University Mohamed Boudiaf Of M'sila

Laboratory Life Skills

Http://abs.univ-msila.dz/lls

skillslife250@gmail.com

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

المديرية العامة للبحث العلمي والتطوير التكنولوجي

المؤسسة: جامعة محمد بوضياف بالمسيلة

مخبر المهارات الحياتية

1985

جامعة محمد بوضياف - المسيلة

Université Mohamed Boudiaf - M'sila

شهادة مشاركة

يشهد رئيس الندوة الوطنية بمخبر المهارات الحياتية بأن الأستاذ : أ. حريزي بوجمعة – جامعة محمد بوضياف بالمسيلة

قد شارك في فعاليات الندوة الوطنية التي نظمتها فرقة مهارات التفكير والاستكشاف

" القياس النفسي والتربوي بين النظرية الكلاسيكية والنظرية المعاصرة " يوم 07 فيفري 2017

وهذا بمداخلة عنوانها : الصدق والثبات وفق نظرية القياس الكلاسيكية

رئيس الندوة الوطنية

مدير المخبر

المسيلة في: 2017/02/07

مدير مخبر المهارات الحياتية

د/ مجاهدي الطاهر

برنامج أشغال الندوة الوطنية الأولى:

القياس النفسي والتربوي بين النظرية الكلاسيكية والنظرية المعاصرة

الافتتاح على 08:30 صباحا
القرآن الكريم
كلمة رئيسة الندوة
كلمة مدير المخبر
كلمة العميد
المحاضرة الافتتاحية: أ. د. رابح قدوري
رفع الالتباس عن بعض الأخطاء في تناول مصطلحات وأدوات الإحصاء والقياس.

الجلسة الأولى: **رئيس الجلسة أ. د. رابح قدوري**		**10:30 -09:00**
01	القياس النفسي والتربوي (قضايا ومفاهيم أساسية) أ. بشير حبيش جامعة عمار ثليجي – الأغواط أ. سامرة خنفار جامعة عمار ثليجي – الأغواط	09:10 -09:00
02	استخدام تقنية شبكة التداعيات الترابطية في قياس التصورات الاجتماعية د. حسين مشطر جامعة 8 ماي 1945– قالمة	09:20 -09:10
03	مرتكزات نظرية الاستجابة للمفردة وأسباب ظهورها أ. عمار سويسي جامعة قسنطينة 2 أ. جمال ياحي جامعة الجزائر 2	09:30 -09:20
04	نظريات القياس النفسي التقليدية والحديثة: نقاط التلاقي وأوجه الاختلاف د. نصيرة لمين جامعة محمد بوضياف –المسيلة أ. عبد الحق بحاش جامعة محمد بوضياف– المسيلة	09:40 -09:30
05	الإطار المفاهيمي للقياس النفسي والتربوي ومستوياته أ. الزهراء بعيسى جامعة سطيف 2 أ. عقيلة ريغي جامعة سطيف 2	09:50 -09:40

06	صعوبات استخدام الإحصاء والقياس في إعداد مذكرات التخرج لدى طلبة الماستر 2 والليسانس بقسمي علم النفس والتربية البدنية أ. سليم عمرون جامعة محمد بوضياف – المسيلة أ. العيد قرين جامعة محمد بوضياف – المسيلة	10:00 -09:50
07	نماذج النظرية الحديثة في القياس: نظرية الاستجابة للمفردة أ. عبد الله عمارني جامعة عبد الحميد بن باديس– قسنطينة	10:10-10:00
08	الاتجاه الكلاسيكي والاتجاه المعاصر في القياس النفسي والتربوي د. سي محمد سعدية جامعة أكلي محند أولحاج –البويرة د. عفيفة جديدي جامعة أكلي محند أولحاج –البويرة	10:20-10:10
09	الصدق والثبات وفق نظرية القياس الكلاسيكية د. حمود طه جامعة محمد بوضياف– المسيلة أ. بوجمعة حريزي جامعة محمد بوضياف– المسيلة أ. فرحات بن ناصر جامعة محمد بوضياف– المسيلة	10:30-10:20
	مناقشة محتوى المداخلات	10:50 -10:30

10:50- 11:00 استراحة قهوة

	الجلسة الثانية: رئيس الجلسة: د. حمود طه	12:30 -11:00
01	مبادئ وأسس القياس في ظل النظرية الكلاسيكية أ. فاطمة عميرات جامعة قاصدي مرباح – ورقلة أ. توفيق بن يمينة جامعة أبو القاسم سعد الله – الجزائر 2	11:10-11:00
02	أهمية المعايير في النظرية الكلاسيكية في القياس النفسي أ. الزهرة بومهراس جامعة قاصدي مرباح– ورقلة أ. عبد القادر الأنصاري جامعة أحمد دراية – أدرار	11:20-11:10
03	من النظرية الكلاسيكية للاختبارات إلى نظرية إمكانية التعميم د. فاروق طباع جامعة مولود معمري – تيزي وزو	11:30-11:20

04	مزايا وانتقادات كل من النظرية الكلاسيكية والنظرية الحديثة في القياس د. وردة العزيز جامعة الجيلالي بونعامة – خميس مليانة د. صليحة لعزالي جامعة يحي فارس – المدية	11:40-11:30
05	القياس النفسي والتربوي أ. د. يامنة اسماعيلي جامعة محمد بوضياف – المسيلة أ. فيصل نويوة جامعة محمد بوضياف – المسيلة	11:50-11:40
06	أساسيات البحث في الخصائص السيكومترية للمقاييس النفسية والتربوية دراسة سيكومترية على مقياس ممارسة الحرية الأكاديمية للأستاذ الجامعي أ. خير الدين بن خزور جامعة لونيسي علي –البليدة 2	12:00-11:50
07	النظرية الكلاسيكية والمعاصرة بين المميزات والعيوب أ. ساعد بن سبع جامعة البليدة 2 أ. حمزة خوجة جامعة محمد بوضياف – المسيلة	12:10-12:00
08	مفاهيم أساسية حول القياس النفسي أ. أسماء لجلط جامعة قاصدي مرباح – ورقلة أ. عبد المليح نقبيل جامعة قاصدي مرباح – ورقلة	12:20-12:10
09	الصدق والثبات وفق نظرية القياس المعاصرة أ. د. رابح قدوري جامعة محمد بوضياف – المسيلة د. سامية ابراهيمي جامعة محمد بوضياف – المسيلة	12:30-12:20
	مناقشة محتوى المداخلات	13:00 -12:30

الفترة المسائية

	الجلسة الثالثة: **رئيس الجلسة: أ. د. زين الدين ضياف**	16:00 -14:00
01	مفاهيم أساسية في القياس النفسي أ. نزيهة عزري جامعة باجي مختار – عنابة	14:10-14:00
02	جدلية المعايير بين وضعية القياس والوضعية الاختبارية العيادية د. حدة ميمون جامعة محمد بوضياف –المسيلة أ. فاطمة الزهراء بوعلاقة جامعة محمد بوضياف– المسيلة	14:20-14:10

03	مدخل إلى نظرية الاستجابة للمفردة أ. ابتسام بوراس جامعة لونيسي علي – جامعة البليدة 2	14:30-14:20
04	مقارنة بين النظرية التقليدية للقياس ونظرية الاستجابة للمفردة (أنموذج راش) أ. عبد الكريم رحالي جامعة أبو القاسم سعد الله – الجزائر 2	14:40-14:30
05	مفهوم القياس النفسي والتربوي في ظل النظرية الكلاسيكية في القياس د. سعاد مخلوف جامعة محمد بوضياف – المسيلة أ. نسيمة محرز جامعة محمد بوضياف – المسيلة	14:50-14:40
06	نظرية السمات الكامنة أ. جاب الله سليم جامعة محمد بوضياف – المسيلة أ. محمد فارح بوزيد جامعة محمد بوضياف – المسيلة	15:00-14:50
07	بناء اختبار مهارات التفكير العلمي في مجال الظواهر الضوئية لمقرر العلوم الفيزيائية والتكنولوجيا مع حساب الخصائص السيكومترية له د. لخضر بن حامد جامعة أكلي محند أولحاج– البويرة أ. حسين عدلي جامعة أكلي محند أولحاج– البويرة	15:10-15:00
08	النظرية المعاصرة في القياس د. عاشور علوطي جامعة محمد بوضياف – المسيلة أ. محمد بن كيحول جامعة محمد بوضياف – المسيلة	15:20-15:10
09	نماذج نظرية الاستجابة للمفردة: نموذج راش أ. حمزة بركات جامعة محمد بوضياف – المسيلة أ. حنان فنيش جامعة الحاج لخضر – باتنة 1	15:30-15:20
مناقشة محتوى المداخلات		16:00 -15:30

الصدق والثبات وفق نظرية القياس الكلاسيكية

د. طه حمود

أستاذ محاضر بقسم علم النفس كلية العلوم الإنسانية والاجتماعية جامعة محمد بوضياف–المسيلة

hamd-taha@hotmail.fr

أ. حريزي بوجمعة

أستاذ مؤقت بقسم الإعلام والاتصال الرياضي معهد علوم وتقنيات النشاطات البدنية والرياضية جامعة محمد بوضياف–المسيلة

herizi boudj@hotmail.com

أ. بن ناصر فرحات

أستاذ مؤقت بقسم علم الاجتماع كلية العلوم الإنسانية والاجتماعية جامعة محمد بوضياف–المسيلة

bennaceurferhat@gmail.com

الملخص

حظي ميدان القياس منذ نشأته وحتى وقتنا الحاضر باهتمام كبير وجهد متواصل من قبل المختصين في القياس النفسي والتربوي، مما تمخض عنه إحداث نقلة نوعية وتطور كبير ومستمر في إجراءاته وأدواته وفلسفته؛ بغرض الوصول إلى أقصى درجات الدقة في قياس السلوك الإنساني. ومن بين هذه التطورات التي تخللت تاريخه ظهور النظرية الكلاسيكية في القياس. ومن هذا المنطلق تناولت هذه الورقة البحثية هذه النظرية، وذلك من خلال التعرف على مفهومها وأهم افتراضاتها؛ وما تنطوي عليه من مفاهيم ومبادئ تتعلق بالخصائص السيكومترية للاختبار والمتمثلة في الصدق والثبات.

الكلمات المفتاحية: النظرية الكلاسيكية في القياس، الصدق، الثبات.

Abstract

Measurement field received from its inception until the present day with great interest and continuing effort by specialists in educational and psychological measurement, which resulted in significant events of evolution paradigm shift in its procedures, tools and philosophy; to reach maximum precision in the measurement of human behavior. These included the date the emergence of the classical theory of measurement. In this sense this paper dealt with this theory, and that by identifying the most important sense assumptions; and the underlying concepts and principles regarding psychometric properties of test validity and reliability.

Keywords: classical theory of measurement, validity, reliability.

مقدمة:

تعتبر النظرية الكلاسيكية في القياس أولى نظريات القياس، حيث نشأت في بدايات القرن العشرين من إسهامات العالم سبيرمان(Spearman,1927) من خلال تقديمه للكثير من المصطلحات المعروفة كالدرجة

الحقيقية؛ والثبات؛ والصدق؛ ونظريته في الذكاء المسماة نظرية العاملين، بالإضافة إلى إسهامات علماء آخرين، وتبلورت هذه النظرية على يد العالم جاليكسون(Gulliksen, 1950) من خلال نشره لكتاب بعنوان "نظرية الاختبارات العقلية"، ثم قام كل من العالمين لورد ونوفيك(Lord & Novick,1968) بتطوير النظرية واستخدامها بشكل واسع في المجال النفسي والتربوي، واعتُبر كتابهما "النظريات الإحصائية لدرجات الاختبار النفسي" من أهم الكتب التي طورت وجددت النظرية الكلاسيكية في القياس (زكري، 2010، ص.29).

وحظيت هذه النظرية باهتمام المشتغلين بالقياس النفسي والتربوي، بسبب استخدامها كمرجع أساس في بناء الاختبارات والمقاييس النفسية والتربوية، واعتمادها كمنطلق لإصدار أحكام تتعلق بالتحصيل، والأداء، والتشخيص، والذكاء، وغيرها. ورغم هذه المزايا التي اتصفت بها إلا أنها لم تخلُ من عيوب، إذ تعرضت إلى الكثير من الانتقادات من قبل خبراء القياس خلال العقود الماضية، بسبب أوجه القصور التي انطوت عليها، وكان من أبرز عيوبها قلة دقتها وموضوعيتها، ومع ذلك مازالت تستخدم بشكل كبير من قبل الباحثين في العلوم السلوكية، ولاتزال أكثر نظريات القياس انتشارا حتى الآن في كثير بلدان العالم، وخاصة البلدان العربية، نظرا لبساطة مبادئها وسهولة فهمها واستخدامها.

ويطلق على هذه النظرية أحيانا نظرية الدرجة الحقيقية، لكون الأساس النظري الذي تبنى عليه هو نموذج رياضي يعرف بنموذج الدرجة الحقيقية، والذي ينطلق من الافتراض القائل بأن لكل فرد قدرا معينا من السلوك غير الملاحظ يعبر عنه بالدرجة الحقيقية؛ ولا يمكن قياسه، ولكن ما يتم قياسه هو السلوك الملاحظ المتمثل في مجموع الاستجابات الصحيحة للفرد(الدرجة الملاحظة)، والتي يتم قياسها على أساس أنها تتضمن قدرا معينا من الخطأ، والمعبر عنه بدرجة الخطأ، بالإضافة إلى الدرجة الحقيقية. أي أن الدرجة الملاحظة تختلف عن الدرجة الحقيقية، وهذا الاختلاف يسمى بالخطأ العشوائي للقياس(Suen,1990, P.28) . لذلك يجب أن يبنى الاختبار بطريقة تمكننا من قياس الدرجة الحقيقية لكل فرد، بمعنى آخر يجب أن نأخذ كل الاحتياطات التي تقلل من تأثير الخطأ العشوائي؛ بحيث تكون نسبته في الدرجة الملاحظة أقل ما يمكن. (الدوسري، 2000، ص.83)

أهداف الدراسة:

1- التعرف على مفهوم النظرية الكلاسيكية في القياس.

2- التعرف على أهم الافتراضات المتعلقة بنظرية القياس الكلاسيكية.

3- التعرف على الخصائص السيكومترية للاختبار (الصدق، الثبات).

1- مفهوم النظرية الكلاسيكية في القياس:

تعرّف النظرية الكلاسيكية في القياس "Classical Measurement Theory"على أنها: "إحدى نظريات القياس التي ترتكز على مفهوم الدرجة الحقيقية والدرجة الخطأ، إذ يفترض أنه لو أمكن أن نجري الاختبار عدد من المرات على الفرد بعناصر جديدة وتحت ظروف مختلفة، فإننا نحصل على درجات ملاحظة مختلفة متوسطها هو أقرب تقدير غير متحيز لقدرة الفرد أو درجته الحقيقية". (Randall, 1998, p.4)

ويرى العنزي (2010، ص.12) أن النظرية الكلاسيكية في القياس هي: "نموذج بسيط يحتوي تعريف لمفهوم الدرجة الملاحظة والدرجة الحقيقية ودرجة الخطأ العشوائي في القياس وافتراضات تتعلق بينها".

ويذكر الموسوي (2014، ص.15) أن النظرية الكلاسيكية في القياس هي: "إحدى نظريات القياس التي ترتكز على مفهوم الدرجة الحقيقية ودرجة الخطأ، وتفترض إمكانية تطبيق الاختبار عدة مرات على الفرد في ظروف مختلفة، والحصول على درجات ملاحظة يمثل متوسطها درجة الفرد الحقيقية التي تعتبر مقدارا ثابتا يعكس قدرته، ومعامل خطأ يضاف إليها أو ينقص منها".

ومن التعاريف السابقة نجد الدرجة التي يحصل عليها الفرد بعد تصحيح الاختبار تسمى الدرجة الملاحظة "Obtained Score" أو الدرجة الخام(الكلية)، وهذه الدرجة هي عبارة عن مجموع الدرجة الحقيقية "Ture Score" ودرجة خطأ القياس "Error Score"، ومنه فنظرية القياس الكلاسيكية تقوم على مفهوم الدرجة الحقيقية التي تعكس مقدار ما يمتكه الفرد حقيقة من السمة المقاسة، والتي هي عبارة عن الفرق بين الدرجة الملاحظة(الكلية) ودرجة الخطأ في القياس، مع العلم أن درجة الخطأ قد تكون موجبة أو سالبة أو معدومة. وأن الدرجة الحقيقية لا يمكن معرفتها ولا قياسها، بل يُستدل عليها من خلال الدرجة الملاحظة.

ويمكن توضيح ذلك على النحو التالي:

الدرجة الملاحظة = الدرجة الحقيقية + (± درجة الخطأ)

ومنه: **الدرجة الحقيقية = الدرجة الملاحظة – (± درجة الخطأ)**

2- افتراضات النظرية الكلاسيكية:

تقوم نظرية القياس الكلاسيكية على مجموعة من الافتراضات، من أهمها ما حدده كل من: (Hambleton & Zaal, 1991؛ الطريرى، 1997؛ Randall, 1998؛ علام، 2000) المشار إليها في (أبو هاشم، 2006أ، ص.8-9):

1) الدرجة التي يحصل عليها الفرد (الدرجة الملاحظة) هي نتاج نوعين من الدرجات درجة حقيقية ودرجة الخطأ.
2) الدرجة التي يحصل عليها الفرد ليس من الضروري أن تمثل درجته الحقيقية ولذا فالدرجة التي يحصل عليها قابلة للتغيير حسب الظروف الاختبارية.
3) الدرجة الحقيقية للفرد يفترض أن تكون ثابتة، ذلك أنها تمثل قدرة الفرد المقاسة.
4) وجود علاقة عكسية بين درجة الخطأ والدرجة الحقيقية، وهذا يعنى أن انخفاض خطأ القياس يترتب عليه زيادة الدرجة الحقيقية.
5) إن الدرجة الحقيقية يمكن معرفتها من خلال تكرار تطبيق الاختبار عدد كبير من المرات واستنتاج متوسط الدرجات لهذه التكرارات.
6) عدم وجود ارتباط بين الدرجات التي يحققها الأفراد وبين درجات الخطأ.
7) عدم وجود ارتباط بين درجات الخطأ في الاختبارات المختلفة وهذا يرجع إلى الاختلاف في طبيعة الاختبارات.

8) أن درجات الخطأ للأفراد تكون عشوائية وغير مرتبطة ببعضها، وذلك لتطبيقات متوازية للاختبار، ويكون متوسط درجات الخطأ هذه مساوياً للصفر.

9) درجة الخطأ ليست محددة وثابتة في كل المواقف والظروف بل تتغير بتغير هذه المواقف والظروف.

10) درجات الخطأ ليست لها صفة الانتظام، أي أنها لا تتكرر بنفس الصورة وبنفس المستوى في كل الحالات التي يتم بها تطبيق الاختبار.

11) أن الدرجات التي يحصل عليها الفرد في مفردات الاختبار يمكن جمعها كما لو كانت تمثل ميزاناً خطياً، وأن المفردات المتعلقة بالمتغير المراد قياسه تحمل المعنى نفسه لدى جميع المختبرين.

كما حدد كل من (Crocker& Algina, 1986؛ جمحاوي، 2000) الواردة في (زكري، 2010، ص.32-33) مجموعة من الافتراضات:

1) ضرورة تحقق التوزيع الاعتدالي للدرجات على متصل القدرة التي يقيسها الاختبار، فعدد الأفراد الواقع في مستوى معين من القدرة يناظر العدد المتوقع من دالة الكثافة الاحتمالية.

2) أن الدرجة الملاحظة(الخام) هي حاصل جمع الدرجة الحقيقية زائدًا خطأ القياس، حيث أن خطأ القياس يأخذ قيمًا سالبة وموجبة.

3) أن تباين الدرجات الملاحظة يساوي تباين الدرجات الحقيقية مضافًا إليه تباين درجات الخطأ.

4) أن الارتباط بين مجموعتين من الدرجات الخام المستمدتين من اختبارين متوازيين، أو ما يسمى ثبات الاختبارات في مجتمع معين، يساوي تباين الدرجات الحقيقية إلى تباين الدرجات الخام.

5) أن الزيادة أو النقص في طول الاختبار يؤثر في ثبات درجاته.

6) أنه لا يمكن معرفة أو قياس الدرجة الحقيقية، بل يمكن الاستدلال عليها أو تقديرها من خلال الدرجة الملاحظة، وذلك بحساب متوسط الدرجة الملاحظة المستقلة الناتجة من تطبيق الاختبار لعدد كبير من المرات، وتزداد دقة الدرجة الحقيقية بزيادة مرات التطبيق.

7) ليس هناك ارتباط بين الدرجات الحقيقية وأخطاء القياس للمفحوصين على الاختبار، بمعنى أنه ليس ضروريًا أن تكون أخطاء القياس للمفحوصين ذوي الدرجات الحقيقية المرتفعة أصغر من تلك الأخطاء لذوي الدرجات الحقيقية المنخفضة.

8) ليس هناك ارتباط بين أخطاء القياس لاختبارين مختلفين على نفس المفحوص.

9) أن خطأ القياس هو خطأ عشوائي يحدد دقة القياس، أو ما يسمى بثبات الاختبار، بينما الخطأ المنتظم هو خطأ يتعلق بصدق القياس أو الاختبار.

10) تفترض هذه النظرية تساوي تباين أخطاء القياس لجميع المفحوصين، الذين يطبق عليهم الاختبار أو أداة القياس.

3- الخصائص السيكومترية للاختبار Test Psychometric Characteristics:

يرتبط مصطلح الخصائص السيكومترية للاختبار بالموضوعات التي تتناول سمات وخصائص الاختبار الجيد، فبتوفرها يمكن الوثوق في قدرة الاختبار على قياس المتغير الذي بني لقياسه، والحصول من تطبيقه على

قياسات دقيقة نثق في استخدامها. وأهم هذه الخصائص القياسية ما يتعلق بالدرجة الكلية(الملاحظة) للاختبار وتتمثل في الصدق، والثبات. مع العلم أن هناك خصائص قياسية أخرى لا يتسع المجال لذكرها. وسوف نكتفي بتناول الخصائص السيكومترية المذكورة سلفا فيما تبقى من صفحات هذا الموضوع.

أولا: الثبات Reliability:

1- مفهوم الثبات:

يرى علام (2000، ص.131) أن الثبات يقصد به: "مدى خلو درجات الاختبار من الأخطاء غير المنتظمة التي تشوب القياس، أي مدى قياس الاختبار للمقدار الحقيقي للسمة موضع القياس، فدرجات الاختبار تكون ثابتة إذا كان الاختبار يقيس سمة معينة قياسا متسقا في الظروف المتباينة التي تؤدي إلى أخطاء القياس. فالثبات بهذا المعنى يعني الاتساق أو الدقة في القياس".

وترى كل من أنستازي وإربينا (Anstasi & Urbina, 1997, p.84) بأن مفهوم الثبات يعني: "اتساق القياسات التي يتم الحصول عليها من نفس الأفراد عندما يتم إعادة اختبارهم باستعمال نفس الأداة في مناسبات مختلفة وتحت نفس الظروف".

كما أشار أبو حطب وآخرون (2003، ص.101) إلى مفهوم الثبات بأنه: "دقة الاختبار في القياس أو الملاحظة وعدم تناقضه مع نفسه، واتساقه واطراده فيما يزودنا به من معلومات عن سلوك المفحوص".

وذكر أبو هاشم (2006ب، ص.3) أن هناك عدة مفاهيم لمعنى ثبات الاختبار هي كمايلي:

1) أن يعطى الاختبار نفس النتائج تقريباً إذا أعيد تطبيقه على نفس المجموعة من الأفراد. أي أن درجات الاختبار لا تتأثر بتغير العوامل أو الظروف الخارجية، حيث إن إعادة تطبيق الاختبار والحصول على نفس النتائج يعنى دلالة الاختبار على الأداء الفعلي أو الحقيقي للفرد مهما تغيرت الظروف.
2) أن ثبات الاختبار يعنى أيضاً الأداء الحقيقي للفرد والذي يعبر عنه بالدرجة الحقيقية التي يحصل عليها في اختبار ما. والأداء الحقيقي هو جزء من الأداء الكلى الذي يعبر عنه بالدرجة الكلية(الملاحظة) للاختبار، أما الجزء الآخر فهو الأداء الذي يعود إلى أخطاء القياس (الصدفة أو الظروف الخارجية البعيدة عن موضوع الاختبار) ويعبر عنه بدرجة الخطأ. وعلى هذا يمكن أن القول إن تباين الدرجات الملاحظة يساوي مجموع تباين كل من الدرجات الحقيقية وتباين الخطأ. ويمكن التعبير عنها بالمعادلة التالية:

تباين الدرجات الملاحظة = تباين الدرجات الحقيقية + تباين الخطأ

وتعتبر درجات الاختبار ثابتة إذا كان التباين الحقيقي (تباين الدرجات الحقيقية) أكبر ما يمكن بالنسبة للتباين العام (تباين الدرجات الملاحظة) أو أن تباين الخطأ أقل ما يمكن.

كما يمكن التعبير عن مفهوم الثبات إحصائيا من خلال معامل الثبات الذي يساوي النسبة بين تباين الدرجات الحقيقية وتباين الدرجات الملاحظة، وتكون قيمته محصورة بين الصفر والواحد الصحيح، فمعامل الثبات قد يساوي الواحد الصحيح أو قريبا منه إذا كانت أخطاء القياس معدومة (تباين الخطأ يساوي الصفر) أو قليلة في الدرجات الملاحظة، مما يدل على اتساق تام للدرجات، أما عندما تكون

أخطاء القياس كبيرة، فإن الدرجات تكون غير متسقة، وبالتالي فإن معامل الثبات يصبح مساويا للصفر أو قريبا منه. ويمكن توضيح هذه العلاقة بالمعادلة التالية:

معامل الثبات = تباين الدرجات الحقيقية/ تباين الدرجات الملاحظة

3) أن تكون هناك علاقة قانونية بين مفردات الاختبار، ويدل ذلك على التناسق في البناء الداخلي للاختبار، وهذا يعنى أن معامل ثبات الاختبار سوف يتوقف على العلاقة البينية بين كل المفردات من جهة، وبين كل مفردة والاختبار ككل من جهة أخرى. ويتضح من هذا أن تماسك الاختبار أو تناسق بنائه يدل على ثبات درجاته. ومن هذه العلاقة القانونية القائمة بين مفردات الاختبار يمكن حساب معامل الثبات.

2-طرق حساب معامل الثبات:

توجد أكثر من طريقة لحساب معامل الثبات وفق النظرية الكلاسيكية في القياس، والتي تعتمد في حسابها على الدرجات الملاحظة بسب كون الدرجات الحقيقية غير معلومة، وتختلف هذه الطرق عن بعضها باختلاف مصدر أخطاء القياس، والتي تعتمد على طبيعة الاختبار والغرض من استخدام نتائجه. ويمكن عرض هذه الطرق كمايلي:

2-1-طريقة إعادة الاختبار Test Retest Method:

يسمى معامل الثبات بإعادة الاختبار معامل الاستقرار. ويستخدم عندما تكون السمة مستقرة نسبيا، أو عند بناء خطط مستقبلية قائمة على التنبؤ (الغامدي، 2003، ص.18). وتقوم هذه الطريقة جراء تطبيق الاختبار موضع البحث على عينة من الأفراد، ثم إعادة تطبيق الاختبار ذاته على العينة نفسها في وقت لاحق، ويتبع ذلك حساب معامل الارتباط بيرسون بين درجات أفراد العينة على ذلك الاختبار في الفترتين (النبهان، 2004، ص.238). ويتحدد الفاصل الزمنى بين التطبيقين وفق نوع التفسير المطلوب للدرجات، ويكون في حدود أسبوع إلى ستة أشهر، ويتأثر معامل الثبات بعدد من العوامل أهمها مدى استقرار السمة، وعامل الألفة والتذكر، وعامل النسيان والتعلم، والنضج، واختلاف الظروف بين التطبيقين. (أبو هاشم، 2006ب؛ وأبو حطب وآخرون، 1997)

2-2-طريقة الصور المتكافئة Equivalence Forms Method:

يسمى معامل الثبات بالصور المتكافئة معامل التكافؤ. ويستخدم عندما يكون الهدف من الاختبار الاستقراء(الاستنتاج) وفي حالات العلاج النفسي والتقويم (الغامدي، 2003، ص.19). وفي هذه الطريقة يتم إعطاء شكلين متكافئين بالمحتوى والتباينات لاختبار معين لنفس المجموعة وبنفس الوقت (فاصل زمني قصير)، ثم يحسب معامل ارتباط بيرسون بينهما (النبهان، 2004، ص.240).

وتتميز هذه الطريقة بأن مصادر الخطأ المتعلقة بالزمن والتذكر والتعب والنسيان والتعلم أقل ما يمكن مقارنة بطريقة إعادة التطبيق، وتصلح لحساب معامل ثبات اختبارات الذاكرة والاختبارات التحصيلية. ومن عيوب هذه الطريقة صعوبة تصميم اختبارين متكافئين لقياس نفس السمة أو الخاصية. أبو هاشم (2006ب، ص.5)

2-3-طريقة التكافؤ والاستقرار (إعادة الاختبار بصور متكافئة) Equivalence and stability Method:

في هذه الطريقة يتم جمع طريقتي إعادة الاختبار والصور المتكافئة وذلك بتطبيق إحدى الصورتين، وبعد فاصل زمني طويل نسبيا تطبق الصورة الثانية، وعن طريق حساب معامل ارتباط بيرسون للدرجات في الصورتين، فإننا نقدر قيمة لمعامل الثبات تدل على كل من الاستقرار والتكافؤ. (أبو حطب، آمال، وعثمان، 1997، ص.150)

هذه الطريقة تعطي قيم تقديرية لمعامل الثبات أقل من نظيرتيها السابقتين، وذلك لأنها تجمع بين الأخطاء العشوائية للقياس المؤثرة في كل منهما. وقيمة معامل الثبات يعد بمثابة الحد الأدنى. (علام، 2000، ص.151)

2-4-طريقة الاتساق الداخلي Internal Consistency Method:

طريقة الاتساق الداخلي تعتمد على معامل الارتباط بين كل مفردة من الاختبار والاختبار ككل، ولذا فإن هذه الطريقة عادة ما تستخدم لتحديد ثبات الاختبار من ناحية ومدى صلاحية مفرداته من ناحية أخرى (عوض، 1998، ص.57). كما تستخدم في بعض الحالات التي قد يصعب فيها تطبيق الاختبار مرتين، أو يصعب بناء صيغتين متكافئتين، في هذه الحالة يفضل تقدير الثبات باستخدام اختبار واحد وتطبيقه مرة واحدة. (علام، 2000، ص.154)

ويمكن تقسيم طرق الاتساق الداخلي إلى مايلي:

طريقة التجزئة النصفية، طريقة كيودر-ريتشاردسون20 و21، طريقة ألفا كرونباخ.

أ- طريقة التجزئة النصفية Split – Half Method:

تعتمد هذه الطريقة على تقسيم الاختبار إلى نصفين متكافئين، ثم تطبيقه على مجموعة واحدة مرة واحدة. وهناك عدة طرق لتقسيمه، فقد يستخدم النصف الأول من الاختبار في مقابل النصف الثاني، أو تستخدم الأسئلة ذات الأرقام الفردية في مقابل الأسئلة ذات الأرقام الزوجية، وبعد انتهاء تطبيق الاختبار نحصل على مجموعتين من الدرجات تستخدم في حساب معامل الارتباط بيرسون بين المجموعتين، وفي هذه الحالة نحصل على معامل ثبات نصف الاختبار وليس الاختبار ككل، وعليه يتعين علينا تعديل هذا المعامل أو تصحيحه للحصول على معامل ثبات الاختبار ككل. (عبد الرحمن، 2008، ص.180-181)

وفي مايلي توضيح لمعادلات تصحيح الاختبار بطريقة التجزئة النصفية:

- **معادلة سبيرمان – براون Spearman – Brown Formula:**

تفترض هذه المعادلة تساوي تباين كل من نصفي الاختبار (علام، 2000، ص.156)

وصيغة المعادلة كمايلي:

$$r_{11} = \frac{2r_{\frac{11}{22}}}{1 + r_{\frac{11}{22}}}$$

حيث أن:

r_{11} : معامل ثبات الاختبار ككل.

$r_{\frac{11}{22}}$: معامل ثبات نصف الاختبار (معامل الارتباط بين نصفي الاختبار)

ويؤخذ على هذه المعادلة عدم تساوي تباين نصفي الاختبار تماما مما يؤثر على قيمة معامل الثبات الكلي. ولتلافي هذا العيب ظهرت معادلات أخرى تأخذ في الحسبان اختلاف تباين نصفي الاختبار ومن هذه المعادلات مايلي:

- **معادلة رولون Rulon Formula** :

وصيغة المعادلة كمايلي:

$$r_{tt} = 1 - \frac{S_D^2}{S_t^2}$$

حيث أن:

r_{tt}: معامل ثبات الاختبار ككل.

S_t^2: تباين درجات المقياس ككل.

S_D^2: تباين الفرق بين درجات الأفراد في النصف الأول ودرجاتهم في النصف الثاني من الاختبار، ويحسب من المعادلة التالية:

$$S_D^2 = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2$$

D: الفرق بين درجتي الفرد في كل من نصفي الاختبار.

n: عدد الأفراد.

- **معادلة جيتمان Gutman Formula**:

وصيغة المعادلة كمايلي:

$$r_{tt} = 2\left[1 - \frac{(S_a^2 + S_b^2)}{S_t^2}\right]$$

حيث أن:

r_{tt}: معامل ثبات الاختبار ككل.

S_a^2: تباين النصف الأول من الاختبار.

S_b^2: تباين النصف الثاني من الاختبار.

S_t^2: تباين الاختبار ككل.

- **معادلة موزير Mosier Formula**:

تتميز هذه المعادلة عن غيرها بأنها توفر قدرا كبيرا من الجهد المبذول في العمليات الحسابية؛ نظرا لتميزها بسهولة اجراء هذه العمليات؛ لاعتمادها على حساب الارتباط بين أحد النصفين وبين الدرجة الكلية للاختبار. (فرج، 2007، ص.325)

وصيغة المعادلة كمايلي:

$$r_{tt} = \frac{r_{ot}S_t - S_o}{\sqrt{S_t^2 + S_o^2 - 2r_{ot}S_oS_t}}$$

حيث أن:

r_{tt}: معامل ثبات الاختبار ككل.

r_{ot}: معامل الارتباط بين النصف الفردي مثلا والاختبار ككل.

S_t: الانحراف المعياري لدرجات الاختبار ككل.

S_o: الانحراف المعياري لدرجات النصف الفردي.

➢ **معادلة فلانجان Flanagan Formula:**

تعتمد هذه المعادلة على نفس المنطلق المعتمد في معادلة رولون، إلا أن معادلة فلانجان أكثر سهولة؛ إذ تستخدم التباين الخاص بكلا النصفين، حيث يجمع هذا التباين ويقسم على التباين الكلي للاختبار. (فرج، 2007، ص.326)

وصيغة المعادلة كمايلي:

$$r_{tt} = 1 - \frac{S_1^2 + S_2^2}{S_t^2}$$

حيث أن:

r_{tt}: معامل ثبات الاختبار ككل.

S_1^2: تباين النصف الأول من الاختبار.

S_2^2: تباين النصف الثاني من الاختبار.

S_t^2: تباين درجات الاختبار ككل.

➢ **معادلة هورست Horst Formula :**

تستخدم هذه المعادلة في حالة تعذر تقسيم الاختبار إلى قسمين متعادلين من حيث الطول (عدد المفردات)، نظرا لطبيعة تكوين الاختبار أو لطبيعة البحث الذي يقوم به الباحث، وتعتمد على معامل الارتباط بين القسمين ونسبة عدد المفردات لكل قسم إلى كل مفردات الاختبار. (الكناني، وجابر، 1995، ص.164)

وصيغة المعادلة كمايلي:

$$r_{tt} = \frac{r_{12}\left(\sqrt{r_{12}^2 + 4ab(1 - r_{12}^2)} - r_{12}\right)}{2ab(1 - r_{12}^2)}$$

حيث أن:

r_{tt}: معامل ثبات الاختبار ككل.

r_{12}: معامل الارتباط بين درجات الجزء الأول(1) والجزء الثاني(2) للاختبار.

a: نسبة عدد مفردات أحد الجزئين إلى عدد مفردات الاختبار ككل.

b: نسبة عدد مفردات الجزء الآخر إلى عدد مفردات الاختبار ككل.

ب- طريقة كيودر – ريتشارتسون Kuder – Richardson Method :

تسمى طريقة كيودر – ريتشارتسون بمعامل التجانس، وظهرت هذه الطريقة لتلافي الانتقادات التي وجهت لطريقة التجزئة النصفية (الغامدي، 2003، ص.23). وتمثلت في أنه عند تغيير ترتيب الأسئلة أو عند تغيير طريقة التجزئة فإن قيمة معامل الارتباط تتغير وتختلف.

تعتمد هذه الطريقة على تقدير معامل ثبات الاختبار عن طريق حساب الدراجات لكل مفردة من مفردات الاختبار ثم تجمع هذه التباينات. ومعامل ثبات هذه الطريقة هو القيمة التقديرية لمتوسط قيم معاملات الارتباط بين كل نصفي الاختبار لجميع طرق التجزئة الممكنة دون أن نقوم بالتجزئة فعلا. (اللحياني، 2009، ص.64)

وهناك معادلتان في هذه الطريقة هما المعادلة KR-20والمعادلة KR-21.

1) معادلة كيودر – ريتشرتسون20 20 Kuder – Richardson Formula:

هي الصيغة الأكثر استخداما ويشترط في استخدامها أن تقيس جميع مفردات الاختبار سمة واحدة، ألا يعتمد الاختبار السرعة، وأن تكون تصحيح مفردات الاختبار بشكل ثنائية التدرج (صحيحة=1، خاطئة=0).

وصيغة المعادلة كمايلي:

$$r_{tt} = \frac{k}{k-1}\left[1 - \frac{\sum pq}{S_t^2}\right]$$

حيث أن:

r_{tt} : معامل ثبات الاختبار ككل.

k: عدد مفردات الاختبار.

p: نسبة الأفراد الذين أجابوا بشكل صحيح على المفردة (صعوبة الفقرة).

q: نسبة الأفراد الذين أجابوا بشكل خاطئ على المفردة.

pq: تباين الفقرات المصححة بشكل ثنائي.

S_t^2: تباين درجات الاختبار ككل.

2) معادلة كيودر – ريتشرتسون21 21 Kuder – Richardson Formula :

يشترط في استخدامها إضافة إلى الشروط المذكورة سابقا في الصيغةKR-20 تساوي جميع مفردات الاختبار في درجة الصعوبة، أو على الأقل أن يتراوح متوسط درجة صعوبة جميع المفردات 0.50، ولكن نظرا لصعوبة تحقق هذا الشرط في كثير من الاختبارات والمقاييس، لذلك كانت الصيغة(20) أكثر استعمالا. (علام، 2000، ص.164)

وصيغة المعادلة كمايلي:

$$r_{tt} = \frac{k}{k-1}\left[1 - \frac{\bar{x}(k-\bar{x})}{kS_t^2}\right]$$

حيث أن:

r_{tt} : معامل ثبات الاختبار ككل.

k: عدد مفردات الاختبار.

$\bar{x}$: متوسط الدرجات الكلية للاختبار.

S_t^2: تباين درجات الاختبار ككل.

ت- طريقة ألفا لكرونباخ Cronbach Alpha Method:

اشتق العالم كرونباخ (Cronbach, 1951) معادلته من معادلة كيورد ريتشاردسون20 لحساب ثبات الاختبارات ذات البدائل المتعددة وتسمى معامل التجانس؛ حيث يحل مجموع تباينات درجات جميع المفردات محل مجموع حاصل ضرب نسبة الافراد الذين أجابوا إجابة صحيحة على كل مفردة ونسبة الذين لم يجيبوا على المفردة. (أبو حطب، وآخرون، 1997، 119)

وصيغة المعادلة كمايلي:

$$r_{\alpha} = \frac{k}{k-1}\left[1 - \frac{\sum S_i^2}{S_t^2}\right]$$

حيث أن:

r_{tt} : معامل ثبات الاختبار ككل.

k: عدد مفردات الاختبار.

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

S_t^2: تباين درجات الاختبار ككل.

كما يمكننا التذكير بأن هناك طرقا أخرى لحساب معامل الثبات في حالة الاجابة المتعددة أهمها طريقة تحليل التباين وطريقة التحليل العاملي.

3- العوامل المؤثرة في ثبات الاختبار:

هناك عدة عوامل تؤثر في ثبات الاختبار، منها ما يعود إلى أداة القياس نفسها ومنها ما يعود إلى الظروف المحيطة بتطبيق الاختبار، بالإضافة إلى نوعية المفحوصين. (الثبيتي، 1998، ص.121)

وفي مايلي أهم العوامل المؤثرة في ثبات الاختبار.

أ) طول الاختبار:

كلما زاد عدد مفردات الاختبار أي طوله، كلما زادت قيمة معامل ثبات درجاته، ويرجع ذلك إلى زيادة عدد المفردات الذي يسمح بتلاشي الأخطاء العشوائية الموجبة بالأخطاء العشوائية السالبة، مما يجعل الدرجة الملاحظة للفرد تقترب من درجته الحقيقية. (علام، 2000، ص.181)

ب) درجة صعوبة مفردات الاختبار:

إذا كانت مفردات الاختبار غاية في السهولة أو الصعوبة فإن ذلك يؤثر على قيم الثبات، مما لا يسمح لنا باستخدامها من قياس الفروق الفردية، ويجعل توزيع الدرجات منتظما مما يقلل من ثبا الاختبار. (مرجع سابق، ص.181)

ولزيادة ثبات الاختبار يجب أن يبنى الاختبار من مفردات يتراوح مدى صعوبتها بين(0.25 – 0.75)، وافضل المفردات ما كانت مستوى صعوبتها يساوي 0.50.(ملحم، 2005، ص.286)

ج) مدى تجانس أفراد العينة:

إذا كانت مجموعة الأفراد متجانسة في السمة التي يقيسها الاختبار، فإن تباين الدرجات الحقيقية يقل مؤديا إلى انخفاض قيمة معامل الثبات، لأنه يعتمد على مدى الفروق بين أفراد العينة، فكلما قلت الفروق قل تباين الدرجات الحقيقية للمفحوصين وبالتالي قيمة معامل الثبات. (علام، 2000، ص.180)

د) صياغة مفردات الاختبار:

المفردات الغامضة والخادعة والعاطفية والطويلة والموحية بالإجابة تقلل من ثبات الاختبار، بينما المفردات الواضحة والموضوعية والقصيرة تزيد من ثبات الاختبار. (ملحم، 2005، ص.268)

هـ) مستوى قدرات المفحوصين:

ينبغي على الباحث مراعاة انتقاء اختبارات ومقاييس يعتمد تقدير ثبات درجاتها على عينة من الأفراد يكون مستوى قدراتهم في نطاق مستوى المجموعة التي يريد دراستها؛ حتى لا يتأثر معامل الثبات باختلاف مستوى قدرة المفحوصين. (علام، 2000، ص.181)

و) زمن الاختبار:

يزداد معامل الثبات تبعا لزيادة الزمن المستغرق في الإجابة على مفردات الاختبار؛ حتى يصل إلى الحد المناسب للإجابة؛ الذي يصل فيه الثبات إلى نهايته القصوى، ثم يقل الثبات تبعا لذلك كلما زاد الزمن المستغرق عن ذلك الحد. (ملحم، 2005، ص.267)

ز) موضوعية التصحيح:

يشير النبهان(2004) إلى أن المفردات التي تصحح بموضوعية عالية بعيدا عن الذاتية والتحيز تؤدي إلى زيادة معامل الثبات. وترى دروزة (2005، ص.201) أن الاختبارات الموضوعية كالاختبارات من متعدد نادرا ما يتأثر ثبات درجاتها بحكم أو رأي المصحح حين ما يكون تصحيحه على درجة عالية من الموضوعية.

ح) التخمين:

يرى إسماعيل(2009) أن مفهوم التخمين يعني الاستجابة لمفردات الاختبار بشكل عشوائي (الشمراني، 2012، ص.31). ويشير ملحم (2005، ص.268) إلى أن زيادة التخمين تنقص من ثبات أي اختبار، وتعتبر اختبارات الاختيار من متعدد أكثر أنواع الاختبارات تأثرا بالتخمين.

ثانيا: الصدقValidity :

1- مفهوم الصدق:

يرى علام (2000، ص.231) أن "الصدق يعد من الخصائص السيكومترية الأساسية للاختبارات والمقاييس، وأن الصدق يعني مدى فائدة القياسات في اتخاذ قرارات موائمة لغرض معين".

ويرى أبو لبدة (2008، ص.212) أن الصدق أهم خواص الاختبار ويعني ذلك أن يقيس الاختبار بالفعل ما وضع لقياسه.

ويعرف ميسك"Messick" الصدق على أنه:" تقييم شامل يوفر من خلاله الدليل المادي والمبرر المنطقي اللازم لإثبات كفاية وملاءمة ومعنى أي تأويل أو فعل يبنى على درجة الاختبار". (الدوسري، 2000، ص.53)

ويصف كروباخ (Cronbach, 1971) الصدق على أنه: "العملية التي يجمع معد الاختبار أو مستخدمه من خلالها الأدلة التي تدعم الاستدلالات التي استخلصها من درجات الاختبار".(كروكر، والجينا، 2009، ص.291)

ويشير أبو هاشم (2006ب، ص.19) إلى أن هناك عدة مفاهيم أساسية تتعلق بصدق الاختبار بمعنى ألا يكون الاختبار صادقاً إلا إذا توافر ما يلي:

1) أن يكون الاختبار قادراً على قياس ما وضع لقياسه. بمعنى أن يكون الاختبار ذا صلة وثيقة بالقدرة التي يقيسها. فالاختبار الذي وضع من أجل قياس القدرة الرياضية مثلاً يجب أن يكون واضحاً أنه يقيس هذه القدرة، وذلك عن طريق مدى صلته بمكونات القدرة الرياضية وعناصرها.
2) أن يكون الاختبار قادراً على قياس ما وضع لقياسه فقط. بمعنى أن يكون هذا الاختبار قادراً على أن يميز بين القدرة التي يقيسها والقدرات الأخرى التي يحتمل أن تختلط بها أو تتداخل معها. فاختبار في القدرة الرياضية – بجانب قدرته على قياس هذه القدرة – يجب أن يقيسها فقط، بمعنى ألا يتأثر بالقدرة اللغوية مثلاً.
3) أن يكون الاختبار قادراً على التمييز بين طرفي القدرة التي يقيسها. بمعنى أن يميز بين الأداء القوى والأداء المتوسط أو الأداء الضعيف. فإذا كانت درجات الاختبار جميعها تتقارب دل ذلك على صدق ضعيف، لأن الاختبار في حقيقة الأمر لم يقم بالمهمة الأساسية في عملية القياس، وهي عملية إظهار الفروق الفردية بين أفراد العينة. وبالمثل فإن الاختبار الذي لا يميز بصورة واضحة بين طرفي القدرة التي يقيسها ولا يظهر الفروق الفردية، فإنه اختبار ليس بصحيح أو صادق.

2- خصائص الصدق:

1) الصدق صفة نسبية فلا يوجد اختبار عديم الصدق تماما أو تام الصدق.
2) أن الصدق يتعلق بنتائج الاختبار وليس بالاختبار نفسه.
3) يعبر عن الصدق بدرجة وصفية، وقد تستخدم الأرقام لتوضيح الدرجة الوصفية. (ملحم، 2005، ص.270؛ مراد، وسليمان، 2002، ص.355)
4) الصدق صفة نوعية أي خاص باستعمال معين.
5) يتوقف صدق الاختبار على ثباته. (الظاهر، وآخرون، 2002، ص.134)

3- أنواع صدق الاختبارات:

أصدرت اللجنة المشتركة المكونة من الجمعية الأمريكية لعلم النفس APA، والجمعية الأمريكية للبحث التربوي AERA، والمجلس القومي الأمريكي للقياس التربوي NCME، كتيبا متجددا يتضمن معايير الاختبارات التربوية النفسية(1999)، حيث حصرت جوانب الصدق في ثلاث أنواع رئيسية هي: صدق المحتوى، الصدق المرتبط بالمحك، وصدق التكوين الفرضي أو المفهوم. علام (2000، ص.188)

3-1-صدق المحتوى Content Validity:

يسمى صدق المحتوى أحيانا صدق المضمون أو الصدق المنطقي أو صدق عينة الاختبار، ويهتم بمدى تمثيل مجال سلوكي معين ومحدد بشكل دقيق في شكل مجموعة من المفردات بصورة مناسبة. (فرج، 2007، ص.266)

ويدل صدق المحتوى على مدى تمثيل الاختبار للنطاق السلوكي الشامل للسمة المراد الاستدلال عليها، إذ يجب أن يكون المحتوى ممثلا تمثيلا جيدا لنطاق المفردات الذي يتم تحديده مسبقا.

ويستند تقدير صدق المحتوى إلى ثلاث فروض صاغها ليون Lenon كالتالي:

1) ينبغي أن يكون مجال اختبار الأفراد مكون من نطاق سلوكي شامل للمفردات حسب أهميته لهم.
2) يمكن انتقاء عينة من مفردات هذا النطاق السلوكي بطريقة هادفة ومناسبة.
3) يمكن تحديد عينة المفردات وأسلوب المعاينات المستخدم وتعريفها بدقة كافية حتى يتمكن مستخدم الاختبار من الحكم على مدى تمثيل عينة المفردات للنطاق السلوكي الشامل الذي تقيسه. (علام، 2000، ص.188)

ويشير علام (2000، ص.191) إلى أن صدق المحتوى يتطلب أدلة منطقية وليست إحصائية، لذلك فإن أغلب أساليبه تعتمد على الأحكام التقييمية للخبراء(المحكمين)، وتتعلق هذه الأحكام بتقدير مدى التلاؤم بين مفردات الاختبار والمجال السلوكي الذي تمثله هذه المفردات. لذلك يمكن إعداد استمارة بميزان تقدير تشتمل على الأبعاد الرئيسة للنطاق السلوكي، مثل نوع المهارات ومجالاتها، ومحتوى ونوع المفردات، وملاءمتها للمحتوى والمهارة المقصودة، ويقوم المحكم بتقييم كل مفردة، ويفضل الاعتماد على مجموعة من المحكمين.

3-2-الصدق المرتبط بالمحك Criterion-related validity:

يسمى الصدق المرتبط بالمحك بالصدق التجريبي أو المعياري، أو الواقعي أو العملي. ويرى كل عمر وفخرو والسبيعي وتركي (2010، ص.197) أن صدق المحك يشير إلى فاعلية الاختبار في التنبؤ بأداء الفرد في سمة معينة، ولتحقيق هذا الغرض فإن الأداء على الاختبار يراجع في ضوء ارتباطه بالاختبار المحك، بمعنى أن الصدق المرتبط بالمحك يظهر عندما يكون الاختبار فعالا في تقدير أداء الفرد على اختبار خارجي مستقل عن الاختبار والذي يدعى المحك، فالمحك هو اختبار مباشر ومستقل للسمة التي صمم الاختبار للتنبؤ بها أو تشخيصها.

ويمكن التمييز بين نوعين من الصدق المرتبط بالمحك وهما الصدق التنبئي والصدق التلازمي.

أ- الصدق التنبؤي Predictive Validity:

هو الصدق الذي يتعلق بتقدير مدى صلاحية الاختبار في التنبؤ بأداء الأفراد في المستقبل. (علام، 2000، ص.193)

ويمكن تقدير الصدق التنبئي بعدة طرق أهما مايلي:

1) حساب معامل الارتباط بين درجات اختبار تنبؤي طبق على مجموعة من الأفراد، ودرجات اختبار محك خارجي طبق عليهم بعد مرور مدة زمنية طويلة. (بخش، 2007، ص.51)
2) استخدام معادلة الانحدار في التنبؤ بدرجات الاختبار المحك، إذ يمكن استخدام معامل الارتباط بين الاختبار والاختبار المحك، وكذلك قيمة المتوسط والانحراف المعياري لدرجات كل منهما في التوصل إلى معادلة الانحدار، وينبغي مراعاة أن تكون العلاقة خطية بين درجات الاختبار والاختبار المحك، والارتباط بينهما مرتفعا.

ب- الصدق التلازمي Concurrent Validity:

يستخدم للتحقق من صدق الاختبارات التي تحاول وصف الوضع الحالي للمفحوصين، وفيه يتم الحصول على درجات المحك في نفس الوقت تقريبا مع درجات الاختبار. (عمر، وآخرون، 2010، ص.197)

ويمكن تقدير الصدق التنبئي بإيجاد قيمة معامل الارتباط بين درجات الأفراد في الاختبار المراد التحقق من صدقه التلازمي ودرجات الأفراد في الاختبار المحك. (علام، 2000، ص.210)

3-3-صدق التكوين الفرضي (صدق المفهوم) Construct Validity:

يرى عوض (1998، ص.62) "أن صدق المفهوم يعني مدى الارتباط بين الجوانب التي يقيسها الاختبار ومفهوم هذه الجوانب كما وضعها مصمم الاختبار".

كما يقصد بصدق المفهوم (التكوين الفرضي) مدى قياس الاختبار لتكوين فرضي أو مفهوم نفسي، وهذا التكوين سمة غير قابلة للملاحظة، ومن أمثلة هذه التكوينات الفرضية: الذكاء، والطلاقة اللغوية، والعصابية، والقلق، وغيرها. (أبو حطب، وآخرون، 1997)

وقد بين Sax وزيلر Zeller أنه يمكن إيجاز عملية التوصل إلى أدلة تتعلق بصدق التكوين الفرضي في الخطوات التالية على الأقل:

1) تبرير أهمية التكوين الفرضي تربويا ونفسيا على أن يكون معرفا تعريفا إجرائيا ويدل على سمة قابلة للقياس.
2) الاستناد إلى نظرية تربوية أو نفسية أو اقتراح نموذج منطقي يوضح المفاهيم والعلاقات القائمة بينها.
3) التمييز بين التكوين الفرضي والتكوينات الفرضية الأخرى المماثلة له.
4) التوصل إلى أدلة من مصادر متعددة لتأكيد التكوين الفرضي. وعندما تستخدم اختبارات متعددة في هذا الشأن فإنه يمكننا الحصول على الصدق التقاربي للتكوين الفرضي.
5) التوصل إلى أدلة نتأكد منها أن التكوين الفرضي لا يرتبط بعوامل وقتية او دخيلة لكي نحصل على الصدق التمايزي للتكوين الفرضي.

6) إجراء تعديلات مستمرة في التكوين الفرضي بما يتفق والأدلة والمعلومات الجديدة المتجمعة. (علام، 2000، ص.228)

وهناك عدة أساليب للتأكد من صدق التكوين الفرضي:

- **أساليب تعتمد على الارتباطات:**

1) دراسة الفروق بين مجموعتين مختلفتين من الأفراد أو عدة مجموعات في اختبار يقيس سمة معينة، من خلال المقارنة بين درجات الأفراد بطريقة مباشرة، بحيث تكون السمة مرتفعة عند مجموعة ومنخفضة عند مجموعة أخرى، فإذا تبين أن هناك اختلافا بين درجات المجموعتين اعتبر ذلك أحد أدلة صدق التكوين الفرضي للاختبار.
2) إيجاد معامل الارتباط بين درجات اختبار يفترض أنه يقيس تكوينا فرضيا معينا، ودرجات اختبار آخر بينت الأدلة المتعددة أنه يقيس نفس التكوين الفرضي. فإذا كانت قيمة معامل الارتباط مرتفعة يعد ذلك أحد أدلة صدق التكوين الفرضي للاختبار. وقد يتم تناول العلاقات بين الاختبار ومجموعة من الاختبارات التي تقيس نفس السمة، لذلك يستخدم الأسلوب الاحصائي المسمى تحليل التباين، ويهدف هذا الأسلوب لتكوين مصفوفة ارتباط، واجراء تحليل إحصائي لها لتحديد أقل عدد من التكوينات الفرضية التي تفسر التباين في قيم معاملات الارتباط، داخل المصفوفة.

- **أساليب تعتمد على التجريب:**

تركز هذه الأساليب على عملية التجريب لإحداث تغييرات في درجات الأفراد في اختبار معين كوسيلة للتعرف على مدى تأثر الأداء بمعالجات أو متغيرات معينة، مما يساعد في تأكيد بعض التفسيرات المتعلقة بنتائج الاختبار أو رفضها.

- **أساليب تعتمد على التحليل المنطقي:**

تركز هذه الأساليب على الفحص الدقيق للاختبار والأداء الذي يتطلبه، وإحداث تكامل بين نتائج هذا الفحص وبين النظرية التي يرتكز عليها الاختبار وآراء المفحوصين الذين سبق أن اختبروا باختبارات مماثلة، ويترتب على ذلك افتراض تفسيرات ربما تختلف عن التفسيرات التي اقترحها معد الاختبار. فالمحكم الذي لديه خبرة سابقة كبيرة بالأخطاء التي شابت الاختبارات السابقة يمكنه أن يكتشف جوانب القصور في أداة القياس الجديدة. (علام، 2000، ص ص.222-227)

4- العوامل المؤثرة في الصدق:

وذكر كل من أبو لبدة (2008) وملحم(1997) عدة عوامل تؤثر في الصدق أهما مايلي:

1) عوامل متعلقة بالاختبار نفسه: تتمثل في صياغة أسئلة الاختبار كأن تكون غامضة أو مصاغة بلغة فوق مستوى الطلاب، أو سهلة دون مستواهم أو صعبة فوق مستواهم، مما يؤثر على درجات التلاميذ الحقيقة ويقلل من صدق الاختبار.

2) عوامل متعلقة ببيئة الاختبار: كالضوضاء والحرارة الزائدة في حجرة الاختبار، والتهوية والإضاءة غير المناسبة، عدم وضوح طباعة وإخراج الاختبار، وعدم وضوح تعليمات الاختبار، مما يقلل من صدق الاختبار.
3) عوامل متعلقة بالمفحوص: اضطراب المفحوص وقلقه الزائدان، لجوئه إلى الغش والتخمين، كل ذلك يؤثر على مستواه في الاختبار ويقلل من صدق الاختبار.
4) طول الاختبار: يزداد الاختبار بزيادة عدد مفرداته، لأن ذلك يؤدي إلى شمول الاختبار للمحتوى ويقلل من أخطاء القياس.
5) يتأثر صدق الاختبار والمحك المستخدم في تقدير الصدق بمعامل ثباتهما.

ثالثا: العلاقة بين الصدق والثبات:

بالرغم من أن الصدق والثبات يمثلان أهم خصائص الاختبار الجيد إلا أن بينما أوجه اختلاف، حيث يتعلق الثبات بمدى اتساق مفردات هذا الاختبار، بينما يتعلق الصدق بالفائدة والهدف الذي صمم من أجله هذا الاختبار. كما أن الثبات يتأثر بالأخطاء غير المنتظمة، بينما الصدق يتأثر بالأخطاء المنتظمة وغير المنتظمة، ومن جانب آخر فإن الصدق خاصية تتعلق بتفسير درجات الاختبار ولا تتعلق بالاختبار في ذاته.

كما أن هناك أوجه تكامل بينهما، فمفهوم الصدق يرتبط بمفهوم الثبات ارتباطا وثيقا، لأن الاختبار لا يمكن أن يتسم بالصدق في استخدامات معينة دون أن تكون درجاته متسقة، إضافة إلى أن قيمة مؤشر الثبات تعد بمثابة الحد الأعلى لقيمة معامل صدق الاختبار. كما تؤثر قيم معامل الثبات في درجة العلاقة بين الاختبار التنبئي واختبار المحك. كما تسهم قيمة معامل الاستقرار لدرجات اختبار ما في زيادة ثقتنا بصدق التكوين الفرضي للاختبار؛ نظرا لأننا نهتم بالسمات التي تتميز بالثبات النسبي. كما أن معامل الاتساق الداخلي أو معامل التجانس للاختبار يقدم دليلا آخر على صدق التكوين الفرضي. (علام، 2000، ص.230)

خاتمة:

يهتم الباحث في ميدان العلوم السلوكية بالحصول على معلومات وبيانات دقيقة حول الظاهرة السلوكية التي يدرسها، وفى سبيل ذلك يستخدم عدداً من الأدوات كالاختبارات والمقاييس والتي صممت في ظل نظرية القياس الكلاسيكية لقياس الظاهرة المعنية، والتي لا تزال تستخدم إلى يومنا هذا رغم ظهور نظريات جديدة في هذا المجال. والهدف من عملية القياس الحصول على معلومات وبيانات تتصف بالدقة والموثوقية، مما يسمح له باتخاذ قرارات صحيحة وإصدار أحكام سليمة. ولتحقيق هذا الغرض؛ يجب أن يستند استخدام هذه الاختبارات والمقاييس المستخدمة في القياس إلى مجموعة من الخصائص التي يجب أن تشتمل عليها وتتصف بها؛ من أهمها خاصيتي الصدق والثبات. فالصدق يتعلق بمدى أهمية وفائدة القياسات التي يتم الحصول عليها من تطبيق الاختبارات والمقاييس في اتخاذ قرارات مناسبة لغرض معين، بينما يتعلق الثبات بمدى خلو الدرجات التي نحصل عليها من استخدام الاختبارات والمقاييس من أخطاء القياس.

المراجع:

1- أبو حطب، فؤاد، آمال، صادق، وعثمان، سيد، (2003)، التقويم النفسي، مكتبة الأنجلو المصرية، ط3، القاهرة.

2- أبو لبده، سبع محمد، (2008)، مبادئ القياس النفسي والتقييم التربوي، دار الفكر، عمان.

3- أبو هاشم، السيد محمد، (2006ب)، الخصائص السيكومترية لأدوات القياس في البحوث النفسية والتربوية باستخدام SPSS، دراسة مسترجعة بتاريخ 20 نوفمبر 2016 من:
faculty.ksu.edu.sa/70810/.../الخصائص%20السيكومترية%20لأدوات%20القياس.pdf

4- أبو هاشم، السيد محمد، (2006أ، يناير)، دراسة مقارنة بين النظرية التقليدية ونموذج راش في اختيار فقرات قائمة مداخل الدراسة لدى طلاب الجامعة، مجلة كلية التربية بالزقازيق، العدد(52)، ص ص 1-52.

5- بخش، غادة عبد القادر محمد، (2007)، تقويم طرق تقدير الثبات وصدق أدوات القياس المستخدمة في بعض رسائل الماجستير، رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.

6- الثبيتي، علي حامد، (1998)، صدق البناء النظري لأدوات جمع المعلومات في البحوث التربوية والنفسية وانعكاس ذلك على تفسير النتائج، رسالة الخليج، مجلد(19)، عدد(96)، ص ص17-62.

7- دروزة، أفنان نظير، (2005)، الأسئلة التعليمية والتقييم المدرسي، دار الشروق للنشر والتوزيع، عمان.

8- الدوسري، إبراهيم مبارك، (2000)، الإطار المرجعي للتقويم التربوي، مكتب التربية العربي لدول الخليج، ط2، الرياض.

9- زكري، علي بن محمد عبد الله، (2010)، الخصائص السيكومترية لاختبار (أوتيس – لينون) للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية، رسالة دكتوراه غير منشورة، كلية التربية، جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.

10- الشمراني، صالح موسى، (2012)، أثر التفاعل بين نمط الاختبار وجنس المستجيب على ثبات الاختبار وبعض الخصائص السيكومترية للفقرات، رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.

11- الظاهر، زكريا محمد، وآخرون، (2002)، مبادئ القياس والتقويم في التربية، دار الثقافة للنشر والتوزيع، ط3، عمان.

12- عبد الحمن، سعد، (2008)، القياس النفسي النظرية والتطبيق، هبة النيل العربية للنشر والتوزيع، ط5، الجيزة، مصر.

13- علام، صلاح الدين محمود، (2000)، القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة، دار الفكر العربي، القاهرة.

14- العنزي، محمد طالب، (2010)، أثر شكل الفقرة على معالم الفقرة وثبات الاختبار وفقًا لنظرية استجابة الفقرة(IRT)، رسالة ماجستير غير منشورة، عمادة الدراسات العليا، جامعة مؤتة، المملكة العربية السعودية.

15- عمر، محمود أحمد، فخرو، حصة عبد الحمن، السبيعي، تركي، وتركي، آمنة عبد الله، (2010)، القياس النفسي والتربوي، درا المسيرة للنشر والتوزيع، عمان.
16- عوض، عباس محمود، (1998)، القياس النفسي بين النظرية والتطبيق، دار المعرفة الجامعية، الاسكندرية، مصر.
17- الغامدي، سعيد حسن آل عبد الفتاح، (2003)، مدى اختلاف الخصائص السيكومترية لأداة القياس في ضوء تغاير عدد بدائل الاستجابة والمرحلة الدراسية، رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.
18- فرج، صفوت، (2007)، القياس النفسي، الانجلو المصرية، ط6، القاهرة، مصر.
19- كروكر، ليندا، الجينا، جيمس، (2009)، مدخل إلى نظرية القياس التقليدية والمعاصرة، ترجمة زينات يوسف دعنا، دار الفكر، عمان.
20- الكناني، ممدوح، وجابر، عيسى عبد الله، (1995)، القياس والتقويم النفسي والتربوي، مكتبة الفلاح، بيروت.
21- اللحياني، عفاف بنت راضي مشخص، (2009)، أثر بعض طرق تقدير الدرجات المفردات على ثبات وصدق اختبار تحصيلي في الرياضيات ذي الاختيار من متعدد لدى طالبات الصف الأول الثانوي بمكة المكرمة، رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.
22- مراد، صلاح أحمد، سليمان، أمين أحمد، (2002)، الاختبارات والمقاييس في العلوم النفسية والتربوية، دار الكتاب الحديث، القاهرة.
23- ملحم، سامي محمد، (2005)، القياس والتقويم في التربية وعلم النفس، دار المسيرة للنشر والتوزيع، ط3، عمان.
24- الموسوى، نعمان محمد صالح، (2014)، الخصائص السيكومترية لمقياس مهارات التواصل الإلكتروني للمراهقين في ضوء نظرتي القياس التقليدية والحديثة، مجلة الطفولة العربية، المجلد(15)، العدد(59)، ص ص 9-34.
25- النبهان، موسى، (2004)، أساسيات القياس في العلوم السلوكية، دار الشروق للنشر والتوزيع، عمان.
26- Randall, S, (1998), Comparing Measurement Theories, Paper Present at the Annual Meeting of the American Educational Research Association (San Die go, CA, April 13-17).
27- Suen, H. K, (1990), Principles of Test Theories, Lourence Erlbaum Associates, Inc, Hillsdale, New Jersy.
28- Anastasi, A. and Urbina, S. (1997). Psychological Testing. New Jersey 07458: Upper Saddle River.