



الأسس العلمية والطرق الاحصائية للاختبارات والقياس في التربية الرياضية

د. مروان عبد المجيد ابراهيم

بسم الله الرحمن الرحيم

﴿والوزنَ يومَدِ الحقِ فَمَنْ ثَقَلَتْ مُوازِينَهُ فَأُولَئِكَ
هُمُ الْمُفْلِحُونَ * وَمَنْ خَفَتْ مُوازِينَهُ فَأُولَئِكَ الَّذِينَ
خَسَرُوا أَنفُسَهُمْ بِمَا كَانُوا يَأْيَتْنَا يُظْلَمُونَ﴾

سورة الأعراف - الآية 7 - 8

قياس النفس البشرية ، وقياس السلوك البشري
وتقييمه أو الحكم عليه مهمة صعبة وشاقة تتطلب :

- الصدق مع الله .
- الثبات مع النفس .
- الموضوعية مع الآخرين .

المحتويات

9	المقدمة
الباب الأول		
11	الصدق
16	1.1 أساليب حساب وتقدير الصدق
17	1-1.1 صدق المحتوى
21	1-1.2 الصدق العاملية
22	1-1.3 الصدق الظاهري
24	1-1.4 صدق التكوين
28	1-1.5 صدق التعلق بمحك - الصدق التجاري
28	1-1.5.1 الصدق التنبؤي
31	1-1.5.2 الصدق التلازمي
34	1-2 تصحيح معامل الصدق
36	1-3 الطرق الاحصائية لقياس الصدق
36	1-3.1 طريقة معاملات الارتباط
38	1-3.2 طريقة المقارنة الطرفية
43	1-3.3 طريقة الجدول المرتقب
48	1-4 أمثلة لحساب معامل صدق الأختبار إحصائياً
58	1-5 العوامل المؤثرة في الصدق
58	1-5.1 طول الاختبار

60	1-5-2 ثبات الاختبار.....
62	3-5-1 عينة الأفراد.....
63	4-5-1 التباين.....

الباب الثاني

65	الثبات
68	1-2 معامل ثبات الاختبار.....
69	2-2 مفهوم الثبات.....
74	2-3 القياس والثبات.....
75	2-4 معنى الثبات.....
80	2-5 الثبات والدلالة الاحصائية.....
80	2-6 الطرق الاحصائية لقياس الثبات.....
82	2-6-1 طريقة اعادة الاختبار.....
88	2-6-1-1 طريقة بيرسون.....
89	2-6-1-2 طريقة سبيرمان.....
93	2-6-2 طريقة التجزئة النصفية.....
96	2-6-2-1 طريقة الزوجي والفردي.....
96	2-6-2-2 طريقة جزء الاختبار.....
97	2-6-2-3 المعاملات الاحصائية لطريقة التجزئة النصفية.....
97	2-6-2-3-1 معادلة سبيرمان وبراؤن.....
102	2-6-2-3-2 معادلة رولون.....

104	3-3-2 معايير جثمان	6-2
105	4-3-2-6-2 معايير جلكسون	
108	3-6-2 طريقة تحليل التباين	
110	4-6-2 طريقة الاختبارات المتكافئة	
111	5-6-2 طريقة المزج بين طرفي الاختبارات المتكافئة واعادة الاختبار	
111	6-6-2 طريقة اتفاق الملاحظين	
116	7-2 ثبات درجات المصححين	
116	8-2 شروط الحصول على ثبات عالي	
116	1-8-2 عدد فقرات الاختبار	
117	2-8-2 تجانس فقرات الاختبار	
118	3-8-2 صعوبة فقرات الاختبار	
119	2- العوامل المؤثرة في الثبات	
121	10-2 العلاقة بين الصدق والثبات	
123	11-2 أساس تحديد نوعية الفقرات المستخدمة بالاختبار	
124	12-2 إعداد تعليمات الاختبار	

الباب الثالث

127	التحليل الاحصائي لاسئلة الاختبار
129	1-3 تحليل أسئلة الاختبار
130	1-1-3 فوائد تحليل أسئلة الاختبار
131	1-1-3 خطوات تحليل أسئلة الإختبار

131	1-2-1-3 معامل السهولة
132	2-1-3 معامل الصعوبة
133	3-2-1-3 معامل التمييز

الباب الرابع

151	الموضوعية والتقنيات والمعايير
153	1-4 الموضوعية
155	1-1-4 حساب موضوعية الاختبار
155	1-2-4 طريقة معامل ارتباط الرتب لسييرمان
159	1-3-4 طريقة استخدام دلالة معاملات الارتباط
159	1-4-4 الخطوات المتعددة لاستخدام الجدول
162	2-4 التقنيات
163	3-4 المعايير
165	4-3-4 انواع المعايير
165	4-3-1-4 المعيار الصنفي
167	4-3-2-4 المعيار العمري
169	4-3-3-4 الرتب المئوية
175	4-3-4-1 الدرجة القياسية
180	4-3-2-4 المعيار الثاني
180	4-2-3-4 نشأته و معناه
183	4-2-3-2-4 طريقة حساب المعيار الثاني

المقدمة

إن التخطيط المحكم بما يتمشى مع الظروف المحيطة بتنفيذ برنامج الاختبارات يغنينا عن مشاكل كثيرة قد تواجهنا. فالإعداد الجيد والتحضير والإلمام بكل النواحي الفنية ووسائل المعرفة المتعلقة بموضوع الاختبار يضمن لنا درجة كبيرة من دقة وثبات النتائج ومن هنا كان من الضروري فهم الخطوات الفنية المستخدمة لاداء برنامج الاختبارات.

فالاختبار الذي يوضع لقياس خاصية معينة لدى الأبطال مثلاً لا يصلح لقياس هذه الخاصية عند الناشئين أو عند الفتيات وبذلك نرى أن الاختبار الجيد يجب أن يراعي الخصائص السنية والمهارية والجنس لعينة الأفراد.

كما يجب أن يكون للاختبار هدف واضح ومحدد ومرتبط بالأهداف العامة. وبهذا يجب تحديد المشكلة أولاً بحيث تكون واضحة ومن ناحية أخرى يجب مراعاة قلة تكاليف الاختبار وسهولة تنفيذه وأن تكون خطوات تنفيذه محددة بوضوح ووافية للشرح المبسط التي يمكن أن يفهمها الفرد الذي يؤدي الاختبار.

إن نجاح الاختبارات في تحقيق الغرض من إجرائها أو تنفيذها يتطلب الرجوع إلى ما يعرف بنظرية معايرة الاختبارات والتي تشرط أنه عند استخدام أي اختبار في المجال التطبيقي فيجب أن توفر فيه الخصائص الأساسية والعلمية في الاختبار وهي التقنيات والموضوعية والصدق والثبات والمعايير. ويمكن الحصول على هذه الخصائص من خلال تطبيق الاختبار على عينة من الأفراد ثم استخراج النتائج التي نحصل عليها من هذا التطبيق لتحديد تلك الخصائص.

إن الكتاب الحالي يقع في أربعة فصول مرتبة ترتيباً اعتبارياً.
تطرقنا من خلال الباب الأول إلى الصدق وأساليب حساباته والطرق الاحصائية

لقياسه وقد تناولنا أمثلة إحصائية من الواقع الرياضي ثم تناولنا العوامل المؤثرة فيه .
أما الباب الثاني فقد تناولنا فيه الثبات مفهومه ومعناه والطرق الإحصائية لقياسه
وشروط الحصول على الثبات العالي والعوامل المؤثرة فيه ، ثم تناولنا العلاقة بين الصدق
والثبات وتطرقنا إلى تحديد نوعية الفقرات المستخدمة في الاختبار وأخيراً أعداد تعليمات
الاختبار .

ومن خلال الباب الثالث فقد تناولنا التحليل الإحصائي لأسئلة الاختبار وفوائد
تحليل هذه الأسئلة مركزين على خطوات تحليل الأسئلة المتمثلة بعامل السهولة والصعوبة
والتمييز .

هذا وقد ختمنا الكتاب بالباب الرابع الذي ورد الحديث من خلاله عن الموضوعية
وحسابها وطريقة استخدام دلالة معاملات الارتباط والخطوات المتتبعة لاستخدام الجداول
الإحصائية ، ثم تطرقنا إلى التقنيات ومعايير مركزين بذلك على أنواعها . وأخيراً تناولنا
المعيار الثاني وطريقة حسابه .

هذه هي مفردات الكتاب ومنهجيته وتلك هي غايتها . . .
الله أرجو أن يُعين الدارسين والباحثين وأن أكون قد وُفِّقت في تقديم ما يصبوا إليه
طلابنا الأعزاء في وطننا العربي الكبير .

والله من وراء القصد

الدكتور مروان عبدالمحيد إبراهيم

1419هـ - 1998م

الباب الأول

الصدق

- 1-1 أسلوب حساب وتقدير الصدق
 - 1-1-1 صدق المحتوى
 - 2-1-1 الصدق العاملی
 - 3-1-1 الصدق الظاهري
 - 4-1-1 صدق التكويرين
 - 5-1-1 صدق التعلق بمحك - الصدق التجربی
 - 1-5-1-1 الصدق التنبؤی.
 - 2-5-1-1 الصدق التلازمی.
 - 1-2 تصحيح معامل الصدق.
 - 1-3 الطرق الاحصائية لقياس الصدق.
 - 1-3-1 طريقة معاملات الارتباط.
 - 2-3-1 طريقة المقارنة الطرفية.
 - 3-3-1 طريقة الجدول المرتب.
 - 1-4 أمثلة لحساب معامل صدق الاختبار احصائياً.
 - 1-5 العوامل المؤثرة في الصدق.
 - 1-5-1 طول الاختبار.
 - 2-5-1 ثبات الاختبار.
 - 3-5-1 عينة الأفراد.
 - 4-5-1 التباين.

الباب الأول

١- الصدق / Validity

يعد مفهوم الصدق أحد أكثر المفاهيم الأساسية أهمية في مجال الاختبارات والقياس ورغم هذا فهو أكثرها إثارة للجدل وتنوعاً إن لم يكن اختلافاً في التعريف، ورغم أن الدارس المبتدئ لا يمكن بسهولة من ملاحظة هذه الاختلافات العميقه في تعريفات الصدق وما يترتب عليها وهي اختلافات يدور بعضها حول صلاحية المفهوم نفسه بينما يدور البعض الآخر حول فلسفة المفهوم في صلته بأغراض الاختبارات والقياس ويدور البعض الثالث حول اجرائية التعريفات وتطبيقاتها وقد أكدت كل الاختلافات الى إثراء واضح وتعزيز للمفهوم وللأسس وستتناول بعض هذه التعريفات.

١- تعريف جيلكسن / Guliksen

الصدق هو إرتباط الإختبار ببعض المحکمات.

٢- تعريف كيرتون / Kirtton

يحدد كيرتون الصدق باعتباره تقدير للإرتباط بين الدرجات الخام للإختبار والحقيقة «الثابتة ثباتاً تاماً».

٣- تعريف ليندوكست / Lindquist

درجة الصحة التي نقيس بها الاختيار ما نريد قياسه.

٤- تعريف أديجerton / Ederton

المدى الذي تكون به أداة القياس مفيدة لهدف معين.

5-تعريف كرونياخ / Grohbach

يذكر كرونياخ أنه بقدر اكتمال تفسير درجة الإختيار للتنمية المعنية والثقة في هذا التفسير بقدر صدق الاختبار.

6-تعريف كاتل / Cattell

قدرة الاختبار على التنبؤ بعض الوظائف أو أشكال السلوك المحددة والمستقلة عن الاختبار والتي تعد محاكلاً لصدق الدرجة.

يعتبر الصدق من الخصائص المهمة التي يجب الاهتمام بها في بناء الاختبارات فعندما يريد الباحث تصميم اختبار معين فلا بد أن تكون هناك ظاهرة سلوكية معينة يقيسها الاختبار كأن يقيس ظاهرة التكيف الاجتماعي أو الاتكالية أو القلق أو الذكاء أو التحصيل في موضوع دراسي معين ويقوم الباحث بتحويل هذه الظاهرة السلوكية إلى عبارات يتتألف منها الاختبار وعندئذ يسمى هذا الاختبار باختبار الذكاء أو القلق أو التكيف الاجتماعي وعندما يتتأكد الباحث بطريقة علمية أن الاختبار يقيس الظاهرة التي يريد دراستها أو تشخيصها فعندئذ يعتبر الاختبار صادقاً. فالاختبار الصادق إذن هو ذلك الاختبار القادر على قياس السمة أو الظاهرة التي وضع لأجلها، فإذا اختبار الاستعداد الدراسي يعتبر صادقاً إذا كان قادراً على قياس الإستعداد الدراسي ويكون غير صادق إذا كان يقيس ظاهرة سلوكية أخرى ويعتبر الاختبار صادقاً بالنسبة لمجموعة الأفراد الذين وضع الاختبار لأجلهم وقد يكون أقل صدقاً إذا طبقناه على مجموعة أخرى من الأفراد تختلف في خصائصها عن المجموعة الأولى ، فالإختبار التحصيلي في موضوع التاريخ الذي أعد لطلبة الصف السادس الإعدادي قد لا يكون صادقاً بالنسبة لطلبة الصف السادس الابتدائي وذلك لوجود فروق واضحة بين المجموعتين كما إن الاختبار الصادق يكون في معظم الأحيان إختباراً ثابتاً ولكننا لا نضمن أن يكون الاختبار الثابت إختباراً صادقاً في معظم الأحيان فقد يكون أحد الاختبارات ثابتاً بعطينا التائج نفسها إذا أعدنا تطبيقه على

العينة نفسها لكنه قد لا يكون صادقاً في قياس الظاهرة السلوكية المقصودة.

وفي المجال الرياضي قد لا يكون إختبار الرشاقة لطلاب كلية الاقتصاد صادقاً بينما يكون صادقاً عند إجرائه على طلاب كلية التربية البدنية وذلك لوجود فروق واضحة بين طلاب الكليتين .

عند اختبار عنصر السرعة كإحدى الصفات البدنية الأساسية وذلك بالنسبة لعدائي المسافات القصيرة يجب علينا أولاً تحديد شكل السرعة المراد قياسها حتى يكون القياس صادقاً وخصوصاً إذا ما اختر عدائين على مستوى متقدم لمعرفة ما ينقصهم حتى يركزون عليه فمطاولة السرعة والسرعة القصوى وتعجيل السرعة وكذلك سرعة رد الفعل من أشكال السرعة المعروفة والتي يجب أن تراعي مثل هذه الاشكال المختلفة وما يتطلب منها وذلك عند تطبيق الاختبار أو القياس على مجموعة من العدائين اذا ما تطلب ذلك مثلاً وهذا نطرح التساؤل بالنسبة لقياس مستوى السرعة عند عدائين متقدمين .

- هل يعتبر الزمن المسجل في عدو مائة متر مؤشراً حقيقياً لقياس السرعة للعدائين ؟

وللرد على مثل هذا التساؤل يمكن الإجابة بنعم ولكن فلتتساءل مرة أخرى ما هو شكل السرعة المطلوب قياسها؟

وللإجابة على هذا التساؤل نوضح أنه عند عدو مائة متر هناك أربع مراحل يمر بها العداء :

المرحلة الأولى : وتحكم فيها سرعة رد الفعل عند عملية الإنطلاق من أجهزة البدء .

المرحلة الثانية : مرحلة تزايد السرعة وهي في حدود 40 - 50 الأولى ويمكن زيتها إلى 60م بالنسبة للعدائين ذو المستوى العالي .

المرحلة الثالثة: مرحلة السرعة القصوى أو المطلقة وهي في حدود 30 - 40 م.

أما المرحلة الرابعة: مرحلة تحمل السرعة (مطاولة السرعة) وهي آخر مرحلة من مراحل سباق 100م وبذلك فعند قياس سرعة عداء متقدم يجب توضيح شكل السرعة المراد قياسها فإذا كان المقصود من القياس تحديد السرعة القصوى مثلاً فيمكن بذلك إجراء القياس من البدء الطائر Flying Start ولمسافة تتناسب مع مستوى اللاعب وحتى لا يتاثر القياس بأشكال السرعة الأخرى . . . مثل سرعة «رد الفعل» أو تعجيل السرعة أو «تحمل السرعة» وبذلك يكون الاختبار صادقاً لما يهدفه من قياس، أما عند قياس صفة القوة العظمى لمجموعة من المختبرين يجب أن لا تقع جماعة المختبرين تحت تأثير عناصر أخرى كعنصري التحمل (المطاولة) أو السرعة لارتباط هذين العنصرين بصفة القوة العضلية ارتباطاً كبيراً وعلى ذلك تقادس القوة العظمى بواسطة اجهزة الديناموميتر Dynamometer والخاصة بذلك ، أما قياسات القوة العضلية في اتجاهات السرعة كالقوة المميزة بالسرعة وكذلك في اتجاه «تحمل القوة» أو مطاولة القوة فلهمما الاختبارات الصادقة والخاصة بهما وعند إختبار بعض المهارات الحركية للألعاب المختلفة كاختبار سرعة الكرات المرتدة Re-bound في كرة السلة يجب أن يوضع الاختبار الصادق والذي يتم فعلاً عن تقويم هذه المهارة بواسطة اللاعب عند إمكانية حصوله على الكرات المرتدة وبذلك يعني صدق الاختبار «إن تكون مهمة الاختبار قياس وتقويم الصفة التي وضع من أجلها فعلاً».

1-1 أساليب حساب وتقدير الصدق

إن الصدق يعني أن الاختيار يقيس ما أعد قياسه وباعتبار أن مشكلة الصدق تقتصر على أدوات القياس في مجال العلوم التربوية والسلوكية التي يتوجه اهتمامها لتوفير الدقة والتقرير المثالى في تقدير الظواهر وأن الصدق أمر مطلوب بل يختلف من اختبار لأخر بحيث لا نستطيع أن نقول أن الاختبار إما صادقاً أو غير صادق بل نقول انه صادق بقدر أو بدرجة ما يصبح من المقبول أن نستخدم تعريف ثورندايك وهاجان وهو أن الصدق تقدير

لمعرفة ما إذا كان الاختبار يقيس ما نريد أن نقيسه به وكل ما نريد أن نقيسه به ولا شيء غير ما نريد أن نقيسه به ألا .

وتتعدد أساليب حساب وتقدير الصدق في ضوء هذا التعريف العام فنحصل في بعض الحالات على معامل كمي للصدق بالتغييرات الرقمية أو نحصل في حالات أخرى على تقدير كيفي للصدق بالتغييرات اللغوية ، وبما أن علامات الاختبار يمكن أن يحقق أكثر من فائدة إذن يكون هناك أكثر من شكل من أشكال الصدق .

وفيما يلي نعرض لأكثر من نوع الصدق إنتشاراً واستخداماً في مجال إختبارات التربية البدنية وهي :

1-1-1 صدق المحتوى Content Validity

يطلق عليه صدق المضمن أحياناً أو الصدق المنطقي أو الصدق بحكم التعريف أو صدق عينة الاختبار أو صدق التمثيل ، وتقدير صدق مضمون الاختبار باجراء فحص منظم لمجموع العمليات والفترات والنبهات التي يتضمنها الاختبار لتقدير مدى تمثيلها للمجال الذي اعد الاختبار لقياسه ولهذا فإن أفضل المجالات مناسبة لقبول صدق المحتوى هي مجالات الاختبارات التحصيلية فعندما ترغب في تقدير صدق اختبار لفهم القراءة أو للحساب أو التاريخ فعلينا أن نسأل أنفسنا باعتبارنا خبراء ما مدى حسن تمثيل مادة هذا الاختبار لما نعتبره هاماً من المعلومات في المجال وما نتفق فيه مع أداء أفضل المحكمين أو الخبراء . وفي صدق المحتوى ليس المطلوب أن نعرف إذا ما كانت مكونات الاختبار تعكس السمة المعينة فقط ولكن المطلوب أن نعرف ما إذا كانت هذه المكونات ممثلة لهذا المضمن وعناصره الأساسية أم لا؟ لهذا لا يكفي مجرد فحص فقرات الإختبار للتعرف على ما إذا كانت تؤدي المطلوب أم لا ولا يكفي فقط اعطاء تعريف جيد للمجال الذي يقيسه الاختبار بل لتعيين تحليل المجال الكلي ذاته إلى عدد من المجالات أو الفئات الفرعية التي تمثل كل الجوانب الأساسية في المجال وفحص ما إذا كان هناك عدد مناسب من الفقرات يقيس كل

مجال أو فئة فرعية أم لا .

وكمبدأ عام يجب أن يقتصر اهتمام الفاحص الذي يرغب في الحصول على تقدير الصدق محتوى إختبار معين على مادة هذا الاختبار دون اعتبار على الإطلاق لنوعية أو فئات الأفراد الذين يختبرون . فإذا إتفق المضمون مع تعريف المجال وتصنيف فئاته الرئيسية يصبح الاختبار صادق المضمون بالنسبة للأشخاص مما اختلفت نوعياتهم ويجب دائماً على سؤال صدق المحتويات على أساس ذاتية أو تقدير مهني وتقدم درجة الصدق بتغيرات لفظية مثل صدق (مرتفع) أو (متوسط) أو (منخفض) .

يهدف هذا النوع من الصدق إلى معرفة مدى تمثيل الاختبار للظاهرة السلوكية أو الموضوع الدراسي الذي يهدف إلى قياسه ويعتبر هذا النوع من الصدق أهم الأنواع المستخدمة في الاختبارات التحصيلية فالعلم على سبيل المثال قد يرغب بالتعرف على مدى قدرة التلميذ في القراءة أو تحديد ما يعرفه في الموضوعات الاجتماعية وفي مثل هذه الحالة عليه أن يأخذ الاختبار منها وعندما لا تكون العينة ممثلة لكل المحتوى فإن الاختبار يعتبر غير صادق بالنسبة لذلك الموضوع الدراسي ، والخطوة الأولى في الوصول إلى صدق المحتوى هو القيام بتحليل الموضوع الذي نريد قياسه تحليلًا منتظماً لتعديده مجالاته وتقدير الوزن لكل مجال بناءً على أهميته بالنسبة لبقية المجالات الأخرى ثم إختبار عينة ممثلة لكل مجال من هذه المجالات ، فعندما نقوم بوضع اختبار لموضوع الفيزياء مثلاً فإنه يجب أن نختار عينة من كل مجال من مجالاته التي هي على سبيل المثال : الكهربائية والمغناطيسية والميكانيك والضوء وأن تكون العينة مناسبة مع أهمية المجال بالنسبة للموضوع كما يجب أن تكون هذه العبارات ممثلة للأهداف السلوكية في موضوع الفيزياء والاستيعاب وللتطبيق مثلاً فيجب أن تمثل العبارات - فقرات القياس هذه الأهداف بناءً على الأهمية النسبية لكل هدف ويتطلب في غالب الأحيان من مجموعة من الخبراء في المناهج وتأليف الكتب ومن الذين لديهم خبرة تدريسية في الموضوع أن يقيموا كل عبارة من عبارات الاختبار في ضوء مدى تحقيقها ومدى قياسها للمحتوى الموضوعي الاهداف السلوكية التي يسعى إلى

تحقيقها، وبناء على تقييم الخبراء تحدّف الفقرات غير الصادقة وهي الفقرات التي لا تمثل ما يريد الاختبار أن يقيسه وهذا النوع من أنواع الصدق مهم في الاختبارات التحصيلية على وجه الخصوص وفيه يعني بالمستوى الجيد الذي يقيس عنده الاختبار محتوى المادة الدراسية وأهداف التعلم التي صمم الاختبار لقياس مدى تحقيقها عند المفحوصين ويكتنّ توافر صدق عال عند إعداد الإختبار عن طريق:

أ- تحليل محتوى المادة إلى عناصرها.

ب- تحليل أهداف التعلم وتحديدها بصورة سلوكية.

ج- وضع جدول مواصفات يصف عينة الفقرات التي ستستخدم.

د- بناء إختبار يلائم إلى حد كبير جدول المواصفات المذكورة. وهذه الخطوات التي وصفت بشكل مفصل في موضع سابق من هذا، ويتبع واضعوا الاختبارات المقتنة هذه الخطوات نفسها في بناء الإختبارات التحصيلية ولكن المحتوى وأهداف التعلم السلوكية المضمنة في جدول المواصفات ذات قاعدة أعرض حيث تعتمد على الكتب المدرسية وعلى نصائح الخبراء المتعددين في هذا المجال وعليه فإن اختبار التحصيل المقتن يُكن أن يكون ذي صدق عال من نوع صدق المحتوى. ذلك أنه يمثل المحتوى العام والأهداف في مجال ما. ومع ذلك فهو ذو صدق محتوى منخفض فيما يتعلق بالموقف المدرسي الخصوصي وللمعرفة ما إذا كان اختبار تحصيلي مقن صادقاً لدى استعماله في صف معين لا بد من تقويم الفقرات في ضوء المحتوى والأهداف التي يركز عليها في التعليم وعلى وجه العموم فإن المفتاح في صدق المحتوى هو في اختبار عينة الأسئلة فالاختبار دائمًا هو عينة من الأسئلة الكثيرة التي يمكن أن تسأل، وصدق المحتوى هو قضية تحديد ما إذا كانت العينة مناسبة للمجموع الذي يفترض أن تمثله.

يعتمد صدق محتوى الاختبار» وبصورة أساسية على مدى إمكانية تمثيل الاختبار لمعنيات عناصره وكذلك المواقف والجوانب التي يقيسها تمثيلاً صادقاً ومتجانساً ومعنىوية

عالية لتحقيق الهدف الذي وضع من أجله الاختبار وللوصول إلى ذلك يجب مراعاة ما يلى:

- معرفة المكونات التي يشملها الاختبار.
 - تحديد نسبة كل مكون وتجانس هذه النسب في الإطار العام للإختبار.
 - دراسة كل مكون حتى يثبت صدق أهميته لهدف الاختبار.
 - التأكيد من أن مكونات الاختبار (محتوياته) مجتمعة تمثل الهدف الذي من أجله وضع الإختبار.

عندما أرادت اليزابيث جلوفر Elizabeth Glover وضع بطارية اختبار لقياس اللياقة البدنية لأطفال المرحلة الابتدائية للبنين والبنات من عمر 6-9 سنوات فقد صممت بطارية اختبارات لقياس هذا العرض حيث شملت أربع اختبارات:

- 1- الوثب العريض من الثبات / والغرض منه قياس القوة المميزة بالسرعة للعضلات المادة للساقين.
 - 2- العدو المكوكي / 4 قدم (15 مرة - قياس بالثانية). الغرض منه قياس تحمل السرعة.
 - 3- سباق عجل البحر / (20 قدم - القياس بالثانية) المشي على اليدين مع سحب الساقين على الأرض - الغرض منه قياس القوة المميزة بالسرعة لعضلات الذراعين والمنكبين.
 - 4- الجلوس من رقود القرفصاء/ الغرض منه قياس تحمل القوة (مطاولة القوة»العضلات البطن.

وعلى ذلك نطرح التساؤلات التالية:

- هل هذه المحتويات الأربع صادقة لاختبار محتوى ومضمون اللياقة البدنية عند هؤلاء الأطفال.

- وهل محتويات الإختبارات متتجانسة وصادقة.

وعند الإجابة على هذه التساؤلات بالإيجاب يمكن القول بأن المحتويات التي شملتها بطارية الاختبار «صادقة المحتوى» وبذلك أصبح هذا الإختبار صادقاً.

1-1-2 الصدق العاملـي Factorial Validity

يُعد الصدق العاملـي شكلاً متطروراً ومعقداً من أشكال الصدق ففي هذا الأسلوب تستخدم التحليل العاملـي للحصول على تقدير كمي لصدق الاختبار في شكل معامل إحصائي هو تشبع الاختبار على العامل الذي يقيس المجال المعين ، فنحن نبدأ من مصفوفة ارتباطية بين عدد من الاختبارات التي تقيس مجالاً متتجانساً طبقت على عينة متتجانسة من الأفراد ونخرج من تحليل هذه المصفوفة عاملـياً بعده من الفئات التصنيفية المختصرة هي العوامل التي تعبر على التباين المشترك بين هذه التغيرات .

- إن استخدام التحليل العاملـي كأسلوب لحساب الصدق إعتماداً على دالة تشبع الاختبارات على العامل يتضمن اعتبار بعض اختبارات البطارية بمثابة محركات مستقلة عن الاختبار موضوع الدراسة وبحيث تنتهي من التحليل إلى تقدير حجم التباين الحقيقـي المشترك بين اختبارنا (موضوع دراسة الصدق) وبين مجموعة الاختبارات المحكـية وهو ما جعل من الصدق العاملـي مزيجاً من صدق التكوين وصدق التعلق بمحكـ خارجي على أن يؤخذـ في الاعتـار دائمـاً إن مثل هذا الفهم للصدق العاملـي يتطلبـ أن تكون الاختبارات المحكـية إختبارـ صادقة بمعاييرـ أخرى أو من خلالـ محركـات أخرى وفي دراسـ سابـقةـ فالصدق العاملـي من خلالـه نتعرفـ على مدى الانسجامـ والاتساقـ الداخـلينـ من فقرـاتـ الاختـارـ بعضـها بعـضاًـ منـ جهةـ وبيـنـ كلـ فـقـرةـ منـ فـقـراتـ هـذاـ الاختـارـ منـ جهةـ آخرـ وـمنـ هـنـاـ نـحـكمـ عـلـىـ أيـ فـقـرةـ آنـهـاـ كانـتـ صـادـقةـ فـيـ قـيـاسـ ماـ أـرـدـنـاـ مـنـ فـقـراتـ الآخـرـ أـنـ نقـيـسـهـ وـكـذـلـكـ ماـ أـرـدـنـاـ مـنـ مـجـمـوـعـ إـلـيـختـارـ أـنـ يـقـيـسـهـ.

نستطيعـ أنـ نـسـتـدـلـ عـلـىـ الصـدـقـ العـاـمـلـيـ لـمـجـمـوـعـةـ مـنـ فـقـراتـ إـذـاـ كـانـتـ هـذـهـ

الفقرات ذات درجات تمييز جيدة ونستدل على الصدق العاملية للجسم الكلي للاختبار إذا كان معامل الارتباط بين الفقرات وبينه مرتفعاً.

1-3-1 الصدق الظاهري Apperataince of validity

يستخدم تعبير الصدق الظاهري للإشارة إلى مدى ما يbedo أن الإختبار يقيسه أي أن الاختبار يتضمن فقرات يbedo أنها على صلة بالمتغير الذي تفاص وأن مضمون الاختبار متافق مع الغرض منه ومن المرغوب فيه . بصفة عامة أن يكون الاختبار ذي صدق مظهي إذ يلعب الصدق المظهي دوراً واضحاً في تنمية تعاون المفحوص وتوجيه انتباذه إلى نوع الإجابة المطلوبة منه وعليينا أن نلاحظ أن الصدق الظاهري لا يعد وحده محكماً لصدق قياس الاختبار المراد قياسه ويتعين أن لا يختلط الأمر بينه وبين صدق المحتوى فرغم أن مضمون الإختبار هو الأساس الذي نقدر وفقاً به الصدق الظاهري إلا أن هذا التعبير الأخير لا يعد محكماً لصدق المضمون ولا لحسن تمثيل عينة الفقرات للمجال ولا يعده ، ويحدد وزير (Weizer) أربعة معانٍ مختلفة للصدق المظهي منها الصدق بحكم المفهوم . فاختبار معين يbedo أنه يقيس جانباً أو سمة معينة وتحصيل هدف معين ما دامت المفاهيم الأساسية التي يقوم عليها تدعى أنها تفعل ذلك ، والصدق بحكم التعريف فإذا كان المرغوب فيه على سبيل المثال قياس اتجاهات الطلاب نحو بعض المعلمين فإن تصميم مقياس رتبى لهذا الغرض يمكن أن يعتبر صادقاً بحكم التعريف والمعنى الثالث هو صدق المظهر إذ من المرغوب فيه بصفة عامة أن تبدو الاختبارات مقبولة لدى المفحوصين باعتبارها مقياساً صادقاً لما ندعى أنها تقيسه وأن تبدو أمينة في هذا الإدعاء . . .

المعنى الأخير هو الصدق بحكم الفروض . فقد يقدم مصمم الاختبار الدلائل الأولية للصدق في فرضه وإن كان مجرد وضع هذه الفروض لا يعني أن الاختبار يقيسها بالفعل أو لا بد من توفر دلائل أخرى مستقلة للصدق لذا يتبعن حساب الصدق بوسيلة أخرى مقبولة . ويلاحظ أن المعنى الثالث فقط من بين المعاني الأربع التي يذكرها مويسز هو

الذي يعرف الآن اصطلاحاً بالصدق المظاهري وعلينا أن نضع في اعتبارنا في نهاية الأمر عدداً من الحقائق التي تتعلق بالصدق المظاهري.

الحقيقة الأولى: هي أن الصدق الظاهري لا يكفي لفرد تقدير صدق الاختبار فهو سمة ضرورية في التعامل مع المفحوص ولكنه لا يصلح للفصل في قضية صدق الاختبار التي يقصد بها أساساً تقدير مدى تعبير الاختبار عن مفهوم معين سواء وفق محك خارجي أو من خلال أسلوب آخر لتقدير الصدق.

الحقيقة الثانية: هي أن هذه السمة الضرورية في التعامل مع المفحوص جوهرية في مجال الاختبارات والمقاييس المعرفية فقط فالصدق المظاهري للإختبار يعني إيحاء للمفحوص أنه يقيس المراد قياسه بشكل مناسب لعمره وكفاءته وجدية الموقف بما يؤدي إلى تعاون المفحوص في موقف الاختبار واستمراره وصدقه.

الحقيقة الثالثة: أنه بقدر أهمية الصدق المظاهري في الاختبارات والمقاييس المعرفية بقدر ضرورة تجنبه تماماً أو خفظه لأدنى درجة في مجال مقاييس الشخصية بحيث تتعذر اخفاء الهدف من الاختبار واحفاء الدلالات التربوية للفقرات من آية محاولة قد تحدث لتزييف الإجابة تحت آية دافع. والصدق الظاهري أقل أنواع الصدق أهمية لأنه يحاول أن يتعرف على مدى قياس الاختبار للغرض الذي وضع من أجله ظاهرياً وقد يقوم واضع الاختبار نفسه براجعته من حيث نوع العبارات وأسلوبها ووضوحها ويتناول أيضاً تعليمات الاختبار ومدى دقتها في توضيح كيفية الاجابة على عبارات الاختبار.

يعتبر الاختبار صادقاً ظاهرياً إذا كان صالحًا في ظاهرة وبصورة مبدئية من خلال النظر إلى عنوانه وتعليماته والوظيفة التي يقيسها وتمثل الفقرات مبدئياً للأهداف المقيدة مما

يوجي بأن الاختبار يبدو من حيث ظاهره انه مناسب إلى حد ما للغرض المطلوب ، وعلى الرغم من أن هذا الصدق لا يكفي وحده لاعتبار الاختبار صادقاً إلا أنه لا غنى عنه لأنه يشجع المعنيين على اخذ الاختبار بعين الجد والاعتبار، اما إذا ظهر انه غير صالح فلن يأخذه المراقبون للامتحانات والمفحوصون بالاهتمام الزائد ولن يشعروا بفائدة . قد يبدو أن هناك التباس بين مفهومي صدق المحتوى والصدق الظاهري في بعض الحالات حيث أن الصدق الظاهري يشير إلى إمكانية تحقيق عناصر الاختبار للهدف الذي وضع من أجله وينفس الظروف فعند قياس مهارة لاعب كرة القدم في التصويب على الهدف مثلاً فيجب في مثل هذه الحالة أن توضع عناصر الاختبار ليس فقط التي تهدف إلى تحقيق هدف محتوى الاختبار فقط بل يتعدى ذلك إلى وقوع المختبر تحت ظروف المباريات وحتى يكون الاختبار صادقاً صادقاً ظاهرياً وهنا تلعب النواحي النفسية على المختبرين دوراً كبيراً وخصوصاً على نتائج إختباراتهم في مجال القياس أو عند تأديتهم المباريات أو البطولات وعلى ذلك يُنصح عند تطبيق أي اختبار مهاري على أحد اللاعبين يجب أن يؤدى الاختبار في فترة المنافسات الرئيسية وتحت نفس ظروف المباريات ، ويفضل اجراء اختبار اللاعبين أثناء تأديتهم المباريات وحتى يكونوا بعيدين عن أي مؤثرات خارجية أخرى ما عدا مؤثرات وظروف المباريات نفسها وحيثنة يظهر الصدق الظاهري للاختبار .

4-1-1 صدق التكوين Construct validity

يكاد يكون صدق التكوين أكثر أنواع الصدق قبولاً من وجهة نظر فلسفية ويرى عدد كبير من المختصين أنه يتفق أكثر من غيره مع جوهر مفهوم أبيل للصدق من أن الصدق هو (تشيع الاختبار بالمعنى) ويطلب فحص صدق التكوين مزيجاً من التناول المنطقى والجريء وأساساً تقوم دراسات صدق التكوين لفحص النظرية القائمة خلف الاختبار وينطوي هذا الاجراء على ثلاثة خطوات :

أولاً: يبحث الفاحص في ضوء النظرية عن أي الفروض يمكن وضعها وتقوم

بتفسير الدرجة المرتفعة والدرجة المنخفضة عن الاختبار.

ثانياً: يقوم بجمع بيانات واقعية لاختبار هذه الفروض.

ثالثاً: يقوم في ضوء ما تم جمعه من دلائل بتقديم استدلالات معينة عما إذا كانت النظرية مناسبة لتفسير البيانات التي أمكن جمعها أم لا.

ويقدم كرونباخ وميهل ثلاثة أنواع من الدلائل المتاحة في مجال صدق التكوين

«صدق البناء».

أ- الفروق بين الجماعات:

لأن التكوين سمة أولية مجردة مفترضة بين الناس وتنعكس على أدائهم للاختبارات المختلفة ولأن هذه السمات لا توجد بقدر متساوٍ لدى كل الناس فمن المفترض أن الأفراد يختلفون في ما لديهم من هذه السمات يختلفون بوصفهم أعضاء في جماعات كما يختلفون بوصفهم أفراد أو وبالتالي يمكننا أن نتوقع فروقاً بين الذكور والإناث في قدرات معينة ونتوقع فروقاً بين الجماعات العمرية في أداء أفرادها على بعض الإختبارات الصادقة للقدرات.

وإذا تمكن الإختبار من إبراز هذه الفروق والتوقعات النظرية بناءً على خصائص التكوين المبدئي الذي تقدمنا منه فيكون قد يديننا للإختبار أنه صادق.

ب- «التغيير في الأداء»:

يختلف هذا النوع من الفحص للصدق عن النوع السابق إختلاف الدراسات التبعية عن الدراسات العرضية فهنا يقوم الباحث بدراسة الفروق في الأداء الخاص بالعينة نفسها من الأفراد على مدى فترات زمنية مختلفة ولا يعني ذلك ضرورة ظهور تغير في الأداء بين القياس الأول والقياسات التالية فالامر يعتمد أولاً وأخراً على السمة المقيدة التي تقف خلف الأداء فإذا كان قد افترض مسبقاً إن هذه السمة لا تتغير مع التقدم في العمر فعلينا ألا

نتوقع تغيراً في الأداء على الإختبار الصادق ووجود تفسير في الدرجة على مقياس للإحباط بعد موقف تجرببي تعرض فيه المفحوصون للخجل يؤيد صدق الإختبار ويلاحظ بصفة عامة أن النتائج السلبية سواءً في هذا النوع أو النوع السابق قد تكون أكثر قيمة من النتائج الإيجابية فالنتائج السلبية تؤدي إلى شك واضح في صدق قياس الإختبار للسمة.

جـ- الإرتباط:

لا يعد الإرتباط بين الإختبار وإختبار آخر دليلاً في حد ذاته على الصدق وسواءً أكان هذا الاختبار الآخر محكماً مقبولاً أو مقيساً للسمة نفسها فإننا نتوقع أن نحصل على معامل مرتفع للإرتباط بينهما وما من شك في أن الدراسة العاملية التي تصمم للتعرف على التباين المشترك والتقويم الذي يقيسه التباين بين ارتباطات عدد من الاختبارات أفضل كثيراً من الارتباط بين إختبارين بوصفه مؤشراً لثبات التقويم وعلى هذا فعلينا في حالة الإختبارات الصادقة أن نتوقع إرتباطاً أكبر بين إختبارين يقيسان السمة نفسها بأسلوب مختلف عن الارتباط بين الاختبار واختبار آخر لا يقيس السمة نفسها ولكنه يقيس سمة أخرى بنفس أسلوب قياس السمة موضوع الدراسة.

يقصد بصدق البناء مدى قياس الاختبار لسمة أو ظاهرة سلوكية معينة فالباحث في هذا النوع من الصدق يحاول معرفة طبيعة الظاهرة السلوكية التي يسعى الاختبار إلى قياسها فإذا كان الاختبار إختباراً للذكاء أو الأمانة فإن الباحث يحاول معرفة طبيعة العوامل المكونة له فهو يهتم إذن بالظاهرة التي يقيسها الاختبار ولا يهتم بقدرته على التنبؤ بالسلوك المستقبلي (صدق تنبؤي) للفرد أو السلوك الحالي (صدق تلازمي) له وهناك طرق عديدة يمكن أن نلجأ إليها للوصول إلى هذا النوع من الصدق.

1- معاملات الإرتباط مع اختبارات أخرى

ويتم ذلك عن طريق حساب معامل الإرتباط بين الإختبار الذي نريد الوصول إلى صدق البناء فيه وبين الاختبارات التي تتشابه معه في الوظائف التي يقيسها والتي سبق

التأكد من صدقها فإذا افترضنا أن باحثاً أراد بناء مقياس للطموح بين الطلبة فإنه لكي يتوصل إلى صدقة يقوم بتطبيق اختبار آخر مستقل عنه يقيس «المثابرة» والتي تعتبر صفة من صفات الفرد الطموح وعندما يحصل الباحث على معامل إرتباط عال بين الاختبارين يتوصل إلى أن اختباره قد تحقق فيه صدق البناء.

2- العلاقة بين درجات الفقرات والإختبار

تفترض هذه الطريقة أن الدرجة الكلية للفرد تعتبر معياراً لصدق الإختبار ويحاول الباحث عادة إيجاد العلاقة الإرتباطية بين درجة كل فقرة من فقرات الإختبار والدرجة الكلية وتحذف الفقرة عندما يكون معامل إرتباطها بالدرجة الكلية واطئاً على اعتبار أن الفقرة لا تقيس الظاهرة التي يقيسها الاختبار باكمله أو أنها قد نختار مجموعتين متطرفتين من أفراد بناء على الدرجات الكلية التي حصلوا عليها في الاختبار، ثم نستخرج الفرق بين إجابات أفراد هاتين المجموعتين في كل عبارة من عباراته وتحذف العبارة التي لا تظهر تمييزاً واضحاً بين هاتين المجموعتين وقد نقوم باستخراج العلاقة بين درجات الاختبارات الفرعية (إذا كان الاختبار يتتألف من عدة اختبارات فرعية) وبين الدرجة الكلية للاختبار.

3- تقديرات الخبراء

في هذه الطريقة نلجأ إلى الخبراء لتقدير الأفراد المبحوثين في الصفة السلوكية التي يقيسها الاختبار فلو فرضنا أن أحد الباحثين صمم اختباراً لقياس التكيف الاجتماعي لدى طلبة المرحلة الإعدادية وعرف الباحث التكيف الاجتماعي بأنه قدرة الفرد على ملائمة سلوكه مع ما تتطلبه المواقف الاجتماعية فإنه لغرض أن يعرف فيما إذا كان الاختبار يقيس التكيف الاجتماعي أم لا يختار مجموعة من الخبراء (المدرسون في المدرسة مثلاً)... الذين يجهلون اجابة الطلبة على المقياس ويطلب منهم تقدير سلوك كل طالب في بعض المواقف المدرسية التي تدل على التكيف فإذا كانت درجة الارتباط بين تقديرات الخبراء واجبات الطلبة على المقاييس عالية فإن ذلك يدل على أنه يتتوفر في المقياس صدق البناء.

4- استخدام التحليل العاملی

يهدف التحليل العاملی إلى معرفة مدى قیاس الاختبار للظاهرة التي صمم الاختبار لأجلها وتعتمد فکرته على حساب معاملات الارتباط بين كل عبارة والعبارات الأخرى ويتجمع بين كل مجموعة عامل او أكثر ونتيجة لهذه العملية فان الاختبار يخترل الى عدد صغير من العوامل أو السمات المشتركة والتي يطلق عليها (المكونات الأساسية) للظاهرة التي يقيسها الاختبار ومن خصائص هذه الطريقة أنها تنقى الاختبار من العبارات الضعيفة التي يظهر انها لا ترتبط بأي عامل أو مكون أساسي من مكونات الاختبار وهذا يعني أن هذه العبارة لا تقيس الظاهرة التي وضع الاختبار لقياسها.

1- 1- 5 صدق التعلق بمحك ، الصدق التجربی

يطلق أحياناً على صدق التعلق بمحك اسم الصدق الواقعي والعلمی ويقصد به مجموعة الإجراءات التي يمكن من خلالها من حساب الارتباط بين درجات الاختبار وبين محك خارجي مستقل هو السلوك نفسه أو النشاط الذي يتناوله الاختبار بالقياس وذكر تقدیر اللجنة القومية الامريكية لمعايير الاختبارات أن صدق التعلق بمحك واجراءاته يتم بواسطة مقارنة درجات الاختبار بتغير أو متغيرين خارجين يعتبران مناسبين لتوفير قیاس مباشر للسمة موضوع الاختبار وبالنسبة للاستخدام التنبؤي للاختبار الذي يضمن توقعات على المدى البعيد تجمع بيانات المحك في تاريخ لاحق بيانات الاختبار.

إن صدق التعلق بمحك يقبل التصنيف الى فئتي رئيسين الصدق التنبؤي والصدق التلازمي .

1-5-1 الصدق التنبؤي Ptedictuve validity

يشير الصدق التنبؤي إلى مدى الصحة التي نستطيع أن نقدم بها تخمينات أو توقعات على خاصية معينة لأي الفرد من خلال وسيلة مستقلة بمعنى آخر هو المدى الذي نستطيع أن نقيس به خاصية معينة تحدث في وقت آخر من خلال اختبار يفترض انه يقيس

هذه المعايير.

السمة الهمامة في الصدق التنبئي هي انه يفترض توفر قياس مستقل للخاصية يكن الأخذ به باعتباره قياساً محكياً وهذا القياس المستقل هو المحك .

وعادة ما نلجأ إلى حساب معامل الارتباط المناسب بين الاختبار والمحك الخارجي ، ويعد معامل الارتباط هنا معامل صدق للاختبار ، ويستخدم الصدق التنبؤي في الأغراض العملية ، بصفة خاصة فنحن نرغب في وقت مبكر في معرفة من من الأفراد سيحقق نجاحاً دراسياً أو من اللاعبيين سيكون أداؤه عالياً واختبارات الاستعدادات بصفة عامة تبدو أكثر من غيرها تعلقاً بهذا النوع من الصدق ويقصد بالصدق التنبؤي مدى قدرة الاختبار على التنبؤ بنتيجة معينة ويستخدم هذا النوع من الصدق في اختبارات الاستعداد التي تهدف في الأصل إلى التنبؤ بما يمكن أن ينجزه التلميذ في المجال الدراسي أو المهني في المستقبل ولكي يحصل الباحث على هذا النوع من الصدق فإنه يلجأ إلى معايير مختلفة أهمها:

- التحصيل الدراسي العام

يعتبر التحصيل الدراسي أحد المؤشرات المهمة للتنبؤ بصدق اختبارات الذكاء أو الاستعداد الدراسي ويمكن توضيح ذلك بالمثال الآتي:

لنفرض أن أحد المعلمين أراد أن يعرف مدى قدرة أحد اختبارات الاستعداد القرائي على التنبؤ بتحصيل التلاميذ في موضوع القراءة وقد أعطى لهم اختبار الاستعداد في بدء السنة الدراسية وحصل منه على درجة لكل تلميذ ثم أعطى لهم اختباراً تحصيلياً في موضوع القراءة في نهاية السنة وحصل منه على درجة لكل تلميذ أيضاً فإذا كانت العلاقة الارتباطية عالية بين درجات التلاميذ على اختبار الاستعداد القرائي الذي أعطى لهم في بدء السنة الدراسية ودرجاتهم في الاختبار الذي أعطى لهم في نهاية السنة الدراسية فإن هذا يدل على قدرة اختبارات الاستعداد على التنبؤ.

2- إنجاز الفرد في العمل نفسه

يتطلب هذا المعيار متابعة إنجاز الفرد الفعلي في عمل ما فمثلاً يقوم الباحث بعد إعطاء الاختبار بمحصلة العامل أو الموسيقي أو المدرس في عمله الفعلي ثم توصله درجة على أدائه للعمل وتحقق الصدق التنبؤي في اختبار الاستعداد أو الذكاء أو الشخصية عندما تكون العلاقة الارتباطية بين درجة اجابة الفرد في الاختبار وأدائه في العمل عالية وفي حالة تعذر ملاحظة الفرد أثناء قيامه بالعمل فإنه يمكن الاعتماد على بعض المؤشرات الدالة على إنجازه في العمل منها مقدار ما يتجه الفرد واستمراره في العمل والتقدم فيه رضى رؤسائه عنه أم لا ، أما الصدق التنبؤي يعني مدى صلاحية الاختبار في الاشارة إلى مستوى الانجاز الذي يتظر أن يصل إليه المفحوص في موضوع مماثل أو قريب من حيث نوعه أو صعوبته والوصول إلى هذا الصدق يساعد في التحقق من قدرة الاختبار على تحقيق أحد اغراضه التي وضع من أجلها وهي اتخاذ قرارات حول المستقبل ومثال على هذا النوع من أنواع الصدق الاختبار ما يتعلق بقدرة التنبؤ بعلامات الطلاب في الدراسة الجامعية من خلال علاماتهم في شهادة الدراسة الثانوية واختبارات الاستعدادات كاختبار الاستعداد القرائي .

بعد انتهاء الطلاب من الدراسة الثانوية يتهيئون للدراسة الجامعية وذلك حسب معدلاتهم التي أحرزواها نتيجة للاختبارات (الامتحانات) العامة من جهة ومعدل نتائج بعض مواد التخصص الخاصة بكليات معينة من جهة أخرى فالطلاب ذو المعدلات العالية وخصوصاً في المواد العلمية تكون لهم الأفضلية بالدراسة في بعض الكليات مثل الطب والصيدلة والهندسة ولدراسة هذه الظاهرة تتبع حالة هؤلاء الطلاب في كلياتهم من ناحية المستوى التحصيلي وهل هو موفق في دراسته وهل أحرز تقدماً أم لا ، فإذا دلت الإحصائيات لمجموعة هؤلاء الطلاب على تقدمهم وحصولهم على معدلات كبيرة ثبت بذلك تنبؤات واضعي الاختبارات والتي اعتبرت كمعايير لقبولهم في مثل هذه الكليات وبذلك يمكن استخدام الاختبارات التي خاضها هؤلاء الطلاب في امتحانات الشهادة

الثانوية وما تلتها من قياسات وذلك بالتنبؤ بنجاحهم في الكليات التي درسوا فيها.

1-5-2 الصدق التلازمي concurrent validity

لا تقتصر استخدامات صدق التعلق بمحك على تقدير الخصائص التنبؤية للدرجة على الاختبار ومدى تفسير هذه الدرجة لأداء الفرد في فترات لاحقة بل يتعد هذا الاستخدام للمواقف الراهنة المتعلقة بظروف يتزامن فيها الاختبار ودرجاته والمحك ودرجاته ويصبح الهدف هو معرفة ما إذا كان من الممكن احلال الاختبار باجراءاته البسيطة وغير المكلفة بدلاً من المحك الذي قد يكون مكلفاً في حالات التشخيص المرضي أو اختبار كفاءة الاداء لاعمال معينة فكلا المتغيرين (الاختبار والمحك) يقيسان خصائص قائمة بالفعل في وقت واحد ولكن الاختبار أسهل استخدام وأبسط وأقل تكلفة، ويرى انسازي أن الصدق التلازمي هو أنساب الأساليب ويمكن تقديم التبريرات التي تدعمه كما ترى أن التمييز المنطقي بين الصدق التلازمي التنبؤي لا يقوم على الفروق الزمنية بين الاختبار والمحك في الحالتين ولكن على الأهداف من الاختيار فالفرق يتضح من الاختلاف بين صيغة هذين السؤالين هل مصطفى رياضي (صدق تلازمي).

وهل مصطفى مكن أن لا يكون رياضياً (صدق تنبؤي).

في هذا النوع من الصدق لا تكون هناك فترة بين الاختبار الأول الذي صممه الباحث والذي يريد التعرف على صدقه في قياس الظاهرة وبين المعيار الخارجي ففي الصدق التنبؤي نحصل على درجات المعيار الخارجي بعد فترة من تطبيق الاختبار الجديد قد تصل إلى سنة أو أكثر أما درجات المعيار الخارجي في الصدق التلازمي فتحصل عليها أثناء تطبيق الاختبار الجديد وهناك عدة طرق نستخدمها في الصدق التلازمي .

1- إعطاء اختبار يقيس نفس الظاهرة التي يقيسها الاختبار الأول فعندما يصمم باحث اختباراً لقياس اتجاهات الطلبة نحو عمل المرأة ويريد أن يتعرف على صدقه فإنه يقوم بإعطاء اختبار آخر يتناول نفس الظاهرة إلى عينة الطلبة نفسها ثم يستخرج العلاقة

الارتباطية بين درجات الطلبة في الاختبارين ويكون الاختبار صادقاً صدقاً تلازميًّا عندما تكون الدرجة الارتباطية عالية بين الاختبارين .

2- تقديرات المسؤولين عن العمل لإنجاز العاملين لديهم أو قد يتضمن مثلاً تقديرات المعلمين لظاهره سلوكيه معينة لدى الطالب فعندما يصمم باحث اختباراً لقياس دافعية الطلبه نحو الدراسة ويريد التأكد من صدق الاختبار فإنه يطلب من مجموعة من المعلمين الذين يعرفون عينة الطلبه الذين يطبق عليهم الاختبار أن يقدروا دافعية كل طالب منهم وفق مقياس يندرج من الدافعية العالية جداً إلى الدافعية الواطئة جداً نحو الدراسة ثم يحسب العلاقة الارتباطية بين درجات التلاميذ على الاختبار وتقديرات المعلمين لهم ويكون الاختبار صادقاً إذا كانت العلاقة الارتباطية عالية .

3- المجموعات المتضادة - إن الاختبار الصادق هو الذي يميز بين الأفراد فاختبار الذكاء يجب أن يميز بين الاذكياء والأقل ذكاءً واختبار القلق يجب أن يميز بين الفرد القلق وغير القلق وهكذا وعلى هذا الأساس يختار مصمم الاختبار مجموعتين من الأفراد تختلفان في الظاهرة التي يقيسها الاختبار ويعطي لهما الاختبار فإذا كانت هناك فروق واضحة بين اجابتي المجموعتين في الاختبار فإنه يكون قد تحقق فيه الصدق التلزامي .
ويكن التمييز بين الصدق التنبؤي والصدق التلزامي بما يلي :

أ- وقت الحصول على درجات المعيار

ففي الصدق التنبؤي نحصل على درجات المعيار بعد فترة من تطبيق الاختبار الجديدة وقد تصل هذه الفترة الى سنة أو أكثر أما درجات المعيار في الصدق التلزامي فيحصل عليه أثناء تطبيق الاختبار الجديد .

ب- الفائدة من الاختبار

يستفاد من الاختبار الذي يتحقق فيه الصدق التنبؤي في التعرف على درجة إنجاز

الفرد في المستقبل أما الصدق التلازمي فيستفاد منه في الاختبارات التي تقيّم السلوك الحالي للفرد.

ج- يستخدم الصدق التنبؤي عادة في اختبارات الذكاء أو في إختبارات الإستعداد الدراسي أما الصدق التلازمي فإنه يستخدم في اختبارات الشخصية والتحصيلية أيضاً.

وهناك ثلاثة شروط أساسية يجب توافرها في المعيار الجيد (في كل من الصدق التنبؤي والتلازمي) هي:

١- أن لا يتأثر المعيار بالعوامل الذاتية

فالدرجات المدرسية مثلاً تعتبر أحد المعايير المهمة للتوصل إلى صدق الاختبار سواء كان تنبؤياً أو تلزماً ولكن إستخدامها يتطلب شيئاً من الحذر عندما لا يعتمد تقديرها على اختبارات تحصيلية مفتوحة وذلك لأن درجة التلميذ قد تتأثر بالعوامل الذاتية لدى المصحح كما يعتبر معيار تقييمات المعلمين والمشرفين على العمل والأقران مهما في الصدق التنبؤي والتلزمي لدى تقييم بعض جوانب الشخصية، ومثل هذا المعيار يتأثر أيضاً بالأحكام الذاتية للمعلم أو المسؤول عن عمل الفرد الذي يراد تقييمه.

2- أن تبقى الدرجة التي يحصل عليها الفرد فيه ثابتة إذا أعيد تطبيقه عليه أكثر من مرة ولكي يتتأكد بعض الباحثين من ثبات معيار (تقديرات المعلمين والمسؤولين ، فانهم يطلبون منهم أن يعيدو تقدير الأفراد مرة ثانية فإذا تطابقت إجاباتهم في المرتين فإن ذلك يدل على ثبات المعيار.

3- أن يكون المعيار وثيق الصلة بما يقيسه الاختبار ، فإذا كان الاختبار يقيس استعداد الطالب في موضوع الرياضيات أو الاستقرار الانفعالي فيجب أن يكون المعيار المستخدم (الدرجات المدرسية أو تقديرات المعلمين أو الأقران) متعلقاً بالموضوع نفسه دون غيره ليكون معياراً صادقاً يمكن الاعتماد عليه في الحكم على الاختبار الجديد.

4- يحكم على الاختبار بأن فيه صدقاً تلزماً من خلال الارتباط والتلزام بينه وبين اختبار آخر ثبت صدقه ويتم التأكد من هذا الصدق بأن تطبق الاختبارين الاول والثاني على مجموعة من المفحوصين ونضع لكل مفحوص علامتين أحدهما على الاختبار الاول والثانية على الاختبار الثاني ثم يحسب معامل الارتباط بين نتائجتي المفحوصين على الاختبارين فإن كان معامل الارتباط عالياً كان الاختبار صادقاً والعكس صحيح.

يعتبر الصدق التلزامي نوع من أنواع الصدق التجاري حيث أنه يبين الارتباط بين حالة اللاعب الراهنة ونتيجه في الاختبار فمثلاً يمكن للمدرب تقدير مستويات لاعبيه الحالية والتي من خلال التدريب ثم يوجد معامل الارتباط بين حالاتهم ونتائج المسابقات التي قد يخوضونها في نفس المرحلة.

(مثال) مدرب ألعاب القوى يقوم بتدريب خمسة لاعبين في رمي القرص ومستوياتهم متقاربة أعطى رأياً في مستوى لاعبيه وأعطى لكل منهم درجة وفي يوم إقامة البطولة تسابق اللاعبون وحققوا مستويات قريبة إلى رأي مدربهم وباستخدام معامل ارتباط الرتب بين تقديره والمستوى الذي تحقق في البطولة وجد أنه معنوي (0.92٪) هذا يدل على الصدق التلزامي.

وهذا النوع من الصدق يمكن التعبير عنه عادةً عند استخدامنا لاختبار مادون التتحقق من صدقه ولكن تحقق وجود معامل ارتباط وبدرجة معنوية عالية بينه وبين اختبار آخر يقيس نفس الصفة أو السمة وعلى ذلك تعتبر كلا الاختبارين متناظرين ويمكن الاعتماد عليهم واختبار نفس الصفة وبدرجة واحدة وبذلك يمكن قبول نتائج الاختبار الذي لم يتحقق في معامل صدقه واعتباره في مثل هذه الحالة صادقاً بالتطابق.

2-1 تصحيح معامل الصدق

أشرنا إلى أهمية ثبات المحك وأن هذا الثبات يؤثر إلى حد كبير في معامل الصدق أي الارتباط بين المحك والاختبار غير أن هناك أيضاً مشكلة لثبات المحك التي تلعب الدور

نفسه في تقدير صدق الاختبار لهذا يصبح من الضروري إعادة تقدير معامل صدق الاختبار في ضوء معامل الثبات الخاصة بكل منها . فإذا أردنا إجراء تصحيح شامل لمعامل الصدق في ضوء ثبات المحك والاختبار فإننا نستخدم المعادلة الآتية /

$$\frac{\text{س ص}}{\sqrt{\text{س س} \times \text{ص ص}}} = \text{د ح 1}$$

حث:

ر ١ = الإرتباط بين المكونات الحقيقية أو التباين الحقيقي في الاختبار والمحك .

$\text{س، ص} = \text{معامل الارتباط المحسوب بين المحك والاختيار.}$

(س س) ص ص = معاملات الشّيات لـكل من س ، ص أي المحك والاختيار.

ويستخدم معامل الثبات المناسب والذي يتولى تقدير تباين الخطأ الأقرب لأنواع التحيزات التي تحدث في المحك وان كان جيلفورد يذكر أن معامل الثبات في هذه المعادلة ما زال موضوعاً للجدل ووجهات النظر المختلفة غير أنها لا تحتاج في الواقع الأمر لإجراء هذا التصحيح الكامل لكل من ثبات المحك وثبات الاختبار فالاختبار يفترض أن به درجة من الخطأ.

$$\frac{رس ص}{رس ص} = ح_1$$

(وكل الرموز بمعانٍها السابقة في المعادلة السابقة)

وعلينا أن نلاحظ عند تطبيق معادلة لتصحيح الصدق أن المعادلة لا تصحيح أخطاء العينة ولا أخطاء القياس التي يمكن أن تكون لها دورها في التأثير على معامل الصدق الذي يحصل عليه في شكل ارتباط بين المحك والاختيار، درجات الطلبة في معاهد المعلمين

ودرجاتهم على مقياس للتبؤ بنجاحهم في مهنة التعليم فكيف يفسر معامل الصدق الواطئ هذا الذي حصل عليه؟

يمكن أن تفسر هذه النتيجة على أساس أن المعلومات التي درسها الطلبة في المعهد ليست كافية للنجاح في مهنة التعليم أو ربما أن الدرجات الأكادémie التي حصل عليها الطلبة لا تقيس من الناحية الفعلية معلومات الطلبة في الموضوعات الدراسية أو ربما أن الاختبار الذي صمم لتشخيص المعلم الناجح وتميزه عن غير الناجح اختبار غير سليم أو ربما أن العينة متميزة وغير ممثلة لعموم الطلبة في معاهد المعلمين.

- ومن العوامل التي تزيد من صدق الاختبار ما يلي :

- التوازن بين الجوانب التي تتضمنها فقرات الاختبار وبين الجوانب التي يقيسها الاختبار.
- وجود تعليمات واضحة ومحددة للاختبار.
- وضوح اللغة والطباعة.
- جدية المفحوصين في الإجابة عن فقرات الاختبار.
- الدقة والحزم في تطبيق الاختيار وعدم السماح بالغش لأن الغش يعمل على إفساد صدق الاختبار.
- الدقة في تصحيح الاختبار.
- شمول فقرات الاختبار وتمثيلها لجوانب السلوك المختلفة المراد قياسها.
- تحديد طبيعة الصفة التي يقيسها الاختبار وتحديد انماط السلوك المدرجة تحت هذه الصفة.
- تطبيق الاختبار على عينة الطلاب التي أعد لها.

3-1 الطرق الاحصائية لقياس الصدق:

تلخص أهم الطرق الاحصائية المعروفة لقياس الصدق فيما يلي :

3-1-1 طريقة معاملات الارتباط:

إن معامل الصدق يساوي معامل ارتباط الاختبار بالميزان أيًّا كان نوع هذا الميزان

اختباراً أو عاملأً أو أي مقياس آخر. وبما أن معامل الصدق يدل على مدى صلاحية الاختبار للتنبؤ بدرجات الميزان حتى نستعين بمثل ذلك الاختبار بعد ذلك في قياس الاستعداد للدراسة أو المهنة التي يقيسها ذلك الميزان. إذن فالصدق وحده لا يصلح بصورةه المباشرة للتنبؤ، ولذا يحسب التنبؤ بطريقة الانحدار والمثال التالي يوضح هذه الفكرة:

لنفرض أن الرمز (ص) يدل على درجات الميزان فالمعادلة التي تصلح لاستنتاج درجات الميزان من درجات الاختبار هي معادلة انحدار ص على س وقد سبق أن درسنا هذه المعادلة في الصورة التالية:

$$\text{ص} = \text{ر} \times \frac{\text{س}}{\text{ع}} + (\text{س} - \text{م}_\text{س})$$

وهكذا نستطيع أن نتبأب بدرجة أي فرد في الدراسة أو المهنة المقبلة وذلك بمعرفة درجته في الاختبار الذي حسبنا معامل صدقه بالنسبة لتلك الدراسة أو المهنة. لكن هذا التنبؤ يتأثر بأخطاء العينات، ولذا يجب أن نعرف مدى الدلالة، فعلينا أن نحسب الخطأ المعياري للتنبؤ بدرجات (ص) من درجات (س).

ويحسب الخطأ المعياري للانحدار بالمعادلة التالية:

$$\text{ع}_\text{ص} / \text{س} = \sqrt{1 - \text{ر}^2}$$

حيث يدل الرمز ع_ص / س على الخط المعياري لانحدار س على ص

ويدل الرمز ع_ص على الانحراف المعياري لدرجات الميزان ص ويدل الرمز (ر) على معامل صدق الاختبار. أو يعني آخر معامل ارتباط الاختبار بالميزان.

هذا ويمكن حساب $\sqrt{1 - \text{ر}^2}$ مباشرة من معامل الاغتراب وذلك بالاستعانة بجدول رقم 15 المبين بملحق الجداول الاحصائية النفسية الذي يدل على المقابلات

الاغترابية للارتباط . وقد سبق أن بينما أن الاغتراب $\text{غ} = \sqrt{1 - \rho^2}$ وهكذا نستطيع أن نعيد صياغة المعادلة السابقة في الصورة التالية :

$$\text{ع ص} / \text{س} = \text{ع ص} \times \text{غ}$$

$$\text{فإذا فرضنا أن معامل الصدق } \rho = 0.75$$

$$\therefore \text{معامل الاغتراب } \text{غ} = 0.66$$

$$\text{وفرضنا أن الانحراف المعياري } \text{س} = 6.5$$

$$\therefore \text{ع ص} / \text{س} = 0.66 \times 6.5$$

$$= 4.3 \text{ تقريرياً}$$

أي أن حدود أي درجة من درجات الميزان ص التي تقابل الدرجة س من درجات الاختبار س تمتد من (ص-4.3) إلى (ص+4.3) ، واحتمال وقوع الدرجة في هذا النطاق إلى احتمال وقوعها خارج هذا النطاق يساوي 2 إلى 1 كما سبق أن بينما ذلك في تفسيرنا لمعنى الدلالة الاحصائية للخطأ المعياري .

3-2 طريقة المقارنة الظرفية:

عندما تدل نتائج الاختبار على أن الأقواء في الميزان أقوىاء في الاختبار صادقاً، ويزداد الصدق تبعاً لزيادة هذا الاقتران ويتناقص تبعاً لتناقص هذا الاقتران، ولذا نرى الأهمية الظرفية لمستويات الميزان في هذه المقارنة . ومن أبسط الطرق التي تستخدم لتحقيق هذه الفكرة مقارنة متوسطات درجات الأقواء بمتوسطات درجات الضعاف ثم حساب دلالة الفروق بين هذه المتوسطات ، وعندما تصبح لتلك الفروق دلالة احصائية واضحة نستطيع أن نقرر أن الاختبار يميز بين الأقواء والضعف في الميزان ، وبذلك نطمئن إلى صدقه ، وعندما لا تصبح لتلك الفروق دلالة احصائية واضحة فإننا لا نستطيع الاطمئنان إلى صدق مثل هذا الاختبار .

أي أن هذه الطريقة تدل على صدق الاختبار ولا تدل بطريقة عددية أكيدة على مقدار هذا الصدق ولذا يقتصر استخدامها على الاحكام السريعة التمهيدية التي تفصل الاختبارات

المختلفة الى ما هو صادق وما هو غير صادق بالنسبة لميزان ما ، وتصلح أيضاً لترتيب تلك الاختبارات ترتيباً يدل على مدى صدقها بالنسبة للميزان . هذا ولا غنى للباحث عن هذه الطريقة عندما لا يستطيع الحصول على ترتيب جميع الافراد بالنسبة لمستويات الميزان المختلفة بل يستطيع فقط الحصول على الافراد الممتازين والضعاف . والجدول يوضح طريقة حساب فروق المتوسطات الطرفية والكشف عن دلالتها الاحصائية :

الفئات درجات الاختبار	متصفات الفئات	متصفات الفئات	متصفات الفئات	متصفات الفئات	متصفات الفئات	متصفات الفئات القوي
						تكرار المستوى الميزياني القوي X
	0	52	1	52	54.50	
	0	114	2	57	59.55	
	0	62	1	62	64.60	
	0	335	5	67	69.65	
216	3	576	8	72	74.70	
462	6	308	4	77	79.75	
574	7	0	0	82	84.80	
435	5	0	0	87	89.85	
276	3	0	0	92	94.90	
291	3	0	0	97	99.95	
2254	مج = 27	مج = 1447	مج = 21			
$\frac{2254}{27} = 1$		$\frac{1447}{21} = 1$				
83.48 =		68.9 =				
7.43 = 1 ع		6.81 = 1 ع				

طريقة حساب المتوسطات الطرفية وانحرافاتها المعيارية

ويدل العمود الأول في هذا الجدول على فئات درجات الاختبار وبذلك تمتد الفئة الأولى من 50 إلى 54 والثانية من 55 إلى 59 وهكذا حتى تمتد الفئة الأخيرة من 95 إلى 99. وتدل درجات العمود الثاني على متصفات تلك الفئات فتتصف الفئة الأولى 52 ومتتصف الفئة الثانية 57، ومتتصف الفئة الأخيرة 97.

لقد رصدنا في العمود الثالث تكرار أفراد المستوى الضعيف في الميزان كل أيام درجته في الاختبار، وبذلك يدل السطر الأول في هذا العمود على أن فرداً واحداً من أفراد المستوى الضعيف في الميزان حصل على درجة في الاختبار تقع في الفئة الأولى لدرجات هذا الاختبار التي تمتد من 50 إلى 54، ويدل السطر الثاني على أن 2 من أفراد هذا المستوى حصلا على درجات في الاختبار تقع في الفئة التي تمتد من 55 إلى 59. وهكذا بالنسبة للفئات الأخرى.

ويدل العمود الرابع على حاصل ضرب متتصف كل فئة من فئات الاختبار في التكرار المقابل لها، وبذلك يبين السطر الأول في هذا العمود حاصل ضرب $52 \times 1 = 52$ وبين السطر الثاني حاصل ضرب $257 = 114$ وهكذا بالنسبة لبقية الفئات، وقد حسب متوسط درجات أفراد هذا المستوى وذلك بقسمة مجموع الدرجات المساوي لـ 1447 على درجات أفراد هذا المستوى الذي يساوي 21، وبذلك أصبح المتوسط مساوياً لـ 68.9.

يدل العمود الخامس على تكرار أفراد المستوى القوي في الميزان بالنسبة لفئات درجات الاختبار، فمثلاً يدل السطر الأخير على أن عدد أفراد المستوى الممتاز الذين حصلوا على درجات في الاختبار تقع في الفئة 95-99 هو 3، ويدل السطر الذي قبله على أن عدد أفراد هذا المستوى الذين حصلوا على درجات في الاختبار تقع في الفئة 90-94 هو 3 أيضاً، وهكذا بالنسبة لبقية تكرار هذا العمود.

ويدل العمود السادس على حساب متوسط هذا المستوى بنفس الطريقة التي أتبناها في حساب متوسط المستوى الضعيف، وبما أن مجموع تكرار هذا العمود يساوي 27،

ومجموع درجات هذا المستوى يساوي 2254 إذن فمتوسط هذا المستوى يساوي 83.48 .
أي أن : متوسط درجات أفراد المستوى الميزاني الضعيف - 68.90 ومتوسط درجات
أفراد المستوى الميزاني القوي - 93.48

ولحساب الدلالة الاحصائية للفرق القائم بين هذين المتوسطين نحسب أولاً الخط
المعياري لكل متوسط وذلك بحساب الانحراف المعياري لدرجات كل مستوى من هذين
المستويين ، ثم نستعين على حساب دلالة الفرق بالنسبة الحرجة .

وقد سبق أن بينا أن

$$\text{النسبة الحرجة} = \frac{1^M - 2^M}{\sqrt{2^M + 1^M}}$$

وذلك بالنسبة للمتوسطات غير المرتبطة ، هذا وتحسب الأخطاء المعيارية
للمتوسطات من المعادلات التالية :

$$\frac{\bar{U}_2 - \bar{U}_1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{U}_1 - \bar{U}_2}{\sqrt{\frac{1}{n_2} + \frac{1}{n_1}}}$$

لكن الانحراف المعياري لدرجات المستوى الميزاني الضعيف $\bar{U}_1 = 6.81$

$$\therefore \text{الخط المعياري المتوسط درجات هذا المستوى } \bar{U}_1 = \frac{6.81}{\sqrt{21}}$$

والانحراف المعياري لدرجات المستوى الميزاني القوي $\bar{U}_2 = 7.43$

$$\therefore \text{الخط المعياري لمتوسط درجات هذا المستوى } \bar{U}_M = \frac{6.81}{\sqrt{27}} = 1.43$$

$$\therefore \text{النسبة المحرجة} = \frac{68.90 - 83.48}{\sqrt{^2(1.49) + ^2(1.43)}}$$

$$= \frac{14.58}{\sqrt{2.2201 - 2.0449}}$$

$\therefore \text{النسبة المحرجة} = 70.4$ تقريرياً.

وبما أن هذه النسبة تزيد على 2.58 تقريرياً.

وبما أن هذه النسبة تزيد على 25.8 درجة معيارية أو على 3، اذن فالفرق القائم بين المتوسطين له دلالة احصائية اكيدة ولا يرجع الى الصدفة أي أن درجات هذا الاختبار تميز تميذاً واضحاً بين المستويات الضعيفة والقوية للميزان سواء أكان هذا الميزان مهنة أو عملاً أو دراسة أي أن هذا الاختبار صادق في قياسه لتلك الصفة التي يقيسها الميزان . هذا ونستطيع أن نحصل على ترتيب جميع الافراد من الميزان ثم نقسم هؤلاء الافراد إلى قسمين : قوي و ضعيف ، ونحسب بعد ذلك معامل ارتباط هذا التقسيم الثنائي للميزان بالتدريج المتتابع للاختبار بطريقة معامل الارتباط الثنائي أو الثنائي الاصليل لنحصل على القيمة العددية مثل هذا الصدق ، وبذلك نطور هذه الطريقة التقريرية إلى دقة الطريقة الأولى التي تعتمد على حساب مثل ذلك الارتباط .

وترجع فكرة هذه الطريقة إلى تقسيم مستويات الميزان بالوسطى الى طرفين : علوى وسفلي أو ما فوق الوسيط وما دون الوسيط ، ثم يحسب بعد ذلك معامل الارتباط لهذا التقسيم الثنائي ويختار من القسم العلوى الى 27٪ الأقوباء ويختار من القسم السفلي الى 27٪ الضعاف ويحسب من ذلك معامل الارتباط من جدول فلانagan (J. FLANGAN) (A. P. JOHNSON) المبين بلحق الجداول الاحصائية النفسية أو بطريقة جونسون السريعة كما سنبي ذلك بالتفصيل في تحليلنا لصدق اسئلة الاختبارات في الفصل الثاني .

3-3 طريقة الجدول المترقب:

تعتمد هذه الطريقة على الافادة من التكرار المزدوج للاختبار والميزان في تقدير صدق الاختبار ، و تؤدي الى الكشف عن معرفة النسب المئوية للنجاح في كل مستوى من مستويات الميزان بالنسبة لكل مستوى من مستويات الاختبار .

وتتلخص خطوات هذه الطريقة في حساب جدول التكرار المزدوج للاختبار والميزان ثم تحويل خلايا هذا الجدول إلى ما يسمى بالجدول المترقب وذلك بحسب النسبة المئوية لكل تكرار ، وبذلك نستطيع تفسير نتائج الاختبار في ضوء هذه النسبة المئوية والمثال التالي المبين بالجدول يوضح خطوات هذه الطريقة .

المجموع	مستويات النجاح في الميزان					جدول التكرار المزدوج للاختبار والميزان	
	5	4	3	2	1		
33		6	11	13	3	59.50	
63		6	30	21	6	69.60	فئات
114	9	24	45	24	12	79.70	درجات
60	12	24	15	9		89.80	الاختبار
30	6	18	6		0	99.90	

التكرار المزدوج لفئات درجات الاختبار ولمستويات النجاح في الميزان

حيث يدل العمود الأول على فئات الدرجات التي تبدأ بالفئة 50-59 وتنتهي إلى الفئة 90-99 .

ويدل السطر الأول على مستويات الاداء والنجاح التي تبدأ بالمستوى الاول الذي

يعد أضعف هذه المستويات ويليه المستوى الثاني الذي يفضله في القوة ثم تنتهي إلى المستوى الخامس الذي يعد أقوى هذه المستويات.

وتدل الخلايا الداخلية لهذا الجدول على التكرار المزدوج للاختبار والميزان، وبذلك نرى أن التوزيع التكراري لمستويات الميزان بالنسبة للفئة الأولى لدرجات الاختبار التي تمت من 50 إلى 59 هو 3 افراد في المستوى الميزاني الأول، 13 فرداً في المستوى الميزاني الثاني، 11 فرداً في المستوى الميزاني الثالث ، 6 افراد في المستوى الميزاني الرابع، وصفر في المستوى الميزاني الخامس أي تكرار النجاح في المهنة بالنسبة للفئة الأولى 50-59 يميل الى التجمع في المستويات الدنيا لهذا الميزان.

إن الفئة الدنيا للاختبار تقترب إلى حد ما بالمستويات الضعيفة للميزان ، وي يكن أن يستطرد فهمنا لخلايا هذا الجدول حتى تصل إلى أعلى فئات الدرجات التي تمتد من 90 إلى 99 فنرى أن التوزيع التكراري لمستويات الميزان يساوي صفرأ في المستوى الميزاني الأول ، وصفرأ في المستوى الميزاني الثاني ثم يرتفع هذا التكرار ليساوي 6 أفراد في المستوى الميزاني الثالث 182 فرداً في المستوى الميزاني الرابع ، 6 أفراد في المستوى الميزاني الخامس ، أي أن الفئة العليا للميزان تقترب إلى حدما بالمستويات القوية للميزان . لكن هذا الجدول بصورةه القائمة لا تدل بطريقة واضحة أكيدة على المقارنة الاقترانية لفئات الاختبار ومستويات الميزان ، ولذا نحسب النسبة المئوية لخلايا للنجاح في كل مستوى من مستويات الميزان بالنسبة لكل فئة من فئات الاختبار . ونحسب هذه النسبة بقسمة كل تكرار على المجموع المقابل له في نهاية السطر ، ثم يضرب الناتج بعد ذلك في مائة .

والخطوات التالية تووضح طريقة حساب هذه النسب:

التكرار المزدوج للفعل 59 وللمستوى الميزاني الاول يساوي 3 ، وبما أن مجموع

تكرار هذا السطر يساوي 33.

$$\text{النسبة المئوية لتكرار هذه الخلية} = \frac{3}{23} \times 100 = 9 \text{ تقريباً.}$$

وهكذا بالنسبة لبقية الخلايا كما يدل على ذلك الجدول الثاني .

المجموع	مستويات النجاح في الميزان					التكرار المزدوج المثوي للختبار والميزان	
	5	4	3	2	1		
99		18	33	39	9	59.50	
101		10	48	33	10	69.60	فئات
100	8	21	39	21	11	79.70	درجات
100	20	40	25	15		89.80	الاختبار
100	20	60	20			99.90	

(الجدول المرتب أو التكرار المزدوج المثوي لفئات درجات الاختبار ومستويات الميزان) .

ويسمى جدول التكرار المزدوج المثوي للختبار والميزان بالجدول المرتب إذ به نستطيع أن نعلم احتمال النجاح في المهنة بالنسبة لكل فئة من فئات الاختبار فاحتمال النجاح في المستوى الرابع للمهنة يساوي 18٪ بالنسبة للفئة الاولى الاختبارية التي تتد من 50 إلى 59 ، واحتمال النجاح في نفس هذا المستوى يصل 60٪ بالنسبة للفئة الاختبارية التي تتد من 90-99 كما يدل على ذلك الجدول المرتب .

وهكذا نستطيع أن نقدر مدى صدق هذا الاختبار بالنسبة لكل مستوى من مستويات الميزان بطريقة عملية سريعة . هنا ونستطيع أن نجمع البيانات العددية للجدول السابق في أربع خلايا تلخص التكرار المزدوج للمستويات الضعيفة والقوية للميزان والفئات الدنيا والعليا للختبار ، وبذلك نكشف بطريقة سريعة عن صدق الاختبار ونستعين بهذا الصدق في تحديد اختبار الأفراد كما يدل على ذلك الجدول .

المجموع	الميزان		مستويات الضعيف	الجدول الرباعي للتكرار المزدوج		
	القوي	من 3 الى 5		من 1 الى 2	الأدنى	فقات
210	(ب)	131	79	79.50	الأعلى	درجات الاختبار
90	(د)	81	9	99.80		

(الجدول الرباعي للتكرار المزدوج للفئات الدنيا والعليا للدرجات وللمستويات الضعيفة والقوية للميزان).

حيث يدل هذا الجدول أن التوزيع التكراري لمستويات الميزان بالنسبة للفئة الدنيا للدرجات الاختبار التي تتد من 50 إلى 79 هو 79 فرداً في المستوى الميزاني الضعيف التي يتد من 1 إلى 2 ، 131 فرداً في المستوى الميزاني القوي الذي يتد من 3 إلى 5 .

ويدل أيضاً على أن التوزيع التكراري لمستويات الميزان بالنسبة للفئة العليا للدرجات الاختبار التي تتد من 80 إلى 99 هو 9 أفراد في المستوى الميزاني الضعيف الذي يتد من 1 إلى 2 ، 81 فرداً في المستوى الميزاني القوي الذي يتد من 3 إلى 5 .

هذا ونستطيع أن نحسب معامل الارتباط الرباعي مباشرة من هذا الجدول وذلك بقسمة حاصل ضرب الخلايا المتشابهة على حاصل ضرب الخلايا المختلفة . تم قراءة الارتباط الرباعي من الجداول الإحصائية النفسية :

$$\frac{\text{حاصل ضرب الخلايا المتجانسة}}{\text{حاصل ضرب الخلايا المحتملة}} = \frac{1 \times د}{ب \times ج}$$

$$\frac{81 \times 79}{9 \times 131} =$$

$$\frac{6399}{1179} =$$

$$5.427 =$$

هذا ويدل جدول الارتباط الرباعي على انه عندما تكون

$$5.388 = \frac{أ}{ب ج}$$

$$\text{يصبح الارتباط الرباعي } = 0.585$$

ويدل هذا الجدول ايضاً على انه عندما تكون

$$5.959 = \frac{أ}{ب ج}$$

$$\text{يصبح الارتباط الرباعي رب } = 0.595$$

$$\text{وبما أن قيمة } \frac{أ}{ب ج} \text{ في مثالنا هذا } = 0.427$$

إذن فالارتباط الرباعي لمثالنا هذا = 0.6 تقريباً.

أي أن معامل صدق هذا الاختبار بالنسبة لذلك الميزان هو 0.6 هذا ونستطيع أن نحوال الجدول الرباعي للتكرار المزدوج التي جدول مرتب وذلك بحساب النسب المئوية للخلايا كما يدل على ذلك الجدول التالي :

المجموع	الميزان		مستويات القوى	مستويات الضعف	الجدول الرباعي للتكرار المزدوج	
	من 3 الى 5	من 1 الى 2			الأدنى	الأخير
100	62	38	79.50		الأدنى	درجات فئات
100	90	10	99.80		الأعلى	الاختبار

(الجدول المرتب أو الجدول المثوي للتكرار المزدوج للفئات الدنيا والعليا للدرجات وللمستويات الضعيفة والقوية للميزان).

وتفسر نتائج هذا الجدول بنفس الطريقة التي فسرنا بها نتائج الجدول المرتب السابق.

٤- أمثلة لحساب معامل صدق الاختبار احصائياً .

نعرض فيما يلي بعضأ للطرق المستخدمة في إيجاد بعض أنواع الصدق والتي نلجم إليها في مجال التربية البدنية والرياضية وذلك على شكل أمثلة .

مثال (١)

لإيجاد معامل الصدق التنبؤي للاختبار :

تعتبر الاختبارات التي يجتازها الطلاب الراغبون في دراسة التربية الرياضية تقييماً

لقدراتهم البدنية والمهارية والشخصية وعلى ذلك نعرض المثال التالي لايجاد معامل الصدق التنبؤي لبطارية الاختبارات التي تجرى على الطالب بكليات التربية الرياضية بالجامعات العراقية.

البطارية مكونة من:

- 1- الوثب العريض من الثبات.
- 2- رمي كرة طبية (2 كغ).
- 3- اختبار المهارات المختلفة.
- 4- عدو 60 متراً.
- 5- ركض 1500 متراً
- 6- اختبار شخصي.
- تسجيل قدرات كل طالب وتحول المعطيات الى درجات وفق جداول معيارية معينة وتكون الدرجة النهائية لهذه البطارية درجة (100).
- وتتبع مجموعة مكونة من عشرة طلاب وحتى تخرجهم حيث حصلوا على الدرجات المبينة في الجدول التالي وكلها من (100) درجة أيضاً.
- والخطوات التالية لايجاد معامل الصدق التنبؤي لبطارية الاختبارات والتي على أساسها يقوم طلاب التربية الرياضية.

مربع فرق الرتب 2f	فرق الرتب f	رتب ص	رتب س	الطلاب		نتائج الاتحاق عند ص الخروج	
				عند ص	عند س		
				الالتحاق			
صفر	صفر	10	10	70	65	1	
1	1-	9	8	72	68	2	
2.25	1.5-	7.5	6	75	85	3	
2.25	1.5	5.5	7	80	80	4	
1.0	1	4	5	82	87	5	
12.25	3.5-	7.5	4	75	88	6	
12.25	3.5	5.5	9	80	66	7	
1.0	1-	3	2	85	90	8	
1.0	1-	2	1	90	95	9	
4	2	1	3	95	89	10	
37	8+	55	55			(10)	مجـ
	8-						

وبتطبيق قانون معامل ارتباط الرتب لسييرمان :

$$\text{رات} = \frac{222}{990} = \frac{37 \times 6}{990} = \frac{6 \text{ مجـ}^2}{n(n^2 - 1)} - 1$$

$$\text{رات} = 0.78 = 0.22 - 1$$

وبالكشف في جدول 10-2 معاملات الارتباط وجد أن الارتباط دال تحت المستوى 0.05 وغير دال تحت المستوى 0.10 أي أن هناك ارتباط معنوي (جيد) وبدرجة ثقة 0.95 في حين أنه لم يحظ بدرجة الثقة 0.99.

مثال (2) للصدق التنبؤي

أراد أحد الباحثين أن يختبر الصدق التنبؤي لاختبار القفز من وضع القرفصاء في زمن قدره 30 ثاً، فأفترض أن عينة بحثه والتي حققت مستويات جيدة في هذا الاختبار سوف تحقق مستويات جيدة أيضاً في سباق 400 متر وذلك اذا ما أختبروا في سباق 400 متر (وهذا ما تنبأ به الباحث) معتمداً على أن اختبار (القفز من وضع القرفصاء في زمن قدره 30 ثاً، له تأثير ايجابي على 400 م، وعلى ذلك اجرى الباحث سباق 400 متر وسجل الاذمنة لكل لاعب وبمقارنة نتائج الاختبار المراد ايجاد صدقه التنبؤي بتتابع سباق (اختبار) 400 متر، وحصلنا على النتائج التالية المثبتة في الجدول التالي :

حيث تمثل (س) نتائج اختبار (400) متر، (ص) نتائج اختبار القفز .

الاسم	ت	ركض 400 م/ث	قفز القرفصاء 30 ث/ص	س ص	س 2	ص 2
ابتسام	1	113	25	2825	12769	625
الهام	2	85	33	2805	7225	1089
آمال	3	81	27	2187	6561	729
بيان	4	99	24	2376	9801	576
جميلة	5	86	30	2580	7396	900
جنان	6	92	30	2760	8464	900
حياة	7	103	28	2884	10609	784

400	9409	1940	20	97	ختام	8
841	7225	2465	29	85	خديجة	9
625	9604	2450	25	98	رجاء	10
784	8281	2548	28	91	رسالة	11
784	6724	2246	28	82	سامية	12
900	7396	2580	30	86	سحر	13
841	7225	2465	24	85	سلوى	14
625	7921	2225	25	89	شهاد	15
441	10000	2100	21	100	ندى	16
576	12996	2736	24	114	عفاف	17
729	8836	2538	27	94	فاطمة	18
900	8281	2730	30	91	كريمة	19
729	7744	2376	27	88	لمياء	20
400	16384	2560	20	128	ناجحة	21
625	10816	2600	25	104	ناهدة	22
784	8281	2548	28	91	ندى	23
676	7569	2262	26	87	نرذت	24
576	8464	2208	24	92	هند	25
900	8836	2820	30	94	هيفاء	26
484	9409	2134	22	97	هيفا	27
900	9025	2850	30	95	قصيد	28
20123	353251	69848	745	2647	28	مج

ويتطبق قانون الارتباط البسيط لبيرسون:

$$\begin{aligned}
 & \frac{\text{مجس} \times \text{مجس}}{n} - \frac{\text{مجس}^2}{n} = r \\
 & \sqrt{\frac{(2(745) - 2.123)(2(2647) - 253251)}{28 \times 28} - \frac{70429.107 - 69848}{300.680 \times 3014.97}} = r \\
 & \frac{581.107 -}{\sqrt{952.124}} = \frac{581.107 -}{\sqrt{106541.1796}} = r \\
 & r = 0.61
 \end{aligned}$$

وبعد الكشف عند درجة حرية (28-2) وتحت المستوى 0.01 وجدنا أن معامل الصدق معنويًا لهذا الاختبار.

مثال (3): للصدق التبئري

لقد عرضت إحدى الباحثات اختبار لقياس مرونة مفصل الكتفين على عينة من طالبات كلية التربية الرياضية وقد حقق هذا الاختبار معنوية عالية من الثبات 0.799

وارادت الباحثة في نفس الوقت ايجاد معامل الصدق التنبؤي لهذا الاختبار حيث تنبأت بأن الطالبات اللاتي حصلن على نتائج جيدة في الاختبار سيخظبن بتفوييم جيد من الخبراء أو عند تأديتهم بعض التمارينات والتي تعتمد على مرونة مفصل الكتفين وعلى ذلك قامت الباحثة بما يلي :

- سجلت الباحثة نتائج الاختبار بالنسبة للطالبات (العينة المختارة).
- قامت نفس المجموعة بأداء حركة على مفصل الكتفين وقوّمت هذه الحركة من قبل بعض الخبراء المختصين.
- حسبت الباحثة المتوسط الحسابي لتفوييم الخبراء.
- أجرت الباحثة معامل الارتباط بين نتائج الاختبار وتفوييم الخبراء . وذلك كما يلي :

ف2	ف	الرتب ص	نتيجة تقويم الخبراء	الرتب س	نتيجة مرونة الكتفين	عدد الطالبات
4	2+	5	7.73	7	16	1
صفر	صفر	3	8.04	3	4	2
1	1+	8	7.21	9	19	3
1	1+	4	7.75	5	11	4
25	5-	9	6.75	4	6	5
صفر	صفر	10	4.53	10	27	6
25	5+	1	9.08	6	14	7
1	1-	2	8.28	1	30	8
صفر	صفر	11	4.28	11	59	9
16	4-	6	7.27	2	1	10
1	1+	7	7.23	8	17	11
74	10+	66		66		11 مج
	10-					

$$R_t = \frac{6 \text{ مجف}^2}{n(n^2 - 1)} - 1$$

$$R_t = \frac{\frac{444}{1320} - 1}{\frac{74 \times 6}{120 \times 11}} - 1$$

$$R_t = 0.66 = 0.34 - 1$$

ويحساب دلالة (11-2) لهذا الاختبار وجد أنه دال تحت المستوى 0.01 وأنه بذلك يتصنف بالصدق التنبؤي .

مثال (4): لصدق المفترى

وضع أحد الباحثين بطارية اختبارات لتقويم مستوى التكتيك لرمي القرص لعشرة لاعبين يقوم بتدريبهم وللتتأكد من صدق محتوى بطارية الاختبارات التي وضعها استعان بآراء بعض الخبراء في تقويمهم للتكتيك الرمي وبذلك يمكن ايجاد معامل الارتباط بين تقويم الخبرير للتكتيك ومتوسط آراء الخبراء وتقويمهم للتكتيك اللاعبين والذي يدل على صدق محتوى البطارية بعرض تقويم التكتيك .

والجدول التالي مبين فيه الدرجات الحاصل عليها هؤلاء اللاعبين من قبل الباحث (س) وكذلك الدرجات الحاصلين عليها من قبل الخبراء (ص)

أوجد معامل صدق المحتوى للعناصر التي شملتها البطارية لقياس مستوى التكتيك .

اللاعبين مائة (س)	درجة تقويم الخبير من الخبراء من 25 ص	معدل تقويم	الرتبة س	الرتبة ص	ف	ف2
محمد	15	10	4	7	3-	9.00
علي	60	11	5	4.5	0.5+	.25
ابراهيم	85	9	2	9	7-	49.00
محمود	90	10	1	7	6-	36.00
قيس	55	10	6.5	7	0.5-	.25
ناجي	70	12	3	2.5	0.5+	.25
سعيد	45	12	9	2.5	6.5+	42.25
محسن	35	11	10	4.5	5.5+	30.25
عبدالي	55	13	6.5	1	5.5+	30.25
علاوي	50	2	8	10	2-	40.00
مج			55	55	18.5- 18.5+	201.5

$$R_t = \frac{6 \times \text{مجم}^2}{n(n^2 - 1)}$$

$$R_t = \frac{1209}{990} - 1 = \frac{201.50 \times 6}{(1 - 100) \times 10}$$

وحساب دالة معامل الارتباط وباستخدام جدول دالة معاملات الارتباط الاحصائية أمام درجة الحرية والتي تساوي $n - 2$ وتحت النسبتين 0.05، 0.01 وجد أن

القيمة غير دالة عند المستويين 0.05، 0.01 أي أن بطارية الاختبارات التي اعتمد عليها الباحث في تقويم التكنيك غير صادقة المحتوى.

مثال (5) : للصدق الظاهري

يقوم مدرب بتدريب عشرة لاعبين وثبت عالي واجرى لهم اختبار لقياس قدرة الوثب وهو عبارة عن حساب الفرق في الوثب للعلى وذلك عن طريق حزام لانكوف. ثم قارن ذلك بنتائج الوثب العالى في احدى البطولات اللاحقة.

وكان المعطيات كما يلى :

ف ₂	ف	رتب ص	رتب س	مستوى اللاعبين		عدد اللاعبين
				بالوثب ع سم	بالحزام	
1	1-	4.5	3.5	195	200	1
صفر	صفر	1	1	205	210	2
6.25	2.5	2.5	5	200	195	3
20.25	4.5-	8	3.5	180	200	4
16	4-	6	2	190	205	5
30.25	2.5	2.5	8	200	180	6
9	3-	9	6	175	190	7
9	3-	10	7	170	185	8
4	2	7	9	185	175	9
1	1-	11	10	165	170	10
42.25	6.5	4.5	11	195	165	11
139	16.5+ 16.5-	66	66			مج 11

$$R_t = 1 - \frac{6 \text{ مجف}^2}{n(n^2 - 1)}$$

$$0.22 = \frac{834}{1320} - 1 = \frac{139 \times 6}{(1 - 121) \times 11} - 1$$

$$0.63 = \frac{486}{1320} - 1 = \frac{486}{1320} - 1$$

وبالكشف عن درجة حرية (11-2) وتحت المستوى 0.05، 0.01 وجد أن الاختبار دال معنوياً أي أن الاختبار صادق ظاهرياً مع هؤلاء اللاعبين.

1-5 العوامل المؤثرة في الصدق:

تلخص أهم العوامل التي تؤثر على الصدق فيما يلي:

1-5-1 طول الاختبار:

إن عدد عبارات الاختبار يؤثر في صدقه وكلما زاد عدد العبارات إرتفع معامل الصدق وذلك لأن أية ظاهرة كالقلق مثلاً تتتألف من مكونات أساسية وهذه المكونات بدورها تتتألف من خصائص وصفات كثيرة ويفترض من الناحية النظرية أن يكون المقياس الذي نبنيه لقياس هذه الظاهرة (القلق مثلاً) محتواً على كل تلك الخصائص والصفات مصاغة على شكل (فقرات). وبما أن هذا غير ممكن من الناحية العلمية لأن الاختبار في هذه الحالة يكون طويلاً جداً تصعب الإجابة عليه لذلك تؤخذ عينة من هذه الفقرات وبناء على ذلك فإن أي اختبار ليس الا عينة من الفقرات التي تمثل مجالاً واسعاً يحتوي على عدد كبير من الفقرات بمجموعها تمثل الظاهرة التي تُريد قياسها لذلك يعمد واضعوا

الاختبارات إلى زيادة عدد الفقرات عندما يتبيّن لهم انخفاض معامل الصدق.

يزداد صدق الاختبار تبعاً لزيادة عدد أسئلته وبذلك يزداد معامل ارتباط الميزان، وترتفع القيمة العددية لمعامل صدق الاختبار. هذا وبما أن الصدق يعتمد على الثبات وبما أن الثبات يعتمد على طول الاختبار، إذن فالصدق أيضاً يعتمد على هذا الطول كما يدل على ذلك المعادلة التالية:

$$R_{(n,s)} = \frac{R_s}{\sqrt{\frac{1 - R_s^2}{n + R_s}}}$$

حيث يدل الرمز (n, s) على معامل ارتباط الاختبار (s) بالميزان (n) وذلك عندما يزداد الاختبار (n) من المرات، ويدل الرمز R_s على معامل ارتباط الاختبار (s) بالميزان (n) قبل الزيادة.

ويدل الرمز R_s على معامل ثبات الاختبار s . ويدل الرمز (n) على عدد المرات التي يزداد بها طول الاختبار.

فإذا كان معامل صدق الاختبار قبل الزيادة $R_s = 0.6$

وكان معامل ثبات الاختبار $R_s = 0.8$

ثم زاد طول الاختبار لاربع امثلة $n = 4$

اذن فالزيادة في الصدق تحسب بالتعويض في المعادلة السابقة.

$$\therefore R_{(4,s)} = \frac{0.6}{\sqrt{\frac{0.8 - 1}{0.8 + \frac{4}{4}}}}$$

$$\frac{0.6}{\frac{0.85}{1}} = \frac{0.6}{\frac{0.2}{0.8 + \frac{4}{4}}} =$$

$$\therefore r(4s) = 0.65$$

أي أن القيمة العددية لمعامل صدق الاختبار يرتفع من 0.6 إلى 0.65 عندما يزداد طول هذا الاختبار الى اربع امثاله.

وبنفس هذه الطريقة يمكن أن نحسب زيادة الصدق تبعاً لأي زيادة في طول الاختبار وبذلك تتغير القيم العددية لمعامل الصدق تبعاً للتغيير قيم ن أي تبعاً للتغيير طول الاختبار.

1-5-2 ثبات الاختبار.

يؤثر ثبات المقياس في صدقه فانخفاض معامل الثبات يكون دليلاً على وجود عيب في الاختبار فيقدم بذلك مؤشراً على انخفاض صدقه وتجدر الاشارة في هذا المجال الى أن معامل الثبات العالي لا يكون دائماً دليلاً على صدق الاختبار حيث أنه من الممكن أن يكون الثبات عالياً والصدق واطئاً في الوقت نفسه.

يتأثر الصدق بالقيمة العددية لمعامل ثبات الاختبار تأثيراً مباشراً مطرداً، فيزداد الصدق تبعاً لزيادة الثبات لكن الثبات يتأثر ايضاً بطول الاختبار تأثراً مباشراً مطرداً، ولذا يزداد الصدق تبعاً لزيادة طول الاختبار كما سبق أن بيننا ذلك في تحليلنا لأثر إطالة الاختبار على الصدق، ويصل هذا الثبات الى أقصاه عندما يصل طول الاختبار إلى مانهاية. ويمكن أن نحسب صدق الاختبار لهذه الحالة التي تدل على الحد العلوي للثبات المفروض بالزيادة اللانهائية لطوله وذلك بالتعريض عن قيمة (ن) التي أصبحت تساوي ما لا نهاية

في معادلة إطالة الاختبار وذلك بالطريقة التالية:

$$\frac{\text{رس س}}{\frac{1 - \text{رس ص}}{\text{رس ص} + \sqrt{\frac{n}{}}}}$$

لكن $n = \infty$ مالا نهاية

$$\frac{\text{رس س}}{\frac{\text{رس س}}{\infty} + \frac{1 - \text{رس س}}{\infty}} = \text{رس } (\infty \text{ س}) \text{ ص}$$

لكن: $\frac{1 - رس ص}{\infty} = 0$ لأن نتيجة قسمة أي عدد على مالانهاية تساوي صفرأ.

$$\frac{\text{رس ص}}{\text{رس س}} = \text{ر}(\infty \text{س}) \text{ ص}$$

حيث يدل الرمز (S) ص على القيمة التنبؤية لمعامل الصدق عندما يصل طول الاختبار إلى ما لا نهاية ويدل الرمز S ص على معامل صدق الاختبار الأصلي أو التجربى . ويدل الرمز R ص على معامل ثبات الاختبار الأصلي أو التجربى .

فاما كان رس ص = 0.60

وکان رس س = 0.81

$$0.67 = \frac{0.60}{0.90} = \frac{0.60}{\underline{\underline{0.81}}} \therefore r(\infty)_{ص}$$

اذن القيمة التنبؤية للصدق عندما يصل طول الاختبار إلى ما لا نهاية تساوي 0.67 في مثالنا هذا.

فإذا فرضنا أن هذه القيمة التنبؤية تأثرت أيضاً بالعوامل الأخرى المساعدة في زيادة الصدق تأثراً يرتفع بكل عامل من تلك العوامل إلى صورته المثلثي، فإن هذه القيمة تساوي الواحد الصحيح، أي الارتباط التام الموجب.

$\therefore r(s)_{ص} = 1$ في هذه الحالة

$$\text{لكن } r(50s)_{ص} = \frac{رس_{ص}}{\sqrt{رس_{س}}} =$$

$$\frac{رس_{ص}}{\sqrt{رس_{س}}} = 1 \quad \therefore$$

$$\therefore r(s)_{ص} = \sqrt{رس_{س}}$$

أي أن صدق هذه الحالة المثالية يساوي الجذر التربيعي لمعامل ثبات الاختبار، وربما أن هذه الحالة حالة فرضية لا تقترب في الأغلب والاعم بالتطبيقات التجريبية، لذلك لا يتحمل أن تساوي قيمة الصدق التجاريقي قيمة الجذر التربيعي لمعامل الثبات إلا في النادر الشاذ الذي يرجع إلى الأخطاء التجريبية أكثر مما يرجع إلى التائج الصحيحة العلمية. إذن فالحد العلوي أو النهاية العظمى للصدق لا يمكن أن تزيد في هذه الحالة عن الجذر التربيعي لمعامل ثبات الاختبار.

3-5-1 عينة الأفراد

عندما يكون أفراد العينة متجانسين في الخاصية التي يقيسها الاختبار فإن ذلك يؤدي

إلى انخفاض معامل الصدق وعليه فإنه من الضروري أن يكون هنالك تباين بين أفراد العينة في الظاهرة المقاسة.

4-5-1 التباين

سبق أن بينا مدى تأثير معامل ثبات الاختبار بالانحراف المعياري للدرجات أو تباين تلك الدرجات ، لكن الثبات في جوهره معامل ارتباط . وهكذا ندرك أثر زيادة أو نقصان الفروق الفردية على معاملات الارتباط المختلفة . وبما أن الصدق صورة من صور الارتباط القائم بين الاختبار والميزان ، إذن فالصدق أيضاً يتاثر بتلك الفروق الفردية .

وهكذا نرى أن التباين الضعيف يقلل من أثر هذا الصدق ، وأن التباين القوي يزيد من القيمة العددية لذلك الارتباط .

يصل الصدق إلى نهايته الصغرى عندما يصل تباين الاختبار والميزان الى النهاية الصغرى أيضاً ، أي عندما تزول الفروق القائمة بين الأفراد في درجات الاختبار ودرجات الميزان . ويحصل في بعض الأحيان أن يكون للمعيار في الصدق التجريبي علاقة واضحة بالاختبار الذي وضعه الباحث ولكن عدم الدقة في ضبط العوامل الدخيلة بين الاختبار والمعيار يؤثر على معامل الصدق .

فالصعوبة أو السهولة المتأتية للاختبار أو عدم وجود رغبة لدى الطلبة في الاجابة على الاختبار أو ضعف الفقرات في قياس الصفة السلوكية كل ذلك يعتبر من العوامل الدخيلة التي تؤثر كثيراً أو تعمل على خفض معامل الصدق ولكي يمكن تفسير معامل الصدق تفسيراً دقيقاً وخاصة عندما يكون المعامل منخفضاً فإنه من الضروري أن تؤخذ بنظر الاعتبار كل العوامل التي أشرنا إليها .



الباب الثاني

الثبات

- 2-1 معامل ثبات الاختبار
- 2-2 مفهوم الثبات
- 2-3 القياس والثبات
- 2-4 معنى الثبات
- 2-5 الثبات والدلالة الاحصائية
- 2-6 الطرق الاحصائية لقياس الثبات
 - 1-6-2 طريقة اعادة الاختبار
 - 1-6-2 طريقة بيرسون
 - 1-6-2 طريقة سبيرمان
 - 2-6-2 طريقة التجزئة النصفية
 - 2-6-2 طريقة الزوجي والفردي
 - 2-6-2 طريقة جزء الاختبار
- 2-6-2 المعاملات الاحصائية لطريقة التجزئة النصفية.
- 3-2-6-2 معادلة سبيرمان وبراؤن.
- 3-2-6-2 معادلة رولون
- 3-2-6-2 معادلة جثمان
- 3-2-6-2 معادلة جلكسون
- 3-6-2 طريقة تحليل التباين
- 4-6-2 طريقة الاختبارات المتكافئة
- 5-6-2 طريقة المزج بين طرفيتي الاختبارات المتكافئة واعادة الاختبار.
- 6-6-2 طريقة اتفاق الملاحظين.
- 7-2 ثبات درجات المصححين
- 8-2 شروط الحصول على ثبات عالي.
 - 1-8-2 عدد فقرات الاختبار.
 - 2-8-2 تجانس فقرات الاختبار.
 - 3-8-2 صعوبة فقرات الاختبار.
- 9-2 العوامل المؤثرة في الثبات.
- 10-2 العلاقة بين الصدق والثبات.
- 11-2 أساس تحديد نوعية الفقرات المستخدمة بالاختبار.
- 12-2 إعداد تعليمات الاختبار.



الباب الثاني

الثبات Reliability

يعد مفهوم الثبات من المفاهيم الجوهرية في القياس ويتمثل مع الصدق أهم الاسس ويتغير توافرها في القياس لكي يكون صالحًا للاستخدام ومفهوم الثبات اشمل من مفهوم الصدق بمعنى اننا نستطيع أن نقول أن كل اختبار صادق ثابت بالضرورة ولكن ليس كل اختبار ثابت صادق بالضرورة وذلك بالطريقة نفسها التي نقول بها أن كل عراقي اسيوي ولكن ليس كل اسيوي عراقياً بالضرورة فإذا كان الاختبار صادقاً اي يقيس ما أعد لقياسه بالفعل فإن الدرجة عليه ستكون معبرة عن الاداء الحقيقي أو القدرة أو الوظيفة العملية كائنة ما تكون هذه الوظيفة أو الفترة ، وما دامت الدرجة على القياس الصادق تعتبر عن هذه الوظيفة بدقة فإنها تكون في الوقت نفسه ثابتة أي منسقة في تعبيرها عن مقدار الوظيفة ومستقرة عبر الزمن في تعبيرها وتقديرها لهذه الوظيفة غير أن كون الثبات مفهوماً أشمل من الصدق لا يعني انه يمكن الاكتفاء بتقديرات الصدفة الخاصة باختبار معين باعتبارها تتضمن بالضرورة تقدير ثباته فحتى في المرحلة الراهنة من التقدم العلمي لا تتوفر مقاييس ذات صدق تام كما إن صدق القياس يخضع لاعتبارات عملية متعددة تختلف من حالة إلى أخرى ومن مجتمع إلى آخر وبالتالي يتغير حسم وتقدير كل من صدق وثبات المقاييس باستقلال كل منها عن الآخر مثل استخلاص نتائج تجريبية منها تتعلق بمجتمع معين أو عينة معينة .

يعرف الثبات بأنه الاتساق في النتائج ويعتبر الاختبار ثابتاً إذا حصلنا منه على نفس النتائج لدى إعادة تطبيقه على نفس الأفراد وفي ظل نفس الظروف فلو افترضنا أن أحد

الباحثين يريد القيام ببحث لمعرفة متوسط اعمار المعلمين في العراق ، فإنه يعد استماره لهذا الغرض ويختار عينة تمثل المعلمين ، ويزع علىها تلك الاستمارة ، ويطلب منهم أن يدونوا فيها تاريخ الولادة ثم يقوم بعد ذلك بحساب معدل العمر . ولكي يتأكد الباحث من ثبات اجابات المعلمين ، فإنه يقوم بارسال الاستمارة اليهم مرة اخرى بعد فترة من الزمن ويطلب منهم أن يدونوا نفس المعلومات التي طلبها منهم في المرة الأولى وبذلك يصبح لدى الباحث اجابتان لكل معلم فإذا اتفقت إجابات المعلمين في المرة الأولى مع اجاباتهم في المرة الثانية فإن ذلك يدل على ثبات اجاباتهم على تلك الاستمارة .

ويستخدم في حساب معامل الثبات عادة معامل الارتباط بين الدرجات وهو يستخرج من ايجاد علاقة الارتباط بالدرجات التي حصل عليها الافراد في المرة الاولى والدرجات التي حصلوا عليها في المرة الثانية ويطلق على النتيجة التي نحصل عليها مصطلح «معامل الثبات» ويتراوح معامل الثبات بين الدرجتين (صفر - 1) ويعتبر الصفر أدنى معامل الثبات اما الدرجة (1) فتمثل اعلى معامل للثبات .

2-1 معامل ثبات الاختبار

توجد بعض المرادفات الانجليزية الاخرى لمصطلح ثبات الاختبار Reliability of test والتي نجدها في بعض المراجع C الانجليزية وهي Dependabilitit أو Stability of test وكل هذه المصطلحات ما هي الا تعبير عن معنى الثبات والوثق والاعتماد على نتائج الاختبار المستخدم ، وقد يجد البعض تداخلاً في المعنى بين ثبات الاختبار وصدق الاختبار الامر الذي يدعونا إلى توضيحه قبل أن نتكلم عن ثبات الاختبار فصدق الاختبار يعني كما قلنا أن يكون الاختبار والقياس أميناً وصادقاً فيما يدعى أن يقيسه ، اما ثبات الاختبار فيعني أن تكون نتائج الاختبار الواحد اذا ما كررت اكثر من مرة فتكون متشابهة اي لا توجد فروق كبيرة بينها اي فروق «غير معنوية» ويتوقف ثبات الاختبار على مدى هذه الفروق والتي تعمل على تقليلها وتلافيها دائماً ، وعلى ذلك فإن اجهزة القياس المستخدمة تعتبر اجهزة

ثابة Reliable حيث أن معطياتها لا تتغير عندما تكرر نفس عملية القياس ومن أمثلة ذلك اجهزة الديناموميتر المختلفة لقياس القوة العضلية العظمى وجهاز الاسببروميتر لقياس السعة الرئوية وضغط الدم . ومن هذه المقاييس المقتنة وبالمثل يجب أن يكون ذلك عند القياس النفسي والسوسومترى والتي يجب أن لا يتغير معطياتها عندما تكرر نفس عملية القياس ، قد يستخدم بعض مدرسي التربية البدنية بعض الاختبارات لقياس قدرات طلابهم البدنية أو المهارية أو النفسية دون التتحقق من ثبات نتائج هذه الاختبارات معللين بأن هذه الاختبارات ليست بالجديدة في تطبيقها أو استخدامها . فقد استخدمت من قبل آخرين على مجموعات مماثلة وطبقت منذ سنوات طويلة ولا داعي للتأكد من ثباتها مرة أخرى ، هذا امر قد تكون نتائجه وخيمة حيث أن ثبات الاختبار يجب التأكد منه على نفس العينة المختارة وذلك بإجراء تجربة استطلاعية تمثله تمثيلاً صادقاً للمجتمع .

2-2 مفهوم الثبات :

إن الأدوات المختلفة المستخدمة في القياس تمثل مرحلة متطرفة وليست نهائية من مراحل التقرير المتالي نحو الحقيقة العلمية ونتيجة لعدم التحكم التام والضبط الدقيق لمواصف القياس والتي تسرب خلالها بعض التغيرات الداخلية المؤثرة في الظاهرة نتيجة لكل هذا تصبح الدرجة التي نخرج بها من القياس معبرة عن الظاهرة التي نقيسها سواء أكانت سمة أو قدرة أو غيرها وتتضمن الدرجة دائماً قدرأً من الخطأ سواء أكان خطأ موجباً في شكل زيادة في الدرجة عما يتحقق الشخص نتيجة لقدرته الحقيقة أو نقصاناً في هذه الدرجة يمثل تقييماً لاداء الفرد أقل من الواقع .

معنى هذا أن الدرجة المستخلصة من اي اختبار لا تعبر تماماً عن الاداء الحقيقي أو التباين الحقيقي للفرد نتيجة لتسرب هذه الشوائب الداخلية الناجمة عن عدم أحکام ودقة القياس أو عدم ضبط الظروف سواء أكانت موجبة أو سالبة أي في شكل زيادة في الدرجة أو نقصان فإن هذه الشوائب هي التي نطلق عليها اسم تباين الخطأ . وحتى نتمكن من

التعرف على الدرجة الحقيقة أو التباين الحقيقي للمقياس فلا بد من حساب ثبات المقياس ويشير مفهوم الثبات إلى نسبة التباين الحقيقي في الدرجة المستخلصة من اختبارها.

الثبات إذن هو النسبة من تباين الدرجة على المقياس التي تشير إلى الأداء الحقيقي للمفحوص، ويعرف بأنه الاتساق في النتائج ويعتبر الاختبار ثابتاً إذا حصلنا منه على نفس النتائج لدى إعادة تطبيقه على نفس الأفراد وفي ظل نفس الظروف.

ثبات الاختبار يعني انه لو اعيد تطبيق الاختبار على الافراد انفسهم فإنه يعطي النتائج نفسها أو نتائج متقاربة وهذا يشير إلى أن الثبات يعني اتساق نتائج الاختبار مع نفسها لو كررت مرة أو عدة مرات اخرى ويمكن أن يقصد بالثبات الاستقرار اي انه لو اعيد تطبيق نفسه على الفرد الواحد فإنه يعطي شيئاً من الاستقرار في النتائج . ومعامل ثبات الاختبار هو معامل ارتباط بين نتائج المرات المختلفة لاجرائه اي بين الاختبار نفسه .

والثبات نسبي ، ذلك انه قد يوجد اختلاف بين الفرد ونفسه في المرات المختلفة لاجراء الاختبار فالفرد نفسه يتغير وهذا يؤدي إلى اختلاف في النتيجة بين تطبيق وأخر للاختبار.

أن ثبات الاختبار يعتبر واحد من اهم المعاملات للوثوق بنتائجها، ويمكن تعريف ثبات الاختبار انه اذا ما أعيد الاختبار مرة او مرات اخرى على نفس العينة أو على عينات اخرى بنفس المواصفات وتحت نفس الظروف يعطي نتائج معنوية اي وجود معامل ارتباط كبير بين نتائج الاختبار في كل مرة يجري فيها، ويرى مبيرز Meyers أنه عندما تتوفر في اي اختبار الشروط التالية فإنه ينصف بدرجة عالية من الثبات:

الشرط الأول : أن يعطي الاختبار نفس النتائج عند اعادته مره أو مرات لاحقة.

الشرط الثاني: أن يكون الاختبار مناسباً لعينة المختبرين.

الشرط الثالث : اختبارات أو أكثر لقياس نفس الغرض تعطي نفس النتائج.

أما أرنولدج لين Arnold J. lien فيرى أن هناك ثلاث عوامل رئيسية تؤثر على معامل ثبات الاختبار في مجال «التعليم» مع مراعاتها بحذر هي :

العامل الأول : موضوعية الاختبار فكلما كان الاختبار موضوعياً كلما زاد من ثباته.

العامل الثاني : عوامل تتعلق بالقائم على الاختبار مثل :

- اعداد المكان و المناسبة .

- الدافعية للاختبار .

- قلة التوجيهات و سهولة الاختبار .

- دقة الاشراف على الاختبار .

العامل الثالث : بالنسبة للمختبرين :

- مستوى المختبرين كلما كان المستوى مرتفعاً و متجانساً كلما ادى ذلك إلى ثبات الاختبار .

- كلما كان المختبرين بدون مشاكل اجتماعية و نفسية الخ ، كلما ادى إلى ثبات الاختبار .

- كلما كانت هناك علاقة وطيدة بين المختبر (المعلم) و عينة المختبرين (الطلاب) كلما ادى ذلك إلى ثبات الاختبار .

- كلما كانت هناك الرغبة إلى التعلم من قبل المختبرين (الطلاب) كلما ادى ذلك إلى ثبات الاختبار .

عند اختبار قدرة عضلات الساقين (القوة المميزة بالسرعة للعضلات المادة للساقين) يمكن استخدام اختبار الورب العريض من الثبات - وبذلك يسجل لمجموعة المختبرين المسافة مرتين متتاليتين مع إفتراض وجود فترة راحة كافية بين كل محاولة و اخرى لنفس

الشخص المختبر ثم يحسب المعدل الوسطي لمجموعة المختبرين في كل مرة ، ويحسب على ضوء ذلك معامل الارتباط بين نتائج المرتين والذي يدل على معامل ثبات الاختبار وذلك إذا كان الارتباط معنواً .

إن تطبيق مثل هذه الطريقة على الاختبارات النظرية والتي تعتمد على تذكر مجموعة المختبرين بنتائج اجاباتهم في المرة الأولى في الاختبارات ، قد تتأثر نتائجهم التالية بما استفادوا من نتائج المرة الأولى ، وعلى ذلك يتضح في مثل هذه الحالة أن يكون هناك وقت كاف للفصل بين المرتين والذي يتضمن أن ينسى أن اختبارات أخرى لا تتطلب المغالاة في طول هذه الفترة الزمنية وذلك عند اختبار القوة العضلية مثلاً . وحتى لا يستفيد مجموعة المختبرين من تلك الفترة من اكتسابهم غوا خاصا يؤثر على نتائجهم بسبب طول هذه الفترة . وعلى ذلك نعرض فيما يلي بعض العوامل الهامة إلى جانب ما عرضناه من أداء بعض الباحثين والتي تؤثر في معامل ثبات الاختبار .

أولاً : تعدد مفردات الاختبار وعدم تجانسها :

إن الابتعاد عن تعدد الموضوعات ، التي يشملها الاختبار وعدم تجانسها ، يعطي فرصة لمجموعة المختبرين بالابتعاد عن عامل الصدفة والتتخمين في اجاباتهم ويفسر ذلك بوضوح في القياس والاختبارات النظرية والتي تعتمد على اجابات الاسئلة ، مع ملاحظة عدم المغالاة في الاطالة حتى لا تتأثر الاجابات بظاهرة التعب والملل . فعند وضع اختبار الذكاء على سبيل المثال لمجموعة من المختبرين يجب أن تكون الاسئلة المكونة للاختبار بالقدر الكافي محددة وبحيث يكون نصيب الصدفة أو التخمين من جانب المختبرين شيء لا يذكر ، هذا من جهة ، ومن جهة أخرى عدم المبالغة في كثرتها حتى تتجنب ظاهرة التعب أو الملل (الا اذا رأى واضع الاختبار سبباً لذلك) وحتى لا يؤثر الملل أو التعب تأثيراً سلبياً على نتائج اجابات هذه الاسئلة وبذلك لا تتصف بالثبات في نتائجها .

ثانياً : الانسجام بين مستوى الاختبار ومستوى المختبرين :

يجب على واضح الاختبارات أن يختار القياس المناسب لمستوى مجموعة المختبرين سواء كانت هذه الاختبارات بدنية أو مهارية أو نفسية فإذا وضع الاختبار الغير مناسب لعينة المختبرين وذلك بالنسبة لل اختبارات العملية الميدانية ، تظهر بذلك نتائج غير متوقعة وغير حقيقة حيث نجد الفرق بين نتائج هذه القياسات لفرد الواحد متباعدة اذا ما اعيدت مرة أو مرات تالية فيقل بذلك ثبات الاختبار ويكون غير معنوي . أما في الاختبارات النظرية والتي تعتمد على اجابات الاسئلة فيكثر فيها التخمينات نتيجة لصعوبة الاسئلة والتي لا تناسب مستوى المختبرين فإذا اعيدت هذه الاختبارات على نفس الفرد نجد ثباتاً كبيراً بين نتائج الاختبارات الاولى والثانية والتي تشير اليها دائمًا بعدم وجود ارتباط بين النتيجتين مما يقلل من معنوية معامل ثبات الاختبار .

ثالثاً : العوامل في مستوى المجموعة المختبرة :

عند إجراء اي اختبار على عينة من المختبرين في أي مدرسة من المدارس يجب أن لا يكون بين أفراد هذه العينة تبايناً واضحاً من جميع الوجوه وعلى ذلك يجب أن تصنف هذه العينة حسب بعض العناصر الهامة التي يراها واضح الاختبار واجبة ومتطلبة لصحة نتائج الاختبار وذلك التصنيف أما أن يكون حسب اعمارهم أو صفو فهم الدراسية أو مستوى اهم البدنى أو المهارى ، وعندما يقتضي الأمر يمكن أن تصنف عينة المختبرين حسب هذه العناصر سالفة الذكر مجتمعة وهنا تعمل عملية التصنيف على تدريب التباين الكبير بين المختبرين بالإضافة إلى أهميتها التربوية الكبيرة .

إن عملية التصنيف لا تتوقف على هذه العناصر فقط بل تتعداها الى التصنيف بالنسبة للطول أو الوزن وذلك في بعض الاختبارات والقياس البدنى .. وحتى التصنيف الجغرافي بالنسبة للمسكن أو التصنيف الاجتماعي الخ ، (وسوف تتكلم عن عملية التصنيف بأسهاب كما ذكرنا) أن التباين في العمر أو المستوى الخ ، الموجود بين المختبرين

يؤثر تأثيراً سلبياً على نتائج ثبات اختباراتهم اذا ما اعيدت نفس هذه الاختبارات عليهم مرة أخرى وبذلك نحصل على نتائج غير معنوية والمثال التالي يبين لنا هذه الحقيقة .

أراد احد مدرسي التربية البدنية في احدى المدارس المتوسطة تقويم عنصر السرعة لطلاب مدرسته ، فبادئ ذي بدء وضع اختبار السرعة المناسب لمستوى وقدرات هذه المرحلة من العمر (عدو 3 متر من البدء الطائر) وبدأ في تنفيذ الاختبار دون إجراء عملية التصنيف ثم اوجد المعدل الوسطي للسرعة ، وبعد تكرار الاختبار مرة ثانية وتحت نفس الظروف الاولى حصل على معدل وسطي بعيد إلى حد ما عن المعدل الوسطي السابق وباختبار نتيجة معدلى السرعة للاختبارين احصائياً وجد انها غير مرتبطة ، هذا دليل على عدم معنوية ثبات الاختبار على هذه المجموعة ، وبعد اسبوع من اجراء المحاولة الاولى لقياس سرعة الطلاب اجرى نفس الاختبار بعد تصنیف الطلاب حسب صفوهم واعمارهم ومستوياتهم وسجل المعدل الوسطي لسرعتهم وبأعادة نفس الاختبار تحت نفس الظروف السابقة حصل على معدل وسطي اقرب ما يكون إلى القياس السابق اي بمعنى عالية من الثبات .

2-3 القياس والثبات :

إن الاختبارات وسائل لقياس النواحي المختلفة كما يقيس المتر النواحي الطولية والكيلو النواحي الوزنية والساعة النواحي الزمنية وتعتمد صحة القياس على مدى ثبات نتائجه وصدقها . فالمقياس الثابت يعطي نفس النتائج اذا قاس نفس الشيء مرات متتالية فإذا قسمت طول قطعة من القماش ودل القياس على أن طولها 1.5 مترأ ثم أعدنا عملية القياس ودللت النتائج للمرة الثانية على أن الطول يساوي 1.5 مترأ ! استنتجنا من ذلك أن نتائج هذا القياس ثابتة وبما أن المقياس المتر يقاس الاطوال ولا يقاس شيئاً آخر غير هذه الاطوال فهو إذن صادق فيما يقاس لانه يقاس الصفة التي يهدف إلى قياسها .

إذا قاس المتر صفة الوزن بدل قياسه لصفة الطول لم يصبح صادقاً في قياسه للطول

وصدق المقاييس المادية اوضح من أن يدرس علمياً لكن صدق المقاييس النفسية يحتاج إلى كثيراً من الدراسة والتحليل فقد لا ندرى مثلاً مدى صدق اختبارات الذكاء في قياسها لصفة الذكاء الا اذا أقمنا الدليل العلمي على صحة هذا الرعم وذلك حساب وتقدير تلك الاختبارات .

4- معنى الثبات :

اذا أجرى اختبار ما على مجموعة من الافراد ورصدت درجات كل فرد في هذا الاختبار ثم اعيد اجراء نفس هذا الاختبار على نفس هذه المجموعة ورصدت ايضاً درجات كل فرد ودللت النتائج على أن الدرجات التي حصل عليها الطالب في المرة الاولى لتطبيق الاختبار هي نفس الدرجات التي حصل عليها هؤلاء الطلبة في المرة الثانية . استنتاجنا من ذلك أن نتائج الاختبار ثابتة تماماً لأن نتائج القياس لم تتغير في المرة الثانية بل ظلت كما كانت قائمة في المرة الاولى . وخير طريقة لمقارنة هذه الدرجات هي حساب معامل ارتباط درجات الاختبار في المرة الاولى بدرجات هذا الاختبار في المرة الثانية وعندما ثبتت الدرجات فتصبح واحد في المرتين يصبح معامل الارتباط مساوياً للواحد الصحيح . لكن المقاييس النفسية لا تصل إلى هذه الدقة المثالية التي قد تقترب منها في قياسنا العلمي الصفات المادية المختلفة كالطول والوزن والزمن ولذا يقترب معامل ارتباط الاختبار بنفسه من الواحد الصحيح لكنه لا يساوي هذا الواحد الصحيح وينشأ هذا الفرق من الاخطاء المختلفة التي تصل من قريب أو بعيد بنتائج المقاييس النفسية والتي لا تخضع في جوهرها الضبط العلمي أو التحكم الدقيق في الظاهرة التي تخصصها للقياس ، وذلك لأن نتائج القياس تتأثر إلى حد ما بالحالة النفسية للفرد وبحالته الجسمية وبالتغيرات الجوية والاصوات المفاجئة ويعبرها من العوامل التي تؤثر بطريق مباشر في ثبات تلك النتائج .

وعندما نحسب معامل ارتباط الاختبار بنفسه ويحصل على قيمة عددية تدل على هذا الارتباط فإننا بذلك نحسب الجزء الثابت من هذا الاختبار اي الجزء الذي لا يتأثر بتلك

الامور الخارجية. وهكذا نستطيع أن نقسم درجة اي فرد في هذا الاختبار إلى جزئين جزء جوهري ثابت لا يتأثر بالعوامل الخارجية المختلفة وجزء لا يتأثر بهذه العوامل وبما أن هذا الجزء الاخير الذي لا يتأثر بالعوامل الخارجية يختلف تبعاً لاختلاف هذه العوامل اذن فهو لا يرتبط ببعضه في المرات المتتالية التي تجري فيها هذا الاختبار على نفس الفرد انه الجزء الخاطئ من الدرجة الذي يتلاشى ويختفي عندما تحسب معامل ارتباط الدرجات اي أن معامل ارتباط تلك الاجزاء الخاطئة يساوي صفرأ. أو بمعنى آخر الدرجة التجريبية = الدرجة الحقيقية + الدرجة الخاطئة

أي أن

$$\text{س ج} = \text{س ق} + \text{س خ}$$

حيث يدل الرمز س ج على الدرجة التجريبية التي نحصل عليها فعلاً عند إجراء الاختبار ويدل الرمز س ق على الدرجة الحقيقية التي نفترض ثباتها. ويدل الرمز س خ على الدرجة الخاطئة التي نفترض تغيرها. وعندما نعيد اجراء هذا الاختبار على نفس هذا الفرد فإن الدرجة التي يحصل عليها في المرة الثانية عن قيمتها في المرة الاولى وهكذا بالنسبة لنفس هذا الاختبار وبما أن هذا الانحراف يقاس بالانحراف المعياري وبحسب هذا الانحراف المعياري المسمى بالتبابين. اذن فتبابين الاختبار ينقسم إلى التبabin الحقيقي للدرجات والتي تبادل خطأ المقياس.

$$\text{تبابين درجات الاختبار} = \text{تبابين حقيقي للدرجات} + \text{تبابين الخطأ}$$

$$\therefore \text{ع}^2 \text{ ج} = \text{ع}^2 \text{ ق} + \text{ع}^2 \text{ خ}$$

حيث يدل الرمز ع² ج على التبabin التجاري للدرجات ويدل ع² ق على التبabin الحقيقي لهذه الدرجات ويدل الرمز ع² خ على تبabin الخطأ وهكذا يعرف الثبات: « انه الجزء الحقيقي من التبabin العام للاختبار وهذا الجزء الحقيقي هو الذي يعطينا القيمة العددية لارتباط الاختبار بنفسه »

- معامل الارتباط والثبات :

إن الثبات في حقيقة الأمر عبارة عن النسبة من التباين الكلي التي مصدرها الأداء الحقيقى على الاختبار اي المقدار الحقيقي من قدرة المفحوص دون زيادة ناتجة عن ظروف مواتية أو نقصان ناتج عن ظروف معوقة.

فإننا نتوصل إلى تقدير لهذا التباين الحقيقي بحساب الارتباط بين عيتيين من هذا الأداء على الاختبار سواء كانت هاتان العيتيان على الاختبار نفسه عبر المرة الثالثة والرابعة وغير ذلك من المرات المتالية :

$$\therefore \text{س ج } 1 = \text{س ق} + \text{س خ } 1$$

$$\text{س ج } 2 = \text{س ق} + \text{س خ } 2$$

$$\text{س ج } 3 = \text{س ق} + \text{س خ } 3$$

$$\text{س ج } 4 = \text{س ق} + \text{س خ } 4$$

وهكذا بالنسبة لأى عدد من المرات التي يجري فيها هذا الاختبار على نفس هذا الفرد كذلك بالنسبة لأى عدد من الافراد.

وبما أن معامل ارتباط الدرجة الخامسة س خ 1 بالدرجة الخامسة س خ 2 يساوي صفرأً اذن فالارتباط القائم بين س ج 1 ، س ج 2 يعتمد جوهره على س ق التي لم تتغير في المرتين اي أن الثبات يقيس الجزء الحقيقي من الدرجة التجريبية ولذا تعتمد فكرة الثبات على أن

$$\text{س لا تساوي ولا ترتبط بـ س خ } 2$$

$$\text{وان س خ لا تساوي ولا ترتبط بـ س خ } 2$$

وهكذا بالنسبة لبقية الدرجات المختلفة وعندما يقيس في المرتين التي يطبق فيها على نفس مجموعة الافراد فإنه ايضاً يقيس عدم ارتباط الاختبار بنفسه او بمعنى اخر يقيس

الاغتراب . وهكذا تعتمد فكرة الثبات على مدى انحراف درجة كل فرد في التطبيق الاول للاختبار عنها في التطبيق الثاني مرحلة زمنية محدودة أو في الوقت نفسه عبر صورتين مكافتين لاختبار وعند وصف الفقرات على الاختبار الواحد يصبح ضروري أن تربط بين مفهوم الثبات ومفهوم الارتباط ، فرغم اتنا نحسب ثبات الاختبار بوصفة واحدة الا اتنا نعالجه وفقاً لمفهوم الارتباط بوصفه متغيرين هما هاتان العيتان من الاداء مهما كان اسلوب سحب هاتان العيتين تنصيفاً أو تجزئة أو اعادة للاختبار . ومعامل الارتباط في اساسه مقياس للعلاقة بين متغيرين وهو يقدم تقديرأً لمدى اتساق التغيير في هذين المتغيرين وقوة هذه الاتساق .

فمن الناحية الاولى يقدم معامل الارتباط تقديرأً لما اذا كان هناك تلازم في التغيير ام لا ومقدار هذا التلازم أن وجد .

ويقدم من ناحية ثانية تقديرأً لاتجاه هذا التلازم في التغيير وما اذا كان ايجاباً بمعنى أن المتغيرين يتزايدان معاً ويتناقصان معاً او أن الزيادة في احدهما يناظران نقصان في الآخر وفي اي من هاتين الحالتين لا توجد فرض متعلقة بالعملية في العلاقة بين المتغيرين فلا تعني تزايدهما او تناقصهما معاً أن احدهما السبب والآخر نتيجة لتغير ثالث أو متغيرات اخرى تلعب دور السبب . وتتوفر لنا المعلومات الاحصائية الاولية كيف يمكن معالجة الدرجة على الاختبار وتمثلها على منحنى بياني ونستطيع بالاسلوب نفسه أن نستخدم هذه التمثيلات البيانية لايضاح العلاقة بين متغيرين بأن تخصص تقسيمات المحور السيني للمتغير الاول وتقسيمات المحور الصادي للمتغير الثاني وتقع النقطة بين المحورين المعبرة عن درجة فرد ما على بعد من نقطة الاصل على المحور السيني يناظر درجته على المتغير الاول ومرتفعة بمقدار معين مناظر لنقاط المحور الصادي تعبيراً عن الدرجة على المتغير الثاني . ولانا عرفنا ميزات استخدام الدرجات المعيارية فمن الافضل هنا أن نستخدم هذه الدرجات التي تقدم لنا تقديرأً للوحدات المعيارية التي يتعد بها كل فرد عن نقطة الصفر (اي المتوسط) على كل من هذين المتغيرين .

غير انتا لا تعبر عن الارتباط بين متغيرين باعتباره (يكون أو لا يكون) أو صفر بل هناك درجات من الارتباط تتراوح بين +1 - 0.

وتعتبر معادلة بيرسون أو معامل ارتباط الرتب صورة أخرى من هذه المعادلة الأساسية ونص معادلة بيرسون كالتالي:

$$r = \sqrt{\frac{\sum_{i=1}^n (M_{xi} - \bar{M}_x)(M_{yi} - \bar{M}_y)}{\sum_{i=1}^n (M_{xi})^2 - (\bar{M}_x)^2} \cdot \frac{\sum_{i=1}^n (M_{xi} - \bar{M}_x)(M_{yi} - \bar{M}_y)}{\sum_{i=1}^n (M_{yi})^2 - (\bar{M}_y)^2}}$$

ويلاحظ ضرورة اختبار الأسلوب المناسب لحساب الارتباط بين جزئي أو نصفي الاختبار فاستخدام معامل ارتباط بيرسون مقصور على البيانات المتصلة وليس المنفصلة ويستخدم معامل الارتباط في بعض اساليب الثبات أو المعدلة المناسبة لكل حالة على حدة.

وهناك نقطة اخيرة تتعلق باستخدام معامل الارتباط بوصفه تقديرًا للثبات الاداء ففي الاستخدامات المختلفة يتحتم حساب معامل الارتباط للثبت ما اذا كان المعامل الذي لوحظ بين المتغيرين ناجح عن الصدفة أم لا مستوى الثقة في هذا المعامل وما اذا كان هذا المستوى 0.95 أو 0.99، أما في حالة استخدام معامل الارتباط بوصفه معاملًا للثبات فلا يجوز حساب مستوى الثقة له أو احتمالية أن يكون الارتباط ناجح عن الصدفة فالمتوقع في الظروف المثالبة أن يكون الارتباط بين الاختبار ونفسه أو بين جزئي أو نصفي أو كل فقرات الاختبار 0.1 والشوائب والخطأ فقط هي التي تخفض من هذا المعامل فلا يصل إلى 1.5 وعلى ذلك لا يجوز الاشارة إلى مستوى الثقة أو مستوى الدلالة لمعاملات الثبات المختلفة.

2-5 الثبات والدلالـة الـاحصـائـية

ترتبط فكرة الثبات بفكرة الدلالـة الـاحصـائـية وذلك لأنـ الثبات يتأثر بالـاخـطـاء التجـريـيـة كما تـأثرـ بها ايـضاـ الدـالـلـة الـاحـصـائـية لـالمـقـايـيس الـمـخـتـلـفةـ . لكنـ الثـبـات يـدـلـ علىـ اـخـطـاءـ الـقـيـاسـ فيـ تـقـدـيرـهـ لـلـجـزـءـ الـحـقـيقـيـ الـثـابـتـ لـلـاـخـتـبـارـ . وـهـوـ لـهـذـاـ يـعـتـمـدـ فيـ نـتـائـجـ تـطـبـيقـ الـاـخـتـبـارـ اـكـثـرـ مـنـ مـرـةـ عـلـىـ نـفـسـ مـجـمـوعـةـ الـافـرـادـ ايـ انـ يـقـارـنـ مـدـىـ اـخـتـلـافـ نـتـائـجـ الـاـخـتـبـارـ فـيـ الـمـرـاتـ الـمـتـابـعـةـ ، فـهـوـ لـهـذـاـ يـرـتـبـطـ اـرـتـبـاطـاـ مـباـشـراـ بـخـطـأـ الـقـيـاسـ . وـتـقـيـسـ الدـالـلـةـ الـاحـصـائـيةـ خـطـأـ الـعـيـنـاتـ لـانـهـ تـعـمـدـ فـيـ جـوـهـرـهـ عـلـىـ مـقـارـنـةـ مـدـىـ اـخـتـلـافـ نـتـائـجـ الـقـيـاسـ بـالـنـسـبـةـ لـعـدـدـ كـبـيرـ مـنـ مـجـمـوعـاتـ الـافـرـادـ اوـ بـالـنـسـبـةـ لـعـيـنـاتـ كـثـيرـةـ مـنـ الـافـرـادـ لـتـقـيـسـ بـذـلـكـ مـدـىـ اـتـصـالـ هـذـهـ الـعـيـنـاتـ بـالـاـصـلـ الـتـيـ اـنـتـزـعـتـ مـنـهـ . وـبـذـلـكـ تـقـرـرـ الدـالـلـةـ الـاحـصـائـيةـ لـمـتوـسـطـ اـحـدـ الـعـيـنـاتـ الـخـطـأـ الـمـعـيـارـيـ لـهـذـاـ مـتـوـسـطـ وـمـدـىـ اـبـتـعـادـهـ اوـ اـقـرـابـهـ مـنـ مـتـوـسـطـ الـاـصـلـ الـذـيـ اـنـتـزـعـتـ مـنـهـ هـذـهـ الـعـيـنـةـ وـهـكـذـاـ بـالـنـسـبـةـ لـدـالـلـةـ الـمـقـايـيسـ الـاحـصـائـيةـ الـاـخـرـىـ .

2-6 الطرق الـاحـصـائـيةـ لـقـيـاسـ الثـبـاتـ

تـعـدـ اـسـالـيـبـ حـسـابـ الثـبـاتـ وـيـخـصـ كـلـ اـسـلـوبـ مـنـهـ بـتـقـدـيرـ نـوـعـيـةـ مـحـدـدـةـ مـنـ (ـتـبـاـيـنـ الـخـطـأـ)ـ وـهـوـ تـبـاـيـنـ الـذـيـ يـؤـثـرـ فـيـ ثـبـاتـ الـقـيـاسـ الـذـيـ نـحـصـلـ عـلـيـهـ كـلـمـاـ اـسـتـخـدـمـنـاـ مـقـايـيسـ مـخـتـلـفـةـ وـتـصـلـحـ بـعـضـ اـسـالـيـبـ لـحـسـابـ ثـبـاتـ مـقـايـيسـ مـعـيـنـةـ بـيـنـمـاـ لـتـصـلـحـ لـحـسـابـ ثـبـاتـ مـقـايـيسـ اـخـرـىـ وـلـاـ يـعـنـيـ هـذـاـ أـنـ الـقـيـاسـ الـواـحـدـ لـتـصـلـحـ لـهـ اـلـطـرـيـقـةـ وـاـحـدـةـ لـحـسـابـ ثـبـاتـ فـيـ بـعـضـ الـحـالـاتـ يـكـنـ أـنـ نـسـتـخـدـمـ اـكـثـرـ مـنـ اـسـلـوبـ لـلـاـخـتـبـارـ الـواـحـدـ بـهـدـفـ التـعـرـفـ عـلـىـ مـصـادـرـ (ـتـبـاـيـنـ الـخـطـأـ)ـ الـذـيـ يـؤـثـرـ فـيـ اـسـتـقـرـارـ اوـ اـتـسـاقـ الـدـرـجـةـ الـتـيـ نـحـصـلـ عـلـيـهـ مـنـهـ .

ويـتعـيـنـ دـائـماـ درـاسـةـ أـيـ اـسـالـيـبـ اـصـلـحـ لـحـسـابـ ثـبـاتـ الـاـخـتـبـارـ الـمـحـكـ لـكونـ اـسـلـوبـ مـاـ أـفـضـلـ مـنـ اـسـلـوبـ اـخـرـ هوـ أـنـ اـسـلـوبـ يـؤـدـيـ إـلـىـ حـصـولـ عـلـىـ اـعـلـىـ مـعـاـمـلـ ثـبـاتـ مـمـكـنـ .

فإذا كان لدينا ثلاثة أساليب أ. ب . ج يمكن استخدام كل منها لتقدير ثبات مقياس معين وكان «أ» يوفر معامل ثبات قدره 0.90 بينما يوفر «ب» معاملًا قدره 0.80 ويوفر «ج» معاملًا قدره 0.60 فإن أ هو أفضل اسلوب مناسب لتقدير ثبات هذا المقياس . ويعتمد هذا المحك على منطق أن الاسلوب الذي يؤدي إلى أعلى معامل ثبات إنما يقوم على تقدير لاقصى تباين حقيقي للأداء على الاختبار . والنقطة الهامة التي يجب أن تكون واضحة في الذهان هي أن الاختبار نفسه بوصفه مجموعة من البنود أو الأسئلة أو الاعمال المعينة لاثبات له .

إن الثبات هو ثبات الاجابة على الاختبار ثبات الاداء على الاختبار ولهذا فإن ما يحسب ثباته إنما هو عينة استجابات نحصل عليها من مجموعة من الافراد ولأن كل مجموعة من الافراد لها خصائصها فمن الضروري أن نحصل على أكثر من معامل ثبات للاختبار الواحد إلى أكثر من عينة ولا يعني حصولنا على معامل ثبات لاداء عينة من الراشدين على اختبار معين . أن ثبات اداء المراهقين أو الاطفال في المستوى نفسه كما لا يجوز انه مطابق لثبات اداء عينة من الاناث على الاختبار نفسه . وسواء كنا في اي مجال من مجالات البحث يتبعن دائمًا أن نعود إلى معاملات الثبات الخاصة لمجموعات متشابهة للافراد نتعامل معهم أو يدخلون في عيناتنا ويتحدد مقدار الثقة في النتائج التي نخرج بها أو نشير إليها في ضوء معاملات الثبات التي نتوصل إليها كما يتبعن الاشارة في تقاريرنا إلى اسلوب الثبات المستخدم وتقديم التبرير المناسب لاستخدام هذا الاسلوب أو هذه الاساليب دون غيرها ويعتبر الصفر ادنى معامل للثبات أما الدرجة +1 فتمثل أعلى معامل للثبات .

تعتمد جميع طرق حساب نتائج الاختبارات اعتماداً مباشراً على فكرة معاملات الارتباط كما سبق أن اشرنا إلى ذلك في تحليلنا لمعنى الثبات وإذا كان الارتباط يدل على الثبات فإن الافتراض يدل على عدم الثبات أو على الشوائب التي تحول بين الاختبار ودقة القياس .

ويكن أن للشخص أهم الوسائل الاحصائية لقياس الثبات في الطرق التالية:

2-6- طريقة اعادة الاختبار: Test - Retest:

تقوم فكرة هذه الطريقة على اجراء الاختبار على مجموعة من الافراد ثم اعادة اجراء نفس الاختبار على نفس الافراد بعد مضي فترة زمنية وهكذا يحصل كل فرد على درجة في الاجراء الاول للاختبار وعلى درجة اخرى في الاجراء الثاني للاختبار ، وعندما نرصد هذه الدرجات ونحسب معامل ارتباط درجات المرة الاولى بدرجات المرة الثانية فأئنا نحصل بذلك على معامل ثبات الاختبار .

تصلح هذه الطريقة للختبارات الموقوتة ذات الزمن المحدود والتي تعتمد إلى حد كبير على السرعة ، وتصلح ايضاً للختبارات غير الموقوتة التي لا تخضع للتحديد الزمني السابق وتقوم في جوهرها على قياس قوة الاستجابات الفردية اكثر مما تعتمد على قياس سرعة تلك الاستجابات . ولا تصلح هذه الطريقة لحساب ثبات الاختبارات التي تهدف إلى قياس التذكر أو ترتبط ارتباطاً مباشراً بهذه العملية العقلية وذلك لتأثير عملية التذكر تأثيراً مباشراً بالفاصل الزمني الذي يمضي بين اجراء الاختبار للمرة الاولى واعادة اجرائه للمرة الثانية .

وقد دلت نتائج الابحاث التجريبية على أن الحد المناسب للفاصل الزمني الذي يمضي بين اجراء الاختبار في المرة الاولى والثانية يجب أن لا يتجاوز اسابيع قليلة بالنسبة للاطفال أو طلبة المرحلة الاولى وطلبة المرحلة الاعدادية وألا يتجاوز ستة أشهر بالنسبة للكبار البالغين كطلبة المرحلة الثانوية وطلبة الجامعات . ومهما يكن من هذا التحديد الزمني فإن العوامل المؤثرة على الموقف التجاري في الاجراء الاول للاختبار تختلف إلى حد ما عن العوامل المؤثرة على الموقف التجاري في الاجراء الثاني وهذا يؤدي إلى ضعف الضبط التجاري ولذا تتأثر النتائج النهائية لتلك الطريقة بالشوائب الكثيرة التي يصعب اخضاعها للظروف التجريبية الدقيقة وهكذا تدرك مدى قصور هذه الطريقة على مستوى

الدقة العلمية التي يهدف إليها في ابحاثنا المختلفة وقد يعاد عليها أيضاً إنها تكلف الباحث جهداً ومالاً ووقتاً.

ويعد أسلوب إعادة الاختبار من أهم أساليب حساب الثبات ويتلخص هذا الأسلوب في اختبار عينة من الأفراد ثم إعادة اختبارهم مرة أخرى بالاختبار نفسه في ظروف متشابهة تماماً للظروف التي سبق وتم اختبارهم فيها ثم حساب معامل الارتباط المناسب بين أدائهم في المرتين ويعبر معامل الارتباط الذي تحصل عليه عن ثبات الاختبار ويلاحظ أن ما نقوم به هنا هو الحصول على عينة من أداء مجموعة من الأفراد وبعد مضي فترة زمنية معينة نعود للحصول على عينة جديدة من أداء الأفراد أنفسهم في الاختبار نفسه. وإذا افترضنا أن الأداء ثابت ولم يتغير رغم مضي هذه الفترة الزمنية رغم ما قد حصل به من تدخل عوامل مختلفة فإننا تتوقع أن نجد أن الدرجة التي يحصل عليها الفرد في المرتين واحدة وبالتالي يكون معامل الثبات (أي معامل الارتباط بين الأداء في المرتين) عالياً غير أن هناك أموراً تؤدي إلى تباين درجة الفرد واختلافها بين المرتين ومن هذه الأمور كالتالي :

- أ- التذبذب العشوائي في الأداء الواحد في المواقف
- ب- تعرض الفرد للتغيرات مختلفة خلال المدى الزمني المحدود الذي انقضى منذ اختباره وحتى إعادة اختباره وهذه التغيرات قد تكون في حالته الصحية أو المعنية أو في تأثيره بالتغيرات في حالة الجو أو غير ذلك من التغيرات .
- ج- التغيرات التي يمكن أن تحدث في موقف الاختبار نفسه والتي قد يكون المسؤول عنها هو الفاحص أو ظروف المكان أو تغيير الوقت من النهار بين جزئي الاختبار مما قد يبرز بعض عوامل التعب إذا كان الاختبار في نهاية اليوم بينما كان الاختبار الآخر في بداية النهار وهنا تتدخل في الأداء بعض الظروف الطبيعية المتغيرة في الموقف مثل حجم الاضاءة أو تنظيم موقف الاختبار أو غير ذلك من الجزيئات .

الواقع أن هناك عوامل كثيرة تتدخل في تحديد الفترة بين الاختبار وإعادة الاختبار وبصفة عامة يتغير الاختبار عن أسبوع ولا تزيد عن ستة أشهر ويحدد مضمون الاختبار هذه الفترة الزمنية كما تحددها طبيعة العينات المستخدمة في الدراسة. فمن حيث مضمون الاختبار يمكننا أن نتبين أن اختباراً من مضمون يتأثر بسرعة التذكر يجعل إعادة الاختبار بمثابة اختبار للتذكر أكثر منه اختباراً لاستقرار الاستجابة كما قد تكون طبيعة مادة الاختبار من ذلك النوع الذي يرتبط باكتشاف سياق معين للحل مما يؤدي إلى ارتفاع غير منتظم في درجات افراد العينة عند إعادة الاختبار وهو ما يتربّط عليه بالتالي انخفاض في معامل الثبات اذا ما كانت الفترة قصيرة بين الاختبار واعادته . ومن حيث طبيعة العينة المستخدمة في دراسة الثبات فقد تكون هذه العينة في مرحلة عمرية سريعة التأثير بالمتغيرات التعليمية مما يؤثر بشكل حاسم في ثبات الاختبار اذا اعيد بعد فترة طويلة .

واذا كانت مادة الاختبار من النوع الذي يسهل تعلمه أو تذكره أو الاستفادة فيه من التدريب أو يتعرض الافراد فيه لنوع من الالفة فمن الافضل أن تطول الفترة بين الاختبار واعادة الاختبار .

واذا كانت العينات من الاطفال صغار السن أو الاشخاص الذين يعبرون مراحل ارتقائية سريعة وحاسمة أو يعانون بعاقير تؤثر في ذاكرتهم أو انتباهم أو وظائفهم المختلفة فمن الافضل أن تنخفض الفترة بين الاختبار واعادة الاختبار . ويفضل عادة عند حساب معامل ثبات الاختبار بهذا الاسلوب أن لا يكتفي الباحث بحساب الثبات على مدى فترة زمنية واحدة بل على اكثر من فترة زمنية ، من ذلك حساب معامل الثبات لفترة خمسة عشر يوماً ول فترة شهر ول فترة ثلاثة أشهر ولاحتمال تأثير اداء العينة باعادة اختبارها اكثراً من مرة سواء في الاتجاه الايجابي أو السلبي اي من حيث تحسن الاداء نتيجة للمران والتدريب أو سوء الاداء نتيجة للملل أو عدم الاهتمام فمن الممكن أن يستخدم الباحث في هذه الحالة ثلاثة عينات متشابهة تماماً لكن يعد اختبار كل منها بعد فترة مختلفة ويوفر مثل هذا الاجراء تقديرأً مناسباً للثبات ومدى تأثيره بطول الفترة حتى اعادة الاختبار . في هذه

الطريقة يجري الاختبار مرتين على المجموعة نفسها من المفحوصين ويحسب معامل الارتباط بين النتائج في المرتين فيكون هو معامل ثبات الاختبار ولكن تتأثر النتيجة في هذه الطريقة بعدة عوامل يحسن أن نتعرف عليها لتحاشي الانخداع نتيجة هذا الاجراء . فمن المعروف انه عند تطبيق الاختبار لأول مرة تتدخل عوامل تؤثر على النتيجة ولا تتكرر بعض العوامل عند الاجراء الثاني للختبار ، فمثلاً في التطبيق الأول يكون الاختبار جديداً بالنسبة للمفحوص الا انه في الاجراء الثاني يكون قد الفه كما يكون قد تعلم موقف الاختبار وطريقة الاجابة عليه مما يؤثر على نتيجته وبالاضافة إلى ذلك فإن المفحوص عندما يعيد الاختبار للمرة الثانية يكون قد اكتسب خبرة فيه وترن عليه واستفاد من عملية انتقال اثر التدريب من الاجراء الاول إلى الاجراء الثاني . ومن العوامل التي تؤثر كذلك في نتيجة المفحوص هي احتمال تأثير عامل النمو لديه وذلك اذا كانت الفترة بين الاجراء الاول والاجراء الثاني طويلة ويكثر تأثير هذا العامل في حالة الصغار اكثر منه في حالة الكبار ، وفي حالة الاذكياء اكثر منه عند اقل ذكاء . كما يتأثر معامل الثبات عند استخدام طريقة اعادة تطبيق الاختبار بالذكر وذلك اذا كانت بين التطبيق الاول والتطبيق الثاني قليلة او اذا كانت فقرات الاختبار لا تتطلب اعمال وظائف عقلية معقدة . ولكي تحاشى تأثير العوامل التي تقلل من فعالية هذه الطريقة يمكن مراعاة الامور التالية :

- 1- أن لا تكون الفترة بين الاجراء الأول والاجراء الثاني قصيرة لكي تزيد من احتمال تدخل عامل التذكر عند المفحوصين بحيث لا تقل عن اسبوع في حالة الاختبارات الطويلة والاختبارات التي تتطلب اعمال وظائف معقدة ولا تزيد عن شهر في الاختبارات القصيرة والتي تقيس وظائف بسيطة يسهل تذكرها .
- 2- لا تكون الفترة بين الاجراء الأول والاجراء الثاني طويلة بشكل يساعد في تدخل عامل النمو الجسمي أو العقلي أو الاجتماعي لدى المفحوصين والمعروف أن نمو الصغار اسرع من نمو الكبار وكذلك نمو الاذكياء بالنسبة لغير الاذكياء وذلك للتقليل من تأثير عامل النمو لدى المفحوصين .

وتتضمن هذه الطريقة تطبيق الاختبار على عينة مماثلة من الافراد ثم إعادة تطبيق الاختبار عليها مرة أخرى بعد مرور فترة مناسبة من الوقت، ثم يحسب معامل الارتباط بين الدرجات التي حصل عليها افراد العينة في المرة الاولى والثانية، والفرق بين طرفيتي التجزئة النصفية واعادة الاختبار. هو أن معامل الثبات في الطريقة الاولى يشير إلى طريقة التجانس بين العبارات ويقصد بالتجانس أن العبارات تقيس مفهوماً واحداً، بينما يشير معامل الثبات بطريقة اعادة الاختبار إلى درجة استقرار الافراد في اجاباتهم على المقياس عبر فترة مناسبة من الزمن. ويدلنا معامل الثبات العالي بطريقة اعادة الاختبار على وجود استقرار في اجابات الافراد عبر فترة معينة من الزمن، اما معامل الثبات الواطيء فقد يعني أن الظاهرة التي يقيسها الاختبار تتذبذب لدى الافراد الذين يطبق عليهم الاختبار بين وقت وأخر أو أن القياس قد يكون غير دقيق في قياس تلك الظاهرة، أو أن اجابات الافراد تتأثر بظروف لا تتعلق بالاختبار نفسه بقدر تعلقها بالمجيب نفسه أو بالظروف المحيطة به. ومن الضروري في حساب الثبات بطريقة اعادة الاختبار، تقدير الوقت بين تطبيق الاختبار في المرة الاولى وتطبيقه في المرة الثانية بعناية، فإذا كانت الفترة قصيرة جداً بين التطبيقين، فإن هناك احتمالاً كبيراً بأن يتذكر المجيب الاجابة التي ذكرها في المرة الاولى. وعندما تكون الفترة طويلة جداً فقد يحدث تغيير حقيقي في الصفة السلوكية التي يقيسها الاختبار (وذلك بسبب ثو الفرد خلال هذه الفترة).

ولهذا السبب فإنه ليس هناك اتفاق على الفترة الزمنية المناسبة بين تطبيقي الاختبار ويمكن القول بصورة عامة، إن ذلك يعتمد في غالب الاحيان على طبيعة الظاهرة التي يقيسها الاختبار، وفي كثير من الحالات يعتبر اعادة تطبيق الاختبار بعد فترة اسبوع أو اسبوعين فترة مناسبة.

ومن المأخذ على هذه الطريقة أن الدرجات التي يحصل عليها الافراد في المرة الثانية من تطبيق الاختبار تكون أعلى بقليل من درجاتهم في التطبيق الاول، وذلك بسبب الفة المجيب بالاختبار وتذكره لاجابته في التطبيق الأول. كما تتأثر اجابات الفرد في

الاختبارين (التطبيق الأول والثاني) بعوامل عديدة كالصحة العامة والدافعية للاجابة عن الاختبار والحالة الانفعالية لديه وطريقه فهمه لتعليمات الاختبار وأسئلته والظروف البيئية المحيطة به، لذلك فمن الضروري بالنسبة لمصمم الاختبار أن يهتم بهذه المتغيرات الداخلية، العرضية التي يسبب اهمالها إنخفاض معامل الثبات الذي نحصل عليه من طريقة إعادة الاختبار.

يمكن استخدام طريقة إعادة الاختبار وبصورة كبيرة في مجال الاختبارات والقياس العامة وخصوصاً في المجال الرياضي لايجاد معامل ثبات الاختبار والذي يريد استخدامه قبل البدء في تطبيق الاختبار باتباع الخطوات التالية:

- تحديد الاختبار.
- تحديد العينة التي يجري عليها الاختبار.
- اجراء الاختبار للمرة الاولى واحتساب النتائج.
- اجراء الاختبار للمرة الثانية واحتساب النتائج تحت نفس الظروف وعلى نفس المجموعة.
- ملاحظة الفترة الزمنية بين اجراء الاختبار في كلتا المرتين.
- ايجاد معامل الارتباط بين نتائج الاختبار في كلتا المرتين.
- الكشف في جدول دلالة معامل الارتباط للتأكد من دلالة الاختبار تحت احدى المستويين 0.05 أو 0.01 وعلى ذلك فإن الفترة المسموح بها بين اجراء الاختبار في كلتا المرتين، يجب الا تطول كثيراً، حتى لا تستفيد عينة المختبرين من بعض العوامل أو المتغيرات والتي يمكن أن تؤثر على النتائج في المرة الثانية كعامل النمو مثلاً كما هو الحال في اختبارات عناصر اللياقة البدنية كالقوة العضلية، ويجب أن لا تكون قصيرة جداً حتى لا تتأثر العينة من جراء ما اكتسبوه من بعض المعارف نتيجة اجراء الاختبار في المرة الاولى.

حيث يتضح نتيجة ذلك في اختبارات الورقة والقلم بعدم اجراء الاختبار للمرة الثانية بعد المرة الأولى مباشرة او بزمن قصير حيث يمكن أن تستفيد مجموعة المختبرين من معارفهم التي اكتسبوها في المرة الأولى وعلى ذلك فإن معامل ثبات الاختبار يتأثر بهذه المعرف ، ولذلك ينصح بعدم اطالة الفترة بين الاختبار بل في الحالات التي يمكن أن يحدث فيها ثباتاً كما ينصح أن تطول الفترة في اختبارات الورقة والقلم والتي تعتمد على الناحية الذهنية ، وهنا يجب أن لا نغفل العوامل الهامة الأخرى والتي لها تأثير على ثبات الاختبار والتي تكلمنا عنها . كالانسجام بين مستوى الاختبار ومستوى المختبرين وتعدد مفردات الاختبار وعدم تجانسها وكذلك التباين في مستوى المجموعة المجربة والتي تكلمنا عنها في الباب السابق .

ان ما يهمنا في هذه المناسبة هو كيفية احتساب «معامل الثبات» احصائياً وذلك من الناحية التطبيقية .

أولاً : عن طريق قانون معامل ارتباط بيرسون .

$$r = \frac{\sum (M_s - \bar{M}_s)(M_c - \bar{M}_c)}{\sqrt{\left[\frac{\sum (M_s - \bar{M}_s)^2}{n} \right] \left[\frac{\sum (M_c - \bar{M}_c)^2}{n} \right]}}$$

حيث r_t = معامل ارتباط

s = المتغير الأول (والذي يعبر عن القياس الأول)

c = المتغير الثاني (والذي يعبر عن القياس الثاني)

n = عدد الحالات (والذي يعبر عن عدد افراد العينة)

ثانياً : عن طريق قانون معامل ارتباط الرتب لسبيرمان.

$$R_t = \frac{6 \cdot M_f^2}{n(n^2 - 1)} - 1$$

حيث R_t = معامل ارتباط .

M_f^2 = مجموع مربع فروق الرتب .

n = عدد الحالات أو عدد البيانات .

مثال :

لإيجاد معامل الثبات لأحدى الاختبارات (إعادة الاختبار) بطريقة بيرسون :

أراد أحد الباحثين التأكد من ثبات الاختبار « الوثب العريض من الثبات » على عينة معينة من المختبرين والذي يهدف إلى قياس قدرة عضلات الساقين (القوة المميزة بالسرعة) وهذه العينة مكونة من عشرين لاعباً ، وكانت نتائجهم كما هو مبين في الجدول التالي وذلك لاختبارين متتاليين على مدى ثلاثة أيام وقد استخدم الباحث معامل الارتباط بين نتائج الاختبارين من القيم الخام التي حصل عليها للوصول إلى ثبات الاختبار .

اللاعبين الرتبة	نتائج الاختبارين		الثاني (ص)	الأول (س)	اللاعبين
	ص ²	س ²			
1	40000	36100	38000	سم200	سم190
2	36100	32400	34200	190	180
3	36100	38025	37050	190	195
4	38025	40000	39000	195	200
5	44100	42025	43050	210	205
6	30625	28900	29750	175	170
7	32400	30625	31500	180	175
8	46225	44100	45150	215	120
9	44100	46225	45150	210	215
10	55225	52900	54050	235	230
11	28900	27225	28050	170	165
12	32400	36100	34200	180	190
13	36100	38025	37050	190	195
14	48400	44100	46200	220	210
15	50625	48400	49500	225	220
16	36100	34225	35150	190	185
17	40000	38025	39000	200	195
18	40000	46225	43000	200	215
19	46225	48400	47300	215	220
20	27225	27225	27225	165	165
مج(20)	788875	779250	783575	3955	3930

$$r = \frac{\frac{\text{مجس} \times \text{مجس}}{\text{مجس ص}} - \frac{\text{مجس ص}}{n}}{\sqrt{\left[\frac{\text{مجس}^2}{n} - \left(\frac{\text{مجس}}{n} \right)^2 \right] \left[\frac{\text{مجس}^2}{n} - \left(\frac{\text{مجس}}{n} \right)^2 \right]}}}$$

بعد تحويل النتائج إلى أمتار والتقرير:

$$r = \frac{\frac{29.5 \times 39.3}{20} - 78.3}{\sqrt{\left[\frac{2(39.5)}{20} - 78.9 \right] \left[\frac{2(39.3)}{20} - 77.9 \right]}}$$

$$r = \frac{77.6 - 78.3}{\sqrt{(78.01 - 78.9)(77.2 - 77.9)}}$$

$$r = \frac{0.7}{\frac{0.7}{0.79} - \frac{0.7}{0.9 \times 0.7}}$$

وبالكشف في جدول دلالة معاملات الارتباط لمعرفة مدى ثبات الاختبار بدرجة حرية (20-2=18) ويسوى 0.01 وجد أن القيمة الجدولية المنشورة هي 0.549 وهي أقل من القيمة المحسوبة، لذا فإن نتيجة الاختبارات ذات دلالة معنوية بدرجة عالية.

طريقة سبيرمان:

مثال:

ارادت باحثة ايجاد معامل ثبات اختبار لقياس مرونة مفصل الكتف على عينة من

احدى عشرة طالبة من طالبات كلية التربية الرياضية جامعة بغداد، والاختبار عبارة عن مسک عصا بالقبضتين باتساع الصدر تقربياً محاولة تمرين العصا امام الجسم عالياً خلفاً - الذراعان عمودتان بقياس عرض الكتفين والمسافة بين القبضتين .

- التقويم : المسافة بين القبضتين - عرض الكتفين = مؤشره لمرونة مفصل الكتف .

لقد اجريت الباحثة الاختبار على العينة في يوم الاحد الموافق 6/12/1981 ثم اعادت الاختبار على نفس العينة وتحت نفس الظروف يوم 13-12-1981 وحصلت على النتائج التالية وذلك باستخدام قانون ارتباط الرتب لسييرمان .

$$R_t = \frac{2^{M_f}}{F(n-2)}$$

عينة البحث	الاختبار الأول	الاختبار الثاني	ر	ف	ف	ف2
بشرى	16	11	7	7.5	5.-	.25
سنديس	4	8	3	4	1.0	1.5
سمراء	19	22	9	9	9	صفر
عبيز	11	6	5	3	2.0	4.0
هند	6	10	4	5.5	1.5-	2.25
رسالة	27	25	10	10	10	صفر
أمل	14	11	6	7.5	1.5-	2.25
هنا	3-	9-	1	1	1	صفر
عفراء	59	38	11	11	11	صفر
ابتسام	1	2-	2	2	2	صفر
نجاة	17	10	8	5.5	2.5	6.25
(11) مج			66	66	4.5+ 4.5-	16

$$R_t = \frac{1 - \frac{1}{n(n-1)}}{\frac{16 \times 6}{120 \times 11} - 1} = \frac{6 \text{ مجف}}{96}$$

$$R_t = \frac{1 - \frac{96}{1320}}{0.073 - 1} = \frac{1}{12}$$

وبالبحث في جدول دلالة معاملات الارتباط وجد أن الاختبار ثابت حيث انه دال تحت مستوى 0.01 درجات الحرية 12-2=10

2-6-2 طريقة التجزئة النصفية Split- Falf

نقوم في هذه الحالة بقسمة عدد فقرات الاختبار إلى نصفين متساوين فإذا كان الاختبار يتكون من 50 فقرة فإن النصف الاول سيكون عبارة عن الفقرات من (1) إلى (25) والنصف الثاني سيكون عبارة عن الفقرات من 26 إلى 50 ثم نقوم بحساب معامل الارتباط بين نصفي الاختبار على عينة من الافراد ويعبر معامل الارتباط هنا عن ثبات الاداء على عينة نصف الاختبار فقط فإذا كان اختبارنا جيد التصميم متوقع أنه سيظهر ثباتاً في الاداء يعكسه الارتباط بين نصفين . غير أن هناك بعض المقاييس لا تقبل القسمة النصفية لاعتبارات كثيرة وهنا لا يجوز استخدام هذه الطريقة ، وتم هذه الطريقة بإجراء الاختبار مرة واحدة وبعد تصحيحه يحسب للمفحوصين علامتان احدهما على نصف الاختبار والثانية على النصف الثاني .

وأفضل اجراء لذلك هو أن تحسب علامة لفقرات الاختبار الفردية ، والعلامة الثانية للفقرات الزوجية وثم يحسب معامل الارتباط بين العلامتين .

تمتاز هذه الطريقة بأنها توفر الظروف نفسها في أجزاء نصفي الاختبار وتبعد عوامل التذكر والنمو والألفة بموقف الاختبار ، كما مر في الطريقتين السابقتين الا انه يؤخذ عليها أنها نتعامل مع نصف الاختبار كأنه اختبار كامل في طبيعة الحال فإن فقرات هذا النصف

غير كافية لتمثيل جوانب السلوك المقيدة، وقد يلجأ البعض لاجراء بعض المعادلات الاحصائية التي تعالج هذا الضعف. وهناك عدة طرق لتقسيم الاختبار، منها تحديد مستوى صعوبة كل سؤال، وتوزيع الاسئلة بالتساوي على قسمي الاختبار حسب مستوى الصعوبة، وقد توزع الاسئلة على مدى صلاحيتها للتمييز بين الافراد والذين حصلوا على درجات عالية فيها، والافراد الذين حصلوا على درجات واطئة فيها. أو أنها قد توزع إلى مجموعتين بحيث توضع الاسئلة الفردية في قسم الزوجية في القسم الآخر، والطريقة الاخيرة هي أكثر طرق التجربة النصفية استخداماً والمثال الآتي يوضح كيفية تطبيقها.

نفرض أن أحد الباحثين قام بإعداد اختبار يتكون من (12) عبارة أو سؤالاً، وقد وزع الاختبار على عينة من الطلبة بلغ عدد افرادها (10) طلاب وعند التصحيح اعطي (درجة واحدة) إذا كانت اجابة الطالب على العبارة اجابة صحيحة، واعطاه (صفرًا) إذا كانت اجابته على العبارة خاطئة.

وفي الجدول التالي يكون الطالب الذي حصل على المرتبة الثانية قد اجاب عن العبارة (1)، (2)، (3)، (6) اجابة صحيحة، في حين كانت اجاباته عن بقية العبارات خاطئة.

وفي هذه الحالة تعطى له درجتان لا جابته الصحيحة عن العبارات الفردية (1)، (3) وتعطى له درجتان ايضاً لا جابته عن العبارات الزوجية (2)، (6) اجابة صحيحة.

(جدول يوضح كيفية حساب درجات الأسئلة الفردية والزوجية لاحدى الاختبارات).

الأسئلة												ترتيب الطلبة		
درجات الأسئلة الزوجية	درجات الأسئلة الفردية	12	11	10	9	8	7	6	5	4	3	2	1	
0	2	0	0	0	0	0	0	0	0	0	1	0	1	1
2	2	0	0	0	0	0	0	0	0	0	1	1	1	2
2	2	0	0	0	0	0	0	0	1	1	1	1	1	3
3	2	0	0	0	0	0	0	0	1	1	0	1	1	4
2	3	0	0	0	0	0	0	0	1	1	1	1	1	5
2	4	0	0	0	0	0	1	1	0	0	1	1	1	6
3	4	0	0	0	0	0	1	1	1	1	1	1	1	7
5	4	0	0	1	1	1	1	1	1	1	1	1	1	8
5	6	0	1	1	1	1	1	1	1	1	1	1	1	9
6	6	1	1	1	1	1	1	1	1	1	1	1	1	10

ثم يحسب الباحث معامل الارتباط بين الدرجات الفردية والزوجية للطلاب، وفي المثال السابق يبلغ الارتباط (0.89) وتوضح الدرجات الارتباطية الثبات بالنسبة إلى نصف الاختبار فقط وليس الاختبار كله لأننا كنا قد قسمنا الاختبار إلى قسمين. ولأجل حساب ثبات الاختبار بصورة كاملة نلجأ إلى استخدام معادلة سبيرمان براون لتصحيح معامل الارتباط السابق. وفي المثال أعلاه يبلغ معامل الثبات بعد تصحيحه بهذه المعادلة (0.894) وعلى العموم فإن هذه المعادلة ترفع من درجة معامل الارتباط، لأن طريقة التجربة النصفية

ذات فائدة في الاختبارات التي تكون فيها الاسئلة أو العبارات متجانسة. اما اذا قاس الاختبار عدداً من الخصائص او الموضوعات فلا يكون من المناسب استخدام هذه الطريقة لانه لا يمكن من الناحية العلمية المساواة بين فقرات الاختبار في الصعوبة، والمساواة بين متوسطي قسمي الاختبار وانحرافاتها المعيارية. ومن عيوب هذه الطريقة انها اقل كلفة في الجهد والوقت وذلك لانها لا تتطلب اعادة تطبيق الاختبار.

2-6-2-1 الفردي والزوجي :

اسلوب اخر من اساليب التنصيف : هو قسمة فقرات الاختبار نصفين لا من متتصف الاختبار ولكن من بدايته وحيث يتكون النصف الاول من الفقرات الفردية ويكون النصف الثاني من الفقرات الزوجية.

فإذا كان الاختبار يتكون من (50) فقرة فإن النصف الاول يتكون من الفقرات (1، 3، 5، .. إلى 49) بينما يتكون النصف الثاني من الفقرات ارقام (50، 4، 2، .. إلى 6) والميزة التي توفرها هذه الطريقة مقارنة بالطريقة السابقة هي انها تؤدي إلى ضبط متغير دخيل يؤثر في موقف الاداء ولا يدخل في نطاق تباين الخطأ الذي تقدرها طرق الثبات بالتصنيف.

وهذا المتغير الدخيل هو عوامل التعب أو الارهاق أو الملل التي يتعرض لها المفحوص أثناء ادائه فتؤثر في الجزء الاخير من الاختبار فإذا كان هذا الجزء الاخير كله يدخل في احد نصفي الاختبار وفقاً للطريقة السابقة فسيتأثر الارتباط وبين الجزء الاخير نتيجة لذلك. اما اذا كانت فقرات الجزئين موزعة بالتساوي (زوجي في نصف الاختبار وفردي في النصف الآخر) فتوقع أن يتوازن الاداء في النصفين مما يؤدي إلى ارتباط مرتفع بين النصفين .

2-6-2-2 جزء الاختبار:

رغم أن اسلوب التنصيف إلى فردي وزوجي من افضل الاساليب التي تؤدي إلى

الحصول على نصفين متعادلين في خصائصهما، الا أن بعض الاختبارات والمقاييس لا يصلاح معها اسلوب التصنيف إلى فردي وزوجي من ذلك أن بعض الاختبارات تقسم إلى عدد من الفقرات وتقدم كل مجموعة في فقراتها أو كل فقرة على حدة في وحدة زمنية معينة أو يكون المطلوب من المبحوثين تقديم عدد كبير من الاستجابات لفقرة واحدة في فترة زمنية معينة ثم يقدم الفقرة التالية في فترة زمنية جديدة عائلة وهكذا. فالاسلوب الأمثل هنا هو أن نقسم الاختبار إلى جزئين يتضمن كل منها جزء من جزئيه أو يتضمن كل نصف جزئين على حدة اذا كان يتكون من اربعة أجزاء مستقلة حيث يقوم كل منها في الوقت المحدد له وقد نلجم هنا إلى تضمين النصف الأول الجزئين (1 ، 3) بينما يتضمن النصف الثاني الجزئين (2 ، 4) مستفيد من خصائص هذه الطريقة وخصائص طريقة الفردي والزوجي .

2-6-3- المعدلات الاحصائية لطريقة التجزئة النصفية

تتلخص اهم معدلات طريقة التجزئة النصفية فيما يلي :

- 1- معادلة سبيرمان وبراون .
- 2- معادلة رولون .
- 3- معادلة جثمان .
- 4- معادلة جلكسون .

وسنین فيما يلي ميزات كل معادلة من تلك المعدلات وتطبيقاتها المختلفة ونواحي قصورها .

2-6-3-1- معادلة سبيرمان للتجزئة النصفية :

بين سبيرمان C.spearman وبراون Brown سنة 1910 انه يمكن التنبؤ بمعامل ثبات اي اختبار اذا علمنا معامل ثبات نصفه او اي جزء منه فمثلاً اذا امكننا ان نقسم اي اختبار

إلى جزئين متكافئين. ثم حسبنا معامل ارتباط الجزئين فإننا نستطيع أن نستعين بمعادلة التنبؤ لسبيerman وبراون في معرفة معامل ثبات الاختبار الكلي الذي يتكون من هذين الجزءين وهكذا نستطيع أن نتغلب على الصعوبات التجريبية التي حالت بيننا وبين دقة حساب الثبات بالطريقة السابقة التي تعتمد على فكرة اعادة اجزاء الاختبار.

وتعتمد فكرة تكافؤ الاختبارات على تساوي القيم العددية لمقاييسها الاحصائية المختلفة ، فمثلاً اذا امكننا أن نقسم الاختبار إلى ثلاثة اجزاء فإن هذه الاجزاء تصبح متكافئة

عندما تتحقق الشروط التالية :

$$3_m = 1_m$$

$$3_u = 2_u$$

$$r_{21} = r_{31}$$

حيث يدل الرمز 1 على الجزء الأول ويدل الرمز 2 على الجزء الثاني ويدل الرمز 3 على الجزء الثالث وحيث تتساوى ايضاً مستويات صعوبة الاسئلة في هذه الاجزاء ، اي صعوبة السؤال وهذه بدورها تساوي صعوبة السؤال الأول في الجزء الثالث وتتلخص الفكرة العامة لمعادلة التنبؤ في الصورة التالية :

$$r^A = \frac{n_r}{1 + (n - 1)r}$$

حيث يدل الرمز r^A على معامل ثبات الاختبار . ويدل الرمز n_r على عدد الاجزاء ويدل الرمز r على معامل ارتباط هذه الاجزاء او يعني اخر معامل ارتباط اي جزئين لأن $r_{21} = r_{31}$ معامل ارتباط اي جزئين . وتعتمد الطريقة التجريبية العملية لحساب الثبات على تجزئة الاختبار الى جزئين فقط بحيث يتكون الجزء الأول من الدرجات الفردية للاختبار ويكون الجزء الثاني من الدرجات الزوجية وبذلك تحول معادلة التنبؤ إلى الصورة التالية :

$$R_1 = \frac{r^2}{r+1}$$

حيث أن n أصبحت متساوية لـ 2

والجدول يوضح طريقة تجزئة درجات الاختبار إلى نصفين بحيث يقوم النصف الأول على درجات الأسئلة الفردية ويقوم النصف الثاني على درجات الأسئلة الزوجية.

السؤال	درجات الأسئلة الفردية	الأسئلة								الافراد
		8	7	6	5	4	3	2	1	
2	3	0	0	0	1	1	1	1	1	1
3	3	0	0	1	1	1	1	1	1	2
2	2	0	0	0	1	1	0	1	1	3
3	4	1	1	1	1	0	1	1	1	4
2	2	0	0	1	0	0	1	1	1	5
3	3	1	1	0	0	1	1	1	1	6
2	3	0	0	1	1	0	1	1	1	7
3	4	0	1	1	1	1	1	1	1	8
2	2	0	0	0	0	1	1	1	1	9
4	4	1	1	1	1	1	1	1	1	10

(جدول طريقة تجزئة درجات الاختبار إلى جزئين فردي وزوجي)

حيث يدل العمود الأول على الافراد، وتدل اعمدة الاسئلة على اجابات كل فرد على كل سؤال من اسئلة الاختبار، فمثلاً الفرد أجاب اجابات صحيحة على الاسئلة

1، 2، 3، 4، 5 واجاب اجابات خاطئة على الاسئلة 6، 7، 8 اي أن مجموع الاجابات الصحيحة على الاسئلة الفردية يساوي 3 ومجموع الاجابات الصحيحة على الاسئلة الزوجية يساوي 2 وهكذا بالنسبة لبقية الافراد . ومعادلة التنبؤ تصلح لحساب معامل ارتباط الدرجات الفردية بالدرجات الزوجية وهو يحسب في مثالنا هذا بالطريقة التالية :

$$\therefore \text{معامل ارتباط} = \frac{n \times \text{مجس ص} - \text{مجس}^2}{\sqrt{[n \times \text{مجس}^2 - (\text{مجس})^2] [n \times \text{مجس}^2 - (\text{مجس})^2]}}$$

\therefore معامل ارتباط الجزء الفردي بالجزء الزوجي :

$$\frac{26 \times 30 - 82 \times 10}{[676 - 72 \times 10] [900 - 96 \times 10]} = \frac{40}{51.38} = \frac{40}{2640} =$$

\therefore معامل ارتباط = 0.78 تقريرياً

وهكذا نستطيع أن نستعين بارتباط الجزئين الذي يدل على ثبات نصف الاختبار في التنبؤ بمعامل ارتباط الاختبار بنفسه أو يعني اخر معامل ثبات الاختبار ، وذلك بالاستعانة بمعادلة التنبؤ لسييرمان وبروان كما يدل على ذلك التحليل التالي :

$$r^2 = \frac{R^2}{1 + r}$$

و بما أن $r = 0.78$ في مثالنا هذا

$$\frac{1.56}{1.78} = \frac{0.78 \times 2}{0.78 \times 1} =$$

$\therefore R^2 = 0.88$ تقريرياً .

أي أن معامل ثبات الاختبار يساوي 0.88 هذا وقد حسبت معاملات ثبات الاختبار لكل القيم العددية الدالة على معاملات ارتباط النصف الفردي والنصف الزوجي ورصدت هذه القيم في الجداول الاحصائية وبذلك نستطيع أن نقرأ مباشرة معامل الثبات الذي يقابل ارتباط النصفين المساوي لـ 0.78 وسنرى انه يساوي 0.88 وهكذا تصبح عملية حساب الثبات عملية سريعة وسهلة .

ولا تصلح طريقة سبيرمان وبراؤن لحساب ثبات الاختبار التي لا تنقسم إلى أجزاء متكافئة وخاصة عندما تختلف القيم العددية للتباين اختلافاً كبيراً أي عندما تختلف القيمة العددية للتباين الجزء الفردي عن القيمة العددية، للتباين الجزء الزوجي اختلافاً واضحاً وذلك لأن البرهان الرياضي لمعادلة التبؤ يفترض تساوى الاجزاء في بنائه الاحصائي لتلك المعادلة كما يدل على ذلك البحث الذي نشره سبيرمان وبراؤن. ولا تصلح هذه الطريقة أيضاً لحساب ثبات الاختبارات الموقوتة التي تعتمد اعتماداً كبيراً على سرعة الاستجابات لأن كثرة الاسئلة المتروكة في آخر كل اخبار تؤثر على الارتباط بين الجزيئين ويغير بذلك معامل الثبات. وقد حاول هورست Horst P. أن يحسب معامل ثبات الاختبار بطريقة سبيرمان وبراؤن وذلك عندما لا تكون اطوال الاجزاء التي ينقسم لها الاختبار متساوية كأن يمثل الجزء الأول ربع الاختبار وأن يمثل الجزء الثاني ثلاثة أرباع الاختبار واستعan على ذلك بمعادلة جديدة لتحقيق الفكرة وبما أن عملية قسمة الاختبار تخضع لاختبار الباحث فلا ضرورة لهذا التعقيد، اللهم إلا في الحالات النادرة التي قد تدعو إلى مثل ذلك التقسيم.

لقد حاول موسيير Mosier أيضاً أن يحسب معامل ثبات الاختبار بطريقة سبيرمان وبراون وأقام فكرته على معامل ارتباط اي جزء من جزئي الاختبار بالاختبار كله وكان يهدف من هذا إلى حساب معامل ارتباط الجزئين. ومهما يكن من امر طريقة موسيير فهي في جوهرها لا تعدو أن تكون احدى الصور الرياضية لمعادلة سبيرمان وبراون. لكنها لا تسرع بالعملية كما كان يظن موسيير. وقد نجح رولون P.J.Rulon في الكشف عن احدى الصور الرياضية الجديدة التي تؤدي إلى حساب معامل الثبات بطريقة اسهل واسرع من طريقة سبيرمان وبراون.

2-3-2- معادلة رولون المختصرة للتجزئة النصفية:

تهدف هذه الطريقة إلى تبسيط معادلة سبيرمان وبراون وذلك بحساب تباين فروق درجات النصفين وحساب تباين درجات الاختبار.

وتتلخص فكرة رولون P.J.Rulon في المعادلة التالية :

$$\frac{R^2 - 1}{U^2}$$

حيث يدل الرمز R^2 على معامل الثبات. ويدل الرمز U على تباين فروق درجات النصفين. ويدل الرمز R^2 على تباين درجات الاختبار.

والجدول التالي يوضح طريقة حساب معامل الثبات هذه الطريقة .

الافراد	درجات الاستلة الفردية	درجات الاستلة الزوجية	درجات الدرجات الزوجية	درجات الدرجات الفردية	فروق الدرجات الزوجية - الفردية	درجات الاختبار
1	3	4	1-	7	الفردية + الزوجية	فرار
2	5	6	1-	11	فرار	فرار
3	9	7	2+	16	فرار	فرار
4	8	4	4+	12	فرار	فرار
5	2	3	1-	5	فرار	فرار
عدد الافراد	27	24	3	51	المجموع = 51	1
الافراد	729	576	9	26.1	مربع الدرجات = 26.1	2
ن=5	183	126	23	595	مجموع المربعات = 595	3

(جدول حساب معامل الثبات بطريقة رولون)

حيث يدل العمود الرابع على فروق درجات الاسئلة الزوجية من درجات الاسئلة الفردية هذا ولا تختلف النتيجة النهائية لهذه العملية اذا حسبنا فروق درجات الاسئلة الفردية من درجات الاسئلة الزوجية وعلى القارئ أن يقوم بحساب هذه الفروق ليرى أن تباين فروق الحالة الاولى يساوي تباين فروق الحالة الثانية . وبما أن التباين يدل على مربع الانحراف المعياري اذن فتبادر الفروق يحسب بالمعادلة التالية :

$$\text{بما أن الانحراف المعياري} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{مجس}_i - \bar{\text{مجس}})^2}$$

لكن التباين = مربع الانحراف المعياري

$$\therefore \text{التبادر} = \frac{1}{n} \sum_{i=1}^n (\text{مجس}_i - \bar{\text{مجس}})^2$$

$$\therefore \text{تبادر الفروق} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{106}{25}$$

$$x^2 n = 4.24$$

$$\text{وتبادر درجات الاختبار} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{374}{25}$$

$$\frac{374}{25} =$$

$$x^2 n = 14.96$$

$$\therefore \text{معامل الثبات} = \frac{4.24}{14.96} - 1 = 0.2834 - 1$$

$$\text{رأٍ} = 0.72 \text{ تقريباً}$$

وعلي القارئ أن يحسب معامل ثبات هذا الاختبار بطريقة سبيرمان وبراون وسيري انه يساوي 0.80 وهكذا تدرك مدى اقتراب طريقة رولون في حسابها للثبات من طريقة سبيرمان وبراون.

2-6-3-3 معادلة جثمان العامة للتجزئة النصفية :

سبق أن بينا في دراستنا لمعادلة التنبؤ لسبيرمان وبراون لحساب معامل الثبات إلى عدم صلاحية هذه المعادلة لحساب الاختبارات التي لا تتساوى الانحرافات المعيارية لجزءيها وقد توصل جثمان L. Guttman إلى معادلة عامة تصلح لحساب الثبات عندما لا تتساوى الانحرافات المعيارية لجزئي الاختبار، وتصلح أيضاً لحساب هذا المعامل عندما تتساوى هذه الانحرافات المعيارية وتلخص هذه الفكرة في المعادلة التالية:

$$\rho_1 = \frac{U_1^2 + U_2^2}{U^2}$$

حيث يدل الرمز على تباين درجات الاسئلة الفردية ويدل الرمز على تباين الدرجات الزوجية وعندما نحسب معامل ثبات درجات الاختبار المبيبة في الجدول السابق نرى أن :

$$\text{تباین درجات الاسئلة الفردية } U_1^2 = \frac{1}{25} (729 - 183 \times 5)$$

$$\frac{186}{25} = \frac{729 - 915}{25} =$$

$$\therefore U_1^2 = 7.44$$

$$\text{وتباین درجات الاسئلة الزوجية } U_2^2 = \frac{1}{25} (576 - 126 \times 5)$$

$$\frac{54}{25} = \frac{576 - 630}{25} =$$

$$\therefore U_2^2 = 7.44$$

وتبين درجات الاختبار $U^2 = 14.96$

كما سبق أن حسبناه في طريقة رولون

$$\therefore R^2 = \frac{2.16 + 7.44}{14.96} - 1 \times 2 =$$

$$\left(\frac{960}{14.96} - 1 \right) \times 2 =$$

$$(0.6417 - 1) \times 2 =$$

$$0.3583 \times 2 =$$

$$\therefore R^2 = 0.72 \text{ تقريرياً.}$$

وهذه هي نفس النتيجة التي حصلنا عليها لنفس هذا المثال وذلك عندما طبقنا طريقة رولون المختصرة لحساب معامل الثبات.

2-6-3-4 معادلة جلكسون للاختبارات الموقوتة :

تتأثر معادلة التنبؤ لسبيرمان وبراؤن بالزمن المحدد للاختبار ، ولذا لا تصلح هذه المعادلة لحساب ثبات الاختبارات الموقوتة التي تحول بين اغلب الافراد وبين تكميله الاختبار في الزمن المحدد للاجابة. هذا وكلما قل الزمن المحدد للاختبار زادت تبعاً لذلك نسبة الاسئلة المتروكة في آخر الاختبار أو الاسئلة التي لا يستطيع اغلب الافراد الاجابة عنها لضيق الوقت وبذلك يزداد التشابه القائم بين نصفي الاختبار وترفع القيمة العددية لمعامل ارتباط الاسئلة الفردية بالاسئلة الزوجية ويزداد تبعاً لذلك معامل ثبات الاختبار . ولذا يجب أن نصحح القيمة العددية لهذا الثبات حتى يدل على الثبات الحقيقي الذي لا يخضع لهذا العامل الزمني . وقد اقترح جلكسون H.Cullikson المعادلة التالية لثبات الاختبارات الموقوتة :

$$R^A = R^A - \frac{M}{x^2}$$

حيث يدل الرمز R^A على معامل ثبات الاختبارات الموقوفة أو معامل الثبات بعد تصحيح أثر السرعة ويدل الرمز R^A على معامل الثبات الذي حسب بطريقة سبيرمان براون.

ويدل الرمز M على متوسط الاسئلة المتروكة في آخر الاختبار ويحسب هذا برصد عدد الاسئلة المتروكة عند كل فرد، ثم تجمع الاسئلة المتروكة عند كل فرد ويقسم هذا المجموع على عدد الافراد لحساب متوسط الاسئلة المتروكة. ويدل الرمز x على تباین الخطأ ويحسب برصد عدد الاستجابات الخاطئة عند كل فرد ويضاف إلى هذا المجموع عدد الاسئلة المحذوفة اي الاسئلة التي حذفها الفرد أثناء اجابته على الاختبار دون أن يجب عليها ثم يحسب تباین هذه الاعداد بالنسبة لكل الافراد. وبذلك تعتمد فكرة هذه المعادلة على الانواع الرئيسية لاجابات الافراد على اسئلة الاختبارات الموقوفة والتي تتلخص فيما يلي:

- 1- الاجابات الصحيحة على الاسئلة وسنرمز لهذا النوع بالرمز (ص).
- 2- الاجابات الخاطئة على الاسئلة وسنرمز لها هذا النوع بالرمز (خ).
- 3- الاسئلة المحذوفة وسنرمز لها هذا النوع بالرمز (و).
- 4- الاسئلة المتروكة وسنرمز لها هذا النوع بالرمز (ك).

ومثال التالي يوضح هذه الانواع الرئيسية لاجابة الفرد على اختبار.

				الاستلة											
مج	مج خ	مج	مج ص												الافراد
ك	+		ص	8	7	6	5	4	3	2	1				
2	3	3	ك	لـ	هـ	كـ	وـ	صـ	وـ	خـ	صـ	وـ	صـ	صـ	أ

(جدول رصد الانواع المختلفة لاستجابات الفرد على استلة اختبار موقوت)

وعندما نرصد جميع استجابات الافراد بهذه الطريقة نستطيع أن نحسب متوسط الاستلة المتروكة ، وتبين الخطأ.

فإذا فرضنا مثلاً أننا حصلنا على القيم التالية:

$$R = 0.8, M_t = 2, \bar{x}^2 = 10$$

فإننا نستطيع تطبيق معادلة جلكسون في حساب ثبات الاختبار الموقوت بالطريقة

التالية:

$$\therefore R = 0.8 - \frac{2}{10}.$$

$$0.6 = 0.2 - 0.8 =$$

هذا ولا تصلح هذه المعادلة للاختبارات التي تعتمد اعتماداً كلياً على السرعة والتي يقل زيتها عن الزمن المناسب للاختبار لأن القيمة العددية لمتوسط الاستلة المتروكة قد تزداد عن القيمة العددية لتبين الخطأ، وبذلك يصبح الكسر ($\frac{M_t}{\bar{x}^2}$) أكبر من الواحد الصحيح وتحول قيمة R إلى قيمة سالبة. وإذا تستخدم طريقة إعادة الاختبار أو طريقة الاختبارات المتكافئة لحساب ثبات مثل هذا النوع من الاختبارات.

2-6-3 طريقة تحليل التباين Analysis of Variance

استعان كودر G.F.Kuder وريتشاردسون M.W.Richardson في دراستهما للثبات بتحليل استئلة الاختبار ودراسة تباين تلك الاستئلة ولذلك تعتمد طريقةهما على الدراسة التفصيلية لهذا التباين، وقد تمكنا الباحثان من استنتاج بعض المعادلات التي تصلح لقياس الثبات. وتحتاج أغلب هذه المعادلات إلى وقت طويق وجهد شديد لحساب الثبات من المقاييس الاحصائية لاستئلة الاختبار ولذا لم تلق صدى قوياً بين المشغلين بالدراسات الاحصائية، وقد حاول الباحثان تبسيط طريقتها في معادلة عامة لحساب التباين بطريقة سهلة سريعة وتلخص فكرة هذه المعادلة في الصورة التالية:

$$\rho^2 = \frac{n\bar{U}^2 - m(n-m)}{(n-1)\bar{U}^2}$$

حيث يدل الرمز ρ^2 على معامل ثبات الاختبار. ويدل الرمز n على عدد استئلة الاختبار. ويدل الرمز m على تباين درجات الاختبار. ويدل الرمز \bar{U} على متوسط درجات الاختبار.

هذا ويعتمد البرهان الرياضي لهذه المعادلة على الفروض التالية:

- 1- أن تتقارب صعوبة استئلة الاختبار.
- 2- أن يجيب كل فرد على جميع استئلة الاختبار.
- 3- أن يقيس الاختبار قدرة واحدة أو صفة واحدة.
- 4- أن تتساوى معاملات ارتباط الاستئلة. أي أن يصبح معامل ارتباط السؤال الأول بالسؤال الثاني مساوياً لمعامل ارتباط السؤال الأول بالسؤال الثالث وهكذا بالنسبة لباقي ارتباطات الاستئلة. ولذا يضيق النطاق التطبيقي لهذه المعادلة إلى الحد الذي يجعلها غير صالحة في كثير من الاحوال وقد استطاع بيرث C.Burt أن يبرهن على صحة هذه المعادلة

بطريقة تحليل التباين دون أن يخضع برهانه للفروض السابقة، ولذا أصبحت تلك المعادلة صالحة لقياس ثبات الاختبارات الموقوتة وغير الموقوتة بشرط الا يكون عدد الأسئلة المتروكة كبيراً، اي أن يستطيع اغلب الافراد الوصول إلى نهاية الاختبارات في الزمن المحدود له.

وعندما نستعين بهذه المعادلة في حساب معامل الاختبار المبين بالجدول والذي سبق أن حسبنا ثباته بطريقة رولون نرى أن:

$$\text{الدرجات} = 7, 12, 16, 11, 5$$

$$\text{مجموع الدرجات} = 51$$

$$\text{عدد الأفراد} = 5$$

$$\therefore \text{المتوسط} M = \frac{51}{5} = 10.2$$

$$\text{الانحراف المعياري} S = 3.87$$

$$\text{البيان} S^2 = 14.96$$

$$\text{ولنفرض أن عدد الأسئلة} N = 20$$

$$\therefore R = \frac{(10.2 - 20) 10.2 - 14.96 \times 20}{14.96 \times (1 - 20)} = 0.99$$

$$\frac{99.96 - 99.2}{284.24} =$$

$$\frac{199.24}{284.24} =$$

$$\therefore R = 0.70 \text{ تقريباً.}$$

وقد سبق أن حسبنا القيمة العددية لثبات هذا الاختبار بطريقة دولون وبيننا أنها تساوي 0.72 وحسبناها أيضاً بطريقة سبيرمان وبراؤن وبينا أنها تساوي 0.80، وهكذا نرى أن القيمة العددية لمعامل الثبات بطريقة كودرور ينشاردسون أقل قيمة نحصل عليها في قياسنا لهذا الثبات وإن القيمة العددية لثبات نفس هذا الاختبار بطريقة سبيرمان وبراؤن مثل أعلى قيمة نحصل عليها في قياسنا لهذا الثبات.

ولذا يرى بعض العلماء أن طريقة سبيرمان وبراؤن تدل على الحد الأعلى لثبات الاختبار وان طريقة كودرور ينشاردسون تدل على الحد الأدنى لهذا الثبات ولهذه الحدود أهميتها القصوى في صحة الحكم على الثبات.

2-6-4 طريقة الاختبارات المتكافئة Paralled tests

تعتمد فكرة الاختبارات المتكافئة على نفس الفكرة التي اعتمدت عليها طريقة التجربة النصفية لسبيرمان وبراؤن في تقسيم الاختبار إلى اختبارين متكافئين أو أكثر وفي التحقيق من هذا التقسيم بدراسة الفروق القائمة بين الانحرافات المعيارية . وقد سبق أن بينا في دراستنا لتلك الطريقة الشروط الاساسية للتكافؤ ولخصناها فيما يلي :

$$-1 = M_1 - M_2$$

$$-2 = U_2 - U_1$$

$$-3 = R_{32} - R_{21}$$

-4 تماثل تدرج الصعوبة في كل الأجزاء .

وذلك بالنسبة للجزاء الثلاثة التي يمكن أن ينقسم لها الاختبار الاصلي ، وقد بين جلكسون H.Gullikson وثورنديك R.H.Thorndike أن أقل عدد من الأجراء المتكافئة التي يمكن أن ينقسم إليها الاختبار الأصلي هو ثلاثة حتى نتأكد من تساوي معاملات الارتباط .

وعندما نستطيع تقسيم الاختبار الاصلي إلى هذه الاجزاء فإننا نتمكن أن نحسب ثبات أي جزء منها وذلك بحساب معامل ارتباطه بأي جزء من الاجزاء الأخرى وبذلك نحسب ثبات الاختبارات الجزئية مباشرة من معلومات الارتباط وبما أن معاملات ارتباط الاختبارات الجزئية المتكافئة متساوية . إذن فثبات أي اختبار منها يدل على ثبات اي اختبار

آخر. هذا وفي مقدورنا أن نزيد القيمة العددية لمعامل الثبات وذلك يضم اختبارين جزئين معاً في اختبار واحد وحساب معامل ثبات هذا الاختبار الجديد بطريقة سبيرمان وبرانون ونستطيع أيضاً أن نقسم الاختبار الكلي إلى أجزاء متكافئة ونستمر في تقسيمنا هذا حتى يصبح كل سؤال من اسئلة الاختبار جزءاً من هذه الأجزاء.

تقوم هذه الطريقة على أساس أعطاء الطالب - في نفس اليوم صورتين من الاختبار نفسه، بحيث تكون كل صورة مكافئة للأخرى وتساويها من حيث محتوى الاختبار والمتوسط والانحراف المعياري. كما يجب أن تكون عبارات الصورتين متساوية من حيث الصعوبة والعدد، وان تكون البذائل التي يختار منها المفحوص اجابته، وتعليمات الاجابة والتطبيق واحدة في صورتي الاختبار. ومن المأخذ على طريقة الصور المتكافئة في صعوبة اعداد صورتين متكافئتين وفي حالة تحقيق ذلك فعلاً فإنه يضاعف الجهد والوقت والكلفة لبناء الاختبار. وفي هذه الطريقة تعد صورة مكافئة للاختبار الذي نريد التحقق من ثباته، ونجري الاختبار وصورته المتكافئة على المفحوصين ويحسب معامل الارتباط بين النتائج على الاختبار وعلى صورته الكافية فيكون بذلك معامل ثبات أو معامل تكافؤ ومعامل استقرار لهذا الاختبار. ولكن يؤخذ على هذه الطريقة صعوبة بناء صورة متكافئة تماماً للاختبار الاصلي من حيث مستوى ونوع الوظيفة التي تقسيمها فقرات الاختبار وطريقة الصياغة وطول الاختبار وطريقة اجرائه وتصحيحه وتقويمه كما أن عامل انتقال اثر التدريب والالفة بالاختبار يزيد كلما اقتربت الصورة من الاصل.

2-6-5 المزج بين طرفيتي الاختبارات المتكافئة واعادة الاختبار :

وهناك طريقة تحاول التقليل من مشكلات «اعادة الاختبار نفسه» وذلك عن طريق اعداد صورتين من اختبار معين، تعطى الصورة الاولى منه في وقت معين وتعطى الثانية بعد ذلك بفترة، ثم يحسب معامل الارتباط بين درجتي الصورتين وبذلك تجمع هذه الطريقة بين مزايا الطريقتين معاً.

2-6-6 طريقة اتفاق الملاحظين Cooper

تعد طريقة اتفاق الملاحظين في حساب الثبات من اكثر الطرق استخداماً، ويطلب استخدام هذه الطريقة اكثر من ملاحظ في الوقت نفسه للاحظة سلوك المختبر باستخدام

نظام الملاحظة المراد ايجاد ثباته اذا استعان الباحث بلاحظ ثان . وبعد التدريب على عملية الرصد والتسجيل يقوم الملاحظان بتسجيل سلوك احد المختبرين خلال إحدى المحاضرات العملية لكي يحسب ثبات الاداء الملاحظين من خلاله ، وقد استخدمت طريقة موستلر و توكي (Mostler, Tukey, 1949) وقد استخدمت هذه الطريقة في دراسات سابقة (محمد امين، 1988)، (الصفار، 1993).

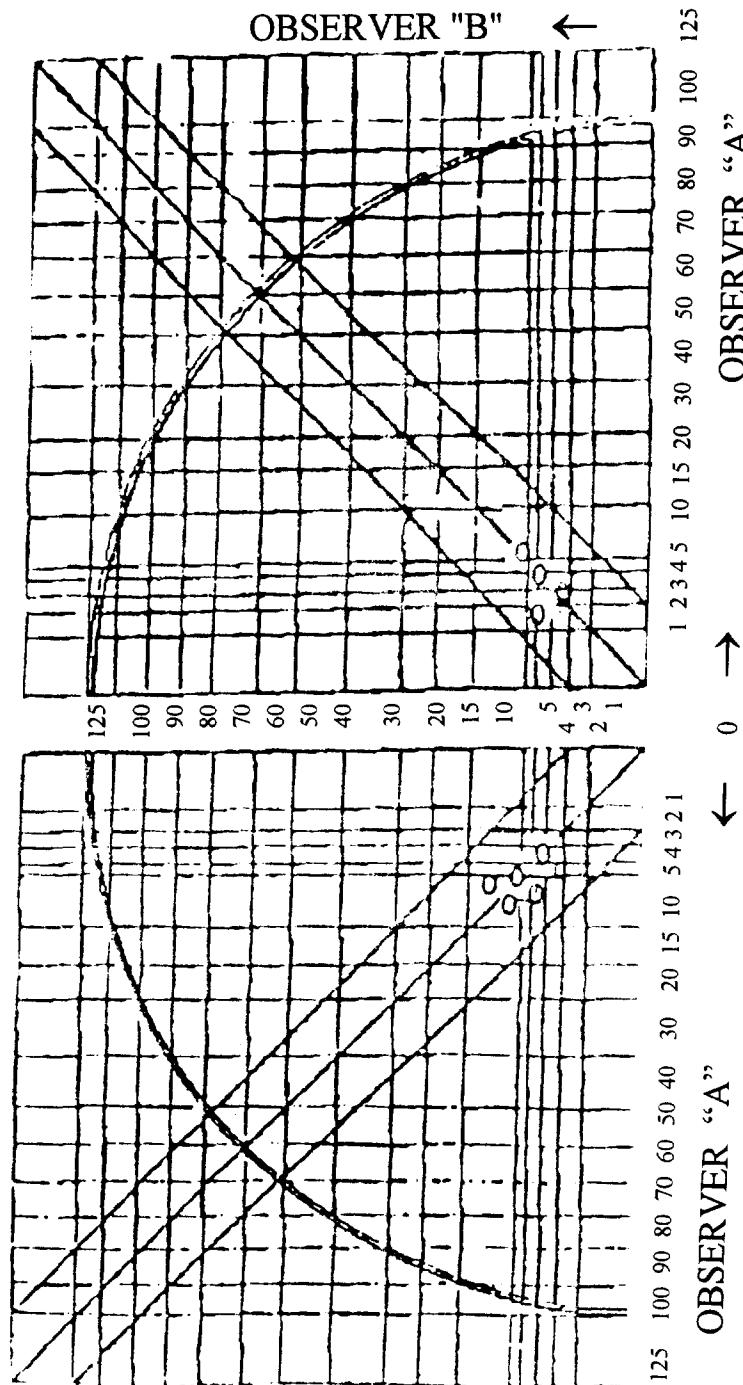
من مميزات هذه الطريقة البيانية انها سهلة الاستخدام وتتوفر سيطرة ثابتة للثبات وذلك بسبب انه عند تعين النقاط على الورقة البيانية يتضح مباشرة ما اذا كان هناك اتفاق او عدم اتفاق بين الملاحظين وكما في الوصف الآتي :

* تستخدم الورقة البيانية التي تكون تقسيماتها بطريقة تكون المسافة من نقطة الاصل هي الجذر التربيعي للحادي .

* يرسم الخط المائل الاوسط الذي يسمى بخط الجزء ، اما الخطان الموازيان له فهما خطان يرسمان على نحو متواز للخط المائل الاوسط ويبعد عنہ بمقدار (اثنين) وحدة بقياس الورقة من كل جهة . اذأن الاحدائي السيني تابع للملاحظ الاول اما الاحدائي الصادي فتابع للملاحظ الثاني ولتعين النقاط على الورقة . نفترض انه في احدى فترات الملاحظة كان الملاحظ الاول قد سجل (4) رصدات ممارسة نشاط في حين سجل الملاحظ الثاني (3) رصدات ممارسة نشاط حيث تقع تقاطع النقطتين (3.4) على الورقة .

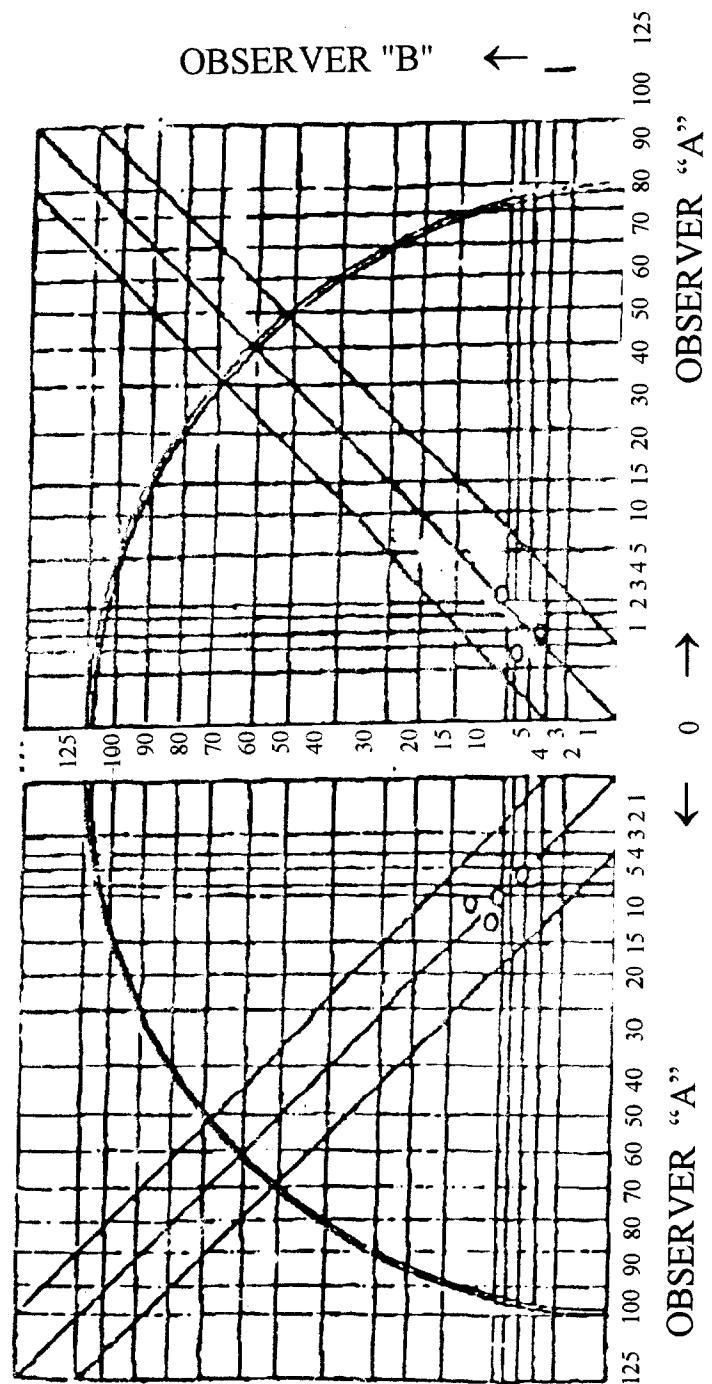
إن مكان تقاطع النقطتين هو الذي سيحدد إذا كان هناك اتفاق أو عدم اتفاق ، ويذكر دايكسون وموسى بأن نقاطات النقاط التي تتكون داخل الخطوط المتوازية مقبولة وإن كانت النقاط خارج المكان المحصورة بين الخطوط المتوازية يعني انعدام الاتفاق بين الملاحظين . ومن الملاحظ للاشكال التالية وفقاً لما تم شرحه فإن الاتفاق قائم بين الملاحظين وقد استخدمت معادلة كوبير لزيادة التأكيد من ثبات الملاحظة التي وضعها كوبير (Cooper, 1974).

$$\text{نسبة الاتفاق} = \frac{\text{عدد مرات الاتفاق}}{\text{عدد مرات الاتفاق} + \text{عدد مرات عدم الاتفاق}} \times 100$$



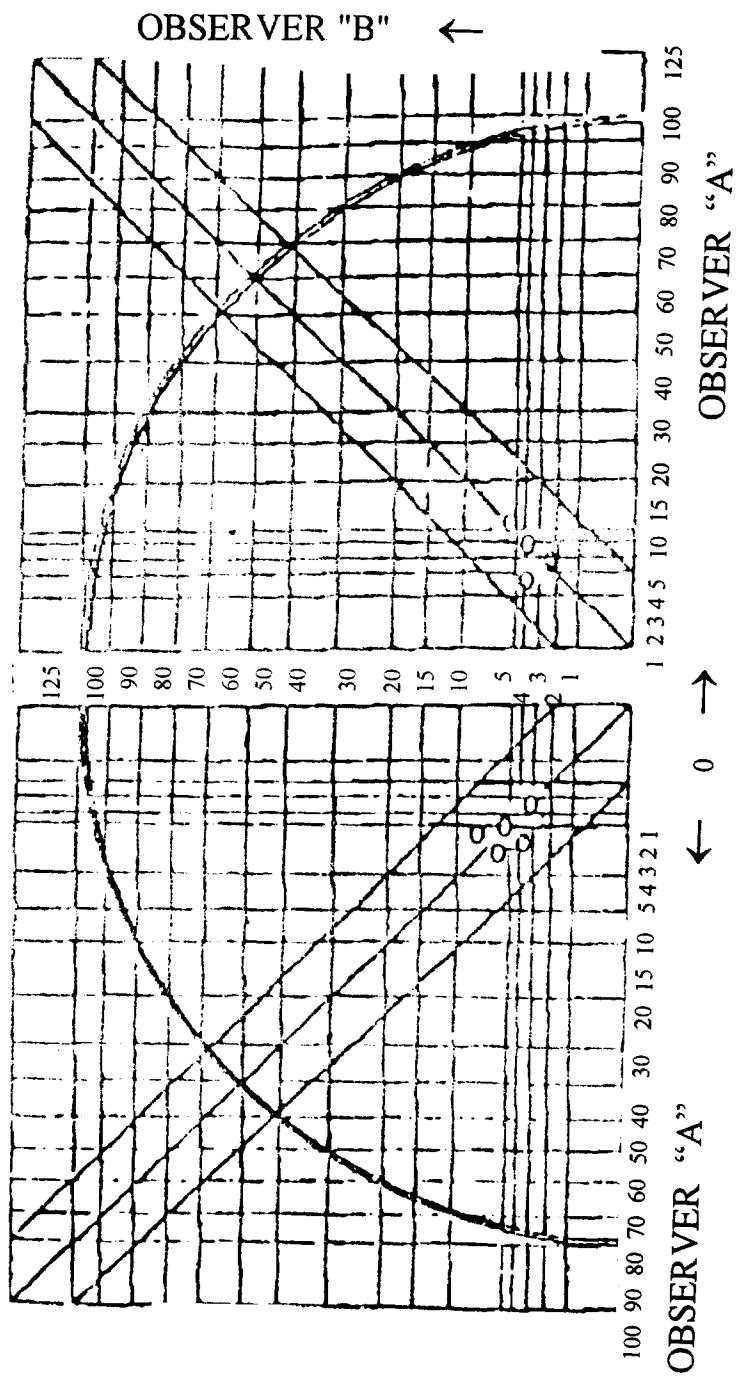
يوضح الاتفاق بين الملاحظين

يوضح الاتفاق بين الملاحظين



يوضح الاتفاق بين الملاحظين
في بند التحرك لأخذ السكان

يوضح الاتفاق بين الملاحظين
في بند أشياء أخرى



يُعرض الاتفاق بين الملاحظين في بند
تقديم المساعدة أو اعطاء المعلومات
في بند الانتظار

يُعرض الاتفاق بين الملاحظين

2-7 ثبات درجات المصححين :

يلاحظ في الاختبارات الرياضية أن تصحيح الاختبار يعتمد إلى حد ما على بعض الجوانب الذاتية للمصحح رغم وجود قواعد يتلزم بها عند التصحيح، ويمكن الحصول على معامل ثبات المصححين عن طريق حساب معامل ارتباط بين درجات مصححين اثنين أو أكثر لنفس الاختبار. وقد يحسب معامل الارتباط بين درجات المصحح نفسه لاختبار واحد عبر فترة زمنية معينة.

2-8 شروط الحصول على ثبات عالي :

هناك بعض الخصائص التي يجب أن تتوفر في الاختبار للحصول على معامل ثبات عال وهي :

2-8-1 عدد فقرات الاختبار :

يمكن الحصول على معامل ثبات عال بزيادة عدد الفقرات ويظهر الجدول الآتي أن معامل الثبات يزداد بازدياد عدد فقرات الاختبار، لانه عندما يكون الاختبار قصيراً نتيجة لقلة عدد الفقرات فيه. فإن أي تغيير في اجابة الفرد على بعض الفقرات لدى اعادة تطبيق الاختبار عليه ثانية، يغير من الدرجة التي حصل عليها في المرة الاولى وهذا بدوره يؤثر على درجة الثبات التي سبق الحصول عليها. بينما يكون تأثير درجة الثبات في الاختبار الطويل، إذا حصل تغيير في إجابة الفرد على بعض الفقرات، أقل من الاختبار القصير لذلك فإن إضافة (60) فقرة إلى اختبار مؤلف من (20) فقرة يرفع من مستوى الثبات من (0.80) إلى (0.89).

ويشترط في الفقرات المضافة أن تتصف بنفس مواصفات الفقرات الأصلية.

جدول يبين العلاقة بين طول الاختبار وثباته:

الثبات	عدد العبارات
0.20	5
0.33	10
0.50	20
0.67	40
0.80	80
0.89	160
0.94	320
0.97	640

٢-٨-٢ تجانس فقرات الاختبار:

يزداد معامل الثبات كلما كانت فقرات الاختبار متجانسة ، ويقصد بالتجانس أن تقيس فقرات الاختبار خاصية واحدة أو موضعًا واحدًا مؤلف من (100) فقرة لقياس الميل الأدبي يكون أكثر ثباتاً من اختبار مؤلف من (100) فقرة أيضاً ولكنه يعطي قياس ميل عديدة ويمكن التعرف على تجانس فقرات الاختبار عن طريق:

أ- إيجاد العلاقة الارتباطية بين إجابة الفرد وإجابته على كل فقرة من الفقرات الأخرى حيث يقوم الباحث بإيجاد العلاقة بين درجة الفرد على الفقرة الأولى ودرجته على الفقرة الثانية ثم بينها وبين درجته على الثالثة وهكذا إلى العبارة الأخيرة ثم يعود ليجد العلاقة بين درجة الفرد على الفقرة الثانية ودرجته على الثالثة فالرابعة وهكذا إلى الأخيرة .

بـ-إيجاد العلاقة بين الإجابة عن كل فقرة وبين الدرجة الكلية للاختبار وذلك على أساس أن الدرجة الكلية تمثل المحتوى السلوكي الذي يقيسه الاختبار ، وأن الفقرة الواحدة تمثل جانباً صغيراً من هذا المحتوى وكلما كانت الدرجة الارتباطية عالية دل ذلك على تجانس الفقرة في قياسها للظاهرة التي يقيسها الاختبار .

جـ- قيام مجموعة من الخبراء بالحكم على صلاحية كل فقرة من فقرات المقياس من حيث مدى علاقتها بالهدف السلوكي الذي يقيسه الاختبار وعلى ضوء تحليل الخبراء تستبعد الفقرات غير المتتجانسة .

وعندما يتضمن الاختبار قياس أكثر من خاصية نفسية أو تربوية فإن ذلك يعني أن مثل هذا الاختبار يتتألف من عدة اختبارات فرعية فلو كان لدينا اختبار لقياس التحصيل في أكثر من موضوع دراسي معين فإنه في مثل هذه الحالة لا بد من حساب الثبات لكل اختبار فرعي يقع ضمن الاختبار الرئيسي .

٤-٨-٣ صعوبة فقرات الاختبار:

يزداد معامل الثبات كلما كانت فقرات الاختبار وكيفية استخدامه واضحة ، ففي اختبارات الاستعداد تكون طريقة إعادة الاختبار هي المفضلة على غيرها من الطرق ويعود السبب في ذلك إلى أن نتائج اختبار الاستعداد تستخدم في التنبؤ البعيد المدى عن مستقبل الطالب التربوي أو المهني ولذلك فإنه من الضروري أن نتعرف على مدى استقرار درجات الاستعداد ويمكن أيضاً استخدام طريقة التجزئة النصفية في اختبارات الاستعداد نظراً لتجانس محتوى فقراتها . ويفضل استخدام طريقة الصور المكافئة في حساب ثبات الاختبارات التحصيلية إذ تميز هذه الطريقة مدى استيعاب الطالب للمعرفة الموجودة في الاختبار التحصيلي ويعتبر معامل الثبات الواطيء دليلاً على أن المعلومات التي لدى الفرد قد تساعده في الإجابة عن سؤال معين من أسئلة أحد الاختبارين ولكنها لا تساعده على الإجابة عن الأسئلة في الاختبار الآخر ويمكن أيضاً استخدام طريقة التجزئة النصفية في

الاختبارات التحصيلية لمعرفة درجة الانسجام في محتوى فقراتها . أما بالنسبة لاختبارات الشخصية ، فإن طرق حساب الثبات تختلف بحسب طبيعة الاختبار ، فإذا أراد باحث أن يستخدم اختباراً للميول المهنية لكي يتبنّأ به عن مدى انتاجية الفرد في مهنة ما ، فلا بد له من استخدام طريقة إعادة الاختبار على أساس أن الميول يجب أن تكون مستقرة . وإذا أراد باحث تصميم مقياس لإحدى خصائص الشخصية فإن بإمكانه استخدام طريقة التجزئة النصفية ، على اعتبار أنه من الضروري معرفة مدى التجانس في الفقرات التي تقسيس ذلك الجانب من الشخصية .

إن استخدام أي من الطرق السابقة يتطلب الحصول على معلومات عن العينة التي تستخدم في حساب الثبات ، من حيث حجمها ومدى تمثيلها للمجتمع الأصلي والمتوسط والانحراف المعياري لدرجات افراد العينة .

٩-٢ العوامل المؤثرة في الثبات:

لكي يكون واضح الاختبار قادرًا على إعداد اختبار ثبات ، فإنه لا بد أن يتعرف إلى العوامل التي تؤثر في ثبات الاختبار أثناء قيامه بإعداد الاختبار.

وهذه العوامل هي:

1- طول الاختبار

يزداد الثبات بزيادة عدد فقراته، على ألا يزيد عدد الفقرات لدرجة تؤدي إلى تعب المفحوص وشعوره بالملل ذلك أن طول الاختبار يتبع فرصة يشموله لمحفوظ المادة الدراسية وأهداف التدريس، وبالتالي، يزيد صدقه والذي يدوره يزيد من معامل ثبات الاختبار.

2 - من الاختبار

إذا لم يكن زمن الاختبار كافياً اضطر المفحوص للتسرع في الإجابة دون رؤية وهذا يقلل من ثباته ولكن إذا خصص وقت كاف للاختبار فإن ذلك يرفع من ثبات الاختبار.

3- مستوى فقرات الاختبار:

الأسئلة الكثيرة الصعوبة أو السهولة على السواء كلاهما يقلل من ثبات الاختبار إذ أن الأسئلة الصعبة قد تجعل المفحوصين يلجأون إلى الغش أو التخمين ويختلف ذلك من تطبيق لتطبيق آخر مما يقلل معامل الثبات . وبالمقابل فإن الأسئلة السهلة تقابل بالاستهان فيجب عنها المفحوصين بدون اكتراث مما يزيد من الخطأ فيها ودون أن يكون هذا الخطأ ناتجاً عن ضعف إمكانيات المفحوص .

4- تجانس المفحوصين:

لو جرب الاختبار على مجموعة متجانسة من المفحوصين فإنه سيظهر كما لو كان ثباته قليلاً ، ذلك أن هذا التجانس يجعل علامات المفحوصين متقاربة فلو أعيد تطبيق الاختبار فإنه ليس من السهل أن تحافظ كل منهم على ترتيبه بين مجموعته مما يظهر الاختبار وكأنه غير ثابت .

أما في حالة عدم التجانس بين المفحوصين فإنهم سيحصلون على علامات بينها مسافات ليست قليلة ، فلو أعيد تطبيق الاختبار عليهم فإنه سيحافظ أكثرهم على ترتيبهم بين زملائهم حتى ولو تغيرت بعض علاماتهم فيبقى معامل الثبات بذلك مرتفعاً .

5- اختلاف طريقة حساب الثبات:

في العادة يكون معامل الثبات المحسوب بالطريقة النصفية أقل من معامل الثبات المحسوب بالطرق الأخرى (لاحظ هنا أن قسم الاختبار إلى نصفين يقلل من طوله مما يقلل من ثباته) ، ولذلك يلجأ عادة في مثل هذه الحالة إلى تصحيح المعامل بتطبيق معادلة خاصة كما ذكرنا سابقاً .

6- صدق الاختبار:

كلما زاد صدق الاختبار زاد ثباته وليس العكس بالضرورة صحيحاً . وعلى العموم

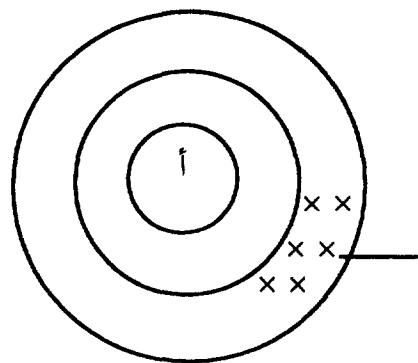
فلو تبين لواضع الاختبار أن معامل ثباته منخفض نوعاً ما فإن بإمكانه رفعه وذلك باتباع الخطوات التالية:

- 1- حذف الفقرات السهلة جداً.
- 2- حذف الفقرات الصعبة جداً.
- 3- زيادة عدد فقرات الاختبار.
- 4- إعطاء الزمن الكافي للاختبار.

2-10 العلاقة بين الصدق والثبات :

يمكن القول أن كل اختبار صادق يكون ثابتاً وليس العكس صحيحاً، أي أن الاختبار الثابت ليس بالضرورة صادقاً، ذلك أنه قد يقيس وظيفة أخرى عدا الوظيفة المخصصة لقياسها، أو وظيفة أخرى إلى جانب الوظيفة التي خصص لها وتوضيح ذلك.

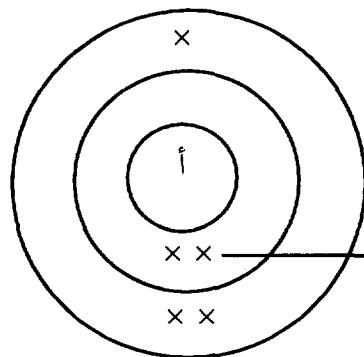
نفترض أن هناك نقطة هدف ولتكن (أ) في الأشكال الثلاثة التالية، وأن هناك تصويباً عليها وأن الإصابات وقعت كما يظهر في هذه الأشكال:



الشكل الأول:

ثبات بدون صدق

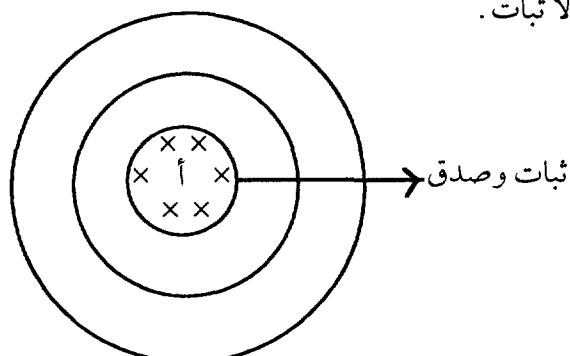
يتضح في هذا الشكل أن الإصابات كلها وقعت بجانب بعضها، الثبات، ولكنها لم تصب الهدف، الصدق، أي أنه يوجد هنا ثبات ولا يوجد صدق.



الشكل الثاني:

لا ثبات ولا صدق

أما الشكل الثاني فنجد فيه أن الإصابات لم تقع بجانب بعضها وأنها مشتتة أي أنه لا يوجد فيها ثبات، كما أنه لم يصيغ الهدف الذي صوبت عليه بمعنى أنه لا يوجد صدق وبذلك لا يتتوفر في هذه الحالة صدق ولا ثبات.



الشكل الثالث:

ويتبين من الشكل الثالث أن الإصابات تجمعت كلها في الهدف (أ) وبذلك توفر فيه الصدق والثبات معاً.

أما عن قبول أو عدم قبول معامي الصدق والثبات فإنه يمكن الاعتماد على ما جاء

في الجدول التالي :

معامل الثبات	معامل الصدق
فأعلى ثبات عالٍ ٠.٧٥	٠.٩٩ - ٠.٨٠ صدق عالٍ
ثبات مقبول ٠.٧٤ - ٠.٥٠	٠.٧٩ - ٠.٥٠ موضع تساؤل
أقل من ٠.٤٩ موضع تساؤل	فأقل غير مقبول ٠.٥٠

2-11 أسس تحديد نوعية الفقرات التي تستخدم في الاختبار:

بعد القيام بتحليل محتوى مادة الاختبار التي سيجري الاختبار فيها، وتحليل أهداف الاختبار لهذا المحتوى وإعداد جدول الموصفات، فإن علينا أن نعد فقرات تقيس مدى تحقيق أهداف الاختبار بشكل مباشر ما أمكن ذلك.

ويجب أن نتذكر أنه ليس هنالك قواعد محددة لاختبار نوعية الفقرات التي سيختارها واضع الاختبار، ماذا يختار؟ خطأ الصواب، الخطأ أم الاختيار من متعدد أم المزاجة أم أسلمة التكميل. وعلى الرغم من عدم وجود قواعد عامة لتحديد نوعية هذه الفقرات إلا أن هناك سمات معينة يجب أن تتوفر في واضع الاختبار حتى ينجح في وضع الاختبار المناسب وأهم السمات:

1- التمكّن من مادة الاختبار:

يجب أن يكون واضع الاختبار متمكنًا من جوانب المادة بشكل كافٍ، بحيث يكون مدركاً لنقط الضعف والقوة فيها.

2- الإيمان بالأهداف التربوية:

إذ من الصعب أن يضع شخص ما اختباراً سليماً، إذا كان ذلك الشخص لا يؤمن أصلاً بأهداف التربية.

3- المعرفة بمستويات المختبرين وقدراتهم: بحيث تكون مفردات الاختبار مفهومة وواضحة من قبل المختبرين وتناسب مستواهم، وتراعي الفروق الفردية بينهم.

4- امتلاك ناصية التعبير والطلاقة اللفظية.

5- المعرفة الواسعة بأمراض الاختبارات وأسس كتابتها.

ومن الجدير بالذكر أن اختيار واضع الاختبار للأسئلة الموضوعية سواء من نوع المزاجة أم الصواب والخطأ أم التكميل أم الاختيار من متعدد، يجب أن يتنااسب مع وظيفة

المادة التي سيعدها الاختبار لها. وهكذا نستطيع القول أن هدف مادة الاختبار وطبيعته يحددان نوعية الفقرات المناسبة إضافة إلى هدف السؤال ذاته، وبصورة عامة يمكن استخدام جميع أنماط الأسئلة مع جميع مواد الاختبارات ولذلك من الأفضل أن تعرف على هدف السؤال نفسه مثال ذلك، إذ أردنا من السؤال أن يذكر لنا المختبر موجزاً عن تاريخ الألعاب الأولمبية، فالأفضل استخدام سؤال مقالى ، أما إذا أردنا من سؤالنا أن ينسب للمختبر توزيع مجموعة من الألعاب ، فيفضل أن نستخدم أسئلة المزاوجة وإذا أردنا أن نتعرف على مؤسس رياضة المعاقين مثلاً فيمكننا استخدام نمط الاختبار من متعدد ، على أن نراعي قواعد الصياغة لهذا النوع من الأسئلة . ومن الجدير بالذكر أن نشير هنا إلى أنه إذا تساوت الظروف فإن أفضل أنواع الأسئلة الموضوعية هو نمط الاختيار من متعدد ، وذلك لأنه يقيس أهدافاً تعليمية متنوعة ومتدرجة من البسيط إلى المركب .

2-12 إعداد تعليمات الاختبار :

يساعد وجود تعليمات واضحة للاختبار على رفع معامل صدق الاختبار وموضوعية ، وثباته . وهذه التعليمات يجب أن تكون بسيطة وموজزة وتتضمن المعلومات المتعلقة بما يلي :

- 1- هدف الاختبار .
- 2- الوقت المحدد للاختبار .
- 3- كيفية تدوين الأجوبة .
- 4- إمكان التخمين عند الشك في الجواب .

وعندما يتضمن الاختبار الواحد نوعين أو أكثر من الفقرات ، يتطلب الأمر وضع تعليمات عامة للاختبار الكلي ثم وضع تعليمات خاصة بكل جزء وفي هذه الحالة يجب أن تتضمن التعليمات العامة معلومات عن هدف الاختبار ، والזמן المقرر ، على أن تبين التعليمات الخاصة أسلوب تدوين الأجوبة الخاصة بالجزء المعنى .

وقد يكون هدف الاختبار معروفاً مسبقاً للمختبر وللمشرف على الاختبار ولكن يفضل مع ذلك ذكر هدف الاختبار في التعليمات، خاصة إذا كان الموقف جديداً، واستخدام مشرف آخر غير واضح الاختبار في مراقبة وإعطاء الاختبار. و بما أننا نعي هنا باختبارات التحصيل فإن مجرد ذكر موضوع الاختبار والصف والمرحلة الزمنية تعبر ضمناً عن هدف الاختبار، وهو قياس التحصيل في الموضوع المعين وللنصف المذكور. وعند تحديد الزمن المتاح للاختبار، فإنه يفضل أن يحدد زمن الاختبار ككل أولاً، ثم الزمن الذي يحتاجه المختبر في كل جزء، وإذا كان يتكون من أجزاء فذلك يساعد المختبر كثيراً في توزيع وقته واستغلاله بشكل فعال. ومن المهم أيضاً أن نبين للمفحوص كيفية تدوين أجابته على فقرات الاختبار، فإذا كان الاختبار يتكون من نوع واحد من الفقرات نذكر طريقة الإجابة في التعليمات العامة مع مثال أو مثالين لتوضيح الطريقة. وإذا تكون الاختبار من نوعين أو عدة أنواع من الفقرات فإن تعليمات الإجابة عن كل نوع والأمثلة التوضيحية لها تسبق النوع الذي تخصه حتى إذا انتقل الطالب من نوع لآخر يقرأ التعليمات الخاصة بالنوع الجديد ثم استمر في الإجابة عن أسئلة ذلك النوع وهكذا. وفي حالة استخدام أوراق إجابة منفصلة يجدر بنا أن نقدم تعليمات مفصلة تتعلق بتدوين الأجوبة، وإذا لم يكن استخدام مثل هذه الأوراق مألفاً لدى المختبرين فيفضل تقديم فقرة مثلاً على الجواب الصحيح مؤشراً عليه.

وإذا كان واضع الاختبار ينوي التصحيح من أثر التخمين فإن عليه أن ينبه في تعليمات الاختبار إلى ضرورة عدم التخمين ولقد قام جدل كبير حول قضية التصحيح من أثر التخمين على علامات الاختبار ، ولكن معظمه أنصب على الاختارات المقنية .

و عموماً إذا أراد واضع الاختبار أن يزيل التخمين أثناء التصحيح فإنه من الضروري
بالتخمين في بعض الأحيان ليس أمراً مكرر و هاماً تماماً من وجهة نظر تربوية (Gronland).
أما اختبارات التحصيل غير الرسمية فلا ينصح بمعالجة أثر التخمين فيها وإن السماح

أن ينبه المختبرين لذلك في تعليمات الاختبار، حيث تطبق لذلك المعادلة التالية في أثناء عملية التصحيح:

$$\text{العلامة} = \frac{\text{عدد الإجابات الصحيحة} - 1}{\text{ن}} \quad \text{الخطأ}$$

حيث (ن) تساوي عدد الأبدال في كل فقرة، وعلى هذا فإن المعادلة في اختبار من نوع الاختبار من متعدد بفقرات ذات أربعة أبدال تظهر كما يلي:

$$\text{العلامة} = \frac{\text{الصواب} - 3}{3} \quad \text{الخطأ}$$

ويستلزم تطبيق معادلة معالجة التخمين حسب عدد الأجوبة الصحيحة وعدد الأجوبة الخاطئة على ورقة الاختبار، وإدراج هذه الأعداد في المعادلة، أما الفقرات التي تركها المختبر فلا تحسب.

فلو فرضنا أن مختبراً أجاب عن (90) فقرة إجابة صحيحة، وعن (6) فقرات إجابة خاطئة وكان عدد فقرات الاختبار (100) فقرة، والاختبار من نوع الاختبار من متعدد ذي الأبدال الأربعة فإن العلامة المعالجة تحسب كما يلي:

$$88 = 2 \cdot 90 = 6 / 3 \cdot 90$$

والافتراض هنا أن المختبر قد خمن على (8) فقرات مفصل على نجاح بالصدفة (يعني صوایین وستة أخطاء) فتقوم المعادلة باستبعاد الجوابين اللذين يمكن اعتبارهما نتيجة نجاح الصدفة الناتج عن التخمين.

الباب الثالث

التحليل الاحصائي لأسئلة الاختبار

3-1 تحليل أسئلة الاختبار.

1-1-3 فوائد تحليل أسئلة الاختبار.

1-2-3 خطوات تحليل أسئلة الإختبار.

2-1-3 معامل السهولة

2-2-3 معامل الصعوبة

2-3-3 معامل التمييز



الباب الثالث

التحليل الاحصائي لأسئلة الاختبار

إن المتخصصين في مجال القياس والتقويم لا يكتفون بهذا القدر من الاعداد في حكمهم وتقييمهم لمصداقية وموثوقية الاختبار، وإنما يقومون بتطبيق ما يسمى بالتحليل الاحصائي للاختبار، ومن ضمن ما يدخل تحت هذا التحليل هو القيام بتحليل اسئلة الاختبار. وتعني جملة تحليل اسئلة الاختبار استخراج ما يسمى بمعاملات السهولة والصعوبة والتميز، وكذلك تحديد فعالية الم Mohamedات أو المشتقات ثم استعمال نتائجة هذا التحليل لتقييم اسئلة بقصد تحسينها إذا وجد بها ضعف في تركيبها أو صياغتها أو للتخلص منها إذا لم تستطع عمل ذلك وعليه تعني جملة تحليل اسئلة بصورة مختصرة تقويم اسئلة الحكم على جودتها أو رداءتها.

3-1 تحليل اسئلة الاختبار :

يحتاج المعلم الماهر والمهم بتحسين وتطوير اختباراته لتطبيق تحليل اسئلة لتحقيق الاهداف التالية :

- 1 - فرز السهل من الصعب من اسئلة الاختبار وتقييم الصالح من الطالع منها حيث تضمن طريقة تحليل اسئلة للكشف عن اسئلة الجيدة ومن ثم الابقاء عليها وكذلك كشف اسئلة الرديئة ومن ثم تحسينها أو تعديلها أو التخلص منها.
- 2 - لتبیان وبویجه عام مدى نجاح أو فشل مجموعة أو فصل الاجابة على سؤال من اسئلة الاختبار كل على حده .

3- يوفر لنا تحليل إجابات عدد كبير من الطلاب على سؤال معين معلومات كثيرة ويعطينا ثقة أكبر في الحكم على جودة السؤال من عدمها. إذ يشير المتخصصون في القياس والتقويم إلى أنه لا يجب أن ينخدع مصمم الاختبار للاسئلة التي يبدو من خلال النظرة العامة إليها أنها جيدة وفعالة بل يجب أن يقوم بعملية تحليلها حتى يتتأكد وبناءً على قواعد علمية مدى صدقها وفعاليتها، فقد يختفي على المتخصصين في طرق تصميم الاختبارات نقائض وعيوب كبيرة في عدد الاسئلة لا تتضح إلا بعد اجراء عملية تحليل الاسئلة.

1-1-3 فوائد تحليل اسئلة الاختبار :

- 1- يعود على هذه الطريقة من حيث أنها تعطي لمصمم الاختبار ما يسمى بالتجذية الراجعة حول فهم وتفكير الطالب وطريقة أخذة للاختبار.
- 2- تعد طريقة تحليل الاسئلة في الاختبارات الموضوعية أحد الوسائل الفعالة والمتميزة في تحسين كفاية ونوعية اسئلة الاختبار.
- 3- تكشف عملية تحليل الاسئلة بطريقة مباشرة على الاسئلة السهلة جداً والأسئلة الصعبة جداً والاسئلة المتوسطة والاسئلة المتميزة وأية اسئلة أخرى بها عيوب لا يسبب من الاسباب .
- 4- تقدم عملية التحليل أحياناً الاجابة على السؤال : لماذا سؤال معين لم يكن فعالاً بالقدر المطلوب والمستهدف منه ، ومن ثم في كيفية تحسينه أو تعديله أو حذفه .
- 5- عادة ما تتمتع الاختبارات المنقحة اسئلتها بواسطة عملية (تحليل الاسئلة) بقدر أكبر من الموثوقية والصدق أكثر من الاختبارات التي لم تجرب عليها هذه العملية .
- 6- ثم أن مصمم الاختبار الذي يختبر ويتحقق اختباراته بطريقة تحليل الاسئلة غالباً ما تتحسن وتتطور مهاراته الفنية والمعرفية في تصميم الاختبارات بسرعة أكثر من غيره الذي لا يعرف هذه الطريقة أو يعرفها ولا يستعملها .

3-1-2 خطوات تحليل الاسئلة :

ما هي الخطوات التي يجب إتباعها لتحليل الاسئلة؟

هناك طرق متعددة يمكن عن طريقها اجراء عملية تحليل الاسئلة من بينها:

1- طريق استخراج معاملات السهولة والصعوبة والتميز.

2- مؤشرات ثبات السؤال وصدقه.

3- منحنيات صافت الزوال

4- طريقة التحليل العاملي.

3-1-2-1 معامل السهولة :

يعرف معامل السهولة على أنه مجموع نسبة المفحوصين في المجموعتين العليا والدنيا الذين أجابوا على السؤال إجابة صحيحة مقسوماً على (2) ويمكن التعبير عنه بالمعادلة التالية:

$$\text{معامل السهولة} = \frac{\text{ن ع ص} + \text{ن ر ص}}{2}$$

حيث تشير م س إلى معامل السهولة، ن ع ص إلى نسبة المفحوصين في المجموعة العليا الذي أجابوا على السؤال إجابة صحيحة، ن د ص إلى نسبة عدد المفحوصين في المجموعة الدنيا الذين أجابوا على السؤال إجابة صحيحة.

ويكون تعريف معامل السهولة على انه عدد الاجابات الصحيحة على السؤال مقسوماً على الاجابات الصحيحة والخاطئة ويعطى بالمعادلة التالية:

$$م س = \frac{\text{ص}}{\text{ص} + \text{خ}}$$

M_S = معامل السهولة .

S_c = عدد الاجابات الصحيحة

X = عدد الاجابات الخاطئة

3-2-2 معامل الصعوبة :

يعرف على أنه مجموعة نسبة المفحوصين في المجموعتين العليا والدنيا الذين أجابوا على السؤال إجابة خاطئة مقسومة على (2) ويمكن التعبير عنه بالمعادلة التالية:

$$M_{Sc} = \frac{N_{X} + N_{Dx}}{2}$$

M_{Sc} = يشير إلى معامل الصعوبة

N_{X} = يشير إلى نسبة عدد المفحوصين في المجموعة العليا الذين أجابوا على السؤال إجابة خاطئة .

N_{Dx} = يشير إلى نسبة عدد المفحوصين في المجموعة الدنيا الذين أجابوا على السؤال إجابة خاطئة .

ويكن تعريفه ايضاً وببساطة على أنه عدد الاجابات الخاطئة على السؤال مقسوماً على عدد الاجابات الخاطئة والصحيحة ويعطى بالمعادلة التالية:

$$M_{Sc} = \frac{X}{S_c + X}$$

حيث تشير :

M_{Sc} = معامل الصعوبة

خ = عدد الاجابات الخاطئة

ص = عدد الاجابات الصحيحة

ومن المعروف أن مجموع معاملي الصعوبة والسهولة للسؤال الواحد لا بد وأن يساوي دائماً الواحد الصحيح وعليه إذا حسب أحد المعاملين لسؤال معين سهل معرفة المعامل الثاني دون اللجوء إلى القيام بعملية حسابه بالمعادلة السابقة حيث يطرح المعامل المعروف التي تم حسابه من الواحد الصحيح فتحصل على المعامل المجهول .

3-1-3 معامل التمييز :

يعني معامل التمييز قدرة السؤال على التمييز بين الأفراد ودليلًا على صدقه، أو بأن السؤال صادق في قياس الشيء المراد قياسه بدليل قوة تميزه بين الأفراد. يعني أن السؤال يكون مميزاً إذا كان قادراً على التمييز بين الطالب الممتاز والجيد والمقبول والضعيف ويمكن تعريفه على أنه النسبة المئوية لعدد الاجابات الصحيحة في المجموعة العليا مطروحاً منها النسبة المئوية لعدد الاجابات الصحيحة في المجموعة الدنيا ويكون التعبير عنه بالمعادلة التالية :

$$م_{ت} = \frac{ن_{ص}}{ن_{دص}}$$

حيث تشير

$م_{ت}$ = إلى معامل التمييز

$ن_{ص}$ = إلى النسبة المئوية لعدد الاجابات الصحيحة في المجموعة العليا .

$ن_{دص}$ = إلى النسبة المئوية لعدد الاجابات الصحيحة في المجموعة الدنيا .

وعليه تكون خطوات عملية تحليل الاسئلة على النحو التالي :

١- بعد القيام بإجراء عملية تصحيح أوراق الاجابة ، رتب أوراق الاجابة ترتيباً تناظرياً متدلياً بالدرجة العليا ومتناهاً بالدرجة الدنيا.

- قسم اوراق الاجابة إلى مجموعتين متساويتين تتكون المجموعة الأولى من اوراق الاجابة ذات الدرجات المنخفضة، فمثلاً إذا كان عدد الطلاب 30 طالباً فقسم اوراق الاجابة إلى مجموعتين عدد كل منهما 15 ورقة اجابة، يحدث أحياناً أن يكون عدد الطلاب في الفصل فردياً 35 أو 41، في هذه الحالة لا يأس من أبعاد أحد الاوراق ول يكن الوسيط وهو في مثالنا هذا رقم 18 في حالة عدد 35 طالباً والطالب رقم 21 في حالة عدد الطلاب 41 حتى يتسعى لنا الحصول على مجموعتين متساويتين عددياً. وقد يحدث أحياناً أيضاً أن نجد في منتصف الدرجات بعد ترتيبها تنازلياً عدداً من الدرجات التي لها قيمة متساوية وفي هذه الحالة يجب أن تحدد عدداً أولأ ثم توزع على المجموعتين العلية والدنيا بالتساوي.

3- عين نسبة عدد الطلاب في المجموعة العليا الذين اجابوا على كل سؤال اجابة صحيحة واشر الى ذلك بالرمز نع ص ويتم ذلك بقسمة عدد الاجابات الصحيحة للمجموعة العليا على عدد الافراد في هذه المجموعة :

$$\text{نوع ص} = \frac{\text{عدد الإجابات الصحيحة في المجموعة العليا}}{\text{عدد أفراد المجموعة الدنيا}}$$

4- عين نسبة الطلاب في المجموعة الدنيا الذين اجابوا على كل سؤال اجابة صحيحة واشر الى ذلك بالرمز ن د ص ويتم ذلك بتقسيم عدد الاجابات الصحيحة في المجموعة الدنيا على عدد الافراد في هذه المجموعة :

$$\text{ن د ص} = \frac{\text{عدد الاجابات الصحيحة في المجموعة الدنيا}}{\text{عدد افراد المجموعة الدنيا}}$$

5- استخراج معامل السهولة بتطبيق المعادلة:

$$م_{ص} = \frac{ن_{ص} + ن_{دص}}{2}$$

6- استخراج معامل الصعوبة بتطبيق المعادلة:

$$م_{ص} = \frac{ن_{ع_خ} + ن_{د_خ}}{2}$$

أو بطرح معامل السهولة من الواحد الصحيح

$$م_{ص} = 1 - م_{س}$$

7- استخراج معامل التميز من المعادلة

$$م_{ت} = ن_{ع_ص} - ن_{دص}$$

مثال تطبيقي

الاسئلة المراد توضيح تحليلها مأخوذه من مادة الاختبارات والقياس.

نفترض أن الفصل الذي يقوم بتعليمه والمكون من ثلاثين 30 طالباً كانت درجاتهم على اختبار في مادة الاختبار والقياس من نوع الخطأ والصواب:

17	9	30	20	14	40
7	26	18	12	36	24
16	13	10	32	2	15
6	8	23	13	13	33
5	25	13	11	34	23

تطبيق عملية تحليل أسئلة الاختبار:

١- ترتيب الدرجات تنازلياً من الأعلى إلى الأدنى على النحو التالي

9	14	18	23	30	40
8	13	18	22	28	38
7	12	17	20	26	36
6	11	16	18	25	34
5	10	15	18	24	32

2- تقسيم الدرجات الى مجموعتين متساويتين العليا والدنيا

الجدول رقم (1) درجات المجموعتين العليا والدنيا على اختبار مادة الاختبارات

والقياس، مر تنازلياً ترتيبياً.

المجموعة الدنيا				المجموعة العليا				
الدرجة	الرتبة	الدرجة	الرتبة	الدرجة	الرتبة	الدرجة	الرتبة	الرتبة
11	24	18	16	25	9	40		1
10	25	18	17	24	10	38		2
9	26	17	18	23	11	36		3
8	27	16	19	22	12	34		4
7	28	15	20	20	13	32		5
6	29	14	21	18	14	30		6
5	30	13	22	18	15	28		7
-	-	12	23	-	-	26		8

3- نفترض أننا اختربنا خمسة اسئلة من هذا الاختبار بغرض توضيح تحليلها ولتكن الاسئلة رقم ، 5 ، 4 ، 3 ،

نوع الاسئلة من نوع الخطأ والصواب .

ضع علامة (صح) أمام العبارة، الصحيحة وعلامة (خطأ) امام العبارة التي تراها خطأ .

جدول رقم (2)

يبين توزيع اجابات المجموعتين العليا والدنيا على الاسئلة الخمسة من اسئلة الوحدة الاولى

الاسئلة	عدد الاجابات الصحيحة	المجموعة العليا	المجموعة الدنيا
السؤال 1	13	6	12
السؤال 2	15	11	3
السؤال 3	12	7	7
السؤال 4	8	3	12
السؤال 5			

4- نستخرج معاملات السهولة والتميز للاسئلة الخمسة بتطبيق المعادلات الواردة في الخطوات 3 ، 4 ، 5 ، 7 كما هو مبين أدناه

السؤال الأول :

أ- معامل السهولة (م س)

$$م س = \frac{13}{15} = 0.866 = 0.87$$

$$\text{ن د ص} = \frac{6}{15}$$

$$0.63 = \frac{1.27}{2} = \frac{0.40 + 0.87}{2} = \frac{\text{ن ع ص} + \text{ن د ص}}{2} = \frac{\text{م س}}{2}$$

ب - معامل الصعوبة (م ص)

$$\text{ن ع خ} = \frac{2}{15}$$

$$\text{ن د خ} = \frac{9}{15}$$

$$0.37 = \frac{0.73}{2} = \frac{0.60 + 0.13}{2} = \frac{\text{ن ع خ} + \text{ن د خ}}{2} = \frac{\text{م ص}}{2}$$

أو من المعادلة : $\text{م ص} = 1 - \text{م س}$

$$\text{إذن } 0.37 = 0.63 - 1 = \text{م ص}$$

ج - معامل التميز (م ت)

$$\text{م ت} = \text{ن ع ص} - \text{ن د س}$$

$$\text{إذن } \text{م ت} = 0.47 = 0.40 - 0.87$$

جدول يبين معاملات السهولة والصعوبة والتفضيل للأمثلة الخمسة

رقم	سؤال	عدد الإجابات الصحيحة	المجموعة	العليا	الدنيا	نوع صن	نوع صن	ن دص	معامل السهولة	معامل الصعوبة	معامل التمييز
1	السؤال للإجابات الصحيحة	13	6	13	1	الدنيا	العليا	ن دص	م	م ص = 1 - م س	م = نوع صن - ن دص
2	0.40 - 0.87 = 0.37 = 0.63 - 1 = 0.47 =	$0.40 = \frac{6}{15}$	$0.87 = \frac{13}{15}$						$0.40 + 0.87 = 2$	$0.37 = \frac{0.40 + 0.87}{2} = 0.63$	
3	0.27	0.87	$0.73 = \frac{11}{15}$	11	15	11	15				
4	0.60	0.50	0.20	0.30	3	12	3				
5	0.60-	0.50	0.47	0.53	7	6	4				
		0.80	0.20	12	3						

تحليل الأسئلة يأخذ الـ 27٪ الأعلى والـ 27٪ الأدنى من الدرجات.

تصلح طريقة تقسيم عدد الأوراق إلى مجموعتين متساويتين في حالة عدد الطلبة في الفصل قليلاً مثل 30 إلى 50، أما إذا كان العدد لمجموعات كثيرة يزيد عن 100 طالب فأكثر، فيوصي المتخصصون في القياس والتقويم بأخذ نسبة مئوية تتراوح بين 10٪ إلى 33٪ من طرفي التوزيع إلا أن كيلي يشير إلى أن الـ (27٪) من المجموعة العليا والـ 27 من المجموعة الدنيا هي أفضل نسبة تحصل بواسطتها على أعلى معاملات تميز، وعادة ما تؤخذ هذه النسبة أيضاً بسبب وجود جداول خاصة مُعدة بناءً على هذه النسبة فيمكن استخراج معاملات السهولة والصعوبة والتميز من خلالها مباشرة.

والآن دعونا نحاول تطبيق طريقة الـ 27٪ الأعلى في مقابل الـ 27٪ الأدنى على مثالنا السابق رغم أن عدد الطلبة في ذلك الفصل لا يزيد عن 30 طالباً لتوضيح مدى تطابق أو اختلاف الطريقتين.

1- رتب الدرجات ترتيباً تناظرياً كما فعلنا في الطريقة السابقة.

2- تقسيم أوراق الإجابة إلى مجموعتين متعارضتين تمثل المجموعة الأولى العليا الـ 27٪ الأعلى وتمثل المجموعة الدنيا الـ 27٪ الدنيا من الدرجات ويتم ذلك بضرب العدد الكلي للطالب (ن) في 0.27 ثم نقرب نتيجة الضرب في حالة وجود كسر إلى أقرب عدد صحيح ونرمز للعدد الناتج من هذه العملية بالرمز (ن) ففي مثالنا السابق تكون:

ن مساوية ل 30

ن مساوية ل 8

$$\text{أي ضربنا } 8.1 = 0.27 \times 30 = 0.27$$

ثم قربنا الرقم 8.1 إلى أقرب عدد صحيح فاصبحت : $n = 8$

3- نبدأ في فرز أوراق الـ 27٪ الأعلى مقابل أوراق الـ 27٪ الأدنى ونضعها في

جدول كما هو مبين أدناه:

جدول رقم (4)

درجات المجموعتين 27٪ الأعلى والـ 27٪ الأدنى في اختبار مادة الاختبار والقياس

الـ 27 الأدنى		الـ 27 الأعلى	
الدرجة	رتبة الدرجة	الدرجة	رتبة الدرجة
12	23	40	1
11	24	38	2
10	25	36	3
9	26	34	4
8	27	32	5
7	28	30	6
6	29	28	7
5	30	26	8

ولنفرض الآن انه بعد فرز الاجابات تبين أن اجابات المجموعتين 27٪ الأعلى والـ 27٪ الأدنى على الاسئلة الخمسة السابقة توزعت على النحو في الجدول التالي رقم (5) :

جدول رقم (5) يبين توزيع اجابات المجموعتين 27٪ الأعلى والـ 27٪ الأدنى

الـ 27٪ الأدنى	الـ 27٪ الأعلى	الاسئلة	عدد الاجابات الصحيحة
			الـ 27٪ الأدنى الـ 27٪ الأعلى
3	7		1
6	8		2
2	6		3
4	4		4
6	2		5

4- نستخرج معاملات السهولة والصعبية والتميز للاسئلة الخمسة بتطبيق المعاملات الواردة في الخطوات 3، 4، 5، 6، 7 في الطريقة السابقة كما هو مبين أدناه:

السؤال الأول:

أ - معامل السهولة

$$\text{ن ع ص} = 0.88 = \frac{7}{8}$$

$$\text{ن د ص} = 0.38 = \frac{3}{8}$$

$$0.63 = \frac{0.38 + 0.88}{2} = \frac{\text{ن ع ص} + \text{ن د ص}}{2} = \frac{m_s}{2}$$

ب - معامل الصعبية :

$$\text{ن ع خ} = 0.12 = \frac{1}{8}$$

$$\text{ن د خ} = 0.62 = \frac{5}{8}$$

$$0.37 = \frac{0.74}{2} = \frac{0.62 + 0.12}{2} = \frac{\text{ن ع خ} + \text{ن د خ}}{2} = \frac{m_{ch}}{2}$$

ج - معامل التميز :

$$m_t = \text{ن ع ص} - \text{ن د س}$$

$$\text{إذن } m_t = 0.5 = 0.38 - 0.88$$

وهكذا مع بقية الاسئلة ، ولتكن تفريغ اجابات الاسئلة الخمسة وطريقة تحليلها كما فعلنا في الطريقة السابقة في الجدول التالي:

جدول يبين معاملات السهولة والصعوبة والتمييز للأسئلة الخمسة السابقة بطريقة ٢٧٪ الأعلى مقابل ٢٧٪ الأدنى

سؤال رقم	المجموعة العليا	المجموعة الدنيا	نوع صن	نوع صن + ندص	معامل السهولة	معامل الصعوبة	معامل التمييز
					$\frac{1}{2}$	$1 - \frac{\text{ندص}}{\text{صن}}$	$M = \frac{\text{صن} + \text{ندص}}{\text{صن}}$ $M = \text{صن} - \text{ندص}$
١							
٢							
٣							
٤							
٥							

وإذا ما قارنا قسم معاملات السهولة والصعبية والتميز بواسطة الطريقتين طريقة تحليل اوراق جميع المفحوصين وطريقة أخذ الـ 27٪ الاعلى مقابل الـ 27٪ الادنى من الاوراق فقط فستجد أن قيم هذه المعاملات قريبة جداً من بعضها كما هو مبين في الجدول التالي:

جدول يبين مقارنة بين الطريقتين الاولى والثانية لاستخراج معاملات السهولة والصعبية والتميز للاسئلة الخمسة.

السؤال	الطريقة الاولى	الطريقة الثانية
	ادخال جميع الاوراق في التحليل الـ 27٪ الاعلى مقابل الـ 27٪ الادنى	
1	$M_{س} = 0.63$ $M_{ص} = 0.37$ $M_{ت} = 0.50$	$M_{س} = 0.63$ $M_{ص} = 0.37$ $M_{ت} = 0.47$
2	$M_{س} = 0.88$ $M_{ص} = 0.13$ $M_{ت} = 0.25$	$M_{س} = 0.87$ $M_{ص} = 0.13$ $M_{ت} = 0.27$
3	$M_{س} = 0.50$ $M_{ص} = 0.50$ $M_{ت} = 0.50$	$M_{س} = 0.50$ $M_{ص} = 0.50$ $M_{ت} = 0.60$
4	$M_{س} = 0.50$ $M_{ص} = 0.50$ $M_{ت} = صفر$	$M_{س} = 0.50$ $M_{ص} = 0.50$ $M_{ت} = 0.60$
5	$M_{س} = 0.50$ $M_{ص} = 0.50$ $M_{ت} = 0.50$	$M_{س} = 0.50$ $M_{ص} = 0.50$ $M_{ت} = 0.60$

ماذا يعني استخراج معامل السهولة والصعوبة والتميز؟

ماذا يعني معامل السهولة والذي قيمته 0.87 وكذلك معامل السهولة والذي قيمته 0.50؟

يعني معامل السهولة الذي قيمته 0.87 أن 87٪ من الطلاب أجابوا على السؤال اجابة صحيحة في حين فشل الاجابة عليه 13٪ منهم.

أما السؤال الثالث أو الرابع أو الخامس فقد اجاب على كل منهم 50٪ من الطلبة وفشل في الاجابة عليهم الخمسين الباقين وقد يحدث أحياناً أن يجيب على سؤال معين كل الطلاب وحينئذ يكون معامل السهولة مساوياً للواحد الصحيح، كما قد يحدث أن لا يجيب على سؤال معين أحد وحينئذ يكون معامل السهولة قد تتراوح قيمته من الصفر إلى الواحد الصحيح إلا أن السؤال الذي يطرح نفسه عند هذه النقطة هو في معرفة القيمة المرغوبة والمفضولة للسؤال الجيد يعني ما هي قيمة معامل السهولة المرغوب؟ يرى البعض أن السؤال الذي يجيب عليه كل الطالب سؤال غير جيد، ويجب حذفه من الاختبار، وكذلك الحال بالنسبة للسؤال الذي لا يجيب عليه أحد وسبب ذلك أن كلا السؤالين لم يعطنا أيه معلومات عن الفروق الفردية بين الطلبة ولم يميز بين الطلاب الجيدين وعكسهم.

س : هل يعني هذا أنه يجب أن تُحذف كل الأسئلة السهلة جداً، مس = 92٪ أو أكثر ، أو الصعبة جداً مس = 20٪ أو أقل مثلاً.

ج : لا يجب الغاء السهلة جداً من أول وهلة إذ يستطيع المعلم الابقاء عليها إذا كان الهدف من ذلك رفع دافعية ومعنوية الطالب بشرط أن تكون تلك الأسئلة في أول اختبار، أو هناك موضوعات لا يتم تقييمها أو تغطيتها إلا بهذا النوع من الأسئلة ، أما الأسئلة الصعبة جداً والتي لم يجب عليها أحد فيستحسن مراجعتها من حيث صياغتها اللغوية مدى وضوحها للتلاميذ أو احتمالية وجود اجابتين صحيحتين للسؤال ، إلى غير ذلك من النواحي الفنية الخاصة بتركيبة السؤال ووضوحه فإذا وجد هذا فيجب اصلاح هذه

الاخطاء ، اما اذا تعذر ذلك فيجب حذف السؤال حيث أن مضاره من الناحية السيكولوجية في هذه الحالة اكثرا من منافعه على عكس السؤال السهل جدا الذي قد تبقى عليه نظراً لفوائده السكلوجية بغرض استعماله للتمهيد للاختبار وكمشط لدافعيه ومعنويات الطلاب وعليه بصورة عامة كلما اقترب معامل السهولة ، والصعوبة من الواحد الصحيح أو الصفر كلما نقصت المعلومات الناتجة عنه بخصوص الفروق الفردية بين المفحوصين ، اما اذا اقترب معامل السهولة أو الصعوبة من الا 0.50 كلما زادت المعلومات أو التميزات في الفروق الفردية بين المفحوصين .

حينما يكون معامل السهولة أو الصعوبة للسؤال = 0.25 تعطينا اكبر عدد ممكن من الفروق والمعلومات بين الطالب $50 \times 50\% = 2500$ فرقاً ، وهكذا مع اي سؤال يقل او يزيد فيه معامل السهولة أو الصعوبة عن 50٪ فإنه سيعطي مقارنات ومعلومات اقل مما اذا كان معامل السهولة أو الصعوبة يساوي 50٪ بالضبط .

الفروق	م ص	م س
0	.0	.100
99	.1	.99
2275	.85	.65
2500	.50	.50
2100	.70	.30
.99	.99	.1
0	.100	.0

وعليه نستخلص مما سبق أن أفضل معامل سهولة هو 0.50 أو القيم القريبة منه ويرى

الن ودين أن معاملات السهولة بين 0.80 ، 0.70 توفر لنا قدرة تمييزية جيدة بين التلاميذ و يمكن قبولها والأخذ بها .

س : ماذا عن معامل التمييز المرغوب ؟

لا يكفي المعلم بحساب معاملي السهولة والصعوبة فحسب بل يجب معرفة نوعية الطلاب الذين اجابوا عن سؤال معين اجابة صحيحة ، هل هم طلبة متازون أم جيدون أو مقبولون ، أم ضعفاء أم أنهم خليط من هذه الفئات الأربع ، فإذا كانت نتيجة معامل السهولة أو الصعوبة لسؤال معين = 48٪ فإن هذا السؤال يعتبر متازاً بسبب قرب هذه القيمة من 50٪ وهي القيمة التي تعطى أكبر قدر من التميزات ، الامر الذي يتطلب فحص نوعية الطلبة الذين اجابوا على هذا السؤال قبل الحكم على هذا السؤال بالجودة أو الرداءة ، فإذا وجدنا أن 12٪ من الطلاب الذين أجابوا على هذا السؤال من الممتازين ، 12٪ من الطلاب الذين أجابوا على هذا السؤال من الجيدين ، 12٪ من المقبولين ، 12٪ من الباقي من الضعفاء فان مثل هذا السؤال رغم تعمقه بمعامل سهولة وصعوبة مرغوبين الا انه من حيث دلالته التمييزية يعتبر رديتاً - حيث انه لم يميز بين هذه الفئات الاربعة المتمايزة في درجات تحصيلها ، ويتبين من خلال حسابنا لمعاملات التميز في مثالنا السابق أن معامل التمييز قد تتراوح قيمته من - 1 إلى + 1 . فقد نحصل على معامل تميز قيمته واحد صحيح حينما يجب كل الطالب في المجموعة العليا على السؤال بينما لا يجب عليه احد من المجموعة الدنيا ، وفي هذه الحالة تكون قيمته مساوية للواحد الصحيح .

$$م ت = ن ع ص - ن د ص$$

$$م ت = 1 - 1$$

وقد يحدث أن يجب على السؤال كل الطالب في المجموعة الدنيا ولا يجب عليه احد من المجموعة العليا ، وفي هذه الحالة تكون قيمته التمييزية مساوية لناقص واحد (- 1)

$$م ت = 1 - 0$$

وقد يحدث أيضاً أن يتساوى عدد التلاميذ في المجموعتين العليا والدنيا على سؤال معين حيث تكون قيمته مساوية للصفر كما هو الحال بالنسبة للسؤال الرابع في مثالنا السابق حيث

$$M_t = 0.50 - 0.50 = 0$$

وعليه يمكن تحديد شروط قبولنا لمعامل التميز المرغوب في ثلاثة نقاط هي :

1- إن يكون موجباً وهذا يعني أن يكون العدد الأكبر من الطلبة الذين أجابوا على السؤال اجابة صحيحة هم من أفراد المجموعة العليا لا العكس.

2- أن لا يكون صفرأ وهذا يعني الا يساوي افراد المجموعتين العليا والدنيا في اجابتهما على سؤال معين كما هو الحال بالنسبة للسؤال الرابع في مثالنا السابق المستخرج بطريقة الـ 27٪ الاعلى مقابل الـ 27٪ الادنى.

3- أن يعطينا أكبر عدد ممكن من التميزات بين الطالب أو أكبر عدد من المعلومات عنهم ، ويحدث هذا حينما تكون القدرة التمييزية للسؤال تساوي 0.50 أو قريباً منها .

يقترح ايبيل (Ebel,1956) بعض القواعد الفعلية للحكم على القيمة التمييزية للسؤال ، والتي بموجبها نقبل معامل التميز أو نرفضه ، وهذه القواعد هي :

معامل التميز	تقييم السؤال
فوق 40٪	السؤال جيد جداً .
30 الى 39٪	السؤال مقبول ولكن من الممكن تحسينه .
20 إلى 29٪	السؤال على الحافة و يحتاج الى تحسين
تحت 19٪	السؤال رديء ويحتاج الى تعديل وتنقیح جديد أو يجب حذفه .

وعليه فإن السؤال الاول في المثال السابق يكون مقبولاً من حيث دلالته صعوبته وقدرته التمييزية ويمكن البقاء عليه في حين ترتفع قيمة معامل السهولة بالنسبة للسؤال الثاني ، وتقل قدرته التمييزية مقارنة بالسؤال فهو يحتاج الى تحسين ، أما السؤال الثالث فيعتبر ممتازاً من حيث مستوى صعوبته وسهولته وقدرته التمييزية ، في حين أن السؤال الرابع رغم انه ممتاز من حيث مستوى السهولة والصعوبة الا انه فاقد تماماً للقدرة التمييزية بين الطلاب حيث كانت (م ت) مساوية للصفر ، أما السؤال الخامس فعلى الرغم من أنه هو أيضاً يتمتع بمستوى صعوبة وسهولة ممتازين الا انه يتمتع بقدرة تمييزية سالبة والتي تعني أن العدد الاكبر من الطلاب الذين أجابوا على السؤال كانوا من المجموعة الدنيا وعليه يجب تنقیح هذين السؤالين إن امكن أو حذفهما نهائياً من الاختبار خاصة اذا كانت عيوب السؤال لا تتعلق بنواعي بنائية .



الباب الرابع

الموضوعية والتقنيين والمعايير

4- الموضوعية

1-1-4 حساب موضوعية الاختبار

2-1-4 طريقة معامل ارتباط الرتب لسبيرمان

3-1-4 طريقة استخدام دلالة معاملات الارتباط

4-1-4 الخطوات المتّعة لاستخدام الجدول

4- التقنيين

4- المعايير

1-3-4 انواع المعايير

1-1-3-4 المعيار الصفي

2-1-3-4 المعيار العمري

3-1-3-4 الرتب المئوية

4-1-3-4 الدرجة القياسية

2-3-4 المعيار الثاني

2-3-4 نشأته و معناه

2-2-3-4 طريقة حساب المعيار الثاني .



الباب الرابع

الموضوعية والتقنيين والمعايير

4-1 الموضوعية :

من العوامل المهمة التي يجب أن تتوفر في الاختبار الجيد شرط الموضوعية والذي يعني التحرر من التحيز أو التعصب وعدم إدخال العوامل الشخصية للمختبر كآرائه وأهوائه الذاتية وميوله الشخصية وحتى تحizه أو تعصبه . فالموضوعية تعني أن تصف قدرات الفرد كما هي موجودة فعلاً لا كما نريدها أن تكون .

إن من أهم صفات الاختبار الجيد أن يكون موضوعياً لقياس الظاهرة التي أعد أصلاً لقياسها ، وإن هناك فهماً كاملاً من جميع المفحوصين بما سيؤدونه . وإن يكون هناك تفسيراً واحداً للجميع وأن لا يكون هناك فرصة لفهم معنى آخر غير المقصود منه ، وعدم تحقيق الموضوعية في الاختبار يؤدي إلى التأثير في صدقه وبالتالي في ثباته ، أي أن موضوعية الاختيار تتحقق باتفاق الجميع وكلما ارتفعت نسبة الاتفاق كلما دل ذلك على موضوعية الاختبار .

إن الباحث الموضوعي يجب أن يلاحظ السلوك نفسه وأن يشمل ملاحظاته أول بأول لكي لا يتأثر استرجاعه بالنسیان أو الخطأ وأن يقدر السلوك متبعاً القواعد نفسها . ولغرض توفير الدقة في كافة التقديرات يجب أن يفترض أن كل درجة يحصل عليها المفحوص تعد تمثيلاً لقدراته الحقيقة متأثراً ببعض الأخطاء التجريبية في عمليات القياس . إن المفهوم العام للموضوعية أن يكون للاسئلة المعنى نفسه عند مختلف أفراد العينة

التي يطبق عليها الاختبار أي أن السؤال لا يقبل التأويل . والمعنى الشائع للموضوعية في مجال الاختبارات يقصد به الاختبارات الموضوعية والتي عادة تمثلها الاختبارات التحريرية والتي تحدد اجابتها عن طريق مفتاح أو نموذج كما في حالة الاجابة عن اسئلة معينة بخطأ أو صواب أو باختبار إجابة سليمة من بين إجابات عديدة . وفي بناء الاختبارات يقصد بالموضوعية درجة التمايز التي يسجل بها مختلف الأفراد نفس الاختبار وبمعنى آخر أن الاختبار التام الموضوعية هو الاختبار الذي لا يثار حوله عدم الموافقة من قبل المتناسفين عند التنفيذ، يعتبر الاختبار موضوعياً إذا قام شخصان أو أكثر وباستخدام نفس الاجهزة والإجراءات وأمكن الحصول على نتائج متشابهة .

إن الاختبارات ذات الموضوعية المرتفعة سوف تكون على درجة عالية من الثبات حيث أن الفرد سوف يتفق مع نفسه أكثر في أداء الاختبارات عملاً لو كان هذا الاتفاق مع شخص آخر . فعند بناء الاختبارات غالباً ما نكتفي بحساب الموضوعية على أساس أنه إذا تمكننا من الحصول على موضوعية محققة للغرض «مرتفعة» فإن الثبات سوف يكون مرتفعاً بصورة آلية .

إن الاختبار قد يكون على درجة عالية من الثبات دون أن يكون له القدر المماطل من الموضوعية وبهذا فإنه يجب أن نراعي عند بناء الاختبارات بذلك اقصى جهد ممكن من أجل موضوعية أفضل حتى يتمكن مدرس التربية البدنية من مقارنة نتائج طلابه مع النتائج التي يمكن لزميله الحصول عليها سواء في مدرسة أخرى أو على المستوى المحلي والدولي .

يمكن أن نحكم على درجة موضوعية الاختبار بایجاد معامل الارتباط بين الدرجة النهائية التي يطبقها حكمان مستقلان كل منهما عن الآخر ومن الطبيعي أنه كلما كان الاختبار يعتمد على تقدير اعتبرادي كلما كان اتفاق أقل .

تبسيط الاختبارات بدرجة موضوعيتها ، فالموضوعية الكاملة في الاختبار هي أن يتوصل كل ملاحظ أو حكم أداء الفرد على الاختبار إلى التسليمة نفسها في تقديره الذي

يقدمه عن ذلك الفرد، وللقيام بهذه المهمة، عليه أن يتتبّع إلى نفس الجوانب من الأداء، وأن يدون ملاحظاته لكي يزيل الاخطاء التي يمكن أن يقع فيها فيما لو اعتمد على ذاكرته، وأن يصحح الاجابة على الاختبار باستخدام نفس القواعد.

إن موضوعية إجراءات التطبيق يجب أن يحكم عليها بواسطة درجة الاتفاق بين الدرجة النهائية التي يقدمها ملاحظان مستقلان عن بعض أو أكثر وكلما كانت الملاحظة والتقييم ذاتين Subjective كلما انخفضت درجة الاتفاق بين الحكمين.

إن الاختبارات التي يختار فيها المفحوصون البديل الصحيح أو البديل الأفضل من بين عدة بدائل (مثل صواب .. خطأ، الاختيار من متعدد) يطلق عليها «الاختبارات الموضوعية» objective tests ، لأن بامكان المصححين كلهم استخدام مفتاح التصحيح Scoring Key والاتفاق على النتائج اتفاقاً كاملاً. وعلى النقيض من ذلك ، فإن اختبارات المقال الاعتيادية تفسح المجال أمام الاختلاف الواسع بين المصححين. ويمكن بواسطة التعليمات الدقيقة للملحوظ للملحوظ أن تصبح اختبارات الاجابة - free response tests ، و اختبارات المقال والملاحظات ، موضوعة إلى حد مقبول .

١-١-٤ حساب موضوعية الاختبار

تعتبر صفة الموضوعية بالنسبة للاختبار من الاسس العلمية للاختبار الجيد ، عرفنا الموضوعية في مجال القياس بأنها عبارة عن تطابق الآراء لأكثر من جيد عند تقويمه للاختبار .

٤-١-١-١-١ طريقة معامل ارتباط الرتب لسيير مان

مثال (1)

بعد الانتهاء من إحدى مباريات الدوري العام بكرة القدم، وضع أحد الباحثين اختباراً خاصاً لتقدير ناحية تكتيكية معينة للمهاجمين للفريق الذي يتسمى إليه، وعرض

الباحث مفردات الاختبار على خبريين من ذوي الاختصاص شاهدًا للمباراة، وقد أعطيا
 تقويمهم كما يلي: في الجدول التالي:

اللاعبين	درجات الخبر الأول	درجات الخبر الثاني	الاول	الثاني	رتب	ف	ف2
1	15	16	3.5	2	1.5	2.25	
2	14	12	5	7	2-	4	
3	10	11	8	8	صفر	صفر	
4	15	13	3.5	5	1.5-	2.25	
5	9	10	9	9	صفر	صفر	
6	12	13	6	5	1.-	1	
7	18	19	1	1	صفر	صفر	
8	6	7	11	10.5	.5	0.25	
9	8	7	10	10.5	.5-	0.25	
10	11	13	7	5	2	4	
11	17	15	2	3	1	1	
(11) مج			66	66	5+	5-	15

وحساب موضوعية هذا الاختبار يطبق قانون العلاقة، معامل ارتباط الرتب
 لسييرمان.

$$R_t = \frac{1}{n(n^2 - 1)} - \frac{6}{M^2}$$

$$R_t = \frac{15 \times 6}{(1 - 121)11} - 1$$

$$R_t = 1 - \frac{90}{1320} = 0.932 - 0.068$$

ولحساب دلالة موضوعية هذا الاختبار وبالكشف تحت المستوى 0.01 من جدول دلالة (11-2) معاملات الارتباط . . . وجد أن هناك دلالة معنوية عالية . أي أن الاختبار الذي وضعه الباحث لتقويم التكتيكي يوحد بنتائجها وبدرجة عالية من الثقة .

مثال (2)

أرادت إحدى الباحثات حساب معامل موضوعية اختبار قياس مرونة الكتفين ولذلك استعانت الباحثة بتقويم خبيرين من ذوي الاختصاص وأوجدت العلاقة بين التقويمين إحصائياً باستخدام معامل ارتباط الرتب لسبيرمان وكالاتي :

ف2	ف	الرتب	درجات الخبر الثاني	الرتب	درجات الخبر الاول	عدد أفراد العينة
9	3	2	7.6	5	7.8	1
16	4-	6	7	2	9	2
صفر	صفر	8	5.3	8	7.2	3
1	1	3	7.5	4	8.2	4
صفر	صفر	9	5	9	6.9	5
1	1-	10	3.6	11	4	6
صفر	صفر	1	9.5	1	9.5	7
2.25	1.5-	4.5	7.3	3	8.6	8
1	1-	11	7	10	4.7	9
صفر	صفر	7	5.9	7	7.3	10
2.25	1.5	4.5	7.3	6	7.5	11
32.5	6.5+	66		66		مج(11)
	6.5-					

$$R_t = \frac{1 - \frac{2}{n}}{(1 - \frac{2}{n})^2}$$

$$\frac{32.5 \times 6}{(120)11} - 1 = \text{ر ت}$$

$$.15 - 1 = \frac{195}{1320} - 1 = ر ت$$

$$0.85 = \text{ر} \cdot \text{ت}$$

ويستخدم جدول دلالة معاملات الارتباط تحت مستوى 0.01 وجد أن الاختبار موضوعي وبدرجة عالية من الدلالة، مما يؤخذ بنتائجها وبدرجة عالية من الثقة.

٤-١-٢-٣ طريقة استخدام دلالة معاملات الارتباط:

عندما نحصل على معامل الارتباط بين ظاهرتين أو أكثر، فأول شيء يدعونا للتساؤل هو :

هل هذا الارتباط معنوي أم لا؟ أو بصورة أخرى هل العلاقة بين تلك الظواهر أكيدة، وبأي درجة فهناك مستويات يمكن الكشف تحتها 0.05، 0.01 وكما هو موضح في الجدول التالي، «جدول دلالة معاملات الارتباط».

٤-١-٢ الخطوات المتبعة لاستخدام الجدول

حساب درجة الحرية وهي عبارة عن حجم العينة ناقص (2)

أي، أن درجة الحرارة = $n - 2$ ،

فمثلاً عينة حجمها 35 نكشف عند درجة حرية α - 2 أي عند درجة حرية

$$.33 = 2 - .35$$

نظر أمام درجة الحرية (33) وكما هو مبين بالجدول وتحت المستويين 0.05، 0.01 ثم
نقارن بين ما حصلنا عليه من ارتباط وبين ما هو موجود فعلاً فإذا كان متتساو أو أكبر ،
فيكون هناك ارتباط ودال معنوياً . . .

– إذا كان الدلالة تحت مستوى 0.05 هذا يدل على أن هناك ارتباط بدرجة ثقة 95% . Significant

- إذا كانت درجة الارتباط التي حصلنا عليها أقل مما هو موجود بالجدول عند درجة الحرية وتحت إحدى المستويين 0.01 ، 0.05 فسيكون الارتباط غير دال معنِّياً ولا يُخذ به.

مثال

حصل أحد الباحثين على معامل ارتباط قدره 0.381 بين مستوى اذكاء ومستوى القوة العضلية لبعض المجموعات العضلية لاحد شعب احدى الصنوف والبالغ عددهم 25 طالب.

$$أولاً: درجة الحرارة = n - 25 = 2 - 23$$

ثانياً: الدلالة عند درجة حرية 23 وتحت المستوى 0.05 كما هو مبين بالجدول تساوي 0.396 أي أن الارتباط غير دال تحت 0.05.

ثالثاً: وحيث أنه غير دال تحت مستوى 0.05 فيكون غير دال أيضاً تحت مستوى 0.01.

جدول دلالة معاملات الارتباط

الدلالة		درجة الحرية (ن - 2)	الدلالة		درجة الحرية (ن - 2)
.01 عند	.05 عند		.01 عند	.05 عند	
.470	.367	27	1.000	.997	1
.463	.361	28	.990	.950	2
.456	.355	29	.959	.878	3
.449	.345	30	.917	.811	4
.418	.325	35	.874	.754	5
.393	.304	40	.834	.707	6
.372	.288	45	.791	.666	7
.354	.273	50	.765	.632	8
.325	.250	60	.735	.602	9
.302	.233	70	.808	.576	10
.283	.217	80	.684	.553	11
.267	.205	90	.611	.532	12
.254	.195	100	.641	.514	13
.228	.174	125	.623	.497	14
.208	.159	150	.606	.482	15
.181	.128	200	.590	.468	16
.148	.113	350	.575	.456	17
.128	.098	400	.561	.444	18
.115	.088	500	.549	.433	19
.081	.062	100	.537	.423	20
			.521	.413	21
			.515	.404	22
			.505	.396	23
			.496	.388	24
			.487	.374	25
			.478	.374	26

Standardization ٤-٢ التقنين

يتضمن التقنين تحديد شروط تطبيق الاختبار تحديداً دقيقاً تبعاً لمبدأ مراعاة ضبط جميع العوامل التي تؤثر في الظاهرة التي تبحث فتووضع تعليمات الاجراء والتصحيح وتذكر المعايير.

إن التقنين هو اختبار حددت إجراءات تطبيقه وأجهزته وتصحیحه بحيث تكون من الممكن اعطاء الاختبار نفسه في أوقات وأماكن مختلفة ويكون من عدد من الوحدات وسبق تطبيقه على عينات مماثلة للمجتمع الاصلي الذي قصد أن يكون الاختبار له بغرض تحديد معايير واضحة له ويسهل طريقة لتطبيق الاختبار ومحتواه تطبيق الاختبار نفسه على أفراد في أماكن وأوقات مختلفة.

وبهذا يمكن مقارنة درجة فرد ما في اختبار مقنن بدرجات أفراد آخرين يجري عليهم الاختبار نفسه، وتزود بعض الاختبارات بجدوال للمعايير تحدد الدرجة التي يحصل عليها الأفراد والاختبارات التي تتضمن على مثل هذه المعايير تدعى الاختبارات المقنتة-objective tests. وهي اختبارات اعطيت من قبل العديد من العينات Samples أو المجموعات تحت ظروف معينة وأشتقت له معايير.

الاختبارات المقنتة هي التي يقوم باعدادها خبراء في القياس والتي تتيح الفرصة لاستخدام طرق وأدوات للحصول على عينات من السلوك باستخدام إجراءات منتظمة ومتقدمة والتي تعنى أن محتوى الاختبار يطبق طبقاً للتعليمات نفسها وطبقاً للتقويم المحدد لللاء. كما أن القيمة الأساسية للاختبارات المقنتة تكمن في اعتبارها كأدوات بحث في استخدامها كوسائل للمقارنة. والاختبارات المقنتة درجة عالية من الموضوعية فأدائها يتم بطريقة موضوعية بسيطة خالية من الغموض ولا تحتمل التأويل ومع هذا فيمكن أن يكون للاختبارات المقنتة عيوب. فقد يكون معامل ثباتها منخفضاً وكذلك معامل صدقها.

إن الاختبار المقنن هو اختبار قد حددت إجراءات تطبيقه، وأجهزته، وتصحیحه

بحيث يصبح من الممكن اعطاء الاختبار نفسه في اوقات وأماكن مختلفة وتزود بعض الاختبارات بجداول للمعايير Norms تحدد الدرجة التي يحصل عليها الافراد. والاختبارات التي تحتوي على مثل هذه المعايير تدعى في بعض الاحيان «الاختبارات المفينة»، وتدعى عملية جمع البيانات لاستناد المعايير «التقني» وقد يكون للاختبار جدول للمعايير، إلا أن اجراءات تطبيقية قد تكون غير محددة، كما أنه قد يكون للاختبار اجراءات تطبيق مفيدة تقنياً جيداً، إلا أنه لم تشتمل على المعايير، ومن الواضح أن استخراج المعايير لا يكون دقيقاً ما لم تقنن الاجراءات تقنياً جيداً. وتحتختلف الاختبارات من حيث مدى اكتمال تقنيتها، لذلك يجب تحديد كل الشروط المؤثرة في أداء المحبب على الاختبار، إذا أريد للختبار أن يكون مقتناً بصدق. وعندما يكون التقني في تمام الفاعلية ، فإن الفحوص يحصل على نفس الدرجة تقريباً إذا أعيد عليه الاختبار بغض النظر عن يجري الاختبار. وهذا غير ممكن تماماً لظهور بعض الصعوبات التي تنشأ بسبب طبيعة العلاقة بين المحبب ومن يعطيه الاختبار.

4-3. المعايير Norms

إن مصطلح المعيار يشير إلى متوسط درجات جماعة معينة من الافراد على أحد الاختبارات . ويطلق على هذه الجماعة من الأفراد اسم (الجماعة المعيارية) Norm group . والمعيار - عادة - ضروري في الاختبار الرياضي أو التحصيلي ، لأن درجة الفرد التي يحصل عليها في الاختبار (وتسمى الدرجة الخام) ليس لها معنى بحد ذاتها ، ولا تصلح للمقارنة مع درجته في اختبارات أخرى ، أو مع درجة شخص آخر على نفس الاختبار ، أو على اختبارات أخرى ، إلا بواسطة المعايير.

فالمعايير إذن مهمة ، لأنها تخبرنا عن كيفية اداء الآخرين على الاختبار ، فتوفر بذلك أساساً للمقارنة .

فلو أن شخصاً حصل على الدرجة (15) في اختبار السرعة ، فإن هذه الدرجة لا

تعني شيئاً بحد ذاتها، فهي لذلك غير قابلة للتفسير، ولكننا لو عرفناا إن الدرجة الكلية في الاختبار هي (20)، وهي أقصى درجة يمكن أن يحصل عليها أفضل المفحوصين، فإن الدرجة (15) في هذه الحالة تمثل (75 بالمائة) من الدرجة الكلية (20). ولكن تبقى الدرجة (15) غير واضحة، لأنها تخبرنا فقط على أن هذا الفرد يتلذق قدرة على السرعة تساوي 75 بالمائة من السرعة التي يقيسها الاختبار. لكن معناها يصبح أكبر إذا علمنا أن هذا الشخص أفضل من متوسط جماعته بدرجة واحدة، فمستواه في السرعة إذن هو فوق الوسط بقليل، وأن متوسط الجماعة (14) هنا هو (المعيار) الذي اعتمدناه أساساً للمقارنة.

ويجب أن لا نخلط بين المعيار والمستوى القياسي Standard أن المعيار يخبرنا عن الاداء الحقيقي للأفراد على الاختبار، أي أنه يمثل الدرجات الفعلية التي حصل عليها الأفراد في الاختبار. أما المستوى القياسي ، فإنه يشير الى المستوى الذي يجب أن يصله اولئك الافراد لكي تعتبر اجاباتهم مقبولة ، فمثلاً أن الدرجة (100) تمثل المستوى القياسي للصحة النفسية في اختبار يقيس القلق ويتألف من 100 فقرة ، وهذا ممكن من الناحية النظرية . غير أننا في الحياة الواقعية لا نجد شخصاً ليس لديه شيء من القلق ، لذلك نجد أن المعيار قد يكون في مثل هذا الاختبار (80) ، وبناءً على ذلك ، فإن الشخص الذي يحصل على درجة 80 فأكثر نعتبره سوياً لأنه يشبه أكثرية الجماعة التي يتمي إليها . ولكي تكون معايير الاختبارات دقيقة ، فإنها يجب أن تقوم على اساس درجات عينات كبيرة ومثلثة من الاشخاص الذين بني الاختبار لاجلهم ، وأن تكون شروط تطبيق الاختبار عليهم موحدة ، وأن تكون أجاباتهم على الاختبار إجابة جدية بنفس المستوى الذي يتوقع من الاشخاص الذين سوف يطبق الاختبار عليهم . غير أن هناك بعض الحالات التي لا يكون فيها استعمال المعايير في الاختبارات ضرورياً ، خاصة عندما يريد أحدهم تشخيص الفروق الفردية داخل جماعة معينة . فعلى سبيل المثال ، تكون المعايير ذات فائدة قليلة لمدير إحدى المصانع الذي يرغب في استخدام أفضل عشرة من بين جماعة من المتقدمين إلى العمل عنده ، إذ إن بامكانه ترتيبهم من الأعلى إلى الأدنى حسب الدرجات التي حصلوا عليها ، ثم اختبار العشرة الأوائل منهم .

٤-٣-١ أنواع المعايير :

Crade Norm ٤-٣-١-١-١- المعيار الصفي

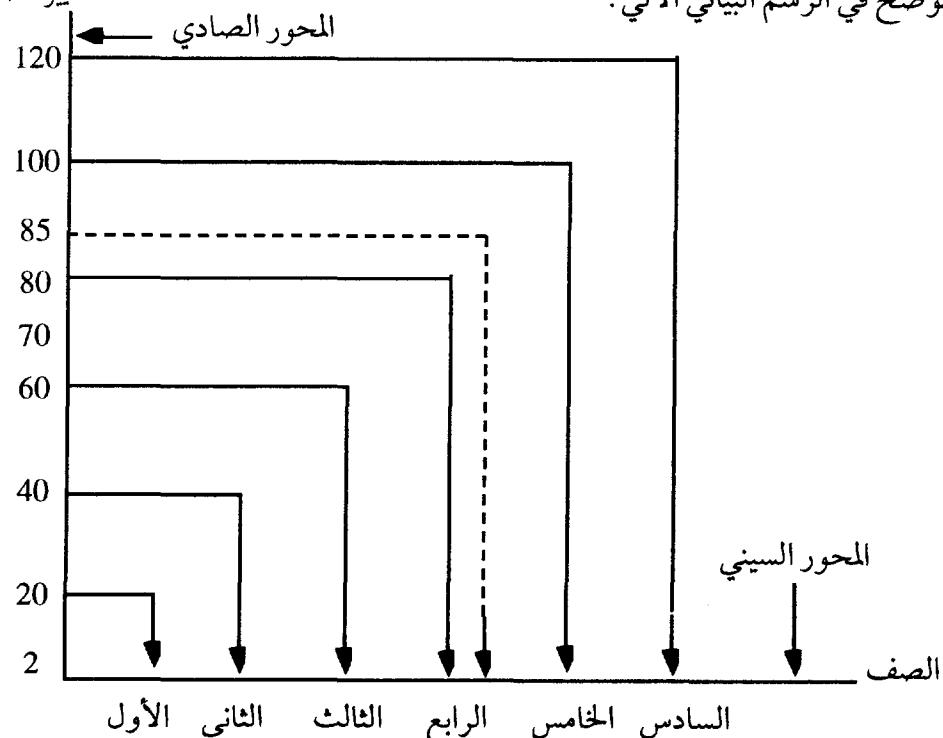
يعتبر المعيار الصفي من الطرق الواسعة الاستخدام في تفسير درجات الأفراد في الاختبارات وخاصة الاختبارات التحصيلية . والمعيار الصفي لأية درجة ما هو إلا المستوى الذي يمثله متوسط تحصيل الطلاب في ذلك الصف ، فلو أن طالباً حصل على الدرجة (65) في اختبار الاحصاء وصادف إن كان متوسط درجات الطلاب في الصف الرابع في نهاية السنة يساوي (65) في الاحصاء ايضاً ، فعندئذ يقال أن هذا الطالب قد حصل في الإحصاء على مستوى الصف الرابع .

خطوات حساب المعيار الصفي :

- 1) نطبق الاختبار الذي نريد حساب المعيار الصفي له على عينات كبيرة من الطلبة في صفوف مختلفة في مستواها الدراسي (ثالث، رابع، خامس .. الخ) ويعطى الاختبار إلى جميع الطلاب في وقت واحد.
- 2) يحسب وسط أو متوسط تحصيل طلبة كل صف.
- 3) نقوم برسم مخطط بياني ، يمثل فيه المحور السيني الصدوف ويمثل فيه المحور الصادي متوسط درجات اطلاب في الاختبار .
- 4) نرسم خطأ بيانياً يمر بال نقاط الممثلة للمتوسطات أو قريباً منها .
- 5) يمكن استخدام هذا الخط البياني في التعرف على المعيار الصفي للطالب .

ولو أن الطالب حصل على الدرجة (85) في اختبار اللغة العربية ، وأردنا معرفة مستوى الصفي ، نقوم برسم عمود من الدرجة (85) الواقعة المحور الصادي ، فيقطع هذا العمود الخط البياني في نقطة ما ، ثم نقوم باسقاط عمود من هذه النقطة على المحور السيني ، فيلقى به في نقطة تمثل مستوى الصفي .

ونجد في هذا المثال الافتراضي أن مستوى الصفي يقع بين الرابع والخامس كما هو موضح في الرسم البياني الآتي:



رسم بياني

(يوضح كيفية استخراج المستوى الصفي لدى معرفتنا درجة الطالب ومعايير الصفة)

ونظراً لبساطة هذا المعيار وسهولة فهمه، فقد شاع استخدامه بين المعلمين في البلدان الأجنبية، وخاصة في المرحلة الابتدائية. ورغم إنتشار استخدام المعيار الصفي ، إلا أنه ليس دقيقاً في تقسيم الطلاب وخاصة في الصفوف المتقدمة من الدراسة الابتدائية أو المتوسطة.

كما أنه لا يجوز استخدامه مع الطلبة في المرحلة الاعدادية ، وكثيراً ما يساء تفسير

هذا المعيار عندما يؤكّد بعض المعلمين الذين يستخدمونه على ضرورة وصول كل طالب إلى الدرجة المتوسطة فما فوق لكي يعتبروا من الناجحين.

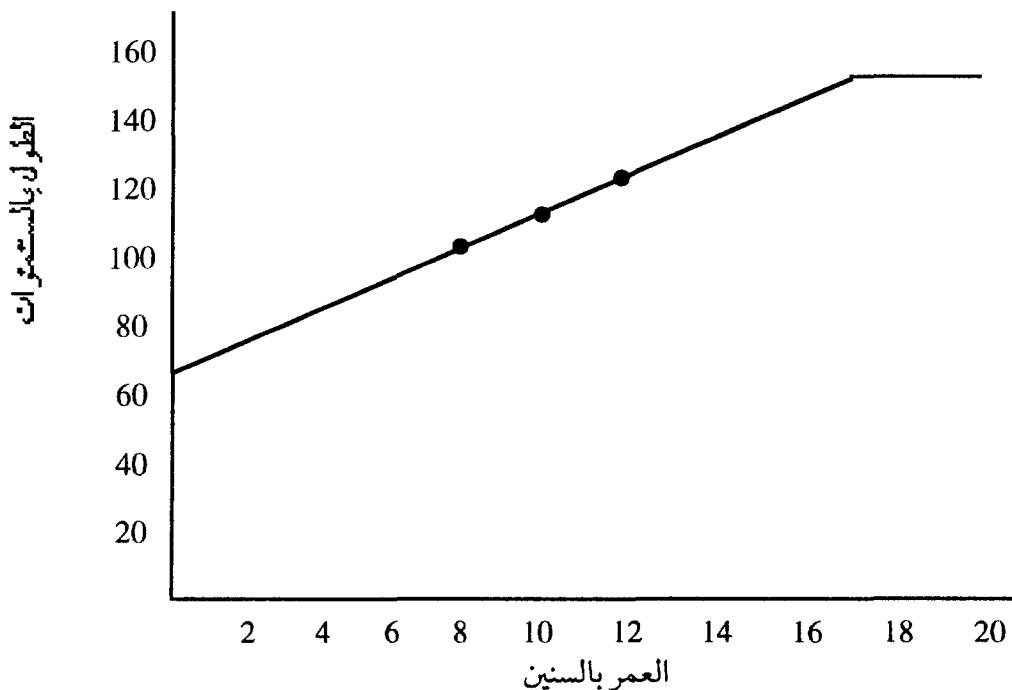
إن مثل هذا الافتراض غير صحيح، لأن هذه المعايير هي متوسطات ، ومن الطبيعي أن يكون هناك بعض التلاميذ الذين لا تسمح له قدراتهم من الوصول إلى المتوسط ، فلا يمكن من الناحية التربوية أن يطلق عليهم «المتأخرن دراسياً».

Agenorm 3-2-1.4 المعيار العمري

يمكن تهيئه مجموعة من معايير العمر لأية خاصية سلوكية تنمو بنمو الفرد ، ويمكن حساب معايير العمر بنفس الطريقة التي حسبت فيها المعايير الصافية. إذ يعطي الاختبار الذي يراد تقسيمه إلى مجموعة كبيرة جداً من الأفراد تمثل فيها الأعمار المختلفة تمثيلاً جيداً، ثم يحسب المتوسط لكل سن ، ويرسم بعد ذلك منحنى بياني بحيث تمثل الأعمار على السيني ، وتمثل متوسطات الدرجات في الاختبار على المحور الصادي ويرسم خط بياني يمر بال نقاط الممثلة للمتوسطات.

إن المعيار لأي عمر من الأعمار هو متوسط درجات أفراد ذلك العمر في الصفة السلوكية المراد قياسها ولنضرب مثلاً على ذلك بالنسبة للطول.

فلو أخذنا عينة ممثلة للبنات في عمر (8) سنوات وقاسينا طول كل واحدة منهم ، ثم حسبنا معدل أطوالهن فإننا نستطيع تحديد معيار الطول لتلك المجموعة العمرية ثم نحسب معدل الطول بنفس الطريقة للبنات في عمر (9) سنوات و (10) سنوات بقية الأعمار الأخرى ، ثم نعين النقاط الممثلة للمتوسطات على المنحنى وهكذا فإنه في الشكل أدناه يتناسب طول الفتاة البالغ (150) سم مع العمر (12) سنة . في حين يتناسب طول الفتاة البالغ (125) سم مع العمر (8) سنوات وبهذا نستطيع تحديد عمر الفتاة بواسطة طولها ، وبالعكس . وهكذا يمكن القول بأن عمر الفتاة التي يبلغ طولها (150) سم هو (12) سنة ، وذلك لأن الفتاة في عمر (12) سنة يكون متوسط طولها (150) سم.



إن معرفتنا لطول الفتاة وعمرها يمكننا من معرفة ما إذا كانت هذه الفتاة طويلة بالنسبة لعمرها، أم أنها اعتيادية (وسط) أو دون الوسط ، وعلى هذا الاساس ، فإن الفتاة التي يبلغ طولها (138) سم وعمرها (8) سنوات ، تكون طويلة بالنسبة لعمرها ، وذلك لأن طولها يتناسب مع العمر (10) سنوات .

ويعتبر المعيار العمري مناسباً للصفات السلوكية التي تظهر استمرارية ونمو مطرد بتقدم العمر ، ورغم منافع المعيار العمري إلا أنه لا يخلو من المساوى ، فهذا النوع من المعايير يفترض أن الصفات والخصائص تنمو باتساق في مختلف الأعمار . غير أنه يحصل في بعض الأعمار أن يتباطأ النمو في بعض الصفات إلى أو طأ درجة وقد يتوقف ، ففي

الشكل السابق نرى أن التباطؤ في الطول يحدث بصورة مفاجئة بعد عمر (14) سنة، كما يحدث الأمر نفسه في بعض الخصائص النفسية، فمثنيات النمو العقلي تظهر تطوراً نسبياً بما هو موضح بالشكل السابق، إذ أنه بعد عمر (14) سنة أو (15) سنة تصبح الزيادة في النمو العقلي أقل إلى أن تلاشى. ويواجه مصممو الاختبارات مشكلات عملية كبيرة في المعايير العمرية، إذ يصعب في الغالب الحصول على عينة ممثلة للافراد في عمر معين، فعند الحصول على افراد في عمر (12) سنة، يتطلب الأمر البحث عن بعضهم في المرحلة الابتدائية والبعض الآخر في المرحلة المتوسطة. وقد يتوفرون في صفوف مدرسية عديدة، وكلما اردننا الحصول على اعمار أكبر، فإن الأمر يتطلب منا الحصول عليهم من المدارس والكليات وأماكن العمل حيث يتواجدون. وعلى العموم، يمكن القول أن الخصائص البدنية كالطول والوزن وظهور الأسنان، والخصائص النفسية كالذكاء العام يمكن أن ينطبق عليها المعيار العمري، أكثر من غيرها من الصفات الأخرى.

4-3-1-3 الرتب المئينية Percentage ranks

الرتبة المئينية، هي قيمة تدلنا على النسبة المئوية من الدرجات في التوزيع التكراري التي تقع دون درجة معينة. فإذا كانت الدرجة (130) في اختبار لقياس الدافعية نحو الانجاز تقابل المئيني (90)، فإن هذا يعني أن صاحب هذه الدرجة (130) يقع دونه (90) بالمائة من زملائه. ونستنتج من هذا بأن «المئيني» نقطة معينة في التوزيع التكراري للدرجات وأن هناك 100 نقطة مثلها نسميها المئينيات تقسم هذا التوزيع التكراري إلى 100 قسم، ومن هنا أتت تسمية (المئيني) بهذا الاسم. ولا يوضح كيفية حساب المئينيات، نضرب المثال الآتي، وهو مثال افتراضي للدرجات طلاب في اختبار لقياس الرشاقة.

(درجات التلاميذ في اختبار الرشاقة (مرتبة من الأعلى إلى الأدنى)).

الدرجة	الدرجة	الدرجة	الدرجة
40	55	75	95
39	54	72	93
36	53	70	93
35	53	68	92
34	50	66	88
34	49	64	86
31	49	64	83
31	47	63	83
29	45	61	82
24	45	59	78
21	42	58	77
19	42	55	76
17			

إن الجدول السابق يحتوي على (50) درجة تمثل (50) طالباً، ولنفترض أننا نريد معرفة ما هي الدرجة التي تقابل المئيني رقم (80)، أي ما هي درجة الطالب التي يقع دونها 80 بالمائة من درجات زملائه الطلبة الآخرين.

نعود الآن إلى جدولنا ونقول لو أنه كان لدينا 100 طالب ، فإننا في هذه الحالة نرتبعهم حسب درجاتهم من الأعلى إلى الأدنى ، ثم نعد الطلاب من (1) إلى (80)، ونتوقف عند الطالب رقم 80 وننظر إلى درجته ، عندما نكون قد عرفنا الدرجة التي تقابل المئيني (80). غير أننا في الجدول السابق لدينا 50 درجة ، ونريد أن نعرف قيمة المئيني رقم ثمانين (أي قيمة الدرجة التي تقابل هذا المئيني) ، نقول في هذه الحالة ، بما أن الـ 100 هي ضعف الـ 50 ، إذن علينا أن نقسم الـ 80 على 20 فنحصل بذلك على الرقم (40).

إن هذا الرقم (40) يمثل المئيني رقم (80) في الجدول من أوطاً درجة إلى أن نصل إلى الرقم (40)، فنجد أن الدرجة التي نصلها هي (78)، ويقع فوقها (10) درجات أخرى، وبهذا نقول أن المئيني (80) تقابل درجة قيمتها 78 في الجدول.

طريقة حساب المئيني (من جدول توزيع كراري)

يوضح المثال الآتي كيفية حساب المئيني بالنسبة لفئات الدرجات، أدناه جدول تكراري لدرجات الطلبة في اختبار الوثب العمودي:

التكرار المجموع	التكرار	فئات الدرجات
52	1	60-58
51	1	57-55
50	1	54-52
49	2	51-49
47	2	48-46
45	3	45-43
42	3	42-40
39	7	39-37
32	8	36-34
24	7	33-31
17	5	30-28
12	4 →	27-25
8	3	24-22
5	3	21-19
2	3	18-17
52		المجموع

فإذا أردنا، مثلاً، حساب المئيني الخامس والعشرين في التوزيع التكراري السابق لدرجات (25) طالباً في الوثب العمودي، نحسب أولاً ترتيب المئيني الخامس والعشرين بالشكل الآتي:

$$\text{المئيني } 25 = \frac{25 \times 25}{100} = 13$$

إن الترتيب (13) يوجد في تكرار متجمع تازل الفتة (28 - 30) ولما كان الحد الأدنى لهذه الفتة هو (27.5) فتكون معادلة المئيني على النحو الآتي:

$$28.1 = 0.6 + 27.5 = 3 \times \frac{12 - 13}{5} = 27.5$$

ولحساب المئيني الخامس والثمانين نحسب ترتيب المئيني بالشكل الآتي:

$$\text{المئيني } 85 = \frac{52 \times 85}{100} = 44.3$$

معادلة المئيني على النحو الآتي:

$$3 \times \frac{42 - 44.2}{3} = 42.5$$

حساب الرتب المئينية:

يوضح المثال الآتي كيفية حساب الرتب المئينية، أدناه جدول تكراري لدرجات (25) طالباً في اختبار «الاختبارات والقياس».

نوات الدوارات	التكرار	التكرار المتجمع الصاعد
96-94	1	25
93-91	3	24
90-88	1	21
87-85	1	20
84-82	3	19
81-79	2	16
78-76	3	14
75-73	3	11
72-70	2	8
69-63	2	6
66-64	3	4
63-61	1	1
المجموع		25

ولنفرض اننا نريد أن حسب الرتبة المئينية لفرد حصل على درجة (74) في اختبار الاختبارات والقياس من الجدول السابق، فتكون طريقة الحساب كما يلي :

إن درجة (74) تقع في الفئة (73-75)، وهناك (8) أفراد درجاتهم أقل من الحد الأدنى للفئة . ونظراً لأن تكرار الفئة (73-75) هو 3 فإن عدد أفراد الفئة (73-75) الذين تقل درجاتهم عن 74 هو :

$$1 = 3 \times \frac{73 - 74}{3}$$

ويكون عدد القيم التي تقل عن 74 في المجموعة هو : $8 = 1 + 8$

ونظراً لأن عدد أفراد المجموعة كلها = 25 لذلك فإن الرتبة المئينية المقابلة للدرجة

(74) تحسب بالشكل الآتي :

$$36 = 100 \times \frac{9}{25}$$

إن هذه النتيجة تعني أن الطالب الحاصل على الدرجة (74) تفوق على 36٪ من الطلبة في اختبار الاختبارات والقياس .

والمئينيات كمعايير تستخدمن على نطاق واسع وقيمتها العملية كبيرة ، فبالإمكان استخدامها ، متى ما أمكن الحصول على العينة المعيارية ، والمئينيات تتتناسب في استخدامها الكبار والصغار ، وفي مجال التربوي وفي غيره من المجالات .

إن المشكلة الأولى التي تواجه الذين يستخدمون معيار المئينيات هي عينة التقنيات (العينة المعيارية) التي يشتق منها هذا المعيار ، فمن الواضح أننا نحتاج إلى عينات تقنيات حسب طبيعة المجموعات التي يراد تطبيق الاختبار عليها .

فالفرد الذي يبلغ من العمر (9) سنوات يجب أن يقيم في ضوء المعايير التي تخص هذا العمر ، والطالب الذي هو في الصف السادس يجب أن يقيم بموجب معايير ذلك الصف ، والمتقدم إلى عمل هندسي ، يجب أن يقيم بموجب معايير ذلك العمل .

إن عينة التقنيات المناسبة لكل حالة ، هي المجموعة التي يتميّز بها الفرد والمجموعة التي في ضوئها تقييم مكانته ، ويعتبر من غير المناسب مقارنة درجات الطالب المتقدم إلى كلية الطب مع معايير عامة غير انتقائية . وعلى هذا فعندما نريد استخدام المعايير المئينية

يتضمن أن تكون لدينا مجموعات متعددة من عينات التقنيات إذ يجب أن تكون لدينا معايير مناسبة لكل مجموعة أو موقف مميز يطبق فيه الاختبار. والمشكلة الثانية بالنسبة للذين يستخدمون المعايير المثنوية تتناول مسألة تساوي الوحدات المثنوية وعلى الأخص عند طرفي التوزيع فهل الفرق بين المئيني الخمسين والخامس والخمسين متساوٍ لفرق بين المئيني التسعين والخامس والتسعين؟

ويمكن الإجابة على هذا السؤال بالقول أن الوحدات المثنوية غير متساوية.

ففي التوزيعات التي تقترب من التوزيع التكراري الاعتدالي وهو ما يحدث في أغلب المقاييس النفسية نلاحظ أن اغلب القيم تتركز عند الوسط بينما يقل عدد القيم المتطرفة في التوزيع. ولهذا نلاحظ أن وحدات المقياس المتميز تضيق في المنطقة الوسطى، بينما تتسع جداً في الطرفين كلما بعدها القيم عن المتوسط.

فالطالب الذي يكون ترتيبه في المئيني الخمسين في الرياضيات والمئيني الخامس والخمسين في القراءة يظهر فرقاً طفيفاً في هاتين القدرتين ، في حين أن الطالب الذي يكون ترتيبه بين المئيني التسعين والخامس والتسعين يظهر فرقاً ملحوظاً في القدرتين .

مجمل القول أن المعايير المثنوية تهيء أساساً لتفسير درجة الفرد في ضوء موقفه من جماعة معينة فإذا أريد للمئيني أن يكون ذا معنى فإن المجموعة التي يراد المقارنة معها يجب أن تكون مجموعة ممثلة له . ونحتاج عادة إلى عدد من الجداول الخاصة بالمعايير المثنوية القائمة على أساس مجاميع مختلفة إذا أردنا أن نطبق اختباراً مع اعمار وصفوف ومهن مختلفة ، إن هذا المعيار يمكن أن يكون قابلاً للتطبيق على مدى واسع إذا هيئت عينات التقنيات المناسبة ويصبح تفسير القيم المثنوية أكثر صعوبة عندما تكون وحدات المقياس المئي صغيرة في وسط المقياس وكبيرة عند طرفيه .

4-3-4. الدرجة القياسية :

تعرف الدرجة القياسية احصائياً بأنها درجة الفرد مطروحاً منها متوسط مجموعته ،

وتقسيم الناتج على الانحراف المعياري عن ذلك المتوسط . وتتراوح قيمتها بين (3+) و (-3) ويقع المتوسط بين هاتين القيمتين ويعطى المتوسط قيمة (صفر) كما في الرسم التالي :

(صفر)



وعند تحويل «الدرجة الخام» وهي الدرجة التي يحصل عليها الفرد في الاختبار الى «درجة قياسية» ، ولنفرض انها كانت (2+) فان ذلك يعني انه يبتعد عن متوسط مجموعة مقدارها (وحتدين) موجبة ، أي أنه أعلى من المتوسط . أما اذا حصلنا على درجة قياسية قيمتها (-1) فان هذا يعني انه يقع دون المتوسط بقيمة مقدارها (وحدة واحدة سالبة .

نفرض أن طالباً حصل على درجة مقدارها (75) في اختبار لقياس الدقة الحركية من اصل (100) درجة فهل يمكن معرفة مستوى الطالب بالنسبة لبقية طلبة صفة؟ أن هذه الدرجة الخام لا تستطيع ان توضح لنا ذلك . فلو كان مستوى طموح طلاب صفة عاليآً فعندئذ قد تكون الدرجة (75) ضعيفة بالنسبة لبقية الدرجات ولو كان مستواهم واطئاً فعندئذ قد تكون الدرجة (75) من الدرجات العالية بالنسبة لبقية الدرجات . وعندما نحصل على المتوسط الحسابي لدرجات هؤلاء الطلاب ، فبالإمكان معرفة مدى تقدم الطالب أو تأخره عن المتوسط ، ولكي نعرف مدى بعد هذا الطالب الحاصل على الدرجة (75) عن المتوسط نحتاج الى ايجاد الانحراف المعياري لدرجات المجموعة التي يتمي اليها ثم نقسم الفرق بين درجته وبين المتوسط على الانحراف المعياري فنحصل على الدرجة القياسية .

وفي المقاييس الرياضية التحميلية كثيراً ما يحاول الباحث أن يقارن بين درجات افرد في عدة اختبارات . وتعتبر الدرجات الخام لل اختبارات غير صالحة للمقارنة ، لذلك لا بد

من تحويل الدرجات الخام إلى درجات قابلة للمقارنة وبواسطة الدرجات القياسية يمكن مقارنة درجات اختبار بأخر مما كان الفرق بينهما في المتوسط الحسابي والانحراف المعياري وذلك لأن متوسط الدرجة القياسية يساوي (صفر) وانحرافها المعياري (1) صحيح وتحسب الدرجات القياسية بالمعادلة الآتية:

$$d = \frac{s - m}{u}$$

حيث أن d = الدرجة القياسية

s = الدرجة الخام.

m = المتوسط

u = الانحراف المعياري

مثال على ذلك:

نفرض أن باحثاً أجرى اختباراً للياقة البدنية لطلبة الصف الثاني متوسط وحصل أحد الطلبة على الدرجة (25) في السرعة والدرجة (75) في المطاولة وحصل طالب آخر على الدرجة (32) في السرعة والدرجة (70) في المطاولة وكان المتوسط الحسابي والانحراف المعياري للاختبارين على نحو الآتي:

المطاولة	السرعة	المتوسط الحسابي	الانحراف المعياري
61.3	20.9		
15.2	8		

عندئذ يمكن استخدام الدرجات القياسية لمقارنة تحصيل كلا الطالبين في الاختبارين أو تحصيل كليهما في الاختبار الواحد.

$$\frac{20.9 - 25}{8} = \text{إن الدرجة القياسية للطالب الاول في السرعة} =$$

$$0.51 = \frac{4.1}{8} =$$

$$\frac{61.3 - 75}{15.2} = \text{والدرجة القياسية للطالب الاول في المطاولة} =$$

$$0.90 = \frac{13.7}{15.2} =$$

ويكن الاستنتاج من خلال الدرجات القياسية أعلاه أن الطالب الاول اكثراً تقدماً في اختبار المطاولة بالمقارنة مع اختبار السرعة.

$$\frac{20.9 - 32}{8} = \text{الدرجة القياسية للطالب الثاني في السرعة} = 1.39$$

$$0.75 = \frac{61.3 - 70}{15.2} = \text{الدرجة القياسية للطالب الثاني في المطاولة} =$$

يمكن الاستنتاج من خلال الدرجتين القياسيتين للطالب الثاني انه اكثراً تقدماً في السرعة بالمقارنة مع تقدمه في اختبار المطاولة.

كما يمكن الاستنتاج من خلال الدرجات القياسية لكلا الطالبين أن تقدم الطالب الاول في اختبار السرعة يقارب تقدم الطالب الثاني في اختبار المطاولة في حين أن تقدم الطالب الثاني في اختبار السرعة افضل من تقدم الطالب الاول في المطاولة.

٤-٣-٤-١ عيوب الدرجات القياسية:

إن الدرجات القياسية تفضل على المثنىات لأن وحدات القياس فيها متساوية إلا أنها لا تخلو من عيوب أهمها:

- 1 لا تصلح لعملية المقارنة إلا إذا كان توزيع درجات الطلبة اعتدالياً أو قريباً من الاعتدالي.
- 2 لا تخلو الدرجات القياسية من الدرجات السالبة التي قد لا يفهمها إلا الخبر بها لأجل ذلك تعدل للتخلص من الدرجات السالبة (تحصل على الدرجات السالبة عندما تكون درجة الطالب أقل من متوسط مجموعته كما مر بنا في تعريف الدرجة القياسية).
- 3 كثيراً ما تحتوي الدرجات القياسية علىكسور عشرية (لأنها تعوق اجراء عملية المقارنة بسرعة وسهولة).

٤-٣-٤-٢ الدرجات القياسية المعدلة:

ولأجل التخلص من الكسور نضرب كل درجة قياسية في ثابت كأن تكون الدرجة (100). كما يمكن التخلص من العلامات السالبة عن طريق إضافة عدد ثابت إلى الدرجة، كأن يكون (500). ويمكن توضيح ذلك بالمثال الآتي الذي بين كيفية تحويل درجات أحد الطلبة إلى درجات قياسية معدلة بحيث يكون متوسطها الحسابي (500) وانحرافها المعياري (100).

الموضوع	الدرجة القياسية	الدرجة القياسية المعدلة
احصاء	0.15	$515 = (100 \times 0.15) + 500$
علم الحركة	0.4	$540 = (100 \times 0.4) + 500$
ميكانيكا	0.6-	$440 = (100 \times 0.6-) + 500$
	0.7-	$430 = (100 \times 0.7-) + 500$

ورغم فوائد الدرجة القياسية المعدلة الا انها لا تخلو من عيوب ، اهمها انها تتأثر بشكل التوزيع التكراري فان كان التوزيع التكراري متواياً التراءاً سالباً او موجباً فانها تصبح غير صالحة للمقارنة . ولذلك فإن الدرجة القياسية المعدلة تستخدم عندما يكون التوزيع التكراري اعتدالياً او قريباً من الاعتدالي . وقبل الانتهاء من موضوع المعايير لا بد من الاشارة إلى أن نتائج الاختبار، سواء أكانت محولة إلى معيار عمري أو صفي أو مئيني أو درجة قياسية ما هي إلا وصف لما هو كائن من خصائص أو صفات أو قدرات لدى الفرد المفحوص وليس لما ينبغي أن يكون ..

إن ما سبق ذكره يعتبر معاييرأ Norms ، وهي تشتق من الجماعة التي يطبق عليها الاختبار لذلك فانها تتأثر بمستوى هذه الجماعة ، فإذا كانت قوية في الخاصية التي يقيسها الاختبار ، فإن المعايير تكون عالية والعكس صحيح ايضاً.

لذلك فإن العاملين في القياس لا يكتفون بهذه المعايير وإنما يلتجأون إلى وضع معاييرهم الخاصة التي على الفرد أن يجتازها لكي يعتبر «ناجحاً» في تلك الخاصية التي يقيسها الاختبار بناءً على متطلبات المهمة التي يراد انتقاء الأفراد لها ويطلق على المعايير الخاصة في هذه الحالة مصطلح «المعايير المرجعية» Criterion referenced norms ويطلق عليها في بعض الأحيان أيضاً مصطلح درجة العبور Passing score .

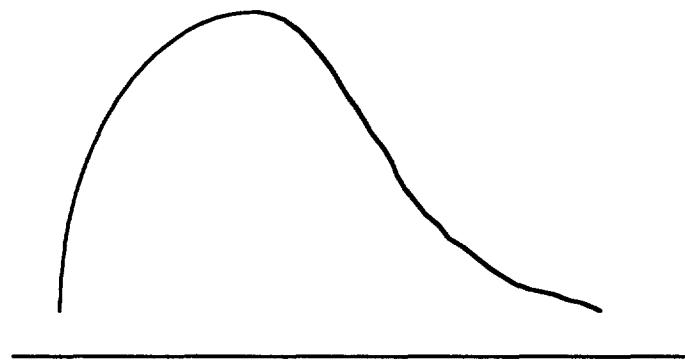
4-3-2 المعيار الثاني

4-3-1 نشأته ومعناه

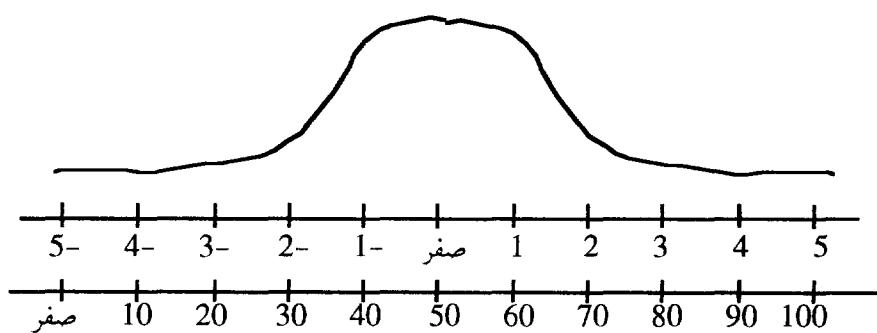
ترجع فكرة هذا المعيار إلى تورنديك L.E Thorndike الذي اقترح على مکال W.A Mc Call سنة 1922 إنشاء معيار نفسي لحساب المستويات المختلفة للقدرة على القراءة ، وقد سمي هذا المقياس بالمعيار الثاني نسبة إلى تورنديك وتيرمان اعترافاً بفضلهما على المقاييس النفسية الحديثة .

وتعتمد فكرته الرئيسية على تحويل التوزيع التجريبي إلى توزيعه الاعتدالي الذي

يصله باصله في صورته العامة، ثم تحويل درجاته الى درجات معيارية متوسطها يساوي صفرأ وانحرافها المعياري يساوي واحداً صحيحاً، ثم تحويل هذه الدرجات المعيارية الى درجات معيارية معدلة متوسطتها 50 وانحرافها المعياري 10 والاشكال التالية توضح مراحل هذه الفكرة.



شكل يمثل التوزيع التكراري المتوازي



شكل يوضح التوزيع الاعتدالي بدرجاته المعيارية التي تمتد من -5 إلى + 5 والدرجات الثانية التي تمتد من صفر إلى 100 .

وعندما نقارن شكل التوزيع التجاري المتبين في الشكل رقم () بالتوزيع الاعتدالي المتبين في الشكل ندرك أهمية المرحلة الاولى في تنسيق التكرار التجاري وتحويله من تكرار العينة التجريبية المحدودة الى تكرار الاصل العام النموذجي الذي تتسمى اليه تلك العينة . وعندما نقارن الدرجات المعيارية التي تقسم قاعدة المنهج الاعتدالي الى 10 أقسام تمتدد من - 5 إلى + 5 الدرجات الثانية التي تقسم قاعدة المنهج الاعتدالي إلى 100 قسم تمتدد من صفر الى 100 ندرك معنى وأهمية الدرجة الثانية في تحويل الدرجات المعيارية السالبة الى درجات موجبة ، وفي تقسيم الأجزاء الكبيرة إلى وحدات صفيرة تساوي كل منها 0.1 من الانحراف المعياري .

فالمسافة التي تمتد من صفر الى + 1 اصبحت تمتد من 50 الى 60 أي انها انقسمت الى 10 اجزاء صغيرة ، وهكذا يصبح المعيار الثاني اكثر حساسية في قياس مستويات الفروق الفردية من الدرجات المعيارية . ويصل بنا هذا التحليل إلى أن الدرجة الثانية درجة معيارية معدلة لتوزيع اعتدالي متوسطة 500 وانحرافه المعياري 10 .

وبما أن الدرجة المعيارية المعدلة

= (الدرجة المعيارية × الانحراف المعياري الجديد) + المتوسط الجديد

$$\text{الدرجة الثانية} = \text{الدرجة المعاشرة} \times 10 + 500$$

$$50 + 0.15 = \text{أيّ أُن ت}$$

حيث يدل ت على الدرجة الثانية.

ويدل الرمز دعلمى، الدرجة المعاشرة.

هذا ويكن أن نستخدم هذه المعادلة في حساب الدرجات التائية المقابلة للدرجات المعيارية المختلفة.

و عندما تصبح الدرجة المعيارية مساوية لـ 5-

$$\text{تصبح الدرجة الثانية} = 50 + (10 \times -5)$$

$$50 + -50 =$$

$$= \text{صفر}$$

وهذه هي الدرجة الثانية التي تحدد بدء القياس.

وعندما تصبح الدرجة المعيارية مساوية لصفر.

$$\text{تصبح الدرجة الثانية} = (\text{صفر} \times 10) + 50 = 50$$

وهذه هي الدرجة الثانية التي تحدد منتصف القياس.

وعندما تصبح الدرجة المعيارية مساوية لـ 50+

$$\text{تصبح الدرجة الثانية} = 15 + (10 \times 5) = 65$$

$$50 + 500 =$$

$$100 =$$

وهذه هي الدرجة الثانية التي تحدد نهاية القياس.

4-3-2- طريقة حساب المعيار الثاني من الدرجات المعيارية

تعتمد الطريقة الاحصائية لحساب درجات المعيار الثاني على جدول المساحات الاعتدالية ، وسنستعين بهذا الجدول في تحويل التوزيع التكراري التجاري الى توزيع تكراري اعتدالي وذلك بحساب التكرار المتجمع التصاعدي النسبي للتوزيع التكراري التجاري، ثم البحث عن الدرجات المعيارية التي تقابل تلك النسب لو كانت اعتدالية أو مساحات اعتدالية . وهذا كفيل بتحويل درجات التوزيع التجاري الى درجات معيارية في التوزيع الاعتدالي المقابل لذلك التوزيع التكراري التجاري . ثم نحوال الدرجات المعيارية الى درجات ثانية بضربها في 10 واضافة 50 إلى حاصل الضرب . والجدول التالي يوضح خطوات حساب الدرجات الثانية :

الدرجة الثانية $(10 \times ذ)$ 50 +	الدرجة المعيارية ذ	النسبة التصاعدي النوعي	المجموع التصاعدي	المجموع التصاعدي	المجموع التصاعدي	الحدود العلية للفئات	فئات الدرجات
26.7	3.3263-	0.010	2	2	59.5	59.55	
33.0	1.6954-	0.045	9	7	64.5	64.60	
38.3	1.1750-	0.120	24	15	69.5	69.65	
46.7	0.3319-	0.270	74	50	74.5	74.70	
54.4	0.4299+	0.670	134	60	79.5	79.75	
62.5	1.2526+	0.895	178	45	84.5	84.80	
67.0	1.6954+	0.955	191	12	89.5	89.85	
75.8	2.5758+	0.995	199	8	94.5	94.90	
		1.000	20	1	100.5	99.95	
				200		المجموع	

جدول يوضح الخطوات الاحصائية لحساب الدرجات الثانية

وتتلخص الخطوات الاحصائية لحساب الدرجات الثانية فيما يلي :

- 1) تكتب فئات الدرجات كما هو مبين بالعمود الاول بالجدول .
- 2) تكتب الحدود الحقيقة العليا لتلك الفئات في العمود الثاني لأنها تحدد المقابلات الخام للدرجات الثانية ، ولأنها تحدد معنى التكرار المجموع التصاعدي النسبي ، فمثلاً تمثل

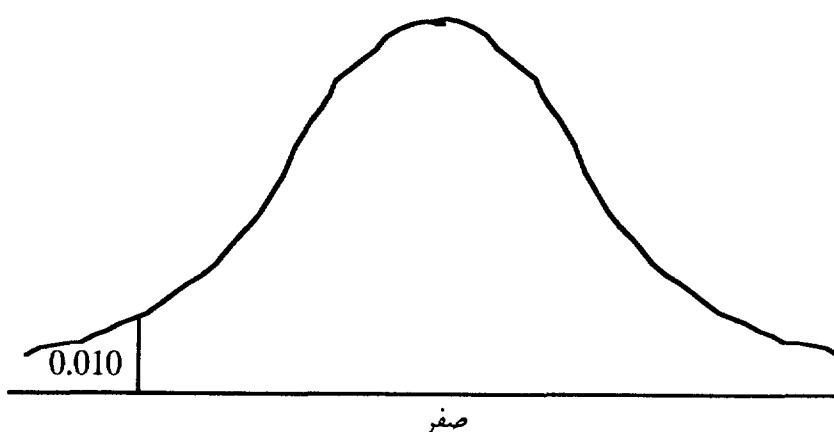
الأفراد الذين حصلوا على درجات أقل من 59.5 تساوي 0.010 كما يدل على ذلك التكرار المتجمع التصاعدي النسبي للفئة الأولى.

(3) يرصد التكرار بالعمود الثالث.

(4) يحسب التكرار المتجمع التصاعدي في العمود الرابع من الجدول السابق.

(5) يحسب التكرار المتجمع التصاعدي النسبي في العمود الخامس وذلك بقسمة كل تكرار متجمع على عدد الأفراد أي أن $200 \div 2 = 0.010$ ، $200 \div 9 = 0.045$ ، $200 \div 24 = 0.120$ وهكذا بالنسبة لبقية الفئات.

(6) نستعين بالتكرار المتجمع النسبي لتحويل التوزيع التجريبي إلى توزيع اعتدالي ، و بما أن هذه النسب تمثل مساحات يقع حدتها الأيسر عند النهاية الدنيا لمساحة ، ويقع حدتها اليسين عند الدرجة المعيارية التي تحدد مستواها العلوي كما هو مبين بالشكل .



شكل يوضح علاقة التكرار المتجمع التصاعدي النسبي بالمساحات الاعتدالية والدرجات المعيارية

إذن نستطيع أن نحسب تلك الدرجات المعيارية التي تقع على الحدود اليمنى للنسبة المختلفة، وذلك بالاستعانة بجدول المساحات الاعتدالية المبين بملحق الجداول الاحصائية.

7) نرصد هذه الدرجات المعيارية في العمود السادس، ونلاحظ عند وصتنا لتلك الدرجات علامتها الجبرية فنكتبها سالبة عندما تقع على يسار المتوسط أي عندما تقل المساحة عن 0.5 ونكتبها موجبة عندما تقع على يمين المتوسط أي عندما تزيد مساحتها على 0.5.

8) نضرب كل درجة معيارية في 10 ثم نضيف 50 إلى حاصل الضرب لنحصل بذلك على الدرجات الثانية المبينة بالعمود الأخير من الجدول السابق.

هذا ونستطيع أن نحسب الدرجة الثانية مباشرة من التكرار المتجمع التصاعدي النسبي دون أن نحسب الدرجة المعيارية ودون أن نعدل لها إلى درجة ثانية، وذلك بالاستعانة بجدول المعيار الثاني المبين بملحق الجداول الاحصائية. وقد رصدنا في ذلك الجدول الدرجة الثانية المقابلة لكل مساحة امتدالية، أي المقابلة لكل تكرار متجمع تصاعدي نسبي، حتى يعتمد عليه القارئ في حساب الدرجات الثانية. وقد أثثنا في مثالنا السابق المبين في الجدول أن تووضح جميع الخطوات الاحصائية لحساب الدرجات الثانية ليدرك القارئ، علاقتها المباشرة بالدرجات المعيارية والدرجات المعيارية المعدلة.

المصادر والمراجع

- 1- القرآن الكريم .
- 2- ابراهيم احمد سلامة: الاختبارات والقياس في التربية الرياضية - الاسكندرية - منار العلم - 1979 .
- 3- ابراهيم احمد سلامة، اللياقة البدنية - دار الفكر العربي - القاهرة - 1976 .
- 4- احمد محمد خاطر وعلي فهمي البيك. القياس في المجال الرياضي - ط ١ - الاسكندرية - 1987 .
- 5- احمد محمد خاطر: الاختبارات والقياس في التربية الرياضية - القاهرة - 1975 .
- 6- بلال عبد الوهاب وحسن عبد العزيز. برامج واختبارات اللياقة البدنية - القاهرة - مطبعة مخيم - 1968 .
- 7- رودثي شتمлер : طرق الاحصاء في التربية الرياضية - ترجمة عبد علي نصيف ومحمود السامرائي - بغداد - 1974 .
- 8- حسن سيد معرض: تقسيم التلاميذ في التربية البدنية - القاهرة - 1975 .
- 9- رودجرز: اختبار رودجرز لدراسة شخصية الاطفال - ترجمة مصطفى فهمي - دار مصر للطباعة - اقاهرة - 1968 .
- 10- سامي الشربيني ومتولي بهائم : وضع مستويات معيارية لطرق السباحة - مجلة دراسات وبحوث - جامعة حلوان - 1986 .

- 11- سبع محمد ابو لبده : **مبادئ القياس النفسي والتقويم التربوي** - عمان - 1979 .
- 12- صفوت فرج : **التحليل العاملي في العلوم السلوكية** - القاهرة - دار الفكر العربي - 1980 .
- 13- عبد الجبار توفيق البياتي وذكريا زكي : **الاحصاء الوصفي الاستدلالي** - بغداد - مطبعة مؤسسة الثقافة العمالية - 1977 .
- 14- عبد الجليل ابراهيم الزويبي : **الاختبارات والمقاييس النفسية** - مطبعة جامعة الموصل - 1986 -
- 15- عبد علي نصيف : **الفحوصات والاختبارات الرياضية** - مجلة الثقافة الرياضية - مطبعة أسعد - بغداد - 1970 .
- 16- عبد الله رمضان الكندري : **مبادئ الاحصاء وأساليب التحليل الاحصائي** - الكويت - منشورات ذات السلسل - 1985 .
- 17- عبد الرحمن محمد عيسوي : **القياس والتجريب في علم النفس والتربية** - دار النهضة العربية - 1974 .
- 18- عبد علي نصيف وآخرون : **الفحوصات والاختبارات الرياضية** - مجلة الثقافة الرياضية - بحث منشور - 1980 .
- 19- عماد الدين محمد سلطان : **التحليل العاملي** - دار المعارف بمصر - القاهرة - 1973 .
- 20- غانم سعيد العبيدي : **التقويم والقياس في التربية والتعليم** - بغداد مطبعة شفique - 1970 .

- 21- قيس ناجي وآخرون: الاختبارات والاحصاء والتقويم في التربية الرياضية - مطبعة جامعة بغداد - 1989.
- 22- قيس ناجي عبد الجبار: اختبارات النمو واللياقة البدنية لطلاب المرحلة الابتدائية 6 - 12 سنة - اطروحة دكتوراة وارشو - 1980.
- 23- فؤاد البهبي السيد: القدرة العددية - القاهرة - 1960.
- 24- فؤاد ابو حطب: بحوث في تقييم الاختبارات النفسية - مكتبة الانجلو المصرية - القاهرة - 1977.
- 25- كمال عبد الحميد و محمد صبحي حسانين: اللياقة البدنية ومكوناتها - القاهرة - دار الفكر العربي - 1978.
- 26- محمد حسن علاوي و نصر الدين رضوان: اختبارات الاداء الحركي - دار الفكر العربي - القاهرة - 1991.
- 27- محمد جاسم احمد الياسري: دراسة مقارنة في اللياقة البدنية - رسالة ماجستير - جامعة بغداد - 1984.
- 28- محمد السيد خيري: الاحصاء في البحوث النفسية والتربوية والاجتماعية - القاهرة - دار النهضة العربية - 1970.
- 29- محمد صبحي حسانين: التحليل العاملی للقدرات البدنية - دار فوزي للطباعة - مصر - 1989.
- 30- محمد عبد السلام أحمد: القياس النفسي والتربوي - ط ١ - القاهرة - 1960.
- 31- مروان عبد المجيد ابراهيم: الالعاب الرياضية للمعوقين - دار الفكر - عمان - 1997.

- 32- مروان عبد المجيد ابراهيم: تصميم بطارية اختبار لقياس اللياقة البدنية لطلاب المرحلة الثانوية بالعراق - بحث منشور - المؤتمر العلمي الرياضي الثاني لكليات التربية الرياضية - العراق - 1986 .
- 33- مروان عبد المجيد ابراهيم: وضع مستويات معيارية لاختبارات اللياقة البدنية لطلاب المرحلة الثانوية بالعراق - بحث منشور - المؤتمر العلمي الرياضي الثالث لكليات التربية الرياضية - العراق - 1988 .
- 34- مروان عبد المجيد ابراهيم: البطارية العراقية لاختبارات اللياقة البدنية - اطروحة دكتوراه - 1994 .
- 35- نزار الطالب و محمد السامرائي: مبادئ الاحصاء والاختبارات البدنية والرياضية - مطبعة جامعة بغداد - العراق - 1981 .
- 36- وديع التكريتي: الاحصاء في بحوث التربية الرياضية - مطبعة جامعة الموصل - 1995 .
- 37-Buskirk E. , **Nutrition for the Athletic**, Acadmie Press - NewYork; 1974.
- 38- Barrow, H.M, and MG Gee. R.A **Practical Approach to Measur ment in Physical education** Leaand Eebiges 1962.
- 39- Clarke; H.H. **Application of measurment to health and Physical Education** 4th ed. Prentice- Hall, ivc. Englewood ciffs. New jersey- 1967.

- 40- Cooper'K.H' **The New Aerobics.** M. Evans co; inc. New York, 1970.
- 41- Cooper, K.H; **Aerobics,** M. Evans co; inc, New Yor, 1968.
- 42- Donald; K,c. **Hand Book of Physical Fitness Activities,** The macmillan. co. N.Y. 1965.
- 43- Eckert, H.M.; **Practical measurement of physical performance** Lea and febiger, Philadelphia. 1974.
- 44- Hockey, R.V, **Physical Fitness The Path Way to healthy full living:** 2nd. ed, The C.V. Mosby. co Sait Lous 1970.
- 45- John, D.W. **Statistical Design and Analysis Of Experiments.** New York. 1971.
- 46- Macloy, C.H. **The measurement of Athlitic dower.,** Ais. Barnes and co. New york. 1932.
- 47- Mathews; D.K; **Measurment in Physical Education;** 2nd-ed. W.B. Saunders. Co. Philadelphia, London- 1964.
- 48- Thyler; L.E. **Test and Measurment,** Prentic- Hall, inc; Englewood Clifs, New jersey, 1963.

المؤلف في سطور

- 1- الإختصاص العام - بكالوريوس - ماجستير - دكتراه تربية رياضية .
- 2- الاختصاص الدقيق - الاختبارات والقياس والإحصاء في التربية الرياضية .
- 3- أستاذ مادة الإحصاء والاختبارات والقياس في كلية التربية الرياضية بجامعة البصرة - العراق - منذ عام 1984 - 1996 . وفي جامعة الفاتح بالجماهيرية العربية الليبية منذ عام 1996 ولحد الآن .
- 4- أنجز عدة بحوث في مجال الاختبارات والقياس في التربية الرياضية .
- 5- أشرف على عدة بحوث لطلبة الماجستير في مجال الاختبارات ولاقياس في التربية الرياضية .
- 6- صمم بطارية اختبار لقياس اللياقة البدنية لطلاب المرحلة الثانوية بالعراق عام 1994 .
- 7- وضع مستويات معيارية لاختبارات اللياقة البدنية لطلاب المرحلة الثانوية بالعراق عام 1988 .
- 8- شارك في عدة مؤتمرات عربية ومحلية قدم خلالها بحوثاً في مجال الاختبارات والقياس والتقويم .