جامعة الجيلالي بونعامة خميس مليانة

كلية العلوم الانسانية والاجتماعية

قسم علم النفس وعلوم التربية

سلسلة دروس موجهة لطلبة السنة الثانية ماستر علوم التربية: إرشاد وتوجيه

المجموعة الثانية

في مقياس

بناء وتكييف الاختبارات والروائز النفسية التربوية

إعداد الدكتورة: لحول فايزة

السنة الجامعية 2022-2023

المحاضرة 03: شروط وإجراءات بناء اختبار نفسي

يرى الباحثون أن هناك اتجاهين في بناء الاختبارات النفسية:

الاتجاه الأول نظري: ويهدف إلى تصميم اختبارات وفق أسس نظرية محددة لتجيب على فروض صاغها الباحث لاختبار ما تذهب إليه النظرية، فالباحث يصوغ بنود اختباره بعناية فائقة لتخدم فروضه.

الاتجاه الثاني عملي: ويهدف إلى مجرد تطوير اختبارات جديدة لتكون أداة في يد الباحث بغض النظر عن اعتبارات نظرية. (سوسن مجيد، 2014، ص69)

إذن، على هذا الأساس يختلف الاتجاه النظري عن العملي في أن الأول تنصب اهتماماته على صياغة فروضه النظرية وهي من أهم خطوات العمل كله وأصعبها، وتتطلب منه الأصالة إذ لا بد أن تكون فروضه متجهة إلى تقديم إضافة حقيقية تتسق مع الحقائق القائمة، ويصمم أداة لتحقيق ذلك مثل اختبار نسبة الذكاء الانفعالي لبار–أون، اختبار ماير وسالوفي للذكاء الوجداني كقدرة، اختبار ثورستون للقدرات العقلية الأولية، قائمة كوستا وماكري لعوامل الشخصية الخمسة الكبرى التي تعتبر اختبارات قامت على أسس نظرية.

أما الاتجاه الثاني فيهدف إلى تحقيق غرض محدد وإيجاد أداة قياس عملية تستخدم من أجل حل مشكلة قائمة بغض النظر عن الفروض مثل اختبار بينيه للذكاء، بطارية مينيسوتا، اختبار ألفا وبيتا للجيش التي تعتبر أدوات عملية قامت لأهداف عملية تفتقر إلى الأصالة في التنظير.

تصميم الاختبار في الاتجاه النظري يتطلب وقتا طويلا للإعداد والتنظيم وتحليل النتائج يبدأ من فروض قوية وواضحة وينتهي بمعارف مهمة تسهم في تطوير النظرية، أما تصميم الاختبار في الاتجاه العملي يعتبر سريع النتائج لا يتطلب وقتا طويلا في الإعداد، يؤدي إلى نتائج عملية مفيدة لكنها ذات قيمة نظرية ضئيلة.

إجراءات بناء اختبار نفسي:

تظهر الحاجة إلى تصميم اختبار جديد نتيجة لتغير المفاهيم في مجال علم النفس وظهور مفاهيم جديدة، كما يمكن أن يجد الباحث أن الاختبارات المتوفرة لا تفي بغرضه مثل خصائص العينة التي قد تختلف عن خصائص العينات التي بنيت الاختبارات الأخرى من أجلها (السن، المهنة، ومثلا قياس القلق عند المصابين بأمراض معينة أو الاحتراق النفسي لدى مستشاري التوجيه المدرسي والمهني بالجزائر، الذكاء الوجداني لدى المعلمين...)

كما أن الباحث قد يتبنى نظرية معينة في دراسة الموضوع، وبالتالي يجب أن يكون اختباره مصمما وفق تلك النظرية من حيث الأبعاد التي تتجلى في بنود الاختبار، مثلا أن يدرس الباحث موضوع التفاؤل من وجهة نظر التعلم الاجتماعي ويجد اختبارات تدرسه من وجهات نظر معرفية وسلوكية، كما أن البيئة الاجتماعية والثقافية لها دور مهم في بنية الاختبار.

سنتطرق فيما يلي إلى أهم الخطوات التي يجب أن يتبعها الباحث لبناء اختبار نفسي:

1- تحديد القدرة أو السمة المراد قياسها والهدف من الاختبار

وهذا من خلال تحديد المجال الذي يهدف الاختبار إلى قياسه، وكذلك الهدف من الاختبار، هل ميدان القياس هو المجال العقلي المعرفي؟ أو المجال الوجداني؟

وهل الهدف من الاختبار هو التوجيه المهني أو التربوي أو التشخيص أو التنبؤ بالسلوك أو تحديد مستويات الأفراد في خاصية معينة، أو التمييز بين الأفراد في سمة معينة وترتيبهم، أو تحديد الصعوبات والاضطرابات التي يعاني منها الفرد؟ وبالتالي تصبح مهمة الاختبار تحليل هذه الصعوبات بشكل تفصيلي وتصنيفي لكي يتم التعرف عليها بشكل محدد، ويختلف كل هدف من هذه الأهداف في طبيعته، فالأساليب التي تهدف إلى تحديد مستوى الأفراد يكثر استخدامها في المجالات التحصيلية أو في قياس القدرات المعرفية، مما يتطلب تصنيف وتحليل العمليات المختلفة في المادة الدراسية المحصلة وحيث يمكن استخدام هذا التصنيف كدليل وصفي لناتج العملية التعليمية ومدى تحقق أهدافها. (سوسن مجيد، 2014، ص70)

2- تحديد خصائص الأفراد المستهدفين بالقياس:

لأن هذه الخصائص تترتب عنها اجراءات خاصة تتعلق بالوقت، شكل الاختبار وطبيعة بنوده، وطريقة تقديمه، منها ما يتعلق بالعمر (أطفال، راشدين، مراهقين...) ومنها ما يتعلق بالمستوى التعليمي (تعليم عالي، تعليم بسيط، أميون) ومنها ما يتعلق بالإعاقة (تخلف عقلي بأنواعه ودرجاته، صمم، مكفوفون...)

3- تعريف القدرة أو السمة تعريفا إجرائيا

ونقصد بالتعريف الاجرائي التعريف العملي أو الوظيفي الذي يمكن أن يستدل منه على العمليات السلوكية التي تتضمنها القدرة أو السمة، والذي يدل كذلك على وظيفتها. (سعد عبد الرحمن، 2008، ص214)

فالمفاهيم العلمية كالذكاء والانبساط والاجتماعية عبارة عن تجريدات لخصائص مفترضة في الأشياء، ويستدل الباحث على هذه الخصائص من وقائع سلوكية محددة سواء في شكل أفكار معبر عنها أو حلول لمشكلات أو استجابة لمنبهات محددة أو أساليب تعامل مع البيئة أو صفات شخصية تتسم بالاستقرار النسبي، فيقوم الباحث بترجمة هذه المفاهيم وتحليلها إلى وقائع سلوكية أو خصائص محددة بصورة تسمح بصياغتها في وحدات معيارية للقياس. (سوسن مجيد، 2014، ص70)

فمفهوم الطلاقة مثلا يمكن ترجمته وتحليله إلى وقائع سلوكية مناسبة إذا وضع له التعريف الاجرائي التالي: الطلاقة هي انتاج أكبر عدد من الكلمات أو الأفكار ذات الدلالة استجابة لمنبه معين، فيتيح التعريف وضع منبهات لاستدعاء استجابات معينة هي التي تكون قابلة للقياس.

وعندما نعرف القدرة اللغوية إجرائيا بأنها القدرة على التعبير شفاهة أو كتابة عن المفاهيم والمدركات باستخدام التراكيب اللفظية الصحيحة والمناسبة، فإن هذا التعريف الاجرائي يساعد الباحث على معرفة العمليات السلوكية التي تشملها القدرة على التعبير عن الفكرة أو المفهوم مثل الوصف، استخدام التركيب اللغوي السليم والمفردات في مكانها المناسب.

فالتعريف الاجرائي إذن هو نوع من التحديد العملي أو الوظيفي للسمة أو القدرة من خلال تحليلها إلى وقائع سلوكية قابلة للقياس تتجلى في بنود الاختبار.

السؤال المطروح الآن: *ما هي المصادر التي يستعين بها الباحث لتحليل الخاصية (سمة/ قدرة) إلى وقائع سلوكية؟*

مصادر تحليل الخاصية إلى وقائع سلوكية: يمكن للباحث أن يستعين بالمصادر التالية:

***الاختبارات السابقة:** يمكن أن يطلع الباحث على بنود الاختبارات السابقة

***الكتابات المتخصصة:** سواء كانت نظريات أو بحوث نظرية أو ميدانية حول الخاصية المراد تصميم الاختبار من أجلها.

***تشخيصات الحالات المرضية:** عندما يريد الباحث أن يصمم اختبارا لقياس خاصية لا سوية، بإمكانه أن يطلع على نتائج تشخيص الأطباء والأخصائيين النفسانيين أو الأرطفونيين.

***تحليل العمل أو المهنة:** عندما يريد الباحث أن يصمم أداة لقياس الأداء أو الاستعداد لمهنة معينة، بحيث يجب أن يحدد المهارات أو المهام التي تتطلبها تلك المهنة مثلا: مهارات التدريس الجامعي التي تشمل مهارات تتعلق بالتخطيط للدرس(التنظيم، الترتيب، وضع خطة عمل، تقسيم الدروس حسب جدول

زمني...)، مهارات تتعلق بتنفيذ الدرس (استخدام استراتيجيات التدريس الحديثة، التنويع في الوسائل التعليمية/التعلمية...)، مهارات تتعلق بتقويم الطالب (التنويع في أساليب تقويم الطالب، تقديم التغذية الراجعة له...)، مهارات التواصل الفعال مع الطلبة (الاحترام، المرح والدعابة...)

***تحليل المقرر الدراسي:** لإعداد اختبار تحصيلي يجب الاطلاع على الوثائق الضرورية كالكتب المدرسية، المطبوعات، المذكرات...

***المقابلات الشخصية مع الأفراد المعنيين:** مثلا عندما يريد باحث بناء اختبار حول اتجاه طالب علم النفس نحو تخصصه، يوجه أسئلة لطلبة علم النفس لوصف مشاعرهم، وتفكيرهم.

***البنود المفتوحة المكتوبة:** يوجه الباحث بندا كتابيا مفتوحا حول خاصية معينة إلى عدد من الأفراد المعنيين، مثلا أن يوجه الباحث البند التالي إلى عدد من مدرسي مادة الرياضيات في مرحلة التعليم المتوسط: تعترض المتعلمين صعوبات مختلفة أثناء تعلمهم لمادة الرياضيات، المطلوب منكم وضع قائمة بهذه الصعوبات ليتسنى التعرف عليها وتحديدها للتغلب عليها، ويمكن ذكر صعوبة أو صعوبتين كنموذج.

4- تحديد مادة الاختبار وزمنه مع شكل وعدد بنوده

يقوم الباحث بتحديد مادة الاختبار أو محتواه وشكله من حيث نوع بنود الاختبار: على شكل ألفاظ مكتوبة أو شفوية (أسئلة/ عبارات) أو أعداد أو على شكل رسوم كاملة أو ناقصة فيكملها المفحوص، أو على شكل رموز، متاهات أو تأتي مادته على شكل أداء يقوم به المفحوص، حسب ما يقتضيه الهدف العام للاختبار، مع مراعاة صلاحية الاختبار لأفراد العينة من حيث أعمارهم، مستوى ذكائهم، مستوى تعليمهم، وقدرتهم على القراءة والكتابة.

وفي هذه الخطوة يضع الباحث تعليمات الاختبار ويحرص على أن تكون محددة بدقة وواضحة.

ويضيف (امطاينوس ميخائيل، 2016، ص59) أن مصمم الاختبار لابد أن يحدد مسبقا وقبل أن يبدأ بإعداد البنود **الزمن** الذي سيعطى للمفحوصين للإجابة عن بنود الاختبار و**عدد هذه البنود.**

وما من شك في أن زمن الاختبار وطوله يتحددان بالأغراض الخاصة له، فإذا كان الاختبار يهدف إلى المسح والمقارنة السريعة وسيطبق على عدد كبير من الأفراد تضاءلت الحاجة إلى زيادة عدد بنوده، أما إذا كان الغرض من الاختبار هو تصنيف الأفراد استنادا إلى مستويات القدرة أو السمة لديهم واتخاذ قرارات مهمة حول مستقبلهم المدرسي أو المهني تزايدت الحاجة إلى زيادة عدد البنود ورفع القدرة التمييزية لهذه البنود، وفي الحالات التي يكون الغرض من الاختبار فيها تشخيصيا تصبح الحاجة ملحة لزيادة عدد

البنود إلى الحدود القصوى المتاحة دون إعطاء أهمية كبيرة للقدرة التمييزية للبنود لأن الاختبار يهدف إلى تعرف الصعوبات ومواطن الضعف والقوة في أداء المفحوص.

كما يجب أن يأخذ الباحث بعين الاعتبار عند تحديد طول الاختبار والزمن المخصص للإجابة الأشكال التي ستأخذها البنود وخصائص المفحوصين.

5- إعداد بنود الاختبار وتحديد الأبعاد الفرعية له

يقوم الباحث باقتراح مجموعة من البنود (الوحدات/المفردات) تغطي جميع جوانب الخاصية المقاسة، حيث تعتبر عينة ممثلة لمجال السلوك المراد قياسه.

يرى (سعد عبد الرحمن، 2008) أن على الباحث أن يقترح عددا كبيرا من البنود أكبر بكثير مما يتوقع أن يحتويه الاختبار في صيغته النهائية، حيث أنه سيتم بعد ذلك الاستغناء عن عدد يتراوح بين 30% إلى 40% من عدد البنود المقترحة.

مثلا قائمة كوستا وماكري لعوامل الشخصية الخمسة الكبرى اشتملت في صورتها الأولية على 160 بندا ليقلص عدد البنود إلى 60 بند في صورتها النهائية.

ومن المهم أن يقوم الباحث **بتقسيم الخاصية (السمة/ القدرة) إلى أبعاد فرعية** تندرج تحت كل بعد مجموعة من البنود.

6- تحليل بنود الاختبار (كميا وكيفيا)

المحاضرة 04: تحليل بنود الاختبار

تأتي خطوة تحليل بنود الاختبار (كيفيا وكميا) بعد عملية اقتراح البنود، وبعد تجميع الاختبار في صورته الأولية، وبعد إعداد التعليمات والامثلة المحلولة لمساعدة المفحوص، تتم عملية تحليل البنود كما يلي:

1- التحليل الكيفي:

وهو يتضمن قواعد كتابة البنود نذكر أهمها:

- يجب أن تكون بنود الاختبار واضحة (غير مربكة)، ولا تحتمل أكثر من معنى أو تفسير واحد.
- تكتب بلغة بسيطة، سهلة القراءة والفهم ولا تتخللها تعبيرات بلاغية أو مصطلحات فنية (تقنية) لا يتحقق فهمها إلا لفئة قليلة، إلا إذا كانت موجهة لهم خصيصا، مثلا مصطلحات خاصة ببداغوجية المقاربة بالكفاءات.
- يجب أن تكون البنود عينة ممثلة للسلوك المراد قياسه وليس بالجوانب الهامشية.
- لا تكون الجمل طويلة ولا شديدة الايجاز والاختصار، وإنما تعبر عن الفكرة بشكل كاف.
- تجنب كتابة بنود تتضمن أكثر من فكرة مثل: إذا صرخ أحد بوجهك، هل ترد عليه؟ أو تنسحب؟
- تجنب استخدام الإطلاق في متن البنود مثل: دائما، أحيانا..(بشير معمرية، 2012)

2- التحليل الكمي:

يهتم القائمون ببناء الاختبارات بكتابة وانتقاء مفردات ذات جودة لقياس السمات قياسا دقيقا، لذلك يراعون كثيرا من الشروط في تكوين هذه المفردات وصياغتها والتحقق بالأساليب المنطقية وأحكام الخبراء من صدق محتوى كل مفردة على حدا، غير أنه مهما بلغت دقة هذه الأساليب والأحكام فإنها لا تغني عن التجريب الميداني للاختبار وتحليل مفرداته باستخدام الأساليب الاحصائية، وهو ما يعرف بالتحليل الكمي لفقرات الاختبار، حيث يفيد ذلك في المراجعة النهائية لمفردات الاختبار وتحسينها، بحيث تسهم كل فقرة إسهاما واضحا فيما يقيسه الاختبار.

تذكر (سوسن مجيد، 2014، ص77) أن هناك اعتبارين أوليين يبرزان عند اختيار فقرات اختبار ما.

الأول هو: هل الفقرة صادقة في قياس السمة التي نرغب في قياسها، بمعنى هل تمكننا هذه الفقرة من التمييز بين ذوي المستويات العليا والأفراد ذوي المستويات الدنيا من السمة أو القدرة محل القياس؟

ويجاب عادة على هذا السؤال من خلال الأساليب الاحصائية الخاصة بحساب **صدق فقرات الاختبار** ويمكن نقول عنها **معاملات التمييز للفقرات.**

أما الاعتبار الثاني فهو: هل مستوى صعوبة الفقرة مناسبة لمجموعة الأفراد الذين سيختبرون بهذا الاختبار؟ ويجاب على هذا السؤال من خلال حساب **معاملات الصعوبة/ السهولة.**

2-1- صدق الفقرة (تمييز الفقرة):

ويعني مدى إمكانية قياس الفروق الفردية بواسطة بنود هذه الاختبارات، وهي خاصية ينبغي أن تتوفر في مفردات الاختبارات مرجعية الجماعة أو المعيار.

ويوجد نوعان من المحكات التي يستند إليها الباحث في تحديد الأفراد ذوي الدرجات العليا في تلك السمة أو القدرة والأفراد ذوي الدرجات الدنيا (محكات داخلية ومحكات خارجية)

2-1-1- حساب معامل تمييز الفقرة (البند/ المفردة) استنادا إلى محكات داخلية: أكثر الأساليب استخداما في حساب معامل تمييز الفقرة بالاستناد إلى محكات داخلية نجد أسلوبا يعتمد على معامل الارتباط، وأسلوبا آخر يعتمد على المقارنة بين مجموعتين طرفيتين نظرا لسهولة تطبيقهما وتفسير نتائجهما.

***أسلوب يعتمد على معامل الارتباط:** من خلال حساب معامل الارتباط بين درجات الأفراد على كل مفردة ودرجاتهم الكلية على الاختبار ككل، بالنسبة لذوي الدرجات العليا وذوي الدرجات الدنيا، يعني تقسيم الدرجات الكلية للأفراد على الاختبار إلى نصفين، نصف خاص بذوي الدرجات الدنيا وأخر خاص بذوي الدرجات العليا، ويسمى معامل الارتباط الناتج بمعامل الارتباط الثنائي المتسلسل الحقيقي Point Biserial Correlation

باختصار، معامل الارتباط الثنائي المتسلسل الحقيقي Point Biserial Correlation هو نوع خاص من معاملات الارتباط مشتق من صيغة بيرسون يمكننا من ايجاد العلاقة بين درجات الأفراد على كل مفردة من الاختبار ودرجاتهم الكلية على الاختبار وذلك بتقسيم توزيع الدرجات الكلية إلى مجموعتين إحداهما تمثل المجموعة العليا والأخرى تمثل المجموعة الدنيا في السمة التي يقيسها الاختبار، القيمة الناتجة تسمى معامل تمييز الفقرة. (صلاح الدين علام، 2000)

ويمكن أن نجد مصطلح معامل الارتباط المنصف للسلسلة الحقيقي حيث تستخدم هذه الطريقة حسب (سوسن مجيد، 2014، ص79) عند تقسيم المتغير المتصل (الدرجات الكلية للاختبار) إلى قسمين متساويين يشتمل أحدهما على الدرجات العليا والآخر على الدرجات الدنيا، حيث ينتج لدينا

متغيران أحدهما ثنائي وهو درجة كل فقرة، والآخر فئوي متصل وهو الدرجات الكلية ولكن تم تقسيمه إلى تغير ثنائي.

***أسلوب يعتمد على المقارنة بين المجموعتين الطرفيتين:**

تعتمد هذه الطريقة على تقسيم الدرجات الكلية للأفراد في الاختبار إلى قسمين، يمثل أحد القسمين المجموعة التي نالت أعلى الدرجات في الاختبار، ويمثل القسم الآخر المجموعة التي نالت أقل الدرجات في الاختبار نفسه، ومقارنة متوسطي درجات الأفراد على كل فقرة باستخدام اختبار "ت" لعينتين مستقلتين عندما تتوفر شروط تطبيقه في بيانات العينة، وإذا لم تتوفر فيمكن استخدام اختبار بديل هو اختبار مان ويتني اللامعلمي (اللابرامتري).

وقد أوصى كيلي Kelley حسب ما يشير إليه (صلاح الدين علام، 2000، ص284) عند تحليل مفردات الاختبار الاعتماد على النسبة 27% من الأفراد في كل من المجموعتين الطرفيتين، واستبعاد نسبة 46% الوسطى، وتتميز هذه الطريقة بأنها لا تتطلب جهدا كبيرا، وتعطي قيما قريبة من القيم التي نحصل عليها باستخدام معامل الارتباط الثنائي المتسلسل الحقيقي.

2-1-2- حساب معامل تمييز الفقرة (البند/ المفردة) استنادا إلى محكات خارجية:

يذكر (صلاح الدين علام، 2000) أنه يندر استخدام محكات خارجية في تقسيم مدى تمييز مفردات الاختبارات، والمحك الخارجي قد يكون اختبارا آخرا مستقلا عن الاختبار لكن بنفس الهدف، ويمكن الاسترشاد بنفس الطرق السابقة في ايجاد قيم معامل التمييز.

كلما ارتفع معامل تمييز الفقرة كان اسهامها أفضل في رفع قيمة تباينها، ويحدد ايبل (Eipel, 1963) مجموعة من المعايير للحكم على تمييز الفقرة

الجدول رقم (01): معايير إيبل للحكم على تمييز الفقرة

مستوى التمييز	تمييز الفقرة
من 0,40 فأعلى	فقرة جيدة جدا
من 0,30-0,39	فقرة جيدة بدرجة معقولة ولكن يمكن تحسينها
0,19-0,29	فقرة هامشية تحتاج إلى تحسين
أقل من 0,19	فقرة ضعيفة تحذف أو تعدل كليا

المصدر: (محمد الخطيب وأحمد الخطيب، 2010، ص49)

المحاضرة 05: التحليل الكمي لبنود الاختبار (معاملات السهولة/ الصعوبة)

2-2- صعوبة/ سهولة الفقرة:

تعد صعوبة مفردات الاختبار من الخصائص المهمة في الاختبارات مرجعية الجماعة أو المعيار، وتؤثر في إجابات الأفراد عن مفرداتها، فالمفردات التي تشتمل عليها هذه الاختبارات ينبغي أن تميز تمييزا دقيقا بين مستويات السمة المراد قياسها، فالمفردة التي يجيب عنها جميع الأفراد، أو التي لا يستطيع أحدهم الإجابة عنها لا تفيد في كشف الفروق بينهم فيما يقيسه الاختبار. (صلاح الدين علام، 2000، ص268)

ويقتصر بطبيعة الحال تحليل مستوى الصعوبة على اختبارات القدرات والاستعدادات والتحصيل، يعني الاختبارات التي تحتمل بنودها إجابات صحيحة وأخرى خاطئة، أو إمكانية حل مشكلة معينة من عدمها.

وترى (سوسن مجيد، 2006، ص82) أن تحليل مستوى صعوبة الفقرات أكبر أهمية بالنسبة لاختبارات الاستعدادات والقدرات، أما اختبارات التحصيل فيمكن أن يكون تقدير خبير في الموضوع أكثر أهمية من حساب مستوى الصعوبة احصائيا، كما أن اختبارات السرعة بكل أشكالها تخرج عن إطار مشكلة الصعوبة، إذ تتضمن أساسا فقرات سهلة، أما في اختبار القوة فيفضل تدرج فقرات الصعوبة من الأسهل إلى الأصعب، وبالتالي احتواء الاختبار على مدى واسع من مستويات الصعوبة.

2-2-1- أهمية استخراج معاملات السهولة والصعوبة:

يلخص (امطاينوس ميخائيل، 2016) فوائد استخراج معاملات السهولة والصعوبة في النقاط التالية:

***ترتيب بنود الاختبار:** استنادا إلى تلك المعاملات يمكن ترتيب بنود الاختبار بحيث تبدأ بالأسهل وتنتهي بالأصعب، ولهذا الترتيب أهميته من حيث أنه يتيح استثارة دافعية المفحوص وينمي ثقته بنفسه بإجابته أولا عن البنود السهلة، ويضعف احتمال تعرضه للصدمة منذ بداية عمله في الاختبار، كما يمنع احتمال إضاعة الوقت والجهد في بنود يعجز المفحوص عن الإجابة عنها أصلا لأنها تتجاوز سقف قدرته.

***انتقاء بنود الاختبار:** لضمها إلى الصورة النهائية للاختبار، وفي هذا الصدد يشير علماء القياس إلى أن المستوى الأكثر ملاءمة لصعوبة البنود هو 50%، لأنها تنتج اكبر قدر من التباين أو التمايز بين الأفراد وتكون حساسة للفروق الدقيقة بينهم، فتباين درجات الأفراد على الفقرة هو حاصل ضرب معامل السهولة في معامل الصعوبة، لذلك فالقيمة العددية للتباين تتناقص كلما ابتعدنا صعودا أو نزولا عن 0,5

***بناء الشكل المكافئ أو عدد من الاختبارات المكافئة:** للتحقق من ثبات الاختبار باستخدام طريقة الصور المتكافئة

***حساب ثبات الاختبار بطريقة كيودر ريتشاردسون الصيغة 20:** حيث يعتمد القانون على حاصل ضرب معاملات السهولة في معاملات الصعوبة لفقرات الاختبار (تباين الفقرات)

***حساب الانحراف المعياري لدرجات الأفراد على فقرات الاختبار:** باعتبار أن الانحراف المعياري هو الجذر التربيعي للتباين، كما أشرنا سابقا تباين درجات الأفراد على الفقرة= معامل السهولة X معامل الصعوبة لتلك الفقرة.

2-2-2- حساب صعوبة/ سهولة الفقرة:

يمكن الحصول على قيم تشير إلى سهولة فقرات الاختبار بإيجاد نسبة عدد الأفراد في جماعة مرجعية محددة الذي يجيبون عن كل مفردة إجابة صحيحة أو يستطيعون حل مشكلة معينة، وكلما زادت هذه النسبة دل ذلك على سهولة المفردة، وكلما قلت دل ذلك على صعوبة المفردة.

ويرتبط معامل الصعوبة بنسبة عدد الأفراد الذين أجابوا إجابة خاطئة على المفردة، أو لم يستطيعوا حل المشكلة المتضمنة في المفردة، وكلما زادت هذه النسبة دل ذلك على صعوبة المفردة والعكس صحيح.

لذلك يكفي أن يحسب الباحث معاملا واحدا ويفسر النتيجة المتحصل عليها.

القانون:

$$\text{معامل سهولة الفقرة} = \frac{\text{عدد الأفراد الذين أجابوا إجابة صحيحة عن المفردة}}{\text{العدد الكلي لأفراد الجماعة المرجعية}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\text{معامل صعوبة الفقرة} = \frac{\text{عدد الأفراد الذين أجابوا إجابة خاطئة عن المفردة}}{\text{العدد الكلي لأفراد الجماعة المرجعية}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

يتضح من خلال (1) و (2) أن: **معامل سهولة الفقرة = 1- معامل صعوبة الفقرة**

يذكر (صلاح الدين علام، 2000، ص269) أنه ينبغي التأكيد على أن خاصية صعوبة/ سهولة المفردة في إطار النظرية الكلاسيكية للاختبارات ليست مقصورة على المفردة، لأنها تعكس أيضا قدرة أفراد

الجماعة المرجعية الذين أجابوا عن تلك المفردة، لذلك يفضل عند تفسير قيمة معامل الصعوبة تحديد الجماعة المرجعية التي استندنا إليها في تحديد هذه القيمة.

مثال: طبق باحث اختبار ذكاء يشتمل على 20 مفردة على مجموعة تتكون من 100 طالب من طلبة السنة الثانية ماستر إرشاد وتوجيه جامعة خميس مليانة للموسم الجامعي 2019-2020، أجاب عن المفردة الأولى 70 طالبا إجابة صحيحة.

معامل سهولة المفردة الأولى= $\frac{70}{100}$ = 0,70 إذن ما نسبته 70% من أفراد العينة (طلبة السنة الثانية ماستر إرشاد وتوجيه جامعة خميس مليانة للموسم الجامعي 2019-2020) أجابوا إجابة صحيحة عن المفردة الأولى.

2-2-3- تصحيح قيمة معامل السهولة من أثر التخمين:

من المعلوم أن بعض الأفراد يتوصلون إلى الإجابة الصحيحة عن طريق **التخمين أي بمحض الصدفة**، لذلك يبرز التساؤل التالي: هل يمكن اعتبار الدرجة الكلية التي يحصل عليها فرد في الاختبار مساوية لعدد إجاباته الصحيحة عن مفردات الاختبار، أم أنه ينبغي تصحيحها من أثر التخمين العشوائي لاستبعاد ذلك الجزء من الدرجة الكلية الذي يفترض حصول الفرد عليه بالصدفة، أي دون أن يعرف بالفعل الإجابة الصحيحة عن بعض المفردات.

والحقيقة أنه لا يوجد اتفاق بين علماء القياس حول الإجابة عن هذا التساؤل، فالبعض يرى أنه من ليس من الضروري إجراء هذا التصحيح وحججهم في ذلك:

- أنه لا يتغير ترتيب الأفراد فيما بينهم إذا أجرينا تصحيحا لدرجاتهم الكلية في الاختبار من أثر التخمين.
- يؤدي تصحيح الدرجات من أثر التخمين للأفراد الذين لا يلجأون إلى التخمين في إجاباتهم إلى خفض درجاتهم الحقيقية، بينما لا يحدث ذلك لمن اعتاد على التخمين في إجاباته.
- تصحيح الدرجات من أثر التخمين قد يؤدي إلى الافتراض الخاطئ بأن الدرجات الناتجة هي الدرجات الحقيقية للأفراد في الاختبار. (صلاح الدين علام، 2000)

والمنطق يقول باختصار أنه: لا يمكن فصل الأفراد الذين استخدموا التخمين في إجاباتهم عن الأفراد الذين كانت إجاباتهم ناتجة عن معلوماتهم ومعارفهم، وتصحيح معامل الصعوبة من أثر التخمين يشمل جميع الأفراد.

لذلك يقترح أصحاب هذا التوجه بعض الارشادات التي يمكن أن تسهم في تقليل أثر التخمين منها:

- تحذير المختبرين من التخمين العشوائي في إجاباتهم عن بنود الاختبار.
- السماح بوقت كاف لكل فرد لكي يتمكن من محاولة الإجابة على بنود الاختبار.
- تضمين البدائل التي تشتمل عليها مفردات الاختيار من متعدد أخطاء شائعة تتعلق بالمجال الذي تقيسه المفردات، باعتبار أن الاختبارات من نوع الاختيار من متعدد هي التي يكثر فيها التخمين العشوائي.

أما البعض الآخر فيرى ضرورة تصحيح قيمة معامل الصعوبة من أثر التخمين العشوائي، وذلك للحصول على تقدير لنسبة الأفراد الذين يجيبون إجابة صحيحة عن كل مفردة استنادا إلى معلوماتهم الفعلية، واستبعاد من يجيبون إجابة صحيحة عن طريق التخمين العشوائي.

$$\text{معامل السهولة المصحح من أثر التخمين} = \frac{\text{ح} - \frac{\text{خ}}{\text{ك}-1}}{\text{ن} - \text{ن}َ}$$

ح: عدد الأفراد الذين أجابوا إجابة صحيحة عن المفردة.

خ: عدد الأفراد الذين أجابوا إجابة خاطئة عن المفردة.

ك: عدد بدائل الاجابة عن المفردة.

ن: العدد الكلي لأفراد المجموعة المختبرين.

نَ: عدد الأفراد الذين تركوا المفردة دون إجابة. (ويمكن للباحث حذف نَ من القانون إذا أراد ذلك)

مثال:

نفترض أن الاختبار الذي يشتمل على 20 بندا أو مفردة اختيار من متعدد، ولكل منها 5 بدائل طبق على 100 طالب من طلبة السنة الثانية ماستر إرشاد وتوجيه جامعة خميس مليانة للموسم الجامعي 2019-2020، وأجاب 75 فرد منهم على المفردة الأولى إجابة صحيحة بينما ترك 10 أفراد المفردة دون إجابة.

$$\text{معامل السهولة المصحح من أثر التخمين} = \frac{75 - \frac{15}{5-1}}{100 - 10} = 0,79$$

نلاحظ أن 79% من أفراد العينة (طلبة السنة الثانية ماستر إرشاد وتوجيه جامعة خميس مليانة للموسم الجامعي 2019-2020) أجابوا عن المفردة الأولى إجابة صحيحة استنادا إلى معلوماتهم الفعلية، في حين البقية (1-79%= 21 %) أجابوا إجابة خاطئة عن المفردة.

بينت الدراسات السيكومترية أو الامبريقية أن الاختبار يمكن أن يميز إلى أقصى حد ممكن بين الأفراد المختبرين إذا كان متوسط الصعوبة التي يشتمل عليها في حدود 0,50، أي يستطيع حوالي 50% من الأفراد الإجابة عن كل مفردة من مفرداته. (صلاح الدين علام، 2000)

لأنها حسب ما ذكرنا سابقا تنتج أكبر قدر من التباين وبالتالي زيادة معامل الثبات، والاختبار الجيد حسب (امطاينوس ميخائيل، 2016، ص74) هو ذلك الاختبار الذي يتضمن تشكيلة واسعة من البنود، من بينها البنود الصعبة التي تتحدى الأقوياء (حوالي 25% من البنود)، السهلة الموجهة للضعفاء (حوالي 25% من البنود) بالإضافة إلى الأسئلة المتوسطة (حوالي 50% من البنود)

7- إجراء تحليل عاملي استكشافي: بعد القيام بالتحليل الكمي والكيفي للبنود يتم الإبقاء على أفضل البنود كمؤشرات للسمة أو القدرة موضع القياس، يمكن للباحث هنا إجراء تحليل عاملي من نوع استكشافي بهدف تحديد مكونات المقياس أو أبعاده، وهذا بالاعتماد على قيم تشبع البنود بالعوامل المستخرجة.
سنتطرق إلى موضوع التحليل العاملي وأهميته في بناء الاختبارات النفسية في محاضرة مستقلة.

8- الخصائص السيكومترية للاختبار: حساب صدق الاختبار وثباته، واستخراج معايير لتفسير النتائج.

المحاضرة 06: الخصائص السيكومترية للاختبار (صدق الاختبار)

يتم حساب الخصائص السيكومترية للاختبار في إطار عملية التقنين، حيث يبين علماء القياس معنيين للتقنين هما:

أولا: أن تكون تعليمات الاختبار وصياغة بنوده والزمن المخصص له، وشروط وإجراءات تطبيقه موحدة في كل المواقف، ومن هذه الناحية يعني التقنين التوحيد، توحيد الشروط المحيطة بالمفحوصين وضبط العوامل والمتغيرات جميعها التي يمكن أن تؤثر في أدائهم الاختباري وإبعاد أثرها إلى أقصى حد، وذلك بهدف قياس المتغير الذي صمم الاختبار لقياسه بحجمه الحقيقي لدى تلك العينة من الأفراد التي سيطبق عليها.

ثانيا: أن يخضع الاختبار للتقنين من خلال تطبيقه على عينة ممثلة للمجتمع الأصلي، بهدف استخراج معايير معينة تحدد معنى الدرجة الخام التي يحصل عليها الفرد في ضوء تمركز أو تشتت درجات أفراد عينة التقنين، والمؤكد منه أنه لا يمكن أن يكون الاختبار مقننا بالمعنى الثاني دون أن يكون مقننا بالمعنى الأول. (امطاينوس ميخائيل، 2016)

أما (صلاح الدين علام، 2000) فيرى أن التقنين يقصد به أن يكون بناء وتصحيح وتفسير نتائج الاختبار أو أداة القياس مستندا إلى قواعد محددة، يحيث تتوحد فيه وتتحدد بدقة مواد الاختبار وطريقة تطبيقه، وتعليمات إجابته وطريقة تصحيحه أو تسجيل درجاته، وبذلك يصبح الموقف الاختباري موحدا بقدر الإمكان لجميع الأفراد في مختلف الظروف.

كما يشمل التقنين حساب الخصائص السيكومترية للاختبار، وهي الخصائص الضرورية المتعلقة بالثبات والصدق، مع استخراج معايير لتفسير النتائج المتحصل عليها من تطبيق الاختبار.

يقوم مصمم الاختبار بحساب الصدق:

يعتبر الصدق من الخصائص السيكومترية للاختبارات النفسية، وهو يمثل إحدى الوسائل المهمة في الحكم على صلاحية هذه الاختبارات، وفيما يلي سوف نتطرق إلى مفهوم الصدق وأنواعه والعوامل المؤثرة فيه.

1-مفهوم الصدق:

تعرف (سوسن مجيد، 2014) الصدق أنه: قياس الاختبار فعلا أو حقيقة للسمة أو الظاهرة التي وضع لأجلها.

بمعنى أن الاختبار الصادق يقيس ما وضع لقياسه فقط فاختبار الذكاء يقيس الذكاء فقط ولا يقيس شيئا آخر مثل سمات الشخصية أو التحصيل أو جوانب انفعالية. (صلاح مراد وأمين سليمان، 2005)

فالصدق هو أن يقيس الاختبار ما وضع لقياسه، ويحقق الهدف الذي وضع من أجله.

يتميز الصدق بالخصائص التالية:

- صفة نسبية: فلا يوجد اختبار عديم الصدق أو تام الصدق، هذا إذا استعملنا الاختبار فيما وضع له، فقد يكون اختبار الذكاء الخاص بتلاميذ السنة الثانية ابتدائي نتائجه مرتفعة الصدق، إذا استعمل مع تلاميذ السنة الثانية ابتدائي، ومتناقصة الصدق إذا استعمل مع عينة أخرى، وعليه فالصدق ليس أمرا مطلقا بل يختلف من اختبار لآخر بحيث لا نستطيع أن نقول أن الاختبار صادق أو غير صادق بل نقول أنه صادق بدرجة ما.
- صفة نوعية: حيث يرتبط الصدق مباشرة بالغرض أو الاستعمال الخاص الذي صمم الاختبار من أجله، فالاختبار الذي أعد بغرض التنبؤ بأداء المفحوصين في فترة مقبلة لا يصلح لغرض التشخيص والكشف عن نقاط الضعف والقوة في أداء المفحوصين، كما يرتبط الصدق مباشرة بالفئة أو الجماعة التي صمم لها، فالاختبار الذي صمم لفئة أو فئات معينة من المعاقين لا يصلح للأسوياء.

- صفة تتعلق بنتائج الاختبار وليس بالاختبار نفسه، ولكننا نربطها بالاختبار من قبيل الاختصار أو التسهيل، والأصوب أن نتكلم عن صدق النتائج، أو بالأحرى صدق تفسيراتنا للنتائج. (امطاينوس ميخائيل، 2016)

فالصدق إذن لا يتعلق بالاختبار نفسه بقدر تعلقه بتفسير الدرجات المستخرجة منه، وهذا يعني مدى فائدة أداة القياس في اتخاذ قرارات تتعلق بغرض أو أغراض معينة، فالصدق هنا هو صدق تفسير النتائج أو تأويلها وليس صدق الاختبار ولا صدق الدرجات.

2- أنواع الصدق:

يرى (بشير معمرية، 2012) أن الجمعية الأمريكية لعلم النفس (1954) صنفت الصدق إلى ثلاث أنواع أساسية هي: صدق المحتوى، الصدق المرتبط بمحك وأخيرا صدق التكوين الفرضي.

2-1-صدق المحتوى:

ويطلق عليه البعض الصدق المنطقي أو صدق المضمون أو الصدق بحكم التعريف، ويشير صدق المحتوى حسب (صلاح مراد وأمين سليمان، 2005) إلى مدى تمثيل بنود الاختبار أو المقياس لمحتوى السمة موضع القياس، ويرتبط صدق المحتوى بمفهوم الشمول حيث يمكن اعتبار هذا النوع من الصدق دليلا على شمول الأداة ودرجة تمثيلها للمحتوى.

بمعنى أن الاختبار يكون صادقا إذا كان المحتوى الذي نريد قياسه ممثلا تمثيلا صادقا من خلال بنود الاختبار، كما أن هذه الطريقة تدخل في المراحل الأولى لتكوين أي اختبار.

ويرى (بشير معمرية، 2012) أن هذه الطريقة تصلح خاصة للاختبارات التحصيلية وبرامج التدريب على الأداء والكفاءة المهنية، كما أنه من الصعب حساب صدق المحتوى للاختبارات النفسية، كسمات الشخصية والاتجاهات والقيم وغيرها.

ففي مجال قياس الشخصية قد يكون بند يقيس الانبساطية / الانطوائية مثلا: هل تفضل الذهاب إلى حفلة مع أصدقائك أو تبقى في المنزل لقراءة كتاب؟ فإذا أجاب قراءة كتاب فهو انطوائي ولكن ربما يجيب هكذا لميله للمطالعة، فهنا يظهر أن البند يقيس الميول، فبالتالي ففي اختبارات الشخصية لا يمكن إصدار حكم على أن بندا ما يقيس الخاصية التي وضع لقياسها.

يندرج ضمن صدق المحتوى:

-**الصدق الظاهري (السطحي):** والذي يتمثل في فحص محتوى الاختبار والتأكد من جودته فيما يقيسه دون فحص تجريبي.

-**صدق المحكمين:** نسبة إلى استخدام مجموعة من الخبراء والمختصين للحكم على جودة الاختبار ومدى تمثيل بنوده للمحتوى، فإذا كان الاتفاق بين آراء المحكمين مرتفعا دل ذلك على الصدق، كما يجب ألا يقل عدد المحكمين عن خمسة، ويجب أن يكونوا من المتخصصين في المجال مع الاستعانة بمتخصصين في القياس النفسي، ولا تقل درجة الاتفاق على كل بند من البنود عن 80%. حسب ما يشير إليه (صلاح مراد وأمين سليمان، 2005)

تتضح أهمية صدق التحكيم في قيمة ما قد يقترحه المحكمون بعد اطلاعهم على الأداة، فقد يزودون الباحث بملاحظات قيمة يكون قد أغفلها عند بناء أداة القياس، كاقتراح عبارات جديدة ذات نوعية جيدة، أو تعديل فيها أو إضافة أبعاد إضافية...لذلك يجب على الباحث أن يولي اهتماما كبيرا لاستمارة التحكيم او كما تعرف باستمارة الخبرة.

حيث ترى (نادية الزقاي، 2017) أنها أداة لجمع معطيات يكون مصدرها مجموعة الخبراء ذوي الاختصاص يجب أن تخضع لشروط بناء مضبوطة، وشروط تطبيق موضوعية، كما أنها لا تحتمل حين تفريغ أو عرض نتائجها أية عشوائية، لذلك يجب تزويد المحكمين بمعطيات وافية عن الاختبار، حيث يجب أن تشتمل الاستمارة على الهدف من الاختبار موضوع التحكيم، المطلوب أو مجموعة المطالب الموجهة للسادة المحكمين، التعريف الاجرائي للخاصية موضع القياس، تحديد الأبعاد الفرعية للاختبار والبنود التي تنتمي إلى كل بعد مع ضرورة تضمين تعريفات إجرائية لأبعاد الاختبار في الاستمارة.

دون أن يهمل الباحث الواجهة التي تذكر فيها المعلومات الأساسية مثل الجهة التي صدرت عنها استمارة التحكيم (الجامعة، الكلية، القسم...) عنوان الاستمارة، السنة، والتعليمات الموجهة للخبراء، وبياناتهم الشخصية (الاسم واللقب، الدرجة العلمية والتخصص)

2-2- الصدق المرتبط بمحك: يعتمد هذا المفهوم على درجة علاقة درجات الاختبار بالأداء الفعلي على محك خارجي، ويقصد بالمحك الخارجي اختبارا آخرا جيدا أو نوعا من الأداء العملي تستخدم فيه السمة موضع الاهتمام، وهناك نوعين من الصدق المرتبط بمحك هما كالتالي:

-**الصدق التلازمي:** يدل الصدق التلازمي (المصاحب) أو التزامني على حجم العلاقة بين درجات الأفراد على الاختبار ودرجاتهم على محك أخر بحيث لا يكون هناك فاصل زمني (أو فاصل زمني قصير) بين أداء الأفراد على الاختبار وأدائهم على المحك، مثلا العلاقة بين درجات الأفراد على اختبار ويكسلر وبينيه.

-**الصدق التنبؤي:** يهتم الصدق التنبؤي باستخدام درجات الاختبار في التنبؤ بالأداء المستقبلي على مقاييس أخرى هي المحكات، فقد تستخدم درجات اختبار الاستعداد للنجاح في المدرسة للتنبؤ بدرجات الطلبة في مواد دراسية معينة، أو درجات التحصيل في الثانوية لمادة الرياضيات كمؤشر للنجاح في كلية الهندسة، ودرجات التحصيل في الثانوية لمادة العلوم الطبيعية كمؤشر للنجاح في كلية الطب مثلا أو علوم الطبيعة والحياة، فيدل هذا الصدق على مدى كفاءة درجات الاختبار في التنبؤ بسلوك المشارك في الاختبار في وقت لاحق، بمعنى أنه يوجد فاصل زمني لا يقل عن 6 أشهر بين تطبيق الاختبار والمحك. (صلاح مراد وأمين سليمان، 2005)

غير أن المشكلة الأساسية في تقدير الصدق المرتبط بمحك حسب ما يذكر (صلاح الدين علام، 2000) تتركز في إمكانية الحصول على محكات مناسبة، فجمع بيانات تتعلق بالمحكات تعد من مشكلات القياس النفسي والتربوي، فالمحك يجب أن يتميز بخصائص تمكننا من **الثقة** فيما يترتب عليه من بيانات، فإلى جانب الأحكام التقييمية للمحك ينبغي أيضا التحقق من ثبات درجاته وخلوها من التحيز ومواءمة المحك للأداء المطلوب (صدقه)، فإذا اعتبرنا مثلا مقياس تقديرات المشرفين على العمل محكا لأداء العامل ولدينا مقياس نريد التأكد من صدقه هو مقياس الأداء المهني للعامل فيجب التحقق من دقة تقديرات المشرفين وعدم تحيزها لكي تكون ذات فائدة في الحكم على صدق القرار الذي يتخذ في ضوء ذلك.

2-3- صدق التكوين الفرضي: ويسمى أيضا صدق البناء أو صدق المفهوم، ومصطلح صدق التكوين الفرضي اقترحه كل من كرونباخ وميهل عام 1950 الذي يتناول العلاقة بين نتائج الاختبارات والمقاييس وبين المفهوم النظري الذي يهدف الاختبار لقياسه مثل مفهوم الذكاء، القلق، الانطواء، الابتكارية...فصدق التكوين الفرضي يهدف لتحديد التكوينات الفرضية التي يعزى إليها تباين الأداء في الاختبارات وليس درجات المحك أو سلوك الفرد.

لكن السؤال المطروح: ما المقصود بالتكوينات الفرضية؟ التكوينات الفرضية هي أفضل ما يتوصل إليه الباحثون في وقت ما لتوضيح طبيعة السمات والقدرات الموجودة لدى الانسان.

فالتكوين الفرضي إذن هو صفة أو خاصة نفسية نفترض وجودها وذلك لتفسير جوانب معينة من سلوك الأفراد، ويتطلب تقدير هذا النوع من الصدق مجموعة من الطرق نذكر أهمها فيما يلي:

-المقارنات الطرفية (الصدق التمييزي):

يقوم الصدق التمييزي حسب ما يطلق عليه البعض على أحد مفاهيم الصدق وهو قدرة الاختبار على التمييز بين طرفي الخاصية التي يقيسها (بشير معمرية، 2012)، ويمكن تلخيص إجراءات تطبيق هذه الطريقة كما يلي:

- ✓ تطبيق الاختبار وإيجاد الدرجة الكلية التي يحصل عليها كل فرد في الاختبار
- ✓ ترتيب الدرجات من أدنى درجة إلى أعلى درجة أو العكس
- ✓ أخذ 27% من الدرجات العليا و 27% من الدرجات الدنيا ليكون المجموع 54% وما يتبقى من نسبة (46%) لا تدخل في عملية الحساب.
- ✓ ايجاد المتوسطات الحسابية والانحرافات المعيارية لدرجات أفراد المجموعتين العليا والدنيا

- ✓ تطبيق اختبار ت لعينتين مستقلتين (بعد التحقق من استيفائه لشروط التطبيق واختبار تجانس النصفين) ومعرفة دلالته الاحصائية ومن ثمة اتخاذ القرار الذي يتعلق بقدرة الاختبار على التمييز بين المجموعتين الطرفيتين وهو ما يعتبر دليلا على صدق الاختبار.

-الصدق العاملي:

إن التحليل العاملي يساعد مصمم الاختبار على تحديد المكونات الأساسية (العوامل) للسمة أو القدرة التي صمم الاختبار لقياسها، والتي يمكن اعتبارها أبعادا للاختبار، بالإضافة إلى تحديد درجة تشبع مفرداته بكل عامل من هذه العوامل، وهذه التشبعات تمثل معاملات الارتباط بين مفردات الاختبار والعوامل، كما يمكن من خلاله اقتراح نموذج يجسد أبعاد الاختبار وما يندرج تحتها من مفردات، والتأكد من مطابقتها لبيانات عينته.

-الاتساق الداخلي:

من خلال حساب معاملات الارتباط بين درجات الأفراد على البنود ودرجاتهم الكلية على الاختبار، أو حساب معاملات الارتباط بين درجات الأفراد على البنود ودرجاتهم على الأبعاد الفرعية للاختبار التي تنتمي إليها تلك البنود، أو حساب معاملات الارتباط بين درجات الأفراد على الأبعاد الفرعية ودرجاتهم الكلية على الاختبار.

-الصدق التقاربي والتمايزي:

تذكر أدبيات القياس النفسي والتربوي ما يعرف بالصدق التقاربي في مقابل الصدق التمايزي، أو الصدق التوافقي في مقابل الصدق التعارضي، حيث أدرجت ضمن (صدق المفهوم الفرضي) حسب عدة باحثين من بينهم (بشير معمرية، 2012) و(امطاينوس ميخائيل، 2016) رغم أنها تعتمد على حجم العلاقة الارتباطية بين الاختبار واختبارات أخرى قد نعتبرها محكات خارجية.

فالصدق التقاربي يعني دراسة العلاقة بين الاختبار الجديد واختبار آخر مشابه، ثبت صدقه في قياس نفس التكوين الفرضي أو السمة التي يهدف الاختبار الجديد إلى قياسها، فإذا كانت قيمة معامل الارتباط بين الاختبارين مرتفعة، فإن ذلك يعد دليلا على أن الاختبار الجديد يقيس ما يقيسه الاختبار الذي ثبت صدقه، أما الصدق التمايزي نعني به حساب معاملات الارتباط بينه وبين الاختبارات الأخرى التي تقيس تكوينات فرضية أخرى مستقلة ومختلفة عما يقيسه الاختبار، ويتحدد صدق التكوين الفرضي من خلال فحص معاملات الارتباط، فإذا كانت قيم هذه المعاملات منخفضة كان ذلك دليلا على أن الاختبار يتمتع بصدق تمايزي عال. (امطاينوس ميخائيل، 2016)

ويطلق عليه (عبد الرحمن الطريري، 1997) أسلوب التشابه والاختلاف مع خصائص أخرى لتقدير الصدق، حيث يتمثل في العلاقة الممكن حدوثها أو عدم حدوثها مع مقاييس يحددها الباحث على أنها تقيس نفس الخاصية أو لا تقيسها.

ويشير (بشير معمرية، 2012) إلى ما يعرف بالصدق التوافقي أو الاتفاقي في مقابل الصدق التعارضي أو الاختلافي حيث يتحدد الصدق بوضوح بواسطة اكتشاف التوافق والتعارض بين الاختبار وغيره من الاختبارات، وفقا للإطار النظري الذي تنتمي إليه المتغيرات محل القياس، مثل ارتباط الاكتئاب باليأس وبالتشاؤم وبالميل للانتحار، وارتباط المثابرة بالدافعية وارتباط القلق بالخوف توافقيا، وكذلك ارتباط الاكتئاب بالسعادة وبالتفاؤل وارتباط الانبساطية بالخجل تعارضيا.

3-العوامل المؤثرة في صدق الاختبارات النفسية:

هناك عدد من العوامل التي تؤثر في مدى صدق الاختبار تتمثل فيما يلي:

- عوامل تتعلق بالاختبار نفسه: مثل لغة الاختبار إذا كانت فوق مستوى المفحوصين، فإن بعضهم سوف يعجز عن فهم السؤال وبالتالي لا يستطيع الإجابة عنه مما يقلل من مستواه في الاختبار، كذلك غموض أسئلة الاختبار تجعل المفحوص يفسرها تفسيرات متباينة وبجيب عنها إجابات خاطئة، تقلل هي أيضا من مستواه الفعلي في الاختبار.

كما أن سهولة أسئلة الاختبار أو صعوبتها أو طريقة صياغتها تجعل المفحوص يحصل على درجات لا يستحقها، فيتأثر حكمنا عليه بأنه مثلا ممتاز أو وسط أو ضعيف وهي لا تمثل في الحقيقة قدرة المفحوص الفعلية في الاختبار.

- طول الاختبار: إذا كان الاختبار قصيرا يضعف تمثيله للمحتوى وبالتالي يضعف صدقه، وإذا كان طويلا جدا فيدخل هنا عامل الملل وعدم الجدية من طرف المفحوص في الإجابة. (امطاينوس ميخائيل، 2016)

وكذلك ترتيب البنود بطريقة غير ملائمة، حيث من المفروض أن يرتب الاختبار وفق مبدأ التدرج في الصعوبة، فإذا وضعت البنود الصعبة في البداية يقضي المفحوص الكثير من الوقت للاجابة عنها، وقد لا يتسع الوقت للإجابة على البنود السهلة، ثم إن هذا الترتيب قد يضعف دافعيته مما ينعكس سلبا على أدائه.

-عوامل تتعلق بتطبيق الاختبار وتصحيحه: فالعوامل البيئية تؤثر على أداء المفحوص تأثيرا يقلل أو يزيد من مستوى صدق الاختبار، مثل الحرارة، البرودة، الضوضاء ....وغيرها.

كذلك طباعة الأسئلة وعدم وضوحها، التعليمات الغير واضحة، استعمال الاختبار في غير ما وضع له وعدم استعماله مع الفئة التي وضع لها، الوقت المخصص للإجابة غير كاف أو أكثر من اللزوم، فيتأثر الأداء الاختباري بالضغط على المفحوصين في الحالة الأولى، وإرباكهم وإتاحة الفرصة للغش والفوضى في الحالة الثانية، كذلك أخطاء التصحيح كلها عوامل تضعف معامل الصدق.

-عوامل تتعلق بشخصية المفحوص المؤثرة في إجاباته: فالتخمين أو الغش أو اضطراب المفحوص في الاختبار تجعله يحصل على نتيجة لا تمثل قدراته الفعلية، وهذا ما يؤثر على صدق الاختبار. (عبد الرحمن الطريري، 1997)

المحاضرة 07: الخصائص السيكومترية للاختبار (ثبات الاختبار)

1- مفهوم الثبات:

يرى (محمد غنيم، 2004، ص61) أن مفهوم الثبات يعبر عن مدى تماسك بنية الاختبار، كما يعبر أيضا عن دقة الاختبار فيما يزودنا به من بيانات عن الخاصية المراد قياسها، فثبات الاختبار يعبر عن علاقة الاختبار بنفسه والتي يمكن تقديرها بإعادة تطبيق نفس الاختبار على نفس العينة أو باستخدام اختبارات متكافئة معه.

أما (صلاح الدين علام، 2000، ص131) فيرى أن ثبات درجات الأفراد على الاختبار هو مدى خلوها من الأخطاء غير المنتظمة (العشوائية) التي تشوب عملية القياس، أي مدى قياس الاختبار للمقدار الحقيقي للسمة التي يهدف لقياسها، فدرجات الأفراد على الاختبار تكون ثابتة إذا كان الاختبار يقيس سمة معينة قياسا متسقا في الظروف المتباينة التي قد تؤدي إلى أخطاء القياس، فالثبات بهذا المعنى يعني الاتساق أو الدقة في القياس.

ويشير (سعد عبد الرحمن، 2008) إلى أن ثبات الاختبار يعني دلالة الاختبار على الأداء الفعلي أو الأداء الحقيقي للفرد، هذا الأداء يعبر عنه بالدرجة الحقيقية التي حصل عليها الفرد في اختبار ما (دح)، والأداء الحقيقي هو جزء من الأداء العام أو الأداء الكلي الذي يعبر عنه بالدرجة الكلية (دك)، وهي الدرجة الملاحظة أو المسجلة على الاختبار والتي حصل عليها الفرد، أما الجزء الآخر فهو الأداء الذي يعود إلى أخطاء الصدفة أو الظروف الخارجية ويعبر عنه بدرجة الخطأ (دخ).

ولأن: الدرجة الكلية = الدرجة الحقيقية + درجة الخطأ

وتبين نظرية الثبات حسب (محمد غنيم، 2004، ص65) أن مفهوم ثبات الاختبار هو خاصية مجموعة من درجات الاختبار وليس خاصية درجة شخص واحد فقط، ولما كان الهدف الأساسي لأي

اختبار هو دراسة الفروق الفردية بين الطلاب، والمصطلح الاحصائي الذي يستخدم لتبيان حجم الفروق بين الدرجات هو التباين، فإن المعادلة الأساسية التي يشتق منها قانون الثبات هي:

التباين الكلي = التباين الحقيقي + تباين الخطا

وعلى هذا الأساس فإن معامل الثبات هو نسبة التباين الحقيقي إلى التباين الكلي

$$\text{معامل الثبات} = \frac{\text{التباين الحقيقي}}{\text{التباين الكلي}}$$

بحيث كلما اقترب التباين الحقيقي من التباين الكلي ارتفع معامل الثبات وهو محصور بين [1,0] فأعلى قيمة يمكن أن يصل إليها معامل الثبات هي 1، وهي قيمة لا نصل إليها في مستوى القياس النفسي.

ولكن تطبيقيا الدرجات الحقيقية للأفراد على الاختبار ودرجات الأخطاء العشوائية غير معلومة، والمعلوم فقط هو الدرجات الكلية للأفراد على الاختبار، فلا يمكن منطقيا تطبيق هذا القانون لحساب معامل الثبات.

2- مصادر الأخطاء العشوائية:

قد يتساءل الباحث عن مصادر الأخطاء العشوائية من أجل التحكم فيها أثناء إعداد الاختبار وعند تطبيقه وتصحيح مفرداته، وذلك لجعل تأثيرها في الدرجات الحقيقية أقل ما يمكن، وبذلك تقترب الدرجات الملاحظة للأفراد في الاختبار من درجاتهم الحقيقية ويزيد تباين الدرجات الحقيقية إلى أقصى حد ممكن، وبالتالي تكون قيم معامل الثبات أكبر ما يمكن.

لذلك سنتطرق فيما يلي إلى أهم الأخطاء العشوائية مصنفة في ثلاث مصادر أساسية:

***مصادر تتعلق بأداة القياس:**

نجد أن مفردات الاختبار قد تكون غاية في الصعوبة، أو أن تكون صياغتها غامضة أو مربكة، أو أن تكون تعليمات الإجابة عن الاختبار غير محددة، أو تشجع على التخمين كما في مفردات الصواب أو الخطأ أو الاختيار من متعدد، غير أن المصدر الحقيقي للأخطاء العشوائية المتعلقة بالاختبار حسب (صلاح الدين علام، 2000) يتعلق بأسلوب معاينة المفردات للنطاق السلوكي الذي يقيسه الاختبار، مما يترتب عليه أن يكون عدد المفردات قليلا، أو أن المفردات لا تمثل مكونات السمة المقاسة تمثيلا كافيا.

فطول الاختبار بحد ذاته ليس مهما، لكن أهميته تكمن في شمول السمة أو القدرة محل القياس من جميع جوانبها، فالاختبار القصير قد لا نتمكن بواسطته من تناول الخاصية بكل جوانبها، ولذا لا يمكن قياسها بدقة ومن ثم سيتأثر الثبات نتيجة لذلك، لذلك فزيادة بنود الاختبار يترتب عليها زيادة في ثباته، لكن الأمر المهم في هذه النقطة هو ألا تكون زيادة بنود الاختبار لمجرد الزيادة بل أن الأمر يجب أن يستهدف تناول السمة أو القدرة من جميع جوانبها. (عبد الرحمن الطريري، 1997، ص188)

***مصادر تتعلق بإجراءات تطبيق الاختبار وتصحيحه:** تعد إجراءات تطبيق الاختبار وتصحيحه مصدرا من مصادر الأخطاء العشوائية، غير أن هذا المصدر يمكن ضبطه أو التحكم فيه لأنه يتعلق بالبيئة الفيزيائية المحيطة بالفرد أثناء الاختبار، فالإضاءة الجيدة لغرفة الاختبار وحسن تهويتها ومنع الضوضاء حولها وتوزيع المقاعد بها وغير ذلك من العوامل الفيزيائية التي قد تؤثر في إجابات الأفراد ينبغي أن يتم ضبطها قبل بدء العملية الاختبارية.

تعليمات الاختبار ينبغي أن تكون واضحة ومفهومة بخاصة في الاختبارات الفردية أو تلك التي تطبق على الأطفال أو الأميين، حيث يفضل أن تكون شفوية ومناسبة لهم.

كما يجب أن يقاس زمن الإجابة عن الاختبارات التي تتطلب السرعة قياسا دقيقا، وأن تكون طباعة الاختبار واضحة وأوراق الإجابة منظمة بطريقة تسمح للفرد بتحديد إجابته بيسر وسهولة.

يجب تصحيح الاختبار بطريقة موضوعية لتقليل أثر الأخطاء الناجمة عن اختلاف التقديرات، وبخاصة إذا كانت مفردات الاختبار تتطلب إجابات مفتوحة كما في أسئلة المقال أو بعض مقاييس الشخصية والاتجاهات، كذلك كيفية رصد الدرجات والدقة في العمليات الحسابية التي قد يقوم بها الباحث للتوصل إلى الدرجة الكلية في الاختبار.(صلاح الدين علام، 2000)

ويضيف (عبد الرحمن الطريري، 1997) أن الاختبارات ذات المحكات الواضحة والمحددة عند تصحيحها تقل فيها فرص خطأ القياس، لكن بالنسبة للاختبارات التي تكون محكات تصحيحها غير محددة تحديدا دقيقا وواضحا فيكون التحيز من قبل مصحح الاختبار واردا في مثل هذه الحالة، وتعتبر الاختبارات الاسقاطية من أكثر الاختبارات عرضة لتحيز وذاتية الباحث، لذلك من بين الانتقادات الموجهة للاختبارات الاسقاطية ما يتعلق بانخفاض ثباتها.

عوامل التحيز مثل أثر الهالة أو عوامل ذاتية أخرى تدخل في التصحيح مثل أن يقوم المعلم في اختبار تحصيلي من نوع المقال بتقدير درجات الطلبة متأثرا بعوامل لا تتعلق بإجاباتهم مثل أخطاء النحو والهجاء أو جودة الخط، أو فكرته الذاتية عن الطالب.

***مصادر تتعلق بالأفراد المختبرين:**

هي مصادر يصعب التحكم فيها لأنها تتعلق بالفرد نفسه، مثلا تذكر الإجابات عن مفردات الاختبار نتيجة الخبرة السابقة أو إعادة تطبيق صيغة مكافئة للاختبار، أو نتيجة لمرض أو انخفاض دافعيته أو اتجاهه السلبي أو تشتت انتباهه، والحالة المزاجية.

فالاختبارات التحصيلية أو اختبارات الذكاء تتطلب من الفرد بذل أقصى جهده للحصول على درجة مرتفعة، غير أن هناك أفرادا يتسمون في معظم الأحيان بانخفاض دافعيتهم للتحصيل (لأسباب مختلفة) أما في مقاييس الشخصية فقد يميل الفرد إلى الإجابة بطريقة تبدو مرغوبة اجتماعيا لكنها لا تعكس شخصيته فعلا. (صلاح الدين علام، 2000)

المحاضرة 08: طرق حساب الثبات

3-طرق حساب الثبات

نظرا لاستحالة معرفة الدرجات الحقيقية ودرجة الخطأ العشوائي في أي عملية قياس مهما حرص الباحث على توفير الظروف الملائمة، ومهما كانت درجة دقة وموضوعية الاختبار المطبق فقد ظهرت معادلات مختلفة يتم من خلالها تقدير الثبات.

3-1- طريقة التطبيق وإعادة التطبيق:

تعرف أيضا بطريقة إعادة الاختبار، ويسمى المعامل الناتج بمعامل الاستقرار أو معامل ثبات التجانس عبر الزمن، حيث يطبق الاختبار على مجموعة من الأفراد ثم يعاد تطبيق الاختبار على نفس الأفراد وتحت نفس الظروف، ويتم ايجاد قيمة معامل ارتباط بيرسون بعد التأكد من شروط تطبيقه، ويكون الفاصل الزمني بين التطبيقين من أسبوعين إلى ثلاث أسابيع، ويمكن أن تصل إلى ستة أسابيع. (صلاح مراد وأمين سليمان، 2005)

يصلح هذا المعامل إذا كانت السمة التي يقيسها الاختبار لا تتغير كثيرا بمرور الزمن، أي تكون مستقرة نسبيا، بحيث أن اختلاف درجات الأفراد في مرتي تطبيق الاختبار يعزى إلى الأخطاء العشوائية.

فمثلا لا يصلح استخدام هذا المعامل في تقدير ثبات درجات اختبار يقيس ذكاء الأطفال الصغار أو إنتاجهم اللغوي، أو درجات مقياس الحالة المزاجية أو حالة القلق، كما لا يصلح في اختبارات التحصيل الدراسي أو الاختبارات التي تهدف إلى قياس التذكر. (صلاح الدين علام، 2000)

مثال تطبيقي:

طبق باحث اختبار لقياس دافعية الانجاز على عينة من مستشاري التوجيه المدرسي والمهني بولاية عين الدفلى تتكون من 10 أفراد، وبقصد التحقق من ثبات درجات الأفراد على الاختبار أعيده تطبيقه على العينة نفسها بعد مرور أسبوعين، فكانت نتائج التطبيقين كالتالي:

أفراد العينة	01	02	03	04	05	06	07	08	09	10
نتائج التطبيق01	22	15	12	14	10	15	14	20	19	25
نتائج التطبيق 02	20	12	11	10	14	15	18	25	22	28

المطلوب: احسب معامل الثبات بطريقة التطبيق وإعادة التطبيق.

ما هي العوامل التي يمكن أن تؤثر في معامل الثبات وفق هذه الطريقة؟

الحل:

لإيجاد معامل ثبات الاختبار نحسب معامل ارتباط بيرسون بين درجات الأفراد في التطبيق الأول (x) ودرجاتهم في التطبيق الثاني (y) ، بافتراض أن شروطه مستوفاة في بيانات العينة (بيانات كمية، خطية العلاقة بين المتغيرين، اعتدالية توزيع بيانات المتغيرين) وفق القانون التالي:

$$R = \frac{N\sum xy_(\sum x)(\sum y)}{\sqrt{[N\sum x^2-(\sum x)^2]\,[N\sum y^2-(\sum y)^2]}}$$

N	X	Y	Xy	x^2	y^2
01	22	20	440	484	400
02	15	12	180	225	144
03	12	11	132	144	121
04	14	10	140	196	100
05	10	14	140	100	196
06	15	15	225	225	225
07	14	18	252	196	324
08	20	25	500	400	625
09	19	22	418	361	484
10	25	28	700	625	784

$\sum$	**166**	**175**	**3127**	**2956**	**3403**

$$R = \frac{10(3127)_(166)(175)}{\sqrt{[10(5952)-(166)^2]\,[10(3403)-(175)^2]}} = 0.85$$

نلاحظ أن الاختبار على درجة عالية من الثبات.

العوامل التي يمكن أن تؤثر في معامل الثبات وفق طريقة التطبيق وإعادة التطبيق هي:

- إذا كان الفاصل الزمني بين التطبيقين قصيرا، فإن عامل الذاكرة سيؤثر في استجابات المفحوصين، حيث أن المفحوص سيتذكر إجابته الأولى ويجيب بنفس الطريقة مباشرة.
- إذا حاول الباحث تجنب ذلك وباعد في المدة الزمنية بين التطبيقين لكي يقلل من تأثير عامل الذاكرة فإنه قد يصادف عاملا جديدا هو عامل النمو والنضج، بالإضافة إلى بعض التغيرات التي قد تطرأ على المفحوصين ما بين التطبيقين مثل: الحالة المزاجية
- قد يتعذر الوصول إلى الأفراد أنفسهم في المرة الثانية لعوامل متعددة.

3-2- طريقة الصور المتكافئة:

تستلزم هذه الطريقة إعداد صورتين متكافئتين للاختبار وتطبيقهما على نفس الأفراد، والمقصود بالتكافؤ هنا: تساوي عدد ونوع البنود وطريقة الإجابة والتصحيح والزمن المخصص للإجابة، والتعليمات إضافة إلى تساوي معاملات الصعوبة والتمييز والتباين والمتوسط الحسابي والانحراف المعياري.

وتطبق إحدى الصورتين على مجموعة من الأفراد ثم تطبق الصورة الثانية بعد فاصل زمني قصير على نفس الأفراد (جلسة اختبارية واحدة أو جلستين يفصل بينهما وقت قصير)، وتصحح الصورتان ثم يحسب معامل ارتباط بيرسون بين الدرجات، ويسمى معامل الثبات الناتج بمعامل التكافؤ. (صلاح مراد وأمين سليمان، 2005)

وتتميز هذه الطريقة بعدم التأثر بالتدريب أوالخبرة من تطبيق لآخر، ويختفي أثر الألفة لاختلاف بنود الصورتين، كما أنها تصلح لحساب معامل ثبات اختبارات الذاكرة والاختبارات التحصيلية.

ومن سلبيات هذه الطريقة صعوبة تصميم اختبارين متكافئين لقياس نفس السمة، إضافة إلى الجهد والوقت المطلوب لذلك.

3-3- معامل التكافؤ والاستقرار:

تجمع هذه الطريقة بين الطريقتين السابقتين، حيث يتم إعداد صورتين من الاختبار، تطبق إحداهما على مجموعة من الأفراد على أن تطبق الصورة الثانية بعد مدة زمنية، ويتم إيجاد معامل ارتباط بيرسون بين درجات الأفراد في الصورتين.

يرى (صلاح الدين علام، 2000) أن هذه الطريقة وبحكم أنها تجمع ما بين طريقة الصور المتكافئة وطريقة التطبيق وإعادة التطبيق فإنها تؤدي إلى قيم تقديرية لمعامل الثبات أقل من نظريتها في أي من الطريقتين السابقتين، وذلك لأنها تجمع بين الأخطاء العشوائية التي تؤثر في كل منهما، فمعامل الاستقرار والتكافؤ يعكس الأخطاء العشوائية الناجمة عن اختلاف مفردات صيغتي الاختبار، وكذلك اختلاف الدرجات نتيجة التغيرات التي تحدث للأفراد المختبرين أو تذبذب السمة المقاسة، ولذلك يعد هذا المعامل الأدق، ويعتبر بمثابة الحد الأدنى لتقدير معامل الثبات.

المحاضرة 09: أساليب الاتساق الداخلي (تابع لطرق حساب الثبات)

3-4- أساليب الاتساق الداخلي:

تتطلب الطرق الثلاث السابقة تطبيق الاختبار مرتين، أو صيغ متكافئة منه بفواصل زمنية متباينة، غير أنه في بعض الأحيان يصعب بناء صيغتين متكافئتين أو قد يصعب تطبيق الاختبار مرتين أو قد يكون من غير المناسب في بعض السمات، في هذه الحالة يفضل تقدير الثبات باستخدام اختبار واحد وتطبيقه مرة واحدة، ولذلك يمكن تقدير ثبات الاختبار باستخدام أساليب الاتساق الداخلي.

3-4-1- التجزئة النصفية:

يطبق الاختبار ثم يجزأ إلى نصفين متكافئين، لذلك فهذه الطريقة تهتم بتقييم الاتساق الداخلي لبنود الاختبار، فيتم تقدير درجات الأفراد على كل من نصفي الاختبار كما لو كان كل منهما اختبارا منفصلا.

يحدد (عبد العزيز بوسالم، 2012) اجراءات تقسيم الاختبار إلى نصفين متكافئين كما يلي:

- ✓ تطبيق الاختبار على العينة.
- ✓ حساب الدرجات التي يحصل عليها الأفراد في كل بند
- ✓ للحصول على نصفين متكافئين يجب حساب معاملات السهولة/ الصعوبة أو التمييز لكل بند، ثم ترتيب البنود وفقا لذلك من أسهل بند إلى أصعب بند (أو العكس)، أو من البند الأقل تمييزا إلى البند الأكثر تمييزا (أو العكس)

- ✓ يتم تقسيم البنود إلى نصفين النصف الأول يحوي البنود ذات الأرقام الفردية 1، 3، 5، 7، 9... والنصف الثاني يحوي البنود ذات الأرقام الزوجية 2، 4، 6، 8، 10...
- ✓ إعداد جدول يحتوي على بنود النصف الأول ودرجات الأفراد عليها، وبنود النصف الثاني ودرجات الأفراد عليها.
- ✓ يتم حساب تباين درجات الأفراد على النصف الأول، وتباين درجات الأفراد على النصف الثاني.
- ✓ يتم اختبار تجانس نصفي الاختبار عن طريق اختبار "ف" بحيث : **ف المحسوبة** = التباين الكبير / التباين الصغير
- ✓ ويتم مقارنة ف **المحسوبة** مع ف **المجدولة** عند مستوى دلالة 0,05 أو 0,01

درجة حرية (1)= ن−1 للبسط

درجة حرية (2)= ن−1 للمقام

إذا كانت ف **المحسوبة** أكبر أو تساوي ف **المجدولة** يتم رفض الفرضية الصفرية ما يعني أن التباينين غير متساويين، وبالتالي: النصفان غير متجانسين.....الحالة (1)

وإذا كانت ف **المحسوبة** أصغر من ف **المجدولة** يتم قبول الفرضية الصفرية ما يعني أن التباينين متساويان، وبالتالي النصفان متجانسان.....الحالة (2)

في الحالة (1) يتم تطبيق معادلة **جتمان Guttman** لحساب الثبات، أو معادلة **رولون Rulon**

$$R_{\text{Guttman}} = 2\left(1 - \frac{S_1{}^2 + S_2{}^2}{S^2}\right)$$

حيث أن $S_1{}^2$: تباين درجات الأفراد على النصف الأول من الاختبار.

$S_2{}^2$: تباين درجات الأفراد على النصف الثاني من الاختبار.

S^2 : تباين درجات الأفراد الكلية على الاختبار.

$$R_{\text{Rulon}} = \frac{S^2 x - y}{S^2}$$

$S_2{}^2$: تباين الفرق بين درجات الأفراد على نصفي الاختبار.

S^2 : تباين درجات الأفراد الكلية على الاختبار.

أما في الحالة (2) يتم تطبيق معادلة **سبيرمان– براون Spearman– Brown** لحساب الثبات.

$$R_{\text{Spearman- Brown}} = \frac{2R}{1+R}$$

حيث أن R : قيمة معامل ارتباط بيرسون بين درجات الأفراد على نصفي الاختبار.

ملاحظة: يذكر (صلاح مراد وآخرون، 2017) أنه: في حالة تساوي حجمي المجموعتين فإنه لا ضرورة لاختبار التجانس، مباشرة نستنتج أنه يوجد تجانس بين المجموعتين.

مثال تطبيقي:

قام باحث بتطبيق اختبار يتكون من 12 بندا لقياس الانتاج الفونولوجي لدى 15 طفلا مستفيدا من جهاز الزرع القوقعي، وكانت نتائجهم كما هو مبين في الجدول التالي:

البنود	01	02	03	04	05	06	07	08	09	10	11	12	∑
01	1	0	0	1	1	1	0	1	0	0	0	1	6
02	1	1	1	0	1	0	0	1	1	1	1	1	9
03	1	1	0	0	1	1	1	0	0	0	0	0	5
04	1	1	0	1	1	1	0	1	1	1	1	1	10
05	0	0	1	1	0	1	1	1	0	0	0	0	5
06	0	1	1	0	1	1	0	1	0	0	0	1	6
07	1	1	0	1	0	1	0	1	1	0	1	1	8
08	0	1	1	1	0	1	1	1	0	0	1	0	7
09	1	0	0	0	1	1	0	1	0	0	0	1	5
10	1	0	1	1	1	1	1	1	0	0	1	0	8
11	1	0	0	0	1	0	1	1	0	0	1	1	6
12	1	1	1	0	0	1	0	1	0	0	1	1	7
13	1	1	1	1	1	1	1	1	0	0	1	1	10
14	1	0	1	1	1	1	1	1	1	1	1	0	10
15	0	1	0	0	0	1	0	0	0	1	1	1	5
∑	**11**	**9**	**8**	**8**	**10**	**13**	**7**	**13**	**4**	**4**	**10**	**10**	
م س	$\frac{11}{15}$	$\frac{9}{15}$	$\frac{8}{15}$	$\frac{8}{15}$	$\frac{10}{15}$	$\frac{13}{15}$	$\frac{7}{15}$	$\frac{13}{15}$	$\frac{4}{15}$	$\frac{4}{15}$	$\frac{10}{15}$	$\frac{10}{15}$	
	0,73	0,60	0,53	0,53	0,66	0,86	0,46	0,86	0,26	0,26	0,66	0,66	

المطلوب: احسب معامل ثبات درجات الأفراد على الاختبار بطريقة التجزئة النصفية.

الحل:

حتى نتحصل على نصفين متكافئين نحسب معاملات السهولة لكل بند، ثم نرتب البنود وفقا لذلك من أسهل بند إلى أصعب بند.

ترقيم البنود	01	02	03	04	05	06	07	08	09	10	11	12
البنود مرتبة من الأسهل إلى الأصعب	06	08	01	05	11	12	02	03	04	07	11	12
درجات الأفراد	13	13	11	10	10	10	9	8	8	7	4	4

- ✓ نقسم البنود إلى نصفين النصف الأول يحوي البنود ذات الأرقام الفردية 1، 3، 5، 7، 9... والنصف الثاني يحوي البنود ذات الأرقام الزوجية 2، 4، 6، 8، 10...
- ✓ نعد جدولا يحتوي على بنود النصف الأول ودرجات الأفراد عليها، وبنود النصف الثاني ودرجات الأفراد عليها.

	النصف1	النصف2
01	13	13
02	11	10
03	10	10
04	09	08
05	08	07
06	04	04

- ✓ نختبر تجانس درجات نصفي الاختبار عن طريق اختبار "ف" بحيث:

$$\text{ف المحسوبة} = \frac{\text{التباين الكبير}}{\text{التباين الصغير}}$$

فيتم حساب تباين درجات الأفراد على النصف الأول، وتباين درجاتهم على النصف الثاني.

$S_1^2 = 9.36$ $S_2^2 = 9.46$

$$\text{ف المحسوبة} = \frac{9.46}{9.36} = 1.01$$

ويتم مقارنة ف المحسوبة مع ف المجدولة عند مستوى دلالة 0,05

درجة حرية (1)= 6-1 =5 للبسط
درجة حرية (2)= 6-1 =5 للمقام

بالرجوع إلى جدول التوزيع الخاص بـ: "ف" نجد ف المجدولة = 5,05

إذن ف المحسوبة أصغر من ف المجدولة، وبالتالي يوجد تجانس بين النصفين.

- ✓ نطبق معادلة سبيرمان- براون لحساب الثبات، والتي تتطلب حساب معامل ارتباط بيرسون بين درجات الأفراد على النصف الأول ونرمز لها (x) ودرجاتهم على النصف الثاني ونرمز لها (y)

N	X	Y	Xy	x^2	y^2
01	13	13	169	169	169
02	11	10	110	121	100
03	10	10	100	100	100
04	09	08	72	81	64
05	08	07	56	64	49
06	04	04	16	16	16
$\sum$	55	52	523	551	498

$$R = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2]\,[N\sum y^2 - (\sum y)^2]}}$$

$$R = \frac{6(523) - (55)(52)}{\sqrt{[6(551) - (55)^2]\,[6(498) - (52)^2]}} = 0.98$$

$$R_{\text{Spearman- Brown}} = \frac{2R}{1+R} = \frac{2(0.98)}{1+0.98}$$

$$R_{\text{Spearman- Brown}} = 0.99$$

نلاحظ أن الاختبار على درجة عالية جدا من الثبات.

مثال تطبيقي 02: باستخدام نفس معطيات المثال السابق، مع افتراض أن نصفي الاختبار غير متجانسين، احسب معامل الثبات بطريقتي جتمان ورولون.

الحل:

أولا: طريقة جتمان

$$R_{\text{Guttman}} = 2\left(1 - \frac{{S_1}^2 + {S_2}^2}{S^2}\right)$$

حيث أن ${S_1}^2$: تباين درجات الأفراد على النصف الأول من الاختبار.

${S_2}^2$: تباين درجات الأفراد على النصف الثاني من الاختبار.

S^2 : تباين درجات الأفراد الكلية على الاختبار.

لدينا: ${S_1}^2 = 9.36$ ${S_2}^2 = 9.46$ $S^2 = 3.69$

$$R_{\text{Guttman}} = 2\left(1 - \frac{9.36 + 9.46}{3.69}\right) = 8.20$$

(الاختبار غير ثابت لأنها ليست الطريقة المناسبة)

ثانيا: طريقة رولون

$$R_{Rulon} = \frac{S^2_{x-y}}{S^2}$$

S^2_{x-y} : تباين الفرق بين درجات الأفراد على نصفي الاختبار.

S^2 : تباين درجات الأفراد الكلية على الاختبار.

N	X	Y	x-y
01	13	13	0
02	11	10	1
03	10	10	0
04	09	08	1
05	08	07	1
06	04	04	0

$$R_{Rulon} = \frac{0.3}{3.69} = 0.08$$

(الاختبار غير ثابت لأنها ليست الطريقة المناسبة)

المحاضرة 10: معادلات كيودر ريتشاردسون وألفا لكرونباخ لحساب الثبات (تابع لطرق لحساب الثبات)

3-4-2- معادلة كيودر ريتشاردسون Kuder_ Richardson

هذه المعادلة قابلة للتطبيق فقط في الاختبارات التي تكون الإجابة على فقراتها إما صحيحة فتأخذ درجة واحدة، أو خاطئة فتأخذ صفرا.

توجد صيغتان لحساب معادلة كيودر ريتشاردسون الصيغة 20 وهي الصيغة العامة، تحسب وفق القانون التالي:

$$K\text{-}R20 = \frac{n}{n-1}\left(1 - \frac{\sum pq}{s^2}\right)$$

n: عدد الفقرات

p: معامل صعوبة الفقرة

q: معامل سهولة الفقرة

S^2 : تباين درجات الأفراد على الاختبار

مثال تطبيقي:

يوضح الجدول التالي نتائج 10 طلاب في اختبار مكون من 5 بنود:

البنود الأفراد	البند1	البند2	البند3	البند4	البند5	الدرجة الكلية
01	1	0	1	0	1	3
02	1	1	0	0	0	2
03	1	1	1	1	1	5
04	0	0	1	1	1	3
05	0	1	1	0	1	3
06	1	1	1	1	1	5
07	1	1	0	1	0	3
08	1	1	1	1	1	5
09	1	0	0	1	1	3
10	1	1	1	1	1	5

المطلوب: احسب الثبات بطريقة كيودر ريتشاردسون

الحل:

N	P	Q	Pq
1	0.8	0.2	0.16
2	0.7	0.3	0.21
3	0.7	0.3	0.21
4	0.7	0.3	0.21
5	0.8	0.2	0.16
∑pq			0.95

$S^2 = 1.34$ $\sum pq = 0.95$ $n = 5$

بتطبيق قانون معادلة كيودر ريتشارسون الصيغة 20 نتحصل على: $K\text{-}R20 = \frac{5}{5-1}\left(1 - \frac{0.95}{1.34}\right)$

وبالتالي: $K\text{-}R20 = 0,36$ نلاحظ أن معامل الثبات منخفض.

3-4-3- معادلة كيودر ريتشاردسون الصيغة 21

تم اشتقاق صيغة أخرى من معادلة كيودر ريتشاردسون الصيغة 20 هي الصيغة 21، تستخدم في حال افتراض أن جميع الفقرات ذات مستوى صعوبة واحد، تحسب وفق القانون التالي:

$$K\text{-}R20 = \frac{n}{n-1}\left(1 - \frac{X(n-X)}{nS^2}\right)$$

n: عدد الفقرات

X: متوسط درجات الأفراد على الاختبار

S^2 : تباين درجات الأفراد على الاختبار

مثال تطبيقي:

أجرى باحث اختبارا للذكاء مكونا من 10 فقرات من نوع صحيح وخطأ (1،0)، وكان متوسط علامات الطلبة 8، وتباين درجاتهم على الاختبار 0,5

المطلوب: احسب معامل الثبات بافتراض أن فقرات الاختبار متساوية الصعوبة.

الحل:

$$K\text{-}R20 = \frac{10}{10-1}\left(1 - \frac{8(10-8)}{10(0.5)}\right)$$

$$K\text{-}R20 = 2{,}44-$$

درجات الأفراد على الاختبار غير ثابتة.

3-4-4- معامل ثبات ألفا لكرونباخ Cronbach Alpha

هي الطريقة التي اقترحها كرونباخ عام (1951) لتقدير ثبات الاتساق الداخلي، وهي تعميم لمعادلة كيودر ريتشاردسون الصيغة 20 عندما لا يتم تصحيح الفقرات بشكل ثنائي (ص:1، خ: 0)، حيث تستخدم هذه الطريقة في تقدير ثبات مقاييس الاتجاهات، القيم، الميول، والشخصية حيث الاستجابة تكون حسب ميزان متدرج مثل: يحدث دائما، يحدث غالبا، يحدث أحيانا، يحدث نادرا، لا يحدث أبدا.

وجدير بالذكر أن طريقة ألفا لكرونباخ تعطي الحد الأدنى للقيمة التقديرية لمعامل ثبات درجات الاختبارات، فإذا كانت قيمة ألفا مرتفعة فهذا يدل بالفعل على ثبات الاختبار، أما إذا كانت منخفضة فالثبات يمكن أن تكون قيمته أكبر من ذلك باستخدام الطرق الأخرى. (موسى النبهان، 2013)

من ناحية رياضية تظهر معادلة ألفا لكرونباخ على أنها صيغة أخرى لمعادلة كيودر ريتشاردسون، سوى أن مجموع معاملات السهولة X معاملات الصعوبة قد استبدل بمجموع تباينات درجات الأفراد على الفقرات، وكما ذكرنا سابقا فإن تباين درجات الفقرة= معامل السهولة X معامل الصعوبة

$$\propto Cronbach = \frac{n}{n-1}\left(1 - \frac{\sum Si^2}{S^2}\right)$$

n: عدد الفقرات

S_i^2: تباين درجات الأفراد على كل فقرة من الاختبار.

S^2 : تباين درجات الأفراد على الاختبار

مثال تطبيقي:

لدينا مقياس للاتجاهات متدرج الميزان ثلاثي التدريج تتم الاستجابة عليه كالتالي: موافق (3)، محايد (2)، غير موافق (1)، يشتمل على 6 بنود وطبق على 5 أفراد، وكانت درجاتهم كما يلي:

البنود الأفراد	البند1	البند2	البند3	البند4	البند5	البند6	الدرجة الكلية
01	2	1	3	2	3	1	12
02	1	2	3	3	2	3	14
03	2	1	2	3	1	2	11
04	2	3	3	1	1	1	11
05	1	2	3	2	3	2	13

المطلوب: احسب معامل ثبات ألفا لكرونباخ.

الحل:

البنود	البند1	البند2	البند3	البند4	البند5	البند6	درجات الأفراد على الاختبار ككل
التباين	0,30	0,70	0,20	0,70	1,00	0,76	1,7

$$\propto Cronbach = \frac{6}{6-1}\left(1 - \frac{3.66}{1.7}\right) \quad = -1.38$$

درجات الأفراد على الاختبار غير ثابتة

4- العوامل المؤثرة في ثبات الاختبار

يتأثر ثبات درجات الأفراد على الاختبار بمجموعة من العوامل مهما كانت الطريقة المستخدمة في حسابه، يمكن تلخيصها فيما يلي:

- **طول الاختبار**: يقصد بطول الاختبار عدد فقراته، حيث يفترض أن زيادة عدد الفقرات يتيح تغطية أكبر للمحتوى، وهذا يحقق صدق المحتوى. (ابراهيم محاسنة، 2013، ص142)

كما أن العدد الأكبر من الفقرات يؤدي إلى الحصول على عينة أكبر من السلوك، وبالتالي يكون من المتوقع أن يمثل بهذه العينة بشكل مستقر العدد الأكبر من مكونات السلوك أو السمة أو القدرة المقاسة والقابلة للظهور في مرتي التطبيق أو في نصفي أو جزئي الاختبار، في حين أن الفقرات القليلة تقلل من استقرار مكونات السلوك أو الصفة المراد قياسها وبالتالي انخفاض معامل الثبات.(سوسن مجيد، 2014، ص133)

والاختبارات الأكثر طولا تعطي درجات أكثر ثباتا، لأن الأخطاء العشوائية الموجبة والسالبة سيلغي أحدهما الآخر، وبذلك تقترب الدرجة الملاحظة من الدرجة الحقيقية، وبنفس الطريقة تعطي الاختبارات القصيرة درجات أقل ثباتا. (موسى النبهان، 2013، ص 309)

ويمكن استخدام معادلة سبيرمان –براون **Spearman– Brown** في تحديد عدد الفقرات التي ينبغي إضافتها للوصول إلى معامل الثبات المرغوب، كما يمكن استخدام هذه المعادلة في تقليل عدد الفقرات بمعامل ثبات معين، حيث أن معادلة سبيرمان –براون تساعد على التنبؤ بالزيادة المنتظرة في معامل الثبات إذا أضفنا إلى الاختبار عددا من الفقرات، والمعادلة هي:

$$R= \frac{k.r}{1+(k-1)r}$$

حيث أن R: هو معامل الثبات المتوقع لاختبار طوله عدد مرات الاختبار الأصلي.

r: معامل ثبات الاختبار الأصلي.

k: نسبة عدد الفقرات في الاختبار الجديد إلى عدد الفقرات في الاختبار الأصلي.

مثال تطبيقي 01:

إذا كانت قيمة معامل ثبات اختبار 0,60، يتكون من 20 فقرة وأردنا زيادة عدد الفقرات لتصبح 60 فقرة، فماذا نتوقع أن تكون قيمة معامل ثبات الاختبار المطول؟

الحل: $r=0.60 \qquad k= \frac{60}{20} =3$

$$R= \frac{3(0.60)}{1+(3-1)0.60}$$

$$R = 0.82$$

نلاحظ أن معامل الثبات قد ارتفع من 0,60 إلى 0,82 بعد زيادة عدد فقرات الاختبار بثلاث أضعاف.

من ناحية أخرى، قد يتمتع اختبار ما بدرجة ثبات عالية، إلا أنه أطول مما نستطيع أن نستخدم، ويمكن استخدام المعادلة في تقدير ذلك.

مثال تطبيقي 02:

إذا كانت قيمة معامل ثبات اختبار 0,82، يتكون من 100 فقرة وأردنا تقليل عدد الفقرات لتصبح النصف (50 فقرة)، فماذا نتوقع أن تكون قيمة معامل ثبات الاختبار الجديد؟

الحل: $r=0,82$ $k=\frac{50}{100}=0,5$

$$R=\frac{0.5(0.82)}{1+(0.5-1)0.82}=0,69$$

$$R=0,69$$

نلاحظ أن معامل الثبات قد انخفض من 0,82 إلى 0,69 بعد تقليل عدد فقرات الاختبار إلى النصف.

كما يمكن حساب قيمة k عن طريق استخدام المعادلة التالية:

$$k=\frac{R(1-r)}{r(1-R)}$$

مثال تطبيقي 03:

إذا كانت قيمة معامل ثبات اختبار يتألف من 30 فقرة هي: 0,60 وبعد عرضه على المحكمين أشاروا بضرورة تطويره ليصبح معامل ثباته أفضل وبما لايقل عن 0,70، فكم فقرة يجب إضافتها للاختبار لتحقيق هذا المستوى من الثبات؟

$r=0,60$ $R=0.70$

$$k=\frac{0.70(1-0.60)}{0.60(1-0.70)}$$

$$k=1,5$$

بناء على ذلك، يتم إطالة الاختبار مرة ونصف، أي يصبح طول الاختبار 45 فقرة (30 X 1,5 = 45)، وبالتالي يجب إضافة 15 فقرة جديدة للاختبار.

- **تجانس/ تباين درجات أفراد العينة:** يزداد معامل ثبات الاختبار إذا ازدادت قيمة تباين درجات أفراد العينة، بمعنى أنه كلما كان أداء أفراد العينة متباينا زاد مقدار ثبات الاختبار.

وعلى العكس، يؤدي التجانس في العينة إلى انخفاض واضح في معامل الثبات، لأن التباين داخل هذه العينة المتجانسة يكون منخفضا بقدر لا يسمح بتقدير التباين الحقيقي، في حين كلما كان كبر حجم العينة زاد تباين درجات أفرادها على الاختبار وكانت بذلك أقرب إلى التوزيع الاعتدالي، وبالتالي يصبح تباين الخطأ صغيرا، فتؤدي هذه النتيجة إلى ارتفاع ثبات الاختبار.(سوسن مجيد، 2014، ص138)

- **صعوبة الفقرات:** بما أن معامل الثبات يعتمد على التباين في درجات المفحوصين، فإن معامل الثبات يزداد تبعا للعوامل التي تحسن من مستوى ذلك التباين، وهذا يتم من خلال الأداء على الفقرات التي تتمتع بمعاملات صعوبة متوسطة، إذ أن الفقرة السهلة جدا أو الصعبة جدا لا تبرز الفروق الفردية، وهي لا تظهر التباين بين درجات المفحوصين، الأمر الذي يجعل الثبات منخفضا. (موسى النبهان، 2013، ص312)
- **موضوعية التصحيح:** تعتبر موضوعية التصحيح من العوامل المؤثرة في ثبات الاختبار، ولاسيما في الاختبارات التي تعتمد على تقدير المصحح كاختبارات المقال التحصيلية واختبارات الإبداع أو الاختبارات الإسقاطية، حيث أن تباين التصحيح يؤدي إلى زيادة تباين الخطأ وبالتالي إلى نقصان معامل الثبات، لذلك ينبغي أن يتضمن الاختبار وصفا دقيقا لمحكات الإجابة والتصحيح وطريقة وضع الدرجة عليها.(سوسن مجيد، 2014، ص136)
- **زمن الاختبار:** يؤثر الزمن المحدد للإجابة على الاختبار على ثباته، فيزداد الثبات تبعا لزيادة الزمن حتى يصل إلى الحد المناسب للاختبار، أي الزمن الذي تتطلبه الإجابة على هذا الاختبار، فيصل الثبات إلى أعلى درجة ممكنة من جراء تأثير هذا العامل، أما إذا كان الاختبار يتطلب وقتا كبيرا لإنجازه فإن ذلك سيؤدي إلى خفض معامل الثبات، لأن ذلك قد يؤدي إلى الإجهاد والتعب وضعف الدافعية على الإجابة، وبالتالي قد تتعرض الفقرات الأخيرة من الاختبار إلى أخطاء في الإجابة، ومن ثم تؤدي إلى زيادة تباين الخطأ.

  وعليه ينبغي تحديد الزمن المطلوب للاختبار وفقا للجهد المطلوب للإجابة عليه، ومدى قدرة المفحوص على الاستمرار في الإجابة دون تعب، وتحديد مدة الاختبار المطلوبة يمكن الوصول إليها عن طريق تجريبه على عينة استطلاعية.
- **أثر التخمين:** حيث أنه كلما خمن المفحوصون في الإجابة يقل ثبات الاختبار، فقد تختلف الإجابة بالتخمين في مرتي التطبيق لنفس الاختبار أو لصورتين متكافئتين منه. (ابراهيم محاسنة، 2013، ص142)

5- عوامل أخرى: مثل ظروف ضبط موقف التطبيق، دافعية المفحوص للإجابة، المؤثرات الفيزيقية والمشتتات المتعددة في موقف الاختبار، الحالة الصحية والانفعالية للمفحوص، حيث ينبغي أن يقوم الباحث بضبط دقيق لكل هذه العوامل حتى يتمكن من الوصول إلى معاملات ثبات مرتفعة. (سوسن مجيد، 2014، ص138)

المحاضرة 11: معايير تفسير النتائج ومحكات الأداء

1- المعـــــايير:

تعد المعايير من المفاهيم الأساسية المتعلقة بتفسير درجات الاختبارات، فالدرجة التي يحصل عليها فرد في اختبار ما والتي تسمى بالدرجة الخام لا يكون لها معنى ويصعب تفسيرها ما لم يتم اسنادها إلى نظام مرجعي، هذا النظام هو الذي يسمح باستخلاص معلومات مفيدة من درجات الاختبار.

هذا النظام المرجعي يعتمد على استخدام معلومات يتم الحصول عليها من جماعة معينة تسمى بالجماعة المرجعية، لذلك تسمى هذه الاختبارات بالاختبارات مرجعية الجماعة أو مرجعية المعيار.

1-1- مفهوم المعايير:

فالمعايير حسب (صلاح مراد وأمين سليمان، 2005) هي نوع من الموازين تستخدم في تفسير الدرجات الخام التي يحصل عليها المفحوص، كما أن إعدادها يعتمد أيضا على الدرجات الخام للعينة المعيارية وهي عينة التقنين أو الجماعة المرجعية، والهدف من استخدام المعايير هو تحديد مستوى أداء الفرد في السمة المقاسة بالنسبة لمتوسط درجات الجماعة المرجعية التي ينتمي إليها (أقرانه في المرحلة العمرية، المرحلة الدراسية...) حيث تفسر على أنها أقل أو أكبر من المتوسط.

ولكي نحدد مركز الفرد بالنسبة للجماعة المرجعية ينبغي تحويل درجته الخام إلى نوع أخر من الدرجات تسمح بعملية المقارنة مثل الدرجات المعيارية، الدرجات المعيارية المعدلة، الميئنيات...

1-2- خصائص عامة يجب أن تتوفر في الجماعة المرجعية:

- تمثيل عينة الأفراد للمجتمع المستهدف: يفضل استخدام العينات العشوائية الطبقية، ومراعاة عدم تحيز العينة والتقليل من أخطاء المعاينات.
- حداثة المعايير: فمعايير الاختبارات تتغير بتغير خصائص الأفراد بمرور الزمن، خاصة إذا كانت اختبارات تقيس خصائص تتأثر تأثرا ملحوظا بما يطرأ على البيئة من تطورات مثلا: المناهج الدراسية، إذ ينبغي تحديث المعايير بصفة مستمرة وفقا لما يطرأ على خصائص الجماعة المرجعية من تغيير.

  ويشير (صلاح الدين علام، 2000) إلى أنه لا ينبغي التحقق من حداثة المعايير بالنظر فقط إلى تاريخ نشر الاختبار، إذ يمكن أن تصدر طبعات جديدة من الاختبار نفسه دون إعادة النظر في المعايير.

- ملاءمة المعايير لمن سيطبق عليهم الاختبار: يجب أن تكون خصائص الأفراد الذين سيطبق عليهم الاختبار متفقة مع خصائص الجماعة المرجعية التي اشتقت منها المعايير مثلا: طلبة الجامعة، موظفين، مراهقين...وفي حالة استخدام أداة القياس لأكثر من غرض عندها ينبغي أن تستمد معاييرها من أكثر من جماعة مرجعية. (موسى النبهان، 2013)

1-3- أنواع المعايير:

يصنفها (صلاح الدين علام، 2000) إلى نوعين: معايير الارتقاء أو المعايير الارتقائية أو المعايير النظرية حسب (عبد العزيز بوسالم، 2014) أو المعايير الطولية حسب (صلاح مراد وأمين سليمان، 2005) وهي التي تتعلق بالعمر الزمني _معايير العمر_، معايير الصف أو الفرق الدراسية أو المستوى الدراسي، معايير العمر العقلي، نسب الذكاء الانحرافية...

النوع الثاني هي المعايير الاحصائية أو ما تعرف بالمعايير المستعرضة مثل الميئنيات، والدرجة المعيارية (Z)، والدرجة المعيارية المعدلة (T)

2- محكات الأداء وهي تخص الاختبارات محكية المرجع:

وهي تقتصر عادة على الاختبارات التحصيلية، ويعتمد استعمالها على تحديد مسبق لمستوى الأداء المقبول من المفحوص، ونقطة الفصل في تحديد هذا المستوى هي التخطيط المسبق لما يجب ألا ينخفض عنه الأداء (المحك)، وعلى ذلك فإن قدرة جميع المفحوصين أو نسبة منهم على بلوغ هذا المستوى أو تجاوزه، أو عدم بلوغ أي منهم له، لا يدخل كعامل محدد لمستوى الأداء المطلوب بوصفه أداء مناسبا، ويعني هذا أننا إزاء مفهوم لا يعتمد على عوامل متغيرة تتعلق بمستوى الصعوبة أو مستوى الأداء أو طبيعة العينة.(صلاح الدين علام، 2000)

فإذا كان لدينا اختبار تحصيلي في مادة القياس النفسي، يتكون من 40 بندا وكان المحك المقبول للأداء هو تمكن الطالب من تقديم 20 إجابة صحيحة على الأقل حتى يعتبر ناجحا.

نلاحظ أن تحديد المحك (مستوى الأداء) تم مسبقا قبل تطبيق الاختبار، وقبل معرفة كم من المفحوصين يمكنهم تقديم هذا العدد من الإجابات أو أكثر.

فالأمر الهام في استعمال هذا النوع من الاختبارات هو تحديد الدرجة المحكية (درجة القطع) للنجاح أو الفشل الذي تحكمه اعتبارات تتعلق بالأهداف المطلوب الوصول إليها مثل تحديد الفاحص لـ 12 درجة أو 60 بالمئة من سقف اختباري يتكون من 20 بند كمحك للنجاح في المادة.

توجد مصطلحات متداولة في هذا المجال أهمها:

***الاختبار محكي المرجع:** هو الاختبار الذي يستخدم لموازنة أداء فرد بمستوى أداء مطلق أو مستوى كفاية معين.

ويرى (ابراهيم محاسنة، 2013، ص177) بأن الاختبارات محكية المرجع لا تعتمد عند تفسيرها لدرجة الطالب على خصائص مجموعته التي طبق عليها الاختبار، ولكنها تعتمد على مقارنة أداء الطالب بمحك أو مستويات أداء متوقعة ومحددة مسبقا، ويتم تحديد هذه المستويات في ضوء الأهداف السلوكية (الإجرائية) المراد قياسها، وتقيس هذه الاختبارات مدى تحقيق الطالب لهذه الأهداف السلوكية، وبذلك يمكن التحقق من كفاءة كل طالب من اكتسابه للمهارات والمعارف المرجوة، ومن ثم تشخيص نواحي الضعف والقوة.

***النطاق السلوكي:** يقصد به مجموعة المعارف والمهارات التي يجب أن يتمكن منها الطالب والتي يقيسها الاختبار، ولتحديده يمكن الاستعانة بمجموعة من الخبراء في المجال الذي يقيسه الاختبار لتحديد الكفايات الرئيسية التي يرون أهمية تحقيقها لدى المتعلمين والتي يمكن تحليلها إلى مكونات أو نواتج للقياس.(صلاح الدين علام، 2000)

***درجة القطع:** هي الدرجة التي تفصل بين المتمكنين وغير المتمكنين، أو بين مستويات مختلفة من التمكن على متصل درجات النطاق السلوكي المشكل للاختبار، وتعرف أيضا على أنها النسبة المئوية لعدد مفردات الاختبار التي يجب أن يعرف الطالب بالفعل إجابتها الصحيحة، وتتعدد الطرق والأساليب المستخدمة في تحديد درجة القطع إلا أنه يمكن تصنيفها في 3 مجموعات رئيسية هي:

-**طرق تعتمد على التحكيم:** وهي تعتمد اعتمادا كليا على أحكام الخبراء، وتتعلق هذه الأحكام بمحتوى الاختبار ومفرداته دون الاسترشاد ببيانات تجريبية مستمدة من التطبيق الميداني لذلك يطلق عليها مصطلح الطرق المطلقة أو الطرق العقلانية منها طريقة نيدلسكي Nedlsky، طريقة إيبل Ebel، الأكثر استخداما طريقة أنجوف Angoff

-**طرق تعتمد جزئيا على التحكيم وتسترشد بالبيانات التجريبية:** وتعتمد هذه الطرق على أحكام الخبراء مع تقديم بيانات تتعلق بالأداء الفعلي للمختبرين لكي يسترشد بها الخبراء في أحكامهم، من خلال بيانات تجريبية تشتمل على معاملات السهولة/ الصعوبة ومعاملات التمييز، وتسمى هذه الطرق بالطرق النسبية لأن أحكام الخبراء أكثر واقعية نذكر منها طريقة أنجوف المعدلة.

-**طرق تعتمد على البيانات التجريبية وتسترشد بالتحكيم:** تعتمد هذه المجموعة اعتمادا أساسيا على البيانات التجريبية المستمدة من تطبيق الاختبار على عينة الأفراد، وتحليل هذه البيانات

باستخدام الأساليب الاحصائية، فمحكات تصنيف الأفراد يتحدد استنادا على أحكام الخبراء في ضوء الأداء الفعلي. (سوسن مجيد، 2014)

***التمكن أو الاتقان:** قدرة الطالب على الحصول على درجة أعلى من أو تساوي درجة القطع المحددة مسبقا للاختبار، فإذا كانت درجة الطالب في الاختبار أعلى من أو تساوي درجة القطع أعتبر متمكنا أو متقنا، وإذا كانت درجة الطالب أقل من درجة القطع اعتبر غير متمكن.

المحاضرة 12: خطوات بناء اختبار تحصيلي

اهتم علماء القياس والتقويم التربوي بالتوصل إلى أساليب متعددة لقياس مدى تحقق الأهداف التعليمية وفعالية البرامج التربوية المقدمة، وكذا لقياس مستوى التحصيل الدراسي للطالب والحكم عليه في ضوء أسس معينة، لذلك فقد ظهرت أنماط مختلفة في بناء الاختبارات التحصيلية وفق الغرض المطلوب.

لذلك نرى أن أول خطوة في بناء الاختبار التحصيلي هي:

1- تحديد الغرض من الاختبار:

يشير (امطاينوس ميخائيل، 2016، ص128) أن عملية بناء الاختبار التحصيلي يجب أن تأخذ بالحسبان استعماله والغرض الخاص منه، فإذا كان الغرض من الاختبار هو تقويم تقدم التلاميذ نحو الأهداف التعليمية المرسومة فلابد أن يوضع الاختبار بحيث يلبي هذا الغرض، ويكشف عن إنجاز التلاميذ ومدى تحقيقهم لتلك الأهداف، وإذا كان الغرض منه هو الكشف عن صعوبات التعلم فلابد أن يتمتع بقدرة تشخيصية عالية، ويكون محدودا في مجاله، كما لابد أن تميل بنوده إلى السهولة وتوضع في ضوء مصادر الخطأ الأكثر شيوعا لدى التلاميذ، وإذا كان الغرض من الاختبار هو قياس التمكن أو الإتقان لوحدة دراسية ما فلا بد أن يختص بهذه الوحدة دون غيرها، ويرتكز على محك للتمكن أو الإتقان يتم في ضوئه تقرير ما إذا كان الطالب أتقن هذه الوحدة أو لا.

وعلى هذا الأساس يحدد نوع الاختبار معياري المرجع، أو محكي.

- **الاختبارات معيارية المرجع:** يكون الغرض منها هو تحديد مقدار المعلومات والمهارات التي سبق أن حصلها الطالب في موضوع معين، وتفسر درجة الطالب في ضوء متوسط أداء الجماعة التي ينتمي إليها، وقد وجه إلى هذا النوع من الاختبارات بأنها لا تقيس الأهداف التعليمية التي يسعى المعلم إلى تحقيقها، كما أنها لا تفيد في تشخيص جوانب القوة والضعف في التحصيل الدراسي للطالب.
- **الاختبارات محكية المرجع:** يكون الغرض منها هو التعرف على مستوى التمكن أو الإتقان للأداء، وبالتالي فإن درجة الطالب تفسر في ضوء محك سابق التحديد، وليس إلى متوسط أداء الجماعة التي ينتمي إليها، وهو بالتالي يحدد مستويات الأداء المطلوب تحقيقها مثل: أن يجيب الطالب على 90% من الأسئلة المقدمة له إجابة صحيحة، أو أن يكتب رسالة لصديقه مكونة من خمسة أسطر بدون أخطاء املائية. (صلاح مراد وأمين سليمان، 2005، ص141)

2- تحديد الأهداف التعليمية سلوكيا (إجرائيا):

تصاغ الأهداف التعليمية بعبارات سلوكية واضحة لتعبر عن التغير المرغوب والمتوقع حدوثه في سلوك المتعلم، وبشكل يمكن ملاحظته وقياسه أثناء أو بعد عملية التعلم. (تيسير كوافحة، 2010)

حيث يتطلب تحديد الأهداف التعليمية سلوكيا أو إجرائيا استخدام أفعال سلوكية تحدد السلوك النهائي (الأداء الناتج) وتتضمن الصياغة شروط حدوث الأداء ومعيار القبول.

تتم صياغة الأهداف التعليمية في صورة سلوكية حسب جرونلاند (Groun_Lund) كما يلي:

فعل (فعل سلوكي في صيغة فاعل الحاضر)		ناتج التعلم (السلوك النهائي الذي يظهره المتعلم بحيث يمكن قياسه)
يرتب	التلميذ	الأرقام من 1 إلى 9 ترتيبا تصاعديا
يميز	التلميذ	بين الأغذية ذات المصدر النباتي والمصدر الحيواني

شروط صياغة الهدف السلوكي:

- أن يصف سلوك المتعلم وليس المعلم.
- أن يصف ناتج التعلم (أي النواتج المباشرة) وليست عملية التعلم.
- أن يكون هدفا بسيطا وليس مركبا، بمعنى أن يتضمن ناتجا تعليميا واحدا وليس مجموعة من النواتج.
- أن يكون الهدف واضحا ومحددا ويمكن قياسه، أي قابلا للملاحظة والقياس.
- أن يكون واقعيا وملائما للزمن المتاح للطالب لمروره بخبرة تعليمية.
- يمكن أن تغطي الأهداف السلوكية الجوانب الثلاث: المعرفية/ العقلية، الانفعالية/ الوجدانية والمهارية/ النفس حركية.

3- تحليل محتوى المقرر الدراسي:

لا شك أن تحليل مكونات المقرر الدراسي قبل القيام بعملية التدريس وعملية التقويم يساعد المعلم في تحقيق الأهداف.

فيتم إعداد قائمة بالموضوعات الرئيسية التي يتضمنها المقرر الدراسي، وكذلك الموضوعات الفرعية المتضمنة داخل كل موضوع رئيسي على حدا.

4- إعداد جدول المواصفات:

يمثل جدول المواصفات حسب (موسى النبهان، 2004، ص76) مخططا لتوزيع فقرات الاختبار التحصيلي في ضوء توزيع أقسام المحتوى السلوكي ومجموعة الأهداف التي يقيسها الاختبار.

ويعرفه (صلاح مراد وأمين سليمان، 2005، ص146) بأنه مخطط تفصيلي ثنائي البعد، أحد أبعاده قائمة الأهداف (نواتج التعلم المرغوب تحقيقها) في جوانبها أو مجالاتها المختلفة (المعرفية/ العقلية، الانفعالية/ الوجدانية والمهارية/ النفس حركية)، والبعد الثاني هو عناصر المحتوى التي يشتملها الاختبار.

ويعرفه (محمد غنيم، 2004، ص 163) بأنه جدول ثنائي الأبعاد تتفاعل فيه البيانات التي توضع أفقيا مع البيانات التي توضع عموديا، حيث تشتمل البيانات الأفقية على موضوعات المحتوى وما يقابلها من أوزان نسبية، بينما تشمل البيانات العمودية الأهداف المحددة وما يقابلها أيضا من أوزان نسبية.

يحدد المعلم (مصمم الاختبار) الأوزان النسبية لموضوعات المحتوى من خلال محكات نذكر منها:

- عدد الصفحات التي يشغلها كل موضوع على حدا بالنسبة لصفحات كتاب المادة ككل. (محك لا ننصح به)
- الزمن المستغرق أثناء عملية تدريس كل موضوع على حدا، والجهد المبذول من طرف المعلم في شرح الموضوع.
- آراء الخبراء في مادة التخصص بشأن أهمية كل موضوع. (صلاح مراد وأمين سليمان، 2005)
- لكن يبقى الرأي الشخصي للمعلم (مصمم الاختبار) وما يراه مناسبا كمادة للاختبار هو الأساس.

أما بالنسبة للأوزان النسبية الخاصة بالأهداف فيذكر (محمد غنيم، 2004) أن المعلم يمكن:

- أن يعتمد على خبرته الشخصية في تحديد وزن كل هدف، أما إذا كان حديث التخرج ولم يكتسب الخبرة الكافية بعد يمكن أن يستعين بزملائه من ذوي الخبرة.
- كما أن بعض موضوعات المحتوى يظهر أنها تتشبع بمستوى معرفي معين مثل التواريخ أو المصطلحات المتخصصة بالتذكر، وبالتالي فإن المعلم يضع لها الوزن النسبي المناسب لهدف في مستوى التذكر، بينما قد تتشبع موضوعات أخرى مثل تحديد الدول الصناعية في خريطة صماء بالهدف المعرفي التطبيق، وهذا يعني أن تحديد الأوزان النسبية للأهداف يرتبط أيضا بطبيعة محتوى المادة في حد ذاتها.

والشكل التالي يوضح جدول مواصفات لاختبار تحصيلي في مقرر بناء وتكييف الاختبارات والروائز النفسية والتربوية للسنة الثانية ماستر تخصص إرشاد وتوجيه.

الجدول رقم (2):جدول مواصفات لاختبار تحصيلي في مقرر بناء وتكييف الاختبارات والروائز النفسية والتربوية للسنة الثانية ماستر تخصص إرشاد وتوجيه.

رقم الموضوع	موضوعات المحتوى	الأهداف %			أوزان المحتوى %
		التذكر	الفهم	التطبيق	
01	مفهوم الاختبار النفسي والمفاهيم المرتبطة به	%5	%5		%10
02	لمحة تاريخية عن ظهور وتطور الاختبارات النفسية	%5			%5
03	تصنيف الاختبارات النفسية	%5	%5	%5	%15
04	خطوات بناء اختبار نفسي	%10	%15	%5	%30
05	خطوات بناء اختبار تحصيلي	%5	%5	%15	%25
06	تكييف الاختبارات		%10	%5	%15
أوزان الأهداف %		%30	%40	%30	%100

بعد أن يتم تحديد طول الاختبار (عدد فقراته) الذي يتحدد غالبا بـ: عمر الطلبة، الزمن المخصص لتطبيق الاختبار، الغرض من الاختبار وشكل الفقرات ونوعها (مقالية، موضوعية...) في هذه الخطوة يتم تحديد عدد فقرات الاختبار التحصيلي أو الأسئلة لكل موضوع وكل هدف حسب النسب المئوية الموضوعة في الجدول.

بافتراض أن عدد فقرات الاختبار هو 20 فقرة، نتحصل في الأخير على الجدول التالي:

الجدول رقم (3): توزيع الفقرات (الأسئلة) لاختبار تحصيلي في مقرر بناء وتكييف الاختبارات والروائز النفسية والتربوية للسنة الثانية ماستر تخصص إرشاد وتوجيه.

رقم الموضوع	موضوعات المحتوى	الأهداف			المحتوى
		التذكر	الفهم	التطبيق	
01	مفهوم الاختبار النفسي والمفاهيم المرتبطة به	1	1		2
02	لمحة تاريخية عن ظهور وتطور الاختبارات النفسية	1			1
03	تصنيف الاختبارات النفسية	1	1	1	3
04	خطوات بناء اختبار نفسي	2	3	1	6
05	خطوات بناء اختبار تحصيلي	1	1	3	5
06	تكييف الاختبارات		2	1	3
مجموع الفقرات (الأسئلة)		6	8	6	20

فوائد إعداد جدول المواصفات:

يمكن تلخيص أهمها فيما يلي:

- توزيع أسئلة الاختبار على مختلف أجزاء المقرر الدراسي بحيث يعطى كل جزء منه وزنه النسبي أي أهميته النسبية عند صياغة الأسئلة.
- عدم تركيز الأسئلة لقياس مستوى محدد داخل المجال كالمجال المعرفي مثلا، بل يؤخذ بعين الاعتبار كل المستويات (المعرفية، الوجدانية، النفس حركية)
- تشجيع الطلبة على الفهم وليس الحفظ بتوقعهم أن الامتحان يشمل أسئلة متنوعة من مختلف جوانب المادة، والتي بدورها قد تعتمد على التطبيق أو الفهم أو الإبداع وليس التذكر فقط.
- توزيع الزمن على الموضوعات وأهميتها، وبذلك يعطي المعلم الوزن الحقيقي لكل جزء من المادة. (صلاح مراد وأمين سليمان، 2005)

5- صياغة فقرات (أسئلة) الاختبار التحصيلي:

إن اختيار شكل الفقرات (الأسئلة) يتحدد بطبيعة الأهداف ونواتج التعلم المقيسة، ومن المعلوم أن كل نوع أو شكل من أشكال البنود الاختبارية قد يكون أكثر ملاءمة لقياس نواتج تعلمية معينة، وأقل ملاءمة لقياس نواتج أخرى، فالاختبار المقالي أكثر فعالية من الاختبار الموضوعي في قياس القدرة على

الربط وتنظيم الأفكار، والاختبار الموضوعي من نوع التكميل أو ملء الفراغ أكثر ملاءمة لقياس القدرة على استدعاء الحقائق، واختبار الاختيار من متعدد يصلح لقياس الكثير من نواتج التعلم ومستوياته الدنيا والعليا.(امطاينوس ميخائيل، 2016، ص137)

6- تحليل فقرات الاختبار التحصيلي:

بعد الانتهاء من تطبيق الاختبار يبدأ المعلم في عملية التصحيح وتقدير الدرجات، ومهما كانت نوعية الأسئلة فعلى المعلم أن يراعي الدقة في تقدير الدرجات والموضوعية والتي تعني استقلالية الدرجة الموضوعة عن ذاتية المصحح، وحتى يكون التصحيح موضوعيا، يجب أن تكون إجراءات التصحيح محددة بطريقة معينة حيث يضع المعلم مفتاحا للتصحيح.

ومن الضروري أن يقوم مصمم الاختبار بإجراء تحليل لفقرات اختباره لمعرفة مستوى سهولتها/ صعوبتها وقدرتها على التمييز.

7- حساب الخصائص السيكومترية للاختبار: من ثبات وصدق.

المحاضرة 13: تكييف الاختبارات النفسية

1- مفهوم تكييف اختبار نفسي:

يتضمن تكييف الاختبار النفسي تطبيق الاختبارات المصممة في مجتمع أصلي ما على مجتمع آخر جديد بشرط أن تتم احترام وتقدير التغييرات اللازمة التي تجعله يتوافق مع ثقافة الوسط الذي سيطبق فيه من جديد، وبهذا فإن عملية تكييف الاختبارات لا تعني فقط ترجمة محتواها في رمز لغوي جديد، ولكن الأمر يتعلق بإبداع جديد، يضم مختلف عمليات التحرير والتعبير والإبدال والإضافات، يضاف إليها مختلف التبريرات النظرية والعمليات الاحصائية والتطبيقية. (مراد نعموني، 2014، ص01)

ويشير مفهوم تكييف الاختبارات النفسية إلى كل الإجراءات التي يتبعها الباحث بدءا من تقديره عما إذا كان باستطاعة الاختبار تقدير التركيبة نفسها عند نقله من ثقافة إلى أخرى، وصولا إلى محاولته الحصول على مفاهيم، مفردات وتعابير متعادلة ثقافيا، لغويا، ونفسيا مع الثقافة الجديدة للاختبار، فتكييف اختبار يأخذ أبعادا أكثر من ترجمة محتوياته من لغة إلى أخرى ليشمل جملة من التعديلات المنطقية المدروسة والمرحلية، والتي تحتاج إلى أدلة علمية لتؤكد أن الاختبار بصورته الحالية صالح للتطبيق، ونتائجه تنطبق على العينة الجديدة وفق خصائصها الثقافية. (عبد العزيز بوسالم، 2017، ص189)

2- إجراءات تكييف الاختبارات النفسية:

توجد مجموعة من الإجراءات لتكييف اختبار نفسي ونقله من بيئته الأصلية إلى البيئة المحلية، يمكن تلخيصها في الخطوات التالية:

2-1- الحصول على نسخة أصلية من الاختبار المراد ترجمته:

تتضمن جميع التفاصيل التي تتعلق بالاختبار مثل الغرض من تطبيق الاختبار، التعليمات، الفئة العمرية التي يصلح لها الاختبار، بنود الاختبار، طريقة التصحيح، اجراءات بنائه وتقنينه من صدق وثبات ومعايير لتفسير النتائج.

2-2- ترجمة الاختبار من لغته الأصلية إلى اللغة المحلية:

يرى هاملتون (Hambleton, 1999) أن الترجمة العلمية للاختبار النفسي هي عملية الوصول إلى مفاهيم، ومفردات وتعابير متعادلة ثقافيا نفسيا ولغويا للغة وثقافة أخرى، انطلاقا من لغة وثقافة معينة.

على أن تكون الترجمة ثقافية وليست حرفية لتجنب التحيز الثقافي، حيث من أسباب التحيز الثقافي هو الترجمة السيئة للبند، واختلاف المعاني الضمنية للكلمات، فعندما نترجم كلمات قد نركز على المعنى الحرفي لها، وحتى إذا ترجمنا المعنى لا بد أن يكون المعنى متضمنا في الثقافة التي نترجم إليها الاختبار، فما يقيسه البند من مؤشرات قد لا يتضمن المعنى نفسه في ثقافات مختلفة. (عبد العزيز بوسالم، 2017)

وفي سياق مرتبط يرى (مراد نعموني، 2014) أن دارسي الترجمة وممارسيها يجمعون على أن من أعظم مشاكل الترجمة هي عجز المترجم في توصيل المعنى الدقيق لأي مفردة في النص الذي يريد نقله إلى لغة أخرى، وترجع هذه المشكلة إلى عوامل أهمها: أن كل لغة لا بد وأنها تنتمي غلى ثقافة معينة، وبالتالي فإن المترجم قد ينقل الكلمة إلى لغى أخرى ولكنه لن يستطيع أن ينقل ثقافة هذه الكلمة بشكل فعال، بحيث ينقل تصور صاحب الكلمة الأصلية إلى اللغة المستهدفة في الترجمة، وقد تؤدي تلك الاختلافات اللغوية أو حتى في اللهجات إلى مشكلات كبيرة.

هذا بالإضافة إلى أن لكل لغة طابع خاص في تركيب الجملة وترتيب مفرداتها (قواعد اللغة)، كما أن كل لغة تحمل في طياتها العديد من المرادفات التي تختلف في معانيها اختلافا طفيفا عن بعضها البعض.

فعلى مترجم الاختبار أن يأخذ كل ذلك بعين الاعتبار.

2-3- التأكد من صلاحية التكييف الثقافي للاختبارات النفسية:

من خلال التأكد من الخصائص السيكومترية للاختبار في البيئة المحلية وهي الصدق والثبات، بنفس الطرق التي تطرقنا إليها سابقا في بناء الاختبار النفسي والتربوي، ما يميز تكييف الاختبار هو ما يعرف **بصدق الترجمة.**

أساليب تقدير صدق ترجمة الاختبارات المكيفة:

يرى (بشير معمرية، 2012) أنه توجد بعض الأساليب لتقدير صدق ترجمة الاختبارات التي تنقل من اللغة الأصلية إلى لغة أخرى منها:

***الترجمة وإعادة الترجمة:** أو ما يعرف بالترجمة العكسية، أو الترجمة الراجعة، وفيها يتم إعداد ترجمة للاختبار من اللغة الأصلية إلى اللغة المحلية، ثم تعاد ترجمته مرة أخرى من هذه اللغة إلى لغته الأصلية، وتعرض الترجمتان مرة أخرى على خبراء في اللغتين لتقييمهما، وهي تسمى بالطريقة القياسية لأنها الأكثر استخداما.

***مدخل الفريق:** حيث يقوم أكثر من مترجم بترجمة الاختبار نفسه بصورة مستقلة، ثم يجتمع هؤلاء لكي يكاملوا بين أعمالهم في صورة نهائية للاختبار من خلال مناقشتهم للترجمات.

***مراجعة العبارات المترجمة:** حيث تعرض الترجمة على مجموعة من الخبراء يتقنون اللغتين لوضع ملاحظاتهم على العبارات غير المناسبة واقتراح صياغة بديلة.

***طريقة أخرى:** بعد أن تتم ترجمة الاختبار تقدم النسختان الأصلية والمترجمة إلى عينة من الأفراد يتقنون اللغتين للإجابة عليهما، ثم يحسب معامل الارتباط بينهما، كلما ارتفعت قيمة معامل الارتباط دل ذلك على صدق الاختبار وفقا لهذه الطريقة. مثلا إذا كان الاختبار في نسخته الأصلية باللغة الانجليزية، ومطلوب نقله إلى اللغة العربية وموجه للراشدين يتم تطبيق هذه الطريقة على عينة من طلبة الجامعة بالجزائر تخصص لغة انجليزية.

2-4- استخراج معايير جديدة لتفسير نتائج الاختبار: تتوافق مع خصائص البيئة المحلية.