دراسة الخصائص السيكومترية لبعض المقاييس والاختبارات فى عينة من البحوث النفسية

دكتور إسماعيل الفقى

كلية التربية - جامعة عين شمس

يهتم الباحثون فى مجالات علم النفس المختلفة بالحصول على معلومات وبيانات دقيقة حول الظواهر السلوكية ، وفى سبيل الحصول على تلك البيانات أو المعلومات يستخدم الباحث عدداً من الأدوات السيكولوجية (المقاييس والاختبارات وقوائم الملاحظة وقوائم التقدير الذاتى Self-Report ومقاييس التقدير Rating Scales) .

ولابد أن تتوافر لهذه الأدوات بعض الخصائص السيكومترية مثل الصدق والثبات والموضوعية والتقنين ومدى الملاءمة لموقف القياس ، وذلك حتى يتمكن الباحث من الحصول على معلومات أو بيانات دقيقة حول الظاهرة موضع القياس فالغرض الأساسى من عملية القياس هو الحصول على معلومات دقيقة تبنى عليها احكام تقويمية يستفيد منها الباحث فى اتخاذ القرارات .

واهتم كثير من علماء النفس فى مجال القياس النفسى ونظرياته المختلفة بهذه الخصائص السيكومترية . فهناك دراسات اهتمت باختبار النماذج المتعددة لحساب الثبات أو طرق حساب الصدق . أو المقارنة بين المعادلات المختلفة فقد قام Peter A. L Lori (1990) بدراسة تناولت المعايير النفسية لصدق وثبات الأدوات المختلفة التى يستخدمها المدرسون ، وذلك كما حددتها الجمعية الأمريكية للبحوث التربوية American Association For Educational Research (AAEA) والمجلس القومى للقياس التربوى National Council of Measurment in Education (NCME) ورابطة علم النفس الأمريكية American Psychological Association (APA) وركزت الدراسة على ثلاثة مؤشرات اعتبرتها كافية للوثوق فى نتائج الاختبارات ، وهذه المؤشرات هى صدق التكوين الفرضي Construct validity والصدق المرتبط بالمحك Criterion -related validity. والثبات Reliability.

وقام Kathy (1978) بدراسة للمقارنة بين صدق وثبات الاختبار من نوع الاختيار من متعدد Multiple-Choice والصواب والخطأ أو البديلين True-False وتناولت هذه الدراسة أثر كل من صيغة الاختيار، ونمط استجابة المفحوص على الخصائص السيكومترية للاختبار (الصدق ، الثبات) . وفى مجال القياس المرجعى المحك Criterion-Refrenced Measurment قامت كل من Susan & Mary (1983) بدراسة مقارنة لتقدير ثبات الاختبارات المرجعة إلى المحك من خلال تطبيق الاختبار لمرة واحدة ولمرتين . واشارت

الدراسة إلى العوامل التى تؤثر على ثبات هذه الاختبارات ، وهى : طول الاختبار ، ودرجات القطع Cutting Scores ، وقدرات التلاميذ ، وحجم العينة ، ومدى تجانس محتوى الإختبار.

ويقدم John (1971) نموذجاً تقويميا لقياس الاتقان An Evaluation Model For Mastry Testing ، ركز فيه على اخطاء القياس الشائعة وطرق تحليل المفردات فى اختبارات الاتقان وطرق حساب صدق وثبات هذه الاختبارات وذلك باعتبار أن قياس صدق وثبات اختبارات الاتقان هو من الحالات الخاصة التى يجب أن تؤخذ فى الاعتبار .

واقترح Subkoviak (1980) تكنيكاً لحساب ثبات الاختبارات المرجعة للمحك وذلك من خلال تطبيق الاختبار لمرة واحده ولا يتأثر هذا التكنيك فى حساب الثبات بقلة حجم التباين فى درجات الأفراد على الاختبار .

هذا عن الاختبارات المرجعة إلى المحك كحالة خاصة فى حساب الثبات والصدق . وهناك العديد من تلك الحالات الخاصة مثل قياس الابتكار Creativity Testing وقياس سمات الشخصية باستخدام الاختبارات الاسقاطية Projective Tests ، واختبارات السرعة واختبارات القوة.

وتهتم الدراسة الحالية ببعض الخصائص السيكومترية لأدوات القياس النفسى وذلك من خلال فحص ومراجعة الاختبارات والمقاييس النفسية المتضمنة فى بعض البحوث والدراسات المنشورة بمجلة الجمعية المصرية للدراسات النفسية *منذ عام ١٩٨٥ حتى عام ١٩٩٩م .

ويرى الباحث أن القاء الضوء على تلك البحوث والدراسات السيكولوجية المنشورة فى هذه المجلة وما أستخدم فيها من اختبارات ومقاييس ، ومدى مناسبة الطرق الاحصائية المستخدمة للتحقق من الخصائص السيكومترية لتلك الأدوات ، ومدى مناسبة حجم عينة البحث أو الدراسة ، وغير ذلك من الإجراءات السيكومترية التي يعتبرها الباحث بمثابة نظرة تقويم ذاتى Self Evaluation لتلك الاجراءات المتعلقة بعملية القياس فى بحوثنا النفسية فى مصر وهو بمثابة نظره تقويمية نستطيع من خلالها تشخيص مواطن القوة فندعمها ، ومواطن الضعف فنعالجها في الأدوات المستخدمة فى البحوث النفسية .

وبذلك فإن مشكلة الدراسة الحالية تتمثل فى السؤال الرئيسى التالى :

ما مدى ملاءمة الأدوات المستخدمة (المقاييس والاختبارات) فى بعض الدراسات المنشورة فى مجلة الجمعية المصرية للدراسات النفسية منذ العام (١٩٨٥-١٩٩٩) لموضوع القياس؟ ويمكن أن يتفرع من هذا التساؤل الرئيسى الأسئلة الآتية :

أولاً : فيما يتعلق بإعداد الأدوات :

هل المقاييس أو الاختبارات المستخدمة فى الدراسة من اعداد الباحث ؟ أم من اعداد باحث آخر ؟

* مجلة علمية محكمة تصدرها الجمعية المصرية للدراسات النفسية .

هل المقاييس أو الاختبارات المستخدمة فى الدراسة مترجمة ، أم تم اعدادها دون الاعتماد على المقاييس الأجنبية ؟

ثانياً : فيما يتعلق بمجال القياس :

ما هى مجالات القياس التى تستخدم فى قياسها أدوات الدراسة (المجال المعرفى ، المجال الوجدانى ، المجال المهارى – النفسحركى) ؟

ثالثاً : عدد المفردات ونمط الاستجابة :

هل هناك اشارة إلى عدد المفردات التى يتكون منها المقياس أو الاختبار ؟

هل هناك اشارة إلى نمط أو طريقة الاستجابة على المقاييس أو الاختبارات المستخدمة؟

رابعاً : الثبات :

ما هى الطرق المستخدمة لحساب ثبات المقاييس والاختبارات ؟

ما مدى مناسبة تلك الطرق لطبيعة تلك المقاييس والاختبارات ؟

خامساً : الصدق :

ما هى الطرق المستخدمة لحساب صدق المقاييس والاختبارات ؟

ما مدى مناسبة تلك الطرق لطبيعة تلك الأدوات ؟

وبذلك فإن الباحث من خلال عرضه لمشكلة الدراسة على النحو السابق، يأمل أن يتناول تلك المفاهيم السيكومترية الخاصة بأدوات القياس النفسى ، تناولاً نظرياً دقيقاً وشاملاً يستعرض من خلاله التأصيل النظرى لتلك المفاهيم – وبخاصة مفهومى (الثبات – الصدق) بدءاً من التعرض لماهية كل مفهوم فى مدارس ونظريات القياس المختلفة وانتهاءً بالوقوف على الجديد والحديث من مفاهيم القياس النفسى . ويرى الباحث أن تناول هذه المفاهيم المختلفة من خلال الدراسات التطبيقية يساعد فى الوقوف على مدى وضوح استخدام كل من هذه المفاهيم فى الدراسات والبحوث التطبيقية استخداماً صحيحاً .

وهناك العديد من الدراسات التي أجريت في هذا المجال ، فقد قام عبد العاطى الصياد (١٩٨٣) بدراسة العينات وأدوات القياس عرض فيها لكيفية اختيار العينة فى البحوث والدراسات الإنسانية والاجتماعية ، وأنواع العينات ، ثم عرض لأدوات القياس (الاستفتاء) وأنواعه وخصائصه ، وقام بالمقارنة بين الاستفتاء والمقابلة . وقام عبد العاطى الصياد (١٩٨٩) فى دراسة أخرى حول نفس الموضوع (جداول تحديد حجم العينة فى البحث السلوكى) حاول من خلالها تقديم حلاً لمشكلة العينات فى البحوث التربوية والنفسية والاجتماعية سواء كانت هذه البحوث وصفية أو مسحية أو تجريبية ، وقد حدد مشكلة البحث

فى سؤال رئيسى : ما هو حجم العينة الضرورى لاجراء البحث ؟ وللإجابة عن هذا التساؤل يرى أن هناك مدخلين رئيسيين يمكن من خلالهما تحديد حجم العينة :

المدخل الأول : قدمه كوهين Chohen (1977) وهو ما يسميه مدخل اختبار الفروض (التجريبى أو شبه التجريبى) .

المدخل الثانى : وهو ما يسميه مدخل البحث المسحى أو الوصفى أو مدخل تقدير المعالم Parameters Estimation فى تحديد حجم العينة . وقد اثبت Brewer (1978) أن هذين المدخلين يمكن أن يعطيا تقديرين مختلفين لحجم العينة وذلك فى حالة اختبار الفرض الإحصائى : متوسط المجتمع (م) يساوى كمية معينة (م) . وبالتالى حدد مشكلة دراسته فى كيفية العثور على طريقة موضوعية مبسطة سهلة الاستخدام (فى صورة جداول)تمكن كلا النوعين من الباحثين ؛ العارفين بالتصميم الاحصائى (التجريبى أو شبه التجريبى) ، أو أولئك الذين لا يعرفون ذلك التصميم (المسحيين أو الوصفيين) من تحديد حجم العينة الضرورى لاجراء بحوثهم ، وقدم الباحث عددا من الجداول التى تساعد الباحثين فى تحديد حجم عينة بحوثهم . وبذلك يكون الباحث قد حدد بشكل اجرائى ما قد سبق أن انتقده فى دراسة سابقة له (١٩٨٨) حول صياغة معظم كتب مناهج البحث حين تناولت موضوع العينات نوعاً واختباراً وتحديداً لحجمها، حيث يرى أن الصياغة فى الغالب صحيحة ولكنها تفتقر إلى الاجرائية واعطى مثالاً على ذلك . . يزداد حجم العينة بزيادة التباين فى المجتمع ، يتناقص خطأ الصدفة بزيادة حجم العينة ، حجم العينة يتحدد بعدة عوامل منها طبيعة المجتمع الأصل ونوع الاختبار الإحصائى ودرجة الدقة المطلوبة ، وكلما زاد حجم العينة كلما زاد ثبات المقياس الاحصائى المسحوب منها ، . . . وعينة حجمها أقل من ٣٠ تعد عينة صغيرة ، ويمكن أن يبلغ حجم العينة ٥ ٪ أو ١٠ ٪ أو ٢٠ ٪ من حجم المجتمع ، . . . وغير ذلك من العبارات التى يرى أنها صحيحة تماماً ولكنها غير اجرائية (عبد العاطي الصياد ١٩٨٨) .

ويقدم عبد الله عبد الغنى وعبد العاطى الصياد (١٩٨٩) سلسلة دراسات حول طريقة امبريقية مقترحة لتقدير وتقويم صدق المحكمين . حيث يرى الباحثان ،أن حساب معامل اتفاق عام بين المحكمين يمثل محك قبول أو رفض العنصر لقياس المحتوى موضع التحكيم . وهذا المعامل قد يكون مضللاً لأسباب أهمها أن المحكم الفرد قد لا يكون متسقاً مع نفسه حينما حكّم العناصر التى تقيس المحتوى موضع التحكيم ، وحتى إذا كان متسقاً مع نفسه فقد لايكون متسقاً مع بقية المحكمين حول عناصر المقياس الضرورية أو فى مجموعها أضف إلى ذلك أن المحكمين أو بعضاً منهم قد يتفقون ولكن اتفاقهم قد يجئ حول القرار الخـــاطئ

بالنسبة لقبول أو رفض العنصر ضمن المقياس،

كما يرى الباحثان أن هذه الطريقة المقترحة تؤدى إلى كشف المحكم أو المحكمين الذين لا يتسقون مع أنفسهم أو مع بعضهم البعض قبل حساب نسبة الاتفاق حول صلاحية العنصر المعين فى قياس محتوى معين ، كذلك حساب نسبة الاتفاق (أو متوسط الاتفاق) بين المحكمين الذين يتميزون بدرجة مقبولة من الإتساق مع أنفسهم كفرادى أو كمجموعات . وفى اطار ما اسماه بمشكلات الاستجابة Problems of response ، والعوامل المتوقع حدوث تأثير لها على الاختبارات والمقاييس النفسية مثل الوقت المحدد Time Limits، والتخمين Gusesing ونمط الاستجابة Response style والتعب Fatigue ، والدافعية Motivation ،...

قام عبد العاطى الصياد (١٩٩٠) بدراسة للكشف عن هل معرفة اسم المستجيب من عدمه لها تأثيرها الجوهرى احصائياً وعملياً على قيمة المعامل (الفا) لكرونباخ ؟؟ وتوصل فى نتائج دراسته أنه عند النص فى تعليمات الاختبار أو المقياس على عبارة «اكتب اسمك اذا رغبت» فإنها تبعث فى المستجيب الطمأنينة ليمارس حريته فى التعبير عن استجابته الحقيقة مما ينعكس ايجابياً على ثبات هذه الأدوات .

وفى دراسة على حامد الثبيتى (١٩٩٨) تناول ثبات أدوات القياس فى العلوم الإنسانية بين النظرية التقليدية للاختبار ونظرية التعميم ، ودور ثبات أدوات القياس فى رفع مستوى القدرة التفسيرية للبيانات . وتكونت عينة الدراسة من (٥٠) دراسة منشورة فى بعض الدوريات العربية المحكمة التى تصدرها بعض كليات التربية بالجامعات العربية . واقتصرت العينة على سنتين فقط ، واوضحت نتائج الدراسة أن أكثر من ٥٥ ٪ من المقاييس المعدة باللغة العربية والتى سبق استخدامها فى دراسات سابقة لم يحسب لها أى نوع من أنواع الثبات مما يقلل بطبيعة الحال من قيمتها التفسيرية . كما لوحظ أن ٣٣ ٪ من هذه المقاييس تم حساب معامل ثباتها بطريقة اعادة الاختبار ، وأن ٤٤ ٪ من مجموع هذه المقاييس لم يشر الباحث لمعامل الثبات . وأن هناك تجاوزاً كبيراً فى عملية حساب معامل ثبات المقاييس المستخدمة فى جمع المعلومات فى العلوم الإنسانية .

وفيما يلى يعرض الباحث عرضاً موجزاً لمفهومى الثبات Reliability والصدق Validity .

أولاً : الثبات .

سوف يقصر الباحث حديثه عن مفهوم الثبات فى ضوء النظرية الكلاسيكية فى القياس النفسى .

النظرية الكلاسيكية Classical Theory :

طور هذه النظرية شارل سبيرمان (١٩٠٤) Spearman ويطلق عليها أحيانا نظرية الدرجة الحقيقية ودرجة الخطأ Theory of True and Error Scores – وقد تم تطوير هذه

النظرية بواسطة عدد من السكيوميترين المعاصرين ورغم ظهور نماذج حديثة فى القياس فلا زالت النظرية الكلاسكيكية سائدة فى هذا المجال .

تبدأ النظرية من منطلق ان درجة الاختبار تنتج عن تاثير عاملين :

١- العوامل التى تسهم فى الاتساق ؛ وهى تتألف من الخصائص المستقرة للفرد والتى يحاول الفاحص ان يقيسيها.

٢- العوامل التى تسهم فى عدم الاتساق وهى تشتمل على خصائص الفرد والاختبار أو المواقف التى ليس لها دور فى السمة أو الخاصية التى تقيسها ولكنها تؤثر على درجة الاختبار.

وبالطبع فانه من الواضح أن العامل الاول مرغوب لأنه يمثل الكمية الحقيقة من الخاصية موضوع القياس بينما يمثل العامل الثانى التشويش الذى لايمكن تجنبه لعوامل الخطأ التى تسهم فى جوانب عدم الدقة للقياس . ويمكن ان نعبر عن ذلك التصور بالمعادلة التالية :

$$س = ق + خ$$

حيث (س) هى الدرجة الناتجه عن الاختبار ، ق هى الدرجة الحقيقية ، (خ) هى درجة الخطأ أو أخطاء القياس وبذلك فان اخطاء القياس ودرجة الخطأ ، تمثل الفروق بين درجة الاختبار والدرجة الحقيقية للقياس ، أى ان $خ = س - ق$

ونلاحظ فى هذه المعادلة أن أخطاء القياس (خ) قد تكون موجبة أوسالبة . فاذا كانت خ موجبة فإن الدرجة التى نحصل عليها (س) ستكون اكبر من الدرجة الحقيقية (ق) وبالعكس اذا كانت (خ) سالبة فان الدرجة المتحصل عليها (س) ستكون أقل من الدرجة الحقيقية (ق)

وعلى الرغم من أنه قد يستحيل أن نلغى كل خطأ القياس فإن واضعى الاختبارات يطمحون إلى أن يقللوا هذا التشويش السيكوميترى من خلال الانتباه الواعى لمصادر خطأ الاختبار المذكورة.

وما ينبغى أن نعرفه هو أن الدرجة الحقيقية لا يمكن معرفتها على الاطلاق ، وكل ما يمكن أن نصل اليه هو تحديد احتمالية أن تقع هذه الدرجة الحقيقة فى مدى معين وأنه من الممكن أن نحصل على أحسن تقدير للدرجة الحقيقة ولكن من المستحيل أن نعرف قيمتها بالتحديد (89 - 88 - Kaplan & Saccuzo (1989 pp.

مصادر خطأ القياس :

كما تدل عليه المعادلة : $س = ق + خ$

فإن خطأ القياس هو كل شئ بخلاف الدرجة الحقيقية ، ويمكن أن تنشأ أخطاء القياس

من عدد غير محدود من المصادر ويقدم ستانلى(1971) Stannly قائمة شاملة بمثل هذه الأخطاء وسوف نعرض فقط لأهمها وهى : اختيار المفردات ، تطبيق الاختبار، تصحيح الاختبار ، والأخطاء المنتظمة للقياس .

اختيار المفردات : Item Selection

يعتبر الاختبار نفسه واحدا من مصادر خطأ القياس . فمعد الاختبار يختار عددا محدودا من المفردات من بين مجموعة من المفردات ، ويتم تحديد أي منها يتضمنها الاختبار ويعتبر هذا أمراً اساسياً فى دقة القياس وكيفية صياغة هده المفردات .

والاختبار الذى أحسن إعداده فان خطأ القياس الناتج عن الحصول على عينة من المفردات يكون أدنى ما يمكن . والاختبار هو فى الواقع عينة من السلوكيات أو المعرفة المتصلة بالخاصية التى نقيسها . ونتيجة لذلك فان اختيار المفردات يعتبر دائما مصدراً لخطأ القياس فى الاختبارات النفسية وأقصى ما يمكن لمعد المقياس عمله هو أن يقلل ذلك التشويش غير المطلوب بالانتباه بعناية لقضايا بناء الاختبار .

تطبيق الاختبار: Test Administretion

على الرغم من اهتمام االفاحصين بتوفير أفضل الظروف وكذلك بيئة معيارية للاختبار فإنه تنشأ عدة مصادر لخطأ القياس من ظروف تطبيق الاختبار ومن بين الظروف التى تؤثر على موقف الاختبار ، عدم ملاءمة درجة الحرارة في غرفة الفحص، وضعف الاضاءة ، ووجود ضوضاء كثيرة . وفى بعض الأحيان يكون من الصعب ان نتوقع خصائص موقف الاختبار التى يمكن ان تسهم فى خطأ القياس .

وقد يكون للتذبذبات اللحظية فى مستوى القلق ، والدافعية والانتباه ومستوى التعب لدى المفحوص اسهامات فى خطأ القياس . فالمفحوص الذى يكون مشتتاً بفعل ألم انفعالى مؤقت قد يضع الاجابات فى مكان خطأ أى تحت عمود خطأ فى ورقة الإجابة وبالتالى فقد يترتب على ذلك أن تكون كل الاجابات فى الموقع غير الصحيح .

تصحيح الاختبار :

إن معظم الاختبارات لها نظام تصحيح موحد مما يؤدى إلى تفادى كثير من الأخطاء المترتبة على التصحيح – لكن حينما يعتمد التصحيح على جوانب ذاتية فإنه يمكن أن يشوبه الخطأ – ويمكن أن يحدث هذا فى الاختبارات الاسقاطية وكذلك فى اختبارات المقال .

الخطأ المنتظم للقياس Systematic Measueement Error

أن مصادر عدم الدقة التى سبق الإشارة إليها تسمي خطأ القياس غير المنتظم Unsystematic measurement error مما يعنى أنه لا يمكن التكهن بآثارها كما أن هذه الاثار غير متسقة . على أن هناك نوعاً أخر من خطأ القياس هو الخطأ المنتظم فى القياس

وهذا النوع من الخطأ ينشأ عندما يقيس الاختبار بغير علم من الباحث بشكل متسق شيئاً آخر غير السمة التى أعد ليقيسها . فمثلا مقياس أعد لقياس الانبساطية يقيس أيضا عن غير قصد القلق بشكل متسق . فى هذه الحالة فإن المعادلة التى تصور العلاقة بين الدرجات المشاهدة والدرجات الحقيقة ومصادر خطأ القياس تصبح .

$$س = ص + خ_م + خ_غ$$

حيث(س) هى الدرجة التى نحصل عليها من الاختبار (الدرجة المشاهدة) ،

(ص)هى الدرجة الحقيقة ، خ م هو الخطأ المنتظم الراجع إلى المكون الفرعى للقلق بينما خ غ هو التأثير المجمع لأخطاء القياس غير المنتظمة المذكورة من قبل ولأنها لا يكتشف وجودها فإن الأخطاء المنتظمة للقياس قد تمثل مشكلة جوهرية فى تطوير الاختبارات النفسية ويمكن تقليل أخطاء القياس المنتظمة إذا اتبع معد الاختبار القواعد الفنية لبناء الاختبارات بدقة كافية. لكن أخطاء القياس المنتظمة تذكرنا أنه من الصعب إن لم يكن من المستحيل أن نقدر سمة من السمات فى صورة نقية بمعزل عن السمات الأخرى .

خطأ القياس وثبات الاختبار :

لكن ما هى العلاقة بين خطأ القياس وبين ثبات الاختبار؟فى الواقع فإن الرابطة الواضحة بين هذين المتغيرين هى أن خطأ القياس يقلل من ثبات أو مدى تكرارية نتائج الاختبار النفسى . فالثبات يحمل فى طياته علاقة احصائية دقيقة مع خطأ القياس . فهما مجرد أسلوبان مختلفان للتعبير عن شئ واحد : ما مدى اتساق الاختبار ؟ والعلاقة المتداخلة بين هذين المفهومين سوف تتضح مع عرض النظرية الكلاسيكية للقياس .

ان احد الافتراضات العامة فى النظرية الكلاسيكية هو أن اخطاء القياس غير المنتظمة تعمل كمؤثرات عشوائية .

أسباب خطأ القياس تعتبر متعددة ومتنوعة للغاية وعلي هذا فان اخطاء القياس غير المنتظمة تسلك مثل المتغيرات العشوائية وتعتبر النظرية الكلاسيكية لقياس تلك العشوائية الاساسية لأخطاء القياس كافتراض اساس . ولان الأخطاء غير المنتظمة أي أن للقياس أحداثا عشوائية فان أمامها فرصة متكافئة أن تكون ذات قيمة موجبة أو سالبة وبذلك ستصل فى متوسطها مع الاعداد الكبيرة إلي صفر . وبذلك فان الافتراض الثانى هو ان متوسط خطأ القياس = صفر .

كذلك فان النظرية الكلاسيكية للقياس تفترض أن أخطاء القياس لا ترتبط مع الدرجات الحقيقية وهذا يعطينا حسا حدسيا : اذا كانت درجات الخطأ مرتبطة بدرجة أخرى كان من الممكن ان توحى بانها تكون منتظمة اكثر من كونها عشوائية .

ويمكن أن نلخص الخصائص الاساسية للنظرية الكلاسيكية للقياس على النحو التالى:

(١) أن اخطاء القياس عشوائية .

(٢) متوسط خطأ القياس = صفر .

(٣) الدرجات الحقيقية ودرجات الخطأ ليس بينها ارتباط أى $ر_{ح خ}$ = صفر

(٤) الاخطاء على المقايس المختلفة ليس بينها ارتباط اى انها مستقلة عن بعضها أى ان $ر_{خ١،خ٢}$ = صفر .

اذا بدأنا بهذه الافتراضات يمكن أن نطور عددا من الاستخدامات العامة للثبات والقياس . (النقاط التالية تقوم على أساس افتراض أن اخطاء القياس المنتظمة قليلة الى ادنى درجة أو غير موجودة بالنسبة لأداة القياس) . ومن المعروف أن أى مقياس يطبق مع مجموعه كبيرة من المفحوصين سيظهر لنا تباينا كبيرا فى الدرجات التى نحصل عليها ويمكن التعبير عن هذا التباين والاختلاف فى صورة تباين $ع^{٢}$ وأهمية نظرية القياس هى انها تسمح لنا بتقسيم التباين الخاص بالدرجات المشاهدة الى مصدرين منفصلين وبصفة خاصة يمكن اثبات ان تباين الدرجات المشاهدة هو ببساطة تباين الدرجات الحقيقية مضافا اليه تباين اخطاء القياس وذلك فى اطار كون الدرجات الحقيقية مستقلة عن درجات الخطأ وأن الارتباط بينهما = صفر .

$$ع^{٢}_{س} = ع^{٢}_{ح} + ع^{٢}_{خ}$$

هذه المعادلة توضح ان درجات الاختبار تتباين نتيجة لعاملين هما : الاختلاف فى الدرجات الحقيقية (ح) ، والاختلاف الراجع لخطأ القياس . والتطبيق الواضح لهذه العلاقة هو ان أخطاء القياس تسهم فى عدم اتساق الدرجات المشاهدة وان النتائج لن تكون مستقرة اذا اعيد تطبيق الاختبار مرة أخرى (Gregory, J. Robert (1996 pp 88 - 90) .

معامل الثبات : The Reliability Coefficient

بعد ان عرضنا للافتراضات التى تقدمها النظرية الكلاسيكية للقياس وامكانية تقسيم تباين درجات الاختبارات الى مصدرين . تباين خاص بالدرجة الحقيقية ، وتباين خاص بخطأ القياس أصبح واضحا أن ثبات الاختبار يعبر عن التأثير النسبى للدرجات الحقيقية ودرجات الخطأ على درجات الاختبار المشاهدة وفى صورة رياضية فان معامل الثبات $ر_{سس}$ عبارة عن نسبة تباين الدرجة الحقيقية الى تباين الدرجة الكلية أى أن :

$$ر_{سس} = \frac{ع^{٢}_{ح}}{ع^{٢}_{س}}$$

وبالتعويض عن $ع^{٢}_{س}$

$$\therefore \quad ر_{سس} = \frac{ع^{٢}_{ح}}{ع^{٢}_{ح} + ع^{٢}_{خ}}$$

ويلاحظ ان مدى القيم الممكنة التى تأخذها ر س س يمكن اشتقاقها من تحليل المعادلة المذكورة أعلاه . وفي الحالة التى يصغر فيها التباين الراجع الى درجة الخطأ (ع٢خ) إلي ما يقرب به من الصفر ، فى هذه الحالة فان معامل التباين ر س س يقرب من قيمة $\frac{ع^{2}_{د} - ع^{2}_{خ}}{ع^{2}_{د}}$ اى من الواحد الصحيح وعلى الطرف الاخر اذا زاد التباين الراجع الى خطأ القياس زيادة كبيرة جداً فإن قيمة معامل الثبات تصبح أصغر وبذلك فإنها تقرب من الحدود النظرية للصفر .

والخلاصة ان الاختبار الذى ينخفض ثباته جدا (خطأ قياس كبير) سوف ينتج لنا معامل ثبات قريب من الصفر بينما الاختبار الثابت تماما حيث لايوجد خطأ قياس ينتج لنا معامل ثبات = +١,٠٠ (واحد صحيح) وبذلك فان المدى الممكن لمعامل الثبات يقع بين صفر، والواحد الصحيح . وفي الواقع العملى فإن كل الاختبارات تنتج لنا معاملات ثبات تقع ما بين الصفر والواحد الصحيح ولكن كلما كانت قيمة هذا المعامل أقرب إلى الواحد الصحيح كان ذلك أفضل .

وبذلك يمكن القول أن المقياس الذى يتمتع بدرجة ثبات عالية هو الذى نقل فيه اخطاء القياس إلى ابعد حد كما أن الثبات ويعبر عنه بالرمز(ر س س) يمكن أن يقع بين صفر ويصل إلى الواحد الصحيح بالنسبة للاختبار التام الثبات Gregory, J. Robert (1996 p. 90) .

طرق حساب معامل الثبات :

عندما نقوم بتطبيق أحد الاختبارات النفسية على عينة من مجتمع فإن الدرجات التى نحصل عليها لكل فرد من أفراد هذه العينة على هذا الاختبار (تعرف بالدرجات المشاهدة) تشتمل بداخلها على الدرجات الحقيقية وعلى جزء من الدرجات المشاهدة الراجعة إلى خطأ القياس لكن ليس بوسع الفاحص أن يفصل بين الدرجة الحقيقية ودرجة الخطأ ولكن من الممكن أن نحصل على معاملات الثبات عن طريق مجموعة من الأساليب المختلفة نعرضها فيما يلى :

١ – طريقة اعادة الاختبار Test- Retest Reliability

هذه الطريقة من طرق حساب معامل الثبات تعتبر طريقة مباشرة – وفيها تستخدم نفس الصورة للاختبار مع نفس المفحوصين فى زمنين مختلفين حيث يطبق الاختبار فى الزمن (أ) ثم يعاد تطبيق نفس الاختبار على نفس المفحوصين فى زمن لاحق – أى بعد مرور فترة زمنية مناسبة : اسبوعين – شهر – شهرين .... إلخ فإذا كان الاختبار يتمتع بثبات كامل فإن كل مفحوص فى التطبيق الثانى سيحصل على نفس الدرجة التى حصل عليها فى التطبيق الأول (أو قريبا منها) وبمعنى آخر فإن درجة الاختبار الثانى لكل شخص

يمكن التنبؤ بها تماما من درجته فى الاختبار الأول – وفى الواقع العملى يمكن أن نتوقع فى كثير من أنواع الاختبارات مثل اختبارات القدرات واختبارات التحصيل أن تكون درجات المفحوصين فى التطبيق الثانى أعلى من درجاتهم فى التطبيق الأول بسبب عوامل الممارسة والنضج وغيرها من المؤثرات التى تقع بين التطبيقين . وفى هذه الطريقة لتقدير الثبات يؤخذ معامل الارتباط بين درجات المفحوصين علي الاختبار فى التطبيق على انه معامل للثبات ويسمى معامل الاستقرار .

٢- **طريقة الصور البديلة (المكافئة)** Alternet (Equivalent) forms Reliabilty.

فى بعض الحالات يقوم الباحث باعداد صورتين من نفس الاختبار . هذه الصور المتبادلة تبنى بشكل مستقل لتقابل نفس المواصفات . والصور البديلة لاختبار يكون لها عادة خصائص احصائية ومعيارية مماثلة وعلى سبيل المثال اذا طبقت الصور البديلة فى موقف مماثل مع نفس المجموعة من الحالات فان المتوسطات والانحرافات المعيارية للصور البديلة تكون متقاربة ويتم تقدير معامل الثبات فى هذه الطريقة بتطبيق صورتي الاختبار لنفس مجموعة المفحوصين وحساب معامل الارتباط بين الدرجات فى التطبيقين ويسمى معامل الثبات فى هذه الطريقة بمعامل التكافؤ . وهذه الطريقة تشترك فى جوانب كثيرة مع طريقة اعادة الاختبار حيث تشتمل الطريقتان على تطبيقين لنفس مجموعة المفحوصين مع وجود فترة بين التطبيقين .

ورغم التشابه فى التصور النظرى للطريقتين فإن هناك اختلافا بين الطريقتين فطريقة الصور البديلة تقيس عينة مختلفة من مفردات الاختبار والفقرات كمصدر اضافى لتباين الخطأ وبذلك فإن بعض المفحوصين قد يكون أداؤه أفضل أو أسوأ علي احدى صور الاختبار لاختلاف المفردات المنتقاة له . فرغم أن الصورتين قد تكونا متساويتين فى الصعوبة ، فإن بعض المفحوصين قد يجدوا احدى الصورتين اصعب قليلا من الثانية (أو اسهل) بسبب كون المفردات المتقابلة ليست متساوية فى مدى الألفة بها لكل مفحوص . وكما يلاحظ فإن الفروق فى معاينة المفردات ليست مصدرا لتباين الخطأ فى طريقة الاختبار – اعادة الاختبار نظرا لأنه فى طريقة اعادة الاختبار تستخدم نفس المفردات فى كلا التطبيقين .

٣- **الاتساق الداخلى** : Internal Consistency

وهنا يتم حساب الثبات باعتباره اتساقا داخلياً . حيث يتم حساب معامل الارتباط بين نصفى الاختبار فيما يعرف بطريقة التجزئة النصفية ، وكذلك الطرق التى تتعامل مع مفردات الاختبار والعلاقة بينها حيث يتم حساب ثبات الاختيار باستخدام بعض المعادلات مثل **معادلة سبيرمان براون** Spearman -Brown Formula

$$\text{ر}_{\text{سن سن}} = \frac{\text{ن ر ع ع}}{\text{ر ع ع} + (\text{ن} - 1)\ 1}$$

حيث ($\text{ر}_{\text{سن سن}}$) = ثبات الاختبار كل

(ر ع ع) = ثبات الاختبار فى صورته المختزلة أو المطولة .

(ن) = عدد الفقرات فى الاختبار المصحح بالنسبة لعدده فى الجزء المسحوب منه أى عدد مرات مضاعفة أو اختزال الاختبار .

وفي حالة حساب ثبات الاختبار بطريقةمن التجزئة النصفية نستخدم صورة خاصة من الصورة العامة لمعادلة سبيرمان براون هى :

$$\text{ر}_{\text{سن سن}} = \frac{\text{٢ ر}_{\frac{1}{2}\ \frac{1}{2}}}{\text{ر ع ع} + (\text{ن} - 1)\ 1}$$

أ -طريقة رولون للتجزئة النصفية :

اقترح رولون (1939) Rulon طريقة لحساب الثبات باستخدام طريقة التجزئة النصفية تتطلب حساب تباين الفرق بين درجتى الفرد على نصفى الاختبار ($\text{ع}^{٢}_{\text{ف}}$) وتباين الدرجة الكلية ($\text{ع}^{٢}_{\text{س}}$) والتعويض بهما فى المعادلة الأتية :

$$\text{ر}_{\text{سن سن}} = \frac{1 - \text{ع}^{٢}_{\text{ف}}}{\text{ع}^{٢}_{\text{س}}}$$

وقد اقترح (جتمان)صورة مبسطة لمعادلة رولون يتم من خلالهاحساب الثبات باستخدام تباين كل من الدرجات على الاختبار والتباين على الدرجة الكلية وهذه المعادلة هى:

$$\text{ر}_{\text{سن سن}} = ٢ \left[1 - \frac{\text{ع}^{٢}_{\text{م}} + \text{ع}^{٢}_{\text{ب}}}{\text{ع}^{٢}_{\text{س}}} \right]$$

ب: معادلة كيودر-ريتشاردسون .

اقترح كيودر -رديتشاردسون مجموعة من المعادلات لحساب الثبات واعطاها ارقاما مسلسلة - وقد اشتهرت منها الصورة رقم ٢٠ والمعروفة اختصاراً بـ K-R 20 . وتستخدم هذه

اذا كانت مفردات الاختبار تصحح بنعم أو لا أى باستخدام الواحد والصفر .

$$\text{ك} - \text{ر} ٢٠ = \frac{\text{ن}}{\text{ن} - ١} \left(١ - \frac{\text{مج ف}}{\text{ع}^{٢}\text{س}} \right)$$

جـ : معامل الفاكرونباخ :

وتعتبر معادلة الفا صورة عامة من المعادلة ٢٠ التى اقترحها كيودر-ريتشاردسون حيث يمكن استخدامها للاختبارا ت التى تصحح باستخدام درجات متصلة والصورة الأساسية لمعامل الفا هى .

$$\text{ر}_{\text{س س}} = \frac{\text{ن}}{\text{ن} - ١} \left[١ - \frac{\text{مج ع}^{٢}\text{ق}}{\text{ع}^{٢}\text{س}} \right]$$

حيث ($\text{ر}_{\text{س س}}$)معامل ثبات الاختبار

(ن) عدد فقرات الاختبار

($\text{ع}^{٢}\text{ق}$) تباين الفقرة حيث ق ترمز إلى رقم الفقرة

($\text{ع}^{٢}\text{س}$) تباين الدرجة الكلية

ومن الواضح أن معادلة الفا هى نفسها من الناحية الرياضية معادلة كيودرريتشاردسون ويمكن النظر إليها باعتبارها متوسطا لمعاملات ثبات كل الانصاف الممكنة للاختبار . وتعنى الدرجة العالية لمعامل الثبات أن هذه الانصاف ترتبط فيما بينها ارتباطات عالية وهذا يتحقق إذا كانت مفردات الاختبار تتمتع بدرجة عالية من التجانس بمعنى انها تتجه إلى أن تقيس سمه واحدة مشتركة أما إذا كانت المفردات متعددة العوامل multifactorial فإن كثيرا من توافيق انصاف الاختبار ستشتمل على نسب عالية من احدى السمات على حساب السمات الأخرى التى يقيسها الاختبار . وفى هذه الحالة فإن النصفين اللذين ينقسم اليهما الاختبار سيقيسان خصائص مختلفة وسيكون الارتباط بينهما ضعيفا وبذلك يمكن اعتبار معامل الفا مؤشرا للدرجة التى يقيس بها الاختبار عاملا واحداً - والاختبارات المختلطة Heterogenious التى تقيس أكثر من سمة تنتج لنا قيما منخفضة لمعامل الفا - أما الاختبارات المتجانسة التى تقيس عاملا واحداً فإنها تعطي قيما عالية لمعامل الفا - ولهذا السبب فإن معامل الفا يشار اليه على أنه مؤشر للاتساق الداخلى .

ثبات الاتفاق بين المصححين Inter-Scorer Reliability

بعض الاختبارات تخضع فى تصحيحها لحكم المصحح ، مثل الاختبارات الاسقاطية وكذلك اختبارات النمو الاخلاقى واختبارات الابتكارية . وإلى الحد الذى يكون فيه المصحح عاملا رئيسيا فى ثبات هذه الاختبارات يصبح تقرير وجود ثبات بين المصححين أمراً ضرورياً .

وثبات المصححين يدعم التقديرات الأخرى للثبات ولكنه لا يحل محلها فمن المناسب أن نحسب الثبات بطريقة اعادة الاختبار عندما يصحح الاختبار بطريقة ذاتية (Anastasi, 1988, pp. 91 - 66) & (Aiken, R., Lewis , 1982 pp. 70 - 75) .

ما هو الثبات المناسب :

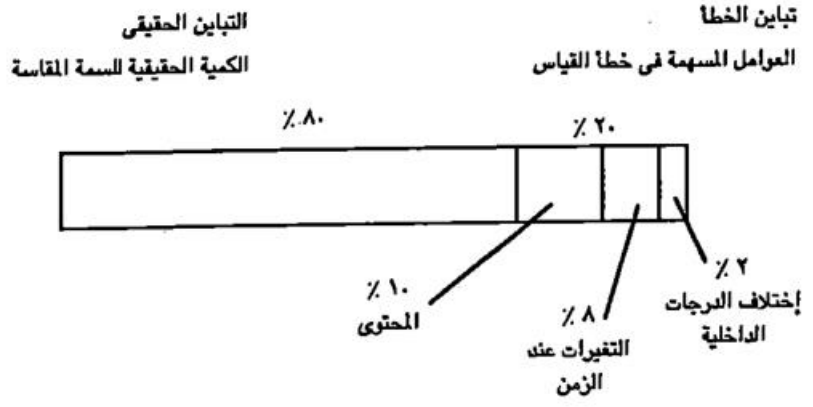
يتضح مما سبق وجود طرق مختلفة لتقدير الثبات فإنه عندما تكون لدينا صورة واحدة من الاختبار يمكن أن نقدر الثبات بعدد من الطرق مثل اعادة الاختبار ، والتجزئة النصفية ، ومعامل ألفا وكذلك ثبات المصححين ، أما الاختبارات التى يكون لها صورتان فيمكن أن نضيف طريقة الصور البديلة **ولكن أى هذه الطرق هو الأفضل ؟ ومتى نستخدم طريقة دون الأخرى ؟** للإجابة على هذه التساؤلات ينبغى أن نحدد طبيعة وغرض الاختبار الذى نقدر له الثبات .

فبالنسبة للاختبارات المعدة لتطبق مع الأفراد أكثر من مرة نتوقع أن يكون للاختبار ثبات عبر الزمن وفى هذه الحالة فإن الثبات المناسب هو ثبات اعادة الاختبار . وبالنسبة للاختبارات التى تهدف إلى أن يكون لها نقاء عاملى فإن معامل الفا يصبح أساسيا لتقدير الثبات ؛ وعلى العكس فإنه فى حالة الاختبارات متعددة العوامل مثل مقاييس الذكاء فإننا لا نستفيد كثيراً من مقاييس الاتساق الداخلى وبذلك فإن معامل الفا ومعادلة كيودريتشادسون لن تعطيا تقديرا جيدا للثبات فى هذه الحالة . ولهذا فإن معامل الفا لا يعتبر مؤشرا جيدا للثبات لكل الاختبارات وأنما هو يناسب الاختبارات التى تشتمل على عامل واحد . كذلك فإن طريقة تقدير الثبات باستخدام التجزئة النصفية تناسب أكثر الاختبارات التى تكون مفرداتها مرتبة بعناية تبعا لمستوى الصعوبة . وبالطبع فإن ثبات المصححين يناسب بشكل أكثر الاختبارات التى تشتمل على جانب ذاتى فى التصحيح مثل الاختبارات الاسقاطية ، واختبارات المقال، ومقاييس الإبتكارية .

وتلخص (Anastasi & Urbina, 1997 p . 101) مصادر تباين الخطأ في علاقته بمعامل الثبات فيما يلي :

نوع معامل الثبات	تباين الخطأ
إعادة تطبيق الاختبار	تعيين الوقت Time Sampling
الصور البديلة عندما تطبق فى آن واحد	تعيين المحتوى Content
الصور البديلة عندما تطبق بعد فترة زمنية	تعيين الوقت + تعين المحتوى .
التجزئة النصفية	تعيين المحتوى .
معادلة كيودر- ريتشاردسن ومعامل الفا كرونباخ	تعيين المحتوى ، وتغاير المحتوى

ويحدد (Gregory J. Robert (1996. pp, 97 - 98 تلك العلاقة في الشكل التالي :



الحالات الخاصة وتقدير الثبات :

إن الطرق التقليدية لتقدير الثبات قد تعطينا نتائج مضللة أو غير مناسبة لبعض التطبيقات . ومن المواقف التى تمثل مشكلات فى هذا الصدد عدم ثبات واستقرار الخصائص، واختبارات السرعة ، وتقيد المدى Ristriction range وكذلك الاختبارات المرجعة إلى محكات Criterion-referenced tests .

١ - الخصائص غير المستقرة

فى بعض الأحيان فإن بعض الخصائص التى نقيسها تعانى من تذبذب شديد أو عدم استقرار باختلاف ظروف القياس أو بعض المتغيرات الفسيولوجية . فمثلا رد الفعل الانفعالى Emotional reactivity كما يقا س باستجابة الجلد الكهربية أو الاستجابة الجلفانية للجلد يعتبر مثالا جيدا . فمثل هذا المقياس يتذبذب بسرعة كرد فعل للضوضاء العالية أو حالة

التفكير أو ضغوط البيئة بل انه ربما الكلام مع شخص أخر يمكن أن يؤدى إلى ارتفاع كبير فى الاستجابة الكهربية للجلد . ونتيجة لهذا التذبذب السريع يصبح تقدير الثبات باستخدام طريقة اعادة الاختبار غير مناسبة لهذه المقاييس .

٢ - اختبارات السرعة واختبارات القوة Speed and power Tests

اختبار السرعة هو الذى يشتمل على مفردات ذات مستوى منخفض من الصعوبة وإذا سمح الوقت فإن معظم المفحوصين يكونون قادرين على حل معظم أو كل مفردات هذا الاختبار . لكن كما يدل الاسم فإن اختبار السرعة يكون له زمن محدد يضمن أن يجيب عدد قليل من المفحوصين على الاختبار كاملا فى الوقت المحدد . ولأن المفردات التى يجاب عليها تكون عادة صحيحة فإن درجة المفحوص على اختبار سرعة تعكس إلى حد كبير سرعة الأداء ، أما اختبار القوة فهو اختبار يسمح بوقت يكفى للمفحوص أن يحاول مع كل مفرداته ولكنه أعد بحيث يصعب على المفحوص أن يحصل على الدرجة الكاملة .

وتشتمل معظم الاختبارات على مكونات سرعة ومكونات قوة . والنقطة الهامة التى نود أن نؤكد عليها حول ثبات اختبارات السرعة هو أن استخدام طريقة التجزئة النصفية Split-half التقليدية (التى تقارن المفردات ذات التسلسل الفردى مع المفردات ذات التسلسل الزوجى) سوف ينتج عنها معامل ثبات عالٍ زائف ؛ فمثلا المفحوص الذى يجيب على ٦٠ سؤالا من اختبار يشمل ٩٠ سؤالا سيحصل غالبا على ثلاثين اجابة فردية صحيحة وعلى ثلاثين اجابة زوجية صحيحة وبذلك يقترب معامل الثبات من الواحد الصحيح إذا كان أداء المجموعة على هذا المستوى . وبذلك فإن ثبات اختبارات السرعة ينبغى أن يحسب بطريقة إعادة الاختبار .

٣ - تقييد المدى : Restriction of Range

أن تقدير الثبات بطريقة اعادة الاختبار سوف يكون منخفضا إذا أُعد على أساس عينة متجانسة من المفحوصين الذين تقيد قدراتهم لمدى محدود . فلن يكون مناسبا أن نقدر ثبات اختبار للذكاء بتطبيقه مرتين لطلاب من المستوى الجامعى . على سبيل المثال فإننا لو استخدمنا أحد الاختبارات على عينة كبيرة من أفراد المجتمع واخذنا عينة من هؤلاء الأفراد (مثلا للتوظف) بناء على هذا الاختبار فإن هذه المجموعة المختارةأصبحت متماثلة تقريباً فى القدرة التى يقيسها الاختبار فعند اعادة تطبيقه على هذه العينة المختارة التى حصلت على أعلى الدرجات على الاختبار وحسابنا معامل الارتباط بين الدرجات التى اعطاها الاختبار لهذه العينة (المتجانسة) فى المرتين سنجد أن هذا المعامل منخفض عن المعامل المسحوب من المجتمع العام (الخليط) .

٤ - ثبات الاختبارات المرجعة لمحك Reliability of Criterion Referenced tests

فى الاختبارات المرجعة لمحكات يقاس الأداء فى صورة تمكن Mastary بدلا من

تقديره فى صورة تحصيل على خط متدرج . وتعد مفردات الاختبار للتعرف على مهارات معينة تحتاج إلى علاج ؛ وبذلك فإن المفردات تكون عادة فى صورة صحيح أو خطأ وتبنى الاختبارات المرجعة لمحك بحيث يكون التباين فى الدرجات بين المفحوصين منخفضا إلى أدنى درجة . وفى الواقع فإنه إذا استخدمت درجات الاختبار لأغراض التدريب واستمر كل فرد فى الدراسة إلى أن يتمكن من كل المهارات الخاصة بالاختبار فإن التباين فى درجات الاختبار يصبح غير موجود . وبذلك فإنه فى اطار هذه الظروف يصبح تقدير الثبات بالطرق التقليدية غير مناسبة لعدم وجود تباين فى الدرجات.

وباستخدام معادلة (Subokouiak 1980) يمكن حساب ثبات الاختبارات المرجعة إلى المحك حيث تعتبر اقل حساسية عندما تكون درجات الافراد على الاختبار اقل تبايناً حيث لاتصلح هنا معادلة جتمان أو استخدام معادلة كيودر – ريتشاردسن أو غيرها. (James s &Anthony 1980).

الثبات المنخفض : ويجيب Kaplan & Saccus (1989 pp 110 - 115) عن سؤال : ماذا يمكن عمله نحو الثبات المنخفض ؟؟ :

– زيادة عدد المفرادت :

حيث طبقاً لنموذج تعيين النطاق .The domain sampling model تعتبر كل مفردة عينة مستقلة من السمة أو القدرة موضع القياس ، وبالتالى فإن زيادة عدد المفرادت يؤدى إلى زيادة ثبات الاختبار . ويقدمان مثالاً على ذلك :

ثم حساب ثبات اختبار يتكون من ٢٠ مفردة فوجد ٠,٧٦ ونريد زيادة ثبات هذا الاختبار إلى ٠,٨٥ باستخدام معادلة للتنبؤ بطول الإختبار Formula :

$$ن = \frac{ر_م (1 - ر_و)}{ر_و (1 - ر_م)}$$

حيث ن = عدد المفردات التى تحقق زيادة معامل الثبات .

$ر_م$ = هو مستوى الثبات المأمول

$ر_و$ = هو مستوى الثبات الذى تم حسابه ويراد زيادته

$$ن = \frac{٠,٨٥ (١ - ٠,٧٦)}{٠,٧٦ (١ - ٠,٨٥)} = \frac{٠,٢٠٤}{٠,١١٤} = ١,٧٩$$

اى انه لزيادة ثبات الاختبار من ٠,٧٦ الى ٠,٨٥ ، يجب ان يزيد طول الاختبار بمقدار ١,٧٩ ضعفًا من العدد الأصلى لمفرادت الاختبار أى ١,٧٩ × ٢٠ = ٣٥,٨ مفردة أى ٣٦ مفردة.

ولكنهما يضعا محاذير على استخدام هذه المعادلة لزيادة ثبات الاختبار تتضح من المثال التالى .

كم يكون عدد مفرادت اختبار نريد زيادة ثباته من ٠,٥٠ إلى ٠,٩٠ وعلماً بأن عدد مفردات الاختبار هو ٤٠ مفردة وباستخدام المعادلة :

$$ن = \frac{(٠,٥٠ - ١)\ ٠,٩٠}{(٠,٩٠ - ١)\ ٠,٥٠} = \frac{٠,٥٠ \times ٠,٩٠}{٠,١٠ \times ٠,٥٠}$$

$$= \frac{٠,٤٥}{٠,٠٥} = ٩$$

أى يجب ان تكون عدد المفردات تسعة أضعاف العدد الأصلى !!

أى ٤٠×٩ = ٣٦٠ مفردة . وهذا أمر صعب تحقيقة .

★ **التحليل العاملى وتحليل المفردة** : Factor and item analysis

حيث يعتمد ثبات الاختبار على أن جميع المفردات التى تقيس نفس الخاصية المراد قياسها إلا أنه يحدث أحياناً الا تقيس بعض المفردات التكوين الفرضى المراد قياسه . وهذه المفردات تؤثر فى انخفاض معامل الثبات الذى يعتمد فى جوهره على الارتباط العالى الداخلى للمفردات . وللتأكد من ان المفردات تقيس بالفعل نفس الشئ هناك مدخلان : الأول هو التحليل العاملى ، حيث يتمتع الاختبار بثبات مرتفع اذا كانت مفرداته مشبعة على عامل واحد Unidimensional

والمدخل الثانى :

هو فحص الارتباط بين كل مفردة والمجموع الكلى لدرجات الاختبار ، فاذا كان الارتباط منخفضاً فهذا معناه ان هذه المفردة تقيس شيئاً آخر عما يقيسة الاختبار ، أو هذه المفردة سهلة أو صعبة جداً وهذا يعنى ان الثبات يكون ذا قيمة عند استبعاد مثل هذه المفردة.

* **تصحيح سبب الخلل** : Correction For attenuation

يعتبر انخفاض ثبات المقياس من المشكلات الحقيقية فى البحوث النفسية حيث يؤدى إلى عدم وجود ارتباطات دالة بين القياسات المختلفة .

فالاختبار غير الثابت يعطى نتائج قليلة أو عديمة القيمة ، وعلى ذلك فان الارتباطات المحتملة بين القياسات تكون ضعيفه بسبب اخطاء القياس ويقدم Kaplan & Saccuzzo) (1989 مثالاً على ذلك باستخدام المعادلة الآتية

$$\hat{r}_{12} = \frac{r_{12}}{\sqrt{r_{11} \; r_{22}}}$$

حيث $\hat{r}_{12}$ = التقدير الحقيقى للارتباط بين الاختبار ١ والاختبار ٢

r_{12} = الارتباط الملاحظ أو المحسوب بالفعل بين ١ ، ٢

r_{11} = ثبات الاختبار ١

r_{22} = ثبات الاختبار ٢

المثال :

معامل الارتباط بين اختبار للمهارة اليدوية وآخر للمهارات الرياضية هو ٠,٣٤ ومعامل ثبات الاختبار الاول هو ٠,٧٥ والثانى هو ٠,٨٢ فإن معامل الارتباط الحقيقى بينها بحسب كما يلى:

$$\hat{r}_{12} = \frac{٠,٣٤}{\sqrt{(٠,٧٥)(٠,٨٢)}} = \frac{٠,٣٤}{\sqrt{٠,٦١٥}} = \frac{٠,٣٤}{٠,٧٨} = ٠,٤٤$$

وبالتالى فإن الارتباط بين الاختبارين قد زاد من ٠,٣٤ إلى ٠,٤٤

ثانياً : الصـــــدق

معني الصدق

تري انستازى (1997) Anastasi أن صدق الاختبار يهتم بماذا يقيس الاختبار وإلي أي درجة من الجودة والكفاءة يقيسه . ويخبرنا بما يمكن الاستدلال عليه من درجات الإختبار.

وتحدد الجميعات العلمية ببقياس فى الولايات المتحدة (الرابطة الامريكية لعماء النفس APA ، والجمعية الأمريكية للقياس واللجنة القومية للقياس APA & AAER & NCME (1985)الصدق بأنه :

«يكون الاختبار صادقا بالدرجة التى تكون الاستدلالات المبنية عليه ملائمة ، ولها معنى ، ومفيدة »

"A test is valid to the extent that inferences made from it are appropriate, meaningful, and useful ".

كذلك فإن مهرنز وليهمان يعرفان الصدق بأنه « الدرجة التى يكون الاختبار قادرا بها على تحقيق اهداف معينة» Mahrens & Lehman (1978 p 109) ومن التعريفات البسيطة المتداولة عن الصدق – أن الاختبار يعتبر صادقا إذا كان يقيس ما يهدف إلى قياسه Allen & yen (1972 P95) .

ويرى المتخصصون فى مجال القياس النفسى أن الصدق هو الخاصية الوحيدة التى تحدد جودة الاختبار وأن الاختبار الصادق هو الذى ترتبط درجاته بدرجة عالية مع السلوك الفعلى الذى كان يهدف إلى قياسه فالاختبار الذى يقيس سمة القيادة تكون درجة صدقه عالية إذا استخدم فى اختيار قادة تثبت جدارتهم فى الأعمال التى توكل اليهم .

أنواع الصدق :

رغم تعدد مسميات الصدق المشار إليها فى المراجع التى تتحدث عن القياس مثل الصدق الظاهرى ، وصدق المحكمين ، والصدق التميزى ، والصدق العاملى ، والصدق التجريبى وغيرها فإنه أصبح من المتفق عليه بين المتخصصين أن الأنواع الرئيسية للصدق تنحصر فى ثلاثة أنواع رئيسية هى

١ – صدق المحتوى Content Validity (صدق المضمون) .

٢ – صدق المحك Criterion Related validity .

٣ – صدق التكوين الفرض Construct validity .

وسوف نتحدث فيما يلى عن كل نوع من هذه الأنواع .

١ - صدق المحتوى Content validity

من بين الأغراض التى يسعى اليها الاختبار هو أن نصف أو نعرف شخصا ما من خلال خصائص أو سمات هذا الشخص . ويعتبر الاختبار عينة Sample إذا كانت مفرداته مسحوبة من مجتمع محدد بشكل واضح . ويتعلق صدق المحتوى بمدى ملاءمة المحتوى فى تمثيله للمجال الذى تجرى الاستدلالات حوله . بل أن لينون يقدم تعريفاً للصدق يدور حول معنى صدق المحتوى فيقول :

أنه الدرجة التى يمكن فيها اعتبار اجابات المفحوص على مفردات الاختبار، عينة ممثلة لاجابات تجمع حقيقى أو افتراضي للمواقف التى تمثل معا المجال الذى يهتم به الشخص الذى يفسر درجات الاختبار (Lennon (1926 P 294.

ويمكن القول إن صدق المحتوى يمكن الحكم عليه باظهار أن السلوكيات التى يعتبر الاختبار عينة منها ممثلة للخاصية التى يقيسها الاختبار . وبذلك فإن صدق المحتوى يتوقف على كل من الاختبار نفسه والعمليات التى تدخل فى الاجابة على مفرداته . وعلى سبيل المثال فإن مقياسا يعتمد على الورقة والقلم حول المعرفة بوظيفة معينة قد لا يوفر مقياسا صادقا عن قدرة الفرد على أداء الوظيفة رغم أنه قد يكون مقياسا صادقا عن معرفته بماذا يفعل فى الوظيفة .

ويعتبر صدق المحتوى مفهوما مفيدا إذا عرفنا قدرا كبيرا من المعلومات حول المتغير الذى يرغب الباحث فى قياسه . وفى هذا الصدد فإنه فى اختبارات التحصيل على وجه الخصوص يكون من الممكن تحديد مجتمع السلوكيات (الظاهرة) التى سيعتبر الاختبار عينة منها بشكل مسبق .

على أنه إذا كان صدق المحتوى من الميسور تحقيقه فى اختبارات التحصيل فقد يكون الأمر أكثر صعوبة عندما يقيس الاختبار خاصية أو سمة غير واضحة التعريف . فكيف يمكن لمعد اختبار لقياس القلق أن يحدد مجتمع المفردات التى يمكن أن يأخذ عينة منها ليضمها إلي اختباره الذى يهدف إلى قياس القلق ؟ وفى مثل هذه الحالات فإن معد الاختبار لن يحاول تحديد مجتمع أو عالم المفردات التى تقيس الخاصة التى يحاول قياسها وإنما فقط بوسعه أن يتحقق من أن مجموعة من الخبراء قد راجعوا المجال الذى يمثله الاختبار وحكموا بأن مفردات الاختبار يتوفر فيها صدق المحتوى . وتشير بعض المراجع إلى هذا النوع من الصدق تحت مسمى صدق المحكمين Judge Validity أو الصدق الظاهرى Face Validity ولأن صدق المحتوى يقوم على أساس أحكام ذاتية فإن تحديد هذا النوع من الصدق يكون أكثر عرضة للخطأ عن غيره من أنواع الصدق . وعلي ذلك فإن وجود صدق المحتوى لاختبار ما لا يعتبر مبررا كافيا لاستخدامه . فقبل استخدام الاختبار يجب أن يثبت الاختبار فاعليته باستيفاء الصدق المرتبط بمحك آخر على سبيل المثال (Allyn & yen (1978 p.96، Gregory J. Roberl (1996, p.108)

تقدير صدق المحتوى

لا يوجد مقياس احصائى دقيق لتقدير صدق المحتوى . ذلك أن صدق المحتوى يعتبر بمثابة حكم حول الدرجة التى يمكن لها لاختبار ما أن يقدم عينة مناسبة من مجال محتوى خاص . والوصف التفصيلى لمجال المحتوى يتيح اطارا للتقويم الدقيق للاختبارات كما يزودنا بطريقة للتقويم المنظم لصدق القياس Murphy and Davidshofer(1994 P 110) . وتشتمل عملية تقدير صدق المحتوى على ثلاث خطوات :

١ – وصف مجال المحتوى .

٢ – تحديد المناطق (المجالات الفرعية) التى تقيسها كل مفردة فى الاختبار .

٣ – مقارنة بنية الاختبار مع بنية مجال المحتوى .

ورغم ما يبدو من بساطة هذه الخطوات إلا أنها تكون صعبة فى تطبيقها – وأكبر صعوبة نواجهها تكون فى الخطوة الأولى المتعلقة بوصف مجال المحتوى . وفيما عدا مجال الاختبارات التى تجرى فى حجرة الدراسة (الإختبارات التحصيلية) فإنه فى الغالب يكون من الصعب أن تصف مجالات المحتوى بشكل تفصيلى .

٢ – الصدق الظاهري : Face Validity

فى بعض الأحيان نجد بعض الباحثين يتحدث عن نوع من الصدق يطلق عليه الصدق الظاهرى . وقد يطلق البعض هذا المسمى على ماقام به بعض الخبراء من عمليات تحكيم حول اختبار معين على النحو الذى عرض فى صدق المحتوى . وفى الواقع فإن الصدق الظاهرى ليس صدقا بالمفهوم العلمى للصدق . وهو لا يشير إلى ما يقيسه الاختبار بالفعل ، وإنما إلى ما يبدو ظاهريا أنه يقيسه . ويهتم الصدق الظاهرى بما إذا كان الاختبار يبدو صادقا للمفحوصين الذين يطبق عليهم الاختبار ، وكذلك الافراد الذين يستخدمونه وكذلك الاشخاص غير المدربين الذين يتعاملون مع الاختبار . وبشكل أساسى الصدق الظاهرى يهتم بالالفة والعلاقات العامة .

والصدق الظاهرى رغم أنه لا يعبر عن صدق حقيقى فإنه يعتبر خاصية مرغوبة فى الاختبار من حيث فائدته فى العلاقات العامة . فلو أن الاختبار بدا غير مناسب من وجهة نظر المفحوصين فسوف لا يستجيبون له بشكل جاد .

وفى بعض الاحيان فإن الفاحص يرغب فى تقليل الصدق الظاهرى ليخفى الغرض الحقيقى للمقياس ويحدث ذلك فى الاختبارات التى تقيس المجال الوجدانى (مثلا اختبارات القلق والخجل والخوف والاكتئاب) والخلاصة أن الصدق الظاهرى ليس صدقا من الناحية الموضوعية وإنما هو يتصل بالشكل الظاهر للاختبار ومرغوبية المستخدم . ورغم أنه ليس صدقا حقيقياً فإنه يؤخذ فى الاعتبار عند بناء الاختبارات . وليس معنى توافره فى اختبار أن

الاختبار صادق فعلاً (Anastasi&Urbina (1997 pp 117-118 .

٣ - الصدق المرتبط بالمحكات :

يتم من خلال حساب ارتباط الاختبار مع محك آخر خارجي مستقل عنه ، ويهتم هذا النوع من الصدق بقدرة الاختبار فى التنبؤ بسلوك الفرد فى المواقف المعينة . ويميز المتخصصون فى القياس وكذلك المعايير التى نشرتها الرابطة الأمريكية لعلماء النفس بين نوعين من صدق المحك هما :

١ - الصدق التلازمى Concurrent validity .

٢ - الصدق التنبؤى Predictive Validity .

ويتم التمييز بين هذين النوعين من صدق المحك باستخدام متغير الزمن الذى تجمع فيه البيانات الخاصة بالمحك فإذا كانت هذه البيانات تجمع تقريبا فى نفس الوقت الذى نحصل فيه على بيانات الاختبار فإننا نكون بصدد الحديث عن الصدق التلازمى أما إذا كانت بيانات المحك تجمع فى وقت لاحق فإننا نكون بصدد الحديث عن الصدق التنبؤى .

ثمة طريقة أخرى للتفرقة بين هذين النوعين من صدق المحك - وهو تمييز منطقى أكثر منه تمييز اجرائى ولا يهتم بالزمن وإنما يركز على الغرض من الاختبار . ففى الصدق التنبوئ يكون اهتمامنا فى الواقع منصبا على فائدة درجة الاختبار فى التنبؤ بأداء ما فى المستقبل . أما فى الصدق التلازمى فإننا نهتم بالسؤال حول ما إذا كانت درجة الاختبار يمكن أن تحل محل طريقة أقل كفاءة لجمع بيانات المحك . وبذلك فإن ما إذا كنا نعبر عن صدق المحك باستخدام الصدق التنبؤى أو الصدق التلازمى يتوقف على ما إذا كنا نهتم أساساً به لتنبؤ أو بتقديرات الحالة الراهنة . وبمعنى أخر فإن الصدق التنبؤى يناسب الاختبارات التى تستخدم للتنبؤ بأداء أو نتائج فى المستقبل (مثل مقاييس التفضيل المهنى) بينما الصدق التلازمى يناسب أكثر الاختبارات التى تستخدم لتشخيص الحالة الراهنة .

Anastasi & Urbina (1997 pp118 - 120)

٤ - صدق التكوين الفرضي : Construct Validity

يعبر صدق التكوين الفرضى عن الدرجة التى يقيس بها الاختبار التكوين الفرضى Construct أو الخاصية النظرية التى أعد لقياسها (Allen & yen 1979 p 108) .

ويهتم المتخصصون فى علم النفس بقياس الخصائص المجردة مثل السعادة ، والذكاء ، والدافعية ، والاجتماعية . وهذه الأشياء لا توجد فى شكل ملموس بل هى خصائص مجردة أو تكوينات فرضية .

والاختبارات النفسية تعد غالبا لقياس تلك التكوينات الفرضية النفسية . وبعض الاختبارات تنتج قياسات صادقة لتكوينات فرضية هامة بينما أخرى تظهر قليلا أولا تظهر

أى تكوين فرضى . ولأن التكوينات مجردة فى طبيعتها ، فإن عملية تحديد ما إذا كان الاختبار يقدم قياسات مناسبة لتكوين فرضى معين أمر معقد – ولكى نصف عملية الحكم على صدق التكوين الفرضى ، يجب أولا أن نناقش طبيعة التكوينات الفرضية .

طرق تقدير صدق التكوين الفرضي :

إن الصدق الخاص بالتكوين الفرضى يعتمد على الوصف التفصيلى للعلاقة بين التكوين الفرضى وعدد من السلوكيات المختلفة – وكلما زادت معرفتنا عن التكوين ، كلما كانت الفرص أفضل فى تحديد ما إذا كان الاختبار يمثل قياسا مناسبا للتكوين . وأحد النتائج هى أنه سوف يصبح من السهل أن نحدد صدق التكوين للمقاييس ذات التكوينات الفرضية جيدة التعريف عنه فى المقاييس ذات التكوينات الفرضية ضعيفة التعريف .

وفيما يلى بعض طرق تقدير صدق التكوين الفرضي :

١ – التغيرات النمائية : (Cronbalch, L-J , (1990)

يعتبر التمايز العمرى محكا رئيسيا يستخدم فى تقدير الصدق لعدد من اختبارات الذكاء . مثل اختبار ستانفورد بينيه ومعظم اختبارات مرحلة ما قبل المدرسة يستخدم الأعمار الزمنية لتحديد ما إذا كانت درجة الاختبار تظهر زيادة متطورة مع التقدم فى العمر . لأن القدرات يتوقع أن تزداد بزيادة العمر فى مرحلة الطفولة ، فإنه من القضايا الجدلية أن درجات الأختبار تزداد بالمثل ، إذا كان الاختبار يتمتع بالصدق . وتصور وجود مقاييس عمرية للذكاء كتلك التى بدأها بينيه تقوم على افتراض أن الذكاء يزداد مع العمر ، على الأقل حتى البلوغ .

وبالطبع فإن التمايز العمرى ليس مناسباً بالنسبة للوظائف التى لا تظهر تغيرات واضحة ومتسقة مع التغير فى العمر . وعلى سبيل المثال فإنه فى قياس الشخصية يكون ذلك الدور ضئيلا . و يكون التمايز العمرى شرطاً ضروريا ولكنه غير كاف لصدق الاختبار . وبذلك فإنه إذا اخفقت درجات الاختبار فى أن تتحسن مع السن فإن ذلك قد يشير إلى أن الاختبار لا يمثل قياسا صادقا للقدرات التى أعد لقياسها .

كذلك فمن النقاط العامة فى موضوع التمايز العمرى ضرورة أن نأخذ بالاعتبار الفروق الثقافية التى قدن فيها الاختبار .

ويستخدم التحليل النمائى أيضاً فى تقدير صدق بعض الاختبارات التى تتعامل مع مراحل معينه فمثلا فى اختبارات بياجيه من يعتبر هذا النوع من التحليل فى المقاييس الرئيسية الخاصة بنظرية بياجيه . فأحد الافتراضات الأساسية فى هذه المقاييس هو تتابع تنميط النمو مثل التحصيل فى المراحل المبكرة ونمو المفهوم باعتباره مطلبا مسبقا لاكتساب المهارات التصورية اللاحقة .

٢ - الارتباط بالاختبارات الأخرى :

إن الارتباطات بين اختبار جديد واختبارات مماثلة قديمة تؤخذ فى بعض الاحيان كدليل على أن الاختبار الجديد يقيس تقريباً نفس المجال العام للسلوك كغيره من الاختبارات التى تحمل نفس المسمى ، مثل اختبارات الذكاء، واختبارات الاستعداد الميكانيكى ، وعلى خلاف الارتباطات التى نحصل عليها فى صدق المحك ، فإن هذه الارتباطات تكون مرتفعة بدرجة متوسطة ولكنها ليست عالية بشكل كبير. وإذا كان الاختبار الجديد يرتبط بدرجة عالية باختبار موجود فعلا بدون مميزات اضافية مثل الاختصارأو سهولة التطبيق ، فإن الاختبار الجديد يصبح تكراراً لا ضرورة له .

وفى بعض الأحيان يكون الحرص ألا يرتبط الاختبار مع بعض الاختبارات الأخرى المختلفة عنه فى المجال ؛ وبذلك فإنه من المرغوب أن يكون معامل الارتباط منخفضا فى هذه الحالة . ورغم ذلك فلا يؤخذ انخفاض الاختبار كدليل على دقة الاختبار . ومما ينبغى أن يذكر هنا أن هذا الأسلوب يمثل أحد الأساليب الأضافية مستخدمة فى حساب صدق المحتوى . Anastasi (1982 p 144 - 145) Factor Analysis .

٣ - التحليل العاملى

طور أسلوب التحليل العاملى اصلا لتحديد السمات النفسية - لكن هذا الأسلوب يعتبر مناسبا بشكل خاص فى عملية تقدير صدق التكوين . وأسلوب التحليل العاملى هو أسلوب احصائى لتحليل العلاقات الداخليه للبيانات الخاصة بالسلوك . وتشير بعض المراجع إلى صدق التكوين الفرضى الذى يقدر باستخدام التحليل العاملى - بالصدق العاملى Factorial Validity والتحليل العاملى اصطلاح يمثل عددا كبيرا من الأساليب الرياضية المختلفة لتحليل العلاقات البينية بين مجموعة من المتغيرات ولتفسير هذه العلاقات فى شكل عدد أقل من التغيرات يطلق عليها اسم «العوامل Factors - والعامل هو متغير فرضى يؤثر على الدرجات فى واحد أو أكثر من المتغيرات المشاهدة .

وعندما يستخدم التحليل العاملى على مصفوفة ارتباطات ، فإن الاختبارات التى تتأثر بعوامل معينة يقال إن لها تشبعات عاملية عالية أو أنها تتشبع بدرجة عالية على هذه العوامل.

٤ - الاتساق الداخلى : Internal Consistency

تذكر بعض الاختبارات المنشورة وخاصة فى مجال قياس الشخصية أن الاختبار قد تم التحقق من صدقه بطريقة الاتساق الداخلى والخاصية الرئيسية لهذه الطريقة هو أن المحك ليس سوى الدرجة الكلية للاختبار نفسه. وفى بعض الاحيان تستخدم طريقة معدلة لطريقة المجموعات المتناقضة Contrasted groups حيث تختار المجموعات المتطرفة على أساس من الدرجة الكلية للاختبار . ويقارن أداء المجموعة الأعلى على المحك على كل مفردة من

مفردات الاختبار مع المجموعة الأدنى فى الأداء . والمفردات التى تظهر نسبة أكبر من نجاح الأداء فى المجموعة الأعلى عنها فى المجموعة الأدنى على المحك يمكن اعتبارها صادقة ، وقد تستخدم أساليب الارتباط ايضا لهذا الغرض . وعلى سبيل المثال فإنه يمكن حساب الارتباط ثنائى السلسلة biserial Correlation بين «ينطبق - ولا ينطبق» على كل مفردة مع الدرجة الكلية على الاختبار والمفردات التى ترتبط مع الدرجة الكلية للاختبار تستبقى . والمفردات التى تختار بهذه الطريقة يقال عنها أنها تظهر اتساقا داخليا حيث أن كل مفردة تميز بين المفحوصين فى نفس الاتجاه كالاختبار كله .

وهناك استخدام أخر للاتساق الداخلى للمحك يشتمل على ارتباطات الاختبار الفرعى مع الدرجة الكلية للاختبار. وكثير من اختبارات الذكاء تتألف من اختبارات فرعية مطبقة بشكل منفصل . (مثل الاختبارات اللفظية ، والحسابية ، وتكملة الصور الخ) حيث نجمع درجاتها فى حساب الدرجة الكلية للاختبار وفى بناء هذه الاختبارات ، فإن الدرجات على كل اختبار فرعى يحسب ارتباطها غالبا مع الدرجة الكلية ، وأى اختبار فرعى يكون ارتباطه مع الدرجة الكلية منخفضا يستبعد وحينئذ يتم اثبات ارتباطات الاختبارات الفرعية الباقية مع الدرجة الكلية كدليل على الاتساق الداخلى للاختبار كله .

ومن الواضح أن الارتباطات الخاصة بالاتساق الداخلى سواء ما يبنى منها على أساس المفردات أو ما يبنى على أساس الاختبارات الفرعية هى اساسا قياسات لتجانس الاختبار . تساعد على تحديد مجال السلوك أو السمة التى تعتبر الاختبار عينة ممثلة لها ، فإن درجة التجانس للاختبار لها بعض معنى لصدقه التكوينى - ومع ذلك فكما تقول انستازى Anastasi 1982 أن اسهام بيانات الاتساق الداخلى فى صدق الاختبار محدودة - وفى غياب البيانات التى نحصل عليها من خارج الاختبار ومحك خارجى ، فإننا نعلم القليل عما يقيسه الاختبار

٥ - الصدق التقاربى والصدق التمييزى

Convergent and Discriminant validity

اشار كامبيل (1960) Compbell فى دراسة تحليلية عن صدق التكوين الفرضى أنه لكى نطمئن على صدق التكوين فإن علينا أن نوضح أن الاختبار لا يرتبط فقط بدرجة عالية مع المتغيرات الأخرى والتى ينبغى أن يرتبط بها نظريا وأنما ايضا أنه لا يرتبط بشكل جوهرى مع المتغيرات التى ينبغى أن يختلف معها . وقد وصف كامبل وفيسك Campbell & Fiske العملية الأولى على أساس أنها الصدق التقاربى Convergent validity والثانية على أنها الصدق التمييزى Discriminent validity وعلى سبيل المثال فإن الارتباط العالى بين اختبار للاكتئاب وأخر للميل العصابى يمكن أن يؤخذ مثالا للصدق التقاربى بينما الارتباط المنخفض بين اختبار الاكتئاب واختبار للمرح يعتبر مثالا للصدق التمييزى .

٦ - التدخلات التجريبية

تعتبر التجارب مصدرا هاما لقياس صدق التكوين وذلك من خلال تأثير متغيرات معينة على درجات الاختبار ، وفى دراسة صدق اختبار مرجع لمحكات لاستخدامه فى برنامج للتعليم الفردى على سبيل المثال ، فإن احدى الطرق المستخدمة تكون بمقارنة الدرجات على الاختبار القبلى والدرجات على الاختبار البعدى - ومنطق هذه الطريقة يفترض أن تكون الدرجات منخفضة على الاختبار القبلى الذى يطبق قبل التدريس وتكون الدرجات مرتفعة على الاختبار البعدى .

٧ - اسهامات علم النفس المعرفى Cognitive psychology

اسفر التعاون بين علم النفس التجريبى والقياس النفسى فى السبعينات من القرن العشرين عن اسهامات جوهرية فى فهمنا للتكوينات المقدرة بواسطة اختبارات الذكاء واختبارات الاستعدادات الأخرى .

وفى مرحلة قبل ذلك فى الخمسينات بدأ اصحاب الاتجاه المعرفى فى استخدام مفاهيم معالجة المعلومات Ienformation processing فى محاكاة التفكير ، بحيث يمكن كتابة برامج لمحاكاة اداء الأشخاص ، وبمثل هذه البرامج يمكن التنبؤ بعدد وأنواع الأخطاء التى تحدث والوقت اللازم للاستجابات المختلفة .

وفى فترة الستينات بدأ قليل من علماء النفس فى استخدام هذه العينات لمعالجة المعلومات والمحاكاة على الحاسب الآلى . لاستكشاف ما تقيسه اختبارات الذكاء . وقد اقترح ستيرنبرج (1977) Starnberg طريقة تحليل المكونات وهى طريقة طموحه ومنظمة للربط بين الأداء على اختبار الذكاء وعلم النفس المعرفى وما يمكن قوله عن اسهامات علم النفس المعرفى فى مجال القياس النفسى تتمثل فى قياس العمليات المعرفية التى تحدث بدلا من الاقتصار على النتائج النهائية وهذا من شأنه أن يثرى فهمنا لما تقيسه الاختبارات Anastasi (1982 p 150 - 152) .

مقارنة أنواع صدق الاختبارات :

فى دراسة الصدق فإن الباحث يهتم بالوصول إلى اجابة عن السؤال هل يقيس الاختبار فعلا ما هدف إلى قياسه ؟ وبصفة عامة فإن صدق طريقة ما يعنى الدقة التى يمكن بها اجراء قياسات مناسبة وذات معنى ؛ بمعنى أن هذه الطريقة تقيس فى الواقع السمات التى تقصد أن تقيسها وبمعنى آخر فإننا عندما نبحث فى الصدق فإننا فى الواقع نبحث فيما إذا كان الاختبار المعروف ثباته يقيس ما أعد لقياسه ؟؟

وبالنسبة لأى اختبار فإنه لا يكون هناك معامل ثابت للصدق لكل غرض يستخدم فيه الاختبار ولا لكل مجموعة من الأفراد يستخدم معها ؛ وإنما يختلف الصدق تبعا للاغراض التى يستخدم فيها الاختبار . وكذلك المجموعة التى يفترض أن يميز بين افرادها . وبالنسبة

لكل اجراء لتقدير صدق الاختبار فيها فإن السؤال الرئيسى الذى نطرحه هو لمن ولأى غرض يكون الاختبار صادقا ؟ وبذلك فإننا نحتاج إلى محكات مختلفة لأغراض الاختبار المختلفة . وعلى سبيل المثال فإن الصدق الذى يتمتع به أحد اختبارات الاستعدادات عند استخدامه فى توزيع الأفراد على تخصصات معينة قد يكون مختلفا تماما عن صدق الاختبار نفسه عندما يستخدم للتنبؤ بنجاح هؤلاء الأفرادوبعد انهاء تعليمهم فى الوظائف التى يحصلون عليها . وبذلك فنحن فى الواقع ندرس مصداقية استخدام المقياس أو القرار الذى يتخذ به أكثر من دراستنا لصدق المقياس نفسه .

وقد عرض الباحث لثلاثة أنواع من الصدق هى صدق المحتوى أو صدق المضمون Content Validity وصـــدق المحك Critarion Related validity بنوعية التنبؤى Predictive والتلازمى Concurrent وأخيرا صدق التكوين الفرضى Construct Validity ومن الملاحظ فإن صدق التكوين الفرضى يعتبر مفهوما شاملا يشتمل على الانواع الأخرى من الصدق فالأساليب التى تستخدم للتعرف على صدق المحتوى وصدق المحك يمكن أن تستخدم مرة أخرى فى دراسة صدق التكوين الفرضى . وكما تقول انستازى (Anastasi 1982) فإن مقارنة أداء المجموعات المتنافرة (المتضادة) Contrasted groups مثل العصابين والاسوياء تعتبر طريقة من طرق فحص صدق التكوين الفرضى لاختبار معد مثلاً لقياس التوافق الانفعالى أو القلق – كما أن الارتباط بين الأداء على اختبار للاستعداد الميكانيكي مع الأداء فى عدد من الوظائف يسهم فى فهمنا للتكوين الفرضى الذى يقيسه الاختبار.

كذلك فإن تقدير الصدق باستخدام محكات عملية يذكر دائما فى أدلة الاختبارات لمساعدة الباحث على فهم ما يقيسه الاختبار رغم أنه قد لا يكون مهتماً بشكل مباشر بالتنبؤ بأي نوع من المحكات المستخدمة – ولكن عن طريق دراسة هذه المحكات فإن مستخدم الاختبار يكون قادرا على بناء مفهوم حول السلوك الذى يمثل الاختبار عينة منه . وإذا تتبعنا هذا النمط من التفكير بشكل أكثر عمقًا فسوف نرى أن كل استخدام للاختبار – وكل تفسير لدرجاته يتضمن صدق التكوين الفرضى وهو أمر يؤكده باحثون كثيرون مثل Guion 1977, Mascick 1980 الذي يري أن اصطلاح الصدق كما يستخدم على الدلالة التفسيرية للاختبار يمكن أن نحتفظ به لصدق التكوين الفرضى وأما الأساليب الأخرى التى ارتبط بها اصطلاح الصدق تقليديا فينبغى من وجهة نظره أن نقرنها بمسميات وصفية نوعية – وبذلك يمكن أن نسمى صدق المحتوى بملاءمة المحتوى وتغطية المحتوى Content relevance and Content coverage للإشارة إلى تحديد المجال وتمثيل المجال على التوالى . أما صدق المحك يمكن تسميته المنفعة التنبؤية dictive uitility ، والمنفعة التشخيصة Diagnostic utility للإشارة إلى الصدق التنبؤى والصدق التلازمى . لكن من المتوقع أن تظل الاصطلاحات القديمة سائدة لفترة زمنية أخرى نظرا لطول مدة استخدامها ووجودها فى

كتبيات الأدلة الخاصة بالاختبارات المنشورة (Anastasi 1982 p 152 - 154) ، ويقدم Messick (1980) وصفاً لمظاهر صدق الأختبار كما يلي :

نوع الصدق	معناه
صدق المحتوي	الصدق المرتبط بالمحتوي – وخصائص المجال – تمثيل المجال .
صدق المحك	الصدق المرتبط بالمحكات
* الصدق التنبؤي	عند استخدام الاختبار بغرض التنبؤ
* الصدق التلازمي	عند استخدام الاختبار في التشخيص
صدق التكوين الفرضي	تفسير شمولية المعني
الصدق التقاربي	يعبر عن مدي التماسك أو الترابط
الصدق التميزي	مدي التمييز
صدق السمة	تطابق السمة
الصدق المنطقي	الصدق المرتبط بالمنطق
الصدق العاملي	التكوين العاملي
الصدق الجوهري	الإتساق الداخلي
صدق البنية	دقة التركيب أو البنية
الصدق الخارجي	الإرتباط بالعوامل الخارجية
الصدق المجتمعي	مدي تعميم النتائج علي المجتمع
الصدق البيئي	التعميم البيئي
الصدق المؤقت	الاستمرار المؤقت عبر مراحل النمو أو التطور .
صدق المهمة	القدرة علي تعميم المهمة

خطوات الدراسة :

قام الباحث بفحص بعض الأعداد المنشورة من المجلة المصرية للدراسات النفسية منذ عام ١٩٨٥ وهو العام الذى بدأت فيه المجلة تصدر بانتظام ،حتى عام ١٩٩٩م * . وقد تم هذا الفحص فى ضوء المحاور التالية :

١ – الإعداد : أى اعداد المقاييس هل قام الباحث باعدادها بنفسه . أم استخدام مقاييس

* هنا سوف نشير إلي أرقام الأعداد وتاريخ صدورها .

معدة من قبل وإلى أى مجال ينتمى الاختبار أو المقياس . المجال المعرفى أم المجال الوجدانى) ، هل ذكر الباحث عدد المفردات التى يتكون منها الاختبار أو المقياس هل حدد الباحث نمط الاستجابة على الاختبار أو المقياس ، هل ذكر الباحث الطرق المستخدمة لحساب ثبات المقياس أو الاختبار هل اشار الباحث إلى الطرق المستخدمة فى حساب الصدق وما مدى ملاءمة هذه الطرق .

٢ – مجال القياس .

٣ – عدد المفردات .

٤ – نمط الاستجابة .

٥ – طرق حساب الثبات .

٦ – طرق حساب الصدق .

نتائج الدراسة وتفسيرها ومناقشتها :

جدول (١) اعداد المقاييس

م	السنة	عدد المقاييس	الإعداد	
			الباحث	باحث آخر
١	١٩٨٥	٢٨	٣	٢٥
	٪	٥,٦	٠,٦٠	٥,٠٢
٢	١٩٨٦	٦٣	١٩	٤٤
	٪	١٢,٧	٣,٨	٨,٨
٣	١٩٨٩	٦٢	٢٩	٣٣
	٪	١٢,٤	٥,٨	٦,٦
٤	١٩٩٠	٧٧	١٥	٦٢
	٪		٣,٠١	١٢,٤
٥	١٩٩١	٢٤	٨	١٦
	٪	١٥,٥	١,٦	٣,٢
٦	١٩٩٢	٤٠	١٥	٢٥
	٪	٨,٠٣	٣,٠١	٥,٠٢
٧	١٩٩٣	٢٨	١٨	١٠
	٪	٥,٦	٣,٦	٢,٠٠

م	السنة	عدد المقاييس	الإعداد	
			الباحث	باحث آخر
٨	١٩٩٤	٣٧	١٤	٢٣
	٪	٧,٤	٢,٨	٤,٦١
٩	١٩٩٥	٢٣	٨	١٥
	٪	٤,٦	١,٦	٣,٠١
١٠	١٩٩٦	٢٤	٨	١٦
	٪	٤,٨	١,٦	٣,٢
١١	١٩٩٧	٢٥	٨	١٧
	٪	٥,٠٢	١,٦	٣,٤
١٢	١٩٩٨	٥٢	٢٠	٣٢
	٪	١٠,٤٤	٤,٠١	٦,٤٢
١٣	١٩٩٩	١٥	٥	١٠
	٪	٣,٠١	١,٠٠	٢,٠٠
	مج	٤٩٨	١٧٠	٣٢٨
	٪		٣٤,١٣٦	٦٥,٨٦٣

يتضح من جدول (١) ان عدد المقاييس والاختبارات التى استخدمت فى البحوث والدراسات المنشورة بالمجلة منذ العام ١٩٨٥ حتى ١٩٩٩م باستثناء عامى ١٩٨٧ ، ١٩٨٨ هو (٤٩٨) مقياساً واختباراً منهم عدد (١٧٠) مقياس من اعداد الباحثين انفسهم وهو يمثل ٣٤,١٤ ٪ من اجمالى العدد ، (٣٢٨) مقياس من اعداد باحثين أخرين بنسبة ٦٥,٢٦ ٪ .

وباستقراء نتائج هذا الجدول نستنتج أن نسبة عدد المقاييس التى استخدامها الباحثون وهى من اعداد باحثين آخرين ٦٥,٨٦ ٪ أى أنها تقريباً ضعف نسبة عدد المقاييس التى قام الباحثين باعدادها بأنفسهم ٣٤,١٤ ٪ وأن هناك عدد من الدراسات اقتصر الباحثين على استخدام تلك المقاييس المعدة من قبل Ready Mad Tests بدون اعادة حساب صدق أو ثبات هذه المقاييس أو حتى الاشارة إلى صدق وثبات المقاييس فى الدراسة الأصلية وسوف يناقش الباحث ذلك فيما بعد .

– وبالنسبة للمقاييس التى اعدها الباحثون ٣٤,١٤ ٪ منها مقاييس تمثل اضافة جيدة لميدان القياس النفسى .

جدول (٢) مجال القياس

م	السنة	عدد المقاييس	مجال القياس		م	السنة	عدد المقاييس	مجال القياس	
			معرفى	وجدانى				معرفى	وجدانى
١	١٩٨٥	٢٨	١١ ١ معلومات	١٦	٨	١٩٩٤	٣٧	٨ ١ + ١ معلومات	٢٧
		٥,٦	٢,٤	٣,٢			٧,٤	٢,٠٠	٥,٤
٢	١٩٨٦	٦٣	٣٢ ٨ معلومات	٢٣	٩	١٩٩٥	٢٣	٨	١٥
		٪	٨,٠٠	٤,٦			٤,٦	١,٦	٣,٠٠
٣	١٩٨٩	٦٢	٢٠	٤٢	١٠	١٩٩٦	١٥	٣ ١ معلومات	٢٠
		٪	٤,٠٠	٨,٤			٪	,٨	٤,٠٠
٤	١٩٩٠	٧٧	١٤	٦٣	١١	١٩٩٧	٢٥	٦ ٢ معلومات	١٧
		٪	٢,٨	١٢,٧			٥,٠٠	١,٦	٣,٤
٥	١٩٩١	٢٤	٥	١٩	١٢	١٩٩٨	٥٢	٩ ٢ معلومات	٤١
		٪	١,٠٠	٣,٨			١٠,٤	٢,٢	٨,٢
٦	١٩٩٢	٤٠	١٣ ١ معلومات	٢٦	١٣	١٩٩٩	١٥	٤	١١
		٪	٢,٨	٥,٢			٣,٠	٠,٨	٢,٢
٧	١٩٩٣	٢٨	٤	٢٤		مجـ	٤٩٨	١٥٤	٣٤٤
		٪	٠,٨	٤,٨			٪	٣٠,٩	٦٩,١

يتضح من جدول رقم (٢) أن عدد المقاييس والاختبارات التى استخدمت لقياس البعد المعرفى (١٥٤) مقياساً بنسبة ٣٠,٩ ٪ بينما عدد الادوات التى استخدمت لقياس البعد الوجدانى (٣٤٤) مقياساً بنسبة٦٩,١ ٪ ، وهناك مقياس للجوانب الفسيولوجية وهو اختبار المسح النيورولوجى السريع (عبد الوهاب كامل ، ١٩٩٤) واستمارة لاستطلاع رأى المعلمين وخبراء التربية (لورانس بسطا ، ١٩٩٤) واستمارة بيانات مثل (استمارة تاريخ الحالة) (حسن مصطفى ، ١٩٨٩) ، واستبيان الكفاءة الجنسية (محمد سمير فرح ، ١٩٩٠) ، واستمارة جمع معلومات (هناء أبو شهبة ١٩٩٠) ، ومنها استبيان مفتوح لقياس الاتجاهات نحو الزى الشرعى (وفاء كمال ، ١٩٩٣) تمثل البعد المعرفى فى قياس الذكاء العام (المتشابهات ، المصفوفات ، الذكاء غير اللفظى) .

القدرات العقلية العامة ، وكسر الاستعداد المكانى . الاشكال المتتابعة ، التفكير التجريدى ، بناء المكعبات ، العلاقات المكانية ، رسم الرجل اختبار الذكاء غير اللفظى ، القدرة العقلية الأولية ، التحصيل التفكير الابتكارى (تورانس ، الصفطى) مرونة الافكار ، القدرة على التفكير الابتكارى – الاستعمالات غير المعتادة ، الاشكال الهندسية ، عناوين القصص التفكير التجريدى – الطلاقة اللفظية ، الطلاقة الفكرية . الإسلوب المعرفى . التسلطيه، الدجماطبقية ، الصلابة والمرونة . الاندفاع ، التروى ، التذكر ، سرعة التذكر ، الذاكرة المكانية ، تذكر الارقام ، الاستعمالات ، المترتبات ، الادراك المكانى ، العلاقات المكانية .

واستخدمت الادوات فى تقدير البعد الوجدانى لقياس العدوان ، مفهوم الذات ، التقبل الإجتماعى ، الرضا الشخصى ، الاتجاهات ، القلق ، موضع الضبط ، الاكتئاب ، الخجل ، التنشئة الاجتماعية ، القيم ، تقدير الذات ، الوعى الدينى ، الدافع للانجاز ، مركز التحكم ، الحكم الخلقى ، الضغوط والاتجاهات الوالدية ، التفضيل الشخص ، الانبساط / الانطواء الاغتراب ، الدوافع ، مفهوم الذات ، العزو ، الطمأنينه ،الميول ، التكيف ، المخاوف ، التوافق.

ومما سبق يتضح أن الادوات (المقاييس والاختبارات) المستخدمة فى قياس الجانب الوجدانى فى الشخصية تمثل حوالى ثلاثة اضعاف الادوات المستخدمة فى قياس البعد المعرفى ويعزو الباحث ذلك إلى :

* سهولة اعداد مقاييس البعد الوجدانى فى الشخصية مما يترتب علية عدم بذل جهد كبير بالمقارنة فى اعداد مقاييس الذكاء والقدرات العقلية والعمليات العقلية المختلفه .

إحجام الباحثين عن خوض تجربة بناء اختبارات ومقاييس للبعد المعرفى تناسب البيئة والثقافة العربية . والمتتبع لحركة القياس النفسى فى مصر يجد أن مقاييس الذكاء المختلفة التى أعدها أو قام بتعريبها رواد علم النفس مثل اختبار القدرات العقلية ، اختبار الذكاء المصور، واختبار الذكاء العالى ، واختبار المصفوفات المتتابعة ، ومقياس تورانس للتفكير الابتكارى ، ..... وغيرها من الاختبارات مازالت تمثل الركيزه الأساسية لادوات القياس فى البحوث والدراسات النفسية ، بل وفى معامل علم النفس فى الجامعات المصرية والعربية .

ذلك على الرغم من تعدد النظرة إلى مفهوم الذكاء وكيفية قياسة فى الفترة الأخيرة ويرى الباحث أنه مع مطلع الفية جديدة لابد من تضافر الجهود وحثها على مزيد من الاهتمام بحركة القياس النفسى وبناء الاختبارات النفسية بما يناسب طبيعة البيئة المصرية والثقافة العربية . ويرى أن السبيل إلى ذلك بشكل مبدئي هو تنشيط حركة ترجمة المقاييس الاجنبية وتقنينها على البيئة المصرية ، على أن يتم ذلك من خلال مشروع Project تتبناه الجمعية المصرية للدراسات النفسية أو المركز القومى للامتحانات والتقويم التربوى أو مركز الارشاد النفسى أو رابطه الاخصائين النفسيين ، أو أى مؤسسة أو مركز للبحوث له اهتمام بمجال القياس النفسى والتربوى.

جدول (٣) عدد المفردات وطريقة الإستجابة

السنة	عدد المقاييس	عدد المفردات	نمط أو طريقة الإستجابة	السنة	عدد المقاييس	عدد المفردات	نمط أو طريقة الإستجابة
١٩٨٥	٢٨	١٣	١	١٩٩٤	٣٧	٢٧	١٤
٪	١٢,٧	٢,٦	٠,٢	٪	٤,٧	٥,٤	٢,٨
١٩٨٦	٦٣	١١	٥	١٩٩٥	٢٣	١٣	٢
٪	٥,٦	٢,٢	١,٠	٪	٤,٦	٢,٦	٠,٤
١٩٨٩	٦٢	٢٤	٢	١٩٩٦	٢٤	١٥	٦
٪	١٢,٤	٤,٨	٠,٤	٪	٤,٨	٣,٠	١,٢
١٩٩٠	٧٧	٣٩	١٧	١٩٩٧	٢٥	١٥	—
٪	١٥,٥	٧,٨	٣,٤	٪	٥,٠	٣,٠٠	—
١٩٩١	٢٤	١٥	٨	١٩٩٨	٥٢	٣١	٣
٪	٤,٨	٣,٠	١,٦	٪	١٠,٤	٦,٢	٠,٦
١٩٩٢	٤٠	١٣	٣	١٩٩٩	١٥	٧	—
٪	٨,٠	٢,٦	٠,٦	٪	٣,٠٠	١,٤	—
١٩٩٣	٢٨	٢٠	٧	مجـ	٤٩٨	٢٤٣	٦٧
٪	٥,٦	٤,٠	١,٤	٪		٤٨,٨	٪ ١٣,٥

يتضح من الجدول (٣) الذي يبين نسبة المقاييس والإختبارات التي اشارت إلي عدد المفردات التي يتكون منها المقياس ، وكذلك طريقة الإستجابة علي المقياس . يتضح أن هناك نسبة كلية تصل إلى ٤٨,٨ ٪ قد أشارت إلى عدد المفردات التى تتكون منها المقياس أو الإختبار ، وهذا يعني أن نسبة تصل إلي ٥٢,٢ لم تشر إلي عدد المفردات ، كذلك هناك نسبة

١٣,٥ ٪ فقط من تلك المقاييس اشارت إلي نمط أو طريقة الإستجابة . ومما سبق الإشارة إليه في الإطار النظري فإن طول الإختبار ، أى عدد المفردات المكونة له – ذو إرتباط مؤثر بطرق حساب صدق وثبات الإختبار . كذلك فإن نمط أو طريقة الإستجابة تمثل أحد العوامل التي تحدد طريقة حساب ثبات الإختبار ، فطريقة تحليل التباين بإستخدام معادلة K.R.20 حينما تكون طريقة الإستجابة Yes & Non أو (١) ، صفر ، أو (نعم) ، (لا) ، أو موافق وغير موافق . بينما الإختبارات أو المقاييس التي تكون الإستجابة علي مفرادتها على متصل ، يفضل معها إستخدام معادلة الفاكرونباخ .

جدول (٤) طرق حساب الثبات

م	السنة	عدد المقاييس	طرق حساب الثبات						م	السنة	عدد المقاييس	طرق حساب الثبات					
			ت ن	عدد المكونة	ت التباين	الإعادة	أخرى	لم تذكر				ت ن	عدد المكونة	ت التباين	الإعادة	أخرى	لم تذكر
١	١٩٨٥	٢٨	١	٢		٤		١٨	٨	١٩٩٤	٣٧	٥			١٣	٤ الفا ٣ س	١١
	٪									٪		-					
٢	١٩٨٦	٦٣	٨	١	٤	١٤	١ س ١ ك	٣٤	٩	١٩٩٥	٢٣	٢		٢	١٢	٣ الفا	٣
	٪	-								٪							
٣	١٩٨٩	٦٢	١٨		١	١٤	٢ جت ١ س ٢ ك ٤ س	٣٦	١٠	١٩٩٦	٢٤	٢	–	١	١٠	٢ س ٢ الفا	٦ ١
	٪									٪							
٤	١٩٩٠	٧٧	٦		٦	١٤	٣ الفا ٤ س	٣٦	١١	١٩٩٧	٢٥	٥	٣	٣	١٣	١ س ٢ الفا	٤
	٪									٪							
٥	١٩٩١	٢٤	٨	١١	–	٧	١ س ٢ الفا		١٢	١٩٩٨	٥٤	٦٠	-		٢٢	١٢ س ٤ الفا ١ ج	٥
	٪			-						٪							
٦	١٩٩٢	٤٠	٣			١٩	١ س ١ مس ٢ الفا	١٦	١٣	١٩٩٩	١٥	٦			٥	٥ الفا ١ ج	١
	٪									٪							
٧	١٩٩٣	٢٨	٦		٢	١١	٣ الفا ٣ س	١٠		مج	٤٩٨	٦٨	١٧	١٩	١٥٨	٣٥ الفا ٣٠ س ٣ ج ٤ مس	١٨٠
	٪									٪		١٣,٧	٣,٤	٣,٨	٣١,٧		٣٦,١

يتضح من جدول (٤) أن اكثر طرق حساب الثبات للمقاييس من الاختبارات المستخدمة فى البحوث هى طريقة اعادة الاختبار Test - Retest حيث قيمة ت =١٥٨ مرة بنسبة مئوية بلغت ٣١,٧ ٪، ويلى ذلك طريقة التجزئة النصفية Split - Half حيث قيمة ت= ٦٨ مرة بنسبة مئوية ١٣,٧ ٪ ويأتى بعد ذلك طريقة تحليل التباين باستخدام معادلة كيودر– ريتشاردسن K. R 20 حيث ت = ١٩ بنسبة ٣,٤ ٪ . وكانت أقل طرق حساب الثبات

استخداماً هى طريقة الصور المتكافئة حيث ت = ٢ بنسبة ٠,٥٤ ٪ .

وقد قام بعض الباحثين بحساب الثبات باستخدام معادلة جتمان Guttman ت = ٣ بنسبة ٠,٦ ٪ ومعادلة الفا كرونباك Alfa Chronbach ت = ٣٥ بنسبة ٧ ٪ وقام اخرون بحساب الثبات بطريقة الاتساق الداخلى Internal Consistency ت = ٢٨ بنسبة ٧,٦٣ ٪ . بينما بلغ تكرار اعداد المقاييس التى لم تذكر شيئا عن طرق حساب الثبات١٨٠ بنسبة٣٦,١ ٪ وبالقطع هى نسبة عالية خاصة فى مجال علم النفس بصفة عامة والقياس النفسى على وجه الخصوص . هذا فضلاً عن عدم اشارة أى من الباحثين إلى أى حالة من الحالات الخاصة فى حساب الثبات وخاصة أولئك الذين استخدموا مقاييس الابتكار أو المقاييس الاسقاطية أو الاختبارات المرجعة إلى المحك . كذلك لم يفكر الباحث فى استخدام أى طريقة من تلك الطرق التى اقترحها المتخصصون فى القياس النفسى لزيادة معامل الثبات أو دراسة العلاقة بين معامل الثبات ومعامل الارتباط بين الاختبارات .

جدول (٥) طرق حساب الصدق

م	السنة	عدد المقاييس	طرق حساب الصدق						
			محكمين	تكوينى	اتساق	عاملى	محك	أخرى	لم يذكر
١	١٩٨٥	٢٨	١	–	–	٢	٢	٢ من ٤ تج	١٨
٢	١٩٨٦	٦٣	٨	–	٣	١٠	١١	٥ ذاتى ١ س ١ م ط	٢٥
٣	١٩٨٩	٦٢	٧	–	–	٥	٩		٣٠
٤	١٩٩٠	٧٧	٧	–	٨	٩	١١	–	٤٧
٥	١٩٩١	٢٤	٨	–	٢	٣	١٢	مفردات ١ م ٠ ط	٦
٦	١٩٩٢	٤٠	٩	١	٤	٢	٤	١ تمييزى	٢١
٧	١٩٩٣	٢٨	٥	–	–	٢ .	٣	٥ تمييزى	٨
٨	١٩٩٤	٣٧	١١	–	–	٦	٩	٢ ذاتى	١٠
٩	١٩٩٥	٢٣	٨	–	٢	٩	٢	١ ذاتى ١ تميز	٥
١٠	١٩٩٦	٢٤	٥	١	٣	٧	١	١ ذاتى ١ تميز ١ ظ	٥
١١	١٩٩٧	٢٥	٥	–	١	٥	٥	١ ظ ١ منط ١ ذاتى ١ تميز	٣
١٢	١٩٩٨	٥٢	١٠	٢	٦	٦	٩	٢ م ط ٢ تج ٦ تميز	٢٣
١٢	١٩٩٩	١٥	٢	–	١	–	٧	–	٤
		٤٩٨	٨٦ ١٧,٣	٤ ٠,٨	٣٠ ٦,٠	٦٦ ١٣,٢	٨٥ ١٧,١		٢٠٥
		مج ٪							

يتضح من جدول (٥) أن اكثر الطرق المستخدمة فى حساب الصدق هى صدق المحكمين ت = ٨٦ بنسبة ١٧,٣ ٪ ثم الصدق المرتبط بالمحك ت = ٨٥ بنسبة ١٧,١ ٪ ثم

الصدق العاملى ، ت = ٦٦ بنسبة ١٣,٣ ٪ ثم الاتساق الداخلى ت = ٣٠ بنسبة ٠,٦ ٪ ثم الصدق التكوينى ت = ٤ بنسبة ٨ ٪ وهناك عدة طرق اخرى أستخدمت فى حساب الصدق منها الصدق الذاتى ت = ٧ بنسبة ١,٤ ٪ والصدق التميزى ت = ١٥ بنسبة ٣ ٪ والمقارنة الطرفية ت = ٤ بنسبة ٠,٨ والصدق المنطقى ٠,٢ ٪ . وبلغ عدد الاختبارات التى لم يشر فيها الباحث الى اى طريقة من طرق حساب الصدق ت = ٢٠٥ بنسبة ٤١,١٦ ٪ وهى نسبة عالية جداً فى مجال القياس النفسى ويعزو الباحث ذلك إلى عدم إهتمام الباحثين وأيضاً عدم معرفة بعضهم بأهمية مفهوم صدق الاختبار وانواعه وطرق حسابه المختلفة . كذلك استخدام بعض الباحثين للطرق السهلة والمألوفة لهم منذ دراستهم في المرحلة الجامعية الأولي . ويتضح ذلك من ترتيب النسب المئوية لطرق حساب الصدق حيث بلغت نسبة صدق المحكمين ١٧,٣ ٪ وهى طريقة لا تتطلب من الباحث بذل أى جهد . وكذلك الصدق الظاهرى الذى يعتمد عليه بعض الباحثين وهو طريقة لاتصلح بمفردها للحكم على صدق الاختبار والاعتماد على النتائج التى يحصل عليها الباحث كما ذكرت انستازى (١٩٨٢) . ثم يأتى بعد ذلك صدق المحك الذى بلغت نسبة ١٧,١ ٪ وهى نسبة عالية ايضاً. ويعزو الباحث ذلك إلى ان بعض الباحثين يستخدم فى دراسته اكثر من اداه للقياس وغالبا ما تكون الخصائص أو السمات المقاسة فى مجال واحد (معرفى او وجدانى) ثم يقوم بحساب الارتباط بين نتائج الاختبارين ويعتبر ذلك صدقاً تلازمياً . ثم الصدق العاملى الذى بلغت نسبته ١٣,٣ ٪ وهى نسبة مرضية وطريقة لا يشوبها أى عيوب حيث يستخدم الباحث البرامج الاحصائية المتخصصة فى الحاسب الالى (التحليل العاملى) وهو أسلوب إحصائى من الأساليب التى أثرت فى نمو وازدهار حركة القياس النفسى على مر التاريخ . ثم تأتى طريقة الاتساق الداخلى بنسبة تصل إلى ٦ ٪ ويرى الباحث أن هذه الطريقة – الاتساق الداخلى – تمثل بيت القصيد فى هذه الدراسة . حيث انه سبق ان استخدمت تلك الطريقة ايضا على انها طريقة من طرق حساب الثبات ، وبلغت نسبة الاختبارات التى تم حساب ثباتها بطريقة الاتساق الداخلى ٦ ٪ أيضًا وهى نفس نسبتها عند استخدامها فى حساب الصدق .

والسؤال المطروح الآن هل الاتساق الداخلي طريقة من طرق حساب الصدق أم طريقة لحساب الثبات ؟؟

وللاجابة عن هذا السؤال من خلال عرض الاطار النظرى لمفهومى الثبات والصدق وانواعها المختلفة وطرق حساب كل منهما . يرى الباحث أن مصطلح الاتساق الداخلى Internal Consistency اكثر ارتباطا بمعنى الثبات حتى أنه من طرق حساب الثبات الرئيسية طريقة الاتساق الداخلى وهى التى تعتمد فى جوهرها على مدى الارتباط بين مفردات الاختبار الذى يحسب بعده طرق مختلفة مثل : معادلة سبيرمان براون للتجزئة النصفية ومعادلة رولون Rulon ومعادلة كيودر ريتشاردسن ٢٠ K-R20 ومعامل الفا كروبناخ – فكل الطرق أو المعادلات أو الصيغ السابقة تعطينا معامل للثبات يعرف على أنه

مؤشر للاتساق الداخلى أما استخدام بعض الباحثين لطريقة الاتساق الداخلى فى حساب صدق الاختبار فتعتمد على حساب الارتباط بين كل مفردة والمجموع الكلى لدرجات الاختبار ، أوالارتباط بين المكونات الفرعية للاختبار ، أو الارتباط بين المفردة ومجموع درجات البعد فى الاختبار ، أو المقارنة بين اداء المجموعات المتطرفة على الدرجة الكلية للاختبار ، و غير ذلك من الطرق ويتضح من خلال ماسبق ان الارتباطات الخاصة بالاتساق الداخلى سواء ما يبني على أساس المفردات أو مايبنى على اساس الاختبارات الفرعية هى اساساً قياسات لتجانس الاختبار وحيث انها تساعد على تحديد مجال السلوك أو السمة التي يعتبر الإختبار عينة ممثلة لها . فإن درجة التجانس هذه يكون لها بعض الدلالة لصدق الاختبار التكوينى..

ومع هذا فإن الباحث يتبنى رأى انستازى (١٩٨٢) حيث ترى ان اسهام بيانات الاتساق الداخلى لصدق الاختبار محدودة . وفى ظل غياب البيانات التى نحصل عليها من خارج الاختبار أو محك خارجى فإننا نعلم القليل عما يقيسه الاختبار – أى عن صدق الاختبار–.

ويسوق الباحث بعض الأمثلة التى يتضح من خلالها ضرورة أن يأخذ الباحث بعين الإعتبار بعض المفاهيم الضرورية المتعلقة بإستخدام المقاييس والإختبارات النفسية .

* قام باحث بحساب ثبات استبيان كائل للشخصية بإستخدام معادلة K.R20 ، وهذه المعادلة تستخدم فقط عندما يكون الإستجابة علي المقياس (نعم أو لا) أو (١ ، صفر) أو (موافق / غير موافق) بينما إستبيان كائل ذي ثلاثة بدائل ، فكان لزاماً علي الباحث إستخدام طريقة أخري لحساب الثبات .

* في بحث آخر بإستخدام معادلة الفاكرونباخ في حساب ثبات استبيان مكون من مقياسين فرعيين كل منهما يمثل بـ ٧ مفردات . وكما عرض الباحث في الإطار النظري فإن هذه المعادلة (الفاكرونباخ) تستخدم مع المقاييس ذات التجانس بين المفردات (المشبعة علي عامل واحد) . وهذا قد يفسر إنخفاض معاملات الثبات بالإضافة إلي قلة عدد المفردات وهو من العوامل المؤثرة فى ثبات المقياس .

* وفي بحث آخر تم إستخدام أربعة إختبارات نفسية ولم يتم الإشارة إلي أي من الخصائص السيكومترية لهذه الإختبارات حيث ذكر إسم الإختبار فقط .

– ويأمل الباحث ان تكون نتائج هذه الدراسة قد وضعت حداً لبعض اشكاليات القياس النفسى وقدمت بعضاً من المفاهيم الخاصة بالقياس ووضعت بعضاً من النقاط على الحروف عند استخدام الادوات والمقاييس النفسية بما يسهم فى تلافى الاخطاء التى يقع فيها بعض الباحثين فى مصر والعالم العربى.

والله ولي التوفيق.

المراجـــــع

١ -عبد العاطى الصياد (١٩٨٣) ، العينات وادوات القياس ، المعهد المصرى لتقويم البرامج ص ص (١ - ١٧).

٢- عبد العاطى الصياد (١٩٨٨) ، الدلالة العملية وحجم العينة المصاحبتين للدلالة الاحصائية لاختبار « تاء » فى البحث التربوى والنفسى العربى ، دراسة تقويمية ، المؤتمر الثامن ، البحث التربوى ، الواقع والمستقبل ، القاهرة ، رابطة التربية الحديثة ص ص (١٩٩ - ٢٣٣)

٣- عبد العاطى الصياد (١٩٨٩) جداول تحديد حجم العينة فى البحث السلوكى ، سلسلة بحوث تربوية القاهرة ، رابطة التربية الحديثة ، ص ص (١ - ٥٤).

٤- عبد العاطى الصياد (١٩٩٠) ، المعامل الفالكروبناخ ونوعية استجابة المستجيب لعبارة «اكتب اسمك اذا رغبت » ، المؤتمر السنوى السادس لعلم النفس فى مصر ، القاهرة ، الجمعية المصرية لدرسات النفسية .

5- Anastasi , A. , (1982) psychological testing (5 th ed) N . Y : Macmillan .

6- Anastasi , A & urbina . s . (1997) , (7 th ed) psychological tes ting , N . Y , prentic - Hall , Inc .

7- Allyn , M . J & yen ,W.M. (1978) , Introduction to measurmant theory , Montorry, california Brooks cole .

8- Aiken R. Lewis, (1982) , psychological testing and Assessment, (4thed) Boston, Allyn and Bacon

9- Brewer , James , k . (1972) , on the power of statistical test in the American Edacational Research .

J . , 9 . pp 391 - 402 (في عبد العاطى الصياد ١٩٨٩)

10- Canpbell , p.T & Fiske, D.W. (1959) , convergent and discoiminant validation , by the multitrait - multimethod matrix . psychological Bulletin , 56 , 81 - 105 .

11 -Cohen , Jacob , (1977) , statistical power analysis for the be havioral sciences , Academic press , Revised ed ., New york .

(فى عبد العاطى الصياد ١٩٨٩)

12- Cronbach , L . J (1990) , Essontials of psycho logical testing (5 th ed) N . Y Harper collins .

13 - Gregory J . R ., (1996) psychological testing , History , principles , and applications , second ed . Boston allyn and Bacon .

14- James p . scheetz & Anthony J . v ., (1980) Measuring criterion - Referenced test Reliability with a single test adminstration . paper presented at the Annual Meeting of the National council on Measurement in Education , Boston , (ERic No ED 194535)

15- John A . Emrick , (1971) , An Evaluation Model for Mastery testing , Journal of Educational Measurement , 8 , 4 , 321 - 26 , win 71 . (ERic No EJO 48077)

16- Kathy , G . (1978) , Multiple choice converted to true - False : Comparative Reliabilities and validities , paper presen ted at the Annual Meeting of the western psychepogical Association (san Francisco , california ,) (Eric No : ED 169573)

17- Kaplan M. Robet & sacuzzo p . Dennis (1989) psychological testing , principles , Applications , and Issues , second ed . California , Brooksl cole publishing Company .

18- Lennon , R . T . (19662) . A Comparison of results of three intelligence tests , In C . l . chase & H . G . Ludlow (Eds.), Readings in educational and psychlogical measurment, Boston : Houghton Mifflin.

19-Mary . M. & susan K .G ., (1983) , A Comparison of Reliability Estimates From single and Double adminstration of criterion - Referen ced tests , paper presented at the Annual Meeting of the American Educational Researeh Association . (ERlc No : ED 230580)

20- Mehrens , W . A & Lehman , I . J (1978) Measrment and Evaluation in education and psychology ,N.Y , Holt Rinebert & winston .

21- Messick , samuel , (1980) , test validity and the ethics of assessment . American psychologist , 35 , 1012- 1027 .

22- Murphy , k . R . & Davidshofeo , C . o , (1994) psychological testing : principles and application , Enslewood cliffs , N . J : prentice - Hall .

23- Peter A . H . & Lori , Rothenberg , L ., Applying the ApA / AERA / NCME " standards " Evidience for the validity and Reliability of three statewide teaching Assessment instruments . paper presented at the Annual Meeting of the American . Educatianal Research Association . (ERic No : E D 318778)

24- Stanley , J . C . (1971) . Reliability . in R . L . Thorndike (Ed) Educational measurment . washington , D C : American Council on Education .

25- Starnberg R.J. (1977) , intelligence , information processing and analogical reasoning :the componetial analysis ot human abilities Hillsdale, N J : Erlbaum . (in) Anastasi , (1982) .

26- Subkoviak , M.J. (1980) the reliability of mastery classification decislons . in R.A. Burk (ed) criterion - referenced medsurment : the state of the art . Baltimore : Johns Hopkins university press .