مفاهيم أساسية في التعامل مع الخصائص السيكومترية لأدوات القياس في البحوث النفسية والتربوية

Basic concepts in dealing with the psychometric properties of measuring instruments in psychological and Educational Research

د . علي سراوي[1،*] ط . د . محمد الجموعي سراوي[2]

[1] **جامعة الشهيد حمة لخضر . الوادي (الجزائر)، serraouiali1975@gmail.com**

[2] **جامعة الشهيد حمة لخضر . الوادي (الجزائر)، serraouidjemoi@gmail.com**

تاريخ الاستلام : 2022/06/22 ؛ تاريخ القبول : 2022/11/27

ملخص: البحوث في مجالات العلوم النفسية والتربوية والاجتماعية شأنها شأن بقية المجالات تسعى لتزويد البشر بالمعلومات والمعارف المكتشفة أو الخاضعة للبحث، وإن كانت هذه البحوث في ظاهرها تبدو كيفية إلا أن خضوعها للقياس ساهم في تكميمها وأضفى عليها شيء من الدقة والموضوعية، فإجابة الباحث عن مشكلة بحثه تتطلب منه جمع وتحليل البيانات والملاحظات المرتبطة بمشكلة بحثه، ومن أجل الحصول على البيانات عليه أن يختار أدوات قياس جيدة ومناسبة لدراسته تمتاز بخصائص سيكومترية ممثلة أساسا في الصدق والثبات وبدرجة أقل استخراج المعايير، ذلك أنه لا يمكن الوثوق في أداة القياس ولا في نتائجها ولا القرارات المتخذة عنها بدون توفر هذه الخصائص، ومن خلال الدراسة الحالية سيتم التطرق لمفهوم الخصائص السيكومترية لأدوات القياس وأهم هذه الخصائص وأنواعها وطرق حساب معاملاتها ومعنى هذه المعاملات المتحصل عليها.

الكلمات المفتاحية : خصائص سيكومترية، صدق، ثبات، معايير.

Abstract : Research in the fields of psychological, educational and social sciences, like other fields, seeks to provide people with information and knowledge discovered or subject to research, Although these researches appear to be qualitative on the surface, their measurement contributed to their muzzling and gave them some accuracy and objectivity, The researcher's answer to his research problem requires him to collect and analyze data and observations related to his research problem, and in order to obtain data, he must choose good measuring instruments suitable for his study that are characterized by psychometric characteristics represented mainly in honesty, constancy and to a lesser extent, extracting criteria, The current study will address the concept of psychometric characteristics of measuring instruments, the most important of these characteristics, their types, methods of calculating their coefficients and the meaning of these coefficients obtained.

Keywords : Psychometric properties, valitity, reliability, Standards

* المؤلف المراسل.

1- مقدمة:

يقدم القياس النفسي للباحثين في مجالات العلوم التربوية والنفسية والاجتماعية خدمة كبيرة تتمثل في توفير أدوات لقياس مختلف السمات والظواهر بهدف جمع البيانات والمعلومات المتعلقة بها، وعملية جمع البيانات هنا ليست غاية في حد ذاتها يسعى الباحث لتحقيقها، بل هي وسيلة لتحقيق أهداف أخرى، ذلك أن عمل الباحث لا يتوقف عند جمعه للبيانات بل ينطلق منها لإيجاد تفسيرات وصياغة استنتاجات واتخاذ قرارات، ومن أجل الوصول إلى نتائج حقيقية ودقيقة ينبغي اختيار وسائل قياس موثوق في نتائجها، الدقة في النتائج تحتاج إلى توفر أداة القياس على عدة خصائص تدعى الخصائص السيكومترية.

وتعد الخصائص السيكومترية من أكثر المفاهيم تداولا واستعمالا في مجال القياس النفسي، وتشير إلى توفر مجموعة مؤشرات وقيم في بيئة معينة هي بمثابة رخصة لاستعمال الاختبارات والمقاييس في الدراسات والوثوق في نتائجها والاعتماد عليها في تفسير النتائج وتعميمها والوصول إلى استنتاجات دقيقة، ورغم تعدد هذه الخصائص إلا أن أغلب الدراسات تركز على خاصيتين أساسيتين هما الصدق والثبات سيتم التطرق لهما، إضافة إلى خاصية جد هامة يتم إهمالها في كثير من الدراسات وهي المعايير والتي تحدد لنا دلالة الدرجة المتحصل عليها في الاختبارات والمقاييس، كما تمثل محك لتصنيف الأفراد في الاختبارات وتحديد مستوى أداء الفرد بالنسبة لمتوسط مجموعته في السمة المدروسة.

2 . الخصائص السيكومترية للمقاييس النفسية:

يقصد بالخصائص السيكومترية المؤشرات الإحصائية المشتقة من إخضاع أداة قياس لسلسلة من الإجراءات التجريبية قصد الكشف عن صلاحيتها لقياس ما وضعت لقياسه، وذكر (بوسالم، 2014، ص 60) أن " الخصائص السيكومترية للاختبار يقصد بها تلك الصفات الضرورية والمتعلقة بمدى فعالية بنود الاختبار وبالصدق والثبات وما يرتبط بهما من معاملات تمييز ومستويات السهولة والصعوبة ومعايير تفسير النتائج التي يتم التحقق منها بعد تطبيق الاختبار تطبيقا تجريبيا على عينة ممثلة للمجتمع تسمى بعينة التقنين "، وعليه فالخصائص السيكومترية عبارة عن مجموعة مؤشرات تعبر عن إمكانية الوثوق بنتائج الاختبار واستقرارها واتساقها والأحكام الصادرة عنها، وتتمثل أهم هذه الخصائص في:

2 . 1 . الصدق (Validity):

يُعد صدق الاختبار أو المقياس من أهم خصائص الاختبارات والمقاييس النفسية والتربوية الجيدة، ذلك لأن الصدق يتعلق بمدى نجاح الاختبار في قياس ما وضع لقياسه وتحقيق الهدف الذي صمم من أجله، وانطلاقا من الدرجات المحصل عليها من تطبيقه يتم التوصل بدقة إلى استدلالات معينة.

وأشار (Tyler & Walsh, 1979, p 29) إلى أن " صدق المقياس يعبر عن قدرة المقياس على قياس السمة التي أُعد لقياسها "، وأضاف (Stanley & Hopkins, 1972, p 101) أن" الاختبار الصادق هو الاختبار الذي يحقق الوظيفة التي وضع من أجلها الاختبار بشكل جيد، وبالدرجة التي يكون

فيها قادراً على تحقيق أهداف محددة "، في حين أوضح (Messick, 1989, p 18) أن" الاختبار يكون صادقا عندما تكون الاستدلالات المشتقة منه مناسبة ولها معنى ومفيدة ".

كما أن صدق الاختبار يرتبط بمدى نجاحه وصلاحيته لقياس ما وضع لقياسه، وبالقرار الذي سيتم اتخاذه تبعا لذلك، فالصدق لا يرتبط بأداة القياس ذاتها بل بطريقة تفسير الدرجات المستخرجة من تلك الأداة.

2 . 1 . 1 . تعريف الصدق: للصدق عدة تعاريف منها:

ما أورده (الطريري، 1997، ص 218 . 219) على لسان عدة باحثين ومنضمات ونذكر منها:

. ما ورد في معايير الاختبارات النفسية والتربوية المعتمدة من طرف كل من (AERA) و(APA) و(NCME) سنة 1985 بأن " الاختبار صادق بالدرجة التي تكون الاستنتاجات المبنية عليه مناسبة وذات دلالة وفائدة ".

. أما (ساكس، SAX.1980) فقد عرف الصدق بأنه " الدرجة التي تكون فيها المقاييس ذات فائدة في اتخاذ القرارات المتعلقة بهدف أو غرض من الأغراض ".

. في حين يصف كرونباخ (Cronbach. 1971) الصدق بأنه " العملية التي يبحث من خلالها مُطور الاختبار عن دليل يدعم به الاستنتاجات التي من الممكن أن يصل إليها من خلال نتائج الاختبار".

أما (عبد الهادي، 1999، 112) فقد عرف الصدق بأنه " قدرة الاختبار على قياس السلوك أو القدرة أو السمة التي وضع من أجل قياسها، وبذلك يعطي درجة دقيقة تمثل القدرة الحقيقية للمفحوص في هذا السلوك ".

وأوضح (بوسالم، 2014، ص 60) بأن " الصدق يشير إلى الاستدلالات الخاصة التي تعطينا إياها درجات الاختبار من حيث مناسبتها ومعناها وفائدتها، وتحقيق صدق اختبار معين معناه قيام الباحث بتجميع الأدلة التي تؤيد هذه الاستدلالات، ولذلك يشير الصدق في إحدى معانيه إلى مدى صلاحية استخدام درجات الاختبار في القيام بتفسيرات معينة ".

وأشار (معمرية، 2012، ص 180 . 181) لوجود ثلاثة مفاهيم أساسية تتعلق بصدق الاختبار هي:

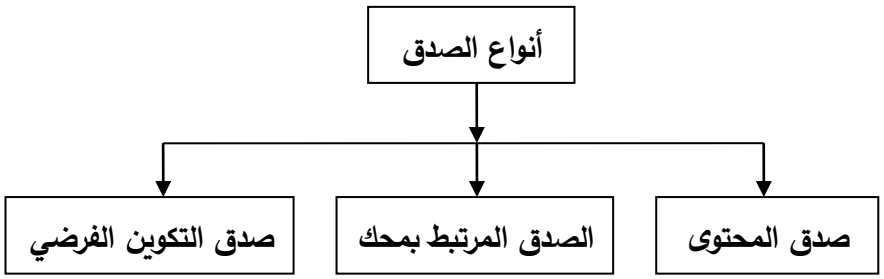
. أن يكون الاختبار قادرا على قياس ما وضع لقياسه، أي أن تكون بنود الاختبار على علاقة وثيقة بالخاصية التي تقيسها.

. أن يكون الاختبار قادرا على قياس ما وضع لقياسه فقط، أي أن يكون الاختبار قادرا على أن يميز بين الخاصية التي يقيسها، والخصائص الأخرى التي يمكن أن ترتبط بها.

. أن يكون الاختبار قادرا على التمييز بين طرفي الخاصية، أي أن يميز بين الأداء القوي والأداء الضعيف لدى أفراد العينة، أما إذا كانت درجات الاختبار جميعها متقاربة دل ذلك على صدق ضعيف للاختبار، أي أن الاختبار لم يقم بالمهمة الأساسية في عملية القياس وهي إظهار الفروق الفردية بين أفراد العينة.

ويمكن القول أن صدق الاختبار يرتبط بصلاحيته لتحقيق الغرض الذي وضع من أجله، وبإمكانية استخدام درجات المقياس في القيام بتفسيرات معينة، إذ لا يمكن الوثوق في نتائج اختبار لا يقيس ما وضع لقياسه.

2 . 1 . 2 . أنواع الصدق: للصدق عدة أنواع أهمها ثلاثة أنواع أساسية حددتها الرابطة الأمريكية لعلم النفس، وقد أشار (عبد الخالق، 1996، ص 52) " لوجود ثلاث طرائق أساسية حددتها معايير القياس التربوي والسيكولوجي الصادر عن الرابطة الأمريكية لعلم النفس عام 1985 وهي: صدق المحتوى والصدق المرتبط بمحك وصدق التكوين "، والشكل التالي يوضح أهم هذه الأنواع:



الشكل(01): أنواع صدق الاختبار

المصدر: بوسالم ، 2014، ص 63

أ . صدق المحتوى: ويسمى أيضا صدق المضمون أو الصدق المنطقي أو صدق عينة الاختبار، حيث تعد بنود (مفردات أو فقرات أو عبارات) الاختبار عبارة عن عينة محددة بشكل دقيق وشامل ومُمَثلة لمضمون وجوانب السمة موضع القياس، فكلما كانت بنود المقياس ممثلة لمختلف جوانب السمة المطلوب قياسها كان محتوى المقياس صادقا.

ويُقصد بهذا النوع حسب (بوسالم، 2014، ص 63) " مدى توافر جوانب السمة في بنود الاختبار، وهذا النوع من الصدق يتطلب تحليلا منطقيا لبنود الاختبار بهدف تحديد الوظائف والجوانب والمستويات الممثلة في بنوده ".

ويعد صدق المحتوى حسب رأي (خطاب، 2001، ص 161) " أكثر ملائمة للاختبارات التحصيلية، بحيث يتم تحديد درجة التطابق بين فقرات الاختبار وجدول المواصفات، فإذا ثبت أن فقرات الاختبار تغطي الموضوعات المحددة بجدول المواصفات وكذلك الأهداف وبالنسب المحددة فإن ذلك يعد دليلا على تمتع الاختبار بصدق محتوى عال ".

ويمكن القول أنه لتحقيق صدق المحتوى ينبغي على واضع الاختبار القيام بتحليل منطقي لبنود الاختبار ومطابقتها مع محتويات الجانب السلوكي المقاس لتحديد الجوانب والمستويات الممثلة فيه ونسبة كل منها في الاختبار أو المقياس ككل، وهذا لضمان تمثيل أو تغطية أكبر قدر ممكن من المجال السلوكي أو السمة المراد قياسها.

ب . الصدق المرتبط بمحك: المحك هو مقياس موضوعي مستقل عن الاختبار يقيس ما صُمم الاختبار لملاحظته أو التنبؤ به أو تشخيصه، ويمكن أن نحدد به مدى صلاحية الاختبار.

وذكر (فرج، 2007، ص 285) بأن " صدق المحك يعتمد على حساب الارتباط بين درجات الاختبار وبين محك خارجي مستقل، وهو السلوك نفسه أو النشاط الذي يتناوله الاختبار بالقياس ".

ويمكن التمييز بين نوعين من الصدق المرتبط بمحك هما الصدق التنبؤي والصدق التلازمي، وذلك بالاعتماد على استخدام محكات أداء مستقبلية أو محكات أداء راهنة.

ب . 1 . الصدق التنبؤي: يهدف هذا النوع من الصدق إلى التنبؤ بأداء الفرد في دراسة أو عمل لاحق يرتبط بما يقيسه الاختبار، ويستخدم لمعرفة مَن مِن الأفراد سيحقق نجاحا دراسيا أو من العمال سيكون أدائه عاليا....

ويتم فيه حساب الارتباط بين نتائج الاختبار ونتائج المحك الذي يتم الحصول عليها لاحقا، ويُستفاد منه لمعرفة قدرة وكفاءة درجات الاختبار وفعاليتها في التنبؤ بنتيجة معينة في المستقبل أو بسلوك الأفراد في وقت لاحق.

ب . 2 . الصدق التلازمي: ذكر (عبد السلام، 1960، 191) أن " هذا النوع من الصدق يُعنى بكشف العلاقة بين الاختبار ومحك تجمع البيانات عليه وقت أو قبل إجراء الاختبار، أي أننا نقارن بين درجات الأفراد على الاختبار ودرجاتهم على مقياس موضوعي آخر يحسب مركزهم فيما يقيسه الاختبار، على أن يعطى الأفراد درجات على المحك في نفس الوقت الذي نطبق فيه الاختبار.

ويتم فيه الكشف عن مدى الاتفاق بين نتائج الأفراد في الاختبار ونتائجهم في مقياس موضوعي أخر يمكننا من الحصول على بيانات عن السمة المُقاسة في نفس وقت تطبيق الاختبار أو في أوقات متقاربة.

ج . صدق التكوين الفرضي: ويُقصد به مدى قدرة وصدق ونجاح الاختبار في قياس خاصية أو مفهوم نظري لسمة معينة يفترض وجودها ولا يمكن ملاحظتها مباشرة، ولابد من تجميع للمعلومات المتعلقة بالسمة أو الخاصية موضع القياس وتحويلها إلى بنود لقياس المكونات الأساسية للسمة أو الخاصية، ويسمى أيضا صدق البناء أو صدق المفهوم.

ويوجد عدة أساليب يستخدمها الباحث للتعرف على صدق التكوين الفرضي والتي من خلالها يتم جمع المعلومات والبيانات التي لها علاقة بالسمة المقاسة منها:

. الارتباط مع مقاييس أخرى: فالحصول على معاملات ارتباط مرتفعة بين المقياس الجديد الذي يفترض أنه يقيس تكوينا فرضيا معينا (الذكاء مثلا) ومقاييس سابقة تقيس التكوين الفرضي ذاته يُعد دليلاً على صدق التكوين الفرضي.

ويُطلق على هذا النوع أيضا الصدق التطابقي أو الصدق التقاربي، لأنه يدرس العلاقة بين المقياس الجديد ومقياس آخر ثبت صدقه في قياس نفس السمة موضع القياس.

. التحليل العاملي: يهتم التحليل العاملي بتحديد درجة تشبع بنود المقياس بالعامل الذي تنتمي إليه، وهذه التشبعات تمثل معاملات الارتباط بين مفردات الاختبار والعوامل، فهو بذلك يعمل على تحديد المكونات الأساسية للمقياس (العوامل المشتركة)، ويطلق المعاملات المتحصل عليها مؤشرات الصدق العاملي.

وقدم (النبهان، 2004، ص 300 . 301) شرحاً للتحليل العاملي بأنه " أسلوب رياضي يمثل عددا كبيرا من العمليات والمعالجات الرياضية في تحليل الارتباطات بين المتغيرات (فقرات المقياس أو الاختبار)، ومن ثم تفسير هذه الارتباطات واختزالها في عدد أقل من المتغيرات تدعى عوامل، ويساعد التحليل العاملي في الحكم على أن السمة التي يقيسها الاختبار ذات بعد واحد أو متعددة الأبعاد، إذ أن هناك معامل ارتباط بين كل فقرة من فقرات المقياس وكل عامل من العوامل الناتجة، وتعرف هذه المعاملات بتشبع الفقرة بالعامل، والمهم في بناء المقياس أن يقل احتمال مساهمة الفقرة في قياس أكثر من عامل مما يضمن تفرد الفقرات في تمثيل العوامل ".

. الاتساق الداخلي: يقصد به الارتباط بين درجات بنود الاختبار، ويستخدم أسلوب الاتساق الداخلي أو التجانس للتأكد من أن جميع مفردات المقياس تقيس سمة أو خاصية واحدة، ويتم التحقق منه من خلال حساب معاملات الارتباط بين المقاييس الفرعية (العوامل أو الأبعاد أو المحاور) المكونة للمقياس والدرجة الكلية للمقياس، وأيضا حساب معاملات الارتباط بين كل درجة بعد والبنود التي تنتمي إليه، وكذلك بحساب معامل الارتباط بين درجات كل بند والدرجة الكلية للمقياس، فالبنود منخفضة الارتباط يستحسن استبعادها.

. الفروق بين المجموعات: أكد (علام، 2000، ص 222) بأن هذا الأسلوب " يُستخدم في الدراسات التي تهدف للمقارنة بين مجموعتين مختلفتين في ضوء الدرجات التي يحصل عليها الأفراد في اختبار أو مقياس معين، فإذا تبين أن هناك اختلافا واضحا بين درجات كل من المجموعتين، فإنه يمكن اعتبار ذلك أحد أدلة صدق التكوين الفرضي للاختبار ويُستبقى عندئذ على المفردات التي ميزت بدرجة أكبر بين المجموعتين "، ومن المنطقي أن نتوقع وجود فروق في السمات والخصائص (مثل القلق) بين الجنسين أو بين الفئات العمرية أو بين الشعوب...

إضافة إلى الأنواع الرئيسية الثلاثة سابقة الذكر توجد أنواع أخرى من الصدق نذكر منها:

الصدق الظاهري (السطحي): يعتمد هذا النوع من الصدق على الفحص المبدئي لمحتوى الاختبار (البنود) وتحديد ماذا تقيس، ثم مطابقة ذلك بالسمة أو الخاصية موضع القياس، فإذا بدا أن بنود الاختبار تقيس أو ترتبط ظاهريا بالسمة التي يهدف إلى قياسها الاختبار كان ذلك دليلا على صدقه ظاهريا (سطحيا).

وذكر (عبد الرحمن، 1998، ص 184) بأن هذا النوع " يقوم على فكرة مدى مناسبة الاختبار لما يقيس ولمن يُطبق عليهم، ويبدو مثل هذا الصدق في وضوح البنود، ومدى علاقتها بالقدرة أو السمة أو البعد الذي يقيسه الاختبار، وغالبا ما يقرر ذلك مجموعة من المختصين في المجال الذي يفترض أن ينتمي إليه الاختبار ".

الصدق الذاتي: وهو عبارة عن الجذر التربيعي للثبات، ولا يعتبر صدقا بالمفهوم الحقيقي للصدق في نظر كثير من العلماء، ويؤكد (معمرية، 2012، ص 241) بأن " الدرجات التجريبية للاختبار بعد تخلصها من أخطاء القياس (عند حساب الثبات) تصبح درجات حقيقية، ويمكن اعتبارها محكاً يُنسب إليه صدق الاختبار ".

2 . 1 . 3 . المفهوم التقليدي لصدق الاختبارات: ذكر (تيغزة، 2008، ص 2) أن " قيام نظرية الصدق التقليدية على فلسفة الأنماط (أنماط الصدق وأنواعه) إذ نظرت إلى مفهوم الصدق باعتباره خاصية وليس عملية ملتحمة بالدلالة النظرية للمفهوم "، وأضاف بأن " الدليل الإرشادي للقياس النفسي والتربوي الأول الذي تم نشره سنة 1954 من طرف الجمعية الأمريكية لعلم النفس لخص هذه الأنماط في أربع أنواع من الصدق: صدق المحتوى والصدق التنبؤي والصدق التلازمي وصدق المفهوم (التكوين الفرضي)، وقلص الدليل الإرشادي الصادر سنة 1966 هذا التصنيف إلى ثلاثة أنواع (التي أُشير إليها سابقا) إذ أبقى على: صدق المحتوى وصدق المفهوم أما الصدق التنبؤي والصدق التلازمي فأُدرجا تحت مسمى الصدق المعياري (المحكي)، وركزت النظرية التقليدية على التمايز الوظيفي لأنواع الصدق، أي أن لكل نوع وظيفة خاصة به تختلف عن وظائف الأنواع الأخرى من الصدق، وترتب عن ذلك أن كل نوع يناسب بعض أنواع الاختبارات والمقاييس ولا يناسب الأنواع الأخرى، فصدق المضمون يناسب اختبارات التحصيل، والصدق المحكي يناسب اختبارات القدرات والاستعدادات، وصدق المفهوم يناسب اختبارات الشخصية ".

2 . 1 . 4 . المفهوم الحديث لصدق الاختبارات: اتفق تيغزة (2008) وبوسالم (2014) على أن " النظرة التقليدية فيها نوع من التفكيك أو التجزئة التي تغيرت جذريا مع نظرية الصدق الحديثة، هذه الأخيرة ركزت على عدة أبعاد وجوانب تتمثل في:

أولا: التركيز على مدى ملائمة عملية تأويل أو تفسير درجات المقياس أكثر من الاهتمام بالبعد الكمي في الصدق، أي أن الصدق أصبح يهتم بطبيعة الأدلة والبيانات التي تقدم للدلالة على مدى دقة تأويل درجات أداء الأفراد على الاختبار أو طريقة تأويلها.

ثانيا: اعتبار الصدق صفة واحدة ونبذ فكرة تعدد أنواع الصدق، فأنواع الصدق أُدمجت كلها تحت نمط واحد سمي بصدق التكوين الفرضي أو البنائي أو صدق المفهوم، وأصبح هذا النوع يمثل شتى أنواع الصدق التقليدية ولا يمثل نوعاً واحداً فقط كما كان سائداً.

ثالثا: إن توحيد مفهوم الصدق أو مجاله لا يعني أن الأدلة أو البيانات الدالة عليه متماثلة أو واحدة في جميع الحالات وجميع أنواع الاختبارات، وهذا يعني أن البيانات التي يجمعها الباحث وهو في إطار تقديم أدلة عن صدق مقياسه إنما تختلف من حيث طبيعتها وطريقة الحصول عليها، لكن في الأخير تصب في معنى واحد هو صدق الاختبار.

رابعا: اهتمام الباحثين بالفروق الثقافية بين الأفراد والجماعات انتقل أثره إلى ميدان القياس النفسي، بحيث أضحى البعد القيمي الثقافي الاجتماعي لعملية القياس يشكل جزءا لا يتجزأ من مفهوم الصدق، ولذلك أضيف بعد جديد للصدق (نمط جديد من البيانات) يتمثل في المآل أو المترتبات أو النتائج الاجتماعية لعملية القياس، وأتخذ هذا البعد الجديد للصدق أحيانا مسمى صدق المآل أو المترتبات.

خامسا: لم يعد الثبات يُنظر إليه كمجال قائم بذاته ومستقل عن الصدق، وإنما أصبح ينظر إليه كنوع من البيانات التي يستدل بها على صدق الاختبار أو وجه من أوجه الصدق، وهو ما أصبح يُصطلح على تسميته بصلاحية الاختبار ".

2 . 2 . الثبات (Reliability): يُعتبر الثبات من أهم الخصائص السيكومترية للاختبار الجيد بعد الصدق، فالثبات يشير إلى مدى استقرار ظاهرة أو سمة معينة، ويتعلق بدقة الاختبار في قياس هذه السمة في أكثر من مرة، فالاختبار لا يمكن أن يكون صادقاً إلا إذا كان ثابتا لأن الدراسات الحديثة أصبحت تنظر للثبات بأنه مؤشر من مؤشرات صدق الاختبار، والثبات يتعلق بمقدار الثقة التي يمكن أن نضعها في نتائج الاختبارات من خلال ثباتها وعدم تغيرها، وقد أشار (علام، 2011، ص 231 . 233) إلى أن " درجات الأفراد على الاختبار ينبغي أن تعكس الفروق الفردية بين الأفراد في السمة المقاسة، وكما هو معلوم فإن درجات الاختبار تتأثر بدرجات متفاوتة بأخطاء القياس، ولذلك فإن درجة الفرد على الاختبار والتي يطلق عليها بالدرجة الملاحظة تتكون من درجتين هما: الدرجة الحقيقية والتي تمثل المقدار الحقيقي للسمة عند الفرد ودرجة الخطأ التي تجعل الدرجة الملاحظة الفعلية للفرد تختلف عن درجته الحقيقية بسبب أخطاء القياس، والفرق بين الدرجة الحقيقية والدرجة الملاحظة يعكس دقة القياس ".

أي أن : الدرجة الحقيقية = الدرجة الملاحظة – درجة الأخطاء العشوائية

أو: الدرجة الملاحظة (X) = الدرجة الحقيقية (T)+ درجة الخطأ(E)

وتتأثر الدرجات الملاحظة للاختبار بأخطاء القياس مما يؤدي إلى خفض ثباته، هذه الأخطاء عشوائية ترجع إلى عدة عوامل بعضها يتعلق بالاختبار مثل غموض تعليماته أو عدم وضوح بنوده، والبعض الأخر يتعلق بظروف الاختبار مثل قلة التهوية وضعف الإضاءة، وعوامل أخرى تتعلق بخصائص الأفراد المُختبرين مثل شعورهم بالتعب والملل أو قلة الدافعية لديهم للإجابة عن الاختبار.

2 . 2 . 1 . تعريف الثبات: يوجد عدة تعريفات للثبات منها:

. تعريف (Anna Anstasi & Susana Urbina, 1997, p 84) بأنه " اتساق القياسات التي يتم الحصول عليها من نفس الأفراد عندما يتم إعادة اختبارهم باستعمال نفس الأداة في مناسبات مختلفة وتحت نفس الظروف ".

. ونقل (Martyn. H, 1987 p 74) تعريف (Jonston & Penny pacpacker 1980) للثبات بأنه " قدرة الأداة على إعطاء نفس قيمة القياس عندما يتم تطبيق الأداة في نفس الظروف ".

. أما (فيصل، 1996، ص 22) فعرفه بأنه " مدى تطابق درجات أفراد مجموعة معينة على اختبار معين في كل مرة يعاد اختبارهم بنفس الاختبار، فهو يشير إلى الاستقرار في درجات الأفراد على نفس الاختبار".

. كما عرف (معمرية، 2012، ص 250) الثبات بأنه "مدى الدقة والاستقرار والاتساق في نتائج القياس، وضمان الحصول على نفس النتائج تقريبا إذا أعيد الاختبار على نفس الفرد أو نفس المجموعة من الأفراد، وهذا يعني قلة تأثير عوامل الصدفة على نتائج الاختبار".

. في حين عرفه (عبد الخالق، 1996، ص 45) بأنه " اتساق درجات الاختبار ودقة نتائجه وتحررها من تأثير المصادفة عندما يطبق على مجموعة محدودة من الأشخاص في مناسبتين مختلفتين يفصل بينهما زمن، أو عند اختبار الأشخاص أنفسهم بمجموعتين مختلفتين من البنود المتكافئة ".

ويتعلق الثبات بدقة الأداة في قياسها للقدرة أو السمة المطلوب قياسها مما يؤدي إلى الوثوق بالدرجات التي نحصل عليها من تطبيق هذه الأداة وبالتالي قلة تأثير عوامل الصدفة أو الأخطاء العشوائية على نتائج الأداة.

وبناء على ما سبق يمكن القول أن الثبات يُقصد به اتساق درجات الاختبار ودقة نتائجه وتحررها من أثر الصدفة عندما يطبق اختبار واحد مرتين على نفس الأشخاص في مناسبتين مختلفتين يفصل بينهما زمن محدد لا يسمح بألفة الاختبار أو حفظ بنوده، أو بإعطاء الأشخاص أنفسهم اختبارين متكافئين، أو بتطبيق اختبار واحد وتقسيمه إلى مجموعتين متكافئتين من البنود وتصحيح كل مجموعة منفصلة عن الأخرى، ويقاس الثبات إحصائيا بحساب معامل الارتباط بين الدرجات المتحصل عليها في الحالتين، ففي حالة ثبوت الدرجات وتطابقها قيل أن الاختبار ثباته عال.

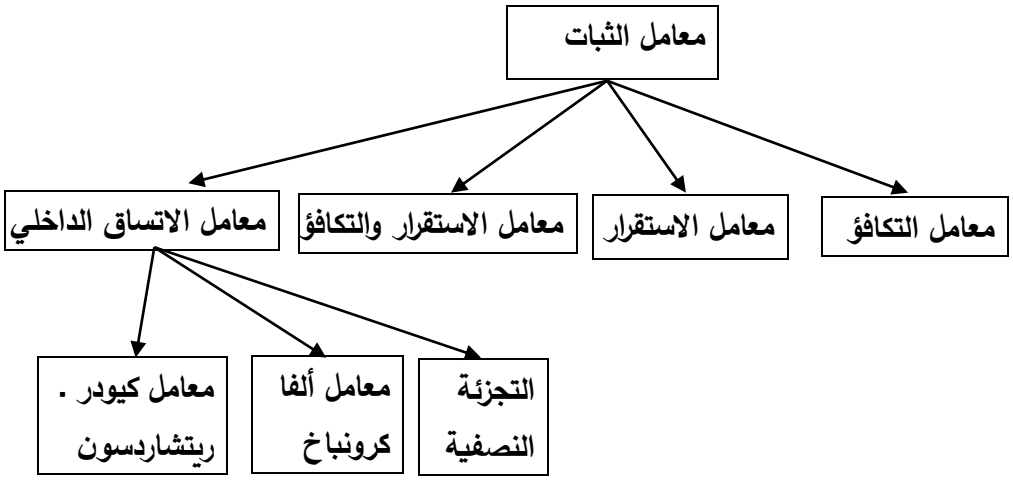
ومعاملات الثبات المرتفعة دليل على ثبات الاختبار واتساق فقراته، أما معاملات الثبات المنخفضة تدعو إلى الشك وعدم الوثوق في الاختبار وعدم استخدام نتائجه والاستدلال بها.

ومعامل الثبات حسب (عبد الكريم، 1996، ص316) هو " معامل نسبي (النسبة بين تباين الدرجات الحقيقية وتباين الدرجات الملاحظة) يختلف باختلاف متغيرات متعددة تتسبب في زيادة أو نقصان تباين درجات الأفراد على الاختبار (أخطاء القياس)، ونُعبر عنه رياضيا بمعامل الارتباط بين درجات الاختبار ونفسه، والتي تتراوح قيمته بين الصفر والواحد الصحيح ".

$$\text{معامل الثّبات} = \frac{\text{تباين الدّرجات الحقيقيّة}}{\text{تباين الدّرجات الملاحظة}}$$

ويشير (علام، 2000، 132) إلى أن " تباين الدرجات الملاحظة يساوي مجموع تباين كل من الدرجات الحقيقية وتباين درجات الخطأ، فإذا كانت الدرجات الملاحظة خالية من الخطأ فإن مقدار تباين الخطأ يساوي صفر، ويصبح مقدار تباين الدرجات الملاحظة مساويا لمقدار تباين الدرجات الحقيقية أي تكون الدرجات متسقة اتساقا تام وتكون قيمة معامل الثبات واحدا صحيحا، أما إذا كانت الدرجات غير متسقة على الإطلاق فمعنى ذلك أن الدرجات لا تمثل سوى أخطاء القياس، والفروق الملاحظة بين الأفراد لا يكون لها علاقة بالفروق الحقيقية بينهم، وتكون قيمة معامل الثبات صفرا ".

2 . 2 . 2 . طرق حساب معامل ثبات الاختبار: يعتمد ثبات الاختبار على مقارنة النتائج المحصل عليها من تطبيق اختبار ما مع نتائج الاختبار نفسه عندما يطبق مرة ثانية أو مع نتائج تطبيق صورة مكافئة له، وتوجد عدة طرق لحساب معامل ثبات الاختبارات يمكن تصنيفها في مجموعتين (طرق الاستقرار الزمني وطرق الاتساق الداخلي)، وأهم هذه المعاملات يمكن توضيحها من خلال الشكل التالي:



الشكل (02): أنواع معاملات الثبات

أ . طريقة إعادة الاختبار (معامل الاستقرار): تقوم هذه الطريقة على تطبيق الاختبار على عينة من الأفراد، ثم تطبيق نفس الاختبار مرة أخرى على نفس العينة بعد فترة زمنية عادة لا تقل عن أسبوع ولا تزيد عن ستة أشهر، ثم يُحسب معامل ثبات الاختبار وذلك بحساب معامل الارتباط بيرسون بين درجات الأفراد في التطبيق الأول والثاني، ويطلق على معامل الارتباط في هذه الحالة بمعامل الاستقرار.

وقد أشار (الدوغان، 2009، ص 61) بأن " معامل الثبات المحدد بهذه الطريقة معرض لكثير من الأخطاء التي تقلل من أهميته إذ أن إجراء الاختبار في المرة الثانية يتأثر بالعوامل التالية:

. تتأثر درجات المفحوصين في الإجراء الثاني بأثر انتقال التدريب والمِران وألفة المفحوص بموقف الاختبار.

. تتأثر درجات المفحوصين في الإجراء الثاني بطول الفترة بين الإجراءين حيث يتدخل عامل النمو خاصة إذا كان المفحوصين صغارا أو أذكياء، وكذلك تتأثر الدرجات إذ قصرت الفترة الزمنية حيث يتدخل عامل الحفظ والتذكر خاصة إذا كان الاختبار قصيرا ".

ب . طريقة الصور المتكافئة (معامل التكافؤ): تقوم هذه الطريقة على إعداد صورتين متكافئتين من الاختبار، بحيث تطبق الصورة الأولى وبعد فترة زمنية وجيزة يتم تطبيق الصورة الثانية على نفس المجموعة من الأفراد، ويحسب معامل الثبات بحساب معامل الارتباط بين درجات الأفراد في الصورتين المتكافئتين، ويطلق على معامل الارتباط في هذه الحالة بمعامل التكافؤ.

ويقوم الباحث بتصميم هاتان الصورتان المتكافئتان ويطبقهما معا على نفس المجموعة، ثم يقارن درجات الأفراد وكلما كانت الدرجات متشابهة والارتباط بينهما كبيرا كان الاختبار ثابتا، لكن هذه الطريقة يندر استخدامها وذلك لما تتطلبه من وقت لبناء صورتين متكافئتين للاختبار.

ج . طريقة إعادة الاختبار بصورة مكافئة (معامل الاستقرار والتكافؤ): يمكن للباحث الجمع بين الصور المتكافئة وإعادة الاختبار بتطبيق إحدى الصورتين وبعد مدة زمنية تطبق الصورة الثانية، وبحساب معامل الارتباط بيرسون للدرجات بين الصورتين يتم تقدير قيمة معامل الثبات، والتي تدل على الاستقرار والتكافؤ.

وتتطلب الطرق سابقة الذكر تطبيق الاختبار مرتين، ويصعب ذلك في كثير من الحالات مما يتطلب تقدير الثبات باستخدام اختبار واحد وتطبيقه مرة واحدة من خلال الطرق التالية:

د . طريقة التجزئة النصفية (Split-Half): تستخدم هذه الطريقة عندما يتعذر استعمال الطرق سابقة الذكر، بحيث يُحسب معامل الثبات من خلال تطبيق الاختبار مرة واحدة، ثم قسمة نتائج هذا التطبيق إلى جزءين متساويين وحساب معامل الارتباط بينهما ويكون هذا المعامل هو معامل الثبات، والتقسيم قد يكون تقسيم أسئلة الاختبار إلى فردي وزوجي، أو تقسيم الاختبار نفسه إلى نصفين متساويين في عدد البنود (النصف الأول والنصف الثاني)، ويتم حساب معامل الارتباط بين نصفي الاختبار باستخدام معامل بيرسون (Persson) وفق المعادلة التالية:

$$ر = \frac{ن \text{ مجـ } (س \times ص) - \text{مجـ } س \times \text{مجـ } ص}{\sqrt{[ن \text{ مجـ } س^2 - (\text{مجـ } س)^2] \times [ن \text{ مجـ } ص^2 - (\text{مجـ } ص)^2]}}$$

حيث:

ر: معامل الارتباط. ن: عدد أفراد العينة.

س: درجات الأفراد في المجموعة الأولى.

ص: درجات الأفراد في المجموعة الثانية.

ويترتب عن هذه التجزئة انخفاض ملحوظ في معامل الارتباط (معامل الثبات)، وذلك لأن معامل الثبات الذي نحصل عليه بهذه الطريقة هو معامل ثبات نصف الاختبار، وتتم معالجة هذا الانخفاض باستخدام معادلات إحصائية تسمى معادلات تصحيح الطول منها:

د . 1 . معادلة سبيرمان براون (Spearman-Brown): تقوم هذه المعادلة على إمكانية التنبؤ بثبات الاختبار إذا علمنا معامل ثبات نصفه، وتفترض تساوي تباين درجات كل من نصفي الاختبار، والمعادلة صيغتها الرياضية كما يلي:

$$ر_{11} = \frac{2\ ر_{21}}{1 + ر_{21}}$$

حيث:

$ر_{21}$: معامل ثبات درجات نصفي المقياس (الارتباط بين درجات نصفي المقياس).

$ر_{11}$: معامل ثبات المقياس الكلي.

د . 2 . معادلة جتمان (Gutman) : تستخدم هذه المعادلة عادة عندما لا يتساوى تباينا النصف الأول والثاني من الاختبار، وهذه المعادلة صيغتها كما يلي:

$$ر_{س\ س} = 2 \left(1 - \frac{ع^2_{أ} + ع^2_{ب}}{ع^2_{س}} \right)$$

حيث:

$ر_{س\ س}$: معامل الثبات.

$ع^2_{أ}$: تباين درجات أحد نصفي المقياس.

$ع^2_{ب}$: تباين درجات نصف المقياس الآخر.

$ع^2_{س}$: تباين الدرجات الكلية في الاختبار.

د . 3 . معادلة رولون (Rulon): تفترض هذه المعادلة أن تباين درجات الأفراد على الاختبار ككل يرجع إلى تباين حقيقي في مستويات القدرة المُقاسة، كما يرجع إلى أخطاء تجريبية في القياس، أي أن السبب في تباين درجات الأفراد على المقياس يعود لتباين قدرات الأفراد واختلاف مستوياتهم الحقيقية في السمة المقاسة (الفروق الفردية) وإلى وجود أخطاء القياس، والمعادلة صيغتها كما يلي:

$$\text{معامل الثبات} = 1 - \frac{ع^2_{ف}}{ع^2_{ك}}$$

حيث:

$ع^2$ ف: تباين الفرق بين درجات نصفي الاختبار.

$ع^2$ ك: تباين الاختبار ككل.

لكن ما يعاب على طريقة التجزئة النصفية أنه وُجد عند تغيير ترتيب الأسئلة أو تغيير طريقة التجزئة فإن قيمة معامل الارتباط تختلف أيضاً.

هـ . طريقة الاتساق الداخلي: تعتمد هذه الطريقة على مدى ارتباط البنود أو الفقرات مع بعضها البعض داخل الاختبار، وارتباط كل بند مع الاختبار ككل، وتقوم فكرة هذا المعامل على تقسيم الاختبار إلى عدد كبير من الأجزاء يساوي عدد بنوده، واعتبار كل بند أو فقرة اختباراً فرعياً، ويتم حساب الارتباطات بين جميع فقرات الاختبار وكلما زاد الاتساق بين هذه الفقرات زاد ثبات الاختبار ككل، ومن المعادلات التي تستخدم لحساب هذا المعامل:

هـ . 1 . معادلة" كيودرريتشاردسون 20: تستعمل هذه الطريقة في حساب معامل ثبات الاختبارات التي تكون درجات مفرداتها ثنائية أي إما واحد صحيح أو صفر مثل مفردات الاختيار من متعدد أو مفردات الصواب والخطأ.

وتعتمد هذه الطريقة حسب (فرج، 2007، 327) " على تجزئة الاختبار إلى عدد كبير من الأجزاء، يتكون كل جزء من بند واحد فقط من بنود الاختبار، وكلما كان هناك اتساق بين هذه الأجزاء أو البنود بقدر ما نحصل على تقدير جيد لثبات الاختبار كله، معنى هذا أن ثبات كيودر وريتشاردسون يفترض منذ البداية أن الاختبار أحادي البعد وشديد التجانس أي يقيس سمة أو وظيفة واحدة فقط "، ومعادلتها كالآتي:

$$\text{معامل التجانس} = \frac{\text{ن}}{\text{ن - 1}} \left(\frac{\text{ع}^2 - \text{مج س ص}}{\text{ع}^2} \right)$$

حيث:

ن: عدد مفردات الاختبار.

$ع^2$: تباين الدرجات الكلية في الاختبار.

س: نسبة عدد الأفراد الذين أجابوا بشكل صحيح على الفقرة (صعوبة الفقرة).

ص: نسبة عدد الأفراد الذين أجابوا بشكل خاطئ على الفقرة.

هـ . 2 . معادلة" كيودرريتشاردسون21 (Kuder– Richardson 21): ورمزها (KR–21)، وقد ذكر (علام، 2000، ص 164) " أن استخدام الصيغة (21) يتطلب توافر شرط أخر إلى جانب الشروط الأخرى التي تتطلبها الصيغة (20)، وهذا الشرط هو تساوي جميع مفردات الاختبارات والمقاييس في درجة صعوبتها أو على الأقل يتراوح متوسط صعوبة جميع المفرادات (0,50)، ولكن نظرا لصعوبة تحقق هذا الشرط في كثير من الاختبارات والمقاييس فإن الصيغة (20) ربما تكون أكثر ملائمة لهذه الاختبارات "، ومعادلة كيودرريتشاردسون21 كالآتي:

$$\text{معامل التجانس} = \frac{\text{ن}}{\text{ن - 1}} \left(1 - \frac{\text{م(ن - م)}}{\text{ن} \times \text{ع}^2} \right)$$

حيث:

م: متوسط الدرجات الكلية للاختبار.

ن: عدد مفردات الاختبار. $ع^2$: تباين الدرجات الكلية في الاختبار.

هـ . 3 . معامل ألفا (α) لكرونباخ: تستخدم هذه الطريقة في تقدير ثبات مقاييس الاتجاهات واستطلاع الرأي وفي مقاييس الشخصية، وتجدر الإشارة إلى أن طريقة ألفا تعطي الحد الأدنى للقيمة التقديرية لمعامل ثبات درجات الاختبارات، فإذا كانت قيمة ألفا مرتفعة فهذا يدل بالفعل على ثبات الاختبار، أما إذا كانت منخفضة فربما يدل على أن الثبات يمكن أن تكون قيمته أكبر من ذلك بإستخدام الطرق الأخرى، وأشار (النبهان، 2004، ص 248) إلى أن " هذه الطريقة اقترحها وطورها كرونباخ (Cronbach) سنة 1951، لتقدير ثبات الاتساق الداخلي للاختبار وهي تعميم لمعادلة (KR-20) عندما لا يتم تصحيح الفقرات بشكل ثنائي.

وأضاف (عبد الرحمن، 2008، ص 186) أن " معامل ألفا يمثل متوسط المعاملات الناتجة عن تجزئة الاختبار إلى أجزاء بطرق مختلفة، وبذلك فإنه يمثل معامل الارتباط بين أي جزئين من أجزاء الاختبار "، وهذه المعادلة صيغتها كما يلي:

$$\text{معامل } (\alpha) = \frac{\text{ن}}{\text{ن} - 1}\left(1 - \frac{\text{مج ع}^2_{\text{ن}}}{\text{ع}^2}\right)$$

حيث:

ن: عدد مفردات الاختبار.

$ع^2_{ن}$: تباين درجات كل مفردة من مفردات الاختبار.

$ع^2$: تباين درجات جميع المفردات.

2 . 2 . 3 . المفهوم الحديث لثبات الاختبارات: أشار (بوسالم، 2014، ص 78) بأنه " ومع النظرة الجديدة لمفهوم الصدق وأدلته أصبح الثبات كخاصية أساسية من خصائص الاختبارات الموضوعية المقننة، حيث أصبح جزءا من الصدق في حد ذاته، بل أصبح نوع من الأدلة الواجب توفيرها في اختبار معين للقول بقبول مستوى صدقه أو ما أصبح يصطلح على تسميته بصلاحية الاختبار للغرض الذي وضع من أجله ".

2 . 3 . المعايير: تعتبر المعايير من أهم العناصر الواجب توفرها في بناء الاختبارات، كونها تمثل محك لتصنيف الأفراد في الاختبارات، وتسمح بمقارنة درجة الفرد في الاختبار بدرجات أفراد عينة التقنين.

2 . 3 . 1 . تعريف المعايير: عرفت (الجلبي، 2005، ص 369 . 370) المعايير بأنها " نوع من الموازين أو المحكات التي تستخدم في تفسير الدرجات الخام التي يحصل عليها الفرد، كما أن إعدادها أيضا يعتمد على الدرجة الخام لعينة معيارية ممثلة للمجتمع الذي أعدت لها لأداة أو الاختبار، والهدف من استخدام المعايير هو تحديد مستوى أداء الفرد بالنسبة لمتوسط السمة".

فيما أكد (مقدم، 2003، ص 163) أن " المعيار مرجع مقنن يسمح لنا بفهم معنى الاختبار ومقارنة درجة الفرد بدرجات أفراد عينة التقنين، أي أن المعايير تحدد لنا دلالة الدرجة في الاختبار، ونطلق

على الدرجة التي يحصل عليها الفرد في المقياس الدرجة الخام ونطلق على درجات المعايير بالدرجة المحولة ".

فالمعايير تسمح بإعطاء معنى لدرجات الاختبار، وتبين وضع الفرد في الاختبار مقارنة بأداء أفراد المجموعة التي أجري عليها نفس الاختبار، حيث يتم مقارنة أداء الفرد بمتوسط أداء مجموعته لفهم دلالة الدرجة التي حصل عليها في اختبار ما.

2 . 3 . 2 . أنواع المعايير: إن الدرجة الخام التي يحصل عليها الفرد في اختبار ما لا تملك معنى معين إلا بعد تحويلها إلى معيار معين يعطيها التفسير المناسب ويُمكننا من تحديد مكانة الفرد أو مركزه بالنسبة لأقرانه في نفس المجموعة (عينة التقنين)، وتوجد عدة طرق لتحويل الدرجات الخام إلى معايير تستخدم في تفسير درجات الأفراد في الاختبار والتعرف على دلالاتها منها:

أ . المعيار الميئيني: تعرف المينيئات بأنها درجات تعبر عن ترتيب الأفراد تصاعديا بالنسبة لدرجاتهم في اختبار ما، والميئيني هو عبارة عن نقطة أو درجة من الدرجات الخام للاختبار يقابلها أو يقع تحتها نسبة معينة من الأفراد الذين طبق عليهم الاختبار، فمثلا الميئيني 80 هي الدرجة من الاختبار التي يقع تحتها 80% من الأفراد المختبرين، والفرد الذي درجته الخام تقابل الميئيني 80 يعني أنه يفوق 80 % من أفراد عينة التقنين، أي أن الميئينيات تحدد المركز النسبي للفرد في مجموعته، والدرجة الميئينية هي النسبة المئوية لعدد الأفراد في عينة التقنين الواقعين أسفل درجة خام معينة.

ب . الدرجة المعيارية: هي درجة مشتقة من الدرجات الخام للأفراد على اختبار ما، وتُعرف بأنها انحراف الدرجة الخام عن المتوسط مقدرا بعدد وحدات الانحراف المعياري.

وقد عرفها (مقدم، 2003، ص 163) بأنها " المسافة التي تبتعد فيها الدرجة الخام عن المتوسط الحسابي التي يعبر عنها بوحدات الانحراف المعياري ".

وتحسب بالمعادلة التالية:

$$\text{الدرجة المعيارية} = \frac{\text{الدرجة الخام} - \text{المتوسط الحسابي}}{\text{الانحراف المعياري}} \quad \text{أو} \quad د = \frac{س - م}{ع}$$

ولذلك فإن الدرجة المعيارية تحتفظ بجميع صفات الدرجات الخام، حيث يتم طرح من كل درجة مقدار ثابت (المتوسط) وقسمة الناتج على وحدة ثابتة (الانحراف المعياري).

لكن ما يُعاب على الدرجة المعيارية أنها قد تكون على شكل قيم سالبة (عندما تكون الدرجات الخام أقل من المتوسط) أو على شكل كسور.

ج . الدرجة التائية: ذكر (عبد السلام، 1960، ص 320) أنه " تم اقتراح درجة معيارية أخرى أُطلق عليها الدرجة التائية T–Score تهدف إلى التخلص من عيوب الدرجة المعيارية ، تقوم على تحويل الدرجات المعيارية إلى درجات معيارية إعتدالية متوسطها الحسابي 50 والانحراف المعياري لها 10 ".

وتحسب الدرجة التائية بالصيغة التالية: الدرجة التائية = 10× الدرجة المعيارية + 50

3 . خاتمة: إن قوة وجودة البحوث النفسية والتربوية مرهونة بدقة وصلاحية أدوات البحث المستعملة فيها والممثلة غالبا في الاختبارات والمقاييس، هذه الأدوات ينبغي أن تتوفر على مجموعة من الشروط التي ترقى لأن تكون خصائص حتى يمكن الوثوق في النتائج المتحصل عليها من خلال تطبيقها والاعتماد على نتائجها وتعميميها واتخاذ القرارات الملائمة في ذلك، وأهم هذه الخصائص والصدق والثبات والمعايير بدرجة أساسية، إذ يشير صدق المقياس إلى قدرته على تحقيق الهدف الذي وضع من أجله، أما الثبات فيقصد به دقة النتائج واتساقها، أما استخراج المعايير فيهدف إلى تحديد مستويات السمة المقاسة بما يتناسب مع كل فئة، ومع ذلك ينبغي أن تتوفر أداة القياس على عدة خصائص أخرى منها الموضوعية والشمولية والحصول على معاملات السهولة والصعوبة والتمييز.

. الإحالات والمراجع:

ـ بوسالم، عبد العزيز، (2014)، القياس في علم النفس والتربية، الأسس النظرية والمبادئ التطبيقية، ط 1، باب الزوار . الجزائر: دار قرطبة للنشر والتوزيع.

ـ تيغزة، أمحمد، (2008)، نظرية الصدق الحديثة ومتضمناتها التطورية لواقع القياس، جامعة الملك سعود، المملكة العربية السعودية، ندوة " علم النفس والتنمية الفردية والمجتمعية ".

ـ خطاب، علي ماهر، (2001)، القياس والتقويم في العلوم النفسية والتربوية والاجتماعية، ط 2، القاهرة: مكتبة الأنجلو المصرية.

ـ الدوغان، عبد الله، وأبو عوف محمد، طلعت، (2009)، حقيبة تدريبية في القياس والتقويم النفسي والتربوي، جامعة الملك فيصل، السعودية: مركز التنمية الأسرية.

ـ شاكر، الجلبي سوسن، (2005)، أساسيات بناء الاختبارات والمقاييس النفسية والتربوية، ط1، سوريا: مؤسسة علاء الدين للطباعة والتوزيع.

ـ الطريري، عبد الرحمن، (1997)، القياس النفسي . نظرياته وأسسه وتطبيقاته، ط 1، الرياض، المملكة العربية السعودية: مكتبة الرشد للنشر والتوزيع.

ـ عباس، فيصل، (1996)، الاختبارات النفسية تقنينها وإجراءاتها، ط 1، بيروت، لبنان: دار الفكر العربي.

ـ عبد الخالق، أحمد محمد، (1996)، قياس الشخصية، ط 1، جامعة الكويت: لجنة التأليف والنشر.

ـ عبد الرحمن، سعد، (1998)، القياس النفسي . النظرية والتطبيق، ط 3، القاهرة: دار الفكر العربي، 1998.

- عبد الرحمن، سعد، (2008) ، القياس النفسي . النظرية والتطبيق، ط 5، مصر: هبة النيل العربية للنشر والتوزيع.
- عبد السلام أحمد، محمد، (1960)، القياس النفسي والتربوي، ط 1، مصر: مكتبة النهضة المصرية.
- عبد الكريم، حبيب مجدي، (1996)، التقويم والقياس في التربية وعلم النفس، ط1، القاهرة: مكتبة النهضة المصرية.
- عبد الهادي، نبيل، (1999)، القياس والتقويم التربوي واستخدامه في مجال التدريس الصفي، الأردن: دار وائل للنشر.
- علام، صلاح الدين محمود، (2000)، القياس والتقويم التربوي والنفسي . أساسياته وتطبيقاته وتوجهاته المعاصرة، ط1، القاهرة: دار الفكر العربي.
- علام، صلاح الدين محمود، (2011)، القياس والتقويم التربوي في العملية التدريسية، ط4، عمان، الأردن: دار المسيرة للنشر والتوزيع.
- فرج، صفوت، (2007)، القياس النفسي، ط 6، القاهرة: مكتبة الأنجلو المصرية.
- معمرية، بشير، (2012)، أساسيات القياس النفسي وتصميم أدواته، القبة، الجزائر: دار الخلدونية للنشر والتوزيع.
- النبهان، موسى ، (2004)، أساسيات القياس في العلوم السلوكية، ط 1، عمان، الأردن: دار الشروق للنشر والتوزيع.

- Anne Anastasi& Susana Urbina, (1997) , Psychological Testing, New jersey, Prentice hall.
- Hammersley, Martyn, (1987), Some Notes on the terms Validity and Reliability.BritishEducational Research Journal.
- Messick , S. (1989): Validity in R.Linn (ed) : Educational Measurement (3rd ed) .N.Y.Macmilla.
- Stanley, C. J. & Hopkins, K. D., (1972): Educational and Psychological Measurement and Evaluation. New Jersey, Prentice-Hall.
- Tyler.L.E.&Walsh,W.B.(1979) . Tests And Measurements . New Jersey .Englewood Cliffs .Prentice-Hall. Inc.