الفصل الخامس

اختيار أدوات القياس

Selection of Measuring Instruments

تمهيد:

سواء أكنت تختبر فروضاً Hypotheses أم تسعى للحصول على إجابات عن أسئلة ، يلزم أن تكون لديك أداة تتسم بالصدق والثبات Valid, Reliable. وثمة نقطة أساسية علينا أن نتذكرها : ليس على الباحث التعرف على الأداة المناسبة لدراسته ، بل يختار أفضل أداة متاحة.

الأهداف :

سيكون في مقدور الدارس (الباحث) بعد دراسة هذا الفصل أن:

1- يعرف ويصف صدق المحتوى (المضمون) Content Validity.

2- يعرف ويصف صدق التكوين (المفهوم) Construct Validity.

3- يعرف ويصف الصدق التلازمي Concurrent Validity.

4- يطبق إجراءات خاصة بتحديد الصدق التلازمي عن طريق اثبات علاقة.

5- يعرف ويصف الصدق التنبؤي Predicitive Validity.

6- يطبق الإجراءات الخاصة بتحديد الصدق التنبؤي.

الثبات Reliability

7- يعرف ويصف الثبات.

8- يطبق الإجراءات الخاصة بتحديد ثبات الإعادة للاختبار Test-retent reliability.

9- يطبق الإجراءات الخاصة بتحديد ثبات الصور المتكافئة Equivalent-forms reliability.

10- يطبق الإجراءات الخاصة بتحديد الثبات النصفي Split-half reliability.

11- يعرف ويصف ثبات التكافؤ النسبي Rationale equivalence reliability.

12- يشرح فارق الثبات بين المصححين وداخلهم Inters scorer and intra scorer reliability.

13- يعرف ويصف الانحراف المعياري للمقياس Standard error of measurement.

14- يصف الهدف من اختبار التحصيل أي يصف ما يقيسه اختبار التحصيل.

أنواع أدوات القياس Types of Measuring Instruments

15- يصف الهدف من أنواع الأدوات الاسقاطية التالية:

أ) قائمة الشخصية ب) مقاييس الاتجاهات

ج) اختبارات الابتكارية د) قائمة الميول

16- يصف الغرض من اختبارات الاستعداد Aptitude.

اختيار الاختبار Selection of a test

17- يبين اثنين من أهم المبادئ التوجيهية أو القواعد فيما يتعلق باختيار الاختبار.

18- يعرف ويصف بإيجاز ثلاثة مصادر لمعلومات الاختبار.

19- يدرج قائمة ، بحسب ترتيب الأهمية ، بالعوامل التي نبيغي أخذها في الاعتبار عند اختيار اختبار واحد من بين عدد من البدائل.

تطبيق (إدارة) الاختبار Test Administration

20- يدرج ثلاثة مبادئ توجيهية لإدارة الاختبار.

الهدف والعملية Purpose and Process

تنطوي كل الدراسات البحثية على جمع البيانات ، وحيث أن كل الدراسات مصممة إما لاختبار الفروض أو إجابة عن أسئلة ، فهي جميعاً تحتاج إلى بيانات لتحقيق غاياتها. ومعظم الدراسات تستخدم نوعاً من أدوات جمع البيانات ، وهي في الغالب منشورة أو مقننة Standardized.

وتوجد في الواقع ثلاثة أساليب رئيسية لجمع البيانات:

1- بتطبيق أداة مقننة.

2- بتطبيق أداة مطورة ذاتياً.

3- تسجيل البيانات المتوافرة (مثل متوسط الدراسات ومعدلات الغياب).

ولا يهتم هذا الفصل ، وذلك لعدة أسباب وجيهة ، إلا باختيار اختبارات منشورة ومقننة.

obeikandl.com

وجمع بيانات متوافرة يتطلب جهداً أقل كما تبدو عملية جذابة جداً ، ومع هذا لا توجد دراسات كثيرة جداً مناسبة لهذا النوع من البيانات ،وحتى إن كانت مناسبة بمعنى إذا أمكن توظيفها من أجل اختبار الفرض المستهدف Indended Hypotheses أو إجابة عن الأسئلة المطلوبة ، فهناك مشكلات متأصلة في هذا النوع من البيانات ، فعلى سبيل المثال لا تمثل هذه الدرجات في ذاتها وبالضرورة نفس مستوى التحصيل حتى في مدرستين مختلفتين من نفس نوع التعليم ، كما أن تطوير أداة تتعلق بدراسة خاصة ينطوي على مصاعب جمة ، إذ يتطلب تطوير أداة جيدة قدراً كبيراً من الوقت والجهد والمهارة ، كما يمكن أن يتزايد الوقت الكلي لتنفيذ دراسة ما ، إذا ما تضمنت تطويراً لأداة ، كذلك يستلزم الأمر أن يتلقى تدريبا ضرورياً في برنامج خاص بالقياس حتى يكتسب المهارات اللازمة بالنسبة لتطوير أداة جيدة.

وعلى الجانب الإيجابي : الوقت الذي يستغرقه اختبار أداة مقننة مناسبة هو بلا جدال أقل من الوقت الذي يستغرقه تطوير أداة لقياس نفس الشيء. فضلاً عن ذلك قد قام بتطوير هذه المقاييس المقننة خبراء يملكون المهارات اللازمة لذلك ، كما أنها من الجهة البحثية ذات ميزة إضافية حيث يمكن مقارنة نتائجها مع دراسات أخرى تستخدم نفس الأداة. ولنفترض على سبي المثال أنه تم تنفيذ دراستين منفصلتين لتحديد ما طريقة التدريس لمادة الجبر التي تسفر عن أعلى تحصيل نهائي.

الطريقة (أ) أم الطريقة (ب) ، لنفترض أن الدراسة (1) استخدمت اختبار تحصيل جبر مطور ذاتياً بينما استخدمت الدراسة (2) تحصيل جبر مقنن.

والآن إذا اكتشفت الدراسة (1) أنه لا فارق بين الطريقة (أ) والطريقة (ب) ، ولكن الدراسة (2) اكتشفت أن الطريقة (ب) كانت الأفضل ، فإن تفسير هذه النتائج المتعارضة يكون صعباً. فربما كان هذان الاختباران يقيسان قدرات مختلفة ، كما يحتمل أن تكون تركيبة (بنية) الاختبار المستخدم في الدراسة (1) غير كافية ، وعلى أي حال ، يصعب تمييز الفروق التي ترجع إلى طريقة التدريس وتلك التي تعزي إلى الاختبار المستخدم. وأخيراً يُسهل استخدام اختبار مقنن إعادة تكرار الدراسة على يد باحث مستقل (آخر).

وهناك آلاف من الأدوات المقننة المتوافرة والتي تسفر عن مجموعة متنوعة من البيانات تتعلق بمجموعة عريضة من الأهداف ومن المجالات الرئيسية التي تشملها أدوات القياس العديدة التي تم تطويرها : التحصيل ، الشخصية ، الذكاء ، الاستعداد ، والتي ينقسم كل مجال منها بدوره إلى عديد من الفئات الفرعية. فأدوات الشخصية كمثال Personality Instruments يمكن أن

تصنف إلى أدوات إسقاطيه وإلى أدوات غير إسقاطيه. وتتضمن الأدوات غير الإسقاطية Non-projective قياسات الاتجاهات والميول Attitude and interest.

ويتضمن اختيار أداة تتعلق بهدف بحثي خاص تعريف واختيار أكثر الأدوات ملاءمة من بين البدائل. والباحث الذي لا يجد أداة تناسب دراسته فالأحرى به أن يختار أحسن ما هو متوافر أمامه. فعملية الاختيار تتضمن تحديد النوع الكثر ملاءمة من الأدوات المتوافرة.

فعلى سبيل المثال : فهم المادة المقروءة ، يتضمن تحليل مقارن للاختبارات المتوافرة في هذا النوع. ولهذا من اجل اختيار ذكي للأداة يتعين أن يكون الباحث على علم بتشكيلة عريضة لأنواع الأدوات الموجودة كذلك أن يكون على علم بشأن المعايير التي ينبغي أن يطبقها عند اختيار أحد الأدوات من هذه البدائل.

خصائص الاختبار المعياري

Characteristics of a Standardized Test

لا يقصد بالاختبار هنا مجموعة الأسئلة التي يجب عنها المفحوص ، إنما المقصود هو أن الاختبار وسيلة لقياس المعرفة أو المهارة أو الشعور أو الذكاء أو الاستعداد لدى فرد ما أو جماعة ما ، وينتج عنها درجات رقمية Numerical Scores يمكن أن تستخدم في تعريب أو تصنيف أو تقييم المُختبرين Test Takers ، فضلاً عن ذلك هناك عدد من الخصائص المشتركة في الاختبارات المعيارية. والمتبع أن يتولى خبراء عملية تطوير الاختبارات المعيارية ، وبالتالي تركيبتها الجيدة ، ويتم تحليل وتنقيح بنود الاختبار الخاص حتى يفي بمعايير جودة معينة ، كما تحدد بدقة التوجيهات بشأن تطبيق هذا الاختبار المقنن وتصحيحه وتفسيره.

والموضوعية Objectivity هي من الخصائص الرئيسية للاختبار المعياري ، وتعني موضوعية الاختبار في الأساس أن الدرجة التي يحصل عليها الفرد هي واحدة (دون تغيير يذكر) بغض النظر عمن يقوم بالتصحيح أي يتم تقدير الدرجة بنزاهة وتجريد.

وصدق البيانات وثباتها Validity and Reliability هما من الخصائص الرئيسية للاختبار المعياري. ونعني بصدق الاختبار أن يقيس الاختبار فعلاً ما يقصد قياسه ، وملاءمته لقياس ما ينوي قياسه ويشير ثبات الاختبار إلى ثبات ما يقيسه الاختبار مهما تكرر القياس (مثل إعطاء الاختبار نفس الدرجات لنفس الأشخاص عند إعادة تطبيقه عليهم.

ومن الخصائص الهامة للاختبار المعياري تحديد مواصفات شروط تطبيقه بما يضمن تنفيذ التعليمات بدقة ، وتطبيق الاختبار دوماً بنفس الطريقة وأن يسفر دائماً عن نفس النوعية من

البيانات فإذا ك انت التوجيهات واضحة وجلية أمكن حتى للباحث المبتدئ تطبيق معظم الاختبارات على نحو صحيح. وتضمن التوجيهات عادة إرشادات لكي يطلع عليها من يؤدى الاخبار Test Taker ، كما تتضمن التحديدات الزمنية ، ومبادئ توجيهية خاصة بمقدار الاتصالات المسموح بها وطبيعتها.

وأخيراً وليس آخراً تتضمن الاختبارات المعيارية التوجيهات الخاصة بالتصحيح وتقدير الدرجات والمبادئ التوجيهية لتفسيرها. وتتضمن توجيهات التصحيح وتقدير الدرجات مواصفات محددة مثل معايير الدرجات المقبولة ، بينما تتضمن المبادئ التوجيهية تفسير درجات الاختبار وجدول المعايير الخاص بذلك.

ومن المألوف تطبيق الاختبار على أكبر عدد مناسب من أفراد محددين مع تليل لدرجات الاختبار الناتجة ، ويعرض جدول المعايير الدرجات الاختبار الناتجة ، ويعرض جدول المعايير الدرجات الخام وعمليات التحويل المتكافئة Equivalent Transformation مثل الرتب المئينية المناظرة Corresponding Percentile Ranks التي تسهل عملية تفسير درجة الفرد بخصوص أداء الجماعة ن وبما أن معظم البحثين يعملون عموماً مع درجات خام ، فهذا النوع من المعلومات هو أكثر قيمة للجماعات المستفيدة (المستهلكة Consumers) الأخرى مثل أفراد المدرسة العامة.

وكما سبقت الإشارة إليه يعتبر صدق الاختبار أمراً لا غنى عنه ، فلا توجد جودة أو فعالية لاختبار ما يمكن أن تعوض عن صدق غير كاف.

صدق الاختبار Validity:

أبسط تعريف للصدق هو أن يقيس الاختبار فعلاً ما يفترض أن يقيسه. ومن المفاهيم الخاطئة والشائعة ، هو أن اختبار ما يكون صادقاً أو غير صادق. فالاختيار ليس صادقاً بذاته Per se بل إنه صادق بالنسبة لهدف خاص أو مجموعة خاصة ، فالسؤال هذا ليس صادقاً ، بل الأحرى صادق بالنسبة إلى من وإلى ماذا؟ Valid for what and for whom? فالاختبار الصادق في تحصيل البيولوجي ليس من المرجح أن يكون صادقاً بالنسبة لاختبار الشخصية. أما بخصوص الصدق المتعلق بمن for whom فالمقياس في المفردات Vocabulary لطلاب الثانوي هو بالتأكيد ليس بمقياس صادق أو صالح لطلاب الصف الثاني الابتدائي.

ولمزيد من التوضيح لمفهوم مصادق إلى ماذا وإلى من ؟ نطبق هذا المفهوم على التدريس ، فعامة المدرس الصادق هو الذي يدرس ما يفترض أن يقوم بتدريسه أو أن يتعلم

طلابه ما يفترض أن يتعلمون ، ولكن المدرس مهما كان تميزه فهو ليس صادقاً بذاته ، بل هو صادق بالنسبة لهدف خاص أي منطقة تعليمية خاصة ، أو سن خاص أو مستوى صف معين .. فمدرس رياض الأطفال الصادق من غير المحتمل أن يكون مدرس ثانوي صادق ، كما أن مدرس الهندسة الثانوي الصادق ليس من المحتمل ان يكون مدرس فيزياء صادق.

والصدق إلى من يضفى أهمية على المجموعة المعيارية حيث يشبه أشخاصها بالأشخاص الذين ترغب تطبيق الاختبار عليهم ، وهذا يمكن أن يؤدى إلى تفسير مناسب للنتائج.

وحيث أن الاختبارات تصمم من أجل تشكيلة من الأهداف ، وحيث أن الصدق لا يمكن تقييمه إلا باعتبار الهدف ، فلا مجال للغربة من وجود أنواع مختلفة عديدة من الصدق : صدق المضمون Content ، صدق التكوين (المفهوم) Construct ، الصدق التلازمي Concurrent والصدق التنبؤي Predictive.

وبسبب اختلاف أساليب تحديدها فهي قد تصنف إما بالصدق المنطقي Logical أو صدق مرتبط بمحك Criterion-Related Validity. وبالصدق المنطقي يتضمن صدق المضمون حيث يحدد أساساً من خلال التقدير Judgment ، ويتضمن الصدق المرتبط بمحك أو الصدق الأمبريقي Empirical الصدق التلازمي والصدق التنبؤي ويسمى بذلك حيث يتحدد الصدق عن طريق ارتباطه بالأداء بناء على محك ، وارتباط الصدق بمحكات يجعله أكثر موضوعية من الصدق المضمون. ويجمع تقييم صدق المفهوم بين كل التقدير ومعايير خارجية. وبالنسبة لأى اختبار من المهم البحث عن دليل فيما يتعلق بنوع الصدق المناسب وهدف أو أهداف الاختبار المستهدفة.

صدق المضمون (المحتوى) Content Validity

أي مطابقة محتوى الاختبار لما يريد قياسه ، فهو يتطلب كلا من صدق المفردات (بنود) الاختبار وصدق المعاينة. فالاختبار الذي يستهدف قياس معرفة الحقائق البيولوجية قد يتضمن صدق مفردات جيد لأن كل المفردات تتعامل فعلاً مع حقائق البيولوجيا ، ولكنه قد يكون غير جيد من حيث صدق المعاينة إذا كانت كل المفردات تتعامل فقط مع الفقريات Vertebrates ، فالاختبار صادق المضمون يضم مفردات ممثلة لمحتوى الاختبار على نحو مناسب.

وأحياناً يستخدم مصطلح الصدق الظاهري Face Validity لوصف الاختبارات ، أي المظهر التام لمحتويات الاختبار من حيث وضوح مفرداته ، أي قياس ما يرمى إلي قياسه. وتستخدم هذه العملية أحياناً كإجراء فرز مبدئي في اختيار الاختبار.

وصدق المضمون يمثل أهمية أولية في اختبارات التحصيل ، فدرجة الاختبار لا يمكن أن تعكس بدقة تحصيل الطالب إذا لم تقيس ما هو مفروض أن يتعلمه. وعلى حين يبدوا هذا جلياً ، فإن صدق المضمون أو المحتوى يشكل مشكلة لعدد من الدراسات البحثية ، فكثير من الدراسات تستهدف مقارنة فعالية طريقتين مختلفتين أو أكثر لتدريس نفس المادة ، وكثيراً ما تعرف الفعالية بالنسبة للتحصيل النهائي لمجموعات المعالجة كما يقيسها الاختبار ، وأحياناً حالة الاختبار المستخدم من ناحية صدق المضمون بالنسبة لمجموعة واحدة أكثر من الأخرى ، وإذا حدث ذلك ، فقد تعزى فروق التحصيل النهائية إلى الاختبار المستخدم وليس لطرق التدريس فقط. وغالبا تحدث هذه الظاهرة إذا ما قورنت طريقة مستحدثة مع طريقة تقليدية حيث يؤكد اختلافها اختلاف مجالات المضمون. والحالة الكلاسيكية الأولى هي الدراسات التي تقارن الرياضيات الحديثة مع الرياضيات التقليدية ، فهذه الدراسات لم تكتشف فروق تحصيل بين المتعلمين في ظل هاتين الطريقتين ، وتمثلت المشكلة في أن الرياضيات الحديثة أكدت على المفاهيم والمبادئ بينما أكدت اختبارات التحصيل على المهارة الحسابية. ولما طورت هذه الاختبارات التي تضمنت عينة كافية من المفردات لقياس المبادئ والمفاهيم ، كشفت الدراسات عن أن طريقتي تدريس الرياضيات أسفرت عن مقدرة حسابية متكافئة أساساً ، إلا أن الرياضيات الجديدة New Math أثمرت عن تفهم للمفاهيم أفضل ، ومغزى ذلك أنه في دراسة تقارن المعالجات وتقيس التحصيل ، يجب أن نمعن النظر إن كان الاختبار يقيس ما تعلمه الأطفال في المعالجة Treatment أي يتأكد الباحث أن الاختبار صادق بالنسبة لدراسته ، وصادق بالنسبة لأفراد عينته.

ويحدد صدق المضمون بواسطة خبير في التقدير ، فليس هناك صيغة تعبيرية حسابية ، أو وسيلة للتعبير عنه كمياً ، إنما يتم ذلك بتقييم يقوم به خبراء فيا لمجال الذي يغطيه الاختبار ، ويقوم هؤلاء الخبراء بمراجعة نقدية دقيقة للعملية المستخدمة في تطوير الاختبار ، وكذلك الاختبار نفسه مع تقدير لمدى تمثيل المفردات الجيد لمجال المضمون والنسب الصحيحة لهذا التمثيل ، وعبارة أخرى مقارنة بين ما يجب أن يتضمنه الاختبار بحسب ما يهدف إليه وبين ما يتضمنه فعلاً ، وعندما يختار الباحث اختباراً لدراسته البحثية ، عليه أن يتقمص دور الخبير في تحديد صدق مضمون الدراسة وأن يقارن ما سوف يتم تعلمه في هذه الدراسة وما يقيسه الاختبار.

صدق المفهوم Construct Validity

يقصد بصدق المفهوم مدى نجاح الاختبار في قياس مفهوم فرضي Hypothetical معين. والمفهوم هو سمة غير ملاحظة كالذكاء الذي يفسر السلوك ، فنحن لا نرى الذكاء الذي يفسر السلوك ، فنحن لا نرى الذكاء ، إنما يمكننا ملاحظة تأثيره. وفي الواقع اخترعت المفاهيم

لتفسير السلوك وإن كنا لا نستطيع إثبات وجودها ، فليس في مقدورنا اجراء جراحة في دماغ شخص ما لنرى ذكاء ، ومن ذلك للمفاهيم وظيفة مدهشة في تفسير فروق معينة بين الأفراد ، فعلى سبيل المثال يلاحظ دائماً أن بعض الطلاب يتعلمون أسرع من غيرهم بل ويتعلمون أكثر ، ويحتفظون بما تعلموه لفترة أطول ، ولتفسير هذه الفروق طورت نظرية الذكاء ، وافترض لذلك أن هناك شيئاً ما اسمه ذكاء يرتبط بالتعلم ، والذي يمتلك كل فرد قدراً منه قل أم كثر ، وطورت الاختبارات بهدف قياس ما يمتلكه الفرد من هذا الذكاء ، وكما هو الحال عليه يشار إلى الطلاب الذين يحصلون على درجات مرتفعة بأن ذكاءهم مرتفع High IQ وأن أداءهم سيكون أفضل في المدرسة في المدرسة وفي بيئات أخرى. ومن المفاهيم الأخرى المفترض وجودها وطورت لها اختبارات من بينها : القلق Anxiety ، الابتكارية Creativity وحسب الاستطلاع Curiosity.

والدراسات البحثية التي تشتمل على مفهوم ما إما كتغير مستقل وإما كمتغير تابع لا تكون صادقة إلا بمدى صدق مقياس المفهوم المتضمن. فالقلق على سبيل المثال يمكن أن يكون متغيراً مستقلا أو متغيراً تابعاً. فقد تستهدف الدراسة تحديد إن كان الطلاب مرتفعي القلق يؤدون المهام الصعبة أفضل من الطلاب منخفضي القلق. فاختبار القلق يستلزم أن يطبق على الطلاب في الدراسة حتى يمكن تصنيفهم إلى مرتفعي القلق أو منخفضي القلق.

وقد تستهدف دراسة أخرى تحديد إن كان التعليم بمساعدة الكمبيوتر يؤدى إلى قلق بين بطيئي التعلم أقل من التعليم التقليدي. لذا قد يستلزم الأمر تطبيق اختبار قلق على المجموعتين في نهاية الدراسة ، وفي الحالتين قد يكون صدق النتائج دالة مباشرة Direct Function على صدق اختبار القلق المستخدم. وإذا لم يقيس الاختبار حقيقة القلق فإن نتائج هذه الدراسة المبنية نتائج هذا الاختبار لا قيمة لها ، فعند اختيار اختبار خاص بمفهوم معين ، على الباحث أن يمعن النظر ، ويقيم الدليل القائم المرتبط بصدق مفهوم الأداء.

وعملية التحقق من صدق اختبار المفهوم ليست بالمهمة السهلة ، فهي تتضمن أساساً اختبار فروض مستنبطة من نظرية خاصة بالمفهوم ، فعلى سبيل المثال ، لو افترضت نظرية القلق أن الأشخاص مرتفعي القلق يبذلون جهداً بشأن مسألة ما أكبر من الأشخاص منخفضي القلق ، إذن لو حصل هؤلاء الأشخاص على درجات أعلى على اختبار قيد البحث ، فإنهم بالتأكيد سينجحون أكثر في مهمة تالية ، وهذا قد يكون دليلاً يدعم مفهوم صدق الاختبار ، وبالطبع إذا لم يبذل الأشخاص مرتفعي القلق جهداً أكبر كما هو مفترض ، فلا يعني هذا بالضرورة أن الاختبار لم يقيس القلق ، فالفروض المرتبطة بسلوك الأشخاص مرتفعي القلق يحتمل ان تكون غير

صحيحة ، وعموماً يستلزم الأمر اجراء عدد من الدراسات المستقلة للبرهنة على قابلية التصديق لاختبار المفهوم Credibility.

الصدق التلازمي Concurrent Validity

ويقصد به نوع العلاقة بين درجات اختبار ما وبين تقديرات محكم صادق متوافر أثناء تطبيق هذا الاختبار. فكثيراً ما يتم تطوير اختبار ما يدعى بأنه يقوم بنفس مهمة اختبار آخر بطريقة أسهل وأسرع ، وإذا كانت الحال كذلك ، فهذا ما يقصد به الصدق التلازمي لاختبار جديد سوف يستخدم في معظم الحالات بدلاً من اختبارات أخرى ، فاختبار الورقة والقلم A Paper and Pencil Test تقوم بنفس الوظيفة لاختبار الأداء وأن اختبار قصير يقيس المظاهر السلوكية كالاختبار الطويل له أفضلية بالتأكيد في دراسة بحثية.

ويحدد الصدق التلازمي عن طريق اثبات وجود علاقة Relationship أو تمايز Discrimination. وتتضمن طريقة العلاقة تحديد العلاقة بين اختبار ما وبين درجات اختبار آخر علاقة ثابتة أو محك صادق Criterion. وفي هذه الحالة تتضمن خطوات تحديد الصدق التلازمي ما يلي:

1- تطبيق اختبار جديد على مجموعة أفراد محددة.

2- تطبيق اختبار سبق اثبات صدقه على نفس المجموعة في نفس الوقت أو بعدئذ بقليل.

3- الربط بين مجموعتي الدرجات.

4- تقييم النتائج.

ويشير الرقم الناتج أو معامل الصدق إلى الصدق التلازمي للاختبار الجديد ، فإذا كان المعامل عالياً كان الاختبار جيداً في صدقه التلازمي. وعلى سبيل المثال ، نفترض أن البروفسور Jeenvus طور اختبار جماعي لقياس ذكاء الأطفال يستغرق تطبيقه خمس دقائق ، وإذا كانت نتائجه ترتبط ارتباطاً عالياً فعلياً بدرجات مقياس وكسلر للذكاء الخاص بالأطفال Wechsler Intelligence Scale الذي يستغرق تطبيقه ساعة على طفل واحد لكان بالتأكيد اختبار Jeenvus هو المفضل في كثير من المواقف.

وتتضمن طريقة التمايز في اثبات الصدق التلازمي تحديد إن كانت درجات الاختبار يمكن استخدامها للتمييز بين الأشخاص بين الأشخاص الذين لديهم سمات معينة وبين من ليست لديهم هذه السمات أو لا يملكون قدر أكبر منها. فعلى سبيل المثال: قد يكون اختبار التوافق الفعلي

ذا صدق تلازمي إذا كان الدرجات الناتجة عنه يمكن استخدامها من أجل تصنيف صحيح لأشخاص مودعين مصحة نفسية أو غير مودعين.

والعادة عند اختيار اختبار لهدف بحثي معين ، يبحث الباحث عن اختبار يقيس ما يرغب فيه بطريقة عالية الكفاءة ، فإذا اختار الاختبار الأقصر والمناسب أكثر والذي يزعم أنه يقيس السلوك المطلوب ، عليه أن يتأكد أن صدقه التلازمي قد ثبت وتمت الرهنة عليه باستخدام محك صادق Valid Criterion.

الصدق التنبؤي Predictive Validity

ويقصد به قدرة وفعالية الاختبار على التنبؤ بأداء الفرد في المستقبل في ضوء درجات الاختبار ، فاختبار الاستعداد لمادة الجبر ذو صدق تنبؤي عالي ، يمكن له التنبؤ بدقة بشأن أي الطلاب سيكون أداؤهم جيداً في مادة الجبر ومن هم دون ذلك.

والصدق التنبؤي في غاية الأهمية بالنسبة للاختبارات التي تستخدم في تصنيف أو اختيار الأفراد. من ذلك اختبارات توزيع الطلاب على فصول التربية الخاصة ، فالقرار بإزاحة طفل ما عن بيئة تعلم عادية ووضعه في فصل خاص ، لأمر خطير. وفي موقف كهذا من المهم أن يستند القرار إلى نتائج مقاييس صادقة Valid.

والصدق التنبؤي الذي يستخدم في توزيع الطلاب على الكليات بحسب مجموع درجاتهم في الثانوية العامة يوضح أن الصدق التنبؤي لأداة معينة يتنوع بحسب عدد من العوامل مثل: المنهج المتضمن ، الكتب الدراسية المستخدمة ، الموقع الجغرافي .. وهكذا إذا ما استخدمت درجات الامتحان كأساس للتنبؤ فمن المهم وصف حالة صدقها مع الموقف الذي تستخدم فيه.

وبطبيعة الحال لا يوجد اختبار ذو صدق تنبؤي كامل ، لذا فإن التنبؤات القائمة على درجات أي اختبار ستكون غير كاملة. ومع هذا التنبؤات التي تستند إلى الجمع بين درجات اختبارات عديدة تكون أكثر دقة من التنبؤات التي تستند إلى درجات اختبار واحد. لهذا إذا تعين القيام بتصنيف هام أو اتخاذ قرارات هامة ، فيجب أن يتم ذلك بناء على بيانات من أكثر من مؤشر. وعلى سبيل المثال : يمكن استخدام متوسط مجموع الثانوية العامة للتنبؤ بنجاح طالب صف أول جامعي ، كما يمكننا استخدام درجة الاستعداد المدرسي أو الترتيب في الفصل للتنبؤ بالنجاح في الثانوية العامة ، فالتنبؤ المبنى على متغيرين هو أكثر دقة من تنبؤ قائم على متغير واحد ، وكذلك سيكون التنبؤ على أساس ثلاثة متغيرات أكثر دقة.

ويتحدد الصدق التنبؤي لاختيار ما عن طريق انشاء علاقة بين درجات هذا الاختبار وبعض مقاييس النجاح في موقف يتسم بالأهمية. ويشار إلى الاختبار المستخدم للتنبؤ بالنجاح بوصفه المتنبئ Predictor ويشار إلى السلوك الذي يتنبأ به بالمحك Criterion.

والخطوة الأولى في إيجاد صدق تنبؤي لاختبار ما هي التعريف والتحديد الدقيق للمحك مما يستلزم أن يكون مقياساً صادقاً للسلوك المتنبأ به. وكمثال : إذا أردنا إثبات صدق تنبؤي لاختبار استعداد لمادة الجبر ، فمن المحتمل اعتبار درجات الامتحان النهائي هي المحك الصادق ولكن لا يمكن اعتبار أيام الغياب أثناء دراسة المنهج.

وعندما يعرف المحك ويحدد يكون أسلوب تحديد الصدق التنبؤي على النحو التالي:

1- تطبيق الاختبار ، المتغير المتنبئ ، على مجموعة ما.

2- الانتظار ريثما يقع السلوك المتنبأ به ، أي المتغير المحك.

3- الحصول على مقاييس المحك فيما يتعلق بنفس المجموعة.

4- الربط بين مجموعتي الدرجات.

5- تقييم النتائج.

ويشير الرقم الناتج أو معامل الصدق إلى الصدق التنبؤي للاختبار ، فإذا كان المعامل عالياً ، كان الاختبار جيد الصدق التنبؤي ، وكمثال إذا كان المطلوب تحديد الصدق التنبؤي لاختبار استعداد لمادة الفيزياء : أولاً- نجري اختباراً على مجموعة طلاب لديهم استعداد فيزيائي أكبر ، ثم ننتظر حتى استكمالهم لبرنامج في مادة الفيزياء والحصول على مقياس لنجاحهم وكمثال درجات الامتحان النهائي. وسوف يحدد الارتباط بين درجات الاختبار ودرجات الامتحان النهائي الصدق التنبؤي للاختبار. فإذا كانت معامل الارتباط الناتجة عالية ، كان الاختبار مرتفع الصدق التنبؤي.

وكما سبقت الإشارة ، تستخدم مجموعة متغيرات تنبؤية من أجل التنبؤ بالمحك ، وفي هذه الحالة يمكن تطوير معادلة التنبؤ Predication Equation حيث ادخال درجات الشخص في عدد من الاختبارات كل على حدة في المعادلة من أجل التنبؤ بما سيكون عليه أداؤه مستقبلاً. وفي هذه الحالة يجب أن يكون صدق المعادلة من خلال الصدق المقارن Cross – Validation الذي يتضمن تطبيق الاختبارات المتنبئة Predictor (المحكات) على عينة مختلفة من نفس المجتمع وتطوير معادلة جديدة.

ويلاحظ أن هناك تشابهاً بين الصدق التنبؤي والصدق التلازمي والفارق الرئيسي بينهما هو في وقت تطبيق المقياس ، فعند اثبات الصدق التلازمي نطبق الاختبار مع المحك في وقت واحد أما في اثبات الصدق التنبؤي ، يجب على الباحث الانتظار فترة طويلة من الوقت قبل تطبيق المحك حتى يمكن جمع البيانات.

ومن حين لآخر يمكن إحلال الصدق التلازمي محل الصدق التنبؤي توفيراً للوقت واستبعاداً لأية مشكلات خاصة بتتبع المفحوصين. وعلى سبيل المثال يمكن تطبيق اختبار الاستعداد الميكانيكي Mechanical Aptitude Test على مجموعة من الميكانيكيين وربط درجاتهم على الاختبار مع مقياس ما لمهاراتهم. والمشكلة مع هذا الأسلوب هي أننا نتعامل فقط مع الذين يعملون حالياً في العمل الميكانيكي ، والذين يحصلون على درجات ضعيفة هم أنفسهم من نتنبأ باعتبارهم فاشلين في عملهم الميكانيكي. فإن النتيجة هي الدليل على صدق الاختبار في قياسه للاستعداد الميكانيكي وذلك للارتباط الوثيق بين درجات الاختبار وبين مدى النجاح الذي هو الأداء الفعلي في عمل الميكانيكي ، فالعامل المتفوق في عمله يكون متفوقاً أيضاً على الاختبار ويل هذا على ان الاختبار صادق ، ونحصل على معامل الصدق في هذه الحالة عن طريق إيجاد معامل الارتباط بين درجات عينة التقنين على المحك ، ودرجاتهم على الاختبار. ومعامل الارتباط 0.50 يكون مقبولاً إذا كان هناك اختبار واحد فقط متوافر يستهدف التنبؤ بمحك معين. ومن جهة أخرى المعامل 0.50 قد يكون غير كاف إذا توافرت اختبارات أخرى ذات معاملات أعلى.

ويرتبط مفهوم الثبات بمفهوم الصدق ارتباطاً وثيقاً حيث يتناول ثبات واتساق درجة السؤال.

الثبات Reliability

يعني الثبات في اللغة الإنجليزية الاعتمادية Dependability أو كون الشيء موثوق به Trustworthiness. ويعني هذا المصطلح أساس نفس الشيء بالنسبة للمقياس. فأساساً يعني ثبات الاختبار مدى الدقة أو الاتساق الذي يقيس بها الاختبار ما يقيسه ، وكلما كان الاختبار أدق وأكثر ثباتا نستطيع أن نحصل على نفس الدرجات لنفس الأفراد عند تطبيقه عليهم بعكس الاختبار الذي لا يعتمد عليه حيث تختلف درجات عينة معينة في كل مرة نعيد فهيا تطبيق الاختبار ، وكمثال لا يكون الاختبار مؤشراً صالحاً للذكاء إذا كانت نسبة ذكاء الطالب اليوم هي 120 وغداً 140 وبعد غد 95 .. والاختبار الدقيق الثابت فهو إذا حصل الطالب على 110 فلن

نتوقع أي تقلب كبير في هذه النسبة إذا أعدنا تطبيق الاختبار فنسبة 150 مثلاً لن تكون عادية كما أن نسبة 145 غير محتملة.

ويعتبر عن الثبات رقمياً Numerically أي بمعامل الثبات ، وكلما كان هذا المعامل مرتفعاً كان الثبات مرتفعاً ، ومعامل الثبات الدقيق هو الواحد الصحيح أي 1.00 ، وهذا يعني أن درجة الطالب تعكس تماماً حالته الحقيقية بخصوص المتغيرات التي نقيسها. ولكن للأسف معامل الارتباط دائماً نسبي وليس له معنى مطلق ، فهو مستمد من التجربة ، ومن القدرات التي نقيسها ومن أدوات القياس المستخدمة ، ومن العينة ، وغيرها من العوامل المؤثرة في التجربة فدرجات الطالب تتأثر بأخطاء المقياس الناجمة عن تشكيلة من الأسباب ، فأعلى ثبات يشير إلى أقل انحراف خطأ ، وتؤثر أخطاء القياس في الدرجات بطريقة عشوائية حيث قد تزيد بعض الدرجات بينما تتناقص بعضها. وقد تنشأ أخطاء المقياس نتيجة لخصائص القياس ذاته (مثل بنود الاختبار الغامضة التي لا يستطيع الطالب تفسيرها تفسيراً واضحاً) وكذلك لظروف تطبيق الاختبار (مثل التوجهات غير المناسبة) ، وأيضا الحالة الراهنة للأفراد الذين يطبق عليهم الاختبار (قد يكون بعضهم متعباً ، والبعض الآخر غير مهتم) ، وكذلك لمجموعة من الأسباب وثبات الاختبار المرتفع يشير إلى قلة مصادر الخطأ لأدنى حد ممكن.

وأخطاء المقياس التي تؤثر في ثبات الاختبار هي أخطاء عشوائية : أخطاء منهجية (منتظمة) Systematic أو أخطاء ثابتة Constant ، تؤثر في صدقه. فإذا كان اختبار تحصيل يشكل صعوبة بالنسبة لمجموعة طلابية معينة فستكون الدرجات منخفضة بطريقة منتظمة ، وبالتالي سيكون صدق الاختبار منخفضاً بالنسبة لهذه المجموعة (تذكر صادق لمن) ، على أن الاختبار قد يسفر عن درجات متوافقة Consistent أي قد تكون ثابتة ، وبعبارة أخرى قد تكون الدرجات منخفضة على نحو منتظم بنفس الأسلوب كل مرة ، فالطالب ذو تحصيل حقيقي 80 قد يحصل على 60 على الاختبار (عدم الصدق) وربما يحصل على نفس 60 كل مرة يعاد فيها الاختبار (ثبات). وهذا يوضح العلاقة المثيرة بين الصدق والثبات .. فالاختبار الصادق ثابت دائماً ، ولكن الاختبار الثابت ليس بالضرورة صادقاً دائماً ، وبعبارة أخرى إذا كان الاختبار يقيس المطلوب قياسه فهو ثابت في كل مرة يطبق فيها ، ولكن الاختبار الثابت قد يقيس ثبات الشيء الخطأ وغير الصادق Invalid.

وقد يجرنا حديثنا عن الأخطاء إلى الاعتقاد بأن المقاييس التربوية غير متقنة نوعاً ما ، وغير حقيقية ، وهذا ليس صحيحاً فهناك العديد من الاختبارات التي تقيس السمات المستهدفة بدقة تامة ، وكما أشار Nunnolly إلى أن المقياس ينطوي في مجالات العلم الأخرى (غير التربية

وعلم النفس) على كثير من الخطأ العشوائي Random Error فمثلاً قياس ضغط الدم ، تلك السمة الفسيولوجية ، هي أقل ثباتاً من معظم المقاييس السيكولوجية ، حيث توجد كثير من الحالات الآتية (اللحظية) التي تؤثر فيها تأثيراً مؤقتاً في ضغط الدم كالفرح الغضب والخوف والقلق .. وغيرها. وهكذا فضغط الدم المقروء إنما هو نتيجة ومحصلة ضغط الدم الحقيقي والخطأ.

وتقييم الثبات أسهل من تقييم الصدق. فهناك عدد من أنواع الثبات المختلفة يتم تحديد كل منها بطريقة ، كما يتعامل كل منها مع نوع مختلف من اثبات ، ويتم تحديد درجات الثبات بدراسات تجريبية واحصائية منها .. ثبات الإعادة Test-Retest Reliability ، ثبات الصور المتكافئة Equivalent Forms ، والثبات النصفي Split-Half Reliability ، وكلها تحدد من خلال الارتباط Correlation ، ويتم ثبات التكافؤ المنطقي (العقلاني) Rationale Equivalence Reliability عن طريق تحديد مدى ارتباط كل بند من بنود الاختبار الأخرى والاختبار ككل. ويعتبر كلاً من الثبات النصفي وثبات التكافؤ المنطقي نوعان من ثبات الاتساق الداخلي Internal Consistency Reliability وهما يقومان كما يفهم من اسمهما على الاتساق الداخلي للاختبار Test Internal Consistency Reliability بينما ثبات إعادة الاختبار ، وثبات الصور المتكافئة يتطلبان مجموعة ما لتطبيق الاختبارين عليها (إما نفس الاختبار مرتين ، أو صورتان من نفس الاختبار). ويمكن تقدير ثبات الاتساق الداخلي بناء على تطبيق اختبار ما على جماعة ما.

ثبات الإعادة Test-Retest Reliability

ويعني ثبات الإعادة مدى اتساق الدرجات مع الوقت أي مرة بعد مرة. كما يشير إلى اختلاف في الدرجات من جلسة اختبار إلى جلسة اختبار أخرى نتيجة لأخطاء المقياس. وبعبارة الذي يثير اهتمامنا هو الحصول على دليل يفيد أن الدرجة التي يحصل عليها الفرد على اختبار ما في وقت ما هي نفس الدرجة أو قريبة منها إذا أعيد تطبيق الاختبار عليه في وقت آخر ، أي معرفة مدى اتساق ما يقيسه الاختبار كلما طبق. وهذا النوع من الثبات هام بصفة خاصة بالنسبة للاختبار الذي يستخدم كمتنبئ Predictor أو اختبار الاستعداد Aptitude.

ويكون تحديد ثبات الإعادة مناسباً إذا لم تتوافر صورة بديلة (متكافئة) للاختبار وعندما لا يكون من المحتمل أن يتذكر الأفراد اجابتهم على الاختبار الأول إذا ما طبق عليهم الاختبار لمرة ثانية. فمن المرجح أن يتذكروا بنوداً من اختبار ملئ بحقائق تاريخية أكثر من تذكرهم لاختبار ملئ بمسائل الجبر.

وطريقة تحديد إعادة الثبات هي أساساً بسيطة على النحو التالي:

1- تطبيق الاختبار على مجموعة مناسبة.

2- تطبيق الاختبار مرتين على نفس المجموعة بحيث يفصل بين المرتين فاصل زمني كأسبوعين مثلاً.

3- حساب الارتباط بين درجات المرة الأولى ودرجات المرة الثانية (مجموعتي الدرجات).

4- تقييم النتائج.

ويشار إلى معامل الارتباط الناتج بمعامل الثبات ، وإذا كان مرتفعاً كان معنى هذا ثبات جيد لإعادة الاختبار. والمشكلة الرئيسية مع هذا النوع من الثبات هي صعوبة تحديد أو معرفة الوقت المنقضي بين جلستي الاختبار. فإذا كان الفاصل الزمني قصيراً جداً تزيد فرصة تذكر الطلاب لإجاباتهم على الاختبار في المرة الأولى ، وكذلك إذا كان الفاصل الزمني طويلاً جداً فقد تزيد قدرة الطلاب على الأداء الجيد في الاختبار نتيجة التعلم الوسيط Intervening أو النضج Maturation ، وهكذا يكون تقدير الثبات منخفضاً على نحو غير حقيقي Artificially ويصعب تحديد هذا الفاصل الزمني بين الاختبار وإعادة الاختبار ، حيث يعتمد الوقت المثالي على نوع الاختبار المتضمن ، كما يعتبر تحديد بأقل من أسبوع مدة قصيرة جداً.

وهذه المشكلات التي ترتبط بثبات الإعادة تنال قدراً من الاهتمام في ثبات الصور المتكافئة.

ثبات الصور المتكافئة Equivalent Forms Reliability

أي تتطابق الصور المتكافئة لاختبارين ما في كل الجوانب ، فهما يقيسان نفس المتغير ، ونفس العدد من البنود (المفردات) ، ونفس التركيبة ، ونفس مستوى الصعوبة ، ونفس تعليمات التطبيق والتصحيح والتفسير.

وفي الواقع إذا طبق الاختباران على نفس المجموعة يجب أساساً أن تكون متوسط الدرجات وكذلك إمكانية تغير الدرجات هي نفسها على الاختبارين. فقط لا تكون القيود النوعية وحدة على الرغم انها تقيس نفس السمات ونفس الأهداف. وفي الأساس نختار أو نقوم بإجراء معاينة لبنود مختلفة من نفس المجال السلوكي ، فإذا كانت البنود مختارة جيداً ، وتمثل كل مجموعة تمثيلا كافياً للمجال محل الاهتمام كان الاختبار الآخر جيداً.

ويشار إلى ثبات الصور المتكافئة أيضا بثبات الصور البديلة Alternate وأحياناً في بعض الدراسات البحثية تطبق صورتي الاختبار على نفس المجموعة صورة كاختبار قبلي Pretest والأخرى كاختبار بعدي Posttest.

وطريقة تحديد ثبات الصور المتكافئة مماثلة جداً لتحديد ثبات الإعادة:

1- تطبق إحدى صورتي الاخبار على مجموعة مناسبة.

2- في نفس الجلسة ، أو بعدها بوقت قصير تطبق الصورة الأخرى على نفس المجموعة.

3- حساب الارتباط بين درجات الصورتين.

4- تقييم النتائج.

ويشار إلى المعامل الناتج بمعامل التكافؤ Coefficient of Equivalence ، وإذا كان عالياً كان الاختبار ذا ثبات صور متكافئة جيد. وإذا تم تطبيق صورتي الاختبار في وقتين مختلفين مناسبين يشار إلى المعامل الناتج بمعامل الثبات والتكافؤ Coefficient of Stability and Equivalence وتجمع هذه النظرية أساساً بين ثبات الإعادة وثبات الصور المتكافئة ، وهكذا يقيم ثابت الدرجات مع الوقت فضلاً عن إمكانية تعميم مجموعات البنود. وبما أن كثيراً من مصادر أخطاء القياس محتملة أكثر من استخدام أي طريقة من الطريقتين على حدة ، من المحتمل أن يكون المعامل الناتج منخفضا نوعاً ما ، وهكذا يمثل معامل الثبات والتكافؤ تقديراً متحفظاً للثبات.

وثبات الصور المتكافئة هو الأكثر استخداماً في تقدير ثبات معظم الاختبارات المستخدمة في الأبحاث ، والمشكلة الرئيسية التي تنطوي عليها هذه الطريقة في تقدير الثبات هو صعوبة تكوين صورتين متكافئتين ، حيث أن فقدان التكافؤ هو من مصادر خطأ القياس ، وعلى الرغم أن ثبات الصور المتكافئة يعتبر أفضل تقدير للثبات فإنه ليس من المفيد دوماً تطبيق صورتين مختلفتين لنفس الاختبار أو أداء نفس الاختبار مرتين. ومن حسن الحظ أن هناك طرقاً أخرى لتقدير الثبات لا تتطلب تطبيقاً للاختبار إلا لمرة واحدة.

الثبات النصفي Split – Half Reliability

الثبات النصفي هو نوع شائع لثبات الاتساق الداخلي ، وبما أن طرق الثبات النصفي تتطلب الاختبار لمرة واحدة فقط ، فإن بعض مصادر خطأ القياس يتم استبعادها كالاختلافات بين ظروف الاختبار التي يمكن أن تقع عند إيجاد ثبات الإعادة. وطريقة الثبات النصفي مناسبة

لاسيما إذا كان الاختبار طويلاً جداً و/أو يصعب تطبيق نفس الاختبار في مرتين مختلفتين أو تطبيق صورتين مختلفتين على مجموعة ما.

وأسلوب تحديد الثبات النصفي كما يلي:

1- تطبيق كل الاختبار على مجموعة ما.

2- تقسيم الاختبار إلى نصفين متكافئين تقريباً ، والأسلوب الأكثر شيوعاً هو تقسيمه إلى نصفين : نصف فردي ، ونصف زوجي. أي نصف يشتمل على بنود الاختبار ذات الأرقام الفردية (مثلاً في اختبار من عشرين بنداً) 1 ، 3 ، 5 ، 7 ، 9 ، 11 ، 13 ، 15 ، 17 ، 19 والنصف الآخر يشتمل على بنود الاختبار ذات الأرقام الزوجية 2 ، 4 ، 6 ، 8 ، 10 ، 12 ، 14 ، 16 ، 18 ، 20 وسيكون البندان 1 ، 2 سهلين ، والبندان 3 ، 4 صعبان وهكذا ..

3- حساب درجة كل مفحوص على النصفين ، وبالتالي يكون له درجتان ، درجة على البنود الفردية ودرجة على البنود الزوجية.

4- حساب الارتباط بين مجموعتين الدرجات.

5- تطبيق معادلة Spearman – Brown لتصحيح المعامل النصفي.

وإذا كان المعامل عالياً كان الاختبار ذا ثبات نصفي جيد. ويمكن استخدام عدد من الطرق المنطقية والاحصائية لتقسيم الاختبار إلى نصفين. واستراتيجية تقسيم الاختبار إلى نصفين فردي وزوجي الأكثر استخداماً.

وحيث أنا الاختبارات ككل هي الأكثر ثباتاً ، وحيث أن الثبات النصفي يمثل فقط نصف الاختبار ، لذا يجب أن نلجأ إلى معادلة تصحيح معامل الثبات النصفي باستخدام معادلة التنبؤ لسبيرمان – بروان Spearman Brown Prophecy Formula.

فعلى سبيل المثال .. لو كان معامل الثبات النصفي لاختبار من 50 بنداً هو 80 ، فهذه النسبة قد تكون مبنية على الربط بين درجات بنود النصف الفردي 25. ودرجات بنود النصف الزوجي 25. وبالتالي هي تقدير لثبات نصفي لاختبار من 25 بنداً وليس لاختبار من 50 بنداً. لذا نحتاج لتصحيح هذا الوضع أن نطبق معادلة سبيرمان – براون لتقدير الثبات (R) لاختبار من 50 بنداً والمعادلة بسيطة على النحو التالي:

$$\text{الثبات الكلي} = \frac{\text{ضعف الثبات النصفي}}{\text{الثبات النصفي} + 1} \quad R_{total} \frac{2\, r_{split\ half}}{1 + r_{split\ half}}$$

وبتطبيق هذه المعادلة على المثال المذكور فإن الثبات الكلي =

$$R_{total\ test} = \frac{2\,(0.80)}{1 + 0.80} \quad \frac{1.60}{1.80} = 89$$

هكذا تم تصحيح تقدير الثبات النصفي إلى 89.

ومن المشكلات المرتبطة بمعادلة التصحيح هذه ، هي اتجاهها إلى تقدير ثبات أعلى مما قد نحصل عليه بطرق أخرى (أي تبالغ في التقدير).

ومن طرق تحديد الاتساق الداخلي .. طريقة التكافؤ المنطقي.

ثبات التكافؤ المنطق (العقلاني) Rationale Equivalence Reliability

ويتحدد ثبات التكافؤ المنطقي عن طريق تطبيق معادلتي Kinder – Richardson التي ينتج عنها تقدير ثبات مكافئ أساساً لمتوسط ثبات نصفي ، وكلا المعادلتين KR20 و KR21 تتضمنن أسئلة ثنائية الإجابة بنعم أو لا صواب أو خطأ ، 1 أو صفر.

واستخدام المعادلة 21 يستلزم وقتاً أقل من أي طريقة أخرى لتقدير الثبات ويسفر تطبيقها أيضاً عن تقدير متحفظ للثبات ولاسيما عند قياس أكثر من سمة والمعادلة على النحو التالي:

$$\text{الثبات الكلي } r_{total\ t} \frac{(K)\,(SD^2) - X\,(K\text{-}X)}{(SD^2)\,(K - 1)}$$

حيث أن K = عدد بنود الاختبار.

SD = الانحراف المعياري للدرجات.

X = متوسط الدرجات (الوسط الحسابي).

وفي الجزء الثاني من الكتاب الحالي (الفصل 12) سوف نتناول هذا الموضوع بشيء من التفصيل.

ويلخص الشكل التالي طرق تقدير ثبات الاختبار ، حيث يشير الرقم (1) إلى زمن تطبيق اختبار واحد (بالثبات النصفي أو KR 21) ، أو تطبيق الاختبارين في وقت واحد (ثبات الصور المتكافئة). وحيث يشير الرقم (2) إلى وجود فاصل زمني (أسبوعين مثلاً) بين تطبيق الاختبارين (ثبات الإعادة وثبات التكافؤ).

ملخص طرق تقدير الثبات		
عدد مرات تطبيق الاختيار	عدد الاختبارات المختلفة	
	اختبار واحد 1	اختباران 2
مرة واحدة 1	ثبت نصفي KR 21	ثبات الصور المتكافئة
مرتان 2	ثبات الإعادة	الثبات والتكافؤ

ثبات المصحح / المقدر Scorer / Rater Reliability

ثمة مواقف أخرى يجب دراسة الثبات فيها. وعادة تحدث هذه المواقف إذا انطوت تقديرات الاختبارات على الذاتية Subjectivity كاختبارات المقال Essay Tests ، والاختبارات قصيرة الإجابة التي تتضمن الإجابة بأكثر من كلمة Short – Answer Test ، ومقياس التقدير Rating Scales ، وأدوات الملاحظة Observation وهي المواقف المرتبطة بثبات المصحح / مقدر الدرجات (Interscorer, interrater, interobserver) أي الثبات بين المصححين Inter judge reliability ويشير إلى ثبات تصحيح الاستجابات على الاختبار وتقديرها لدى اثنين أو أكثر من المصححين المستقلين Independent Scorers.

ويشير ثبات المصحح الداخلي Intra judge reliability إلى ثبات تقدير المصحح كفرد وتصحيح الاختبارات وتقديرها هي من مصادر أخطاء القياس لذا من المهم التأكد من وجود

ثبات أو درجة مقبولة من الاتفاق بين المصححين على تصحيح الاستجابات على الاختبارات وتقديرها. وهذه الدرجة النهائية يمكن الحصول عليها باستخدام التقنيات الارتباطية ، كما يمكن التعبير عنها باتفاق مئوي (بنسبة مئوية) percent agreement. وفي الوقت الذي تكون فيه بعض درجات الثبات غير جيدة ، فقد تم تطوير عدد من الأدوات المقننة التي تجعل ثبات المصححين يبدو جيداً نسبياً وكمثال دراسات المصداقية Validation التي تتضمن نظام او طريقة Flanders المتعلق بملاحظة السلوك اللفظي داخل الفصل ويشير إلى مدى ثبات من 0.85 إلى اقل من 0.90.

معامل الثبات Reliability Coefficients

معامل الثبات الذي يزيد عن 0.90 هو بالطبع يشكل المستوى المقبول للثبات لأي اختبار ، ولكن السؤال سيتعلق بالحد الأدنى المقبول. وبالنسبة لاختبارات التحصيل والاستعداد لا يوجد سبب هام لاختيار اختبار ذي ثبات اقل من 0.90 ويوجد عدد من اختبارات التحصيل والاستعداد تتوافر فيها هذه النسبة من الثبات وبالتالي لا توجد هناك ضرورة لقبول الأقل (وإن كان البعض يفعل ذلك). ولكن عندما يتم تطوير اختبارات جديدة في مجالات جديدة فقد يستقر المرء عادة في البداية على ثبات أقل. وكمثال الاختبارات التي تقيس حب الاستطلاع Curiosity والى هي حالة هامة نسبياً في حقل الاختبارات. حيث لا يتوقع درجة ثبات عالية لهذه الاختبارات الجديدة في هذه المرحلة من تطورها ، والمهم هو التحقق من أن الاختبار يعتمد عليه Reliable بالنسبة لمجموعتك ، وأنه يوفر بيانات ذات قيمة لمطوري الاختبارات والباحثين.

وإذا كان الاختبار يتكون من مجموعتين جزئية Subtests عديدة فيلزم تعميم ثبات كل منها وليس فقط ثبات الاختبار ككل. كما أن فحص ثبات هذه المجموعات الجزئية للاختبار ذات أهمية خاصة إذا كان احداها أو أكثرها سيستخدم في البحث بدلاً من الاختبار ككل. وأنت كباحث يجب أن تكون مستهلكاً (مستفيداً) جيداً a good consumer لمعلومات الاختبار.

الخطأ المعياري للقياس Standard Error of Measurement

يمكن التعبير عن الثبات أيضاً في ضوء الخطأ المعياري للقياس ، ويعني هذا الخطأ أساساً ما يتوقعه الباحث من خطأ محتمل على عينة معينة. وهكذا يدل الخطأ المعياري المنخفض على ثبات عالٍ ، كما يدل الخطأ المعياري المرتفع على ثبات أقل.

لو كان الاختبار كامل الثبات Perfectly Reliable فإن الدرجة التي يحصل عليها الفرد تمثل درجة الحقيقة. ولكن والحالة هذه ، فالدرجة التي يحصل عليها إنما هي تقدير لدرجته

الحقيقية ، وإذا طبقت نفس الاختبار مراراً وتكراراً على نفس المجموعة ، قد تختلف درجة كل فرد ومدى الاختلاف القائم في الدرجات الفردية قد يكون دالة على ثبات الاختبار. فالاختلاف يكون منخفضاً بالنسبة لاختبار عالي الثبات (صفراً للاختبار كامل الثبات) وكبيراً للاختبار منخفض الثبات.

وإذا استطعنا تطبيق الاختبار مرات عديدة ، نستطيع ملاحظة مدى الاختلاف الفعي ، وفي الواقع لا نستطيع ذلك ، فتطبيق نفس الاختبار على نفس المجموعة مهمة شاقة ، ومن حسن الحظ من الممكن تقدير درجة الاختلاف (الخطأ المعياري للمقياس) باستخدام البيانات الناتجة من تطبيق الاختبار مرة واحدة على جماعة ما. وبعبارة أخرى يتيح لنا الخطأ المعياري للمقياس تقدير مدى الاختلاف المحتمل وجوده بين الدرجة التي يحصل عليها الفرد والدرجة الحقيقية ، وحجم هذا الاختلاف هو دالة على ثبات الاختبار.

ويمكن للباحث تقدير الخطأ المعياري للمقياس باستخدام المعادلة الآتية:

$$SE_m = SD\sqrt{1-r}$$

حيث أن SE_m = الخطأ المعياري للمقياس

SD = الانحراف المعياري لدرجات الاختبار

R = معامل ثبات الاختبار

وكمثال بالنسبة لاختبار من 25 بنداً يمكن حساب انحرافه المعياري عند 5 (SD = 5) وثبات نصفي عند 84(R = 0.84) ففي هذه الحالة يمكن حساب الخطأ المعياري للمقياس على النحو التالي:

$$SE_m = SD\sqrt{1-r}$$

$$SE_m = 5\sqrt{1-84}$$

$$= 5\sqrt{0.16}$$

$$= 5\sqrt{(0.4)}$$

$$= 2.0 \qquad \text{(وتعتبر متوسطة)}$$

وكما يوضح المثال فإن حجم SEm الخطأ المعياري للمقياس هو د الة على كل من الانحراف المعياري SD ومعامل الثبات R ومعامل الثبات الأعلى يرتبط بخطأ معياري للمقياس

obeikandl.com

منخفض كما يرتبط الانحراف المعياري المنخفض بخطأ معياري للمقياس منخفض. وهكذا إذا كان الخطأ المقياس = 5 سيكون كبيراً لاختبار من 20 بند ، ولكنه سيكون منخفضاً بالنسبة لاختبار من 200 بند ، وفي المثال المذكور سيكون الخطأ المعياري للمقياس الذي = 2.0 متوسطاً.

أنواع الاختبارات Types of Tests

ثمة أنواع مختلفة من الاختبارات وأساليب عديدة لتطبيق ، وتعتبر الكتب السنوية للقياسات النفسية Mental Measurement Yeabooks مصدراً رئيسياً لمعلومات الاختبار بالنسبة اللباحثين التربويين. وهي تتضمن إضافة إلى عدد من المناهج في اللغة الإنجليزية والرياضيات القراء عناوين رئيسية للشخصية ، والذكاء والاستعداد الدراسي ، والكلام Speech والسمع Hearing ، والمهن Vocations ويعرض الجدول التالي قائمة كاملة بالتصنيفات الرئيسية التي يضمها الكتاب السنوي العشار للقياسات النفسية وكذلك العدد والنسب المئوية لبنود الاختبارات ذات الصلة.

وهذا هو الجدول

التصنيف	العدد	النسبة المئوية
المهن	100	25.3
الشخصية	72	18.2
موضوعات مختلفة	43	10.9
النمو	31	7.8
الذكاء والاستعداد المدرسي	28	7.1
اللغة الإنجليزية	24	6.1
القراءة	24	6.1
الكلام والسمع	22	5.6

obeikandl.com

التربية	20	5.1
التحصيل	12	3.0
الرياضيات	9	2.2
العلوم الاجتماعية	3	0.8
الفنون الجميلة	2	0.5
اللغات الأجنبية	2	0.5
حريك / حسني Sensory / Motor	2	0.50
سيكولوجي عصبي Neuropsychologal	2	0.50
اجمالي Total	396	100.0

التحصيل (الأكاديمي) Achievement

تقيس اختبارات التحصيل الوضع الراهن للأفراد فيما يتصل بمستوى كفايتهم في مجالات معينة للمعرفة أو المهارة ، ويجرى تطوير اختبارات تحصيل مقننة بدقة لكي اشتمل على قياس أهاف تشترك فيها أنظمة مدرسية عديدة ، فهي تقيس معرفة الحقائق والمفاهيم والمبادئ. ويقارن مستوى تحصيل الفرد على الدرجة المعيارية او المتوسطة بالنسبة أو مستوى تحصيل الفرد على الدرجة المعيارية أو المتوسطة بالنسبة لسنة أو مستوى صفة Age or Grade.

وتتوافر اختبار التحصيل المعيارية (المقننة) في مناهج خاصة كالقراءة والرياضات ، وكذلك في صورة بطاريات شاملة لقياس مستوى التحصيل في عدد من المجالات المختلفة ومن الأمثلة المعروفة لذلك اختبار ستانفورد للتحصيل The Stanford Achievement Test ، واختبار متروليتان للتحصيل The Metropolitan Achievement Text وبطارية اختبارات كاليفورنيا للتحصيل The California Achievement Test battery.

وهذه الاختبارات تتضمن مجموعات جزئية تقيس التحصيل في مجالات كالقراءة والرياضيات والهجاء والدراسات الاجتماعية ، والعلوم ، وفهم المادة المسموعة Listening

openkanaan.com

Comprehension ، وقد تستغرق البطارية ساعة واحدة إلى عدة أيام للتطبيق اعتماداً على عدد العوامل كعدد مجالات التحصيل التي يتضمنا الاختبار.

وتستهدف عديد من الدراسات البحثية مقارنة فعالية نظريتين منهجيتين أو أكثر ، أو طرق تعليم. وتعني الفعالية هنا قدرة الطالب على التحصيل في نهاية الدراسة.

وأحياناً تصمم الدراسات من أجل بحث أوجه عديدة للتعليم العلاجي (التدريب العلاجي) وهي تستخدم أدوات تشخيصية لذلك. والاختبار التشخيصي هو نوع الاختبارات التحصيلية يسفر عن درجات متعددة لكل مجال تحصيلي ، وتسهل هذه الدرجات التعرف على جوانب نوعية لعدم الكفاءة. ومن أمثلة هذه الاختبارات : اختبارات ستانفورد للقراءة التشخيصية The Stanford Diagnostic Reading Test ، وهذه الأداة التشخيصية تستخدم على نطاق واسع.

التشخصية Personality

تستهدف اختبارات الشخصية قياس خصائص الفرد على عدد من الأبعاد ، وتقييم المشاعر والاتجاهات نحو الذات ونحو الآخرين ، ومجموعة متنوعة من الأنشطة الأخرى ، والمؤسسات ، والمواقف ، واختبارات الشخصية أكثر الاختبارات استخداما في البحوث التربوية ، وتأتي في صور متعددة [الأداء مقابل الورقة والقلم] Performance Versus paper and Pencil ، الفرد مقابل الجماعة Individual Versus Group ، القدرة مقابل السرعة Power Versus Speed وهناك اختبارات تكاد تقيس كل جوانب الشخصية.

وبرغم التصنيفات العديدة لاختبارات الشخصية ، فإن التمايز المبدئي المنطقي الذي تستخدمه الكب السنوية Year Books هو تقسيم هذه الاختبارات إلى اختبارات غير اسقاطية Non Projective أو اختبارات اسقاطية Projective.

الاختبارات غير الاسقاطية Non-projective Tests

معظم اختبارات الشخصية والطابع المميز لها هي اختبارات غير اسقاطية أو تقرير ذات Self-Report وفي هذه الاختبارات ، يطلب من الفرد الاستجابة إلى سلسلة من الأسئلة أو العبارات. وهذه الاختبارات لمجموعات مختلفة كالمتسربين دراسياً من المدرسة الثانوية (High School Dropouts) وأيضا في الدراسات الارتباطية (كمثال لتحديد العلاقة بين سمات الشخصية المختلفة وغيرها من المتغيرات كالتحصيل الدراسي) وكذلك الدراسات التجريبية (كمثال لبحث الفعالية لبحث الفعالية المقارنة بطرق التعليم المختلفة بالنسبة لأنماط الشخصية المختلفة).

قوائم الشخصية Personality Inventories

وهي تعرض لقوائم من الأسئلة أو العبارات التي تصف خصائص سلوكية لسمات شخصية معينة ويطلب من المفحوص أن يجيب عليها بنعم أو لا أو غير متأكد (محايد) undecided النسبة للعبارة التي تصف. وبعض القوائم تعرض كقائمة تحقق chick list التي تميزه وتحدد درجة الفرد بعدد الاستجابات المميزة للسمة المقاسة. والمتوقع أن يجيب الشخص المنطوي Introvert بنعم على عبارة "القراءة هي أحب هواياتي" وبلا على عبارة "أحب الحفلات الصاخية".

وقد تكون قوائم الشخصية أكثر تحديداً حيث تقتصر على سمة واحدة كالإنطواء والانبساط Introversion – Extroversion ، أو قد تكون عامة وتقيس عدداً من السمات. وعادة المقاييس العامة التي تقيس أكثر من سمة في نفس الوقت طويلة نسبياً ويستغرق إتمامها ساعة على الأقل.

وتتضمن القوائم التي تستخدم عادة في دراسات البحوث التربوية على ما يلي:

- قائمة الصفات Adjective list
- قائمة كاليفورنيا السيكولوجية California Psychological Inventory
- قائمة ادواردز للتفصيل الشخصي

  Edwards Personal preference Schedule
- قائمة مينسوتا متعددة الأوجه للشخصية

  Minnesota Multiphasic Personality Inventory
- قائمة موني للمشكلات Mooney Problem Check List
- مؤشر مايز برجز للنمط Myers Briggs Type Indicator
- استبيان عوامل الشخصية الستة عشر

  Sixteen Personality Factor Questionnaire

وقد استخدمت قائمة مينسوتا متعددة الأوجه للشخصية لوحدها في مئات من الدراسات البحوث التربوية.

obeikandl.com

ومن المشاكل الهامة التي تتعلق بقوائم التقرير الذاتي ، الحصول على إجابات دقيقة فدرجات الشخصية لا تكون صادقة إلا إذا كان المستجيب أميناً ويختار من الإجابات ما يميزه حقيقة.

فالظاهرة العامة هي مفهوم اتجاه الاستجابة Response Set أو ميل الفرد Tendency للاستجابة بطريقة معينة. واتجاه الاستجابة بصفة عامة هي اتجاه الفرد إلى اختيار الاستجابات التي يعتقد أنها أكثر قبولاً من الناحية الاجتماعية وسواء أكان ذلك نتيجة عن دوافع شعورية أو دوافع لا شعورية فهي قد تشوه من تقييم الفرد لبنيته الشخصية.

وإذا كانت استجابات نسبة كبيرة من عينة البحث غير أمنية ، فإن نتائج الدراسة ستكون بلا معنى ، وبالتالي لابد من بذل كل الجهد في الدراسات التي تستخدم هذا النوع من الاختبارات من أجل زيادة احتمالية الحصول على نتائج اختبار صادقة.

مقاييس الاتجاه Attitude Scales

هي وسيلة لمعرفة موقف الفرد إزاء مبدأ أو قصة أو موضوع أو شخص. ويمكن قياس الاتجاهات نحو الذات ونحو الآخرين ومختلف الأنشطة والمؤسسات والمواقف. وهناك أربعة أنواع أساسية من مقاييس الاتجاه ات : مقاييس ليكرت Likert Scales ، مقاييس التمايز الدلالي (دلالة الألفاظ) Semantic Differential Scales ، ومقاييس ثرستون Thurston Scales ، ومقاييس جوتمان Guttmann Scales ، ومقاييس النوعين الأوليين هي الأكثر استخداماً ولكن على الباحث أن يلم بالنوعين الآخرين.

مقياس ليكرت Likert Scale

حيث يطلب من الفرد أن يحدد درجة قبوله أو رفضه لسلسلة من العبارات بالإشارة إلى الموافقة الشديدة Strongly Agrees ، الموافقة Agrees ، الحياد Undecided ، الرفض Disagrees ، الرفض الشديد Strongly Disagrees. وتعطي كل إجابة قيمة رقمية معينة ، وتحدد درجة الفرد بمجموع الدرجات التي حصل عليها في كل عبارة. درجة الموافقة الشديدة = 5 ، الموافقة = 4 ، الحياد = 3 ، الرفض = 2 ، الرفض بشدة = 1.

مقياس التمايز الدلالي (دلالة الألفاظ) A Semantic Differential Scale

حيث يتطلب من المفحوصين أن يعطي تقديراً رقمياً على موضوع مقياس الاتجاه ، على عدد من الصفات ثنائية القطب Bipolar مثل : جيد – سيئ Good – Bad ، ودي – غير ودي Friendly – Unfriendly ، إيجابي – سلبي Positive – Negative.

obeikandi.com

ويشير المستجيب إلى نقطة على متصل Continuum بين طرفين بما يمثل اتجاهه. وكمثال على مقياس بالاتجاهات نحو الضرائب على الممتلكات قد يتضمن البنود التالية:

ضروري Necessary ، غير ضروري Unnecessary

عائلة Fair ، غير عادلة Unfair

وفي الواقع العملي ، يتم تجاهل هذه الأبعاد و/أو تحل محلها أبعاد أخرى يعتقد أنها مناسبة في موقف خاص. وكل موقع على المتصل يرتبط بدرجة ، وبجمع درجات كل البنود يمكن تحديد اتجاهات المستجيب بكونها إيجابية أو سلبية.

ومقاييس التمايز الدلالي ذات 5 إلى 7 فواصل مع اتجاه محايد يخصص بدرجة صفر ، وبالنسبة عالية قد تكون قيم الدرجات على النحو التالي:

غير ضروري - - - - - - - ضروري

Unnecessary3 2 1 0 1- 2- 3-Necessry

غير عادلة Unfair - - - - - - - عادلة Fair

3- 2- 1- 0 1 2 3

والشخص الذي يشير إلى المفاصلة الأولى 3 على كلا البندين قد يدل اتجاهه الإيجابي نحو ضرائب الممتلكات.

مقياس ثرستون A Thurston Scale

حيث يطلب من الفرد أن يختار من قائمة العبارات التي يؤيدها فقط. وكل بند يقع في 11 فئة متدرجة من 1 إلى 11 ، وتكون درجته النهائية على المقياس هي متوسط درجاته على العبارات التي وافق عليها.

مقياس جوتمان A Guttmann Scale

يطلب أيضا من المفحوص ابداء موافقته أو رفضه إزاء عدد من العبارات وبهذا يحاول مقياس جوتمان تحديد إن كان الاتجاه أحادي البعد Uni-dimensional فو ذو اتجاه واحد إذا نتج عنه مقياساً تراكمياً. وفي هذا المقياس: الفرد الذي يوافق على عبارة معينة يوافق أيضا على

obeikandl.com

كل العبارات السابقة المرتبطة بها ، وكمثال إذا وافق المفحوص على عبارة 3 فكأنه وافق على العبارتين 1 ، 2.

ومقاييس الاتجاه نحو الذات يشار إليها بمقاييس مفهم الذات Self-Concept ، وهي تستخدم في كثير من الدراسات البحثية التربوية ولاسيما تلك التي تبحث في:

1- العلاقة بين مفهوم الذات ومتغيرات أخرى كالتحصيل.

2- تأثير مفهوم الذات على المتغيرات مثل الدافعية.

3- تأثير المناهج وطرق التدريس المتنوعة في مفهوم الذات.

مقاييس التقدير Rating Scales

وهي تستخدم أيضا مقياس الاتجاهات نحو الآخرين. وفي مقاييس التقدير يطلب من الفرد أني قدر فرداً آخر على عدد من الأبعاد السلوكية. وهناك نوعان رئيسيان من هذه المقاييس. نوع يتألف من بنود تطلب من الفرد تقدير آخر على متصل (من جيد إلى أسوأ ، ومن ممتاز إلى رديء) والنوع الآخر يطلب من المفحوص أن يقدر شخصاً آخر على عدد من البنود عن طريق اختيار فئة الاستجابة الأكثر ملاءمة [ممتاز ، فوق المتوسط ، متوسط ، تحت المتوسط ، رديء].

وهناك مشكلتان ترتبطان بمقاييس التقدير يشار إليهما بتأثير الهالة Halo Effect وخطأ الكرم Generosity Error.

وأثر الهالة هو انحياز المقدر ذاتياً أثناء حكمه على شخص ما ، كتقدير المدير المدرس بصفته أفضل مدرس على متغيرات لا ترتبط ارتباطاً مباشراً بالتدريس ودون وجود أساسي حقيقي لهذا الحكم.

وخطأ الكرم هو اتجاه المقدر لإعطاء من يقدره مزايا بدون معرفة كافية كلي يجعل حكمه موضوعياً. لهذا نحذر من انقياد الباحث وبعده عن الموضعية نتيجة لتأثير الهالة وخطأ الكرم ، وذلك بإعطاء التعليمات المناسبة للمتدربين.

وتعاني مقاييس الاتجاه من نفس مشكلات قوائم الشخصية فالباحث لا يستطيع أن يتأكد أن الفرد يعبر عن اتجاهه الحقيقي وليس باتجاه مقبول اجتماعياً. ونذكر مرة أخرى أن صدق الدراسة يرتبط ارتباطا مباشراً بصدق استجابات أفراد العينة ، لذا يجب بذل مزيد من الجهد لزيادة أمانة الاستجابة عن طريق إعطاء التوجيهات المناسبة.

اختبارات الابتكارية Creativity Tests

obeikandi.com

وتستهدف هذه الاختبارات في الحقيقة قياس الخصائص الشخصية المرتبطة بالسلوك الابتكاري كالتفكير التباعدي Divergent Thinking وهو يناقض التفكير التقاربي Convergent Thinkers الذي يتركز في التفكير على البحث عن الاستجابة الصحيحة – في حين أن التفكير التبادي يبحث عن حلول كثيرة صحيحة أي بدائل Alternative. ويعتبر Guilford أشهر باحث في هذا المجال ، فقد طور عدداً من الاختبارات تستخدم على نطاق واسع في التفكير التباعدي ، وفي هذه الاختبارات يطلب من الفرد أن يدرج قائمة بالاستخدامات العديدة للطوب العادي بحسب تفكيره فإذا ذكر أن استخدامات الطوب هي: بناء مدرسة ، بناء بيت ، بناء مكتبة ، وهو لا يتسم هنا بتفكير ابتكاري حيث استخداماته واحدة وهي بالتحديد استخدام الطوب في بناء شيء ما ، ولو كان لديه تفكير ابتكاري لأدرج استخدامات مختلفة للطوب مثل كسر نافذة ، اغراق فأر في الماء ، ضرب لص على رأسه.

ومن الباحثين ذوي الشأن في التفكير الابتكاري E.P Torrance وتتضمن اختباراته رسوماً أو صوراً ، وبنوداً مثل البنود اللفظية. وتقدر الدرجات في ضوء عوامل أربعة هي: الطلاقة Fluency ، المرونة Flexibility ، الأصالة Originality ، الحبك أو الربط بين الأجزاء والتفاصيل Elaboration.

اختبارات الميول Interest Inventories

حيث يطلب من الفرد أن يبين ما يحبه وما يكرهه مثل الأنشطة التي يفضل أن يشترك فيها. وعموماً تتم مقارنة الاستجابات على أنماط ميول معروفة.

ومقياس الميول الأكثر استخداماً هو قائمة الميول المهنية Vocational Interest Inventory. وفي مثل هذه الاختبارات يطلب من الفرد عادة أن يبين ما يفضله بالنسبة لنشطة وقت الفراغ كالهوايات. وتتم بعد ذلك مقارنة نمط المستجيب مع أنماط ميول أشخاص ناجحين في مجالات مهنية عديدة. وذلك بسؤالهم فيما يتصل بالحالات التي يكون فيها أكثر سعادة وأكثر نجاحاً.

وقائمتا الميول الأكثر استخداماً هما: قائمة سترونج – كامبل Strong – Campbell interest Inventory التي تقيس الميول في مجال الأعمال والمهن. وسجل كودر للتفضيل المهني Kuder Preference Record Vocational وهو يقيس الميول في مجالات مهنية عريضة كالمجالات الميكانيكية ، العلمية ، الاقناعية Persuasive ، والعلوم الاجتماعية.

الاختبارات الإسقاطية Projective Tests

طورت هذه الاختبارات الإسقاطية في محاولة للتخلص من بعض المشكلات الرئيسية المتأصلة في استخدام مقاييس التقدير الذاتي كاتجاه بعض المستجيبين إلى إعطاء استجابات مقبولة اجتماعياً.

وهذه الاختبارات فيا لعادة لا تبدو واضحة للمفحوصين ، حيث يطلب من الفرد أن يستجيب إلى بنود غامضة. ولأن الهدف منها غير واضح تقل عدم الأمانة في الاستجابة شعورياً ، ويسقط المفحوص مشاعره الحقيقية. وأساساً يستخدم هذه الاختبارات الإسقاطية الأخصائي النفسي الإكلينيكي. ولكنها قليلة الاستخدام من قبل الباحثين التربويين ، ويرجع هذا جزئيا إلى أن صدق الاختبار مشكوك فيه ، وجزئياً أن تطبيق وتقدير وتفسير الاختبارات الإسقاطية يتطلب تدريباً خاصا. ولكن إذا كان ضرورياً أن تستخدم أداة إسقاطية في بحثك ، فيجب أن تتأكد أن من يطبقها هو شخص مؤهل لذلك.

والتقنية الإسقاطية الأكثر استخداماً هي طريقة التداعي Association وفي هذه التقنية يطلب من المستجيب أن يتفاعل مع مثير مثل : صورة A Picture ، بقعة حبر Inkblot ، كلمة Word ولعل تقنية اختبار تداعي الكلمات Word Association هي أشهر أنواع تقنيات التداعي حيث يطلب من المفحوص أن يذكر أول شيء يرد على باله.

وأشهر اختبارين للتداعي هما: اختبار بقع الحبر لرورشاخ Rorschach Inkblot واختبار تفهم الموضوع Thematic Apperception. وقد أجريت ألاف الدراسات التي استخدمت اختبار رورشاخ ، ويعرض هذا الاختبار على المفحوص سلسلة من بقع الحبر ، ويطلب منه أن يحكي ماذا يرى. ويقدم اختبار تفهم الموضوع سلسلة من الصور ويطلب منه أن يقص قصة عن كل صورة.

وحتى فترة قريبة كانت تطبق كل الاختبارات الإسقاطية منفردة ، ومع هذا ثمة جهود حديثة لتطوير مجموعة من اختبارات إسقاطية مثل تقنية بقع الحبر لـ Holtzmen التي تستهدف أساساً قياس نفس متغيرات اختبار بقع الحبر لرورشاخ. وما زالت الأدوات الإسقاطية الجماعية بما في ذلك Holtzmen في مراحلها الأولى إلا أنها تبشر بأن تصبح أدوات إسقاطية أكثر موضوعية.

obeikandl.com

الاستعداد Aptitude

يقيس اختبار الاستعداد ما لدى الفرد من مقدرة كافية ، وهو يستخدم للتنبؤ بأداء الفرد في موقف ما في المستقبل. واختبارات الاستعداد العام General Aptitude متنوعة كاختبارات الاستعداد الدراسي Scholastic Aptitude ، واختبارات الذكاء Intelligence ، واختبارات القدرة الفعلية العامة General Mental Ability.

ونظراً لأن اختبارات الذكاء لم تحظ بسمعة طيبة في السنوات الأخيرة ، فقد أصبح مصطلح الاستعداد الدراسي هو الأكثر شيوعاً. ومع هذا مقاصد هذه الاختبارات هي أساساً واحداً ، للتنبؤ بما سيكون عليه أداء الفرد في مستقبل محدود إذا ما توافر له التدريب والتعليم.

وتتوافر اختبارات في شكل اختبارات فردية في مجالات محددة كالجبر ، أو في شكل بطاريات تقيس الاستعداد في جوانب كثيرة ذات صلة ، وفي واقع الأمر كل اختبارات الاستعداد مقننة وتطبق كجزء من برنامج الاختبار المدرسي ونتائجها مفيدة للمدرسين والمرشدين والمديرين.

الاستعداد العام General Aptitude

تقع تحت هذه الفئة مجموعة من الاختبارات تمثل مجموعة التعريفات المختلفة للاستعداد العام (أو الاستعداد الدراسي أو الذكاء) وبينما الهدف الأساسي لكل هذه الاختبارات هو التنبؤ بالأداء الأكاديمي للفرد في المستقبل إلا أن هناك عدم اتفاق حول العوامل التي يجب أن يتم قياسها وتصلح كعناصر تنبؤ. ومصطلح الاستعداد ذو تعريفات متنوعة حيث يضم متغيرات مثل : التكفير التجريدي ، وحل المشكلات ، والطلاقة اللفظية. واختبارات الاستعداد العام تطلب من تطلب من الفرد أن يقوم بأداء مجموعة من المهام اللفظية وغير اللفظية التي تقيس قدرة الفرد على تطبيق المعرفة وحل المشكلات. وتسفر هذه الاختبارات بصفة عامة عن ثلاث درجات : درجة اجمالية ودرجة لفظية ودرجة كمية. وهي تمثل اجمالي نسبة الذكاء IQ ، ونسبة الذكاء اللفظي ونسبة الذكاء الكمي.

وبينما تستهدف اختبارات الاستعداد العام قياس القدرة الفطرية أو الإمكانية الكافية ، فهي كما يبدو تقيس القدرة الراهنة. وتوجد بعض الشواهد التي تشير إلى تأثر الدرجات إلى حد ما ببيئة الفرد الماضية والحالية ، ولكن بما لها من وظيفة معقولة ومنطقية للتنبؤ بنجاح الفرد الأكاديمي في المستقبل فهي بهذا تقيس إمكاناته وهكذا تكون مفيدة لمن يعمل في المجال التربوي.

openkandi.com

واختبارات الاستعداد العام ذات تطبيق جماعي أو فردي ، ولكل نوع مزايا وعيوبه والاختبارات الجماعية مريحة ومناسبة لمن يطيقها وتوفر وقته وتوفر تقديراً لإمكانية أكاديمية من شأن أن تكون مناسبة لمعظم الدراسات البحثية التربوية. كما تتوافر البطاريات التي تضم عدداً من الاختبارات المناسبة لمختلف الأعمار ومختلف الصفوف. وإذا كانت البطارية ذات تركيبة متماثلة فإنها تسمح بدراسة النمو العقلي مع الوقت ، واجراء مقارنات بين مختلف المستويات. ومن أمثلة البطاريات التي تطبق جماعياً اختبار كاليفورنيا للنضج العقلي California Test of Mental Maturity ذو الستة مستويات ويمكن تطبيقه على كل أطفال المدرسة وكل الأعمار فيها ، وطلاب الكليات ، والراشدين. وهو يتضمن 12 اختباراً فرعياً Subtest تمثل خمسة عوامل : الاستدلال المنطقي ، العلاقات المكانية ، الاستدلال الرقمي ، المفاهيم اللفظية ، الذاكرة. وتتضمن النتائج نسبة ذكاء لغوية ، ونسبة ذكاء غير لغوية ، ونسبة ذكاء اجمالية. وعلى الرغم أن بعض الاختبارات الفرعية ليست جيدة كما ينبغي ، إلا أن الاختبار اجمالاً يعتبر مقياساً للإمكانية جدير بالاحترام.

ومن الاختبارات الجماعية كثيرة التطبيق : اختبار القدرة العقلية لأوتس لينون – Otis Lennon Mental Ability ، وهو يقيس عوامل أربعة : الفهم اللفظي ، الاستدلال الشكلي ، الاستدلال الكمي. وهو يعتبر أيضا مقياساً للإمكانية جديراً بالاحترام.

والعيب الواضح في هذه الاختبارات الجماعية أنها تحتاج إلى كثير من القراءة المتأنية الواعية ، وهكذا فالطلاب ضعاف القدرة على القراءة يتضررون ويحصلون على درجات تعكس – كمثال – مستوى نسبة ذكاء أقل من مستواهم الحقيقي. ومن جهة أخرى الاختبارات الفردية تتطلب القليل من القراءة ومن مميزاتها أنها نظراً لتطبيقها على فرد واحمد فيجب أن يكون الفاحص (المختبر) Examiner واعياً بعوامل المرض والقلق وغيرها التي تؤثر تأثيراً عكسياً في قدرة الفرد على الاستجابة. والعيب الرئيسي للاختبارات الفردية ، ومن جهة نظر الباحث ، صعوبة تطبيقها ، صعوبة تقدير الدرجات ، لذا يحتاج تطبيقها وتصحيحها إلى أشخاص مدربين. ومع ذلك إذا ظهر سبب يشكك في صدق الاختبارات الجماعية بالنسبة لعينة خاصة كأطفال صغار السن جدا ، فنبغي استخدام الاختبار الفردي حتى لو تطلب تخفيض حجم العينة. ولعل أكثر الاختبارات الفردية تطبيقاً مقياس ستانفورد بنيه للذكاء Stanford – Beinet Intelligence Scale ذو المصداقية العالية ، ومقياس وكسلر Wechsler Scales الذي يقيس مستوى الذكاء من سن 4 سنوات إلى الراشدين.

وبينما يثمر مقياس ستانفور بنيه عن درجة نسبة ذكاء واحدة ، سفر أيضا مقياس وكسلر عن عدد من الدرجات الفرعية. ومقياس وكسلر لذكاء الراشدين هو اختبار استعداد خاص يتضمن اختبارات لفظية (مثل المعلومات العامة ، المفردات) واختبارات الأداء (مثل تكملة الصور ، تجميع الأشياء).

وكما سبقت الإشارة إليه ، ثمة دليل يدل على أن درجات نسبة الذكاء تتأثر ببيئة الفرد الماضية والحالية ، لذا فإن مصداقية اختبارات نسبة الذكاء بالنسبة لمجموعة أقلية معينة تكون مثار شك ، كما تتهم هذه الاختبارات بالتحيز الثقافي ، وقد حفزت هذه الانتقادات على تطوير نسبة ذكاء بلا تحيز ثقافي ، فأصبحت لا تحتاج إلى استخدام اللغة فالاختبارات غير اللفظية هي أقل تحيز ثقافي من الاختبارات اللفظية.

الاستعداد الخاص Specific Aptitude

وبحسب المدلول يستهدف إلى التنبؤ بما سوف يكون عليه أداء الفرد في مجال خاص بعد حصوله على قدر من التعليم والتدريب. وهذه الاختبارات متعددة : أكاديمية كالرياضيات ،وغير أكاديمية كالاستدلال الميكانيكي. ويستخدمها المدرسون والمرشدون والمديرون ولنفس الأسباب.

وتستخدم اختبارات الاستعداد الخاص كثيراً في الدراسات البحثية. وربما يعود شيوع استخدامها إلى معادلة الجماعات To Equate Groups عند مقارنتها على اختبار تحصيل ، بعد تلقيها معالجات مختلفة ، لذا قد يعزى فارق التحصيل النهائي إلى الفارق المبدئي وليس إلى فروق في المعادلة. وقد تستخدم درجات الاستعداد لمعادلة المجموعات أما باستخدام المعاينة الطبقية أو من خلال اجراء احصائي يسمى تحليل التغاير Analysis of Covariance. وبينما معظم اختبارات الاستعداد هي اختبارات معيارية ومدونة ، فإن بعضها اختبارات أداء ، والأخيرة مناسبة للطلاب الذين يؤدون اختبارات ولديهم صعوبة في اللغة الإنجليزية كالطلاب الأجانب مثلاً.

وتتوافر اختبارات الاستعداد في المجالات الخاصة كالجبر والموسيقى والقدرة الميكانيكية ، كما يتوافر عديد من بطاريات الاستعداد التي تقيس الاستعدادات تفي مجالات متعددة ذات الصلة من أجل تقييم الاستعدادات الأكاديمية والاستعدادات غير الأكاديمية ، كما تضم عدداً من الاختبارات الجزئية ، من ذلك بطارية الاستعداد الأكاديمي Academic Aptitude Battery التي تشتمل على القراءة والكتابة والرياضيات والعلوم وغيرها. ومن جهة أخرى ، اختبارات الاستعدادات الفارقة The Differential Aptitude Tests التي تشتمل على اختبارات عن

العلاقة المكانية ، الاستدلال الميكانيكي ، السرعة والدقة الكتابية وغيرها توهي مصممة للتنبؤ بالنجاح في مجالات وظيفة مختلفة.

الاستعداد Readiness

وهي اختبارات الاستعداد أو التنبؤ Prognostic ، وتصنف أحياناً كاختبار استعداد ، وأحيانا كاختبار تحصيل. وتبدو أكثر ملاءمة من حيث التقسيم ، ولكن حيث أن اختبار الاستعداد يطبق قبل التعليم أو التدريب في مجال خاص لتحديد استعداد الطالب على الاستفادة من التعلم ، لذا تنال هذه الاختبارات الاهتمام المتزايد ، وهي تشمل عادة على قياسات المتغيرات كالتمييز السمعي (فهم معاني الكلمات) Auditory Discrimination والتمييز البصري Visual Discrimination ، والقدرة الحركية Motor Ability. وقد طورت بطاريات الاستعداد لتقييم الاستعداد في عدد من المجالات منها اختبار متروبوليتان للاستعداد The Metropolitan Readiness Test ، لقياس استعداد الطلاب الجدد على تعلم مادة دراسية في الصف الأول الابتدائي في مجالات مثل القراءة ، الحساب ، الكتابة .. هذا المقياس من أفضل المقاييس في مجاله.

اختيار اختبار ما Selection of A Test

توجد اليوم مصادر متعددة لاختيار الاختبارات ، فهناك شبكة الانترنت ، والأدبيات البحثية المنشورة ، والمقاييس النفسية المنشورة منها والمطبق ، والاختبارات المطبوعة ، والملخصات السيكولوجية أو المجلات المتخصصة ومنها كمثال:

Journal of Applied measurement, Journal of Consulting Psychology, Journal of Educational Measurement, Journal of Education Psychology, Journal of Personal Psychology, and Education and Psychological Measurement.

وكمثال يضم The Journal of Education Measurement عرضاً لأحداث الاختبارات المنشورة. ويجب على الباحث أن لا يختار أول ما يقابله بل أن يتعرف أولاً على الاختبارات العديدة التي تناسب دراسته ويقارن بينهما على العوامل ذات الصلة ويختار الأحسن من بينها.

إدارة الاختبار Test Administration

توجد مبادئ توجيهية عامة بشأن تطبيق الاختبار يجب أن يلم بها الباحث:

أولاً : إذا كان سيطبق الاختبار في بيئة مدرسية ، فعلية أولاً أن يعد الترتيبات مع الشخص المناسب وهو مدير المدرسة عادة وأن يتفق معه على زمان ومكان وظروف تطبيق الاختبار ، ومن يستطيع مساعدته من بين أفراد المدرسة. وعادة يفيده المدير في تزويده بالمعلومات والأوقات المناسبة كأيام الاجتماعات المدرسية ، أو يوم سابق لعطلة رسمية.

ثانياً : يجب على الباحث سواء أكان يطبق بحثه في مدرسة أو غيرها أن يتأكد من ظروف اختبار مثالية ومناسبة وبيئة مريحة وتعاون المفحوصين ، وإذا كان الاختبار سيتم على أكثر من جلسة ، يجب التأكد من تطابق الجلسات بقدر الإمكان.

ثالثا: أن يكون الباحث جاهزاً تماماً ، أي على علم تام بكل إجراءات التطبيق في دليل الاختبار ، وأن يتبع التعليمات بدقة ، وأن يتدرب عليها وأن يطبقها مسبقاً على بعض الأصدقاء إن أمكن حتى يجيدها.

وكما هي سنة الحياة فإن سبيل النجاح هو التخطيط الجيد والاعداد السليم ، فإذا أعدت كل الترتيبات اللازمة تأكد من التعاون الكامل ، كذلك تحتاط لكل الاحتمالات السيئة وكيفية مواجهتها فقد تدفعك إلى المحاولة في يوم آخر أو مع جماعة أخرى.

تم بحمد الله

obeikandl.com