضهم قد يعطي إشارة "صح" باستمرار على أسئلة صح – خطأ، وبعضهم الآخر قد يعطي إشارة خطأ باستمرار بغض النظر عن صحة الجواب أو خطئه. فإذا احتوى الاختبار على عدد كبير من البنود الصحيحة يكون في مصلحة الفئة الأولى وضد الثانية والعكس بالعكس. ومع أن بعض أنماط الاستجابات يمكن مواجهتها والتغلب عليها كأن نجعل عدد الإجابات الصحيحة مساوياً عدد الإجابات الخاطئة فإن أنماطاً أخرى يصعب ضبطها والسيطرة عليها كالميل للعمل بسرعة على حساب الدقة، والميل إلى المجازفة عند الشك، واتباع أسلوب معين في الإجابة عن أسئلة المقال. فهذه الأنماط تضعف الصدق لأنها تدخل عوامل لا علاقة لها أساساً بغرض القياس (Cronbach,1970).

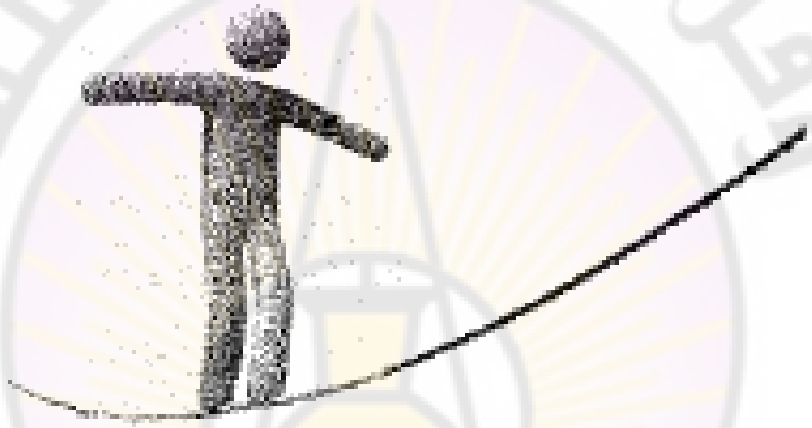
4. طبيعة المجموعة والمحكّ:

فالصدق خاص بمجموعة معينة من المفحوصين. ودرجات اختبار العلوم، على سبيل المثال، قد تتأثر بالفهم القرائي لدى مجموعة من المفحوصين ولا تتأثر بذلك لدى مجموعة أخرى لا تعاني من صعوبات في الفهم القرائي. والواقع أن ما يقيسه الاختبار يتأثر بعوامل عديدة كالعمر، والجنس، ومستوى القدرة، والخلفية التربوية، والخلفية الثقافية، وغيرها . ومن الضروري في تقويم معاملات الصدق مراعاة طبيعة المحكّ المستعمل. فنحن يمكن أن نتوقع أن تترابط درجات اختبار التفكير النقدي ترابطاً عالياً مع درجات اختبار الدراسات الاجتماعية التي تلح على التفكير النقدي أكثر من تلك التي تعتمد على تذكر الحقائق. وكلما ازداد التشابه بين السلوكات التي يقيسها الاختبار والسلوكات التي يمثّلها المحكّ ارتفع معامل الصدق.

وعموماً فإن أي عامل يؤثر في أداء المفحوص ويمنعه عن إظهاره على حقيقته يضعف صدق الاختبار ويحرف الاختبار عن غرضه الأصلي في الكشف عن مستوى الأداء الحقيقي للمفحوص (Gronlund,1971,p.p93 – 96).

الفصل السادس

الثبات



الثبات

معنى الثبات:

يمثّل الصدق المطلب الأساسي لصلاحية المقياس. أما الثبات فهو شرط ضروري أو لازم للصدق يستحيل دونه تأسيس الصدق وتأكيده. فلكي نقول: إن المقياس يقيس ما وضع لقياسه أو إنه ينطوي على درجة عالية من الصدق يتعين علينا التأكد مما إذا كان يقيس بدقة (ذلك الشيء) الذي وضع لقياسه. والسؤال المركزي المطروح بصدد الثبات هو: ما مستوى الدقة الذي توفّره أداة القياس؟ أو كم يتحقق فيها من الدقة والاتساق عند تكرار القياس على مجموعة من الأفراد؟ وبعبارة أخرى، هل يكشف المقياس بصورة ثابتة ومطرّدة الفروق بين الأفراد، وكم تتذبذب نتائجه؟ ومن الواضح أنه كلما تذبذبت النتائج انخفض ثبات الاختبار وضعفت حساسيته للكشف بدقة عن الفروق الحقيقية في السمة أو الخاصية موضع القياس.

يستخدم مفهوم الثبات بالمعنى العام للدلالة على مدى اعتماد الفروق في درجات الاختبار على الفروق الحقيقية في السمة أو الخاصية المقيسة من جهة وأخطاء القياس العشوائية أو الناجمة عن الصدفة من جهة أخرى. ومن المعلوم أن الاختبار في التربية وعلم النفس يستحيل عدّه مقياساً ثابتاً ودقيقاً كالمقاييس الطبيعية (المتر، الكغ، ... إلخ). ففي القياس الفيزيائي يمكن الحصول على قياسات ثابتة (نسبياً) لمجموعة من الأسباب بينها أن الصفات المادية مستقرة نسبياً ويمكن إخضاعها للقياس المباشر، وبينها ما يتصل بأدوات القياس ومستوى الدقة الذي وصلت إليه. ويعدّ القياس النفسي والتربوي أقل ثباتاً من القياس الفيزيائي لأسباب عديدة بينها ما يتصل بعدم استقرار السمات أو الصفات المقيسة بمرور الوقت، وبينها ما يتصل بخطأ العينة أو المعاينة، بالإضافة إلى أخطاء

التطبيق والتصحيح وغيرها. ويمكن إجمال مصادر الخطأ في القياس النفسي والتربوي على النحو التالي:

1. أخطاء في القياس نفسه: فالقياس هو عينة من البنود وقد لا تكون هذه العينة ممثلة لجميع البنود وهذا ما يسمى بخطأ العينة أو المعاينة Sampling Error.
2. الأخطاء الناجمة عن شروط التطبيق: فإذا لم تراع شروط التطبيق كما تنص عليها تعليمات الاختبار أو روعيت بصورة غير كافية فقد تؤدي إلى نتائج غير صحيحة.
3. الأخطاء الناجمة عن التصحيح: وهذه الأخطاء تظهر بخاصة في الأسئلة المقالية التي تفسح مجالاً واسعاً لظهور العوامل الذاتية للمصحح واحتمال نشوء الاختلاف بين مصحح وآخر في تقدير الدرجات، أو حتى عند المصحح نفسه من وقت لآخر.
4. الأخطاء الناجمة عن المفحوص نفسه: وهذه الأخطاء كثيرة ومتنوعة منها ما يتصل بقلق الامتحان ومنها ما يتصل بالتعليمات والميل إلى التخمين والدافعية للأداء ودرجة التأثر بالعوامل الطبيعية كالحرارة والبرودة والضجيج... وغيرها. هذا بالإضافة إلى عدم ثبات واستقرار السمة المقيسة عند المفحوص التي يمكن أن تتأثر بالعديد من العوامل كالتدريب والتذكر والنسيان وغيرها.

إن واحداً أو أكثر من مصادر الخطأ السابقة يمكن أن يؤثر في جميع الطلاب بالدرجة نفسها ويطلق عليه في هذه الحالة اسم الخطأ المنتظم Systematic Error وقد يختلف تأثيره من طالب لآخر سواء من حيث المقدار أو الاتجاه (سلبي أو إيجابي) ويطلق عليه في هذه الحالة اسم الخطأ العشوائي Random Error. ومن الواضح أن بعض العوامل المتصلة بالاختبار نفسه وبشروط التطبيق والتصحيح يمكن ضبطها إلى حد كبير وتقليل الأخطاء الناجمة عنها، وأن بعض العوامل ولاسيما تلك المتصلة بالمفحوص نفسه يصعب

ضبطها والتحكم بها ويتعذّر بالتالي تحاشي احتمالات الخطأ التي يمكن أن تنتج عنها. ومهما يكن من أمر فإن التباين في الدرجات التي يعطيها المقياس يمثّل التباين الحقيقي في أداء المفحوصين كما يمثّل التباين الناتج عن الخطأ أو تباين الخطأ Error Variance. ويمكن توضيح الثبات من خلال مفهوم التباين على النحو التالي:

$$ع^2_{س} = ع^2_{سو} + ع^2_{ع} + ع^2_{عع} \quad \text{(المعادلة 5 – 2)}$$

حيث:

$ع^2_{س}$ = التباين الكلي.

$ع^2_{سو}$ = التباين الحقيقي في علامات الطلاب الناتج عن فروق حقيقية بينهم في السمة المقيمة.

$ع^2_{ع}$ = التباين الحقيقي في علامات الطلاب الناتج عن تباينهم في درجة التأثر بالعوامل المسببة للخطأ المنتظم.

$ع^2_{عع}$ = التباين في الخطأ العشوائي.

فإذا افترضنا أن التباين الناتج عن الخطأ العشوائي يساوي صفراً وهذا مجرد افتراض يستحيل تحقيقه في الواقع في القياس النفسي والتربوي، فإن التباين الكلي يساوي التباين الحقيقي، وهذا يعني أن الاختبار ثابت تماماً أو أنه يقيس بدقة تامة ودون أخطاء. ويمكن القول بلغة الإحصاء: إن الثبات هو نسبة التباين الحقيقي إلى التباين الكلي، مع ملاحظة أن التباين الذي ينطوي عليه مفهوم الثبات يشمل التباين الناتج عن فروق حقيقية في أداء المفحوصين والتباين الناتج عن الخطأ المنتظم. ومن الواضح أن الثبات يزداد كلما نقص تباين الخطأ (أي التباين الناتج عن الخطأ العشوائي) وينخفض كلما ازداد هذا الأخير. ومن

المفيد الإشارة هنا إلى أنه إذا انعدم التباين الناتج عن الخطأ العشوائي والعدم التباين الناتج عن الخطأ المنتظم(وهذا مجرد افتراض) فإن التباين الحقيقي ينسب كلياً إلى السمة المقيسة (أي التباين الحقيقي يعادل التباين الظاهري) وفي هذه الحالة يقيس الاختبار ما وضع لقياسه أي يصدق تماماً وهذا هو المقصود بصدق الاختبار.

وبالإضافة إلى مفهوم التباين ودوره في توضيح معنى الثبات من المفيد ملاحظة النقاط التالية لإلقاء المزيد من الضوء على مفهوم الثبات:

1. يرتبط الثبات بنتائج أداة القياس والتقويم لا بالأداة ذاتها. وبدلاً من الحديث عن ثبات الأداة بحد ذاتها من الأصوب الحديث عن ثبات النتائج. وقد يكون للأداة الواحدة أكثر من معامل ثبات واحد وذلك تبعاً للمجموعة والموقف الاختباري. وبما أن لكل عينة من الأفراد خصائصها فإنه سيكون من المفيد الحصول على أكثر من معامل ثبات للاختبار الواحد لدى أكثر من عينة. فمعامل الثبات المستخرج من أداء عينة من الأطفال قد يختلف عن نظيره المستخرج من أداء عينة من الراشدين، ومعامل الثبات المستخرج من أداء عينة من الذكور أو العاقين قد يختلف عنه لدى الإناث أو العاديين، وهكذا...
2. تقدير ثبات الاختبار يرتبط بالاستعمال الخاص له. فلبعض الاستعمالات قد نهتم بالسؤال عن درجة ثبات نتائج الاختبار عبر فترة من الزمن، ولاستعمالات أخرى قد نهتم بالسؤال عن درجة ثبات النتائج من خلال عينات مختلفة من الأسئلة أو مقدّرين مختلفين. ولاستعمالات أخرى قد نهتم بالاتساق الداخلي للأداة ذاتها (Gronlund,1971,p.101). ولا يعني هذا أن الاختبار الواحد لا تصلح له إلا طريقة واحدة من طرق الثبات. وقد يكون من الأنسب في حالات كثيرة استخدام أكثر من طريقة واحدة للاختبار الواحد.

3. الثبات هو مفهوم إحصائي، بخلاف الصدق الذي يعتمد على التحليل المنطقي وقد يتطلب العمل الإحصائي.

العلاقة بين الصدق والثبات:

تشير المعادلة الخاصة بالتباين الكلي التي وردت سابقاً (المعادلة 5 – 2) إلى انه إذا افترضنا انعدام التباين الحقيقي في السمة المقيسة وانعدام الصدق

(أي $ع^2_{ص ق}$ = صفراً) فإن التباين في العلامات لا يعود إلى فروق حقيقية بين الأفراد في السمة المقيسة بل يعود إلى فروق ناجمة عن الخطأ. وفي هذه الحالة يمكن القول: إن الاختبار "ثابت" ولكنه غير صادق. وهذا يعني إن الاختبار الذي قد يكون " ثابتاً " أو على درجة ما من الثبات قد لا يكون صادقاً أو لا يتمتع بأي درجة من الصدق. ويمكن من الناحية المنطقية إثبات أن الاختبار قد يكون "ثابتاً" أو يتمتع بدرجة من الثبات دون أن يكون صادقاً، فمثلاً اختبار المعلومات التاريخية الذي يقيس فهم المفردات والتراكيب اللغوية قد يعطي نتائج متسقة في حال تكراره ويكون على درجة عالية من الثبات ولكنه غير صادق في قياس المعلومات التاريخية. وإذا أعطى عقرب الساعة إشارة للوقت تزيد بمقدار 50 دقيقة لعدة مرات، أو أعطى ميزان البقال وزناً واحداً خاطئاً لعدة مرات فيمكن القول: إن النتيجة ثابتة مع أنها غير صادقة بالتأكيد.

يعتمد الصدق جزئياً على الثبات ويمثّل الثبات شرطاً ضرورياً وغير كاف للصدق. ومن الواضح أن المقياس الذي يعطي نتائج غير متسقة لا يمكن أن يعطي نتائج صادقة على حين أن المقياس الذي يعطي نتائج صادقة لابد أن يعطي نتائج متسقة ويتمتع بصفة الثبات. وبطبيعة الحال فإن الزيادة في معامل الثبات قد تؤدي إلى الزيادة في معامل الصدق ولكنها لا تحتّم هذه الزيادة. فقد تنتج الزيادة في معامل الثبات عن ازدياد التباين الناتج عن الخطأ المنتظم كما يظهر بصورة واضحة في المعادلة السابقة الخاصة بالتباين.

طرائق الثبات:

يميّز الباحثون عادة بين طرائق عدة في حساب الثبات أهمها طريقة الإعادة Test – Retest Method، وطريقة إجراء مقياس معادل Equivalent Forms، وطريقة التنصيف Split – half Method، وطريقة الاتساق الداخلي باستخدام معامل كودر – ريتشارسون ومعامل الفا Kuder – Richardson Reliability and Coefficient Alpha، بالإضافة إلى الطريقة التي تعرف بثبات المصححين Scorer Reliability. وتتطلب الطريقة الأولى من هذه الطرائق وهي طريقة الإعادة تطبيقين للاختبار، وتتطلب الثانية وهي طريقة الشكل المعادل تطبيقاً للاختبار ولصورة معادلة له، في حين أن الطرائق الثلاث الأخرى تتطلب تطبيقاً واحداً للاختبار. وسوف نتناول فيما يلي هذه الطرائق المختلفة بالدراسة:

1. **طريقة الإعادة:**

وتقوم هذه الطريقة على إعادة تطبيق المقياس نفسه على جماعة واحدة من المفحوصين ثم حساب معامل الثبات وهو معامل الترابط بين العلامات التي ينتهي إليها التطبيق الأول والعلامات التي ينتهي إليها التطبيق الثاني للمقياس أو الاختبار. ويسمى معامل الثبات الذي يحسب بهذه الطريقة معامل الاستقرار Coefficient of Stability. وقد وجّهت لهذه الطريقة انتقادات عديدة منها:

1. صعوبة خفض أثر التغيرات التي قد تقع في الفترة الفاصلة بين التطبيق الأول والتطبيق الثاني للمقياس. فقد يكتسب المفحوصون أو بعضهم خلال هذه الفترة خبرات جديدة. وقد يتعرض بعضهم للنسيان ولاسيما إذا طالت الفترة، وأما إذا قصرت فإن تذكر الطلاب للأجوبة الأولى قد يؤدي دوره وهذا ما يرفع معامل الثبات بصورة مصطنعة. والواقع أنه من الصعب تعيين فترة فاصلة مثلى بين التطبيق الأول والتطبيق الثاني للاختبار وإن كان

المتبّع بصورة عامة هو أن لا تقل هذه الفترة عن أسبوع ولا تزيد عن ستة أشهر. هذا مع الإشارة إلى أن المتبع في اختبارات التحصيل هو أن لا تزيد هذه الفترة على شهر واحد ولا تقل عن بضعة أيام.

2. ثمة عوامل عديدة تؤثر في أداء المفحوصين كما هو معلوم، فإذا أعطي الاختبار في مناسبتين مختلفتين فإن لهذه العوامل فرصتين للتأثير في النتائج، وبالتالي فإن مصادر خطأ عديدة تفعل فعلها وتؤدي نظرياً إلى مضاعفة أخطاء القياس ولاسيما تلك الأخطاء الناجمة عن شروط التطبيق والأخطاء الناجمة عن المفحوص نفسه كالقلق والدافعية والحالة الصحية وغيرها.

3. تتطلب هذه الطريقة جلستين اختباريتين وتعدّ غير اقتصادية من ناحية الوقت، وقد تتعب المفحوص، هذا بالإضافة إلى احتمال انخفاض مستوى الدافعية للأداء الاختباري في الجلسة الاختبارية الثانية.

إن إدراك العيوب التي تعاني منها طريقة الإعادة في حساب الثبات لابد أن يدفع الباحث أو العامل في القياس إلى التصدّي لها والعمل على التخفيف من وطأتها إلى الحدود القصوى المتاحة. وقد يكون من المفضل عند استخدام هذه الطريقة ألا يقتصر الباحث على حساب الثبات على مدى فترة زمنية واحدة، بل على مدى فترتين أو أكثر، كأن يحسب معامل الثبات لفترة عشرة أيام(أي بفاصل قدره عشرة أيام بين التطبيق الأول والتطبيق الثاني للاختبار)، ولفترة شهر واحد، ولفترة ثلاثة أشهر. ولمواجهة احتمال تأثر أفراد العينة بإعادة الاختبار لأكثر من مرة سواء بالاتجاه الإيجابي أم السلبي (كأن يتحسن أداؤهم نتيجة التدرب على الاختبار، أو يضعف نتيجة للملل أو اللامبالاة)، فقد يكون من المناسب استخدام عينتين متجانستين (أو أكثر) يطبق الاختبار ثم يعاد تطبيقه على كل منهما بفواصل زمنية مختلفة (كأن تكون الفترة الزمنية الفاصلة بين التطبيق الأول والثاني للاختبار على أفراد العينة الأولى عشرة أيام، وعلى أفراد العينة الثانية شهراً واحداً وهكذا...).

ويمكن القول، إن اختيار الأسلوب المناسب والفترة المناسبة لإعادة الاختبار ولحدود التجانس المطلوبة في العينة "يعتمد في الجانب الأكبر منه على تحليلنا السيكولوجي لما يقيسه الاختبار، وكيفية قياسه، ولمعلوماتنا التي يوفّرها لنا التراث عن الوظائف التي يقيسها، ونموها، وارتقائها، وتأثرها بالمتغيرات المختلفة" (فرج، 1980، ص353).

ولعل من بين الأمور التي لابد من أخذها بالحسبان عند دراسة الثبات على أية حال هو أن طريقة الإعادة تصلح لدراسة السمات التي تتسم بشيء من الثبات والاستقرار النسبي، ولاتصلح، أو لا تصلح بالدرجة نفسها، عند دراسة السمات التي تفتقر إلى الثبات والاستقرار، أو تتذبذب بفعل عوامل معينة. ومن نافلة القول أن هذه الطريقة ستكون محدودة الفائدة، وربما عديمة الفائدة، عند استخدامها في تقدير ثبات درجات مقياس للدافعية، أو للحالة المزاجية، أو لحالة القلق، أو للذكاء لدى صغار الأطفال، أو غيرها من السمات أو الحالات التي يكون لعامل الزمن دوره وأثره الواضح فيها.

2. طريقة الأشكال المتعادلة:

وترتكز هذه الطريقة على تحديد درجة الاتساق في أداء المفحوصين على عينة من البنود أو المهمات وعينة أخرى من البنود أو المهمات سحبت من المجتمع الأصلي ذاته (مجتمع البنود) الذي سحبت منه العينة الأولى. فإذا كان الاختبار عينة ممثلة للبنود المحتملة في مجال معين (أو المجتمع الأصلي للبنود) فهو مقياس ثابت للمحتوى في ذلك المجال. والطريقة الأيسر لتقدير ما إذا كان يقيس عينة ممثلة للمحتوى هي بوضع شكلين متعادلين (أو أكثر) للاختبار وحساب معامل الترابط بينهما، فإذا كان الترابط عالياً بينهما فهذا يدل على أن كلا الشكلين يقيس المحتوى ذاته، ولذا فهما عينتان ثابتتان لمحتوى المجال المقيس. وبطبيعة الحال فإن المقياس الثاني يجب أن يحتوي على العدد نفسه من الأسئلة وأن تتعادل الأسئلة في كلا المقياسين من حيث مستوى الصعوبة وتمثيلها

للمحتوى وأن تكون صياغتها متماثلة، كما تتماثل أو تتعادل من حيث التعليمات والزمن المعطى للاختبار.

يميّز الباحثون عادة بين طريقتين لحساب الثبات باستخدام الأشكال المتعادلة وتتفرعان عن الطريقة الأم مدار البحث. تستخدم في الطريقة الأولى من هاتين الطريقتين أشكال متعادلة "فورية" تطبق في جلسة (أو جلسات) اختبارية متعاقبة دون وجود فترات زمنية تفصل بينها (أو بوجود فواصل ضئيلة منعاً للتأثر بعوامل التعب أو الملل، وما شابه). ومن الواضح أن هذه الطريقة تستبعد أثر التباين الناجم عن عامل الزمن، وتمتاز عن طريقة الإعادة بأنها تلغي آثار الفترة الزمنية الفاصلة بين التطبيقين الأول والثاني للاختبار، التي قد تظهر في اكتساب خبرات جديدة، أو نسيان ما تمّ تعلّمه، أو تذكر الإجابات التي أعطاها المفحوص في التطبيق الأول أو غيرها. ويطلق على معامل الثبات المستخرج بهذه الطريقة اسم معامل التعادل أو التكافؤ.

أما الطريقة الثانية لحساب الثبات باستخدام الأشكال المتعادلة فيمكن تسميتها بالأشكال المتعادلة المتعاقبة لكونها تقوم على استخدام أشكال متعادلة تطبق في جلسات اختبارية متعاقبة وبوجود فواصل زمنية طويلة نسبياً. ويطلق على معامل الثبات المستخرج بهذه الطريقة اسم معامل التعادل والاستقرار. ومن الواضح أن هذه الطريقة تتيح تقدير الاتساق في أداء المفحوصين عبر الزمن، كما تتيح تقدير الاتساق في أدائهم على عينات متجانسة من البنود. وتنطوي هذه الطريقة على كل مصادر الخطأ المحتملة بحسب جرونلند، من مثل استقرار إجراءات القياس، وثبات السمات المقيسة، وتمثيل العينة، كما تعدّ الطريقة الأكثر صرامة لحساب الثبات(Gronlund ,1971,p.105).

ومن عيوب طريقة الأشكال المتعادلة في دراسة الثبات صعوبة وضع مقياس ثان معادل للمقياس الأول سواء من حيث تمثيل البنود للسمة المقيسة، ومستوى صعوبة البنود، والصياغة، والتعليمات، والوقت المعطى للمقياس وغيرها. هذا

بالإضافة إلى أن هذه الطريقة تتطلب مضاعفة الجهد والوقت والنفقات بطبيعة الحال.

3. طريقة التنصيف:

وتقوم هذه الطريقة على تقسيم الاختبار بعد تطبيقه في جلسة اختبارية واحدة إلى نصفين يفترض أنهما متكافئان، ثم حساب معامل الارتباط بين الدرجات التي حصل عليها المفحوصون على النصف الأول والدرجات التي حصلوا عليها على النصف الثاني. ويطلق على معامل الثبات المحسوب بهذه الطريقة اسم معامل ثبات التنصيف (أو الشطر النصفي) كما يطلق عليه أحياناً اسم معامل الاتساق الداخلي.

وتشبه هذه الطريقة طريقة الأشكال المتعادلة من حيث أنها تشير إلى ما إذا كانت عينة البنود ممثلة للمحتوى المقيس، كما تغفل التغيرات التي يمكن أن تحدث لدى المفحوصين بنتيجة العامل الزمني(الثبات الزمني). وتختلف عنها من حيث أنها تطبق في جلسة اختبارية واحدة وتعدّ أكثر اقتصادية من ناحية الوقت والجهد والنفقات.

والصعوبة الكبرى التي تعترض هذه الطريقة تتمثل في تقسيم الاختبار إلى نصفين متعادلين مع مراعاة مستوى الصعوبة ودرجة التمثيل وشكل البنود وغير ذلك. والواقع أن معظم الاختبارات يصعب تقسيمها إلى نصفين متعادلين والافتراض أن أداء الطلاب على النصف الأول يعادل أداءهم على النصف الثاني بمجرد شطر الاختبار إلى نصفين حسب التسلسل الرقمي للبنود. ومن الطرائق المتبعة في عملية التنصيف جعل جميع البنود ذات الأرقام الفردية في أحد النصفين وجعل جميع البنود ذات الأرقام الزوجية في النصف الآخر. ويرى ثورانديك وهيجن أن هذا الأسلوب مناسب ما دام الاحتمال كبيراً بأن البنود التي تتشابه في الشكل والمحتوى والصعوبة تتجمع بعضها مع بعض في الاختبار

للمحتوى وأن تكون صياغتها متماثلة، كما تتماثل أو تتعادل من حيث التعليمات والزمن المعطى للاختبار.

يميّز الباحثون عادة بين طريقتين لحساب الثبات باستخدام الأشكال المتعادلة وتتفرعان عن الطريقة الأم مدار البحث. تستخدم في الطريقة الأولى من هاتين الطريقتين أشكال متعادلة "فورية" تطبق في جلسة (أو جلسات) اختبارية متعاقبة دون وجود فترات زمنية تفصل بينها (أو بوجود فواصل ضئيلة منعاً للتأثر بعوامل التعب أو الملل، وما شابه). ومن الواضح أن هذه الطريقة تستبعد أثر التباين الناجم عن عامل الزمن، وتمتاز عن طريقة الإعادة بأنها تلغي آثار الفترة الزمنية الفاصلة بين التطبيقين الأول والثاني للاختبار التي قد تظهر في اكتساب خبرات جديدة، أو نسيان ما تمّ تعلّمه، أو تذكر الإجابات التي أعطاها المفحوص في التطبيق الأول أو غيرها. ويطلق على معامل الثبات المستخرج بهذه الطريقة اسم معامل التعادل أو التكافؤ.

أما الطريقة الثانية لحساب الثبات باستخدام الأشكال المتعادلة فيمكن تسميتها بالأشكال المتعادلة المتعاقبة لكونها تقوم على استخدام أشكال متعادلة تطبق في جلسات اختبارية متعاقبة وبوجود فواصل زمنية طويلة نسبياً. ويطلق على معامل الثبات المستخرج بهذه الطريقة اسم معامل التعادل والاستقرار. ومن الواضح أن هذه الطريقة تتيح تقدير الاتساق في أداء المفحوصين عبر الزمن، كما تتيح تقدير الاتساق في أدائهم على عينات متجانسة من البنود. وتنطوي هذه الطريقة على كل مصادر الخطأ المحتملة بحسب جرونلند، من مثل استقرار إجراءات القياس، وثبات السمات المقيسة، وتمثيل العينة، كما تعدّ الطريقة الأكثر صرامة لحساب الثبات(Gronlund ,1971,p.105).

ومن عيوب طريقة الأشكال المتعادلة في دراسة الثبات صعوبة وضع مقياس ثان معادل للمقياس الأول سواء من حيث تمثيل البنود للسمة المقيسة، ومستوى صعوبة البنود، والصياغة، والتعليمات، والوقت المعطى للمقياس وغيرها. هذا

بالإضافة إلى أن هذه الطريقة تتطلب مضاعفة الجهد والوقت والنفقات بطبيعة الحال.

3. طريقة التنصيف:

وتقوم هذه الطريقة على تقسيم الاختبار بعد تطبيقه في جلسة اختبارية واحدة إلى نصفين يفترض أنهما متكافئان، ثم حساب معامل الارتباط بين الدرجات التي حصل عليها المفحوصون على النصف الأول والدرجات التي حصلوا عليها على النصف الثاني. ويطلق على معامل الثبات المحسوب بهذه الطريقة اسم معامل ثبات التنصيف (أو الشطر النصفي) كما يطلق عليه أحياناً اسم معامل الاتساق الداخلي.

وتشبه هذه الطريقة طريقة الأشكال المتعادلة من حيث أنها تشير إلى ما إذا كانت عينة البنود ممثلة للمحتوى المقيس، كما تغفل التغيرات التي يمكن أن تحدث لدى المفحوصين بنتيجة العامل الزمني(الثبات الزمني). وتختلف عنها من حيث أنها تطبق في جلسة اختبارية واحدة وتعدّ أكثر اقتصادية من ناحية الوقت والجهد والنفقات.

والصعوبة الكبرى التي تعترض هذه الطريقة تتمثل في تقسيم الاختبار إلى نصفين متعادلين مع مراعاة مستوى الصعوبة ودرجة التمثيل وشكل البنود وغير ذلك. والواقع أن معظم الاختبارات يصعب تقسيمها إلى نصفين متعادلين وافتراض أن أداء الطلاب على النصف الأول يعادل أداءهم على النصف الثاني بمجرد شطر الاختبار إلى نصفين حسب التسلسل الرقمي للبنود. ومن الطرائق المتبعة في عملية التنصيف جعل جميع البنود ذات الأرقام الفردية في أحد النصفين وجعل جميع البنود ذات الأرقام الزوجية في النصف الآخر. ويرى ثورانديك وهيجن أن هذا الأسلوب مناسب ما دام الاحتمال كبيراً بأن البنود التي تتشابه في الشكل والمحتوى والصعوبة تتجمع بعضها مع بعض في الاختبار

الواحد. " ففي اختبار طويل نسبياً يصل إلى (60) بنداً أو أكثر تميل التجزئة بهذه الطريقة إلى تحقيق الموازنة في عوامل من نوع شكل الفقرة وشمول محتواها ومستوى صعوبتها. وتكون الاحتمالات حسنة أن يكون نصفا الاختبار المتشكلان بهذه الطريقة متعادلين" (ثورندايك وهيجن، 1989، ص77). ومن الطرائق المتبعة في تقسيم الاختبار بالإضافة إلى هذه الطريقة، تحديد مستوى السهولة أو معامل السهولة لكل سؤال ثم توزيع الأسئلة على نصفي الاختبار تبعاً لتكافؤ مستوى السهولة مع مراعاة التشابه في المضمون. وهذه الطريقة أدق من الطريقة السابقة وتتطلب جهداً زائداً.

بيد أن معامل الارتباط المحسوب بطريقة التنصيف يعطينا ثبات نصف الاختبار وليس الاختبار بكامله. ومن المعلوم أنه كلما كبرت عينة السلوك المقيسة كان قياسها أكثر ثباتاً ودقة. لذا لابد من تصحيح معامل الثبات المحسوب بهذه الطريقة باستخدام معادلة سبيرمان – براون على النحو التالي:

$$\text{ر}_{11} = \frac{2\,\text{ر}}{1+\text{ر}}$$

حيث يشير الرمز (ر١١) إلى معامل ثبات الاختبار بكامله والرمز (ر) إلى معامل ثبات نصف الاختبار. فإذا بلغ معامل ثبات نصف الاختبار 0.60 فإنه يصبح بعد تصحيحه باستخدام المعادلة السابقة 0.75 وهو معامل ثبات الاختبار ككل.

وهذا يشير إلى أن معامل ثبات درجات الاختبار ككل زادت قيمته عن قيمة معامل ثبات درجات كل من نصفيه على حدة.

4. طريقة كودر – ريتشاردسون ومعامل ألفا:

وتقوم طريقة كودر – ريتشاردسون على تقسيم الاختبار الواحد إلى عدد كبير من الأجزاء (بدلاً من تقسيمه إلى جزئين او نصفين متعادلين فقط)، وبحيث يتكون كل جزء من هذه الأجزاء، من بند واحد فقط من بنود الاختبار. وتتطلب هذه الطريقة حساب ارتباط كل من هذه الأجزاء أو البنود مع بعضها البعض، وكذلك ارتباط كل منها مع الاختبار ككل. وبذلك فإن الفكرة الموجهة لهذه الطريقة تتلخّص في دراسة التجانس أو الاتساق في إجابات

المفحوصين على سائر البنود التي يتألف منها الاختبار، وهو ما يعرف بدراسة التجانس او الاتساق الداخلي بين بنود الاختبار Interitem Consistency. ومن الواضح أن هذه الطريقة تفترض أن الاختبار احادي البعد ويقيس سمة أو وظيفة واحدة فقط، وان بنوده جميعها تقيس هذه السمة الواحدة.

من المعلوم ان الاتساق بين البنود يتأثر بمصدرين اثنين من مصادر تباين الخطأ وهما:

1. خطأ العينة او المعاينة والذي يرتبط بمدى تمثيل البنود للمحتوى المقيس (وهذا ما يظهر عند استخدام طريقة الشكل المعادل وطريقة التنصيف في حساب الثبات).
2. عدم تجانس مجال السلوك الذي تتصدّى له بنود الاختبار. فكلما كان مجال السلوك متجانساً كان الاتساق بين البنود التي أعدت لقياسه عالياً، وكلما افتقر هذا المجال إلى التجانس ضعف الاتساق بين البنود التي أعدت لقياسه. ولتوضيح ذلك لنفترض أن لدينا اختبارين يتألف كل منهما من 20 بنداً، وان الأول منهما يتضمّن بنوداً في الجبر فقط، في حين أن الثاني يتضمّن بنوداً في الجبر واخرى في الهندسة ((10)بنود للجبر و(10) بنود للهندسة) فإن الاتساق الداخلي بين بنود الاختبار الأول سيكون أعلى منه بين بنود الاختبار الثاني على الأرجح. وقد يكون أداء أحد التلاميذ في الاختبار الثاني أعلى في الجبر منه في الهندسة، في حين أن أداء الآخر في الهندسة قد يكون أعلى منه في الجبر وهكذا. وبالمثل فإننا لو افترضنا ايضاً، اننا طبقنا اختبارين يتألف كل منهما من 30 بنداً، وان الاختبار الأول منهما يقيس القدرة على الفهم اللفظي فقط، في حين أن الثاني يقيس القدرة على الفهم اللفظي، كما يقيس القدرة العددية، والقدرة المكانية والإدراكية (بمعدل 10 بنود لكل قدرة). ولنفترض أيضاً أننا طبقنا هذين الاختبارين على عشرة أفراد وحصلوا على علامات تقع في المدى من صفر إلى 30 في كل من هذين الاختبارين فهل يمكن أن نتوقع اتساقاً في أداء

المفحوصين على الاختبار الثاني كما نتوقع اتساقاً في أدائهم على الاختبار الأول طالما أن الاختبار الثاني توزعت بنوده إلى ثلاث مجموعات، وأن هذه المجموعات الثلاث تقيس ثلاث قدرات مختلفة أو متمايزة، في حين أن الاختبار الأول ببنوده الثلاثين يختص بقدرة واحدة فقط، وهي القدرة على الفهم اللفظي كما أسلفنا؟

إن الإجابة عن السؤال السابق ستكون بالنفي حتماً. ذلك لأننا سنتوقع درجة عالية من الاتساق في أداء المفحوصين على الاختبار الأول فقط بوصفه اختباراً متجانساً يتناول خاصية أو قدرة واحدة، ولن نتوقع مثل هذا الاتساق في أداء المفحوصين ونتائجهم في الاختبار الثاني الذي صمّم أساساً لقدرات متعدّدة. في الوقت نفسه فإن درجات المفحوصين في الاختبار المتجانس تكون أقل غموضاً منها في الاختبار غير المتجانس أو الأقل تجانساً. فلو حصل اثنان من المفحوصين في الاختبار الثاني "غير المتجانس" على الدرجة نفسها (ولتكن الدرجة 20 مثلاً) فهل يمكن أن نستنتج أن أداء الأول يعادل تماماً أداء الثاني؟ والجواب هو بالطبع: لا. فقد تكون الدرجة 20 التي نالها المفحوص الأول هي حصيلة إجاباته الصحيحة عن البنود العشرة الخاصة بالفهم اللفظي، والبنود العشرة الخاصة بالقدرة المكانية والإدراكية فقط ودون أن يعطي أي إجابة صحيحة عن البنود العددية(أي 10+10=20). هذا في حين أن الدرجة 20 التي نالها المفحوص الثاني قد تكون حصيلة إجاباته الصحيحة عن البنود العشرة الخاصة بالقدرة على الفهم اللفظي، والبنود العشرة الخاصة بالقدرة المكانية والإدراكية دون أن يعطي أي إجابة صحيحة عن البنود العددية (أي 10 + 10 + صفر=20). وبالطبع فإن العلامة 20 يمكن أن تكون حصيلة جمع أي علامات ينالها المفحوص في كل من المجموعات الثلاث من البنود بدءاً بالعلامة صفر حتى العلامة 20.

وفيما يتصل بالاختبار الأول المتجانس نسبياً فإننا لو افترضنا حصول أحد المفحوصين على العلامة 20 فسوف تعني هذه العلامة على الأرجح أن هذا المفحوص أجاب عن البنود العشرين الأولى التي يتكون منها الاختبار، إذا كانت

بنود هذا الاختبار مرتبة وفق مبدأ التدرج في الصعوبة. وبالطبع هناك احتمال في أن يخفق هذا المفحوص في اثنين أو ثلاثة من البنود الأسهل الواقعة بين البند الأول والبند العشرين، ويعطي إجابات صحيحة عن اثنين أو ثلاثة من البنود الأصعب الواقعة بعد البند العشرين، غير أن هذه التذبذبات في أداء الفرد ستكون طفيفة بالمقارنة مع التذبذبات التي يمكن توقعها حين يكون الاختبار أقل تجانساً. ومن الواضح أن الدرجة 20 التي نالها الفرد في هذه الحالة يمكن مقارنتها بالدرجة 20 التي حصل عليها زميله مثلاً وتوضح أنهما متساويان في القدرة على الفهم اللفظي، كما أن هذه الدرجة قابلة للمقارنة مع أي درجة أخرى ينالها أي من المفحوصين الآخرين لكونها تشير إلى مستوى محدد من القدرة يعلو عن مستويات القدرة التي تعبر عنها درجات من نالوا أقل من الدرجة 20، ويهبط عن مستويات القدرة التي تعبر عنها الدرجات الأعلى من الدرجة 20.

لقد وضع كودر وريتشاردسون عدداً من المعادلات الخاصة بحساب الاتساق الداخلي بين بنود الاختبار. وهذه المعادلات جميعها تقوم على حساب درجة الاتساق في أداء الفرد من بند لآخر، وتتطلب تطبيق الاختبار لمرة واحدة فقط، ولا تتطلب إعادة تطبيقه، أو شطره إلى نصفين متعادلين. ومن أهم هذه المعادلات معادلة كودر – ريتشاردسون 20 التي تعرف اختصاراً بـ (ك – ر20)، ونصها كما يلي:

$$\text{ر} = \frac{\text{ن}}{\text{ن}-1} \times \frac{\text{ع}^2 - \text{مج ص ع}}{\text{ع}^2}$$

حيث يشير الرمز ر إلى معامل ثبات الاختبار.

والرمز ن إلى عدد البنود (الأسئلة).

والرمز ع^2 إلى تباين درجات الاختبار.

والرمز ص إلى نسبة الإجابات الصحيحة عن كل بند (معامل السهولة).

والرمز خ إلى نسبة الإجابات الخاطئة عن كل بند (معامل الصعوبة).

ولا تصلح طريقة كودر – ريتشاردسون للاختبارات الموقوتة أي الاختبارات التي تعتمد على السرعة في الأداء، بل تصلح لاختبارات القوة. ويمكن استخدام هذه الطريقة حين تكون الإجابة على البند ثنائية وتعطى الدرجة 1 أو صفر كأن تكون الإجابة بـ "نعم او لا"، "موافق" او "غير موافق"، "صح" او "خطأ".

بالإضافة إلى ذلك يمكن استخدام هذه المعادلة حين تكون بنود الاختبار متجانسة بمعنى انها تتناول سمة احادية البعد Unidimensional او تقيس متغيراً واحداً. اما في الحالات التي يتصدّى فيها الاختبار الواحد لمجموعة من السمات أو القدرات، ويشتمل على اختبارات فرعية غير متجانسة فإنه لابد من حساب ثبات الاتساق الداخلي لكل من هذه الاختبارات الفرعية على حدة، وبصورة مستقلة عن الآخر. ويطلق على معامل الثبات المستخرج في هذه الحالات معامل التجانس Homogeneity Coefficient.

ومن الواضح انه إذا استخدمت طريقة كودر – ريتشاردسون في حساب ثبات الاختبار الذي يقيس أكثر من سمة واحدة ستهبط قيمة معامل الثبات المحسوب هبوطاً شديداً.

غير ان طريقة كودر – ريتشاردسون تصلح للاختبارات ثنائية الدرجة، أي التي يعطى البند الواحد فيها الدرجة 1 او صفر كما سبقت الإشارة، ولا تصلح للاختبارات وأدوات القياس "غير الثنائية" أي الأدوات التي يعطى البند الواحد منها الدرجة "1" او "2" او "3" او أكثر، كمقاييس الاتجاهات، وبعض استخبارات الشخصية، إضافة إلى الاختبارات التحصيلية المقالية، وسلالم التقدير، وغيرها. وقد تمكن كرونباخ من وضع معادلة تصلح لهذه الأدوات "غير الثنائية" اشتقها من معادلة كودر – ريتشاردسون 20 التي سبقت الإشارة إليها، وأطلق عليها معامل ألفا. ويمثّل معامل ألفا متوسط قيم المعاملات الناتجة عن تجزئة الاختبار إلى أجزاء بطرق مختلفة، وبذلك فإنه يمثّل معامل الارتباط بين أي جزئين من أجزاء الاختبار. والمعادلة الخاصة بحسابه هي التالية:

$$\text{معامل ألفا } a = \frac{ن}{ن-1}\left(1 - \frac{\text{م مج ع}^2\ ن}{\text{ع}^2\ ك}\right)$$

حيث يشير الرمز ع^2 إلى تباين كل بند من بنود الاختبار.

والرمز مج ع^2 ن إلى مجموع تباينات درجات جميع البنود.

والرمز ع^2 ك إلى تباين الاختبار ككل.

والرمز ن إلى العدد الكلي لبنود الاختبار.

5. ثبات المصححين:

لقد سبقت الإشارة إلى أن هناك العديد من أدوات القياس النفسي والتربوي التي تعاني من مشكلة تدخل العوامل الذاتية في التقدير ووضع الدرجات، التي تؤدي إلى ظهور فروق بين مصحح وآخر، بل حتى عند المصحح نفسه من وقت لآخر في هذا التقدير. والواقع أن هذه الفروق في التقدير تعدّ أحد المصادر الرئيسة لتباين الخطأ في الدرجة على الاختبار، وتمثّل العيب الأكبر من عيوب الاختبارات وأدوات التقويم التقليدية. كما تمثّل عيباً لسائر الاختبارات والأدوات "غير الموضوعية" التي تعتمد على تقدير المصحح، وليس على مفتاح للتصحيح يعطي عادة الدرجة 1 للإجابة الصحيحة وصفر للإجابة الخاطئة. ومن الواضح أن اختبارات التحصيل المدرسية ومقاييس القدرة الإبداعية، والكثير من الأدوات التي تستخدم في قياس الشخصية، كالأدوات الإسقاطية، والمقابلة، وسلالم التقدير وغيرها تفسح مجالاً واسعاً لتدخل العوامل الذاتية (أو الشخصية) في التقدير، وتعاني بالتالي من هبوط، وأحياناً من هبوط حاد في مستوى ثباتها.

تظهر مشكلة الذاتية في التقدير بصورة واضحة للعيان في مقاييس القدرة الإبداعية حيث يتعين على الفاحص في هذه المقاييس أن يحدّد بنفسه مدى أصالة الاستجابة التي يعطيها المفحوص ومرونته في التفكير على الرغم من وجود "معايير" للتقدير يمكنه الاستئناس بها بطبيعة الحال. وتظهر هذه المشكلة ذاتها في الأدوات الإسقاطية حتى حين تستخدم محكات معينة يمكن اعتمادها أساساً في تفسير أو تأويل استجابة المفحوص والحكم على أدائه. ذلك لأن تأويل استجابة المفحوص والحكم على ما إذا كانت تنطبق على المواصفات أو الشروط التي يتضمّنها المحكّ أمر ذاتي يعود إلى الفاحص نفسه، وليس هناك ما يمنع احتمال ظهور فروق في هذا التأويل عند اثنين من الفاحصين يستخدمان المحكّات ذاتها في تقييم استجابات المفحوصين، بل حتى عند الفاحص نفسه من وقت لآخر.

إن السؤال الذي يطرح نفسه في الحالات السابقة وغيرها من الحالات المشابهة هو التالي: ما مدى الثقة بتقديرات الفاحصين؟ وبعبارة أدق، ما مقدار تباين الخطأ الذي يعود مصدره إلى اختلاف تقديرات الفاحصين أو المصححين؟

وللإجابة عن السؤال السابق سيكون من المفيد استخدام مقياس إحصائي يكشف عن مدى الاتساق (أو الثبات) في تقديرات الفاحصين أو المصححين، وهو ما يدخل في نطاق ما يعرف بثبات المصححين Scorer Reliability. والمقياس الإحصائي الذي يمكن استخدامه في هذه الحالة هو معامل الارتباط، أو معامل ثبات المصححين، والذي يشير إلى درجة الارتباط بين التقديرات أو الدرجات التي يعطيها أحد المصححين لإجابات مجموعة من المفحوصين على أسئلة الاختبار موضع الدراسة، والدرجات التي يعطيها مصحح آخر (أو أكثر) لهذه الإجابات ذاتها.

تظهر فائدة استخدام معامل ثبات المصححين في سائر الاختبارات التي تعاني من مشكلة الذاتية في التقدير ولاسيما تلك الاختبارات التي تجري لأغراض بحثية، أو الاختبارات التي قد تعتمد نتائجها أساساً في الحكم على

الأفراد واتخاذ قرارات معينة بشأنهم. غير ان معامل ثبات المصححين لا يمثّل الوسيلة الوحيدة لتقدير ثبات المصححين، وقد يكتفي في حالات معينة بتحديد النسبة المئوية للاتفاق بين المصححين أو المقدرين أو الذين يتولوّن إجراء المقابلة الشخصية للأفراد، ولاسيما حين يكون الهدف من عملية التقدير هو الانتقاء أو تصنيف الأفراد إلى مجموعتين من مثل "مقبول" و "غير مقبول" و"ناجح" و "راسب" و "منطوي" و "منبسط" وهكذا.

طرق أخرى لدراسة الثبات:

بالإضافة للطرق السابقة في دراسة الثبات هناك طرق أخرى تجدر الإشارة إليها، لعل من أهمها طريقة التحليل العاملي Factor analysis، والطريقة التي تعتمد على تحليل التباين Analysis of variance. ومن المعلوم أن التحليل العاملي يستهدف الوصول إلى أقل عدد من العوامل التي يمكن أن تفسر الأداء على الاختبار استناداً إلى دراسة الترابطات بين البنود التي يتكون منها هذا الاختبار، والكشف عن تلك التجمعات من البنود التي يتشبّع كل منها بعامل معين دون غيره من العوامل. ومن خلال تحديد العوامل التي تتشبّع بها تجمعات معينة من البنود يمكن تصنيف هذه البنود في مجموعات متجانسة تختص كل منها بعامل او سمة معينة. وبطبيعة الحال، فإنه في الحالات التي يسفر فيها التحليل العاملي عن وجود عامل مشترك تتشبّع به سائر البنود يمكن النظر إلى هذه البنود على أنها تؤلف مجموعة واحدة متجانسة، وإلى الاختبار ككل على أنه على درجة لا بأس بها من الثبات أو الاتساق. وبذلك يتيح التحليل العاملي تحديد مدى التجانس بين بنود الاختبار، ويمثّل طريقة لدراسة الاتساق بين البنود والذي يمثّل جوهر مسألة الثبات.

وفيما يتصل بتحليل التباين فإنه يمثّل بدوره طريقة أخرى لدراسة الثبات أو الاتساق تقوم على إرجاع التباين الكلي لدرجات الاختبار إلى ثلاثة مصادر للتباين وهي: التباين المتعلق بالأفراد المفحوصين، والتباين المتعلق بالبنود،

والتباين الخاص بالعلاقة التفاعلية بين الأفراد والبنود. ويمكن تقدير التباين المتعلق بالأفراد المفحوصين بحساب مجموع مربعات انحرافات درجات أولئك الأفراد عن المتوسط العام لدرجاتهم. أما تباين الخطأ فيقدّر عن طريق مجموع المربعات المحسوبة لتفاعل الأفراد مع البنود. وبذلك يمكن حساب معامل ثبات الاختبار باستخدام المعادلة التالية لتحليل التباين:

$$\text{معامل الثبات} = \frac{\text{التباين المتعلق بالأفراد} - \text{التباين المتعلق بتفاعل الأفراد مع البنود}}{\text{التباين المتعلق بالأفراد}}$$

ويشير معامل الثبات المحسوب إلى مدى التجانس بين بنود الاختبار، أو اتساقه الداخلي.

هذا وقد اقترح جاكسون معامل ثبات من نوع آخر وأطلق على هذا المعامل اسم معامل الحساسية. والمعادلة الخاصة بحساب معامل الحساسية بحسب جاكسون هي:

$$\text{معامل الحساسية} = \frac{\text{التباين المتعلق بالأفراد} - \text{تباين التفاعل بين الأفراد و البنود}}{\text{تباين التفاعل بين الأفراد والبنود}}$$

ويتطلب تفسير معامل الحساسية العودة إلى مستويات الدلالة الإحصائية التي تتحدّد في ضوء فرضية التوزيع الاعتدالي للدرجات.

معامل الثبات والخطأ المعياري للقياس:

يشير الثبات كما أسلفنا إلى درجة الاتساق والدقة التي يتمتع بها الاختبار في قياس ما يقيسه، أو الدرجة التي يقيس عندها الاختبار الشيء نفسه مرة بعد أخرى وبنداً بعد بند. وليس هناك حد أدنى مقبول لدرجة الثبات التي يجب أن تتوفر في الاختبار كما يعبر عنها بمعامل الثبات وإن كان المفضل بطبيعة الحال ارتفاع هذا المعامل إلى الحدود القصوى المتاحة. ويمكن القول: إن

السؤال حول معامل الثبات المطلوب يرتبط مباشرة بنوع القرارات التي ستتخذ استناداً إلى نتائج الاختبار، فكلما كان القرار هاماً تطلب ذلك ارتفاع الثبات، وعندما نكون بصدد اتخاذ قرارات نهائية وغير قابلة للمراجعة يترتب علينا البحث عن الأداة الأكثر ثباتاً إذ لا يصح اعتماداً على مقاييس ضعيفة الثبات اتخاذ قرارات بتقديم منح دراسية أو رفض متقدّمين للدراسة في الجامعة مثلاً. وعموماً "توفّر اختبارات التحصيل المقننة وكذلك اختبارات الذكاء أعلى معاملات الثبات حيث لا تقل هذه المعاملات فيها عن 0.85 على حين لا تتجاوز في بعض اختبارات الشخصية 0.50 أما الاختبارات الصفية التي يعدّها المعلمون عادة فتصل معاملات الثبات فيها في احسن الأحوال إلى 0.60".

(Ahmann and Glock,1975,p.255)

يرتبط مفهوم معامل الثبات ارتباطاً وثيقاً بمفهوم الخطأ المعياري للقياس Standard error of measurement (SEM). فإذا أخذنا في الاعتبار أن الخطأ المعياري للقياس هو الانحراف المعياري لأخطاء القياس، وأن معامل الثبات هو ذلك الجزء من التباين الكلي الذي لا ينتج عن أخطاء القياس أمكن القول: إنه كلما تناقص التباين الناتج عن الخطأ، أي انخفض الخطأ المعياري للقياس، ارتفع معامل الثبات، وكلما ازداد تباين الخطأ بالنسبة للتباين الكلي انخفض معامل الثبات. ويمكن استخدام معامل الثبات في تقدير الخطأ المعياري للقياس ومن ثم تحديد مدى دقة كل علامة على حدة طالما أن الخطأ المعياري للقياس يشير إلى زمرة الخطأ المحيطة بكل علامة. وهذا يعني أن الخطأ المعياري للقياس يشير إلى المدى المحتمل للخطأ في علامة الفرد، وبالتالي فهو يفيد في تقدير العلامة الحقيقية للفرد. ويحسب الخطأ المعياري للقياس بالمعادلة التالية:

$$ع_{خ} = ع \times \sqrt{1 - ر_{ث}}$$

حيث يشير الرمز ع خ إلى الخطأ المعياري للقياس،

والرمز ع إلى الانحراف المعياري للدرجات،

والرمز ر ث إلى معامل ثبات الاختبار،

أي أن الخطأ المعياري للقياس = حاصل ضرب الانحراف المعياري للدرجات في الجذر التربيعي لمعامل الثبات مطروحاً من الواحد الصحيح. فإذا كان الانحراف المعياري للدرجات هو 5، وكان معامل ثبات الاختبار هو 0.91 فإن الخطأ المعياري للقياس سيكون:

بالتعويض عن المعادلة السابقة:

$$ع_{خ} = 5 \times \sqrt{0.91 - 1}$$

$$ع_{خ} = 5 \times \sqrt{0.09}$$

$$ع_{خ} = 5 \times 0.3 = 1.5$$

ويمكننا القول في هذه الحالة إن الدرجة الحقيقية للفرد لا تختلف عن الدرجة التي حصل عليها إلا بمقدار درجة ونصف زيادة أو نقصاناً، وذلك في 68٪ من الحالات أو بنسبة 2: 1 من المرات. ذلك لأن 68٪ من المساحة تحت المنحنى الاعتدالي تنحصر بين + 1ع و – 1ع، أي بين انحراف معياري واحد (أو وحدة خطأ معياري واحد) إلى يمين المتوسط وانحراف معياري واحد (أو وحدة خطأ معياري واحد) إلى يسار المتوسط.

فلو كانت الدرجة التي حصل عليها الفرد هي 50،

فإن درجته الحقيقية في 68٪ من الحالات ستتراوح بين 50 +1.5 و50 – 1.5.

وهذا يعني أننا على ثقة بنسبة 68٪ بأن الدرجة الحقيقية لهذا الفرد تقع في المدى من 51.5 إلى 48.5.

ولو أردنا حساب درجة الفرد الحقيقية بدرجة أعلى من الثقة، كأن تكون بنسبة 95٪ من الثقة مثلاً، فإنه سيكون بإمكاننا أن نفعل ذلك استناداً إلى خواص المنحنى الاعتدالي حيث تنحصر 95٪ تقريباً من المساحة تحت المنحني الاعتدالي بين +2ع و – 2ع، (أي بين وحدتين للخطأ المعياري فوق المتوسط ووحدتين تحته). وفي هذه الحالة سنضرب قيمة الخطأ المعياري للقياس بـ2ع بدلاً من 1ع، أي 1.5×2 = 3 وسوف تتراوح الدرجة الحقيقية لهذا الفرد بين 50 + (1.5×2) و 50 – (1.5×2) أي بين 53 و 47 وسيمكننا القول: إننا على ثقة بنسبة 95٪ (أو في 95٪ من الحالات) بأن الدرجة الحقيقية لهذا الفرد لا تختلف عن الدرجة التي حصل عليها إلا بمقدار 3 درجات زيادة أو نقصاناً، أي أنها تقع في المدى من 53 إلى 47.

ويتبين مما سبق أن الخطأ المعياري للقياس يشير إلى مقدار التذبذب (أو الخطأ) المحتمل في درجة الفرد إذا ما أعيد تطبيق الاختبار عليه عدداً كبيراً من المرات. في الوقت نفسه، فإن المدى الذي تقع فيه الدرجة الحقيقية للفرد يزداد بزيادة درجة الثقة المطلوبة. فلو أردنا زيادة درجة الثقة لتصل إلى ما يقارب 100٪ من الحالات (أي في المدى من +3ع إلى – 3ع حيث تنحصر 99.7٪ تقريباً من مساحة المنحنى الاعتدالي)، فإن المدى الذي ستقع فيه درجة الفرد في مثالنا سيتراوح من 50± (1.5×3) إلى 50 – (1.5×3) أي من 54.5 إلى 45.5. هذا مع الإشارة إلى أنه يكتفى عادة بالمدى من +1ع إلى – 1ع، وفي حالات قليلة فقط يفضل توسيع هذا المدى كأن نرغب في مقارنة الدرجات التي يحصل عليها الفرد الواحد في اختبارات متعدّدة من خلال رسم الصفحة النفسية له، أو غيرها من الحالات.

ومن المفيد الإشارة إلى أن استخدام الخطأ المعياري للقياس والتعبير عن أداء الفرد بصورة زمرة من الدرجات تقع ضمنها درجته الحقيقية يقف في وجه التشديد المفرط على درجة رقمية واحدة تعبر عن أداء الفرد وما قد ينطوي عليه من مدلولات. ولعل مما يؤكد ذلك تزايد عدد الاختبارات المنشورة حالياً التي توفّر ثبتاً للدرجات، لا بصورة أرقام مفردة، بل بصورة زمر من الدرجات التي يمكن أن تقع ضمنها الدرجة الحقيقية للفرد. وقد يشير هذا الأمر بالذات إلى أحد ملامح التطوّر الذي تشهده حركة القياس في الظرف الراهن.

العوامل المؤثرة في الثبات:

ثمة عوامل عديدة تؤثر في ثبات المقياس منها:

1. عدد البنود:

فكلما ازداد عدد البنود ارتفع معامل الثبات. وهذا يعود إلى أن الاختبار الأطول يقدّم عينة أكثر تمثيلاً للسلوك المقيس ويعكس الفروق الحقيقية في هذا السلوك بصورة أفضل، وتكون الدرجات المتحصّلة على هذا الاختبار أقل تأثراً بالتخمين.

2. انتشار الدرجات:

كلما ازداد انتشار الدرجات ارتفعت قيمة معامل الثبات وتضاءل الأثر الذي تحدثه أخطاء القياس على مركز الفرد داخل المجموعة. ومن الواضح أن كل ما من شأنه أن يضعف احتمال تذبذب مراكز الأفراد داخل المجموعة من قياس لآخر يرفع الثبات فإذا ازداد انتشار الدرجات تضاءل هذا الاحتمال. لاحظ المثال التالي لمجموعتين من الدرجات حصلت عليها مجموعتان من الأفراد:

88	88
81	87
72	85
68	82
56	81
48	79
41	78
33	77

فإذا أعدنا القياس للمرة الثانية يزداد احتمال تغير مراكز الأفراد في المجموعة الأولى أكثر من الثانية وبالتالي يضعف الثبات فيما يتصل بنتائج المجموعة الأولى.

3. صعوبة البنود:

فالاختبارات الصعبة جداً أو السهلة جداً تعطي درجات ضعيفة في ثباتها لأنها تؤدي إلى انتشار محدود للدرجات وتظهر فروقاً طفيفة بين الأفراد. وعموماً إذا شاء واضع المقياس أن يزيد من ثباته عليه أن يضع أسئلة يتراوح مدى صعوبتها بين 0.30 و0.70 وبحيث يقترب معامل صعوبة الاختبار ككل من 0.50.

4. الموضوعية في التصحيح:

فلكي يعطي المقياس نتائج ثابتة يجب ألا تختلف الدرجات من مصحح لآخر أو من وقت لآخر. ومن المعلوم أن الاختبارات الموضوعية عموماً تحقق قدراً عالياً من الثبات بخلاف الاختبارات المقالية التي تفسح المجال لتدخل العوامل الذاتية للمصحح.

- جدول رقم ٣ تفسير قوة العلاقة لمعاملات الثبات

معامل الارتباط	تفسير العلاقة
أقل من ٢٠%	العلاقة ضعيفة الأهمية – الارتباط غير جدير بالاهتمام
من ٢٠% إلى ٤٠%	الارتباط منخفض – واقع العلاقة صغير
من ٤٠% إلى ٧٠%	الارتباط معتدل – العلاقة أساسية
من ٧٠% إلى ٩٠%	الارتباط مرتفع – العلاقة قوي
من ٩٠% إلى ١	الارتباط عالٍ جداً – العلاقة معتمد عليها جداً

الفصل السابع

تقنين المقاييس والاختبارات

يمثل التقنين المرحلة النهائية من مراحل بناء الاختبار الجاهز للاستعمال ، وعليه تبرز أهمية عملية تقنين المقاييس في جدولة الدرجات الخام التي تستحصلها عينات التقنين نتيجة إجرائه واشتقاق المعايير ، ممهدة بذلك لجعل الاختبار المقنن في حيز التنفيذ وإجراءه على عينات مماثلة المجتمع التقنين واستعماله وسيلة في تحقيق أهداف البحوث كالكشف عن الظواهر وفي توفير أوقات الباحثين وجهودهم التي سيبذلونها لوضع اختبارات قد تحقق أهداف واضعيها.

يشير مصطلح التقنين إلى توحيد الشروط المحيطة بالمفحوصين كافة وضبط العوامل والمتغيرات جميعها التي يمكن أن تؤثر أدائهم الاختباري وتحييد أثرها إلى الدرجة القصوى ، وذلك بهدف قياس المتغير الذي صمم الاختبار لقياسه بحجمه الحقيقي لدى تلك العينة من الأفراد التي سيطبق عليها ودون أن يكون أي من المتغيرات الأخرى أثره في أداء أفراد تلك العينة ، وقد يكون من الخطأ النظر إلى التقنين على أنه خطوة أخيرة ومتممة لعملية بناء الاختبار أو لاحقة بها ، والأصح أن ننظر إليه على أنه عملية مواكبة لها أو جزء لا يتجزأ منها ، ويميز علماء القياس بين معنيين للتقنين وهما:

١. أن تكون تعليمات الاختبار وصياغة بنوده والزمن المخصص له وشروط تطبيقه وطريقة تصحيحه موحدة كل المواقف وبما يسمح بإمكان الحصول على النتائج نفسها في حال إعادة إجراء الاختبار ، ومن هذه الناحية يعني التقنين التوحيد ، ويفقد الاختبار اساسه العلمي والموضوعي إذا لم يكن مقننة بهذا المعنى .

٢. أن يخضع الاختبار للتقنين من خلال تطبيقه على عينة ممثلة للمجتمع الأصلي بهدف استخراج معايير معينة تحدد معنى الدرجة الخام التي يحصل عليها الفرد في ضوء تمركز وتشتت درجات افراد عينة التقنين .

وغني عن البيان أنه لا يمكن أن يكون الاختبار مقننا بالمعنى الثاني دون أن يكون مقننا بالمعنى الأول ، وأن عملية استخراج معايير الاختبار ، وهي المرحلة الأخيرة من مراحل تطويره ، تحقق فوائد وأغراض عديدة منها تعرف الوضع النسبي للفرد ضمن المجموعة ، وتفسير الدرجة التي يحصل عليها ، بالإضافة إلى ذلك فإن هذه العملية - عملية استخراج المعايير هي عملية كبيرة ومجهدة من الناحية التجريبية والإحصائية وتستلزم تعاون عدد غير قليل من المختصين والخبراء ، كما أنها مكلفة مادية وتستغرق وقتا غير قليل ، ومن المفيد الإشارة هنا إلى أن الاختبار المقنن بالمعنى الأول ، أي الذي الخضع للتقنين دون أن تستخرج معاييره ، هو أداة قياس علمية ودقيقة ويمكن أن تحقق أغراض وفوائد عديدة في مجالات التشخيص والبحث العلمي ، بل والحالات جميعها التي لا يتركز فيها اهتمام الفاحص على تفسير الدرجة الخام التي يحصل عليها الفرد المفحوص وتعرف وضعه النسبي ضمن المجموعة .

والتقنين يقصد به " أن يكون بناء وتصحيح وتنظير نتائج الإختبار أو أداة القياس مستنداً إلى قواعد محددة بحيث تتوحد فيه وتتحد بدقة مواد الإختبار وطريقة تطبيقه وتعليمات إجابته وطريقة تصحيحه أو تسجيل درجاته ".

والتقنين هو " رسم خطة شاملة وواضحة ومحددة لجميع خطوات الاختبار واجراءاته وطريقة تطبيقه وتفسير درجاته وتحديد السلوك المطلوب من الفرد والشروط المحيطة به اثناء تطبيق الاختبار بالاضافة الى وجود معايير لتفسير النتائج " .

ويعرف الاختبار المقنن بانه " الاختبار الذي حددت أجهزته واجراءات تطبيقه وتصحيحه ، فيصبح من الممكن عندئذ إعطاؤه أو إجراؤه في أوقات وأماكن مختلفة وتكون له معايير مشتقة من عينات ممثلة للمجتمع الأصلي ".

ويعرف ايضاً بانه " الإختبارات التي يتم إعدادها بمعرفة باحث أو فريق من الباحثين المختصين في أي مجال من مجالات القياس التربوي او النفسي أو غيرها وتتميز الإختبارات المقننة بأنَّها مصممة بعناية فائقة وأنَّه قد تم تجريبها مرات كثيرة للتحقق من مدى صلاحيتها ".

والاختبار المقنن له تعليمات تحدد طريقة تطبيقه وتسجيل نتائجه ، و يفترض به أن يكون مطبق على (عينات التقنين) فمن صفاتها أنها تمثل المجتمع المبحوث اصدق تمثيل بغية تحديد المعايير الواضحة لهذا الاختبار ، ومن سمات الاختبار المقنن الدرجة العالية من الموضوعية.

والباحثون يستخدمون الاختبارات المقننة للأسباب التالية:

١. توفر عليهم بذل الجهد والوقت لوضع اختبارات قد تحقق او لا تحقق الهدف الذي يسعون اليه.

٢. تمكنهم من اجراء مقارنة بين اداء الأفراد عندهم مع افراد اخرين طبق عليهم الاختبار نفسه .

١- خصائص عينة التقنين :

١- يجب أن تمثل عينة التقنين ، المجتمع الأصلي المراد دراسته تمثيلا صادقة من حيث الحجم والتركيب والنسب ومستوى القدرة أو الصفة المراد قياسها .

٢- كلما كبر حجم العينة (أي كلما زاد عدد أفرادها) زاد الاعتماد على نتائج الاختبار عليها ... وحجم العينة يعتمد على :

- حجم المجتمع الأصلي المقصود دراسته .
- طبيعة الاختبار ، ونوع الصفة المقاسة .
- إمكانيات الباحث ، ومدى تعاون الجهات الأخرى معه .

٣- يجب أن يتساوى متوسط وتشتت أفراد العينة مع متوسط أفراد المجتمع الأصلي .

٤- كلما صغر حجم القطاع (من حيث التجانس والمسافة وعدد الأفراد) المراد قياسه ، تمكن الباحث من تمثيله في العينة وصدقت معاييره .

٢- **مميزات الاختبارات المقننة :**

١. أنها تتمتع بمعاملات صدق وثبات مقبولة بالنسبة لعينة التقنين التي أعدت لها في الأصل.

٢. أن لها كراسة تعليمات توضح كافة المعلومات اللازمة لتطبيق الاختبار.

٣. أن لها معايير تعكس مستويات اداء عينة (مجموعة) التقنين الأصلية .

٤. أن لها مفتاح تصحيح يوضح اتجاهات العبارات (الأسئلة) الموجبة والعبارات السالبة لتعيين الدرجات الخاصة بكل إجابة .

٥. تعرف هذه الاختبارات في معظم الأحيان باسم الاختبارات المنشورة وذلك لكونها تحظى بالنشر من قبل الدوريات والمراجع العلمية المتخصصة والتي تتمتع بسمعة محلية وعالمية.

٣- **الشروط الواجب توفرها في الاختبارات المقننة:**

١. أن يكون للاختبار تعليمات تعطى لجميع المفحوصين بطريقة موحدة وعلى أن تشمل هذه التعليمات الهدف من الاختبار ، الشروط الخاصة بالأداء ، الأخطاء الشائعة ، الزمن المخصص للاداء ، كيفية الاجابة (الاداء) على الاختبار.

٢. أن يكون للاختبار مفتاح تصحيح (اختبارات الورقة والقلم) لتعيين الاجابات الصحيحة على أسئلة أو وحدات للاختبار

٣. أن يكون للاختبارات معاملات ثبات وصدق معلنة وصريحة بالنسبة لمجموعات الأفراد الذين اعد لهم الاختبار في الأصل (مجموعات او عينات التقنين المرجعية) .

٤. أن يكون للاختبار معايير (جداول مستويات) تظهر بوضوح درجات ومستويات اداء عينة التقنين الأصلية على الاختبار.

٤- **خطوات تقنين المقاييس :**

لكي يتم التقنين يجب أتباع الخطوات الآتية:

- ✓ أولا : تطبيق المقياس : يتم تطبيق المقياس بصورته النهائية بعد أكمال عملية البناء على عينة التقتين ويتم التأكيد على ضرورة قراءة تعليمات وفقرات المقياس بدقة والإجابة عن جميع فقراته بصدق وأمانة .
- ✓ ثانيا : تصحيح مقياس : إتباع الإجراءات نفسها في تصحيح المقياس المتبعة في التجربة الأساسية ، بغية الحصول على الدرجة الكلية لكل فرد من عينة التقنين ، بموجب مفتاح التصحيح المعد للمقياس ، وبعد استخراج النتائج النهائية يفضل وضع التوزيعات التكرارية لهذه الدرجات وأيضا المدرج التكراري لها المعرفة هل الدرجات تتوزع قريبا من التوزيع الطبيعي مما يعطي لنا إمكانية تعميم النتائج على مجتمع البحث ، لذا

يجب استخراج الوسط الحسابي والانحراف المعياري والوسيط والمنوال ومعامل الالتواء لمعرفة مدى قرب الدرجات أو بعدها من التوزيع الطبيعي.

- ✓ ثالثا : اشتقاق المعايير للمقياس : إن البحث العلمي لا يهدف فقط إلى محاولة تطبيق المقاييس المستعملة حالياً ، بل يسعى إلى بناء العديد من المقاييس الجديدة التي تتوافر فيها المعايير لتطبيقها في المجالات المتعددة .
- ✓ رابعا : تحديد المستويات للمقياس : إن المعايير ليست مطلقة أو مستقرة بل هي معايير نسبية تعتمد اعتمادا أساسيا على جماعات مرجعية معينة ، وتتأثر كثيرا بتغير خصائص هذه الجماعات التي يقارن بها الفرد في سمة معينة أو مجموعة من السمات التي يقيسها الاختبار.

٥- مفهوم المعايير:

تعد المعايير أساسا لتفسير درجة الفرد على اختبار ما في ضوء الأداء الفعلي للأفراد الآخرين الذي ينتمي إليهم هذا الفرد لتعطي اختبارات قابلة للمقارنة بين الأفراد أو بين الاختبارات ، فالدرجة الخام التي يحصل عليها المفحوص في الاختبار ليس لها معنى بحد ذاتها ولا يمكن تفسيرها إلا بمقارنتها بمعيار معين ، وبذلك تعتمد المعايير كجداول تستخدم لتفسير درجات الاختبار ، ولتسهل على واضع الاختبار لتدله عما إذا كانت درجات العينة في المستوى المتوسط أو فوق المتوسط أو أقل من المتوسط بالنسبة لعينة التقنين التي استخدمت في بناء المعايير ، وبذلك توفر لنا أساسا للمقارنة ، لذا فالمعايير تعد أحد الأهداف الأساسية التي تهدف إليها عملية التقنين.

ونظراً لأهمية المعايير في عملية تقويم نتائج الاختبارات ، وكذا تقويم المختبرين في الصفات أو الظواهر المقاسة ، عليه نجد من الأهمية تحويل الدرجات الخام التي يحصل عليها واضع الاختبار من جراء تنفيذ الاختبارات الى درجات معيارية لكي يكسب النتائج دلالة ومعنى واضحين ، وينظر إلى المعايير على أنها درجات معيارية ، أو أنها جداول تستخدم لتفسير درجات الاختبار (أي الدرجات الخام التي تمثل النتيجة المستخلصة من جراء تطبيق الاختبار دون اخضاعها للمعالجة الاحصائية) والمعايير اشكال منها : (الدرجة الذائية (ذ أو ص) والدرجة التائية (ت) وقد تسمى الذائية المعدلة ، الميئينات والرتب المئينية ، التساعيات ... الخ ، ولا يمكن الحصول على هذه الأشكال الا من خلال تحويل الدرجات الخام الى درجات معيارية بواسطة استخدام أساليب احصائية معينة..وبغية تأشيراهمية تحويل الدرجات الخام الى درجات معيارية بواسطة استخدام اساليب احصائية معينة.

والمعايير هنا قيم تمثل أداء مجتمع خاص في اختبار معين ، أي أنها تصف أداء مجموعة متعددة على اختبار أو قائمة (استبيان) ، والمعايير وصفية الأنماط موجودة من الأداء ، ولكن من الخطأ اعتبار المعايير مستويات ، وذلك أن المعايير معلومات تدلنا على كيفية الأداء الفعلي للأفراد . أما المستويات معلومات تدلنا على ما يجب أن يؤديه الفرد.

٦-أهمية تحويل الدرجات الخام إلى درجات معيارية:

١. يقصد بالدرجة المعيارية هي الدرجة التي يعبر فيها عن درجة كل فرد على اساس عدد وحدات الانحراف المعياري لدرجته عن المتوسط .

٢. أن الدرجات المعيارية تجعل وحدات القياس موحدة بعد ان كانت في الدرجات الخام مختلفة في وحدات قياسها (سنتميتر / ثانية ، كفم اعدد المرات..الخ) مما يساعد في اعطاء الفرد المختبر درجة كلية تمثل مجموع ماحصل عليه في كل من مفردات (بطارية) الاختبار.

٣. للدرجة المعيارية القدرة في تحديد مكان الفرد بين مجموعته التي اشتقت لها المعايير (او مجتمع البحث) حيث أن المعيار يحدد معنى للدرجة التي حصل عليها الفرد المختبر.

٧-شروط استخدام معايير الاختبارات:

- أن تكون المعايير حديثة .
- أن تكون (عينة التقنين) ممثلة للمجتمع الأصلي .
- أن تكون المعايير مناسبة للاستخدام ، (الصلاحية) .
- أن تكون الشروط الخاصة بتطبيق الاختبار واضحة .

٨-الأنواع الرئيسية للمعايير:

أ- المعايير القومية / ومثالها : الاختبار الدولي للقدرات المعرفية TIMEES.

ب- المعايير الخاصة بمجموعة خاصة / ومثالها : معايير خاصة بصف دراسي.

ت- المعايير المحلية / ومثالها : المعايير الخاصة بمدنية معينة ، نادي ، مدرسة ، ... الخ .

ث- المعايير المدرسية (معايير الصفوف) / ومثالها : مقارنة أداء صف دراسي بالنسبة للأداء الكلي للمدرسة .

٩-الدرجة المعيارية:

يتزايد استخدام الاختبارات الحديثة للدرجات المعيارية التي تعد أكثر أنواع الدرجات المشتقة قبولا من معظم وجهات النظر ، وعبر الدرجات المعيارية عن المسافة بين درجة الفرد والمتوسط من حيث الانحراف المعياري للتوزيع ، ويمكن الحصول على الدرجات المعيارية بتحويل خطي أو غير خطي للدرجات الخام الأصلية ، وعندما توجد الدرجات المعيارية بوساطة تحويل خطي ، فإنها تحافظ على العلاقات العددية المضبوطة للدرجات الخام الأصلية ، نظرا لأنه يتم حسابها بطرح مقدار ثابت من كل درجة خام ، وقسمة الناتج على مقدار ثابت آخر ، والمقدار النسبي للفروق بين الدرجات المعيارية المشتقة بمثل ذلك

التحويل الخطي يناظر تماما المقدار النسبي للفروق بين الدرجات الخام ، وتتكرر جميع خصائص التوزيع الأصلي للدرجات الخام في توزيع هذه الدرجات المعيارية ، ولهذا السبب فإن أية عمليات حسابية يمكن إجراؤها على الدرجات الخام الأصلية ، يمكن أيضا إجراؤها على الدرجات المعيارية الخطية دون أن تتأثر النتائج .

وتعرف الدرجة المعيارية بانها " الدرجة التي يحتاجها الباحث أو الإحصائي لكي يصف موقع أهمية درجة معينة بالنسبة إلى مجموعة من الدرجات في نفس التوزيع أو لمقارنتها مع درجة أخرى في توزيع أخر من خلال طريقة إحصائية يوجد بها وحدة قياس الدرجة ".

١٠-أنواع الدرجات المعيارية :

هناك أنواع متعددة لاشتقاق المعايير من الدرجة الخام منها (الذائية والتائية والمئينيات والرتبة المئينية والجيمية والتساعية ....الخ) وسنتطرق إلى الأنواع الأكثر استخداماً وهما :

١. الدرجة الذائية (ذ) .

٢. الدرجة التائية (ت) .

٣. الرتبة المينية .

١. الدرجة الذائية (ذ):

هي درجة تمتاز بسهولة حسابها وتفسيرها وفهمها ، وذلك إذا فهمنا معنى الانحراف المعياري للقيم الخام ، وتدل قيم (ذ) على المكانة بالنسبة لتوزيع الدرجات وفقا لوحدات الانحراف المعياري التي تقع أعلى أو أسفل المتوسط الحسابي ، وتستخدم هذه الدرجة كمقياس مفيد في حالة أقرب توزيع البيانات عن الاعتدالية ، وتظهر قيم هذه الدرجة عند حسابها في هيئة أعداد صحيحة وكسور ، وهذه القيم تكون موجبة أو سالبة ، والدرجة المعيارية (ذ) تمتد عادة من (٣-) الى (٣+) انحراف معياري ، ومتوسطها (صفر) ، وانحرافها المعياري (١) ، ويستخدم لحساب الدرجة (ذ) المعادلة التالية :

$$\text{الدرجة المعيارية المحسوبة (ذ)} = \frac{\text{الدرجة الخام } (س) - \text{المتوسط الحسابي } (سَ)}{\text{الانحراف المعياري } (ع)}$$

وتدل الدرجة (ذ) المحسوبة كمياً على عدد وحدات الانحراف المعياري التي تقع أعلى أو أسفل المتوسط الحسابي لدرجات الاختبار ، فالفرد الذي تقع درجته عند متوسط درجة اختبار معين ، فإنه يحصل على درجة (ذ) = (صفر) ، والفرد الذي يحصل على (٠,٥) درجة معيارية أسفل المتوسط الحسابي يحصل على درجة (ذ) = (٠,٥) .

وتفيد الدرجات (ذ) في التعبير عن مركز الفرد بالنسبة لتوزيع ما ، وذلك فيما يتصل بمتوسط وتباين الدرجات الأصلية ، ولذا فإن المتوسط الدرجة (ذ) (صفر) ، وانحرافها المعياري (١) .

٢. الدرجة التائية (ت) :

تعد هذه الدرجات هي الصورة الافضل لتحويل الدرجات الخام التي يحصل عليها عن المتوسط مقسوماً على الانحراف المعياري ، وبذلك فهي تحتفظ بالعلاقات العددية الدقيقة للدرجات الخام الاصلية، وتستعمل الدرجة المعيارية المعدلة (التائية) لسببين مهمين هما:

١-التخلص من الكسور التي قد تنتج من الدرجات المعيارية.

٢-التخلص من الدرجات المعيارية التي قد تكون سالبة

فقد أجرى ثورندايك تعديلاً لتخلص من هاتين المشكلتين ويشمل التعديل على:

- العمل على التخلص من الكسور بضرب الدرجة المعيارية×١٠.
- العمل على التخلص من الاشارة السالبة وذلك بإضافة ٥٠ الى الدرجة المعيارية بعد ضربها×١٠.

وهذه الدرجة تبني على أساس خواص المنحنى الاعتدالى ، وقد سميت هذه الدرجة بالمعيار التائي (ت) ترجمة الحرف T وهو الحرف الأول الذي يبدا به اسم كل العالمين (ثورنديك واتيرمان) وذلك اعترافا بفضلهما في استخدام علم الإحصاء في مجال التربية وعلم النفس ، والدرجة التائية عبارة عن درجة معيارية متوسطها (٥٠) ، وانحرافها المعيارى (١٠) ، وتستخدم عادة في تحويل الدرجات الخام إلى درجات يمكن جمعها ، بغرض مقارنتها وتسهيل تفسيرها ، وتمتاز هذه الدرجة بأنها لا تتضمن قيماً سالبة ، وان الدرجة التائية للدرجات الخام ترتفع كلما ارتفع مستوى هذه الدرجات عن المتوسط الحسابي ، ويستخدم الحساب الدرجة المعيارية (ت) المعادلة التالية :

$$\text{الدرجة المعيارية المحسوبة } (\text{ت}) = \frac{\text{الدرجة الخام}(\text{س}) - \text{المتوسط الحسابي}(\text{سَ})}{(\text{ع})\ \text{الانحراف المعياري}} \times ١٠ + ٥٠$$

٣. الرتبة المئينية (م) :

تستخدم الرتبة المئينية لمقارنة أداء الفرد بأداة غيره من نفس أفراد مجموعة ، لتحديد موقعه بالنسبة للمجموعة ، وتستخدم الرتبة المئينية للإجابة على بعض التساؤلات مثل : (ما هي أعلى درجة في الصف أو الفريق ؟ وما هو ترتيب هذه الدرجة ؟ وهل ترتيب الدرجة هو الثالث أم الرابع) .

وأن الرتبة المئينية سهلة في حسابها وذات دلالة واضحة وشائعة الاستخدام في المقاييس والاختبارات التربوية والنفسية وتوفر جداول معيارية لتفسير الدرجات الخام التي يحصل عليها اي فرد ينتمي لمجموعة مماثلة للعينة المرجعية ، ولحساب الرتب المئينية حسب القانون الاتي:

$$\text{الرتبة المئينية } (\text{م}) = \frac{\text{التكرار المتجمع حتى الدرجة المطلوبة } (\text{ك م})}{\text{عدد العينة } (\text{ن})} \times ١٠٠$$

١١-المستويات :

هي أسس للحكم على الاداء في ضوء الأداء ذاته ، أي هي ما يجب ان يكون عليه الاداء الاختباري وليس ما هو كائن ، وهو ما يعرف بالاختبارات المرجعة الى محك ، فالمحك هذا ليس اساساً خارجياً وانما هو اساس داخلي من الاختبار ذاته.

ويجب أن لا نخلط بين المعيار وبين المستوى القياسي ، إذ إن المعيار يخبرنا عن الأداء الحقيقي للأفراد على الاختبار ، أي أنه يمثل الدرجات الفعلية التي حصل عليها الأفراد في الاختبار ، أما المستوى القياسي فإنه يشير إلى المستوى الذي يجب أن يصله أولئك الأفراد ، فمن الخطأ فهم المعايير على أنها مستويات ذلك لأن المعايير معلومات فمن تدلنا على كيفية الأداء الفعلي للأفراد في حين أن المستويات تدلنا على ما يجب أن يؤديه الأفراد.

أهمية المستويات :

١-أسس داخلية للحكم على الظاهرة .

٢- تأخذ الصورة الكيفية .

٣-تتحدد في ضوء ما يجب أن تكون عليه الظاهرة .

٤- يتم أعدادها على أفراد مدربين ذوي مستويات مثالية ، كما يتم أعدادها بعد التعلم والتدريب والممارسة بهدف التحصيل أو تطوير الصفة أو الخاصية للوصول لدرجات تعكس المستوى الأمثل للصفة أو الخاصية .

مثال : مقياس السلوك العدواني نجد ان اعلى درجة نظرية للمقياس هي (٧٢) واقل درجة نظرية هي (٣٦) ، المطلوب ايجاد خمس مستويات معيارية للمقياس مع تمثيله على منحنى جاوس .

خطوات الحل :

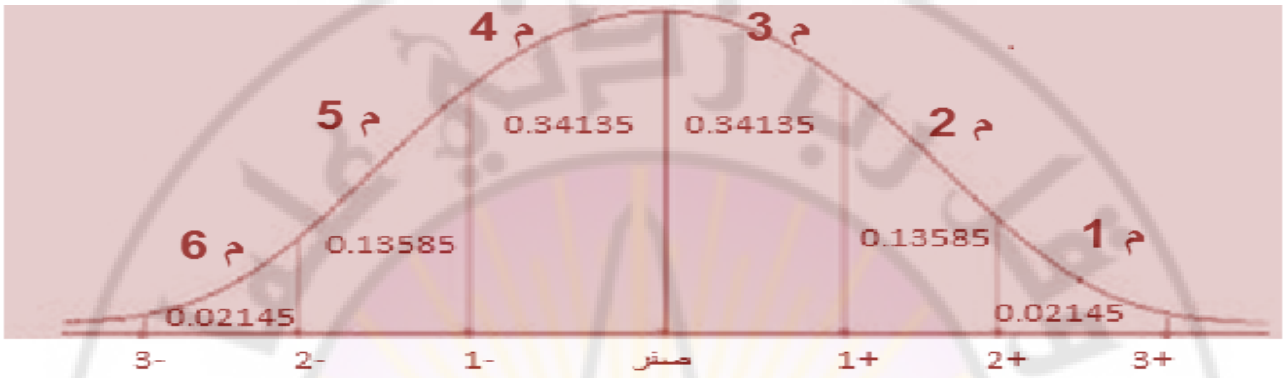
- نستخرج المدى من خلال المعادلة الاتية :

المدى = أعلى درجة نظرية – اقل درجة نظرية = ٧٢ – ٣٦ = ٣٦

- استخراج طول قاعدة كل مستوى من خلال المعادلة الاتية :

طول قاعدة المستوى = $\frac{\text{المدى}}{\text{عدد المستويات المطلوبة}} = \frac{36}{5} = ٧,٢$

- نوضح المستويات على منحنى كاوس والذي يتكون من ستة مستويات .
- ندمج المستوى الثالث والرابع في منحنى كاوس ليصبح المستوى الثالث من (+١) الى (-١) ، ليكون لدينا خمسة مستويات (علماً انه هناك طرق اخرى لعمل المستويات) .



شكل (٣) مستويات المعيارية للمقياس