

المحاضرة الثالثة

معايير تصديق الاختبار (الصدق، الثبات، التمييز)

قبل التطرق للخصائص السيكومترية للمقياس من صدق وثبات يجب أن نتطرق للخطوات السابقة

لهذه المرحلة والتي تعتبر مرحلة التقنين جزئية منها، كما أن تمييز الفقرات يعتبر مرحلة سابقة لها،

لذلك سنفصل في مراحل بناء المقاييس النفسية ونرتبها لتغطية كل محتويات المحاضرة

- خطوات بناء المقاييس النفسية:

تتفق المقاييس النفسية في خطوات إعدادها، لكنها تتفاوت في درجة الاهتمام ببعض الخطوات، فالمقاييس

المقننة التي تنجز من قبل فريق من المختصين في مراكز متخصصة ذات خصائص تختلف في جودتها

عن تلك التي تطور من قبل فرد واحد، لأن عملية إعداد وبناء المقاييس تتطلب وقتاً طويلاً وتحتاج إلى

عدة محاولات من التجريب، كما تتطلب مهارة كبيرة من طرف المختص ومعرفة جيدة بالميدان المراد

قياسه، وسنقوم من خلال هذا العنصر بالتطرق بالترتيب إلى مراحل بناء المقاييس النفسية، والتي تتفق

إلى حد كبير مع خطوات تكييف المقاييس المنجزة سابقاً على بيئات غير التي أعدت لها (وهو هدف بحثنا

الحالي)، مع توضيح مكامن الاختلاف.

1- تحديد الأهداف:

قبل البدء ببناء مقياس ما يجب أن نعرفه بالتحديد ما المراد قياسه بهذا المقياس، وما الغرض المراد الوصول إليه من خلال تطبيق هذا المقياس، ومن هم الأفراد المراد تطبيق المقياس عليهم، ومن أي سن، ومن أي مستوى تعليمي، وما هي الطريقة التي يراد بها تطبيق المقياس، وهل ستلقى أسئلته شفهيًا أو أنها تحريرية، وهل سيطبق بطريقة جماعية أو بطريقة فردية، وكيف سيعبر المفحوص عن استجاباته.

وتعتبر هذه الخطوة هي الخطوة الأولى في بناء المقاييس التي يجب أن يهتم بها الباحث، فتحديد الهدف الذي يسعى إليه الباحث من خلال إنشاء المقياس يوضح الخطوط العريضة لبناء المقياس. ويشير ثرونديك Thorndike و"هايقن Hagen" (1977) في هذا الصدد إلى أن الأهداف الجيدة يجب أن

تتوفر على الخصائص التالية: (مقدم عبد الحفيظ، 2011)

- أن يوضع السلوك المراد قياسه في عبارات.
- أن تبدأ بجملة فعلية تشير إلى السلوك المراد إبرازه.
- أن يوضع السلوك الملاحظ في عبارات.
- أن توضع بطريقة محددة ودقيقة، فالهدف يجب أن لا يكون غامضا يحتمل أكثر من معنى.
- أن توضع بطريقة وحدوية، أي أن كل هدف يرتبط بعملية واحدة فقط.
- أن تتميز بدرجة ملائمة من العمومية أي أنه لا يجب أن تكون عامة جدا ولا محصورة أو محدودة جدا.

- أن تشمل النواتج المتوقعة من المقياس.

ويفيد تحديد أهداف المقياس في صورة إجرائية دقيقة في توجيه معده إلى الأهداف التي يرغب في قياسها عن طريق مقياسه؛ فعلى الرغم من أننا نبدأ من مفهوم علمي مقبول يمكن اختباره بواسطة مقياس معين، إلا أن تحديد الهدف من المقياس يتطلب مزيداً من التفصيل، حيث أن المفاهيم تختلف في طبيعتها، فقد يكون الهدف من مقياس ما هو تحديد مستويات الأفراد وفقاً لمقدار امتلاكهم لخاصية ما، وقد يكون الهدف هو التمييز بين الأفراد وفقاً لترتيبهم على الخاصية أو غيرهما من الأهداف المتعلقة بالخاصية والتي تختلف في طبيعتها، لذلك على الباحث أن يحدد أهدافه بدقة كخطوة أولى في بناء مقياسه (بشرى إسماعيل، 2004، ص66). وتنقسم تلك الأهداف إلى أهداف عامة وأهداف خاصة (محسن لطفي أحمد، 2007).

2- محتوى المقياس:

في ضوء الهدف الذي وضع المقياس من أجله يمكن أن يتحدد دور المقياس أو مادته أو مجال أسئلته ويمكن في ضوء ذلك توزيع أسئلته على أبعاده أو جوانبه المتعددة (بشرى إسماعيل، 2004، ص66)؛ حيث أنه لا بد للخاصية المقاسة أن تستند إلى أساس نظري يبرر مشروعيتها تناولها ويعرفها، وقد يكون المقياس معد في الأصل للتأكد من مدى جدوى النظرية التي تفسر الخاصية المقاسة، والنتيجة المستخلصة قد تفيد النظرية أو تعديلها، كما هو الحال بالنسبة لمقاييس القدرات الإبداعية التي صممت من أجل التحقق من

نظرية "جيفورد Guilford" في الإبداع (Guilford J, 1954)، فضلا على أن الاطلاع على الأطر النظرية للخاصية المستهدفة بالقياس يزود الباحث بالأهمية النسبية للعناصر الفرعية المكونة للخاصية، ومن ثم تمثل العناصر في بنود المقياس فيما بعد وفقا لتلك الأهمية وتحدد تعليمات تطبيقه، ويتضمن المقياس عموما ما يلي:

1-2 بنود المقياس: حيث ينبغي أن تغطي أسئلة المقياس أبعاده المختلفة، مع مراعاة شرط التوازن بين عدد أسئلة كل مجال أو بعد من أبعاد المقياس، وأن تكون الصياغة اللغوية واضحة؛ وهناك عدة أنواع للأسئلة أو المفردات التي يتألف منها المقياس منها (بشرى إسماعيل، 2004):

- **أسئلة الاختيار من متعدد:** حيث يتكون السؤال من مشكلة في صورة سؤال مباشر أو عبارة ناقصة مع إعطاء قائمة من الحلول المقترحة تسمى البدائل الاختيارية ويطلب من المفحوص قراءة السؤال وقائمة البدائل وانتقاء البديل الصحيح أو الأفضل.

- **أسئلة المزوجة:** وتتألف من عمودين متوازيين يحتوي كل منهما على مجموعة من العبارات أو الرموز والكلمات، أحدهما يسمى المقدمات والثاني يسمى الاستجابات وعلى المفحوص أن يزوج بين كل عنصر في قائمة المقدمات وما يناظره في قائمة الاستجابات.

- **أسئلة البديلين:** وتتطلب اختيار استجابة واحدة من بديلين، مثل (نعم، لا).

- الأسئلة التفسيرية: وفيها يختار المفحوص من متعدد، حيث ظهر هذا النوع من المقاييس للتغلب على بعض مشكلات الاختيار من بديلين، خاصة مشكلة التخمين، مع ترك مجال يفسر فيه المفحوص اختياره.

- أسئلة الترتيب: وفيها يقوم المفحوص بإعادة ترتيب عناصر أو خطوات أو مراحل أو أحداث أو إجراءات أو تواريخ في تسلسل طبيعي منطقي.

- أسئلة الإجابة القصيرة (الاستدعاء): ويتطلب هذا النوع من الأسئلة أن ينتج المفحوص استجاباته وليس مجرد التعرف عليها ولذلك تسمى أحيانا أسئلة التكميل.

- أسئلة الإجابة الطويلة (المقال): رغم الاستخدام الواسع لأنماط الأسئلة السابقة الذكر، إلا أن هناك مواقف لا تصلح لها إلا الأسئلة التي تتيح للمفحوص مجالاً للتعبير.

- الأسئلة العلمية: تستخدم الأسئلة العلمية كوسائل موضوعية لتقدير الكفاءة التي تؤدي بها أعمال المهارة، أي أن أسئلتها تتعلق بمجال معرفي معين، تهدف إلى قياس مدى إحاطة المفحوص بهذه المعارف العلمية في هذا المجال.

- الأسئلة الشفهية: وهي أسئلة تقدم للمفحوص بطريقة شفهية مع تقديم بدائل الإجابة، ويدون الفاحص الاستجابات، وتستخدم عادة مع الذين لا يجيدون القراءة، أو فئة معينة لديها مشاكل عضوية أو عقلية.

2-2 نظام الأسئلة وترتيبها: يلاحظ في ترتيب البنود تنوع مجالاتها بحيث لا نضع أسئلة مجال واحد وراء بعضها البعض، وكذا سهولتها وصعوبتها (في حالة قياس القدرات والاستعدادات) ويتطلب ذلك

تحليلاً كفيًا للتأكد من توافر شروط الصياغة الجيدة للبنود، وكذلك تحليلاً كمياً يهدف لتقدير مستوى صعوبة البنود واختيار البنود ذات الصعوبة المناسبة لأهداف الاختبار، ويوضح هذا النوع الأخير من التحليل علاقة صدق البنود بثباتها، إذ أن صدق وثبات المقياس يعتمد أساساً على خصائص البنود.

وهناك عدد من القواعد العامة من حيث الشكل والمضمون في كتابة البنود تساعد إلى الوصول إلى

مستوى صدق جيد للمقياس، هي ما يلي (بشرى إسماعيل، 2004، ص 69، 68):

■ قواعد الشكل:

- يتعين أن تتسم بنود المقياس بسهولة القراءة وبساطة التعبير.
- يجب تجنب التعبيرات الخادعة والمضللة التي يمكن أن تؤدي إلى نتائج متناقضة.
- يجب أن يكون البند وحدة اختبار صغيرة، سواء كانت هذه الوحدة في شكل معلومة أو سؤال
- أن تكون واضحة أو غير غامضة ولا تحتمل أكثر من تفسير من الأفراد المختلفين.

■ قواعد المضمون:

- يجب أن يتعلق البند بجانب هام من جوانب السلوك المقاس ليس بالأعراض أو بالجوانب الهامشية شديدة التغير وغير المميزة.

- يجب أن يكون كل بند مستقلاً عن الآخر.

- يجب أن يكون للبند الواحد إجابة واحدة فقط صحيحة أو أفضل إجابة مقبولة يتفق عليها الخبراء في حالة قياس القدرات والاستعدادات؛ حيث يجب كقاعدة عامة أن لا يتضمن البند أكثر من فكرة/معلومة واحدة فقط، ومن ثم فإن التصميم الغير الجيد للبنود سوف يؤثر لاحقا بإحداث خلل واضح في النتائج التي نخرج بها منه، وفي تفسير الدرجات أيضا.

3-2-3 صياغة تعليمات المقياس: تنقسم تعليمات المقياس إلى قسمين رئيسيين هما (محسن أحمد، 2007، ص5) :

▪ **تعليمات المطبقين:** وهم الذين يقومون بتطبيق المقياس، وتتضمن شرحا وافيا للمقياس والخاصية التي يتم قياسها، وإجراءات التطبيق بالتفصيل، والزمن، وطريقة تسجيل الاستجابات، والمواقف التي يحتمل مواجهتها أثناء التطبيق، وحدود الشرح والتوضيح المسموح به للمفحوصين.

▪ **تعليمات المفحوصين:** وتتضمن عدة محاور منها:

- فكرة مبسطة عن المقياس والهدف من وراء تطبيقه.

- طريقة الاستجابة والزمن المحدد إن وجد.

- تقديم بعض النماذج المحلولة إن تتطلب الأمر.

- يجب على مصمم المقياس أن يراعي خصائص العينة أثناء صياغة التعليمات وما إذا كان سيستخدم

اللغة الفصحى أم الدارجة.

ولا بد من إعداد مفاتيح لأسئلة المقياس والتي يتم من خلالها إعطاء الدرجة لكل التي تقابل كل استجابة.

3- تحليل البنود:

بعد أن ينتهي مُعدّ المقياس النفسي من إنجاز المراحل الأساسية لتصميم المقياس سابقة الذكر، يصل إلى مرحلة تحليل البنود؛ حيث يستخدم هذا الأخير كإجراء لعزل أنواع معينة من البنود أو حذفها، حيث يمكن تحليل المفردات كفيها في ضوء المحتوى والشكل، كما يمكن تحليلها كميًا في ضوء خصائصها الإحصائية. ويشمل التحليل الكيفي للمفردات ما يسمى بصدق المحتوى، كما يشمل تقويم المفردات في ضوء فعاليتها من حيث صياغتها وصلاحتها، حيث تراجع المفردات من حيث النوع لمعرفة مدى توفر الخصائص اللازمة سواء كانت من نوع الاختيار من متعدد، أو الاختيار من بديلين، أو التكملة، أو التفسير، أو إعادة الترتيب... أما التحليل الكمي للمفردات فيتضمن على وجه الخصوص تحديد سهولة (شيوخ) المفردات، وتحديد صدقها بالإضافة إلى تحديد الثبات، وقد نحتاج إلى هذه المعاملات في ذاتها إذا كانت مقاييسها من نوع المفردات المتصلة، كما أنها لازمة للمقاييس التي تتألف من مجموعات من المفردات (مقاييس فرعية) لأن صدق المقياس وثباته يعتمد إلى حد كبير على خصائص مفرداته (صالح الداهري، 2001)، ويمكن توضيح كل من التحليل الكيفي والكمي لبنود المقياس كالتالي:

3-1 تحليل البنود كفيًا (قواعد كتابة البنود):

هناك قواعد لكتابة البنود (مذكورة أعلاه) تساهم في الوصول إلى مستوى صدق جيد للمقياس، وتتضمن

هذه القواعد التحليل الكيفي للبنود شكلا ومضمونا.

2-3 تحليل البنود كميا: ويتم ذلك من خلال:

▪ معامل الصعوبة: Difficalty Fndex

إن تحديد مستوى صعوبة كل مفردة من مفردات المقياس يعد ضروريا، حيث تلعب دورا مهما في المقياس وتؤثر في استجابات الأفراد عن مفرداتها، فالمفردات التي تشتمل عليها هذه المقاييس ينبغي أن تميز تمييزا دقيقا بين مستويات السمة المراد قياسها، فالمفردة التي يفهمها جميع أفراد العينة، أو التي لا يستطيع أن يفهما أحد لا تفيد في الكشف عن الفروق بينهم فيما يقيسه المقياس، وقد بينت الدراسات النفسية أن الاختبار يمكن أن يميز إلى أقصى حد ممكن بين أفراد العينة إذا كان متوسط مستوى صعوبة المفردات التي تشتمل عليها 0.05 تقريبا، أي يستطيع أن يجيب عليه 50% منهم على كل مفردة من مفرداتها، وقيمة معامل الصعوبة تؤثر على تمييز المفردات بشكل مباشر حيث تحصل على أعلى قيمة للتمييز عندما يكون معامل الصعوبة حوالي 0.05 ويقل التمييز بابتعاد معامل الصعوبة عن هذه القيمة، أي أن البنود السهلة جدا أو الصعبة جدا لا تسهم بقدر كبير في التمييز بين الأفراد وتوافر السمة لديهم، وبالتالي يتعين مراجعتها لتنتقيحها أو حذفها عند الضرورة، ويتم إيجاد معامل الصعوبة لكل فقرة من فقرات

الاختبار، عن طريق معرفة نسبة من أجابوا إجابة صحيحة من المفحوصين على كل فقرة (بوسالم عبد العزيز، 2014).

ويحسب معامل سهولة المفردات بصفة عامة بالمعادلات التالية:(صالح الداھري، 2001)

$$S = \frac{ص}{ص+خ}$$

حيث أن:

س: معامل سهولة المفردة

ص عدد الإجابات الصحيحة.

خ: عدد الإجابات الخاطئة

ويمكن أن تصبح المعادلة السابقة أكثر عمومية حين تستخدم لحساب معاملات شيوع المفردات:

$$ش = \frac{أ}{أ + ب}$$

حيث أن:

ش: معامل شيوع المفردة

أ: عدد الذين أجابوا على المفردة في الاتجاه (أ)، وليكن (نعم) في اختبارات الشخصية، أو صواب في

اختبارات القدرات ...

ب: عدد الأفراد الذين أجابوا على المفردة في الاتجاه (ب) أي عكس الاتجاه (أ)

▪ معامل التمييز : Discrimination Index

من الخصائص الأخرى والمهمة التي ينبغي أن تتوفر في مفردات المقياس التمييز، ونعني بذلك مدى إمكانية قياس الفروق بواسطة مفردات المقياس، فإذا اخترنا مجموعة من الأفراد باختبار مهني مقنن في مجال معين، ووجدنا أن عدد الأفراد الأقوياء الذين أجابوا إجابة صحيحة عن إحدى المفردات قليل، فإن هذه المفردة لا تصلح لقياس الفروق الفردية في هذا المجال المهني، حيث أنها لم تميز مطلقاً بين مجموعتي الأفراد، وينطبق كل ما سبق على أي نوع من أنواع الاختبارات والمقاييس النفسية والتربوية والعقلية وحتى الشخصية (بوسالم عبد العزيز، 2014).

▪ صدق المفردات (البنود):

توجد عدة طرق لحساب صدق المفردات أهمها، طريقة حساب معامل الارتباط بين درجات المفردة ودرجات المقياس الكلي (كمحك داخلي)، أو درجات أي محك خارجي، أو طريقة المقارنة الطرفية، وذلك بتقسيم درجات المحك إلى مستويين ثم المقارنة بين أداء أفراد المجموعتين في كل من المستويين في المفردة، وأفضل مجموعتين للمقارنة بينهما كما حددها فلانجان مجموعة (27%) العليا في المحك والـ (27%) الدنيا فيها، وحساب معاملي السهولة في المستويين واستخدام جداول فلانجان التي تحدد معاملات الارتباط بين المفردة والمقياس (صالح الدايري، 2001).

ثبات المفردات (البنود): تصلح لحساب ثبات المفردات الطرق المعتادة لحساب ثبات الاختبار ككل،

والتي من بينها إعادة التطبيق، إلا أن من أفضل الطرق المستخدمة في حساب ثبات المفردات، ما يسمى

بالاحتمال المتوالي وهي من اقتراح (جتمان) ومعادلتها مايلي:

$$r = \frac{n}{1-n} \left(\frac{1}{n} - n \right)$$

حيث:

ر: معامل الثبات

ن: عدد الاحتمالات الاختيارية للمفردة

ل: الاحتمال المتوالي، أي أكبر تكرار نسبي لأي احتمال اختياري من الاحتمالات التي تحتوي عليها

المفردة. (صالح الداهري، 2001).

وبصفة عامة يمكن القول أن التأكد المسبق من الخصائص السيكومترية للبنود يساهم بشكل كبير في

توافرها في المقياس ككل.

1- شروط الاختبار الجيد:

يقصد بالخصائص السيكومترية للاختبار تلك الصفات الضرورية والمتعلقة بمدى فعالية بنود الاختبار،

وكذلك بالصدق والثبات وما يرتبط بهما من معاملات تمييز ومستويات السهولة والصعوبة ومعايير

تفسير النتائج، والتي يتم التحقق منها بعد تطبيق الاختبار تطبيقاً تجريبياً على عينة ممثلة للمجتمع تسمى بعينة التقنين، وتعتمد جودة الاختبار وموضوعيته على مدى توافر درجات مناسبة لهذه الخصائص (بوسالم عبد العزيز، 2014)، وسوف نتطرق من خلال هذا العنصر لكل من صدق درجات الاختبار، ثبات درجات الاختبار، تحليل بنود الاختبار، استخراج المعايير.

1-4 الصدق: تعريفه، أنواعه، طرق حسابه، العوامل المؤثرة فيه

▪ تعريف الصدق:

يعتبر الصدق من الخصائص المهمة التي يجب الاهتمام بها في بناء المقاييس النفسية، فعندما يريد الباحث تصميم اختبار معين فلا بد أن تكون هناك ظاهرة سلوكية معينة يقيسها الاختبار (ظاهرة التكيف الاجتماعي، أو القلق، أو الاحتراق النفسي، أو الذكاء، أو التحصيل الدراسي...)، حيث أن الباحث بعد بناء اختباره يجب أن يتأكد بأن هذا الاختبار يقيس فعلاً الظاهرة التي صمم لقياسها عندئذ يعتبر صادقاً إذا كان قادراً على قياس الاستعداد الدراسي ويكون غير صادق إذا كان يقيس ظاهرة سلوكية أخرى (سوسن شاكر، 2010).

ويعرف الصدق على أنه الاتفاق بين درجات الاختبار أو الاختبار والخاصية التي يعتقد أنه يقيسها، من خلال الإجابة على السؤال: هل يقيس الاختبار ما يفترض أن يقيسه (بركات حسن، 2008).

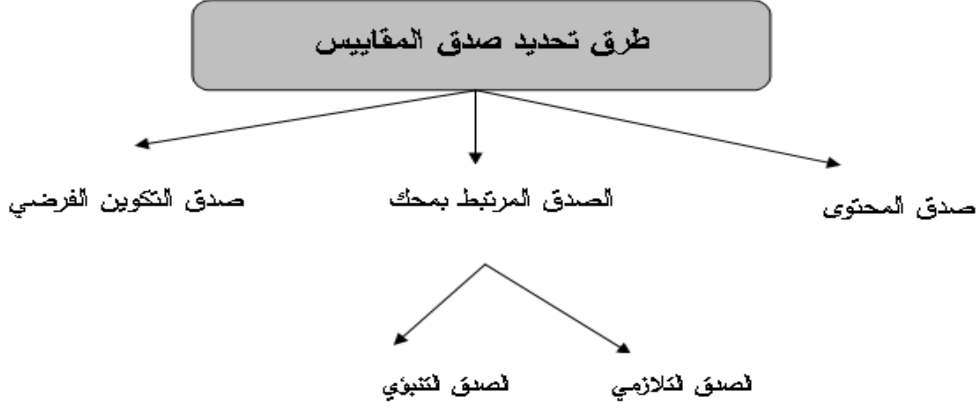
كما يشير الصدق إلى الاستدلالات الخاصة التي نخرج بها من درجات الاختبار من حيث مناسبتها ومعناها وفائدتها، وتحقيق صدق الاختبار معناه تجميع الأدلة التي تؤيد مثل هذه الاستدلالات (رجاء محمود أبو علام، 2005).

ورغم أننا ننسب الصدق إلى أدوات القياس إلا أن الصدق في الواقع يشير إلى الدرجة التي يمكن بها تفسير نتائج الاختبار، أي أننا في الواقع نتكلم عن صدق التفسيرات التي نخرج بها من النتائج، والتي يتم التوصل إليها باستخدام عدة طرق.

▪ طرق تحديد صدق درجات الاختبار:

منذ نشأة مجال القياس النفسي، بذلت الكثير من المحاولات لتقديم أساليب تقوم بوظيفة تقدير وحساب الصدق، وأسفرت المحاولات العديدة لعلماء النفس عن ظهور فئات وتصنيفات وأساليب متعددة للصدق، وقد تم توحيد هذه الجهود في اجتماع اللجنة المشتركة للقياس النفسي التي تشكلت من خلال اندماج ثلاث هيئات هي: (رابطة البحوث التربوية الأمريكية (AERA)، رابطة علم النفس الأمريكية (APA)، المجلس الوطني للقياس التربوي (NCME)) عام 1985، حيث تم خلال اجتماعها الأول الوصول إلى تعريف محدد للصدق، والذي عرف بأنه: "الصدق هو دليل على الاستنتاجات المستخلصة حول الدرجة على الاختبار، وتوجد ثلاث أنواع من الأدلة هي: أدلة مرتبطة بالمفهوم، أدلة مرتبطة بمحك، أدلة مرتبطة بمحتوى" (بركات حسن، 2008).

ويمكن توضيح هذا التقسيم لطرق الحصول على صدق درجات الاختبار كالتالي:



الشكل(..): طرق تحديد صدق المقاييس.

• **صدق المحتوى (Content Validity):**

يقصد بصدق المحتوى مدى توافر جوانب السمة في بنود الاختبار، وهذا النوع من الصدق يتطلب تحليلاً منطقياً لبنود الاختبار، وذلك بهدف تحديد الوظائف والجوانب والمستويات الممثلة في بنوده مع تحديد نسبة كل منها في الاختبار ككل، أي أن تحديد الصدق بالطريقة المرتبطة بالمحتوى تعكس مدى تمثيل مفردات الاختبار للمجال الذي نريد قياسه، ويمكن على هذا الأساس القول بأن الطريقة المرتبطة بالمحتوى تعتمد على تمثيل مفردات الاختبار تمثيلاً سليماً للمجال الذي نريد قياسه (رجاء علام، 2005)، لذلك فإن

تحقيق صدق الاختبار بطريقة المحتوى يتطلب القيام بخطوتين هما:

- تحديد المجال الذي نريد قياسه تحديداً واضحاً، مع تحديد عناصره.

- بناء مجموعة ممثلة من المفردات أو الأسئلة لهذا المجال.

والصدق بالطريقة المرتبطة بالمحتوى يعكس مدى تمثيل مفردات الاختبار للمجال الذي نريد قياسه؛ فارتفاع درجة صدق محتوى اختبار معين مؤشر على درجة تمثيله للمحتوى، ومنه تعد أدلة الصدق المرتبطة بالمحتوى مهمة بوجه خاص عندما نريد أن نصف كيفية أداء الفرد في مجال من المجالات التي يمثلها الاختبار، بمعنى أن الطريقة المرتبطة بالمحتوى تعتمد على مدى تمثيل بنود الاختبار تمثيلا سليما للمجال الذي نريد قياسه، لذلك فإن تحقيق صدق الاختبار بهذه الطريقة يتطلب القيام بالخطوات التالية:

- تحديد مجال المحتوى الذي نريد قياسه تحديدا واضحا ودقيقا بجميع مكوناته وعناصره.
- إعداد جدول المواصفات وتحديد حجم (عدد) البنود ومستوياتها.
- صياغة مجموعة من البنود الممثلة لهذا المجال.
- تقديم المحتوى والبنود إلى مجموعة من الخبراء ليقوموا بفحص الفقرات منطقيا وتقدير مدى صلاحيتها أو مدى تمثيلها للمحتوى المراد قياسه (بوسالم عبد العزيز، 2014، ص64).
- ويقدر صدق المضمون أو المحتوى لاختبار ما بالتحليل المنطقي لمحتوى الاختبار ومطابقتها مع محتويات الجانب المقاس، أو عن طريق استشارة الخبراء

✓ **طريقة استشارة الخبراء:** يقدر صدق المحتوى بإجراء مراجعة منظمة لمجموع البنود التي يتضمنها الاختبار من قبل الخبراء لتقدير مدى تمثيلها للخاصية التي أعد الاختبار لقياسها، وبما أن أي خاصية تتحدد من خلال تعريفها، فإن صدق المحتوى يصبح دالة لتعريف الخاصية موضوع القياس، والطريقة

الشائعة لتقدير صدق محتوى الاختبار، هي استشارة مجموعة من الخبراء (المحكمين) الذين يكونون من ذوي الخبرة والكفاءة في موضوع الاختبار، أو في مجاله، وعلى مُعدِّ الاختبار أن يقدِّم اختباره في استمارة تتضمن ثلاثة أعمدة:

العمود الأول: ويتضمن الأهداف.

العمود الثاني: ويتضمن الأسئلة التي تقيس كل هدف.

العمود الثالث: ويتضمن مكانا فارغا يضع الخبراء فيه ملاحظاتهم.

حيث يقرأ الخبير بأسلوب أفقي الهدف ثم السؤال الذي وُضع لقياسه ثم يضع ملاحظته في المكان الفارغ الخاص بذلك أمام السؤال، مع وضع الملاحظات (الأحكام) (يقيس أو لا يقيس)، أو تكون هاتان الكلمتان مكتوبتين في العمود الثالث بأعلى الاستمارة، ويقوم الخبير بوضع علامة معينة في المكان الخاص لذلك (معمرية بشير، 2012).

وبعد أن يستعيد مُعدِّ الاختبار الاستمارات من الخبراء (المحكمين)، يفرِّغ تكرارات تقديراتهم لكل بند على حدى، ثم يحوّلها إلى نسب مئوية بتقسيم عدد الخبراء الذين وافقوا على صلاحية السؤال أو البند على العدد الكلي للخبراء، وضرب الناتج في 100، ثم يختار البنود التي حصلت على نسبة تقديرات تساوي 80 % فأكثر، لأن أخذه بهذه النسبة المرتفعة يجعله يحصل على الشعور بالارتياح والثقة من حيث

صدق المحتوى، أما إذا كانت نسبة الاتفاق أقل من 80 %، فينبغي مراجعة البند بحذفه أو تعديله (معمرية بشير، 2012).

وتجدر الإشارة إلى أنه ينبغي عدم الاكتفاء بتقدير صدق المقاييس النفسية بطريقة استشارة الخبراء، والاعتماد أكثر على الطرق الإحصائية التي سوف نذكرها لاحقاً، وعندما يريد مُعدُّ الاختبار أن يقدّر صدق اختباره عن طريق استشارة الخبراء، لا بد أن يكون صارماً بالتزامه بالشروط الموضوعية العلمية والفنية لوضع أدوات القياس النفسي من عدة نواحي كما يلي:

✓ من ناحية معد الاختبار أن يكون ملماً علمياً وفنياً بشروط إنجاز أدوات القياس النفسي، وملماً علمياً بالإطار النظري للمتغير لموضوع القياس.

✓ ومن ناحية المصادر التي يعتمد عليها لوضع البنود، ينبغي أن تكون من المراجع المتخصصة في المتغير الذي يقيسه الاختبار، فيعتمد على الكتب المتخصصة والبحوث التي أنجزت حول المتغير والمقاييس الأخرى التي تم تصميمها لقياس المتغير.

✓ أما من ناحية الخبراء (المحكمين)، فينبغي أن يكون عددهم كبيراً (لا يقل عن 30 خبيراً) ومن الذين يتوفرون على خبرات علمية وفنية في القياس النفسي وفي المتغير لموضوع القياس.

وبهذه الطريقة يمكن أن نثق في صدق المحتوى للمقاييس النفسية.

• الصدق المرتبط بالمحك: (Criterion-related validity)

يطلق أحيانا على الصدق المرتبط بمحك اسم الصدق الواقعي أو العملي، ويقصد به مجموعة الإجراءات التي نتمكن من خلالها من حساب الارتباط بين درجات الاختبار وبين محك خارجي مستقل هو السلوك نفسه أو النشاط الذي يتناوله الاختبار أو الاختبار بالقياس (محمد يونس، 2008).

ويشير المحك إلى اختبار موضوعي مستقل عن الاختبار، نتحقق من خلاله من صدق اختبار آخر، أو هو ميزان نحدد به مدى صلاحية اختبار معين، فالصدق المرتبط بالمحك أو الميزان يدل على مدى قدرة الاختبار على التنبؤ بسلوك الأفراد في مواقف محددة أو تشخيص هذا السلوك (بوسالم عبد العزيز، 2014).

من هنا نستطيع القول أن الصدق المرتبط بالمحك يستند إلى الأسلوب الإمبريقي في دراسة العلاقة بين درجات الاختبار المطبق ومقاييس أخرى تعد بمثابة مئينات ودرجات مقاييس خارجية مستقلة تعد بمثابة محكات، ويمكن الاعتماد على الفاصل الزمني بين تطبيق المحك وتطبيق الاختبار وأهداف القياس لنحدد نوعين من الصدق المرتبط بالمحك وهما: التلازمي، والتنبؤي.

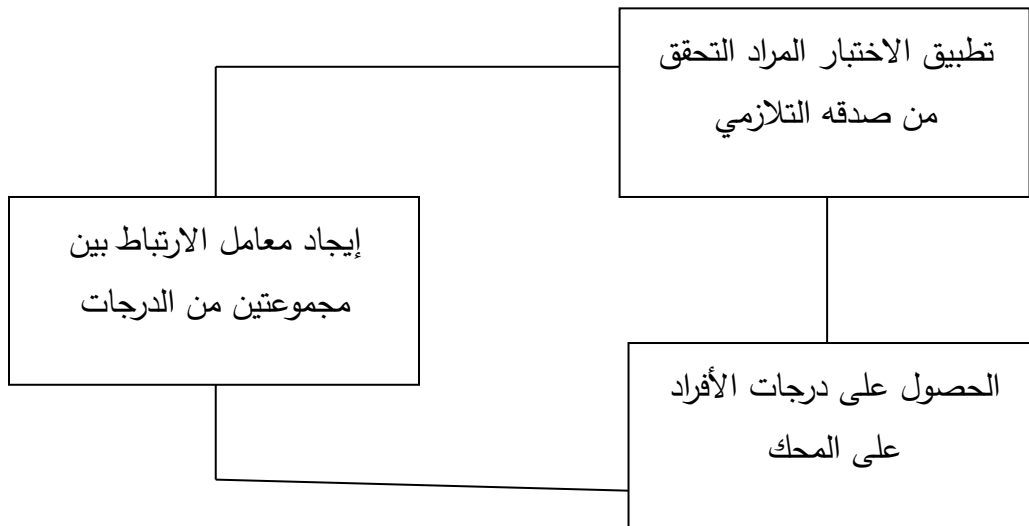
▪ الصدق التلازمي (Concurrent Validity):

الصدق التلازمي يقترن بدرجة اقتران درجات الاختبار مع درجات الاختبار المحك في الوقت نفسه (متلازمين) (صلاح علام، 2000، ص208)؛ كما يشير الصدق التلازمي إلى مدى اقتران تباين درجات الاختبار بتباين درجات محك آخر يطبق في الوقت نفسه على عينة نفسها، وهناك أنواع عدة من المحكات يمكن استخدامها في تقدير هذا الصدق، ومن هذه المحكات الشائعة الاستخدام نجد المقاييس

النفسية التي قام ببنائها باحثون آخرون وطبقت سابقا وتم التحقق من صدقها، شرط أن تكون متعلقة بنفس السمة أو سمة مرتبطة بها وفق أدلة نظرية وإمبريقية نفسها، كما يمكن الاعتماد على التقديرات الخارجية للأفراد الخبراء في مجال السلوك أو المجال المهني أو الدراسي الذي يقيسه الاختبار كمؤشر للصدق التلازمي، أو يمكن الاعتماد على التقديرات الذاتية التي تستخدم في حساب صدق المقاييس الشخصية (بوسالم عبد العزيز، 2014، ص66).

ويستخدم الصدق التلازمي للتحقق من صدق الاختبار انطلاقا من درجات الفرد في الاختبار، والتي تحاول وصف الوضع الراهن للفرد، وفيه يتم الحصول على معلومات المحك في نفس الوقت تقريبا الذي يتم فيه الحصول على درجات الاختبار (محمود عمر وآخرون، 2010).

ويمكن توضيح خطوات حساب معامل الصدق التلازمي في الشكل التالي:



الشكل (..): خطوات تقدير الصدق التلازمي للاختبار.

المصدر: صلاح الدين محمود علام، 2000

▪ الصدق التنبؤي (Predictive Validity):

يشير الصدق التنبؤي إلى مدى الصحة التي نستطيع أن نقدم بها تخمينات معينة لدى الفرد من خلال خاصية أخرى، بمعنى المدى الذي نستطيع أن نقيس به خاصية معينة من خلال اختبار أو اختبار يفترض أنه يقيس هذه الخاصية، والسمة المميزة للصدق التنبؤي هو أنه يفترض توافر قياس مستقل للخاصية هو المحك، وعادة ما نلجأ إلى حساب معامل الارتباط بين الاختبار أو الاختبار والمحك الخارجي، ويتم حساب معامل الارتباط في ضوء طبيعة الدرجات أو التقديرات على كل من الاختبار والمحك، ويعد معامل الارتباط هنا معامل صدق للاختبار (محمد يونس، 2008).

حيث نقوم بإيجاد العلاقة بين الدرجات على الاختبار والدرجات على الاختبار المحك الذي يطبق بعد إجراء الاختبار محل الاختبار، وبعد جمع البيانات عن المحك وحساب الدرجات على الاختبار نقوم

بإيجاد العلاقة بينهما بأحد الطرق الثلاثة وهي: (بوسالم عبد العزيز، 2014)

- طريقة النسب المئوية.

- طريقة المتوسطات.

- طريقة معامل الارتباط لبيرسون.

وهي أشهر الطرق وأكثرها استعمالاً من طرف الباحثين للتحقق من الصدق المحكي لأي اختبار حيث تعتبر قيمة معامل الارتباط ما بين نتائج تطبيق الاختبار ونتائج تطبيق المحك بمثابة معامل صدق وتفسر على أساس اقترابها أو ابتعادها عن قيمة الواحد الصحيح (+1)، وعلى الباحث عند حساب

الصدق التنبؤي أن يأخذ بعين الاعتبار العوامل التالية: (بوسالم عبد العزيز، 2014)

- ضرورة تحديد السلوك الذي نريد التنبؤ به من خلال الاختبار، إذ أن الاختبار الذي يتنبأ بسلوك معين بطريقة جيدة لا يعني بالضرورة أن الاختبار يصلح للتنبؤ بكل السلوكيات المشابهة أو المتقاربة من حيث المؤشرات السلوكية الدالة عليها.

- ضرورة التحكم في المتغيرات الدخيلة التي تواجه تحديد معيار أو محك السلوك الذي يرتبط به الاختبار، مثل عدم موضوعية تسجيل السلوك، ووجود عوامل متعددة تؤثر على هذا السلوك أثناء حدوثه ويصعب ضبطها أو التحكم فيها من طرف الباحث.

- يجب أن تكون المعايير أو المحكات السلوكية المستخدمة شاملة لتغطي جوانب السلوك الذي ادعى الاختبار أنه يتنبأ بها.

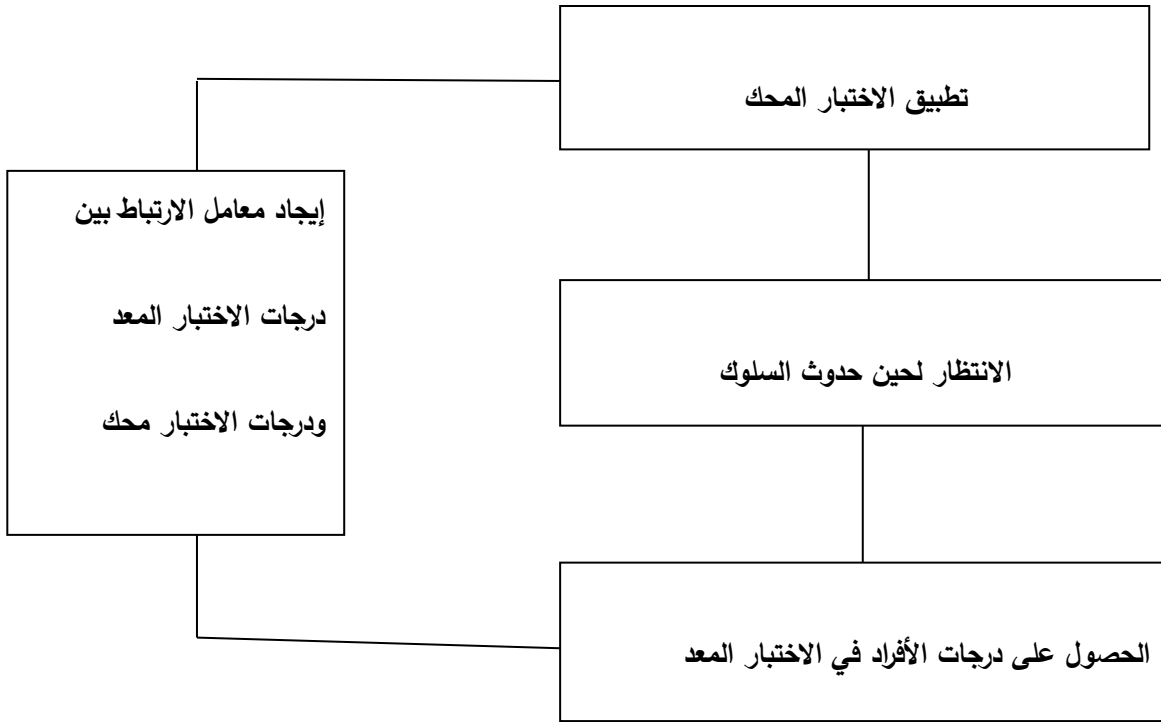
- ضرورة توفير درجات مقبولة من الثبات في الاختبار الذي نعتمد عليه كمعيار أو محك في حالة إذا لم يكن المحك سلوك واقعيًا.

ونشير إلى أن الفرق الأساس بين الصدق التلازمي والصدق التنبؤي هو أنه في الصدق التنبؤي نبحت عن مدى اتفاق درجات الأفراد على الاختبار مع أدائهم في المستقبل، إذ لا بد أن تمضي فترة زمنية بعد إجراء هذا الاختبار وانخراط المختبرين في المهنة أو الدراسة أو نمط علاج أو طريقة أو مجال من مجالات الحياة بصفة عامة ولمدة كافية، بينما في الصدق التلازمي الحصول على درجات الأفراد في الاختبار يكون متزامنا أو متقاربا جدا من الحصول على الدرجات على المحك؛ أي أن المحك في الصدق التلازمي يكون أنيا، أما الصدق التنبؤي فيكون المحك في المستقبل (محمد يونس، 2008)

وتتعدد طرق تقدير الصدق التنبؤي، غير أننا سوف نناقش بإيجاز بعض الطرق شائعة الاستخدام في هذا الشأن.

✓ طريقة تعتمد على الارتباط بين الاختبار المعد والاختبار المحك: يمكن توضيح خطوات هذه

الطريقة بالشكل التخطيطي التالي:



الشكل (09): خطوات تقدير الصدق التنبؤي للاختبار

المصدر: صلاح علام، 2000.

حيث يتم هنا حساب معامل الارتباط بين درجات الاختبار المعد ودرجات محك الأداء، ويتم تفسير معامل الارتباط باستخدام مفهوم التباين المشترك لكل من درجات الاختبار المعد والاختبار المحك الذي نحصل عليه بترتيب قيم معامل الصدق التنبؤي، فإذا كانت قيمة هذا المعامل تزيد عن (0.71) فإن هذا يعني أن (50%) من تباين درجات الاختبار المحك (أي نصف مقدار التباين) يعزى إلى درجات الاختبار المعد، وينبغي الإشارة هنا إلى أنه لا ينبغي أن نتوقع الحصول على قيم مرتفعة لمعامل الارتباط بين درجات الاختبارين (المعد والمحك) في دراسات الصدق التنبؤي، وذلك لتعدد وتشابك السلوك الإنساني، الأمر الذي يجعل التنبؤ بأداء الأفراد في أكثر من موقف غير واقعي، لذلك يرى نانلي Nunnally أنه من

الأفضل تفسير قيم معامل الصدق في ضوء التحسن المحتمل في متوسط أداء الأفراد (صلاح علام، 2000).

✓ طريقة تعتمد على استخدام معادلة الانحدار في التنبؤ بدرجات الاختبار المحك:

حيث أنه باستخدام معادلة الانحدار (**Regression Equation**) التي تعتمد على قيمة معامل الارتباط بين المتغيرين، يمكن التنبؤ بدرجة الفرد في الاختبار المحك بمعلومية درجته في الاختبار التنبؤي اعتماداً على معامل الصدق التنبؤي، غير أن هذا يتطلب أن تكون العلاقة بين درجات الاختبارين مستقيمة، ويمكن التوصل إلى معادلة خط الانحدار في مرحلة دراسة صدق الاختبار المستخدم في التنبؤ، حيث يتم استخدام معامل الارتباط بينه وبين الاختبار المحك وكذا استخدام كل من المتوسط، والانحراف المعياري (صلاح علام، 2000).

✓ طريقة تعتمد على جداول التوقع:

تستخدم هذه الطريقة في وصف العلاقة بين درجات الاختبار التنبؤي ودرجات الاختبار المحك، وتعتمد على جداول يمكن تكوينها تسمى جداول التوقع **Tables Expectancy**، وتنظم بشكل يسمح بتقدير أرجحيه حصول فرد على درجة أو تحقيقه لمستوى معين من الكفاية في اختبار محك إذا علمنا درجته في الاختبار التنبؤي، أي أن هذه الجداول تفيد في تقدير احتمالات نجاح الأفراد الذين حصلوا على درجات معينة في الاختبار التنبؤي، كأن يحدد المعلم توقعات لدرجات طلبته ثم بعد تطبيق الاختبار يحدد درجاتهم

الفعلية، أو أن يحدد المسؤول درجات الأداء لعماله المتوقعة ثم يحدد درجات أدائهم الفعلية بعد إنجاز العمل، ولتكوين جدول التوقع ينبغي تحديد كل من الدرجات المتوقعة (درجات الاختبار التنبؤي) والدرجات الفعلية (درجات المحك)، وتقسم درجات كل من الاختبارين إلى عدد من الأقسام المتساوية في العدد، ومن ثم تمثيل درجات الاختبار التنبؤي على المحور الأفقي، وأقسام درجات اختبار المحك على المحور الرأسي، وتدوين النسب المئوية(صلاح علام، 2000).

✓ طريقة تعتمد على نظرية المنفعة:

تقوم هذه الطريقة على تقدير المنفعة، حيث يتم تقدير الصدق التنبؤي من خلال مقارنة التكلفة المتوقعة بالتكلفة الفعلية (صلاح علام، 2000).

مما سبق حول الصدق المرتبط بالمحك نلاحظ أن التمييز بين كل من الصدق التلازمي والصدق التنبؤي على الفترة الزمنية التي تربط بين الحصول على درجات الاختبار والحصول على درجات المحك، لكنه يعتمد على الهدف من الاختبار، فالصدق التلازمي يستخدم مع المقاييس التي تشخص الوضع الراهن بدلاً من التنبؤ بالنواتج المستقبلية.

▪ صدق التكوين الفرضي (Construct Validity):

إلى غاية بداية الخمسينات من القرن العشرين، كانت حركة القياس النفسي مهتمة بالصدق التنبؤي وصدق المحتوى، وهذان النوعان من الصدق يفيدان في بحوث علم النفس التطبيقي، أما مشكلات البحوث

الأساسية في علم النفس، فلا تصلح لها هذه الأنواع من الصدق (بشير معمرية، 2012، ص179)؛ لكن مع بداية الخمسينات ظهر ما يعرف بصدق التكوين الفرضين والذي يعرف أيضا بصدق البناء، وهو تصور فكري متعلق بالمفهوم وأبعاده، ويعتبر هذا النوع من الصدق أفضل ما توصل إليه العلماء في وقت ما لتوضيح سمة معينة.

وصدق التكوين الفرضي يركز على وجود مؤشر أو دليل على أن درجات أداة القياس تقيس المفهوم النظري أو السمة المراد قياسه، فعلى سبيل المثال نجد النضج الاجتماعي على أنه مفهوم نظري لا يمكن قياسه بشكل مباشر، ونجد أن التكوين الفرضي لهذا المفهوم هو تعريفه الإجرائي، والذي يضم عناصر ومكونات تعريفه الإجرائي التي نعتقد انه بإمكاننا أن نقيس أو نصف المفهوم من خلالها (مساعدة الذات، التوجيه الذاتي، المهنة، الاتصال، العلاقات...) (سنة إبراهيم أبو دقة، 2012)؛ وهذا النوع من الصدق مرتبط بالمقاييس النفسية بصفة عامة، حيث أن معظم السمات في علم النفس افتراضية، حيث لا يوجد اتفاق على تعريف إجرائي موحد لها، وصدق المفهوم يهتم بجميع الدلائل الضرورية والتي يتم بواسطتها التعرف على قدرة درجة الاختبار في أن تكون مؤشرا على السمة التي يفترض أن يقيسها الاختبار، أي يتناول العلاقة بين نتائج المقاييس وبين المفهوم النظري والذي يهدف الاختبار لقياسه (سوسن شاكر، 2010).

وكلمة المفهوم تدل على شيء قد يكون محسوسا (كالشجرة) أو مجردا (كصفة الذكاء والقدرة الموسيقية)، وهو يتمثل في الارتباط بين الجوانب التي يقيسها الاختبار وبين مفهوم هذه الجوانب، ويسمى صدق

المفهوم لأنه يقوم على تحديد المفاهيم أو البنى المقدمة للظاهرة المقاسة، وبعبارة أخرى فهو نجاح الاختبار في قياس سمة أو قدرة معينة كالذكاء أو الانطواء أو الانبساط. ويتوقف هذا النوع من الصدق على مقدار ما نحصل عليه من معلومات عن هذه السمة وخصائصها ومكوناتها، حيث يقوم الباحث عندما يصمم اختبارا يقيس هذه الصفة أو القدرة بتعريف الصفة التي يريد قياسها منطلقا من إطار نظري، ثم يبدأ بتحليل المجال أو الناحية التي يريد قياسها منطلقا من إطار نظري، ثم يحدد المجال أو الناحية التي يريد قياسها تحليلا يكشف عن عناصرها المختلفة الرئيسة، ثم يحلل كل قسم ويجزئه إلى أجزائه، ثم يقدر النسب المئوية لكل جزء من كل قسم من هذه الأقسام، وهذا ما فعله ألفرد بينيه عامي (1867،1911)، وهو أول من صمم اختبارا للذكاء متبعا في ذلك الخطوات السابقة الذكر (سوسن شاكر، 2010).

وقد أورد "ساكس و زيلر zeller & Sax" الخطوات التي يمكن من خلالها التوصل إلى دلائل تتعلق بصدق التكوين الفرضي (المفهوم) وهي:

- تبرير أهمية التكوين الفرضي من الناحية النفسية أو التربوية على أن يكون التكوين الفرضي معرفا تعريفيا إجرائيا ويدل على سمة قابلة للقياس.
- الاستناد إلى نظرية نفسية أو تربوية أو اقتراح نموذج منطقي يوضح المفاهيم والعلاقات القائمة بينها.
- التمييز بين التكوين الفرضي والتكوينات الفرضية الأخرى المماثلة لها.

- التوصل إلى أدلة من مصادر متعددة باستخدام الأساليب الارتباطية والتجريبية والمنطقية لتأكيد التكوين الافتراضي، وعندما تستخدم مقاييس متعددة في هذا الشأن فإنه يمكننا الحصول على الصدق التقاربي للتكوين الفرضي.

- التوصل إلى أدلة نتأكد منها أن التكوين الفرضي لا يرتبط بعوامل وقتية أو دخيلة لكي نحصل على الصدق التمايزي للتكوين الفرضي.

- إجراء تعديلات مستمرة في التكوين الفرضي بما يتفق والأدلة والمعلومات الجديدة المتجمعة. ويتطلب تقدير هذا النوع من الصدق دراسات متعددة وأساليب منطقية وإحصائية وتجريبية. ويمكن القول أن هذا النوع من الصدق يشكل المرحلة النظرية أو التمهيدية في تطوير المقاييس، وهو موجه لخدمة الاختبار نفسه وذلك بمحاولة الانتقال من الشك في أن الاختبار يقيس السمة التي أعد لقياسها إلى التيقن.

ومن ثم فصدق التكوين الفرضي هو مدى قياس الاختبار لتكوين فرضي معين أو سمة معينة، ويعتبر هذا النوع من الصدق أكثر من غيره على تقديم وصف أشمل، لذلك فهو يتطلب معلومات أكثر عن الظاهرة موضوع القياس نحصل عليها من مصادر مختلفة، ومن هذه المصادر (محمد أبو هشام

حسن، 2006)

- تمايز العمر لمعرفة ما إذا كانت الدرجات على الاختبار المعني تتزايد بتقدم العمر، ويتطلب هذا بالطبع افتراض أن الظاهرة موضوع القياس ظاهرة ارتقائية.

- معاملات الارتباط بالمقاييس الأخرى، وذلك لتحديد مدى ارتباط الاختبار الحالي بالمقاييس الأخرى المماثلة له، ومدى تحرره من العوامل التي لا علاقة لها بما يقيسه والتي تقيسها مقاييس أخرى.

- التحليل العاملي ويتضمن تحليل العلاقات بين البيانات وإيجاد معاملات الارتباط لتحديد الأسس الإحصائية التي يمكن أن تصنف إليها المقاييس.

- معاملات الارتباط بين كل سؤال والدرجة الكلية فيه.

- المقارنة بين المجموعة التي تحصل على أعلى الدرجات في الاختبار والمجموعة التي تحصل على أدنى الدرجات فيه بالنسبة لأداء كل منها في كل بند من بنود الاختبار.

- دراسة اثر المتغيرات التجريبية في درجات الاختبار، وذلك باستخدام المنهج التجريبي المعتاد، حيث تصمم التجارب للتحقق من صحة فروض عديدة تتعلق بما يقيسه الاختبار.

✓ طرق حساب صدق التكوين الفرضي:

فيما يلي بعض المؤشرات في مجال صدق التكوين الفرضي (صفوت فرج، 2007، 264)، (مقدم عبد الحفيظ،

1993، 150) وطرق حسابها.

أ- الفروق بين الجماعات:

الخصائص السيكولوجية رغم أنها مفترضة ومجردة، إلا أنها تتعكس في سلوك الأفراد وفي أدائهم على المقاييس وأدوات القياس النفسي الأخرى، ولأن هذه الخصائص لا توجد بكميات متساوية عند الأفراد، فمن المنطقي أن نفترض أنهم يختلفون في كم وجودها لديهم، سواء بصفتهم أفراداً أو جماعات، وبالتالي يمكن أن نتوقع فروقا بين الجنسين في خصائص معينة كالقدرات والسمات الشخصية، ونتوقع فروقا بين الجماعات العمرية في نمو خصائص نفسية عبر مراحل العمر المبكرة خاصة (كالنمو المعرفي والنمو اللغوي والنمو الاجتماعي والنمو الانفعالي وغيره...)، ونتوقع كذلك فروقا بين الأسوياء وغير الأسوياء على استبيانات تقيس جانبا من السلوك اللاسوي، ونتوقع كذلك فروقا بين ذوي الضبط الداخلي وذوي الضبط الخارجي في إدراكهم لمصادر التعزيز الذي يحصلون عليه، ونتوقع كذلك فروقا بين العصائبيين وغير العصائبيين في سمة الخجل وغيره، وإذا تمكن الاستبيان من إبراز هذه الفروق والتوقعات النظرية، بناء على خصائص التكوين المبدئي للخاصية، سيكون تقديرنا للاستبيان أنه صادق.

ب- التغيير في الأداء على خاصية سلوكية:

يقوم الأخصائي في القياس النفسي هنا بدراسة الفروق في الأداء على المقاييس/الاختبارات في ثلاث حالات كما يلي:

- **الأولى:** عندما يفترض أن خاصية سلوكية (مثل المثابرة أو الانطواء...) لا تتغير مع التقدم في العمر، وبالتالي فإذا قام بقياسها لدى نفس العينة في فترات زمنية مختلفة، عليه ألا يتوقع تغيراً في الأداء المرتبط بهذه الخاصية السلوكية إذا قاسها باختبار صادق.

- **الثانية:** عندما يفترض أن خاصية سلوكية مثل (اللغة والذكاء...) تتغير مع التقدم في العمر، وبالتالي فإذا قاسها لدى عدد من العينات المختلفة في الأعمار الزمنية، فإنه يتوقع أن توجد بينها فروق حقيقية، مما يشير إلى صدق الاختبار.

- **الثالثة:** عندما يفترض أن تعرض الشخص لموقف معين، سوف يؤدي إلى تغير درجته على اختبار يقيس خاصية سلوكية ترتبط بهذا الموقف، فمثلاً عندما يتعرض فرد لموقف الامتحان، فإن درجته سوف ترتفع على استبيان يقيس قلق الامتحان أو القلق العام، وإذا تعرض شخص للإهانة أو للخزي، فإن درجته سوف تتغير على استبيان يقيس الإحباط، وهاتان النتيجةتان تؤيدان صدق الاستبيانين.

وبالإضافة إلى التصنيف السابق للجمعية الأمريكية لطرق تقدير صدق المقاييس يمكن التطرق إلى الطرق التالية:

ج- الصدق التمييزي (أسلوب المقارنات الطرفية):

الصدق التمييزي يقوم على أحد مفاهيم الصدق، وهو قدرة الاختبار على التمييز بين طرفي الخاصية التي يقيسها، ويتم ذلك من خلال إتباع المراحل التالية:

- ترتيب درجات أفراد العينة على الاختبار تنازليا أو تصاعديا.
- سحب نسبة 27% من الدرجات العليا والدنيا للأفراد.
- حساب المتوسطات الحسابية والانحرافات المعيارية لكل مجموعة.
- تقدير الفروق بين متوسطي المجموعتين المتطرفتين في الاختبار، ومدى دلالتها، أي مدى توافر الاختبار على القدرة التمييزية بين العينتين المتطرفتين في الخاصية المراد قياسها (بشير معمرية، 2012).

د- حساب الارتباطات بين المتغيرات توافقيا وتعارضيا: يرى كرونباخ أن الصدق يتحدد بوضوح بواسطة اكتشاف التوافق والتعارض معا بين الاختبار وغيره من المقاييس، وفقا للإطار النظري الذي تنتمي إليه المتغيرات محل القياس، حيث يقوم الأخصائي القياسي بحساب معاملات الارتباط بين المتغيرات التي تَبَيَّن نظريا، أو من نتائج البحوث النفسية المتواترة، أنها ترتبط فيما بينها في الاتجاه الإيجابي أو السلبي.

هـ- طريقة التحليل العاملي:

ترتبط أنستازي الصدق العاملي بصدق البناء، وفي هذا الأسلوب نحصل على تقدير كمي لصدق الاختبار في شكل معامل إحصائي، هو تشعب الاختبار أو البنود على العامل الذي يقيس خاصية معينة، ويهدف التحليل العاملي إلى التعرف على بنية الخاصية النفسية من خلال تحليل العلاقات بين البيانات ووضعها في صورة معاملات ارتباط، فنقطة البداية في التحليل العاملي هي حساب معاملات الارتباط بين درجات كل بند ودرجات البنود الأخرى التي ينبغي أن تكون قد طُبِّقَتْ على نفس المجموعة من المفحوصين،

وتجمع معاملات الارتباط في جدول يسمى فنيا "مصفوفات الارتباط Correlation Matrix" (صفوت فرج، 2007).

وعند تطبيق طرق إحصائية لتحليل المصفوفة عامليا، نخرج بعدد من تجمعات معينة بين المقاييس (البنود)، هذه التجمعات هي العوامل التي تعبر عن التباين المشترك بين الاختبار والتباين الخاص بالعامل، والذي يعبر عن إسهام البنود المشتركة في قياس خاصية معينة، أو تكوينات فرضية معينة، فمثلا إذا كانت المقاييس من نوع المفردات والمتشابهات والأضداد وتكملة الجمل تربطها معاملات ارتباط قوية، بينما تربطها معاملات ارتباط منخفضة مع المقاييس الأخرى التي تقيس العلاقات الميكانيكية والمكانية والهندسية منخفضة، نستنتج من ذلك وجود عامل إحصائي تشعب عليه المجموعة الأولى من المقاييس تشعبا عاليا، وتشعب به المجموعة الثانية تشعبا منخفضا، ولأن تشعب الاختبار على العامل هو في حقيقته معامل ارتباط البند أو الاختبار بالعامل، فمثلا تشعب اختبار مقاومة الضغوط على العامل المتعلق بالتكيف بمقدار (0.72) مثلا، يعني أن هذا الاختبار يقيس هذه القدرة بمعامل صدق عاملي قدره (0.72) ونستطيع أن نتعرف على المكونات العاملة للاختبار بحساب تشعباته على العوامل المختلفة التي خرجنا بها من تحليل مصفوفة ارتباطيه لخاصية متجانسة (بشير معمرية، 2012، ص195).

وفي أسلوب آخر من التحليل العاملي، نستعمل مقاييس صادقة معترف بها في قياس خاصية معينة كمحكات مستقلة، ثم نطبقها مع الاختبار الجديد الذي صمم لقياس نفس الخاصية في نفس الوقت على

أفراد العينة، ثم نحسب معاملات الارتباط بين درجة كل اختبار وآخر، الاختبار المحك والاختبار الجديد، لتكوين مصفوفة ارتباط، وبعدها نقوم بإجراء إحصاءات التحليل العاملي، لنصل إلى تقدير حجم التباين الحقيقي المشترك بين الاختبار الجديد موضوع الصدق العاملي، ومجموعة المقاييس المحكية، فيصبح التحليل العاملي مزيجاً من صدق البناء والصدق المرتبط بالمحك الخارجي (تلازمية)، بشرط أن تكون المقاييس المحكية صادقة بأساليب أخرى، أو من خلال محكات أخرى وفي دراسات سابقة (مقدم عبد الحفيظ، 1993).

لكن من الملاحظ أن أسلوب التحليل العاملي، نجح فقط في استعمالاته الأولى التي توصل السيكولوجيون الإنجليز والأمريكان من خلاله إلى بناء نماذج نظرية، خاصة في الذكاء وفي الشخصية، مثل الإنجليزي "تشارلز سبيرمان Charles Spearman" في بنائه لنظرية العاملين في الذكاء، والأمريكي "لويس ثرستون Lewis Thurston" في وضعه لنظرية العوامل المتعددة في الذكاء، كذلك الأمريكي "جوي جلفورد Joey Gilford" الذي أبدع نظريته في البناء الشامل للعقل، وفي مجال الشخصية نذكر الإنجليزي "هانز إيزنك Hans Aaznk" في بنائه لنظرية الأبعاد الأساسية للشخصية، و"ريموند كاتل Raymond Cattell" في

استخراجه للعوامل الستة عشرة في الشخصية وغيرهم... (Philippe Cibois,2006)

أما عن استعمالاته خارج هذه الأطر النظرية، فلا يمكن أن نقول أنه أسلوب جيد وملائم لتقدير صدق الاختبار، حيث أن الهدف العام لأسلوب التحليل العاملي، هو التعرف على البنية العاملية للخاصية المراد

قياسها، وذلك على اعتبار أن الخصائص البشرية متعددة الأبعاد والعوامل، وليست وحيدة البعد أو وحيدة العامل، ولكن السؤال المطروح هنا هو كيف نتأكد من أن العوامل المستخرجة تشير إلى صدق الاختبار؟ في ظل غياب المرجعيات التي نعود إليها للحكم على صدق الاختبار من عدمه مما يجعلنا لا نعرف بالضبط ما نوع الصدق الذي تم تقديره بهذا الأسلوب، غير أنه ما دام التحليل العاملي أسلوباً إحصائياً لتقدير صدق البناء، كما يعتبره كل من (أنستازي، كرونباخ)، فإن الأمر يتطلب وضع فروض حول البنية العملية للخاصية المراد قياسها، وفقاً للإطار النظري الذي يتناولها، ثم تخضع الفروض للاختبار بأسلوب التحليل العاملي. وتصبح صحتها أو عدم صحتها نقطة مرجعية للحكم على أن الاختبار صادق أم لا، وهذا هو الأسلوب الذي اتبعه مبتكرو ومستعملو أسلوب التحليل العاملي.

من خلال عرضنا السابق لأنواع تفحص صدق درجات الاختبار يتضح لنا أن الطرق المذكورة آنفاً ليست مستقلة عن بعضها البعض، فالتحقق من صدق المحتوى ينبغي أن يسبقه عادة الصدق المرتبط بالمحك، والتوصل إلى أدلة صدق التكوين الفرضي يتطلب جميع البيانات المستمدة من تلك المستخدمة في التحقق من صدق المحتوى، والصدق المرتبط بالمحك؛ كما أوضح كرونباخ (1971، Cronbach) أيضاً أنه يمكن التحقق من صدق المحتوى إمبريقياً بإيجاد الارتباط بين درجات صيغ الاختبار نفسه إذا استند بناؤها إلى نفس نطاق المحتوى (Content Domain) (صلاح الدين محمود علام، 2000، ص111)؛ كما أن هناك جملة

من العوامل المؤثرة على صدق درجات الاختبار والتي يجب أن يأخذها الباحث بعين الاعتبار سواء أثناء بناء أو تكييف اختبار ما، أو حتى أثناء تطبيقه وتصحيحه.

▪ العوامل المؤثرة في صدق الاختبار:

يتأثر صدق الاختبار بمجموعة من العوامل التي من الواجب أن يأخذها معد أو مطور المقاييس بعين الاعتبار، كي لا يظهر أثرها السلبي على مصداقية الاختبار وموثوقية نتائجه؛ ومن بين هذه العوامل نجد (بشير معمرية، 2007):

✓ **طبيعة عينة التقنين:** لفهم طبيعة معامل الصدق، ينبغي وصف طبيعة عينة التقنين وصفا مفصلا. لأن الاختبار الواحد قد يقيس عمليات نفسية مختلفة إذا طبق على عينات تختلف في العمر أو الجنس أو المهنة أو المستوى التعليمي أو السواء واللاسواء، أو أي متغير آخر يرتبط بمفهوم الصدق. فمثلا اختبار الأداء عند فئة المهندسين لا يكون صالحا لقياس الأداء عند العمال المهنيين.

✓ **تجانس عينة التقنين:** من العوامل التي تؤثر كذلك في معامل الصدق سلبيا تجانس أفراد عينة التقنين، بحيث كلما زاد التجانس بين أفراد العينة في الخاصية محل القياس أدى ذلك إلى انخفاض ملموس في مدى درجات أفراد عينة التقنين على الاختبار، الأمر الذي يترتب عليه انخفاض واضح في قيمة معامل الارتباط بين درجاتهم على الاختبار ودرجاتهم على المحك، لذا ينبغي أن تتميز عينة التقنين بعدم التجانس والاختلاف في الخاصية موضع القياس، فإذا كانت الخاصية المراد حساب صدق الاختبار

الذي يقيسها هي الدافعية فإنه ينبغي أن تكون العينة متباينة في مستويات الدافعية (من أقل من المتوسط إلى متوسط إلى فوق المتوسط إلى مرتفع).

✓ **الفترة الزمنية بين تطبيق الاختبار وتطبيق المحك:** تعد الفترة الزمنية الفاصلة بين تطبيق الاختبار

وجمع المعلومات عن المحك من العوامل التي تؤثر على الصدق، فعندما تطول هذه الفترة الزمنية ينخفض معامل الصدق، ويعزى ذلك إلى عامل الصدفة التي تؤثر في درجات الأفراد على المحك، وإلى عوامل أخرى مثل نقص أفراد العينة بسبب غيابهم، أو انتقالهم إلى أماكن أخرى، أو عدم تعاونهم، لذا ينبغي جعل هذه الفترة الزمنية قصيرة حتى يتم تجنب التأثيرات السلبية على معامل الصدق.

✓ **العلاقة الخطية بين الاختبار والمحك:** من المعروف أن حساب معامل الارتباط بين الاختبار والمحك

بطريقة كارل بيرسون يشترط افتراضاً جوهرياً، وهو أن تكون العلاقة بين المتغيرين (الاختبار والمحك) علاقة خطية على امتداد مدى الفروق الفردية بينهما، وإذا لم تتحقق فرضية العلاقة الخطية تنخفض قيمة معامل الارتباط بين درجات الأفراد على الاختبار ودرجاتهم على المحك، لذا ينبغي التأكد أولاً من خطية العلاقة بين الاختبار والمحك أو استعمال طرق إحصائية أخرى لا تقوم على فرضية العلاقة الخطية بين المتغيرات.

وخلاصة القول أنه لا توجد طريقة واحدة للتحقق من صدق المقاييس/الاختبارات التربوية والنفسية، وذلك

لأن نوع الأدلة المرجوة يعتمد على كيفية استخدام هذه المقاييس، فالاختبار لا يكون صادقاً أو غير

صادق بصفة عامة، وإنما يمكن تحديد ذلك في ضوء أغراض بناء الاختبار واستخداماته، فمن المعلوم أن الاختبار أو الاختبار ينبغي أن تتميز درجاته بالثبات قبل الحكم على صدقه، والصدق ليس خاصية من خصائص هذا الاختبار على وجه الخصوص، وإنما يتعلق بالاستدلال الذي يمكن التوصل إليه استناداً إلى الدرجات التي يتم الحصول عليها من خلال تطبيقه ، لذلك فإن الاهتمام ينبغي أن يوجه بدرجة أساسية إلى صدق التفسير المستند من هذه الدرجات، ونظراً لأن البيانات المستمدة من دراسات الصدق تتعلق بمواقف معينة فإن نتائج هذه الدراسات سوف تعتمد على خصائص هذه المواقف، ويترتب على ذلك أن يكون للاختبار أكثر من صدق واحد بقدر المواقف التي تتم دراسة الصدق في ضوءها.

2-4 الثبات: تعريفه، طرق حسابه، العوامل المؤثرة فيه

■ مفهوم الثبات:

نشأ مفهوم الثبات نتيجة لظهور أعداد كبيرة من الأدوات التي تقيس مختلف نواحي السلوك الإنساني وذلك بعد الحرب العالمية الثانية، وبالتالي نشوء الشركات التي أخذت تتبنى هذه الأدوات وتقوم بتسويقها، وهنا واجهتهم مشكلة تتمثل في الأعداد الكبيرة من هذه الأدوات حتى أصبح لكل مفهوم أو مصطلح نفسي عدد كبير من الأدوات التي تقيسه، وبالتالي بات المشكل المطروح هو أي هذه المقاييس أجود وأفضل من حيث دقته في قياس الظاهرة التي يراد قياسها؟ وهو التساؤل الذي جعل شركات القياس تتجه إلى إيجاد المعايير الأساسية التي يحتكم إليها لتحديد جودة الأداة، وكان من أهم المعايير والتقنيات مفهوم الثبات،

حيث بدأ علماء القياس في حسابه بطرق عديدة تقوم الفكرة الأساسية لمعظمها على معامل الارتباط (سوسن شاكر، 2010).

حيث أن الثبات يفيدنا في الكشف عن مصادر الخطأ التي أثرت على الاختبار، وعن طريق عمليات الثبات التي نجريها يمكن معرفة حجم التباين الحقيقي، ونسبته من التباين الكلي في الدرجات التي نحصل عليها عند تطبيق الاختبار. وتجدر الإشارة في هذا المقام إلى أن الطرق المختلفة لحساب الثبات لا يمكن أن تستعمل كلها في نفس الوقت ومع ذات الظروف للكشف عن مصادر الخطأ في أداة القياس، حيث يجدر بالباحث أن يحدد أسلوب التأكد من ثبات درجات أداة القياس المختلفة تبعا لطبيعة أدواته والغرض منها وفترة تطبيقها وكل المتغيرات الأخرى المتعلقة بها.

ويمكن تحديد مفهوم الثبات نظريا وإحصائيا كالتالي:

✓ مفهوم الثبات نظريا: يوضح ملحم (2000) بأن الثبات ليس صفة للاختبار النفسي بحد ذاته، بل هي صفة تتعلق بثبات الدرجات أو النتائج التي نحصل عليها عند تطبيق الاختبار على مجموعة من الأفراد، فكلما كان الاختبار ملائما أو مناسباً للسمة المراد قياسها زاد ثبات النتائج المحصل عليها، وذلك إما بتطبيق الاختبار نفسه على الأفراد أنفسهم في ظروف تطبيق مشابهة وفي فترات زمنية متباعدة وهو ما يسمى بثبات التكافؤ، أو عن طريق حساب التناسق الداخلي للبنود سواء بين البنود كلها، أو بتقسيم درجات البنود إلى قسمين (فردية وزوجية، عليا ودنيا) وحساب قيمة الثبات، فكلما ارتفعت قيمته زادت

درجة ثبات الاختبار، وللحكم على اختبار ما بأنه ثابت لا بد أن يتصف بالموضوعية، بمعنى حصول الفرد المطبق عليه الاختبار على الدرجة نفسها أو درجة قريبة جداً منها أياً كان المختص الذي يطبق عليه الاختبار ويصححه (بوسالم عبد العزيز، 2014).

مما سبق حول مفهوم الثبات نظرياً نستخلص أن الثبات هو مدى استقرار الدرجة التي يحصل عليها الفرد في اختبار يقيس لديه سمة معينة إذا ما أعيد تطبيقه عدة مرات، كذلك قد يشير إلى الاتساق الداخلي للاختبار الذي يمنحه درجة عالية من الدقة في قياس السمة المراد قياسها من خلال تطبيقه.

✓ مفهوم الثبات إحصائياً: يشير الثبات إحصائياً إلى أنه إذا ما تم تطبيق الاختبار على الشخص أكثر من مرة، فإنه يسجل نفس النتائج في كل مرة؛ هذا التجانس الذي يعرف بمعامل الثبات، يوفر حسب (Ebel, 1972) المؤشرات الإحصائية للخاصية موضوع القياس، والتي من خلالها يحكم على دقة الاختبار، وقدرته المستقرة على قياس السمة موضوع الدراسة. هذه القدرة التي لا يمكن أن تكون تامة، فثبات درجات الاختبار لا يمكن أن تصل إل الواحد (1) باعتبار أن الظاهرة في العلوم السلوكية تتأثر دائماً بعدد من المتغيرات الخارجية (الحالة النفسية والجسدية للمفحوص، درجات التعب ومستويات القلق أو الدافع تختلف من يوم لآخر) والعوامل البيئية (مثل: التغيرات في درجة الحرارة، الإضاءة، الضوضاء...)، وتعرف هذه العوامل وغيرها باسم (أخطاء القياس العشوائية)، أي أن درجة الفرد على الاختبار تعبر عن التباين الحقيقي للفرد (True Variance) وتباين الخطأ (Error Variance)، لذلك يؤكد "جيفورد Guilford" على

ضرورة حساب ثبات الاختبار كي تحدد الدرجة الحقيقية أو التباينات الحقيقية للاختبار، فمعامل الثبات

يوضح نسبة التباين الحقيقي في الدرجة المحسوبة على الاختبار. كما تشير أناستازي (Anne Anastasi)

(1976) إلى أن مصدر الثبات هو الاتساق في الدرجات التي يتم الحصول عليها من نفس الأفراد إذا ما

أعيد عليهم الاختبار نفسه في أوقات مختلفة، مع احتساب خطأ القياس لدرجة الفرد.

وينفق "سكس Sax" مع أناستازي في التأكيد على خطأ القياس في ثبات الاختبار، حيث يرى أن درجة

الفرد على الاختبار إنما هي درجته الحقيقية مضافا إليها درجة الخطأ، وهذا يتطلب تقدير درجة الخطأ،

وهو ما تسعى إليه مختلف طرق تقدير الثبات، هذه الأخيرة التي تسعى إلى قياس الجزء الحقيقي من

التباين العام للاختبار، وهذا الجزء هو الذي يعطينا القيمة العددية لارتباط الاختبار، فمعامل الثبات يزودنا

بتقدير كمي للعلاقة بين الفروق الفردية في الدرجة الحقيقية، والفروق الفردية في الدرجات الملاحظة،

وعندما نحصل على درجة ملاحظة خالية من الأخطاء فإن الثبات يكون تاما (+1)، أي أن الفروق بين

الدرجات الملاحظة والحقيقية متساوية، ولهذا فقيمة معامل الثبات تتراوح بين (0،+1).

كما يمكن القول أنه نظرا لاستحالة معرفة الدرجة الحقيقية ودرجة الخطأ في أي عملية قياس مهما حرص

الباحث على توفير الظروف الملائمة، ومهما كانت درجة دقة وموضوعية الاختبار المطبق، فقد ظهرت

معادلات مختلفة يتم من خلالها تقدير ثبات المقاييس وهو ما يعرف بطرق تقدير الثبات، أي أنه كلما

كانت درجات الاختبار خالية من الأخطاء العشوائية، وكانت قادرة على قياس المقدار الحقيقي للسمة في ظروف مختلفة ومتباينة كان الاختبار عندئذ اختباراً على درجة عالية من الثبات.

ويشير "ساكس Sax" إلى أن خطأ القياس أو درجة الخطأ هي نتيجة الاختلاف بين الدرجة التي يحصل عليها الفرد على الاختبار (الدرجة المحققة أو الملاحظة) والدرجة الحقيقية، وعندما يكون خطأ القياس صفراً فإن الدرجة التي يحصل عليها الفرد على الاختبار تكون هي درجته الحقيقية (بوسالم عبد العزيز،

(2014)

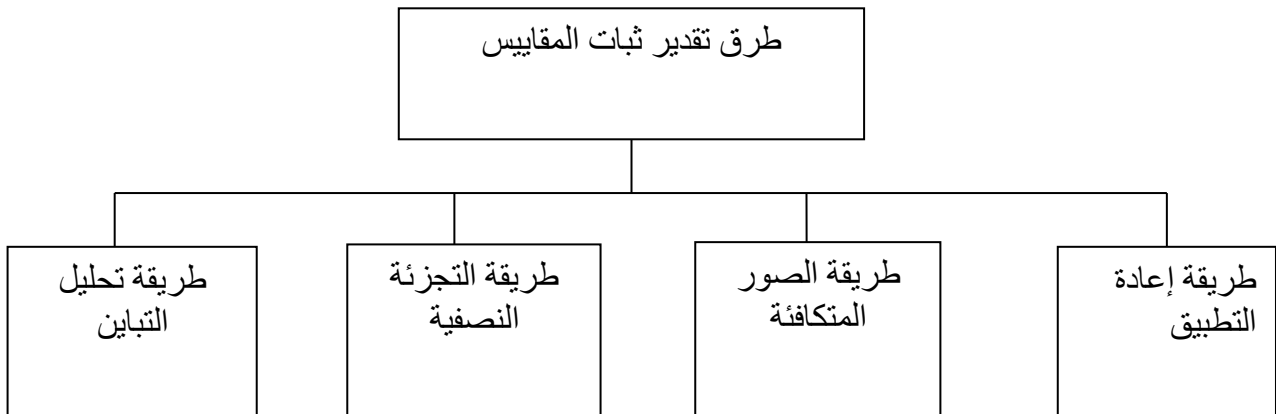
حيث أن:

$$\text{score Error score} = \text{Obtained True score} - \text{الدرجة الحقيقية} = \text{الدرجة الكلية} - \text{درجة الخطأ}$$

مما سبق يمكن القول أن الثبات خاصية تتعلق بدقة عملية القياس بصرف النظر عما يقيسه الاختبار، ويقاس الثبات إحصائياً من خلال حساب معامل الارتباط بين الدرجات التي يحصل عليها الأفراد في المرتين الأولى والثانية، أو من خلال معامل الارتباط بين صورتين لنفس الاختبار أو بين درجات نصفي الاختبار، أو عن طريق التحليل الداخلي لبنود الاختبار، وهي الطرق المختلفة للتحقق من ثبات الاختبارات بصفة عامة .

▪ طرق تقدير معامل الثبات:

يمكن تلخيص طرق تقدير ثبات درجات الاختبار من خلال الشكل التالي:



الشكل(10): طرق تقدير ثبات الاختبار

المصدر: جرد الباحثة

• طريقة إعادة التطبيق (Test-retest):

طريقة إعادة التطبيق تعني أن يعطى الاختبار مرتين لنفس المفحوصين، ويكون بين المرتين فاصل زمني قد يكون طويلا أو قصيرا حسب الظروف التي تواجه عملية القياس أو تتطلبها الحالة. وتكمن أهمية هذا النوع من أساليب تقدير ثبات المقاييس في أنه يقيس مستوى الاستقرار "Stability" الذي يتمتع به المفحوصين في السمة المراد قياسها، أو الخاصية المقاسة عبر الفترة التي تم خلالها تطبيق الاختبار مرتين(عبد الرحمن الطيري، 1994)؛ وتتوقف المدة الزمنية الفاصلة بين التطبيقين على طبيعة السمة التي يقيسها الاختبار، فإذا كان يقيس سمة سريعة التذبذب والتغير مثل: نكاء الأطفال الصغار، أو السمات المزاجية، فإن هذه الطريقة لا تكون مناسبة ويستحسن استخدام طريقة أخرى؛ أو يدخل تأثير الذاكرة قصيرة الأمد على نتائج التطبيق في المرة الثانية، لذلك يجب العناية باختيار المدة الزمنية الفاصلة اعتمادا على طبيعة السمة المراد قياسها، وخصائص مجتمع الدراسة، حتى يكون باستطاعتنا إرجاع التغيرات التي تحدث في درجات الأفراد خلال التطبيقين إلى أخطاء القياس(صلاح علام، 2012)، حيث أنه لا قيمة لهذا النوع من التحليل إلا في حالة قياس السمات أو الخصائص التي لا تتغير عبر الزمن، فرغم سهولة عملية

حساب الثبات عن طرق إعادة التطبيق إلا أننا يجب أن نضع في اعتبارنا تفاصيل أخرى كثيرة بجانب

طريقة حساب معامل الثبات هي: (بوسالم عبد العزيز، 2014)

- إذا ما تلا التطبيق الثاني مباشرة التطبيق الأول للاختبار، فإنه في هذه الحالة يتعرض للتأثر بعامل

الذاكرة، بمعنى أن المفحوص سوف يتذكر إجاباته في المرة الأولى ويكررها كما هي في المرة الثانية دون

بذل محاولة جديدة.

- إذا حاول الباحث تجنب الخطر الأول بأن يجعل الفترة الفاصلة بين التطبيقين فترة طويلة نسبياً لكي

يقلل من تأثير عامل الذاكرة، فإنه يتعرض بذلك لتأثيرات عامل جديد هو عامل النمو والنضج، أي سوف

تتدخل تلك التغيرات وعمليات النمو والتعلم التي لا بد وأن تطرأ على الوظيفة أو السمة المقاسة خلال

الفترة الطويلة الممتدة بين التطبيقين، وتؤثر في السمة فتتغير السمة وتظهر النتائج أنها مختلفة ومعامل

الارتباط بين التطبيقين القبلي والبعدي يكون منخفضاً، مما يوحي بانخفاض معامل ثبات الاختبار في

حين أن السمة هي التي تغيرت بين تطبيقين.

- قد يتعرض الفرد لبعض التغيرات أو المشاكل خلال الفترة الزمنية بين التطبيقين (نفسية، صحية،

اجتماعية، وغيرها)، مما يجعله قلقاً أو متعجلاً في أدائه مرة ومثانياً مرة أخرى، وهو الأمر الذي يؤثر

على درجته في الاختبار ويجعل قيمة الارتباط بين نتائج التطبيقين منخفضة.

- صعوبة الوصول إلى الأفراد أنفسهم في المرة الثانية لعوامل موضوعية متعددة، فقد يجد الباحث صعوبة في قبول التطبيق في المرة الأولى من طرف بعض المؤسسات التي ينتمي إليها أفراد عينته (تلاميذ في مدرسة، عامل في مصنع، موظفين في إدارة، مرضى في مستشفى)، وبالتالي يصعب عليه تطبيق الاختبار مرتين، فيختار طريقة أخرى لحساب الثبات من غير طريقة التطبيق وإعادة التطبيق.

في ضوء النقاط سابقة الذكر حول الشروط الواجب مراعاتها في تقدير الثبات بطريقة التطبيق وإعادة التطبيق، نلاحظ أنه يجب على الباحث التأكد من الضبط الدقيق لموقف الاختبار حتى لا تؤثر العوامل السالفة الذكر على معامل الارتباط بين التطبيقين مما يؤثر في قيمة معامل الثبات.

وعند حساب معامل الارتباط بين درجات التطبيقين الأول والثاني يجب على الباحث مراعاة طبيعة البيانات، فيمكن استخدام معامل ارتباط بيرسون إذا كانت البيانات من النوع المتصل، أما إذا كانت البيانات من النوع المنفصل فيفضل استخدام معامل ارتباط سبيرمان للرتب، إلا أنه في حالة وجود درجات أو قيم مماثلة كثيرة، أي نفس الدرجة تتكرر لدى عدد الأفراد فإنه لا يفضل استخدام معامل سبيرمان بل يفضل استخدام معامل ارتباط "كاندل" (سوسن شاكر، 2010).

وعموماً فاستخدام طريقة التطبيق وإعادة التطبيق في تقدير معامل الثبات يستخدم في الاختبارات والمقاييس المتعلقة بالتنبؤ بالاستعدادات، الاختبارات الموقوتة، مقاييس الاتجاهات والميول، لأنها تتطلب

تقدير تغيرها عبر الزمن، ولا تصلح للاختبارات والمقاييس التي تهدف إلى قياس التذكر أو ترتبط به ارتباطاً مباشراً، كما أنها لا تناسب اختبارات التحصيل لأن التحصيل سريع التغير .

• طريقة الصور المتكافئة (Equivalent forms):

طريقة الصور المتكافئة تعتبر من بين الطرق التي تستخدم لتقدير معامل ثبات درجات المقاييس والاختبارات النفسية، وفي هذه الطريقة يشترط توافر صورتين من الاختبار الذي يقيس السمة موضوع الدراسة، كما ويفترض تكافؤهما في المحتوى، أي أن تكون البنود في كلا الصورتين متكافئتين في السهولة والصعوبة وكذا في طبيعة المواضيع التي تتناولها هذه الصور، ومن أجل تقدير معامل الثبات بهذه الطريقة يتم استخراج معامل الارتباط بين درجات المفحوصين الذين تم تطبيق الصورة الأولى عليهم، والذين طبقت عليهم الصورة الثانية من الاختبار، وتجدر الإشارة إلى أن طريقة الصور المتكافئة تبرز الاتساق في الاستجابة على عينات البنود والفقرات المقدمة من خلال صورتين أو أكثر، كذلك يتبين من خلال هذه الطريقة الاستقرار عبر فترة زمنية معينة، ومن أجل الحصول على معامل ثابت مرتفع يفترض أن تكون الاستجابة على الصورتين متماثلة، حيث أنه كلما زاد التباين كلما تأثر معامل الارتباط، وبالتالي ثبات الاختبار (عبد الرحمن الطيري، 1994) .

وتسند هذه الطريقة على فرضية أن كلا الصيغتين من الاختبار تكونان في الحقيقة متكافئتين من حيث المحتوى، ونوع المفردات، وعددها، وصعوبتها، والمتوسط، والانحراف المعياري لدرجات كل منهما، حيث

ترجع أخطاء القياس في هذه الحالة إلى اختلاف مفردات أو أسئلة كل من الصيغتين، فكلما زاد الاختلاف بينهما انخفض الثبات، أي أن أخطاء القياس في هذه الحالة ترجع إلى اختلاف عينات الأسئلة في كل من الصيغتين، وليس إلى التغيرات التي تحدث للأفراد كما هو الحال في معامل الاستقرار (صلاح الدين محمود علام، 2000).

ومن الطرق الإحصائية التي نتحقق بها من صحة ودقة التكافؤ بين الصورتين أن يكون متوسط الدرجات والانحراف المعياري للصورة الأولى مقاربا لها في الصورة الثانية، ولتقدير معامل الثبات تطبق الصورتين المتكافئتين للاختبار إما في جلسة واحدة أو في جلستين منفصلتين على نفس مجموعة الأفراد، ثم نقوم بحساب معامل الارتباط بين درجات الاختبارين، ويسمى معامل الثبات في هذه الحالة بمعامل التكافؤ (Coefficient of equivalence) (محمود أحمد وآخرون، 2010)، وبخصوص الفاصل الزمني بين تطبيق الصورتين المتكافئتين يشير "أهمان وكلك Clock & Ahman" إلى أن هذا الفاصل الزمني يجب أن لا يقل عن اليوم الواحد ولا يزيد عن أسبوع (Ahman & Clock, 1981) ؛ في حين يشير بوسالم (2014) إلى أنه يتم تطبيق الصورتين المتكافئتين من الاختبار المراد تطبيقه على المجموعة نفسها في الوقت نفسه أو في وقتين متقاربين مع ظروف متشابهة (بوسالم عبد العزيز، 2014).

وعادة نحتاج لاستخدام الصور المتكافئة من الاختبار في بعض الحالات مثل حالة تطبيق البرامج التدريبية على مجموعة من الأفراد، في هذه الحالة نكون بحاجة إلى صورتين من الاختبار للقيام بالقياس

القبلي والبعدي، وكذا عند دراسة الآثار المحتملة لإدخال عامل تجريبي معين على أداء الأفراد، وتجدر الإشارة إلى أن إعداد الصور المتكافئة عادة يكون جد مكلف، حيث يتضاعف الجهد والوقت والتكلفة المتعلقة بتصميم الاختبار، لذلك يوجد عدد ضئيل نسبيا من هذه النوعية من المقاييس، وتعد هذه الطريقة في تقدير ثبات الاختبار من أحسن الطرق من حيث السرعة في الوصول إليه.

• طريقة التجزئة النصفية (Spilt-Half Reliability):

تستخدم هذه الطريقة عادة عندما تكون إعادة تطبيق الاختبار مكلفة جدا، أو أن يكون هناك احتمال لتضليل استجابات المفحوصين عند تطبيق الاختبار مرتين أو تطبيق صورتين؛ لكن التحدي المطروح هنا والذي يواجه طريقة تقدير الثبات من خلال التجزئة النصفية للاختبار هو كيفية تجزئته إلى نصفين متكافئين، حيث أن معظم المقاييس خاصة تلك التي ترتب مفرداتها وفقا لمستوى الصعوبة يكون نصفها الأول أيسر من نصفها الثاني، إضافة إلى تدخل عوامل انخفاض الدافعية والتعب والتي يكون لها تأثير على الأداء في الجزء الثاني من الاختبار، ومن ثم فإن الطريقة الأكثر شيوعا في الاستخدام، والأكثر ملائمة للحصول على نصفين متكافئين هي مقارنة درجات البنود الفردية للاختبار بدرجات البنود الزوجية التي تم ترتيبها وفقا لمستوى صعوبتها (محمود أحمد وآخرون، 2010، ص225).

كما أوضح بوسالم (2014) طريقة تقسيم الاختبار إلى قسمين والإجراءات المتبعة في ذلك وهي: (بوسالم

عبد العزيز، 2014)

- تطبيق الاختبار المراد التحقق من ثباته على عينة تجريبية أو ما يسمى بعينة التقنين.
- رصد الدرجات التي يتحصل عليها كل فرد في الاختبار.
- يتحصل كل فرد على درجة في كل فقرة، وكل فقرة لها درجة بالنسبة لكل الأفراد.
- يتم حساب معاملات السهولة أو الصعوبة أو التمييز لكل فقرة من فقرات الاختبار (حساب معاملات التمييز يكون بالنسبة للاختبارات التي تقيس سمات لا تتضمن مفهوم السهولة والصعوبة بمعنى عدم وجود إجابات صحيحة أخرى خاطئة).
- يتم ترتيب الفقرات تصاعديا او تنازليا من أسهل فقرة إلى أصعب فقرة أو العكس، أو حسب معامل تمييز كل فقرة.
- يتم ترقيم الفقرات من الأول إلى الأخير.
- تقسيم فقرات الاختبار إلى نصفين، الأول مجموع الفقرات ذات الأرقام الفردية والثاني لمجموع الفقرات ذات الأرقام الزوجية.
- بعدها ترصد درجات الأفراد على الفقرات الفردية وتعتبر بمثابة نصف الاختبار، ثم درجات الأفراد على الفقرات الزوجية ويعتبر بمثابة نصف ثان للاختبار.
- قبل تحديد طريقة من بين طرق التجزئة النصفية الملائمة لطبيعة الاختبار لاستعمالها للتحقق من الثبات، ويجب أولا التحقق من مدى تجانس درجات النصفين عن طرفي اختبار ف (F.Test) حيث:

$$F = \frac{\text{التباين الأكبر}}{\text{التباين الأصغر}}$$

- ثم نقارن ف المحسوبة بقيمتها النظرية وهنا نميز بين حالتين:

✓ ف المحسوبة أكبر من قيمتها النظرية يعني أن نصفي الاختبار غير متجانسين، ولحساب معامل

ثبات الاختبار نستعمل إما طريقة جتمان (Guttman Spilt Half)، أو طريقة رولون (Rulon Spilt Half).

✓ ف المحسوبة اقل من قيمتها النظرية يعني أن تباين نصفي الاختبار متجانس، ولحساب معامل ثبات

الاختبار نقوم بتطبيق معادلة سبيرمان وبراون (Spearman-Brown Spilt Half) التصحيحية بعد حساب

معامل الارتباط بيرسون بين درجات النصفين.

وسوف نتطرق لهذه المعادلات بالتفصيل كالتالي:

▪ معادلة سبيرمان وبراون (Spearman-Brown Spilt Half): بما أن الثبات بالتجزئة النصفية هو عبارة

عن ثبات نصف الاختبار وليس الاختبار كله، لذلك ينبغي تصحيح معامل الارتباط الذي يمثل معامل

الثبات في التجزئة النصفية بمعادلة (سبيرمان-براون) من أثر التجزئة، أي أننا نستخدم معامل الارتباط

بين النصفين للتنبؤ بمعامل ارتباط الاختبار الكلي مع نفسه (محمود أحمد وآخرون، 2010، ص 225)

حيث يتم حسابها وفق المعادلة التالية:

$$\text{معامل الثبات للاختبار ككل} = \frac{2 \times \text{معامل الارتباط بين نصفي الاختبار}}{1 + \text{ثبات نصف الاختبار}}$$

بيد أن هناك بعض الانتقادات الموجهة لهذه المعادلة في تصحيح معامل الثبات الذي تم إيجاده منها:

(سوسن شاكر، 2010).

- افتراضها تشتت واحد لنصف الاختبار، ونادرا ما يكون الانحراف المعياري متساويا في النصفين.

- تحتاج إلى عمليات حسابية كثيرة، لأنها تحتاج على حساب معامل الارتباط بين جزئي الاختبار.

- مبالغتها في رفع قيمة معامل الثبات النصفي.

- لا تصلح مع الاختبارات الموقوتة (المرتبطة بالزمن).

▪ معادلة رولون (Roulon): تعتبر معادلة رولون طريقة مختصرة لحساب معامل الثبات الكلي للاختبار

ولا تحتاج لإيجاد معامل الارتباط بين جزئي الاختبار كما هو الحال مع معادلة (سييرمان براون) (سوسن

شاكر، 2010، ص88)؛ حيث افترض رولون أن تباين درجات الأفراد على الاختبار ككل يرجع الى تباين

حقيقي في مستويات القدرة المقاسة، كما يرجع إلى أخطاء تجريبية في القياس (محمد شحاتة، 2008، ص92):

ويمكن توضيح هذه المعادلة كالتالي:

$$r = 1 - \frac{\sigma_{\epsilon}^2}{\sigma^2}$$

حيث أن:

σ^2 هو تباين فروق درجات النصفين.

σ_{ϵ}^2 التباين الكلي لدرجات الاختبار.

▪ معادلة جتمان (Guttman):

تعتمد على نفس المنطق الذي اعتمدت عليه معادلة رولون والفارق بينهما لا يخرج عن كونه فارقا في

أسلوب الحساب الذي يجعل من معادلة جتمان أكثر سهولة، إذ يستخدم فيه التباين الخاص بكلا

النصفين، حيث نقوم بجمع هذا التباين وقسمته على التباين الكلي للاختبار، كما لا تحتاج معادلة جتمان

لاستخدام معادلة أخرى لتصحيح الطول، ونص المعادلة كالتالي:

$$r_k = 2 \left(\frac{e_1^2 + e_2^2}{e_k^2} - 1 \right)$$

حيث أن:

r_k : ثبات الاختبار.

e_1^2 : تباين الجزء الأول من الاختبار.

e_2^2 : تباين الثاني من الاختبار.

e_k^2 : تباين الاختبار كله.

وعموما يمكن القول أن تقدير ثبات الاختبار بطريقة التجزئة النصفية غير مناسب للمقاييس

والاختبارات الموقوتة (اختبارات السرعة)، كما لا تصلح مع المقاييس والاختبارات التي لا يمكن

تقسيمها إلى نصفين متكافئين.

• طريقة الاتساق الداخلي (Inter item consistency):

والثبات هنا هو اتساق الأداء من بند إلى آخر في الاختبار نفسه، ومن أهم المعادلات المستخدمة في

حساب معامل الثبات بتحليل تباين (ع2) درجات الأفراد هي:

▪ معادلة كيودر ورتشاردسون (Richardson & Kuder) (1937): تهدف طريقة كيودر ورتشاردسون إلى

التوصل إلى قيمة تقديرية لمعامل ثبات الاختبارات والمقاييس غير الموقوتة أي اختبارات ومقاييس القوة

(Power Tests)، والتي تكون درجات مفرداتها ثابتة، أي إما واحد صحيح أو صفر، مثل مفردات الاختيار

من متعدد أو مفردات الصواب والخطأ، بمعنى أن إجابة الفرد على أي بند في الاختبار أو الاختبار سوف

يمنح لها إما الواحد صحيح أو الصفر.

وتتسم طريقة كيودر ورتشاردسون بعدم تنصيفها للاختبار إلى نصفين فقط بل تجزئه إلى عدد كبير من

الأجزاء، حيث يتكون كل جزء من بند واحد فقط من بنود الاختبار، ويقدر ما يكون هناك اتساق بين هذه

الأجزاء (البنود) بقدر ما يكون تقدير ثبات الاختبار كله جيد، معنى ذلك أن الثبات هنا يفترض منذ البداية

أن الاختبار أحادي البعد، أي يقيس خاصية واحدة، وأن كل بنوده تقيس هذه الخاصية الوحيدة ولا تقيس

غيرها (بشير معمرية، 2007)

وقد وضع كيودر ورتشاردسون عددا من المعادلات التي تعالج هذه الحالة، وتقوم جميعها على حساب

الاتساق الداخلي بين بنود الاختبار وفقا للاعتبارات المقدمة. ومن أهم هذه المعادلات:

▪ الصيغة (20) لكيودر و رتشاردسون (Richardson & Kuder Formula 20):

تستخدم الصيغة (20) إذا كانت درجات المفردات ثنائية (صفر، واحد)، وتوافرت الشروط السالفة الذكر،

ونصها كما يلي:

$$R_k = \left(\frac{N}{1 - N} \right) \left(\frac{E^2 K - M S S}{E^2 K} \right)$$

حيث أن:

ر ك: ثبات الاختبار كله

ن : عدد بنود الاختبار

ع²ك: تباين الدرجة الكلية على الاختبار.

س: نسبة الإجابات الصحيحة على البند

ص نسبة الإجابات الخاطئة على البند. (بشير معمري، 2007)

أي أن: ص = 1 - س

مج س ص مجموع تباين درجات مفردات الاختبار.

▪ الصيغة (21) لكيودر و رتشاردسون (Richardson & Kuder Formula 21): (بوسالم عبد العزيز، 2014)

يتطلب استخدام الصيغة (21) توافر شرط آخر إلى جانب الشروط الأخرى المطلوبة في الصيغة (20)، وهذا الشرط هو تساوي جميع بنود الاختبار في درجة صعوبتها، أو على الأقل أن يقترب متوسط درجة صعوبة جميع المفردات من (0.05)، ولكن نظرا لصعوبة توافر هذا الشرط في الكثير من الاختبارات والمقاييس فإن الصيغة (20) ربما تكون أكثر ملاءمة لها.

وهذه الصيغة مشتقة جبريا من الصيغة (20)، وهي كالتالي:

$$رك = \left(\frac{س}{ن} - 1 \right) \left(\frac{ن}{ن - 1} \right)$$

حيث:

س: متوسط الدرجات الكلية في الاختبار.

ن: عدد بنود الاختبار.

نلاحظ مما سبق أن الصيغتين اللتين تم التطرق إليهما تستخدمان في المقاييس ثنائية الدرجة، أما إذا كانت استجابات الاختبار متدرجة على ميزان من موازين التقدير (مثل اختبارنا الحالي الذي تتدرج استجاباته على سبعة بدائل)، أو مقاييس الاتجاهات والميول واستطلاع الرأي أو بعض مقاييس الشخصية التي يستجيب الفرد لعباراتها على ميزان ثلاثي أو خماسي التدرج أو تدرج آخر، عندئذ لا يمكن استخدام

أي من صيغتي كيودر وريتشاردسون في تقدير ثبات درجات هذه المقاييس والاختبارات، لأنها تتطلب وجود معطيات في شكل إجابات.

▪ معامل الثبات (α) لكرونباخ (cronbach's alpha):

يعتبر معامل ألفا لكرونباخ (1951) والذي يرمز له عادة بالحرف اللاتيني (α) من أهم مقاييس الاتساق الداخلي للاختبار، ومعامل ألفا يربط بين ثبات الاختبار ككل بثبات بنوده، فازدياد قيمة تباينات البنود بالنسبة إلى التباين الكلي يؤدي إلى انخفاض معامل الثبات، وانخفاضها (تجانسها) يؤدي إلى ارتفاع معامل الثبات (بشير معمرية، 2007).

حيث قام كرومباخ باشتقاق صيغة عامة من الصيغة (20) السابقة الذكر، لتقدير ثبات درجات أنواع الاختبارات والمقاييس المختلفة، ويسمى أيضا معامل التجانس، حيث أن معامل (α) كالتالي:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum c^2}{E^2} \right)$$

حيث أن:

E^2 : تباين درجات كل مفردة من مفردات الاختبار.

$\sum c^2$: مجموع تباين درجات جميع المفردات.

ن: عدد المفردات. (بوسالم عبد العزيز، 2014)

وقد أشار صلاح علام (2002) في هذا الصدد بأن تقدير الثبات باستخدام معامل (a) لكرومباخ نحصل من خلاله على أقل ثبات ممكن في حالة ما إذا استخدمنا معادلات أخرى، ويعني ذلك أنه إذا كانت قيمة معامل (a) لكرومباخ عالية فذلك يعني أنها عالية بالفعل، أما إذا كانت منخفضة فإنه ستكون أكبر من ذلك لو استخدمنا معادلات أخرى (صلاح علام، 2002).

وتجدر الإشارة إلى أن المعادلات المستخدمة في تقدير ثبات الاختبار متعددة إذا ما كانت الإجابة متعددة لكنها تتفق في كونها قائمة على تحليل التباين والتحليل العاملي.

وأيا كانت طريقة تقدير معامل الثبات فإنها تبقى قيم تقديرية تتأثر بالأخطاء العشوائية وعوامل متعددة أخرى يجب مراعاتها عند تصميم الاختبار.

▪ العوامل المؤثرة على ثبات الاختبار:

بعد أن تطرقنا لمختلف الطرق المستخدمة في تقدير الثبات، ومجال وحدود استعمال كل طريقة منها، ينبغي أن نبين كذلك العوامل التي تؤثر في ثبات الاختبار، حتى نتمكن من التحكم فيها، وخفض تدخلها في الثبات.

ويمكن عرض هذه العوامل كما أوردها معمريّة (2007) كالتالي:

- **طول الاختبار:** يقصد بطول الاختبار زيادة عدد فقراته، حيث يفترض أن زيادة عدد الفقرات يتيح تغطية أكبر للمحتوى، وهذا يحقق صدق المحتوى (ابراهيم محاسنة، 2013)؛ كما لاحظنا من خلال معادلة سبيرمان - براون لتصحيح الطول، أن العامل الأول الذي يؤثر في ثبات الاختبار هو طوله، أي عدد البنود التي يتكون منها، فكلما زاد عدد البنود ارتفع الثبات، والمنطق وراء ذلك هو أن العدد الكثير من البنود يؤدي إلى الحصول على عينة أكبر من السلوك (الخاصية)، وكلما حصلنا على عينة أكبر من السلوك كلما كان من المتوقع أن يمثل القدرة أو السمة محل القياس في مرتي التطبيق أو نصفي الاختبار.

- **الاستقلال بين بنود الاختبار:** ويعني الاستقلال هنا أن لا تؤدي إجابة معينة على بند إلى إجابة على بند آخر، مثل أن تؤدي الإجابة بالنفي أو بالإثبات على بند معين إلى الإجابة بنفس الطريقة على بنود أخرى، لأن هذا يؤدي ضمناً إلى انخفاض عدد البنود، ويقلل من الفروق بين الأفراد، مما يؤدي إلى انخفاض معامل الثبات.

- **زمن الاختبار:** يؤثر الزمن المحدد للإجابة على الاختبار بشكل مباشر على الثبات، فمقاييس واختبارات السرعة أو الموقوتة، تكون معاملات ثباتها مرتفعة، مقارنة بالمقاييس التي تمنح متسعاً من الوقت، وعلى ذلك ينبغي على مصمم الاختبار أن يحدد الوقت المناسب للإجابة، دون أن يعطي متسعاً من الوقت للضعفاء في الإجابة حتى ولو كان ليس موقوتاً.

- **تجانس العينة:** يؤدي التجانس الشديد لعينة الأفراد التي يحسب لها الثبات من خلال أدائها إلى

انخفاض ملموس في معامل ثبات الاختبار، لأن التباين داخل هذه العينة المتجانسة يكون منخفضاً بقدر

لا يسمح بتقدير التباين الحقيقي للاختبار أي ثباته، فإذا حسبنا على سبيل المثال ثبات اختبار على عينة

من العمال المتجانسين في العمر والسن والجنس...، فمن المتوقع أن يكون معامل الثبات هنا أكثر

انخفاضاً مما لو حسبناه على عينة أقل تجانساً (يتنوع أفرادها من حيث خصائصهم).