



وزارة التعليم العالي والبحث العلمي
جامعة الشهيد الشيخ العربي التبسي- تبسة
كلية العلوم الإنسانية والاجتماعية
قسم علم النفس



مطبوعة بيداغوجية في مقياس:

بناء وتكليف الاختبارات والروائز النفسية والتربوية

موجهة لطلبة السنة الثانية ماستر الإرشاد والتوجيه

إعداد الدكتور

زياد رشيد

أستاذ محاضر - ب -

السنة الجامعية: 2022-2023

تقديم: مطبوعة بيداغوجية محكمة موجهة لطلبة السنة الثانية ماستر شعبة علوم التربية تخصص الإرشاد والتوجيه، المطبوعة عبارة عن مجموعة من المحاضرات المنسجمة وفق محاور مقياس بناء وتكيف الاختبارات والروائز النفسية والتربوية، بحسب البرنامج الرسمي المعتمد من طرف وزارة التعليم العالي والبحث العلمي.

عنوان الماستر: علوم التربية تخصص الإرشاد والتوجيه

السداسي الثالث

اسم الوحدة: المنهجية.

اسم المادة: بناء وتكيف الاختبارات والروائز النفسية التربوية.

الرصيد: 3

المعامل: 2

أهداف التعليم:

- يتمكن للطلاب من خلال هذا الموضوع من بناء الاختبارات النفسية والتحكم في طرق تكيفها وتوظيفها في عملية التوجيه والإرشاد.

- يستطيع الطالب من خلال هذا الموضوع أن يقوم بعملية بناء الاختبارات والروائز النفسية كما يمكن له تكيفها مع المحيط المحلي في مجال الإرشاد والتوجيه.

المعارف المسبقة المطلوبة:

محتوى المادة:

- شروط بناء الاختبارات النفسية.

- شروط بناء الروائز النفسية.

- تقنيات تكيفها مع المحيط المحلي.

العمل الشخصي:

- اختيار بعض المقاييس وتطبيق اجراءات التكيف.

- بناء نموذج لرائز نفسي تربوي.

طريقة التقييم: متواصل، امتحان.

تأشيرة رئيس القسم

الدكتور: زياد رشيد

فهرس المحتويات

04 مقدمة

المحاضرة الأولى: الاختبارات النفسية والتربوية

05 تمهيد

05 أولاً: ما الاختبار النفسي أو التربوي

08 ثانياً: تعريف الاختبار النفسي والتربوي

08 ثالثاً: أهداف الاختبارات النفسية والتربوية

المحاضرة الثانية: الخصائص السيكومترية للمقاييس والاختبارات النفسية والتربوية

10 تمهيد

10 أولاً: خصائص الاختبارات والمقاييس الجيدة

10 ثانياً- الشروط العامة للاختبارات النفسية والتربوية

11 ثالثاً- الشروط التجريبية للاختبارات النفسية والتربوية

11 1- الصدق Validity

14 2- الثبات Reliability

المحاضرة الثالثة: بناء وتطوير المقاييس والاختبارات النفسية والتربوية

22 تمهيد

22 أولاً: تصنيف الاختبارات والمقاييس النفسية والتربوية واستخداماتها

26 ثانياً: أهمية الاختبارات والمقاييس النفسية

27 ثالثاً: الحاجة إلى بناء اختبار نفسي جديد

28 رابعاً: خطوات بناء الاختبارات والمقاييس النفسية والتربوية

29 خامساً: المبادئ التوجيهية لتطوير الاختبارات أو أدوات القياس

المحاضرة الرابعة: التقنين والمعايير لدرجات الاختبارات النفسية والتربوية

35 تمهيد

35 أولاً: التقنين Standardization

38 ثانياً: المعايير Norms

المحاضرة الخامسة: الصفحة أو البروفايل النفسي والتربوي

50 تمهيد

أولاً: تعريف الصفحة أو البروفايل النفسي أو التربوي.....50

ثانياً: الخطأ المعياري للقياس وتفسير الفروق في الدرجات.....56

المحاضرة السادسة: الأساليب الإسقاطية

تمهيد.....60

أولاً: الاختبارات الإسقاطية.....60

ثانياً: تصنيف الأدوات الإسقاطية.....61

ثالثاً: الفروق الأساسية بين الأدوات الإسقاطية والمقاييس الموضوعية.....62

رابعاً: الاختبارات الإسقاطية الأكثر شيوعاً.....63

خامساً: صدق وثبات الاختبارات الإسقاطية.....69

سادساً: تقييم الأساليب الإسقاطية.....70

سابعاً: ميزات الأساليب الإسقاطية.....71

ثامناً: عيوب الأساليب الإسقاطية.....71

المحاضرة السابعة: ترجمة وتكيف الاختبارات النفسية والتربوية

تمهيد.....73

أولاً: الدليل الإرشادي لترجمة وتكيف الاختبارات.....74

القسم الأول: الشروط الأولية (التمهيدية).....75

القسم الثاني: تطوير أو بناء الاختبار.....77

القسم الثالث: التأكد والتحقق من صدق التحليلات التجريبية.....80

القسم الرابع: تطبيق أو إدارة الاختبار.....88

القسم الخامس: لتسجيل الدرجات والتفسير.....90

القسم السادس: توثيق الاختبار.....91

ملحق رقم (01): قائمة (ITC) لترجمة وتكيف الاختبارات النفسية والتربوية.....93

قائمة المصادر والمراجع.....٩٤

مقدمة:

ظهرت البوادر الأولى للاختبارات والمقاييس النفسية والتربوية الحديثة في أوائل القرن العشرين في فرنسا عام 1905، عندما نشر ألفريد بينيه (Alfred Binet) وزميل له أول اختبار للذكاء مصمماً للمساعدة في تسكين أطفال المدارس في باريس في الفصول الخاصة بهم. انتشر اختبار "ألفريد بينيه" للذكاء بعد عقد من الزمان، ليتم بعدها إعداد نسخة من الاختبار باللغة الإنجليزية للاستخدام في المدارس في الولايات المتحدة الأمريكية.

عندما أعلنت الولايات المتحدة الحرب على ألمانيا ودخلت الحرب العالمية الأولى في عام 1917، احتاج الجيش الأمريكي إلى طريقة سريعة لفحص أعداد كبيرة من المجندين بحثاً عن الاضطرابات العقلية والانفعالية المحتملة في صفوف الجنود، وبالتالي تم الاستعانة باختبار الذكاء "لبينيه". قدمت الاختبارات النفسية بهذه الطريقة كذلك خلال الحرب العالمية الثانية، من خلال اعتماد الجيش الأمريكي بشكل أكبر على الاختبارات النفسية لفحص المجندين للخدمة العسكرية.

بعد الحرب تم تطوير واستخدام المزيد من الاختبارات والمقاييس التي تهدف إلى قياس مجموعة متزايدة من المتغيرات النفسية. كانت هناك عدة اختبارات للقياس ليس فقط لقياس الذكاء ولكن أيضاً لقياس جوانب الشخصية، وكيفية عمل وأداء الدماغ، والأداء في العمل، والعديد من الجوانب الأخرى للأداء النفسي والاجتماعي.

إن تقبل العالم لاختبار "بينيه" للذكاء في أوائل القرن العشرين لم يولد المزيد من الاختبارات فحسب، بل المزيد من مطوري الاختبارات، والمزيد من ناشري الاختبارات، والمزيد من مستخدمي الاختبارات، وظهور ما أصبح يعرف باسم مؤسسات ناشري ومطوري الاختبارات النفسية والتربوية. وبظهور هذه المؤسسات وانتشارها حول العالم ظهرت الحاجة لمزيداً من الاهتمام حول عملية بناء وتصميم الاختبارات والمقاييس النفسية والتربوية، من خلال استخدام تقنيات وأساليب إحصائية وسيكومترية متقدمة من أجل تحقيق أعلى جودة ودقة لأدوات القياس النفسي والتربوي، إلا أن هذه العملية لم تخلوا من مشكلات أهمها مشكلة ترجمة وتكييف الاختبارات والمقاييس ومدى صلاحيتها واستخدامها في سياقات ثقافية مختلفة، وتحقيقها لأهم شرط من الشروط السيكومترية وهو الصدق. يتضمن محتوى المطبوعة البيداغوجية لمقياس بناء وتكييف الاختبارات والروائز النفسية والتربوية المقرر على طلبة السنة الثانية ماستر إرشاد وتوجيه على سبعة محاضرات حول عملية بناء وتصميم الاختبارات النفسية والتربوية وأهم الشروط السيكومترية التي يجب توفرها في هذه الأدوات، وكذلك شرح لخطوات التقنين والمعايرة، والتطرق للصفحة أو البروفايل النفسي والتربوي، وكذلك الأساليب الاسقاطية في القياس، وأخيراً ترجمة وتكييف الاختبارات النفسية والتربوية.

المحاضرة الأولى

الاختبارات النفسية والتربوية Psychological and Educational tests

تمهيد:

يرغب الباحثون في علم النفس وعلوم التربية والمختصون في بعض مجالاته كعلم النفس الإكلينيكي (العيادي) والإرشادي ومجالات الشخصية في جمع معلومات معينة عن الأداء الذهني للشخص كالقدرة اللفظية مثلاً. يحصلون على هذه المعلومات بواسطة أخذ عينة من أداء الشخص باستخدام اختبار نفسي. وفي معظم الأحيان، وليس كلها، تكون الاختبارات كتابية، من نوع اختبارات الورقة والقلم، كاختبارات الذكاء المعتادة. ومن أهم أنواع هذه الاختبارات اختبارات التحصيل التي تقيس مدى إتقان الشخص لمهارة معينة (الاختبارات التي طبقت عليك أثناء دراستك بالمدارس هي اختبارات للتحصيل الدراسي). ومن الأنواع الأخرى اختبارات الاستعداد Aptitude tests. وهي تقيس إمكانيات الشخص وقابليته لتعلم مهارات متنوعة. وهناك أيضاً قوائم الميول Interest Inventories التي تعد للكشف عن ميول الشخص المختلفة، واختبارات الشخصية المصممة لبحث مختلف جوانب الشخصية الإنسانية. لقد أصبحت الاختبارات النفسية في السنوات الأخيرة جزءاً وطرذاً من الطرود الهامة والواقعة تحت تصرف فئة المختصين بالفحص وتصنيف البشر. وما أكثر الذين يتعرضون لعمليات القياس والتقويم لأجل الترقية أو الحصول على وظيفة. (دويدار، 1995، ص 44-46)

أولاً: ما الاختبار النفسي أو التربوي؟

تري "آنستازي" أن الاختبار النفسي بالضرورة هو مقياس موضوعي ومقنن لعينة من السلوك. والاختبارات النفسية تماثل الاختبارات في أي علم آخر (2015، ص 23). وبصرف النظر عن طبيعة الاختبارات النفسية المختلفة فإن لها جميعها قاسماً مشتركاً: تعتمد الاختبارات على عينة صغيرة نسبياً من السلوك، في حين نتائجها تستعمل لإلقاء الضوء على مجتمع، أو أصل سلوكي أكبر وأوسع مجالاً. ومن هذه الزاوية فإن وظيفة الإختبار أقرب ما تكون إلى وظيفة المسح اللهم فيما يلي: إذا استخدمنا اختباراً فإننا ندرس عينة من أشكال سلوك شخص واحد ثم نحاول اشتقاق استنتاجات عن مجتمع بشري أكبر. (دويدار، 1995، ص 44-46)

وحتى يتمخض الإختبار عن نتائج مفيدة فإنه يجب أن تتوفر فيه صفتان أساسيتان: يجب أن يكون ثابتاً (موثوقاً) Reliable ويجب أن يكون صادقاً (صالحاً) Valid.

ويشير مصطلح الثبات Reliability، بدرجة أساسية إلى الاتساق، أي معنى ثبات الاختبار هو اتساق الدرجات التي نحصل عليها من الأشخاص أنفسهم عندما يعاد اختبارهم بالاختبار نفسه أو بصيغة مكافئة للاختبار. (أنستازي، 2015، ص22)

فالثبات أو الموثوقية في أبسط معانيها إلى كفاءة الإختبار في إعطاء نفس النتيجة بطريقة مطردة. فلو وقف شخص ما على ميزان، ووجد أن وزنه يساوي 85 كيلوغرام، ثم نزل وبعد برهة من الزمن وقف على الميزان مرة ثانية ووجد أن وزنه صار هذه المرة 95 كيلوغرام، فمن المحتمل أنك تستنتج أن هناك خللاً بالميزان وأنه غير ثابت (غير موثوق، أو لا يعتمد عليه). وبنفس الكيفية فإن تماثل النتائج التي نحصل عليها من إجراء الإختبار، ومن تكرار إجراءاته تعد من أهم الأمور التي يجب مراعاتها في أي اختبار نفسي. فإذا ما اخترنا نفس الشخص عدة مرات، خلال عدة أسابيع، ووجدنا أن هذا الشخص يحصل على درجات مختلفة في كل مرة، فإننا سنشك بقوة فيما إذا كان الإختبار ثابتاً بدرجة تكفل تقديم معلومات مفيدة وجديرة بالإعتماد عليها. (دويدار، 1995، ص44-46)

أما الصفة الثانية، الصدق validity فالسؤال الأهم بالطبع والذي ينبغي أن يطرح بشأن أي اختبار نفسي يتعلق بصدقه، أي الدرجة التي يقيس بها الاختبار بالفعل ما يهدف لقياسه. ويقدم الصدق تحققاً مباشراً عن درجة جودة الاختبار في أداء وظيفته. (أناستازي، 2015، ص23)

فالصدق إذن، يتعلق بحقيقة مؤداها أن الإختبار يعتمد بالضرورة على عينة صغيرة للسلوك ليس إلا. قد يستغرق اختبار استعداد، مصمم لتحديد ما إذا كانت إمكانيات المتقدم تؤهله لتعلم عمل مكتبي ما، أقل من ساعة. ويحتوي على عدد من المهام التي تستعمل فيها الورقة والقلم، وذات العلاقة بأدوار الموظف. ولا يمكن، في حدود الوقت المتاح، أن تعرض على المفحوص كل النشاطات التي يقوم بها الموظف المكتبي في الحياة الواقعية. وبالمثل فإن اختبارات الورقة والقلم لا يمكن لها أن تعكس تماماً كل مهام الحياة الواقعية المتباينة، والتي يمكن أن يضطلع بها الموظف أثناء أدائه لعمله الفعلي. وعلى هذا الأساس، تصبح قضية الصدق متلبورة فيما إذا كان اختبار الاستعداد يقيس بالفعل الإمكانيات المطلوبة للعمل المكتبي. صدق الإختبار يناظر وظيفياً الصدق الخارجي للتجربة. فالقضية المشتركة في الحالين تتلخص في إمكانية التعميم من عينتنا المحدودة من السلوك أو الناس، على التوالي، إلى مجتمع أوسع من السلوك أو الناس. (دويدار، 1995، ص44-46)

توجد أنواع مختلفة من صدق الاختبارات، ولكنها تدور كلها حول سؤال جوهرية، وهو: " ما هو الشيء الذي وضع الإختبار لقياسه؟". والسؤال المماثل الذي يتعلق بثبات الإختبار هو: " ما مدى الدقة التي يقيس بها الإختبار ذلك الشيء الذي وضع لقياسه؟" ويجب مقارنة نتائج الإختبار بمستوى أو محك

ما لتقرير ما إذا كان هذا الاختبار صادقاً أم لا. يميز المختصون بين أنواع الصدق المختلفة طبقاً لنوع المحك أو المعيار المستخدم لتقويم الإختبار. (دويدار، 1995، ص44-46)

1- مصطلحات ومفاهيم الرائز، الاختبار، المقياس:

يرد استخدام مصطلح اختبار محل رائز وهو استعمال شائع في الكتابات والاستعمالات العربية القديمة لا سيما منها المصرية ذلك أن استعمال رائز قد أصبح في المشرق والمغرب العربيين مستعملاً بكثرة وإن ليس بصفة مطلقة، وبقينا نحن في الجزائر رهناً للماضي العربي. والغريب أن مصطلح رائز الذي يعتبره البعض غريباً ويصر على استعمال مصطلح اختبار بدله. فالحقيقة أن هذا المصطلح يستخدم كثيراً في جنوبنا الكبير، فأبناء هذه المناطق مثلاً عندما يريد أحدهم أن يشتري لباساً ويكون بحاجة لمعرفة إذا ما كان يلائمه من ناحية القامة (Taille) يطلب من التاجر أن يسمح له بقياسه قائلاً: "أريد أن أرزوه". أي أن أقيسه باستعمال فعل "راز، يروز، Tester" ولا يقول أبداً أختبره لأنه بصدد قياس كمي لا كفي، بعبارة أخرى فإن استعمال كلمة الـ *Testage* بالانجليزية أو *Testage* بالفرنسية، هنا مقصود منها الأبعاد (القياس) وليس نوعية اللباس.

وهناك فرق بين الرائز والاختبار في المفهوم، ففي استعمال الرائز ثمة اهتمام بالنتائج وهو عمل خفيف على الفاحص والمفحوص مقارنة باستعمال الاختبار عند "بياجيه" الذي يبذل فيه جهد إضافي من قبل الفاحص وهو تتبع الفاحص لإجابات الطفل ومن قبل الطفل حيث هو مطالب بالاستجابة للمحاجات *les contre argumentations* المقدمة له من قبل الفاحص والتي عليه أن يبرهن على أقواله أو أفعاله إثباتاً أو نفيًا. (قدوري، 2018، ص13)

والأمر لا يتعلق بهذين المصطلحين فقط وإنما هناك لبس آخر بخصوص مصطلح *Echelle* بالفرنسية أو *Scale* بالانجليزية. إذ أن الاستعمال السائد هو مقياس في العربية بدل سلم، ذلك أن المقياس، في رأيي، كلمة عامة ولا يمكن اعتبارها مصطلحاً لأنها تعبر عن أي أداة قياس، كما هو معروف في قواعد اللغة العربية من أوزان لأسم الآلة. فمن فعل قاس نأخذ اشتقاقاً "مقياس على وزن مفعال"، حيث تقول العرب: مسبار من سبر، ومحراث من حرث... الخ. ومن ثم فإن الرائز مقياس والاختبار مقياس والسلم مقياس و القائمة مقياس... وكلها أدوات قياس. "إذ يقصد بمقياس أحياناً السلم وأحياناً الرائز وأخرى الاختبار أو حتى الاستبيان بل في أحيان كثيرة المادة المدرسة أو الوحدة. ويتضح من أوجه الاستعمال هذه المختلفة مدى اللبس وعدم الدقة في المصطلح لا سيما أن المعالجات للبيانات المحصل عليها تختلف تبعاً للأداة ومستوى القياس الذي تستعمل فيه الأداة". (قدوري، 2018، ص13)

وعلى ذلك فإنه بالرغم من أن التداخل بين هذين المفهومين ليس كاملاً إلا أنه لا يزال في إمكاننا أن نقول أن معظم الاختبارات هي وسائل قياس، وأن معظم المقاييس النفسية يمكن أن تستخدم كاختبارات. (ليون، أ، تايلر، 1998، ص 47-48)

ثانياً: تعريف الاختبار النفسي والتربوي:

إن تعدد السمات واختلاف طبيعتها وتباين أنواعها يتطلب أن تتباين أدوات قياس أو تقدير هذه السمات، لذلك ربما لا نجد تعريفاً مناسباً ينطبق على جميع هذه الأدوات. الاختبار النفسي عموماً هو مقياس في علم النفس، وهو عبارة عن مجموعة منظمة من المثيرات أعدت لتقيس بطريقة كيفية أو كمية، بعض العمليات العقلية أو سمات معينة في الشخصية أو دراسة الشخصية ككل، بمختلف جوانبها الدينامية. فمثلاً يعرف (Sax, 1974) الاختبار النفسي والتربوي بأنه مطلب أو مجموعة من المطالب تستخدم للحصول على ملاحظات منظمة يفترض أنها تمثل سمة نفسية أو تربوية. ويعرفه (Nitko, 1983) بأنه إجراء منظم لملاحظة ووصف سمة أو أكثر من سمات الفرد بالاستعانة بميزان قياس أو نظام تصنيف معين. أما (Brown, 1976) فيعرفه بأنه إجراء منظم لقياس عينة من السلوك. وبنفس التعريف تقريباً عرفته (Anastasi, 1976) بأنه مقياس موضوعي ومقنن لعينة من السلوك. وعرفه (Cronbach, 1970) بأنه طريقة منظمة لمقارنة سلوك شخصين أو أكثر، أو هو عبارة أخرى عينة مقننة من السلوك تستخدم في المقارنة بين فردين أو أكثر. في حين قدم (chase, 1985) تعريفاً أكثر عمومية لاشتماله على المكونات الأساسية المشتركة بين أدوات القياس. فقد اقترح (chase) أن الاختبار النفسي والتربوي هو أداة قياس مقننة أو أسلوب منظم يصمم للحصول على قياس موضوعي لعينة من السلوك بهدف موازنة أداء الفرد بمعياري أو مستوى أداء محدد.

وبذلك أصبحت عملية القياس النفسي والتربوي نوعاً من الاختبار يخضع له الفرد، وتكون نتيجته تقويمياً لقدراته العقلية أو تقويمياً لأوجه شخصيته المختلفة. غير أن هذه الاختبارات ترتبط بشروط معينة بالنسبة لتكوين إطارها الأساسي، وبالنسبة لطريقة استخدامها، وبالنسبة لأهدافها، هذه الشروط تعرف بالتقنيات. (علام، 2000، ص 28)

ثالثاً: أهداف الاختبارات النفسية والتربوية:

تعتبر الاختبارات بأنواعها المختلفة والمتنوعة من أهم الأدوات المستخدمة في علم النفس للتوصل إلى فهم أعمق ومعرفة أكثر شمولاً لشخصية الفرد، من حيث قدراته العقلية ومشكلاته الانفعالية. ويمكن تحديد دور الاختبارات في الأهداف التالية:

- تقييم قدرات الفرد وإمكانياته من حيث ذكائه العام وقدراته العقلية الخاصة.

- كشف الجوانب المختلفة من شخصية الفرد ومشاعره وأفكاره ورغباته واتجاهاته (اختبارات الشخصية).
- تقييم ديناميات السلوك لدى الفرد وكشف الدوافع الواعية واللاواعية التي تحرك هذا السلوك، أي كشف أنواع الصراعات أو المازم النفسية (الإسقاطية).
- تشخيص الاضطراب أو المرض النفسي أو العقلي الذي يعاني منه الفرد. فالتشخيص Diagnostic هو التقييم العلمي الشامل لحالة الفرد، ومن حيث قدراته وإمكانياته المشكلات التي يواجهها ويعاني منها. وعلى هذا الأساس يبرئ التشخيص الطرق والأهداف لتخطيط برنامج علاجي شامل كما يتضمن التشخيص التنبؤ والافتراضات المتوقعة للحالة المعنية. (فيصل، 1996، ص18)

المحاضرة الثانية:

الخصائص السيكومترية للمقاييس والاختبارات النفسية والتربوية

Psychometric Characteristics of Psychological and Educational Scales and Testings

تمهيد:

تعتمد جودة ودقة القياس النفسي والتربوي على جودة أدواته، ولتحقيق هذه الجودة من الضروري توفر شروط أو خصائص سيكومترية. وتشير هذه الخصائص إلى مفهومي من المفاهيم الأساسية التي تتعلق بالاختبارات والمقاييس التربوية والنفسية هما مفهوم ثبات درجات الاختبار Test Reliability، وصدق الاختبار Test Validity ونظراً لأهمية هذين المفهومين، فإن المحاضرة ستناقش مفهومي الصدق والثبات في الاختبارات مرجعية الجماعة أو المعيار، وسوف نتناول الطرق المختلفة لتقدير كل من المفهومين، وكيفية تفسيرهما والعوامل المؤثرة في كليهما، وعلاقة كل منهما بالآخر.

أولاً: خصائص الاختبارات والمقاييس الجيدة:

عندما نستخدم اختباراً من أجل الحصول على معلومات تساعدنا في اتخاذ قرار ما، فإننا نواجه مشكلة أساسية تتعلق باختيارنا للاختبار الذي يمكن أن يفيدنا حقيقة في اتخاذ القرار، قد يكون هناك عدة بدائل معروضة للاختبارات يمكن استخدامها لأغراض القرار، ولكن السؤال: أي هذه الاختبارات هو الأفضل في تقديم معلومات مفيدة. وما هي الأسس التي نستند إليها في الاختبار؟.

هناك أمور كثيرة يمكن أن تؤخذ بعين الاعتبار عند تقويم جودة اختبار ما، ولكننا سنجملها في ثلاث مجموعات وهي: (أبو هاشم، 2006، ص02)

المجموعة الأولى: الشروط العامة للاختبار النفسي وتتضمن: الشمول - التقنين - الموضوعية.

المجموعة الثانية: الشروط التجريبية للاختبار النفسي وتتضمن: الثبات - الصدق - المعايير.

المجموعة الثالثة: وهي بعض الاعتبارات العملية والقابلية لاستخدام الاختبار.

ثانياً- الشروط العامة للاختبارات النفسية والتربوية:

1-1- الشمول: وهو تمثيل وشمول الاختبار لجميع مكونات القدرة أو السمة التي يقيسها.

1-2- التقنين: وهو أن يكون تفسير نتائج الاختبار بنفس الطريقة حتى وإن اختلف الفاحصون وتعدد المحققين وتنوعت مواقف الاختبار وتكررت مرات القياس.

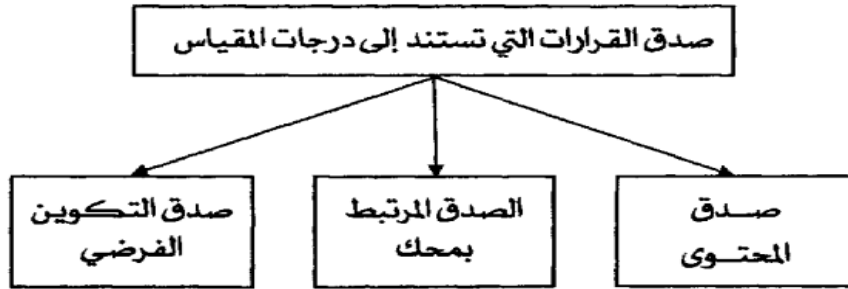
3-1- الموضوعية: ونعني بها ضمان عدم التحيز في عملية تطبيق الاختبار وعملية التصحيح، والشيء الموضوعي هو الشيء القائم على حقائق خارجية وليس ذاتية، بحيث يمكن للأخريين التحقق منها.

ثالثاً- الشروط التجريبية للاختبارات النفسية والتربوية:

1- الصدق Validity:

يعرف صدق الاختبار بأنه المدى الذي يقيس به الاختبار ما وضع من أجل قياسه ويكون بالتالي صدق المفردة هو أن يقيس الهدف الذي وضع من أجل قياسه.

ومن الضروري جمع معلومات عن صدق الاختبار بشكله النهائي للتأكد من صلاحيته وإمكانية تطبيقه، وهذا يعني أن تطبيق الاختبار في عدة حالات، وعلى مجموعات مختلفة قبل الاستخدام النهائي للاختبار، وكذلك معرفة الخصائص التي يقيسها الاختبار أو بماذا يتنبأ. (الناشف، 2001، ص38) وهناك أنواع مختلفة من الصدق والتي يمكن تلخيصها في ثلاثة أنواع رئيسية على الشكل التالي:



شكل (01) يوضح طرق التحقق من الصدق. (علام، 2000، ص189)

1-1- صدق المحكمين أو الخبراء:

1-1-1- صدق الظاهري (السطحي) Face Validity:

ويعني البحث عما يبدو أن المقياس يقيسه وهو المظهر العام للاختبار أو الصورة الخارجية له من حيث نوع المفردات وكيفية صياغتها ومدى وضوح هذه المفردات وكذلك يتناول تعليمات الاختبار ومدى دقتها ودرجة ما تتمتع به من موضوعية ودقة تحديد الزمن المناسب للاختبارات الموقوتة التي تعتمد على السرعة وعلى تحديد مستويات الصعوبة للاختبارات غير الموقوتة التي تعتمد على القوة وعلى نوع الأسئلة ومدى صلاحيتها للإشارة والاستجابة المناسبة من المختبرين.

ولتوضيح الفكرة فإن اختبار السنة الأولى الابتدائي مثلاً يمكن أن يكون صدقاً ظاهرياً إذا كانت جميع مفرداته ذات صلة بمادة الرياضيات للسنة الأولى الابتدائي ولا يقيس قدرته اللغوية مثلاً. إن هذا النوع من الصدق يعد أقل أنواع الصدق جودة، حيث إنه من الملاحظ أن هذا النوع ليس إلا صدقاً ظاهرياً لا يلمس إلا سطح المقياس. (سوسن، 2014، ص103)

2-1-1- صدق المحتوى Content Validity:

ويبحث في مدى أداء الفرد في الحالات التي تعتبر الاختبار عينة، فمثلاً تستخدم العلامات على اختبار في مادة ما للاستدلال على معرفة الطالب الكلية في تلك المادة أو المجال الذي يغطيه الاختبار، ويمثل الاختبار عادة محتوى المادة بحيث يكون شاملاً لكافة جوانبها.

ويمكن إيجاد صدق المحتوى كالتالي:

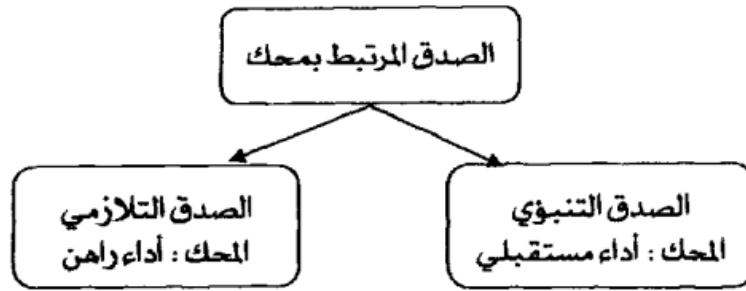
- كتابة محتوى المادة المراد اختبار الأفراد بها.

- كتابة الأهداف المراد تحقيقها.

- صياغة مفردات الاختبار بحيث تتلاءم مع الأهداف المكتوبة ثم عرض المحتوى، والأهداف، ومفردات الاختبار على لجنة من المحكمين (الخبراء) لمعرفة آرائهم في مدى مطابقتها للمحتوى، والأهداف ومفردات الاختبار لبعضها بعضاً. (الناشف، 2001، ص39)

2-1- الصدق المرتبط بالمحك Criterion – Related Validity:

يرتبط هذا النوع بالأدوات التي نريد من خلالها تقدير مدى قدرتها على التنبؤ بأداء لاحق أو مستقبلي. وينقسم هذا النوع من الصدق إلى صدق تنبؤي وصدق تلازمي ويعتمد على معامل الارتباط بيرسون لتقدير معامل الصدق في كلا النوعين. (سوسن، 2014، ص109)



شكل (02) يوضح طرق الصدق المحكي.

1-2-1- الصدق التلازمي Concurrent Validity:

وتستخدم هذه الطريقة لإيجاد صدق اختبار معين بإيجاد معامل الارتباط بين أداء المجموعة على هذا الاختبار المراد التحقق من صدقه وأدائها على اختبار آخر يقيس نفس السمة أو الصفة ولكن له صدق مقبول ومعروف، فمثلاً إذا طور اختبار ذكاء فإنه يطبق أولاً على مجموعة من الأفراد ثم يطبق ثانياً على نفس المجموعة اختبار ذكاء له صدق معروف كأنه يكون اختبار ستانفورد-بينيه المعروف والمعرب مثلاً ويتم إيجاد معامل الارتباط بين الاختبارين، فإذا كان معامل الارتباط عالياً كان اختبار الذكاء المطور عالي الصدق والعكس صحيح. (الناشف، 2001، ص40)

2-2-1- الصدق التنبؤي Predictive Validity:

يمكن إيجاد الصدق التنبؤي لاختبار ما عن طريق تطبيق هذا الاختبار على مجموعة من الأفراد ثم متابعة الأفراد بعد فترة زمنية طويلة نسبياً بتطبيق اختبار آخر يقيس نفس السمة للتحقق من السلوك الذي يقيسه الاختبار الأول ثم حساب معامل الارتباط بين أداء الأفراد على الاختبار المراد التحقق من صدقه، وأداءهم على الاختبار الآخر أو المحك الذي يقيس السلوك الذي يتنبأ به الاختبار الأول وكلما كان معامل الارتباط أعلى كان الاختبار أكثر صدقاً.

يتم استخدام علامات امتحان ما للتنبؤ عن الأداء على محك معين، ولذا يعتبر المحك هنا فائق الأهمية في حين تعتبر علامة الاختبار مهمة فقط من زاوية كونها تتنبأ على المحك، فمثلاً يستخدم اختبار معين للتنبؤ عن المعدل المتوقع للطالب في كليته أو في أي عمل.

وكمثال آخر تستخدم نتائج شهادة البكالوريا مثلاً للتنبؤ بنجاح الطالب في الجامعة وفي تخصصات معينة، كما يمكن أن تتنبأ امتحانات الذكاء بالنجاح الأكاديمي، وتستطيع امتحانات القبول في الجامعات أن تتنبأ بقدرة الطالب على النجاح في الجامعة. (الناشف، 2001، ص39)

3-1- الصدق البنائي أو الفرضي أو المفهوم Construct Validity:

يجيب صدق الاختبار البنائي على السؤالين التاليين:

ما هي السمة أو الصفة التي يقيسها الاختبار؟ وما هو المدى الذي يقيس به الاختبار هذه الصفة؟ وتكمن الأهمية في الصدق البنائي للاختبار في قدرة الاختبار على قياس السمة الكامنة لدى الشخص مثل الميول والاتجاهات، والتفكير الإبداعي، والتفكير الناقد، والذكاء، وغيرها. خلال جمع المعلومات من دلائل مختلفة مثل نوعية مفردات الاختبار، وتجانس الاختبار، وثباته من خلال جمع المعلومات الكمية. (الناشف، 2001، ص40)

وهناك عدة طرق يمكن أن يلجأ إليها الباحث للوصول إلى هذا النوع من الصدق وأهمها:

1-3-1- إيجاد معاملات الارتباط مع اختبارات أخرى: ويتم ذلك عن طريق حساب معامل الارتباط بين الاختبار الذي نريد الوصول إلى صدقه البنائي وبين الاختبارات التي تتشابه معه في الوظائف التي يقيسها والتي سبق التأكد من صدقها. فإذا افترضنا أن باحثاً أراد بناء مقياس للطموح لدى الطلبة فإنه لكي يتوصل إلى صدقه يقوم بتطبيق اختبار آخر مستقل عنه يقيس المثابرة مثلاً والتي تعتبر صفة من صفات الفرد الطموح وعندما يحصل الباحث على معامل ارتباط عالي بين الاختبارين يتوصل إلى أن اختباراه قد تحقق فيه صدق البناء.

2-3-1- إيجاد العلاقة بين درجات المفردات والاختبار (الاتساق الداخلي): وتفترض هذه الطريقة أن الدرجة الكلية للفرد تعتبر معياراً لصدق الاختبار، ويحاول الباحث عادة إيجاد العلاقة الارتباطية بين

درجة كل مفردة من مفردات الاختبار والدرجة الكلية للاختبار. وتحذف كل مفردة يكون معامل ارتباطها منخفض على اعتبار أن المفردة لا تقيس الظاهرة التي يقيسها الاختبار الكلي.

3-3-1- تقديرات الخبراء: وفي هذه الحالة نلجأ إلى الخبراء لتقدير الأفراد المبحوثين في الصفة السلوكية التي يقيسها الاختبار. (سوسن، 2014، ص115-119)

4-3-1- التحليل العاملي الاستشكافي والتوكيدي: ويهدف إلى معرفة مدى قياس الاختبار للظاهرة التي صمم الاختبار لأجلها. وتعتمد فكرته على حساب معاملات الارتباط بين كل مفردة والمفردات الأخرى المكونة للاختبار. وتتجمع نتيجته المعاملات الارتباطية "مصفوفة ارتباطات" تنقسم إلى تجمعات، ويجمع بين كل مجموعة عامل أو أكثر، ونتيجة لهذه العملية فإن الاختبار يختزل إلى عدد صغير من العوامل أو السمات المشتركة والتي يطلق عليها "المكونات الأساسية" للظاهرة التي يقيسها الاختبار.

4-1- العوامل المؤثرة على الصدق:

1-4-1- عوامل تتعلق بالمفحوص: مثلاً، القلق، الغش كلها عوامل تؤثر على صدق الاختبار.

2-4-1- عوامل تتعلق بإدارة الاختبار: مثلاً، العوامل الفيزيائية (الحرارة، الضوضاء، الإضاءة...)، سوء طباعة الاختبارات والأخطاء المطبعية، وكيفية صياغة المفردات، التعليمات غير الواضحة.

3-4-1- عوامل تتعلق بالاختبار نفسه: صعوبة وسهولة المفردات، غموض المفردات، صياغة المفردات، طول الاختبار، ثبات الاختبار، التباين، صدق مقياس المحك، مدى تمثيل الاختبار، تجانس عينة التقنين. (سوسن، 2014، ص115-119)

2- الثبات Reliability:

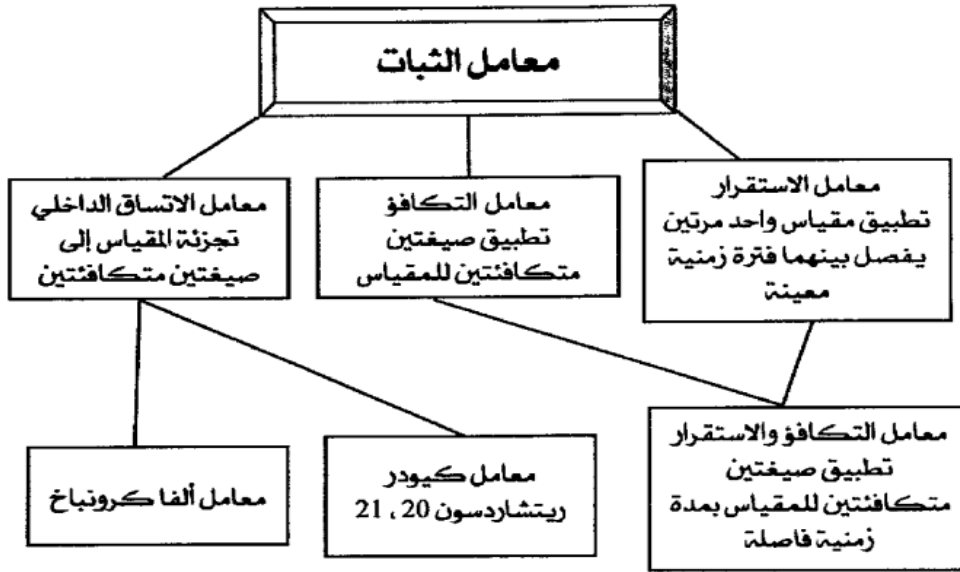
يقصد بمفهوم ثبات درجات الاختبارات مدى خلوها من الأخطاء غير المنتظمة التي تشوب القياس، أي مدى قياس الاختبار للمقدار الحقيقي للسمات التي يهدف لقياسها، فدرجات الاختبار تكون ثابتة Reliable إذا كان الاختبار يقيس سمات معينة قياساً متسقاً في الظروف المتباينة التي قد تؤدي إلى أخطاء القياس. فالثبات بهذا المعنى يعني الاتساق أو الدقة في القياس.

1-2- الثبات والدرجات الحقيقية: (علام، 2000، ص131)

إن الدرجة التي يحصل عليها الفرد في مقياس معين تسمى الدرجة الملاحظة Observed Score، غير أن هذه الدرجة تكون في كثير من الأحيان مشوبة بأخطاء القياس التي إذا أمكن تحديد مقدارها فإننا نحصل على درجة الخطأ Error Score، وإذا طرحنا درجة الخطأ من الدرجة الملاحظة فإننا نحصل على درجة الفرد خالية من الأخطاء العشوائية، وهذه تسمى الدرجة الحقيقية True Score، أي أن:

الدرجة الحقيقية = الدرجة الملاحظة - درجة الأخطاء العشوائية
الدرجة الملاحظة = الدرجة الحقيقية + درجة الأخطاء العشوائية

ويمكن حساب ثبات الاختبار بطرق عديدة يمكن تلخيصها في الشكل التالي:



شكل (03) ملخص لطرق التحقق من الثبات. (علام، 2000، ص144)

2-2- الاختبار وإعادة الاختبار أو معامل الاستقرار Test Retest Method:

ويتم في هذه الطريقة تطبيق الاختبار واستخراج نتائجه، ثم الانتظار لفترة زمنية معينة حيث يعاد تطبيق نفس الاختبار على نفس المجموعة (أو عينة عشوائية ممثلة لها) واستخراج النتائج، ثم إيجاد معامل الثبات بينهما باستخدام معامل الارتباط بيرسون.

- ومن سلبيات هذه الطريقة:

- تأثير الفترة الزمنية بين الامتحانين على النتائج، فإذا كانت الفترة الزمنية قصيرة يتذكر المفحوصون الإجابات، وإذا كانت طويلة فإن النسيان أو النضج يكون لهما تأثيراً كبيراً على مدى فهم المفحوصين وبالتالي على النتائج.

- تتطلب الكثير من الوقت والجهد.

- قد يختلف عدد المفحوصين في مرتي التطبيق نتيجة ظروف معينة طارئة.

- اختلاف ظروف تطبيق الاختبارين. (الناشف، 2001، ص41)

تصلح هذه الطريقة لحساب ثبات جميع الاختبارات، عدا اختبارات الذاكرة، شريطة أن يتراوح الفاصل الزمني بين أسبوعين إلى ستة أسابيع، ومرعاة نفس الظروف. (أبو هاشم، 2006)

3-2- طريقة الاختبارات المتكافئة Equivalent or Parallel Forms Method:

وتتطلب هذه الطريقة تطوير اختبارين متكافئين من ناحية الصعوبة والسهولة وتحقيق الأهداف وغيرها، ثم تطبيقها على مجموعة من الأفراد بفارق أقصر فترة زمنية ثم إيجاد معامل الارتباط بين أداء الأفراد على الاختبارين (الناشف، 2001، ص42). وهذا التكافؤ يشتمل على الجوانب التالية:

- عدد مكونات الوظيفة التي يقيسها الاختبار.
- عدد المفردات التي تخص كل منهما.
- مستوى صعوبة المفردات.
- طول الاختبار وطريقة إجرائه وتصحيحه وتوقيته.
- تساوي متوسط وتباين درجات الأفراد على الاختبارين. (أبو هاشم، 2006)
- ومن سلبيات هذه الطريقة لإيجاد الثبات ما يلي:
- صعوبة إيجاد صورتين متكافئتين تماماً للاختبار.
- قد يرهق المفحوص ويميل نتيجة إعادة الاختبار خاصة إذا كانت الفترة الزمنية قصيرة.

4-2- طريقة التجزئة النصفية (الثبات النسبي) Split – Half Method:

يقسم الاختبار في هذه الطريقة إلى قسمين متكافئين من حيث الأهداف، وصعوبة المفردات، وعددها، وزمن كل منها، وغيرها. ثم تؤخذ المفردات الفردية للاختبار كجزء وتعطى علامة وتؤخذ المفردات الزوجية كجزء آخر وتعطى علامة أيضاً ثم يحسب معامل الارتباط بين أداء الأفراد على الجزئين حيث يشير معامل الارتباط هذا إلى معامل ثبات نصف الاختبار. (الناشف، 2001، ص42) وهناك عدة معادلات تستخدم لتصحيح معامل ثبات نصفي الاختبار منها:

1-4-2- معادلة سبيرمان- براون Spearman – Brown:

وفيها يتم التعويض قيمة معامل الارتباط بين نصفي الاختبار لنحصل على معامل ثبات الاختبار ككل، وهي شائعة الاستخدام وبخاصة في اختبارات التحصيل والقدرات تحت ظروف محددة. ولإجراء تصحيح لهذه القيمة لتعكس معامل الثبات للأداة الكلية في عينتنا يتم حساب التصحيح، الذي يعرف باسم تصحيح سبيرمان، وبراون على النحو، التالي:

$$SB = \frac{2r_{H1,H2}}{1+r_{H,1H2}},$$

حيث: $r_{H2 \cdot H1}$ هو معامل الارتباط بين نصفي المقياس.

2-4-2- معادلة رولون Rulon:

وتعتمد على حساب تباين درجات الاختبار ككل، ثم حساب تباين الفروق بين درجات الأفراد في النصف الأول، ودرجاتهم في النصف الثاني ثم تطبيق المعادلة نحصل على معامل ثبات الاختبار ككل.

2-4-3- معادلة جتمان Guttman:

وفي هذه المعادلة يتم حساب تباين درجات النصف الأول، وتباين درجات النصف الثاني، وتباين درجات الاختبار ككل. أي أنها تضع في الاعتبار احتمال اختلاف تباين درجات النصف الأول للاختبار عن تباين درجات النصف الثاني " وهذا لا يتحقق في المعادلتين السابقتين". (أبو هاشم، 2006) وتعتبر طريقة التجزئة النصفية أكثر طرق حساب الثبات شيوعاً حيث يطبق الاختبار مرة واحدة فيختصر الوقت والجهد ثم يتم حساب الثبات بتجزئته إلى نصفين أو أكثر. (الناشف، 2001، ص43)

2-5-5- طريقة ثبات الاتساق الداخلي Internal Consistency:

وتستخدم هذه الطريقة لإيجاد مدى الارتباط بين أي مفردة من مفردات الاختبار وبين العلامة الكلية للاختبار أو إيجاد مدى الارتباط بين مفردة ومفردة أخرى في الاختبار.

2-5-1- معادلة كودر-ريتشاردسون (20) Kuder – Richardson :

وتعتمد على توفر بيانات عن تباين كل مفردة من مفردات الاختبار. وفي حالة عدم توفر هذه البيانات يمكن استخدام الصيغة رقم (21) بنفس المعادلة وهي تتميز بالسهولة والسرعة في حسابها حيث إنها لا تحتاج لمعرفة تباين المفردات، ولكن يعيها أنها أقل دقة من الصيغ السابقة. وتعطى معادلة كودر ريتشاردسون (20)، كالتالي:

$$\left(\frac{K}{K-1} \right) \left[1 - \frac{\sum_{k=1}^p (p_k q_k)}{S_{Total}^2} \right]$$

حيث: K = إجمالي عدد المفردات على الأداة.

pk = نسبة الأفراد المجيبين إجابة صحيحة على المفردة.

qk = نسبة الأفراد المجيبين إجابة خاطئة على المفردة.

S^2_{Total} = تباين الدرجات في الاختبار.

وقد وضع كودر-ريتشارسون شروطاً لاستخدام هذه المعادلة وهي:

- أن تكون درجة أسئلة الاختبار (صفر أو واحد).

- ألا يكون عدد الأسئلة المتروكة كبير.

- تقارب مستوى صعوبة الأسئلة.

- تساوي معاملات الارتباط بين درجات الأسئلة. ومن الواضح أن هذه الشروط لا توجد في اختبارات

المجال الوجداني ولذلك ننصح بعد استخدامها في هذا النوع من الاختبارات.

مثال تطبيقي:

وليكن لدينا اختبار مكون من خمسة مفردات، طبق على ستة طلبة وكل طالب يجيب على كل

مفردة من الاختبار من خلال بديلين (0 - 1)، وكانت النتائج كالتالي:

جدول (01) يوضح توزيع درجات الطلبة على مفردات الاختبار.

المجموع	المفردة 5	المفردة 4	المفردة 3	المفردة 2	المفردة 1	
4	1	1	1	0	1	الطالب 1
5	1	1	1	1	1	الطالب 2
2	0	0	1	0	1	الطالب 3
1	0	1	0	0	0	الطالب 4
5	1	1	1	1	1	الطالب 5
4	1	1	0	1	1	الطالب 6
$S^2 = 2.25$	0.6667	0.8333	0.6667	0.5	0.8333	p_k
	0.3333	0.1667	0.3333	0.5	0.1667	q_k
	0.2222	0.1389	0.2222	0.25	0.1389	$p_k \times q_k$

وبتطبيق معادلة كودر-ريتشارسون نتحصل على النتائج التالية: (Reynolds et al, 2021, p.128)

$$\sum p_k \times q_k = 0.1389 + 0.25 + 0.2222 + 0.1389 + 0.2222$$

$$\sum p_k \times q_k = 0.972$$

$$KR-20 = \frac{5}{4} \left(1 - \frac{0.972}{2.25} \right)$$

$$= 1.25 (1 - 0.432)$$

$$= 1.25 (0.568)$$

$$= 0.71$$

2-5-2- معادلة ألفا كرونباخ Cronbach Alpha:

ويعتبر معامل ألفا كرونباخ حالة خاصة من معادلة كيودر-ريتشارسون. ويمثل معامل ألفا متوسط المعاملات الناتجة عن تجزئة الاختبار إلى طرق مختلفة، وبذلك فإنه يمثل معامل الارتباط بين أي جزئين من أجزاء الاختبار. ويتم حساب تباين كل مفردة من مفردات الاختبار ثم مجموع التباينات، وكذلك تباين الدرجة الكلية للاختبار، وتشتراط أن تقيس مفردات الاختبار سمة واحدة فقط. وتستخدم هذه المعادلة في المقاييس والاختبارات متعددة الاختيارات (ليكرت). (أبو هاشم، 2006، ص09) وتعطى معادلة ألفا كرونباخ، كما يلي:

$$\left(\frac{K}{K-1} \right) \left[1 - \frac{\sum_{k=1}^p S_k^2}{S_{Total}^2} \right]$$

حيث: K = العدد الكلي لمفردات المقياس.

$$S_k^2 = \text{تباين المفردة } K.$$

$$S^2 = \text{التباين الكلي للمقياس.}$$

مثال تطبيقي:

وليكن لدينا اختبار مكون من خمسة مفردات، طبق على ستة طلبة وكل طالب يجيب على كل مفردة من الاختبار من خلال خمسة بدائل ليكرت من (1 - 5)، وكانت النتائج كالتالي:

جدول (02) يوضح توزيع درجات الطلبة على مفردات الاختبار.

المجموع	المفردة 5	المفردة 4	المفردة 3	المفردة 2	المفردة 1	
21	5	5	4	3	4	الطالب 1
14	3	3	2	3	3	الطالب 2
10	1	2	2	3	2	الطالب 3
20	4	3	5	4	4	الطالب 4
14	3	2	4	3	2	الطالب 5
10	3	1	2	2	2	الطالب 6
$S^2 = 18.81$	1.4722	1.5556	1.4722	0.3333	0.8056	S_k^2

وبتطبيق معادلة ألفا كرونباخ نتحصل على النتائج التالية: (Reynolds et al, 2021, p.129)

$$\begin{aligned} \text{Coefficient Alpha} &= \frac{5}{4} \left(1 - \frac{0.8056 + 0.3333 + 1.4722 + 1.5556 + 1.4722}{18.81} \right) \\ &= 1.25 \left(1 - \frac{5.63889}{18.81} \right) \\ &= 1.25 (1 - 0.29978) \\ &= 1.25 (0.70) \\ &= 0.875 \end{aligned}$$

ملاحظة هامة: وهي أن يتوجب على الباحث التأكد من صدق الاختبار أولاً ثم من ثباته لأن الاختبار الصادق دائماً ثابت في حين أن الاختبار الثابت ليس من الضروري أن يكون صادقاً، كأن تعطى علامات الطلبة حسب الطول حيث يتضمن ذلك الثبات ولا يتضمن الصدق، وإذا لم يكن الاختبار صادقاً أو ثابتاً يتوجب على واضعه إعادته أو تعديله، ويتم التأكد من صدق الثبات بتطبيق لاختبار على عينة استطلاعية لا علاقة لعينة الدراسة بها.

- ويلاحظ كذلك أنه تقبل معاملات ثبات منخفضة نسبياً للاختبارات الاتجاهات والميول مثلاً وبشكل عام للمقاييس الانفعالية في حين يتوجب الحصول على معاملات ثبات عالية نسبياً للاختبارات التحصيلية.

- ويلاحظ أنه يتوجب على الشخص التأكد من صدق الاختبار أولاً ثم من ثباته لأن الاختبار الصادق دائماً ثابت في حين أن الاختبار الثابت ليس من الضروري أن يكون صادقاً، كأن تعطى علامات الطلبة حسب الطول حيث يتضمن ذلك الثبات ولا يتضمن الصدق، وإذا لم يكن الاختبار صادقاً أو ثابتاً يتوجب على واضعه إعادته أو تعديله، ويتم التأكد من صدق الثبات بتطبيق لاختبار على عينة استطلاعية لا علاقة لعينة الدراسة بها. (الناشف، 2001، ص43)

ويلاحظ كذلك أنه تقبل معاملات ثبات منخفضة نسبياً للاختبارات الاتجاهات والميول مثلاً وبشكل عام للمقاييس الانفعالية في حين يتوجب الحصول على معاملات ثبات عالية نسبياً للاختبارات التحصيلية.

6-2- العوامل المؤثرة على الثبات:

يتأثر معامل ثبات الاختبار بالعديد من العوامل أهمها:

6-2-1- طول الاختبار: تزداد قيمة معامل ثبات الاختبار بزيادة مفرداته، ويرجع ذلك إلى أن زيادة عدد مفردات الاختبار تؤدي إلى شمول أكثر للمحتوى ومن ثم صدق المحتوى مرتفع.

2-6-2- تباين درجات الأفراد: يقل معامل الثبات بزيادة تجانس المجموعة المطبق عليها وكلما ارتفع تباين الأفراد كلما ارتفع معامل الثبات.

2-6-3 مستوى القدرة في مجموعة الاختبار: وهذا يرتبط بالدقة بمستوى القدرة لدى الأشخاص، حيث يختلف مستوى دقة الاختبار باختلاف مستويات الصعوبة، ومن ثم يجب أن يحتوي الاختبار على عدداً كبيراً من المفردات المتوسطة الصعوبة التي تميز بين مستويات القدرة في المدى المتوسط، وعدد قليل من المفردات السهلة جداً والصعبة جداً.

2-6-4- نوع العمليات المستخدمة في حساب الثبات: إن مدى الارتفاع في معامل الثبات الذي يمكن الحصول عليه يعتمد كذلك على نوع العمليات التجريبية المستخدمة في حساب الثبات. فطرق حساب الثبات تعالج مصادر مختلفة للتباين بأساليب مختلفة، وأن طريقة الصور المتكافئة بوجود فاصل زمني بين التطبيقين هي وحدها التي تشمل معالجة جميع مصادر التباين في الخطأ. أي أن هذه الطريقة في حساب الثبات تمثل تعريفاً. هو الأكثر دقة لقدرة الاختبار على الإتيان بنفس الدرجة. ويجب على الفرد عندئذ أن يظهر اتساقاً في الأداء من مهمة إلى أخرى وكذلك من يوم لآخر.

فقد أظهرت الأمثلة التي توضح معاملات الثبات لنفس الاختبار عند حسابها بطريقتين مختلفتين أن معاملات الارتباط بطريقة الصور المتكافئة أقل من معامل الارتباط بين نصفي الاختبار. وهذا يؤكد أن طريقة الصور المتكافئة تفرض شروط أكبر من حيث الدقة التي يجب أن تتوافر في الاختبار.

2-6-5- الموضوعية: تظهر الاختبارات الموضوعية ثباتاً أكثر من الاختبارات المقالية أو غير الموضوعية، ويرجع ذلك إلى ذاتية تقدير الدرجات في الاختبارات المقالية. ولكن ليس المقصود من ذلك هو عدم إعطاء اختبارات مقال حتى نحصل على اختبارات ثابتة، فهذا من ناحية أخرى يؤثر على الصدق الذي يعتبر أهم صفة من صفات الاختبار الجيد. ولكن الحل الأمثل هو استخدام وسيلة القياس الأنسب لقياس مخرجات التعلم، وبعد ذلك نحاول زيادة موضوعية الاختبار، بإتباع مجموع من القواعد سواء في الإعداد أو التصحيح.

2-6-6- زمن أداء الاختبار: زيادة الزمن تؤدي بالفرد إلى الحصول على أعلى درجة متسقة مع قدرته، إلا أن زيادة الزمن بدرجة أكبر من اللازم قد تؤدي إلى الارتباك في الإجابة ومن ثم خفض معامل الثبات ويمكن القول أن معامل الثبات يرتفع بزيادة الوقت الذي يستغرقه الاختبار ولكن إذا طالت المدة بشكل كبير ينخفض الثبات. ويختلف أثر طول المدة من اختبار لآخر. (أبوهاشم، 2006، ص11)

المحاضرة الثالثة:

بناء وتطوير المقاييس والاختبارات النفسية والتربوية

Construction and Development of Psychological and Educational Scales and Testing

تمهيد:

تطوير المقاييس والاختبارات مجال يتطور بسرعة مع ظهور مناهج جديدة واكتساب المزيد من القوة. وقد نشأت هذه التغييرات من مجموعة متنوعة من المصادر، بما في ذلك الجهود الكبيرة لتوفير المزيد من أدوات القياس الموحدة بشكل أكبر عبر مجموعة واسعة من السياقات. أولاً: تصنيف الاختبارات والمقاييس النفسية والتربوية واستخداماتها: هناك أنواع عديدة من الاختبارات تصنف بحسب استخداماتها وفوائدها.

1- في ضوء تصحيح ووضع العلامات:

تصنف إلى الاختبارات إلى:

1-1- اختبارات موضوعية: حيث يكون الاختبار موضوعياً إذا إعطاء العلامة للسؤال أو الاختبار موضوعياً. وهذا مرتبط بخصائص وقواعد يعبر عنها الاختبار الموضوعي.

1-2- اختبارات المقال أو الاختبارات الذاتية: وهي الاختبارات التي تتضمن أسئلة مقال كالإنتاجيات الفنية أو الأعمال الخلاقة. حيث تدخل الذاتية عند إعطاء العلامة. فإذا صحح إثنان أو أكثر مقالة أو رسماً أو تلوينة. فإنهم سيختلفون في وضع العلامة بل قد يقدم المصحح نفسه علامتين مختلفتين في مناسبتين مختلفتين، وهذا النوع مرتبط أيضاً بخصائص وقواعد ومزايا يعبر عنها الاختبار المقالي أو الذاتي. (ملحم، 2012، ص52)

2- إنطلاقاً من درجة التقنين أو التعيير:

فإنه يمكن تقسيم الاختبارات إلى:

1-2- اختبارات مقننة أو معيرة: وهذا يعني اختباراً يعد من قبل الأخصائيين في الاختبارات ومواد التخصص بطرق معيرة ومبلورة لإدارة الاختبار وملاحظته وإعطاءه العلامات. كما أن هذا النوع ينتجه الناشرون والتجار يرون بإشراف الأخصائيين في الاختبار وذوي العلاقة. مع مراعاة إدارتها وتصحيحها وتفسيرها تحت ظروف معيرة ومحددة.

2-2- اختبارات غير رسمية يعدها المعلم: ويمثل هذا النوع من الاختبارات الغالبية العظمى للاختبارات في مجال التربية والتعليم. وهي اختبارات تصمم من قبل المعلمين. وتتميز بأساليب غير معيرة لكنه يمكن مقارنة نتائجها مع مجموعة أخرى من الاختبارات. وتمتلك عدة خصائص ومزايا تتلاءم وأهداف المعلم المحددة لعينة من الطلاب الذين يقوم المعلم بقياس قدراتهم في موضوع دراسي ما. (ملحم، 2012، ص52)

3- استناداً إلى أسلوب إدارة الاختبار:

كما يمكن تصنيف الاختبارات إلى قسمين رئيسيين هما:

3-1- اختبارات فردية: تصمم لتدار على فرد واحد في جلسة اختبار خاصة. كالفحص الإكلينيكي أو العيادي حيث تتاح للفاحص فرصة إقامة صلة ألفة مع المفحوص تجعله أكثر قرباً منه وأقوى على تحقيق أهدافه.

3-2- اختبارات جماعية: تصمم هذه الاختبارات لتجرى على مجموعة كبيرة من الناس في نفس الوقت. وهي الأكثر شيوعاً واستخداماً في مجال التربية حيث يمكن لاختبار جمعي أن يدار من قبل شخص واحد يقدم فيه مفحوصون بأعداد كبيرة اختباراً ما. (محمود وآخرون، 2009، ص104)

4- في ضوء الاهتمام المعطى للغة في عملية الاختبار:

فإنه يمكن تقسيم الاختبارات إلى:

4-1- اختبارات لفظية: تعطى الأهمية لملاحظة استجابات المفحوصين اللغوية مثل: التشابه والاختلاف بين المفاهيم، واستقصاء التناظرات الشفوية، وتعريف كلمات ومفاهيم، وكتابة إجابات لأسئلة... الخ.

4-2- اختبارات أدائية أو اختبارات المهارات العملية: وتصمم لمشاهدة الاستجابات غير اللفظية مثل أداء مهمة أو تجميع وتفكيك جزء من آلة والنشاطات الرياضية وأية نشاطات نفسية حركية للمفحوصين.

5- في ضوء الاهتمام المعطى لعامل الوقت:

يمكنك تصنيف الاختبارات إلى:

5-1- اختبارات السرعة (القوة): حيث يكون التركيز فيه على سرعة الإجابة وضبطها أكثر من التركيز على المحتوى أو المعرفة. حيث يتمكن الطلاب من خلالها في أداء مهمة بدرجة معينة من الضبط والتي تمثل الهدف الحقيقي للقياس. وعليه: ففي هذا النوع من الاختبار فإن السمة التي يجب قياسها هي سرعة الأداء.

5-2- اختبارات الدقة: بهدف تقدير حدود القدرة أو العمق أو التوسع المعرفي في مجال موضوع معين.

(ملحم، 2012، ص53)

6- في ضوء طبيعة مجال القياس (المعرفي مقابل العاطفي):

يمكن تقسيم الاختبارات إلى:

1-6- اختبارات الأداء الأقصى: من أجل تحديد حجم قدرات المفحوص والتي بها يتم محاولة ما يمكن أن يعمل عندما يبذل قصارى جهده بحيث يتوقع من المفحوص بذل قصارى جهده ليحصل على أقصى علامة ممكنة في الاختبار.

2-6- اختبارات الأداء الطبيعي: وتهدف هذه الاختبارات إلى تقدير الاستجابات العادية والسلوكيات الطبيعية وتحديد ما يمكن أن يقوم به الشخص بصورة تلقائية طبيعية تحت ظروف عادية بحيث يتوقع من المفحوص أن يعبر عن مشاعره الحقيقية أو اتجاهاته أو عاداته أو أدائه أو تفصيلاته أو اعتقاداته أو آرائه... الخ. (ملحم، 2012، ص53)

7- بناء على تحديد فقرات الاستجابة لنوع القدرات العقلية:

فإنه يمكنك تصنيف الاختبارات إلى:

1-7- اختبارات تزويد أو استدعاء الاستجابة: بهدف اختبار الذاكرة بحيث يقدم الشخص الإجابة من ذاكرته ويستدعي الاستجابة للسؤال.

2-7- اختبار اختيار الاستجابة: بحيث يقدم الاستجابة أو الإجابة الصحيحة لكل فقرة. وما على الطالب إلا أن يحدد أو يختار الاستجابة من عدة بدائل معطاة. مثال ذلك الاختبارات من نوع الصح والخطأ، اختبارات المزاجية، أسئلة الاختيار من متعدد. (ملحم، 2012، ص54)

8- في ضوء استخدام الاختبارات للتعليم الصفي أو لنوع القرار:

يمكن تقسيم الاختبارات إلى:

1-8- اختبارات تحديد المستوى: تستخدم لتحديد المهارات التي تعد كمتطلبات سابقة أو مدى درجة إتقان أهداف لمساق ما. أو اختبار أفضل نموذج تعليمي يناسب الطلاب. مثال ذلك: اختبارات الاستعداد والقابلية والاختبار القبلي وتقرير استبيانات الذات... الخ.

2-8- الاختبارات المرحلية أو البنائية أو التكوينية: وهي تلك الاختبارات التي تستخدم للتقويم المرحلي وتهدف إلى:

- تحديد مدى تقدم تعلم الطالب.

- تزويد المعلم والطلاب بتغذية مرتدة.

- معالجة أخطاء التعلم.

3-8- الاختبارات العلاجية: وهي الاختبارات التي تستعمل لتحديد صعوبات تعلم معين بهدف علاجها.

4-8- الاختبارات الختامية أو النهائية: وهي الاختبارات التي تستخدم للتقويم الختامي وهدفها الرئيسي هو تحديد التحصيل في نهاية مقرر دراسي من أجل تعيين مدى إتقان الطلاب للأهداف واعطائهم علامات مناسبة. وأكثر الاختبارات استخداما هي اختبارات التحصيل التي يقوم بإعدادها المعلم. (ملحم، 2012، ص54)

9- بناء على كيفية تفسير درجات الاختبار:

يمكن تقسيم الاختبارات إلى:

9-1- اختبارات معيارية المرجع: يصنف الاختبار بأنه معياري المرجع إذا صمم ليزود الفاحص بعلامات تفسر بمقارنتها مع علامات مجموعة محددة جيدا في اختبار معين. وعلى هذا الأساس يمكن تفسير علامات الفرد باتخاذ مجموعته معيارا والذي يمكن أن يكون معدل (وسط، وسيط، منوال) كما يمكن مقارنتها مع علامة كل فرد لتحديد موقعها النسبي أو رتبتها المئينية بين المجموعة كلها.

9-2- اختبارات محكية المرجع: يسمى الاختبار بمحكي المرجع إذا صمم لينتج علامة اختبار تفسر في ضوء أداء محكي مستقل. وفي هذه الحالة يكون مستوى الأداء محددًا وموصوفًا بوضوح. وهذا المستوى من الأداء يخدم كمرجع مقابل كل علامة اختبار سنقارن أو نقوم. (ملحم، 2012، ص54)



شكل (04) يوضح تصنيف الاختبارات والمقاييس النفسية والتربوية واستخداماتها.

(ملحم، 2012، ص55)

ثانياً: أهمية الاختبارات والمقاييس النفسية:

إن الاختبارات النفسية هي أداة ووسيلة في عمليات تقدير إمكانيات الفرد، فإنها تهدف إلى تشخيص والتنبؤ بالحالة، وذلك له أهمية كبرى في عملية رسم استراتيجية التوجيه والعلاج من أجل تحقيق التوافق الذاتي للفرد.

ثالثاً: الحاجة إلى بناء اختبار نفسي جديد:

قبل البدء في سرد خطوات الاختبار قد يكون من المفيد أن نتساءل: لماذا نقوم ببناء اختبار جديد ما دام أن هناك اختبارات جاهزة؟ (عزيزان وبجاش، 2018، ص 87)

فعلى الرغم مما يكتنف بناء الاختبارات من مشقة كبيرة إلا أن الأخصائي أو الباحث النفسي قد يجد نفسه إزاء موقف يلزمه بأن يقوم بهذا العمل أو أن يسهم فيه. فالأخصائي النفسي قد يقوم بعمليات الاختيار التربوي أو المهني، أو الإرشاد النفسي أو عمل يدخل في نطاق الخدمات النفسية الفردية أو الجماعية، وهو في كل هذا يحتاج إلى أداة موضوعية مقننة لتقدير الأفراد الذين يتعامل معهم بطبيعة عمله. وهو قد يجد تلك الأداة في اختبارات جيدة تناسب الوظيفة التي استخدمها، إلا أن مثل هذه الأختبارات وانتشارها الواسع، قد تكون مفرداتها معروفة وهنا يفقد الاختبار الموجود قيمته.

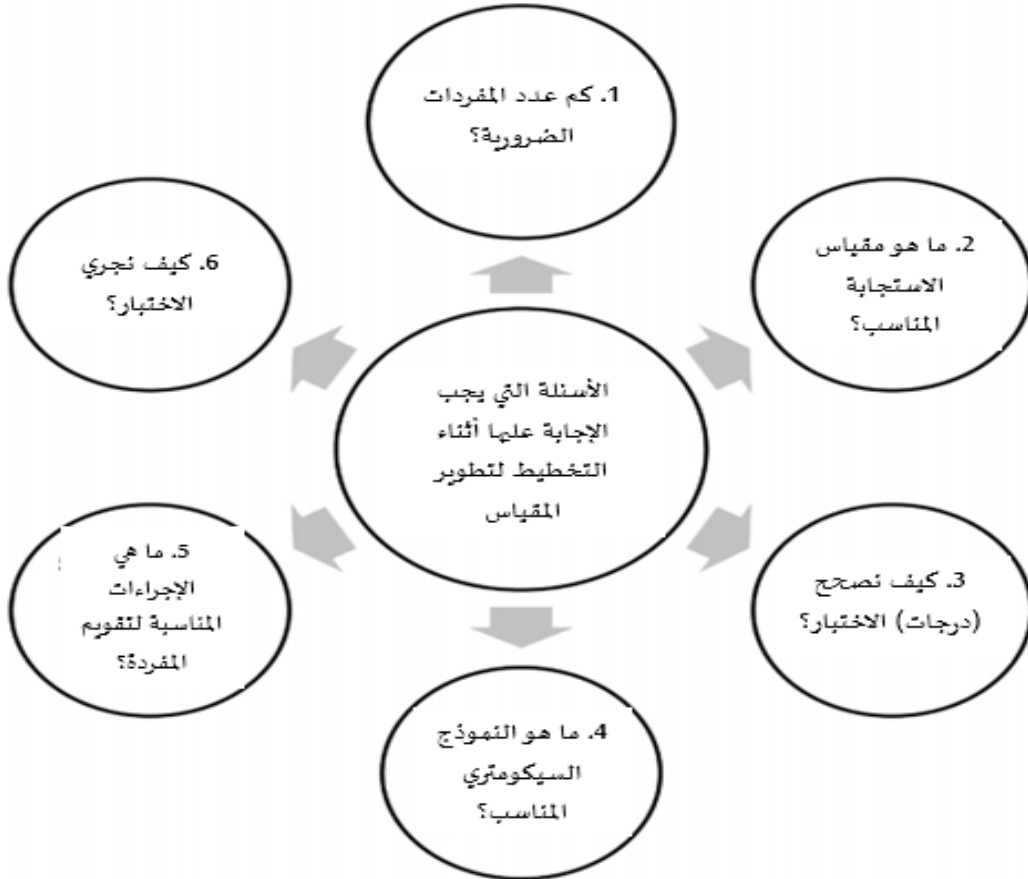
- قد يقيس الاختبار وظيفة أخرى غير المطلوبة، أو نفس الوظيفة ولكن من زاوية أخرى،
- قد يكون الاختبار قد قنن واشتقت معاييره من جماعات تختلف في طبيعتها عن الجماعات المراد دراستها.
- قد يتطلب الاختبار الموجود في إجراءاته أو إقتنائته نفقات التي يستطيع الباحث تحملها، وقد يكون الاختبار الموجود أميل إلى الذاتية مما يجعله أداة لا يمكن الاعتماد عليها.
- قد تظهر مفاهيم نظرية أو ظواهر جديدة تستدعي بناء اختبارات جديدة. كما قد توجد اختبارات تصلح للتطبيق على فئة من الأشخاص ولا توجد من الدلائل أنها تصلح للتطبيق على الفئة موضع الاهتمام، وقد تظهر نظرية جديدة بمفاهيم جديدة تتطلب اختبارات جديدة لتلك المفاهيم.
- قد يقوم بسبب عدم اقتناع الباحثين بالاختبارات المتوفرة وبأنها لا تناسب الثقافة التي يعملون فيها، وفي بعض الأحوال لا يكون لدى الباحث اختبار يقيس المفهوم أو الأبعاد التي يرغب في قياسها، أو تكون الاختبارات الموجودة موضعاً للشك من حيث تمتعها بالشروط العلمية من صدق وثبات ومعايير.
- قد يكون الاختبار المتوفر لا يرقى إلى تقييم الأفراد الذين وصلوا إلى مستوى علمي أو مهني معين.
- قد يقيس الاختبار المتوفر الوظيفة من جهة أخرى. أو أن يكون الاختبار الموجود لا يفي بالغرض في حدود الزمن المحدد، كأن يتطلب وقتاً يزيد عما يسمح به عمل المختص، أو تلف مواد الاختبار عندما تنقل من جهة إلى أخرى، أو صعوبة نقلها بطريقة علمية. (عزيزان وبجاش، 2018، ص 87)
- ولا يعد تصميم وإعداد اختبار هدفاً في حد ذاته، فالباحث لا يلجأ عادة لهذا العمل إلا إذا دعت الحاجة إلى ذلك ويتضح له أن الاختبارات المتوفرة لا تفي بغرضه، ودون وجود الدواعي الضرورية

لتصميم اختبار جديد يظل استخدام الاختبارات المتوفرة في المجال أو تطويرها أفضل كثيراً حتى يمكن الاستفادة من الدراسات والنتائج التي أمكن التوصل إليها. (سوسن، 2014، ص 68)

ففي ضوء الطبيعة التراكمية للعلم يعد تراث البحوث التي استخدم فيها اختبار ما بمثابة رصيد من المعارف العلمية التي يتعين تنميتها أو تطويرها لا إهمالها بين الحين والآخر بابتكار أدوات وأساليب قياس جديدة تبدأ من نقطة الصفر، الأولوية دائماً لتصويب وتعديل وتطوير الأدوات المتوفرة مهما تضمن تراثها من إشارات نقدية أو سلبية. (صفوت، 2007، ص 127)

رابعاً: خطوات بناء الاختبارات والمقاييس النفسية والتربوية:

يخضع تصميم وإعداد الاختبار الجيد لعدد من الاعتبارات والقواعد المنهجية، كما تتطلب فنية تصميم الاختبار دراية واسعة وعميقة بالأساليب المختلفة ومدى مناسبة كل منهما للأغراض المختلفة. إلا أن هناك مجموعة من الأسئلة التي يجب أن يجيب عليه المصمم للاختبار قبل البدء في عملية تصميم الاختبار، كما هي موضحة في الشكل رقم (05):

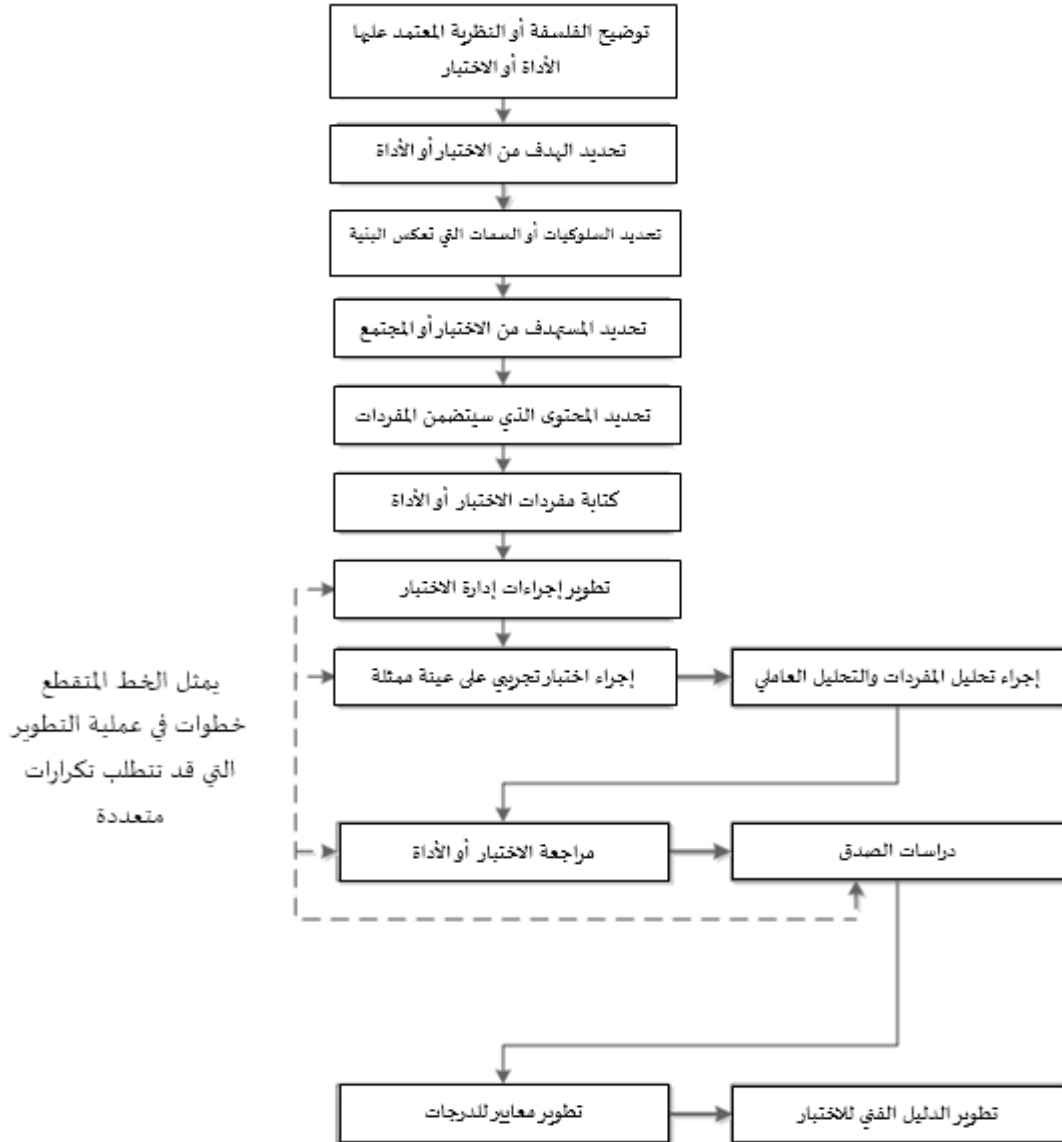


شكل (05) يوضح الأسئلة التي يجب الإجابة عليها أثناء التخطيط لتطوير المقياس.

(Irwing & Hughes, 2018, pp. 9-10)

خامساً: المبادئ التوجيهية لتطوير الاختبارات أو أدوات القياس:

تصف المبادئ التوجيهية التالية المكونات الرئيسية والاعتبارات التقنية للاختبار الفعال و بناء أدوات القياس والتي هي موضحة في الشكل رقم (06). وبالإضافة إلى توفير أسلوب متماسك، يوفر تطبيق الإطار التالي (المبادئ التوجيهية) أدلة على الحجج المتعلقة بكفاية أدلة الصدق أو الصلاحية المتعلقة بالاستخدام المراد للدرجات التي تم الحصول عليها من استخدام الاختبارات أو أدوات القياس. هذه المعلومات هي مكون أساسي للتقارير عن أدلة الصدق أو الصلاحية الشاملة على النحو الموصى به في معايير الاختبارات التربوية والنفسية التي نشرتها الجمعية الأمريكية للبحوث التربوية (AERA) والجمعية الأمريكية لعلم النفس (APA) والمجلس الوطني للقياس في التربية (NCME) (1999).



شكل (06) يوضح خطوات وإجراءات تطوير الاختبارات النفسية والتربوية. (Price, 2016, p.167)

المبدأ التوجيهي رقم 01: تحديد الأساس النظري للاختبار.

لا بد للخاصية المقاسة أن تستند إلى أساس نظري يبرر مشروعيتها تناولها ويعرفها، وقد يكون المقياس معد في الأصل للتأكيد من مدى جدوى النظرية التي تفسر السمة أو الخاصية المقاسة، والنتيجة المستخلصة قد تفيد النظرية أو تعدلها، فضلاً على أن الاطلاع على الأطر النظرية للخاصية المستهدفة بالمقياس قد يزود الباحث بالأهمية النسبية للأبعاد الفرعية المكونة للخاصية ومن ثم تمثل الأبعاد في مفردات تكون المقياس فيما بعد وفقاً لتلك الأهمية. (الزاملي، 2017، ص50)

ينبغي أن يوفر الأساس النظري للاختبار أو الأداة علاقة منطقية وذات معنى بين ما يعتقد الاختبار قياسه وبنية الاختبار أو الأداة، وتشكل مفردات الاختبار علاقة بالاطار النظري والتي تصبح كمجال لمحتوى الاختبار. (Price, 2017, p165-181)

المبدأ التوجيهي رقم 02: تحديد الهدف أو الغرض من الاختبار.

تحتوي الأدبيات المتعلقة بتطوير الاختبارات على العديد من المقاربات لتحديد الغرض من الاختبار، وتحديد الهدف أمر أساسي لعدة أسباب. (Price, 2017, p165-181) وتقوم هذه الخطوة بدور الموجه الذي يعين مصمم المقياس خلال الخطوات التالية على إعداد مقياساً يفي بالغرض المطلوب، ويقصد بتلك الخطوة تحديد الخدمة المطلوب من المقياس أن يقدمها، أو الهدف المراد تحقيقه من وراء المقياس، وتنقسم تلك الأهداف إلى نوعين هما:

أ- أهداف عامة مثل:

- سد عجز في الأدوات التي تتصدى لقياس الخاصية المراد قياسها.

- التأكد من مدى فعالية نظرية ما.

- التعرف على درجة امتلاك الأفراد لخاصية ما.

ب- أهداف خاصة مثل:

- الاستخدام بغرض الاختيار أو التوجيه المهني أو التعليمي.

- الاستخدام بغرض التشخيص.

- الاستخدام بغرض التقويم.

- الاستخدام بغرض اختبار الفروض العلمية كما يحدث في البحوث. (الزاملي، 2017، ص50)

المبدأ التوجيهي رقم 03: تحديد السمات أو السلوكيات التي تعكس بنية الاختبار.

ولعل أحد أكثر الجوانب تحدياً في عملية بناء الاختبار أو الأداة يتعلق بتحديد السمات التي تمثل بدقة البناء الذي يستهدف الاختبار قياسه.

المبدأ التوجيهي رقم 04: تحديد مجتمع الممتحنين.

المعينة هي تحديد و اختيار المفردات وفقاً للقواعد المنصوص عليها من المجتمع المحدد. في تطوير الاختبار تكون مفردات العينة هي الممتحنين أو الأشخاص الذين يطبق عليهم الاختبار أو يستجيبون لمفردات الأداة.

المبدأ التوجيهي رقم 05: تحديد محتوى المفردات.

قبل تحديد محتوى مفردات الاختبار، يجب اختيار البنية والعمل عليها بطريقة تكون بمثابة أساس للاختبار أو الأداة. ذكر في المبدأ التوجيهي 3 أعلاه أن اتخاذ قرار في بنية الاختبار يستند عادة إلى استعراض الأدبيات ذات الصلة، إلى جانب استشارة الخبراء في الموضوع.

المبدأ التوجيهي رقم 06: كتابة مفردات الاختبار.

تطوير مفردات الاختبار لقياس البنية المستهدفة ينطوي على عدة اعتبارات، هي: أولاً: هو اختيار شكل المفردة أن يكون مناسب لمهمة القياس وفعال للممتحنين. ثانياً: يجب اختيار الأشخاص وتدريبهم فيما يتعلق بتقنيات كتابة المفردات الفعالة. ثالثاً: مهمة توليد المفردات أو كتابتها.

رابعاً: يجب مراقبة عملية كتابة المفردات لضمان جودة الاختبار. (Price, 2017, p165-181)

و يتم اختيار إحدى الطرق المعروفة سواء طريقة (ثريستون) أو طريقة (ليكرت) في صياغة مفردات المقياس والأفضل طريقة ليكرت لأنها تعطي معلومات أشمل عن المفحوص لأنها تشمل الجانبين الإيجابي والسلبي والأفضل يجب أن تكون المفردات بصياغة المتكلم وتقبل تفسير واحد، وأن لا تكون المفردة سهلة جداً أو صعبة جداً. إن نوع المفردة يتحدد تبعاً لعدة اعتبارات أهمها:

- الوظيفة أو الوظائف التي يقيسها المقياس.

- الغرض من المقياس.

- نوع العينة المراد قياسها.

- حدود الوقت والتكاليف.

- صدق الفقرات وثباتها.

- الإجراءات المتبعة في تصميم المقياس.

أما مواصفات كتابة أو صياغة مفردات المقياس تكون:

- قصيرة ولا تزيد عن عشرين كلمة.

- غير مصاغة في الماضي.

- لا تعبر عن حقيقة أو تفسير على شكل حقيقة.

- تحتوي على فكرة واحدة بسيطة غير مركبة.
 - مكتوبة بلغة سهلة واضحة المعاني.
 - أن تكون جملاً اعتقادية إنفعالية بحسب طبيعة الموضوع.
 - تعكس تعريف الاتجاه المراد قياسه. (الزاملي، 2017، ص50)
- تحديد شكل الاستجابة:

توجد عدة أشكال للاستجابة على مفردات التي يتكون منها المقياس ويتوقف اختيار أيها منها على هدف المقياس ويجب أن يراعى في اختيار بدائل الإجابة على المستوى الثقافي والدراسي للمفحوص، فلا يمكن أن نعطي نفس بدائل الإجابة لعينة تلامذة المرحلة الابتدائية والطالب الجامعي في قياس سمة معينة. مثلاً عندما يكون السؤال أو الفقرة موجهة لطالب في المرحلة الابتدائية أو المتوسطة أو شخص لم يحصل على فرصة كبيرة في التعليم فلا يصح أن نعطي بدائل إجابة متقاربة المعنى مثل (إلى حد ما، بدرجة قليلة، غالباً، عادةً، أبداً تقريباً، دائماً تقريباً) فهذا الشخص عادة لا يميز بين هذه الفروق القليلة وهكذا. ويمكن لمصمم المقياس أن يختار من بين هذه الأشكال كيف ما يشاء لتحقيق الغرض من القياس، كما يستطيع أن يستخدم أكثر من شكل في نفس المقياس، ومن بين هذه الأشكال:

1. اختيار إجابة واحدة من بين إجابتين، مثل: نعم أو لا.
2. الاختيار بين بدائل على متصل، مثل: موافق بشدة، موافق، محايد، معارض، معارض بشدة.
3. المطابقة، مثل: كل أسئلة التوصليل.
4. التكملة، مثل: كل العبارات الناقصة.
5. الاستجابة الحرة، مثل: التداعي على الصور، أو الكلمات.
6. إعادة الترتيب. (الزاملي، 2017، ص51)

المبدأ التوجيهي رقم 07: تطوير إجراءات إدارة الاختبار.

تتضمن إجراءات إدارة الاختبار (1) الوقت المتاح، (2) طريقة التطبيق الجماعي أو الفردي، و (3) منصة التطبيق (الكمبيوتر أو الورق / القلم). إن تحديد إطار زمني مناسب للممتحنين لإجراء الاختبار أمر بالغ الأهمية، وهناك عدة عوامل مثيرة للاهتمام. أولاً، يجب تحسين الغرض من الاختبار ومدة الاختبار لضمان دقة (أي صلاحية) الدرجات. على سبيل المثال، بالنظر إلى الهدف من الاختبار، ما هو الحد الأدنى لعدد المفردات التي يمكن إدارتها مع قياس القدرة المستهدفة بشكل كاف؟ عمر الممتحن هو أيضاً عامل مهم، في حالة الممتحنين الأصغر سناً يجب تخصيص مدة زمنية أطول لإدارة الاختبار. التعب هو عامل آخر يجب وضعه في الاعتبار.

المبدأ التوجيهي رقم 08: إجراء الاختبار التجريبي على عينة ممثلة للمجتمع.

يعتبر التطبيق التجريبي للاختبار فرصة ممتازة للباحثين للحصول على معلومات من الفحوصات المتعلقة بسلوكهم أثناء إجراء الاختبار. والهدفان الرئيسيان لمرحلة الاختبار التجريبي هما الحصول على معلومات إحصائية عن المفردات والحصول على تعليقات واقتراحات من الممتحنين بعد إجراء الاختبار في ظل الظروف الواقعية. (Price, 2017, p165-181)

فبعد إكمال عملية صياغة مجالات وفقرات المقياس تعرض بصيغتها الأولية على الخبراء، لإقرار صلاحيتها وإبداء ملاحظاتهم بحذف قسم من الفقرات وتعديل أخرى وتجزئة قسم منها، ثم تأخذ نسبة إتفاق آراء الخبراء أو المحكمين، ونستطيع أن نأخذ نسبة إتفاق (50%) فما فوق، وبعد إجراء التعديلات اللازمة على الفقرات، نبقى على الفقرات المميزة على ضوء آراء الخبراء أو المحكمين لإقرار صلاحيتها النهائية. (الزاملي، 2017، ص51)

المبدأ التوجيهي رقم 09: إجراء التحليل الإحصائي للفقرات والتحليل العاملي.

وتشمل تحليل المفردات مجموعة من التقنيات الإحصائية التي توفر أساساً لاختيار أفضل المفردات. يتيح إجراء تحليل المفردات طريقة للباحثين للكشف عن المفردات (1) الغامضة، (2) لا تكتب بشكل صحيح، (3) السهلة أو الصعبة جداً، و (4) المفردة التي لا تميز بشكل جيد بين المفحوصين. أهداف الاختبار محرك المعايير التي يعتبر عنصر تحليل المفردة الأكثر أهمية.

المبدأ التوجيهي رقم 10: وضع معايير أو درجات تفسيرية.

وفي كثير من حالات الاختبار، إن لم يكن معظمها، فإن الممارسة المتكررة هي توفير معنى معياري لتحديد مقياس متصل درجات السمة. الدرجات المعيارية والمعروفة أيضاً باسم المعايير هي إحصاءات وصفية تمكن من إجراء مقارنات لدرجة معينة مع الدرجات التي حصل عليها أشخاص آخرون في مجموعة محددة جيداً. وعموماً، تستخدم المعايير بطريقتين. أولاً، تستخدم المعايير لتمكيننا من تصنيف الممتحنين أو الأشخاص إلى فئات مثل المقبولية أو المرغوبية. ثانياً، غالباً ما تستخدم المعايير لتصنيف الممتحن أو الشخص وفقاً لمعيار محدد أو مثال إكلينيكي.

المبدأ التوجيهي رقم 11: كتابة الدليل التقني والوثائق المرتبطة به.

تطوير الدليل التقني هو نشاط ختامي في عملية تطوير الاختبار ويوفر توثيقاً شاملاً لجميع العمليات والإجراءات المستخدمة لتطوير الاختبار. عملية كتابة الدليل الفني يجبر المرء على إعادة النظر بدقة وتقويم جميع الإجراءات المستخدمة في تطوير اختبار. وعلاوة على ذلك، تساعد الأدلة التقنية في التقويمات الخارجية لأي اختبار يقوم به باحثون مستقلون. ويوفر الدليل التقني مصدراً

موجزا للخصائص السيكومترية وأدلة صدق الأختبار ، وينبغي أن يكتب بالتفصيل الكافي للسماح للقارئ بتشكيل حكم واضح حول صرامة وكفاية الإجراءات المستخدمة في كل خطوة. يتضمن الدليل التقني كذلك التوثيق المنهجي لجميع المكونات الهامة لعملية تطوير الاختبار. وينبغي أن تتضمن مكونات الدليل التقني على الأقل (1) موجزا لعملية تطوير الاختبار، بما في ذلك الأساس السيكومتري والاستخدام المقصود للدرجات؛ (2) إجراءات تطبيق الاختبار، بما في ذلك التدريب الضروري للمشرفين أو مستخدمي الاختبار؛ (3) التدرج باستخدام النظرية الكلاسيكية في القياس أو نظرية الاستجابة للمفردة (IRT) و معلومات عن إجراءات التدرج؛ و(4) المعلومات المعيارية وتقارير عن الدرجات أي جداول المعايير والتحليلات التكميلية مثل دراسات الصدق (الصلاحية) ودراسات الثبات (الموثوقية) ونتائج التحليل العاملي. (Price, 2017, p165-181)

المحاضرة الرابعة:

التقنين Standardization والمعايير Norms لدرجات الاختبارات النفسية والتربوية

تمهيد:

يحتل مفهوم التقنين والمعايير مكانة هامة في أي عملية لتطوير وبناء الاختبارات والمقاييس النفسية والتربوية، فالتقنين هو محاولة لتوحيد اجراءات التطبيق والتصحيح والتفسير لدرجات الاختبارات والمقاييس، أما المعايير هو إعطاء معنى للدرجات الخام المتحصل عليها من عملية القياس، لهذا سنتطرق في هذه المحاضرة لمناقشة هذين المفهومين، وإلى أهم الإجراءات التي تستخدم في عملية بناء وتطوير الاختبارات والمقاييس النفسية والتربوية من خلال توظيف مفهوم التقنين والمعايير.

أولاً: التقنين Standardization:

ويقصد به أن يكون بناء وتصحيح وتفسير نتائج الاختبار أو أداة القياس مستنداً إلى قواعد محددة بحيث تتوحد فيه وتتحدد بدقة مواد الاختبار وطريقة تطبيقه، وتعليمات إجابته وطريقة تصحيحه أو تسجيل درجاته، وبذلك يصبح الموقف الاختباري موحداً بقدر الإمكان لجميع الأفراد في مختلف الظروف. والأسلوب الموضوعي المنظم في بناء الاختبار يعني أن يتم اختيار مفرداته أو كتابتها بطريقة منظمة تتفق مع مواصفات محددة للاختبار، وأن تمثل هذه المفردات السمة المراد قياسها. ومفهوم التقنين ينطبق على الظروف المثالية التي يستطيع فيها مصمم الاختبار أو أداة القياس ضبط جميع المتغيرات ما عدا المتغير الذي يهدف للاختبار لقياسه، والطبع يصعب عملياً تحقيق ذلك بهذا القدر من الدقة.

ولعل مفهوم التقنين يشبه البحث العلمي الذي يتميز بقابلية التكرار وتداول النتائج وعموميتها، فإذا أردنا مقارنة النتائج التي يحصل عليها باحثين يجب أن تكون أدوات القياس وتعليمات استخدامها موحدة للجميع لكي يتسنى جمع الملاحظات العلمية في ظروف مضبوطة، وهو ما يعرف في البحث العلمي بالضبط التجريبي.

فالاختبار أو أداة القياس النفسي والتربوي يمكن تمثيلها بتجربة علمية مصغرة تضبط فيها جميع جوانب الموقف الاختباري واستبعاد العوامل المغتربة لجميع الأفراد، فإذا ما انتظمت هذه الشروط، فإنه يمكن أن نستنتج بقدر كبير من الثقة أن الدرجة التي يحصل عليها الفرد في الاختبار تمثل مقدار السمة أو القدرة المقاسة لديه. (علام، 2000، ص29)

1- أهم الشروط الواجب توفرها في الاختبارات والمقاييس المقننة:

1-1- أن يكون للاختبار تعليمات تعطى لجميع المفحوصين بطريقة موحدة وعلى أن تشمل هذه التعليمات الهدف من الاختبار، الشروط الخاصة بالأداء، الأخطاء الشائعة، الزمن المخصص للأداء، كيفية الإجابة، الأداء على الاختبار.

2-1- أن يكون للاختبار مفتاح تصحيح (اختبارات الورقة والقلم) لتعيين الإجابات الصحيحة على أسئلة أو وحدات للاختبار.

3-1- أن يكون للاختبارات معاملات ثبات وصدق معلنة وصريحة بالنسبة لمجموعات الأفراد الذين اعد لهم الاختبار في الأصل مجموعات أو عينات التقنين المرجعية.

4-1- أن يكون للاختبار معايير (جداول مستويات) تظهر بوضوح درجات ومستويات أداء عينة التقنين الأصلية على الاختبار. (الزامي، 2017، ص80)

2- مميزات الاختبارات والمقاييس المقننة:

1-2- إنها تتمتع بمعاملات صدق وثبات مقبولة بالنسبة لعينة التقنين التي أعدت لها في الأصل.

2-2- إن لها كراسة تعليمات توضح كافة المعلومات اللازمة لتطبيق الاختبار.

3-2- إن لها معايير تعكس مستويات أداء عينة (مجموعة) التقنين الأصلية.

4-2- إن لها مفتاح تصحيح يوضح اتجاهات العبارات (الأسئلة الموجبة والعبارات السالبة) لتعيين الدرجات الخاصة بكل إجابة. (الزامي، 2017، ص80)

5-2- تعرف هذه الاختبارات في معظم الأحيان باسم الاختبارات المنشورة وذلك لكونها تحظى بالنشر من قبل الدوريات والمراجع العلمية المتخصصة والتي تتمتع بسمعة محلية وعالمية.

والباحثون يستخدمون الاختبارات المقننة للأسباب التالية:

1 - توفر عليهم بذل الجهد والوقت لوضع اختبارات قد تحقق أو لا تحقق الهدف الذي يسعون إليه.

2 - تمكنهم من إجراء مقارنة بين أداء الأفراد عندهم مع أفراد آخرين طبق عليهم الاختبار نفسه.

3- خطوات عملية التقنين:

1-3- تحديد عينة التقنين.

2-3- تطبيق المقياس.

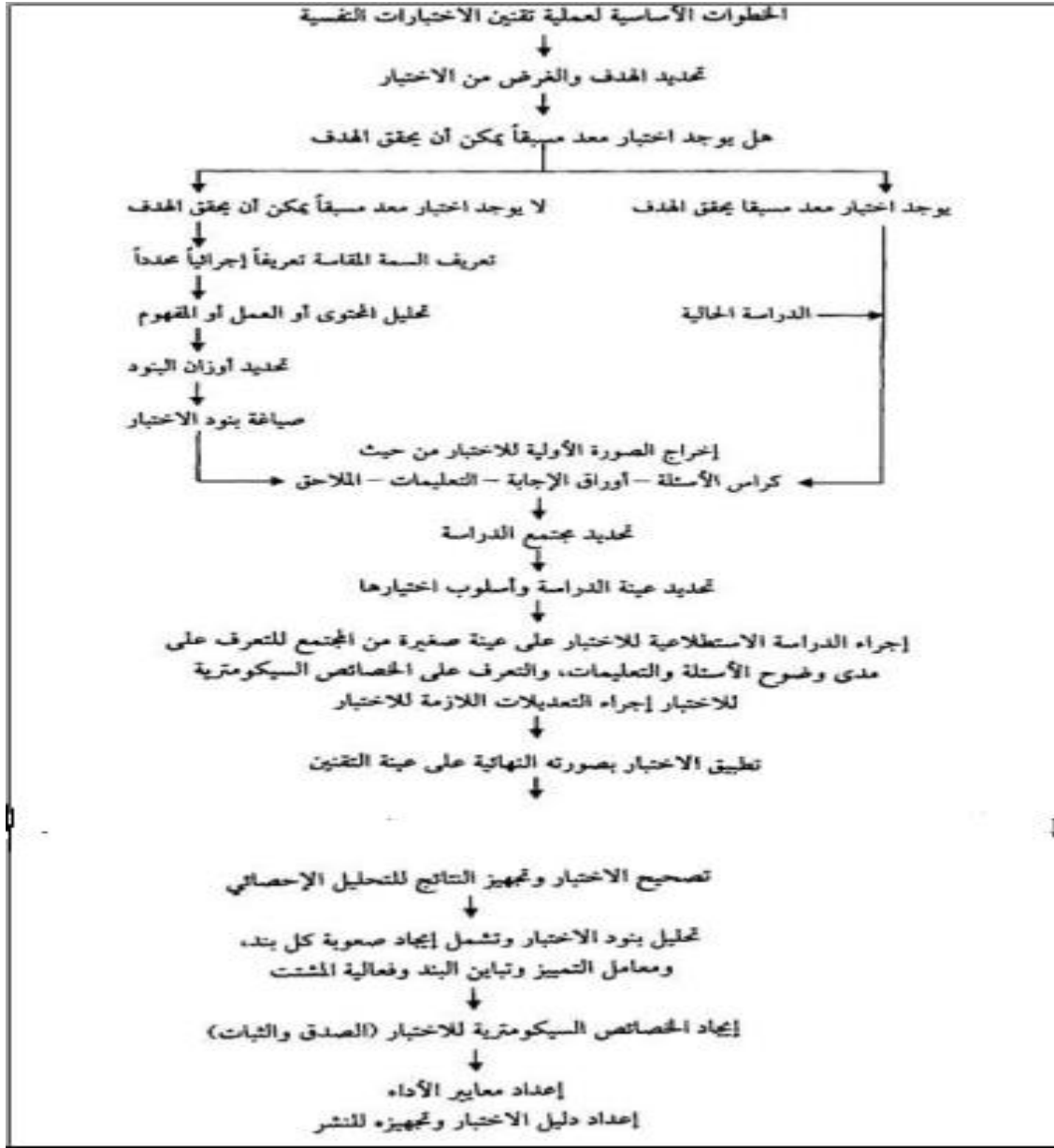
3-3- تصحيح المقياس.

4-3- استخراج القوة التمييزية للفقرات.

5-3- المعاملات العلمية للمقياس (الصدق والثبات والموضوعية).

6-3- الخطأ المعياري للمقياس.

3-7- اشتقاق المعايير للمقياس. (الزالمي، 2017، ص 81)



شكل (07) يوضح الخطوات الأساسية لعملية تقنين الاختبارات النفسية والتربوية.

(سوسن، 2014، ص 89)

ثانياً: المعايير Norms:

يحتاج الباحثين في مجال قياس السلوك البشري في مرحلة من مراحل بناء وتقنين الاختبارات للبحث عن وحدات ذات معنى ودلالة يمكن تفسيرها ومقارنتها، فالمشكلة هنا هي في وضع مستويات نستطيع بواسطتها أن نقارن ونفسر ونقوم السلوك الإنساني بهذه الوحدات والتي نسميها بالمعايير.

1- الحاجة للمعايير:

الحصول على الدرجات الخام من نتائج الطلاب على اختبار ما وكذلك نسبة الطلبة الناجحين في الاختبار إلى المجموع الكلي ليس لها في ذاتها معنى أو دلالة. فهي لا تحدد مركز الشخص بالنسبة لمجموعته أو بالنسبة لعينة التقنين، ولهذا لا بد من الرجوع إلى معيار يحدد معنى هذه الدرجة أو هذه النسبة المثوية. وهكذا فإن هذا المعيار سيدلنا على مركز الشخص بالنسبة للجماعة، هل هو متوسط، فوق المتوسط، أقل من المتوسط وما مدى هذا البعد. (سوسن، 2005، ص 147)

2- خصائص المعايير:

1-2- أن يكون للدرجة الواحدة معنى موحد من اختبار لاختبار.

2-2- أن تكون هذه الوحدات متساوية بحيث أن (10) درجات على جزء من اختبار تدل على نفس الشيء الذي تعنيه (10) درجات على جزء آخر من نفس الاختبار. (سوسن، 2005، ص 147)

3- المعايير والتقنين:

يشق المعيار من عينة التقنين التي تمثل المجتمع الأصلي المدروس في الخطوات التجريبية التي يمر بها المقياس في صورته النهائية. فتكون الدرجات المحولة للدرجات الخام لأفراد هذه العينة مصدراً للمعيار. أي أننا عندما نقارن درجة المفحوص في اختبار ما نعود إلى من يكافئه في عينة التقنين. وبناءً على ذلك فإن المعايير هي نتائج إجراء الاختبار على عينات التقنين. وعند نقل الاختبار من بيئة إلى أخرى يجب تقنينه على عينات موازية لعينات التقنين الأصلية لاشتقاق معايير جديدة. (سوسن، 2005، ص 147) فقد تطرأ على الأفراد جراء الظروف الثقافية والتكنولوجية والاقتصادية... الخ، مما يتطلب تقويم هذه الاختبارات والمقاييس من وقت لآخر وعلى فترات زمنية.

4- أهداف المعايير:

1-4- تحديد مركز الفرد بالنسبة لعينة التقنين.

2-4- يمكن من خلالها مقارنة الفرد على مقياس بمركزه على مقياس آخر.

وهناك طريقتان تربط بهما درجة الفرد بإطار عام وهي:

1- المقارنة بسلسلة متدرجة من الجماعات ونبحث عن أتباعه التي تكافؤه عن طريق متوسطها وكل جماعة تمثل إما صفا دراسياً معيناً أو عمراً زمنياً معيناً.

2- نحدد موضع الفرد في جماعة معينة بالإشارة إلى النسبة المئوية من الجماعة التي يمتاز بها أو الرجوع إلى متوسط الجماعة وانحرافها المعياري. (سوسن، 2005، ص148)

5- الحاجة إلى التعبير:

التعبير هو تقنية رياضية تتمثل في توزيع أو تقسيم الدرجات المتحصل عليها عند تطبيق الاختبار أو المقياس على مجتمع محدد إلى عدة أقسام، ويسمح التعبير بتحديد مكانة الفرد داخل المجتمع، آخذين العلامة أو الدرجة المتحصل عليها في الاختبار بعين الاعتبار، كما يتم تجميع درجات جميع أفراد المجتمع في هذا التعبير عن طريق التكميم والتعبير الانحرافي.

ومن الأفضل أن تتمتع الوحدات أو المعايير بالخصائص الآتية:

• أن يكون للدرجة الواحدة معنى موحد من اختبار إلى آخر، حتى نوفر أساساً يمكن به أن نقارن درجات الأفراد في اختبارات مختلفة.

• أن تكون هذه الوحدات متساوية بحيث ان العدد المعين من الوحدات وليكن 10 درجات مثلاً، على جزء من اختبار ما يدل على نفس الشيء الذي تعنيه 10 درجات على جزء آخر من نفس الاختبار، بحيث يمكننا أن نقول إن الفرد الأول يزيد عن الفرد الثاني بـ10 درجات مثلاً، وهذه الزيادة تساوي الشخص الرابع عن الشخص الثالث.

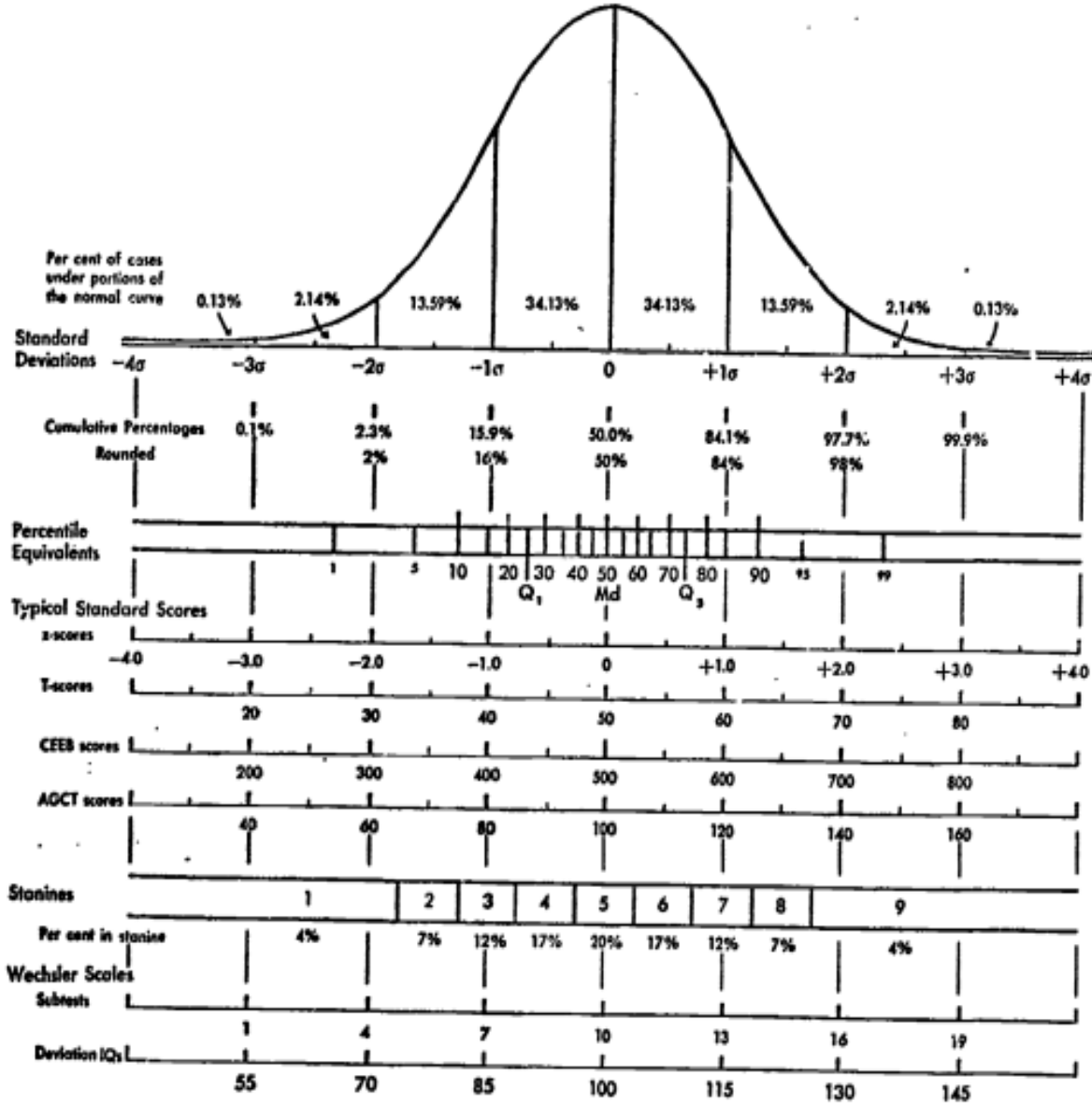
• أن تكون هناك نقطة صفر حقيقية، صفر مطلقة، تعبر عن لا شيء من الصفة التي نقيسها، بحيث نستطيع أن نقول إن درجة معينة ضعف درجة أخرى وأن ثلاثة تعتبر ثلثي درجة رابعة. (سامية، 2018)

وقد أثبتت الدراسات في هذا السياق بأنه إذا غيرنا في ظروف تطبيق الاختبار فإننا نؤثر بهذا على النتائج التي تصبح فيما بعد غير قابلة للتحقيق ولا للمقارنة بنتائج أخرى نكون قد تحصلنا عليها في ظروف أخرى مغايرة.

6- أنواع المعايير أو الدرجات المحولة:

تستخدم الدرجات المحولة في تحديد المركز النسبي للفرد في توزيع ما، بحيث يمكن وصف أداؤه بالنسبة لأقرانه في اختبار يقيس سمة معينة. ولعل هذا مرتبط ارتباطاً وثيقاً بالاختبارات مرجعية الجماعة أو المعيار Norm-Referenced Tests، فهذه الاختبارات تهدف إلى تقدير أداء الفرد بالنسبة لأداء الأفراد الآخرين في السمة أو القدرة التي يقيسها الاختبار، ويكون الاهتمام هنا منصباً على ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم في الاختبار، ويكون للدرجة تفسيراً فقط في ضوء متوسط أداء الجماعة أو بما يسمى بمعيار الجماعة Group Norm، وتصبح الدرجة لا معنى لها ويصعب تفسيرها إذا أغفلنا ردها إلى هذا المعيار. (علام، 2000، ص97)

وتوجد أنواع مختلفة من الدرجات المحولة لعل أكثرها استخداماً في القياس التربوي والنفسي الدرجات المعيارية Standard Scores ، والدرجات التائية T-Scores. ويعتمد هذان النوعان من الدرجات على التحويلات الخطية لدرجات الاختبار باستخدام المتوسطات الحسابية والانحرافات المعيارية معاً. والتي يمكن أن نلخصها في الشكل (08) التالي:



الشكل (08) يوضح أنواع الدرجات المعيارية المحولة ممثلة على منحنى التوزيع الطبيعي. ويمكن تلخيص بعض أنواع الدرجات المعيارية المحولة في الجدول التالي:

جدول (03) يوضح الدرجات المعيارية والدرجات المعيارية المحولة أو التائية ودرجة نسبة الذكاء الانحرافية (DQI).

Écart à la moyenne (z)	-3σ	-2σ	-1σ	Moyenne 0	+1σ	+2σ	+3σ	
	99.72 % de l'échantillon							
	95.44 % de l'échantillon							
			68.3 % de l'échantillon					
%	0.13 %	2.14 %	13.52 %	34.13 %	34.13 %	13.52 %	2.14 %	0.13 %
Note T	20	30	40	50	60	70	80	
Notes NS	1	4	7	10	13	16	19	
Notes QI	55	70	85	100	115	130	145	

(Bernaud, 2014, p133)

1-6- معايير العمر الزمني Age Norms:

تسمى هذه المعايير أحياناً العمر التعليمي Educational Age، أو مكافئ العمر الزمني Age Equivalent. وتفيد هذه المعايير إذا توقعنا أن نمو الفرد يحدث نتيجة لعوامل النضج. فإذا كان محمد تلميذاً في الصف الرابع الابتدائي وطوله 138 سم فربما نعتبره أكثر طولاً بالنسبة لتلاميذ هذا الصف، غير أنه ربما يكون من الأفضل مقارنته بالأطفال الذين في نفس مستواه العمري بدلاً من مقارنته بتلاميذ فرقته. فإذا كان عمره الزمني 9 سنوات، 11 شهراً، ومتوسط الطول لأطفال هذا العمر (الجماعة المرجعية) 140 سم فإنه يعد أقصر من متوسط طول أقرانه لذلك إذا كان لدينا معايير العمر الزمني أو معايير مكافئ العمر للأطوال، فإنه يمكننا تحديد المجموعة العمرية المناظرة لطول محمد وبالتالي يمكن تقييم طوله. فلكي نحدد مكافئ العمر لطول محمد، ينبغي أن نحدد جماعة مرجعية متوسط. فإذا افترضنا أن عمر أفراد هذه المجموعة 9 سنوات، 5 شهور، فإن محمد حينئذ يكون 9 سنوات، 5 شهور، وتكتب (9 - 5). (علام، 2000، ص252)

ومن عيوب هذه المعايير أن بعض الصفات لا تتناسب فيها الزيادة بزيادة العمر كالوزن والذكاء والطول. كما أنها لا تفيد في تقدير سمات الأفراد المتطرفين زيادة أو نقصاناً. كما لا يمكن مقارنة منحنى النمو للسمات المختلفة مع بعضها لاختلاف سرعة النمو من صفة لأخرى لدى الفرد.

2-6- معايير الفرق الدراسية Grad Norms:

و تسمى كذلك بمعايير الصف الدراسي وهي لا تختلف في جوهرها عن معايير العمر، وتحدد هذه المعايير متوسطات درجات أي اختبار أو مقياس ما بالنسبة للفرق الدراسية المتتابة، والخطوات التالية توضح طريقة حساب هذه المعايير:

- يطبق الاختبار على عينة شاملة لطلبة الفرق الدراسية المتتابة، كأن يطبق مثلاً على طلبة الفرق الأولى والثانية والثالثة بالمرحلة الثانوية.
- يحسب متوسط الدرجات لكل فرقة أو صف دراسي، أي متوسط درجات طلبة السنة الأولى ومتوسط الدرجات لكل فرقة أو صف دراسي، أي متوسط درجات طلبة السنة الأولى ومتوسط درجات طلبة السنة الثانية ومتوسط درجات طلبة السنة الثالثة.
- نرسم خطاً بيانياً نبين فيه العلاقة بين الفرق الدراسية ومتوسطات الدرجات بحيث يدل المحور الأفقي على الفرق الدراسية ويدل المحور الرأسي على متوسطات الدرجات.
- يستخدم الرسم البياني السابق لقراءة المعايير الدراسية لطلبة المرحلة الثانوية بالنسبة لذلك الاختبار.

وهكذا نرى أن هذه الطريقة لا تختلف عن الطريقة الزمنية إلا في نسبتها لمتوسطات الدرجات إلى الفرق الدراسية يدل إن كانت تنسب إلى الأعمار الزمنية. (علام، 2000، ص254)

3-6- المعايير المئينية Percentile Norms:

تعتبر المعايير المئينية أو الرتب المئينية **Percentile Ranks** من أكثر المعايير استخداماً وخاصة في اختبارات القدرات العقلية وتتميز بأنها أكثر مرونة وأوسع مدى في تطبيقها عملياً. وهي تحدد النسبة المئوية للحالات التي تقع بعد درجة معينة، فإذا كانت درجة الفرد في اختبار ما هي 60 درجة وحصل 80% من أفراد عينة مماثلة له على درجات أقل من 60 تكون درجته المئينية هي 80 وهذا معناه أنه أفضل من 80% من أفراد العينة المماثلة. وتحسب الرتبة المئينية للدرجة الخام باستخدام المعادلة التالية:

$$\text{الرتب المئينية لدرجة خام} = \frac{\text{عدد الطلبة الذين تقل درجاتهم عن الدرجة الخام أو تساويها}}{\text{العدد الكلي للطلبة}} \times 100$$

فمثلاً لدينا درجات لمجموعة من الطلبة في إحدى الصفوف الدراسية في اختبار مادة الرياضيات لكي نحسب الرتبة المئينية:

48، 40، 50، 38، 27، 30، 25، 33، 22، 20، 45، 35، 42، 34، 29، 31، 47، 19، 26، 21.

ولحساب الرتبة المئينية للدرجة 34 نرتب الدرجات الخام تصاعدياً على النحو التالي:

19، 20، 21، 22، 25، 26، 27، 29، 30، 31، 33، 33، 34، 35، 38، 40، 42، 45، 47، 48، 50.

ثم نحسب عدد الطلبة الذين تقل درجاتهم عن الدرجة 34 أو تساويها، فتبين أن عددهم يساوي 12 طالباً.

$$60 = 100 \times \frac{12}{20} = (34) \text{ إذن الرتبة المئينية المقابلة للدرجة الخام}$$

وهذا يعني أن الطالب الذي حصل على الدرجة 34 هو أعلى من 60% من طلاب صفه وأقل من 40% منهم. كما أن الطالب الذي رتبته المئينية 70% أفضل أداءً أو تحصيلاً من الطالب الذي رتبته 60%. ومع ذلك فإن معرفة أفضلية أداء طالب عن طالب يظل ناقصاً لقيمة هذه الأفضلية وهو ما تعالجه القيم المعيارية والقيم التائية. (أبونا هية، 1994، ص 399)

ويمكن حساب الرتبة المئينية المقابلة لدرجة معينة من بيانات مجمعة في توزيع تكراري باستخدام المعادلة التالية:

$$\text{الرتبة المئينية} = \frac{\text{التكرار المتجمع } T + \left(\frac{S - T}{f} \right) \times 100}{N}$$

حيث أن التكرار المتجمع T = التكرار المتجمع الصاعد للفئة السابقة للفئة التي تحتوي الدرجة S .
 S = الدرجة المطلوبة لإيجاد الرتبة المئينية المقابلة لها.

S = الدرجة المقابلة للحد الأدنى الحقيقي للفئة التي تحتوي الدرجة S .
 f = طول الفئة.

T = عدد الحالات الواقعة في الفئة التي تحتوي الدرجة S .

مثال: حصل أحد المعلمين بعد تطبيق اختبار تحصيلي في مادة العلوم على 35 طالباً من طلاب صفه. وسوف نستخدم هذه البيانات في حساب الرتبة المئينية للدرجة 10.

جدول (04) يوضح توزيع الدرجات التي حصل عليها 35 طالباً في اختبار تحصيلي.

الفئات	الحدود الحقيقية	التكرار	التكرار المتجمع الصاعد
3 – 1	3 – 0.5	4	4
6 – 4	6 – 3.5	5	9
9 – 7	9 – 6.5	6	15
12 – 10	12 – 9.5	8	23
15 – 13	15 – 12.5	5	28
18 – 16	18 – 15.5	4	32
21 – 19	21 – 18.5	3	35

$$46.66 = 100x \frac{8x(\frac{9.5-10}{3}) + 15}{35} = 10$$

ومن الطبيعي أنه يجب أن يكون هناك أكثر من معيار مئيني للاختبار الواحد إذا كانت الصفة أو السمة التي يقيسها الاختبار تتأثر بالسن أو المستوى التعليمي أو نحو ذلك، يمثل كل معيار منها إحدى سنوات العمر المتتالية التي يستخدم الاختبار في حدودها أو احد المستويات التعليمية... الخ، حتى تقاس درجة الفرد على المعيار المئيني الذي يمثل سنه أو مستواه. (أبو ناهية، 1994، ص 400)

وللحصول على المئيني نستخدم نفس الطريقة المستخدمة في تعيين الوسيط والأربعي، فكما قسمنا المجموعة إلى قسمين لتحديد النقطة الوسطى في حالة الوسيط، وكما قسمناها إلى أربعة أجزاء في حالة الأربعي، يمكن تقسيمها إلى مئة جزء في حالة المئيني، يكون المئيني هو النقطة التي تحدد هذه الأجزاء فالمئيني 10 هي النقطة التي يقل عنها 10% من القيم والمئيني 20 هو النقطة التي يقل عنها 20% والمئيني 50 هو النقطة التي يقل عنها 50% من القيم. ومتى تحددت القيم المقابلة لكل مئيني فإنها توضح عادة في شكل جدول للاستفادة منها في معرفة المئيني الذي يقابل أية قيمة يعطها الاختبار وبهذه الكيفية تستطيع تحديد موقع الفرد بالضبط بالنسبة لمجموعته.

مثال: لكيفية إيجاد المئيني 75%:

لدينا درجات في مادة الرياضيات لعشرة تلامذة كالاتي: 8: 6: 4: 12: 10: 12: 13: 20: 19: 17.
أولاً: ترتيب الدرجات تصاعدياً: 4: 6: 8: 10: 12: 12: 13: 17: 19: 20.

$$R = P/100(N+1)$$

بتطبيق المعادلة التالية:

حيث أن: P = المئيني الذي نريد حسابه.

N = عدد أفراد العينة.

$$R = 75/100(10+1) = 8.25$$

بتطبيق المعادلة نحصل على:

- نهتم بالرقم الذي بعد الفاصلة وهو 0.25.

- نبحث عن الدرجة التي رقمها 8 والتي تحصلنا عليها من خلال تطبيق المعادلة وهي الدرجة 17.

- نبحث عن الدرجة التي رقمها أعلى مباشرة من الرقم 8 وهي الدرجة رقم 9 وهي الدرجة 19.

- إيجاد الفرق بين الدرجتين السابقتين رقم (8:9) وهما 19 - 17 وتساوي 2.

- ضرب القيمة الأخيرة 2 في الرقم الذي بعد الفاصلة الذي عثرنا عليه بتطبيق المعادلة وهو 0.25.

$$0.5 = 2 \times 0.25$$

- وأخيراً نضيف القيمة الأخيرة المتحصل عليها وهي 0.5 إلى الدرجة التي رتبها رقم 8.

أي $17.5 = 0.5 + 17$. فالدرجة 17.5 هي إذن قيمة المئيني 75% . (Harsh and Dheeraj, 2022)

وبالتالي فالطالب الذي درجته 17.5 هو أعلى من 75% من أفراد العينة الذين ينتمي إليهم. فالمعايير المئينية تتميز بسهولة استخدامها فضلاً عن تحديد ترتيب الفرد ضمن المجموعة بدرجة كبيرة من الدقة هما السبب في انتشار استخدام المئينيات في الاختبارات النفسية والتربوية المختلفة التي يلائمها هذا النوع من المعايير.

ويرتبط مفهوم الرتب المئينية مفهوم آخر هو المئين، والمئين هو درجة تقل عنها أو تقابلها نسبة مئوية معينة من الأفراد، فدرجة المتعلم التي تقابل المئين الخامس والعشرون بالنسبة لطلبة صفه تشير إلى أن هذا المتعلم يتفوق على 25% من طلبة صفه ويقل عن 75% منهم.

ويمكن التوضيح بين المصطلحين على النحو التالي: عندما يكون المطلوب حساب الرتبة المئينية للقيمة (70) يكون المطلوب إيجاد النسبة المئوية من مجموع القيم التي تقع تحت القيمة (70). وعندما يكون المطلوب حساب المئين 70، يكون المطلوب إيجاد القيمة أو النقطة أو الحد الفعلي الأعلى الذي يقع تحته 70% من القيم. (النيمان، 2013، ص394).

4-6- الدرجات المعيارية Standar Score Norms:

ويمكن تعريف الدرجة المعيارية على أنها المسافة التي تبعد فيها الدرجة الخام عن الوسط الحسابي معبراً عنها في وحدات من الانحراف المعياري. وتسهل الدرجة المعيارية مقارنة المجموعات مع بعضها حتى وإن اختلفت متوسطاتها والتي تحسب بالمعادلة التالية:

الدرجة الخام - المتوسط الحسابي

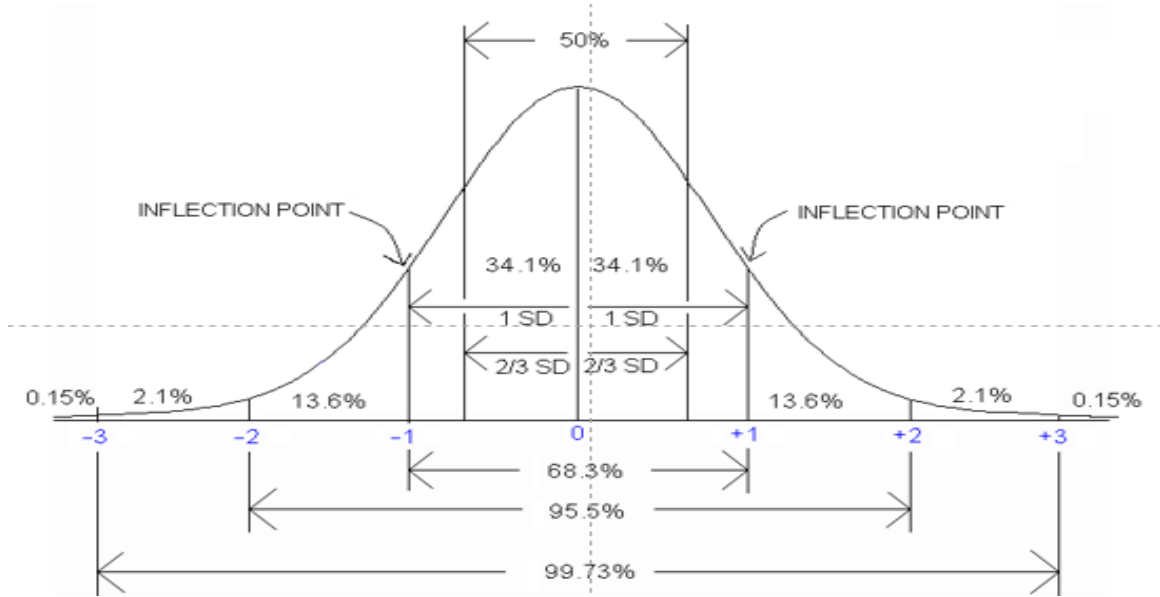
الدرجة المعيارية =

الانحراف المعياري

ومن هنا يتضح أنها تستخدم الانحراف المعياري في حين أن المعايير السابقة لم تستخدمه وغالباً ما تقع درجات التوزيع بين (3+) و(3-) درجة معيارية، فالدرجة الخام التي تكون أعلى من المتوسط تكون إشارتها موجبة والأقل من المتوسط تكون إشارتها سالبة. وبالتالي فإن الدرجة المعيارية إذن هي المسافة التي تبعد بها الدرجة عن المتوسط كما يعبر عنها بوحدات من الانحراف المعياري. والشكل

(02) التالي يوضح التوزيع الاعتمادي المعياري والنسب: (سوسن، 2005، ص151)

ومن مزايا هذا المنحنى الاعتمادي المعياري Standard Normal distribution أنه يمكن استخدامه في تحويل الدرجات المعيارية إلى مئينيات Percentile، أي نسبة مئوية من عدد الأفراد الذي تقل درجاتهم عن درجة معيارية معينة.



شكل (09) يوضح التوزيع الاعتدالي المعياري Standard Normal distribution والنسب

ويمكن إعطاء المثال التالي لتوضيح العملية:

فمثلاً لدينا الدرجات الخام على اختبار ما وهي: 35-30-40-50-55 ، والمتوسط الحسابي لها هو 40 والانحراف المعياري 10. بتطبيق المعادلة السابقة لحساب الدرجة المعيارية فإننا نحصل على البيانات الموضحة في الجدول التالي:

جدول (05) يوضح كيفية حساب الدرجة المعيارية المقابلة للدرجة الخام.

الدرجة المعيارية	الدرجة الخام- المتوسط	الدرجة الخام
1.5	15	55
1	10	50
صفر	صفر	40
1-	10-	30
0.5-	5-	35

يتضح من خلال الدرجات المعيارية المتحصل عليها في الجدول أعلاه نجد أن بعضها موجب وبعضها سالب، بالإضافة إلى الصفر أيضاً، فكيف نفسر الدرجة المعيارية صفر سواء للطالب أو لولي الأمر على أساس أنها تمثل الدرجة المتوسطة. وهناك مشكلات أخرى تتمثل في كون بعض الدرجات المعيارية تظهر في صورة كسر عشري، بالإضافة إلى ضيق مدى توزيع الدرجات المعيارية الذي لا يزيد عن 6 درجات مما يقلل من الفروق الفردية بين المتعلمين ويجعلها تافهة أو معدومة وخاصة في

العينات الكبيرة. وهذا كله يزيد من صعوبة استخدام الدرجات المعيارية، ويكمن الحل في هذه الحالة باستخدام الدرجات المعيارية المعدلة. (أبوناهاية، 1994، ص 395)

ورغم فوائد الدرجة المعيارية فهي لا تصلح لعملية المقارنة إلا إذا كان التوزيع إعتدالياً أو قريباً منه ونظراً لكثرة درجاتها السالبة فقد عمد إلى تعديل الدرجة المعيارية للتخلص من الإشارات السالبة وتسوية المنحنى، حيث تكون جميع الدرجات المعيارية موجبة، باستخدام الدرجات المعيارية المعدلة مثل:

1-4-6-الدرجات المعيارية المعدلة أوالدرجات التائية:

الدرجة المعيارية المعدلة تساعدنا على التخلص من المشكلات التي تنشأ من استخدام الدرجة المعيارية والفرق بينهما يكمن في أننا نختار للدرجة المعيارية المعدلة متوسطاً جديداً بدلاً من الصفر (المتوسط الحسابي للدرجات المعيارية) وإنحرافاً معيارياً جديداً بدلاً من الواحد الصحيح (الانحراف المعياري للدرجات المعيارية) أيضاً للتخلص من الأعداد السالبة والكسور العشرية. ويتم ذلك بضرب الدرجة المعيارية في 10 للتخلص من الكسور، وإضافة القيمة 50 للمتوسط الحسابي للتخلص من الإشارات السالبة، وبذلك تصبح الدرجات المعيارية درجات معيارية معدلة (T.Score) المبينة في الجدول التالي:

جدول (06) يوضح الدرجات المعيارية المعدلة.

الدرجة المعيارية المعدلة	الدرجة الخام المعيارية	الدرجة المعيارية
65	$50+10 \times 1.5$	1.5
60	$50+10 \times 1$	1
50	$50+10 \times 0$	0
40	$50+10 \times 1-$	1-
45	$50+10 \times 0.5-$	0.5-

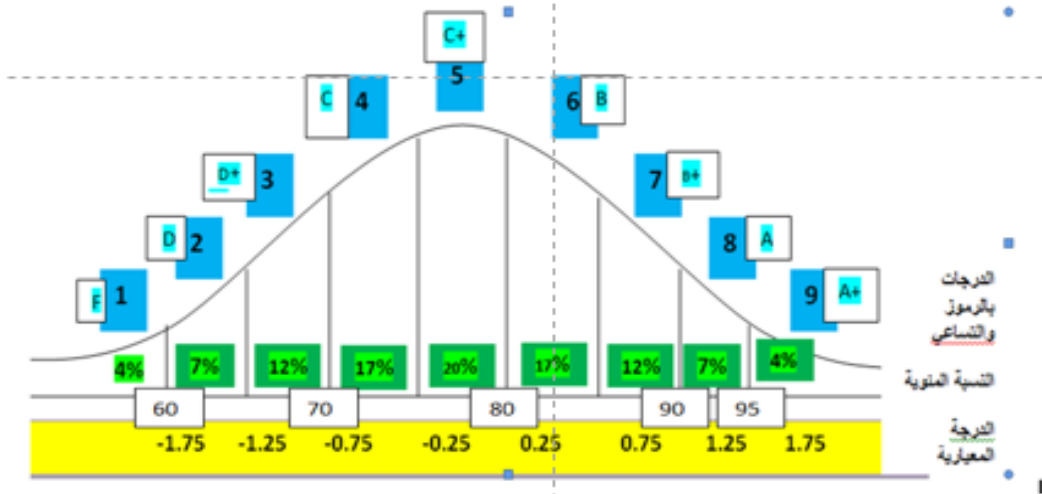
ومع ذلك يبقى أن نشير إلى أنه بالإمكان اختيار متوسط آخر جديد وكذلك انحراف معياري جديد، على أن هذا الاختيار يبقى محكوماً بفكرة أن الحصول على الدرجة المعيارية المعدلة يتم بضرب الدرجة المعيارية بمقدار ثابت للتخلص من الكسور العشرية وإضافة مقدار ثابت آخر لإزالة الإشارة السالبة. وهناك العديد من التحويلات إلا أن الدرجة التائية تبقى أحد أهم طرق تحويل الدرجات الخام إلى درجات معيارية معدلة، ومن بين كذلك التحويلات نذكر:

2-4-6- الدرجات الجيمية:

وهدفها هو إيجاد درجات معيارية تناسب بعض الاختبارات التي تقيس وظائف تكون الفروق بين الأفراد واضحة. وتعديل فيها الدرجات المعيارية بحيث يكون المتوسط (5) والانحراف المعياري (2)، وقسم التوزيع فيها (11) قسماً. فإذا كانت الدرجة المعيارية لفرد هي 0.4 فتصبح الدرجة الجيمية له (4) $(5.8=5+2 \times 0.4)$

3-4-6- الدرجات التساعية Stanine:

وتقوم على أساس الدرجة الجيمية حيث أن مستوياته أقل من تقسيم الدرجة الجيمية فبدلاً من أن يكون (11) قسماً قسم إلى (9) أقسام إذ أنه يجمع بين الدرجة الجيمية المساوية للصفر والدرجة الجيمية المساوية لواحد صحيح، كما يجمع بين الدرجة الجيمية (9) والدرجة (10) في درجة واحدة هي (9) ويوضح هذا المعيار إذا كان الأفراد المتطرفون متقاربين سلباً أو إيجابياً. (سوسن، 2005، ص 152) والذي يمكن أن نوضحه في الشكل التالي:



الشكل (10) يوضح التساعي stanine وتوزيع النسب وفق المنحنى الاعتيادي الى تسع فئات.

وتبرز أهمية هذه التحويلات في النقاط التالية:

- 1- تجعل المتوسط الحسابي والانحراف المعياري والدرجة كلها أعداد موجبة.
- 2- تجعل جميع الدرجات واقعة على تدرج مئوي ما بين الصفر والمائة.
- 3- توحيد المتوسط الحسابي والانحراف المعياري لجميع الدرجات يسهل المقارنة بين المتعلمين.

- مثال تطبيقي في تحويل الدرجات الخام إلى درجات معيارية معدلة أو درجات تأئية:

فإذا افترضنا مثلاً أن لدينا درجات طالبين من طلاب الصف المدرسي في اختبارات الحساب والعلوم واللغة العربية وكذلك المتوسط الحسابي والانحراف المعياري لدرجات هذه الاختبارات نتيجة

تطبيقها على جميع طلاب الصف، وهذه النتائج موضحة في الجدول التالي والسؤال المطروح هو أي الطالبين تحصيله أفضل؟

جدول (07) يبين درجات طالبين في ثلاث اختبارات تحصيلية

الطالب	الحساب	العلوم	اللغة العربية
الياس	72	92	50
مريم	70	68	75
المتوسط الحسابي	60	56	55
الانحراف المعياري	12	12	10

لإجراء المقارنة بين الطالبين لابد من تحويل الدرجات الخام إلى درجات تائية بحيث يكون المتوسط الحسابي لكل اختبار 50 والانحراف المعياري 10. ومع ذلك فمن مراجعة درجات كل طالب نجد أن الطالب الياس حصل على مجموع قدره 214 درجة في حين أن الطالبة مريم حصلت على مجموع قدره 213 درجة وبالتالي فمجموع الدرجات الخام لطالب الياس أعلى من مجموع الدرجات الخام للطالبة مريم، فهل هذا يعني أن أداء الياس أفضل من أداء مريم؟ هذا ما سيوضحه حساب الدرجات التائية:

الطالب (الياس):

$$60 = 50 + 10 x \frac{60-72}{12} = \text{الدرجة التائية لمادة الحساب}$$

$$80 = 50 + 10 x \frac{56-92}{12} = \text{الدرجة التائية لمادة العلوم}$$

$$45 = 50 + 10 x \frac{55-50}{10} = \text{الدرجة التائية لمادة اللغة العربية}$$

مجموع الدرجات التائية للطالب الياس تساوي $185=45+80+60$
الطالبة (مريم):

$$58 = 50 + 10 x \frac{60-70}{12} = \text{الدرجة التائية لمادة الحساب تقريباً}$$

$$60 = 50 + 10 x \frac{56-68}{12} = \text{الدرجة التائية لمادة العلوم}$$

$$70 = 50 + 10 x \frac{55-75}{10} = \text{الدرجة التائية لمادة اللغة العربية}$$

مجموع الدرجات التائية للطالبة مريم تساوي $188=70+60+58$

وهذا يعني أن تحصيل الطالبة مريم أفضل من تحصيل الطالب الياس. (أبو ناهية، 1994، ص 398)

المحاضرة الخامسة

الصفحة أو البروفايل النفسي والتربوي

Psychological Profile and Educational Profile

تمهيد:

يتضمن دليل أي اختبار نفسي أو تربوي بما يسمى الصفحة النفسية أو البروفايل وهو شكل يلخص أداء الأفراد على الاختبارات الفرعية لتسهيل المقارنة لأداء الفرد بالنسبة لجماعته وكذلك مقارنة الفرد مع ذاته.

أولاً: تعريف الصفحة أو البروفايل النفسي أو التربوي:

الصفحة أو البروفايل النفسي أو التربوي هو وصف شامل أو ملخص للمكانة النسبية للفرد والخصائص التي يتم الحصول عليها من خلال مقارنة نتائج تطبيق سلسلة من الاختبارات أو الاختبارات الفرعية. فعلى سبيل المثال، قد يُظهر بروفايل الشخصية البيانات التي تم جمعها حول اختبارات الشخصية وغيرها من الاختبارات ذات الأهمية، ويستخدم البروفايل لتقييم الفرد في المجالات المتعلقة بحياته الشخصية والتعليمية والمهنية، بما في ذلك المزاج وأساليب اتخاذ القرار وأسلوب الاتصال، والاتجاه العام نحو العمل والحياة. (APA, 2022)

تشتمل بعض الاختبارات على عدة اختبارات فرعية تقيس سمات أو قدرات أو مهارات متعددة تشكل جميعها بطاريات من الاختبارات Tests Batteries وتستخدم في أغراض محددة مثل الانتقاء أو التسكين أو التشخيص. ومن أمثلة ذلك بطارية الاستعداد الأكاديمي (SAT)، واختبار ويكسلر لذكاء الراشدين (WAIS)، واختبار ويكسلر لذكاء الأطفال (WISC)، واختبار الشخصية متعدد الأوجه (MMPI)، واختبار الميول المهنية (SVIB)، واختبار الاستعدادات الفارقة (DAT)، وغيرها من الاختبارات والمقاييس النفسية والتربوية.

ويكون الغرض من استخدام هذه البطاريات ليس فقط معرفة أدائه النسبي في اختبار على حدة، وإنما أيضاً مقارنة أدائه في الاختبارات المختلفة. (علام، 2000، ص 257)

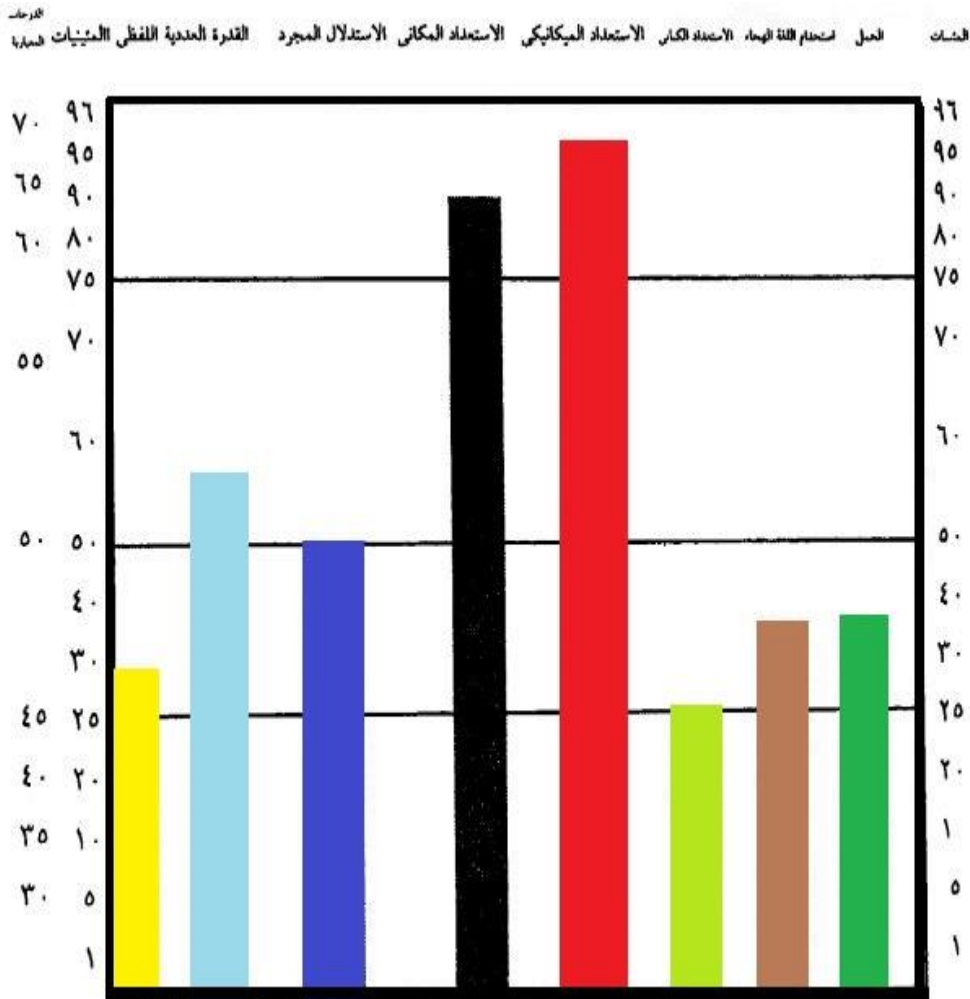
ولتوضيح الصورة العامة للأداء وتيسير عملية المقارنة يزود ناشرو الاختبارات مستخدمي هذه البطاريات ببطاقة أو ورقة رسم بياني معدة بطريقة خاصة تفيد في التمثيل البياني لدرجات الاختبارات

التي تشتمل عليها بطارية معينة؛ وذلك لتوضيح موقع الفرد أو مستوى أدائه في هذه الاختبارات التي تمثل أبعاداً نفسية أو معارف ومهارات تربوية معينة، وتحديد جوانب القوة والضعف. ويسمى هذا التمثيل النفسي أو التربوي الصفحة النفسية (البروفيل النفسي) Psychological Profile، أو الرسم البياني التربوي (البروفيل التربوي) Educational Profile. (علام، 2000، ص258)

وتبنى الصفحة النفسية أو التربوية لمجموعة من درجات الاختبارات لفرد ما بتمثيل الاختبارات على المحور الأفقي، ويمثل مقدار كل درجة بنقطة على المحور العمودي. ثم يصار إلى توصيل النقاط بخط مستقيم أو بأعمدة بيانية، كما يتم تحديد قيم المئينيات أو الدرجات المعيارية على المحور العمودي لتوضيح عملية المقارنة. (النهمان، 2013، ص410)

ويوضح الشكل (01) التالي صفحة نفسية لدرجات أحد الطلاب في بطارية اختبارات الاستعدادات الفارقة (DAT) Differential Aptitude Tests مستمدة من دليل هذه البطارية الذي تم إصداره من طرف المؤسسة السيكولوجية الأمريكية سنة 1963. (علام، 2000، ص259)

والغرض الأساسي من هذه البطارية تقديم معلومات تتعلق ببروفيل قدرات متعددة لاستخدامه في التوجيه والإرشاد التربوي للطلاب في المرحلتين المتوسطة والثانوية لكي يتخذ كل منهم في ضوء ذلك قرارات تتعلق باختياراته المهنية المستقبلية. (نهي ورعدة)



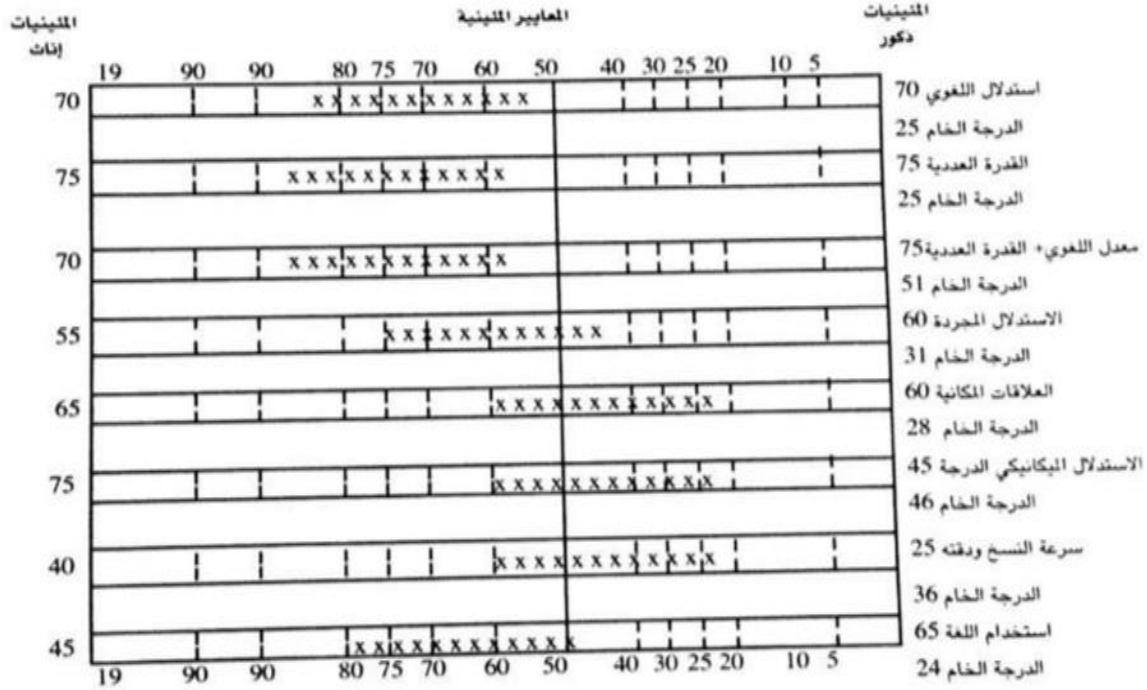
شكل رقم (11) يوضح صفحة نفسية لدرجات طالب في اختبار الاستعدادات الفارقة (DAT).

ويتضح من خلال الشكل (11) أن الصفحة النفسية تحدد المئينيات والدرجات المعيارية المقابلة للدرجات الخام للطلاب في الاختبارات الفرعية الثمانية التي تشتمل عليها بطارية الاستعدادات الفارقة (الاستدلال اللفظي)؛ القدرة العددية؛ الاستدلال المجرد؛ الاستدلال الفضائي (المكاني)؛ الاستدلال الميكانيكي؛ السرعة الكتابية والدقة الإملائية؛ التهجئة؛ استخدام اللغة)، وتمثل الأعمدة الرأسية هذه الدرجات. (علام، 2000، ص259)

ففي الصفحة النفسية توجد ثلاثة خطوط أفقية تمتد من النقط المقابلة لكل من المئيني 25، والمئيني 50، والمئيني 75، أي الأرباعيات الأول والثاني والثالث. وذلك ليتسنى مقارنة درجات الطالب فيما بينها بمجرد النظر إليها، وكذلك يمكن مقارنتها بدرجات الطلاب الآخرين. وتفيد هذه الصفحة

النفسية في تحديد جوانب القوة والضعف في استعدادات الطالب، والاستناد إليها في عملية التوجيه والإرشاد التربوي.

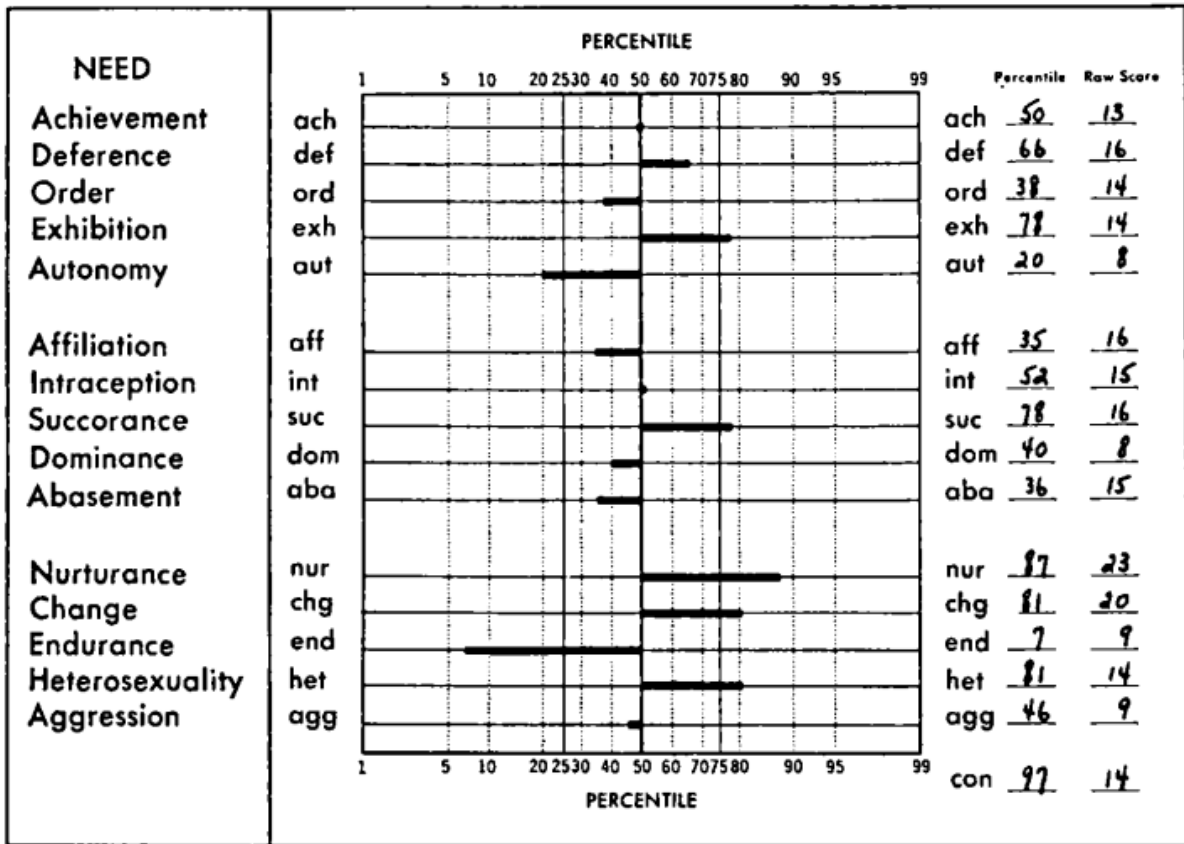
فإذا نظرنا إلى الصفحة النفسية للطالب في الشكل (01) نلاحظ أنه حصل على أعلى الدرجات في الاختبار الفرعي للاستعداد الميكانيكي، حيث تقابل الرتبة 95، تليها درجته في الاختبار الفرعي للاستعداد المكاني المقابلة للرتبة المئينية 87 تقريباً، وحصل على أقل الدرجات في الاختبار الفرعي للاستعداد الكتابي، حيث تقابله الرتبة المئينية 25 أي الإرباعي الأول، وبذلك تتبين جوانب قوته وضعفه. (علام). وهناك مثال ثاني يوضحه الشكل (02) لبروفيل أحد طلاب الصف الأول ثانوي في بطارية اختبارات الاستعدادات الفارقة (DAT)، والذي يعتمد على معايير المئينيات وفق الجنس (ذكور، إناث)، عكس البروفيل السابق الذي ضم المئينيات وكذلك الدرجات المعيارية للعينة الكلية بدون تفصيل في المعايير بحسب الجنس (ذكور، إناث). (نهي ورغدة)



شكل (12) يوضح بروفيل أحد طلاب الصف الأول ثانوي في بطارية اختبارات الاستعدادات الفارقة (DAT).

وتتميز الصفحات النفسية أو الرسوم البيانية التربوية بخصائص ثلاث وهي: المستوى Level، والانتشار Scatter، والشكل Shape. "المستوى" يعني الارتفاع العام للأعمدة في الرسم البياني وهو يدل على مستوى الأداء العام للفرد أو المتوسط العام لدرجاته المعيارية في الاختبارات الفرعية، و"الانتشار"

يعني درجة تشتت مختلف النقط التي تمثل الدرجات حول الموقع المتوسط (المستوى) لدرجات جميع الاختبارات الفرعية، ويدل ذلك على مدى اختلاف درجات الفرد فيما بينها. أما "الشكل" فيعني نمط الاستعدادات المتمثل في ترتيب درجات الاختبارات التي حصل عليها الفرد. وهذه الخصائص يمكن أن تختلف إلى حد ما فيما بينها كل على حدة، إذ يمكن أن تتفق صفحتان نفسيتان لفردين في "المستوى" و"الانتشار"، ومع ذلك تختلف كل منهما عن الأخرى في "الشكل"، أي في ارتفاع وانخفاض النقط التي تمثل درجاتهم في اختبارات البطارية. وهناك مثلاً كذلك لبعض الصفحات النفسية، نوردها كآلاتي:

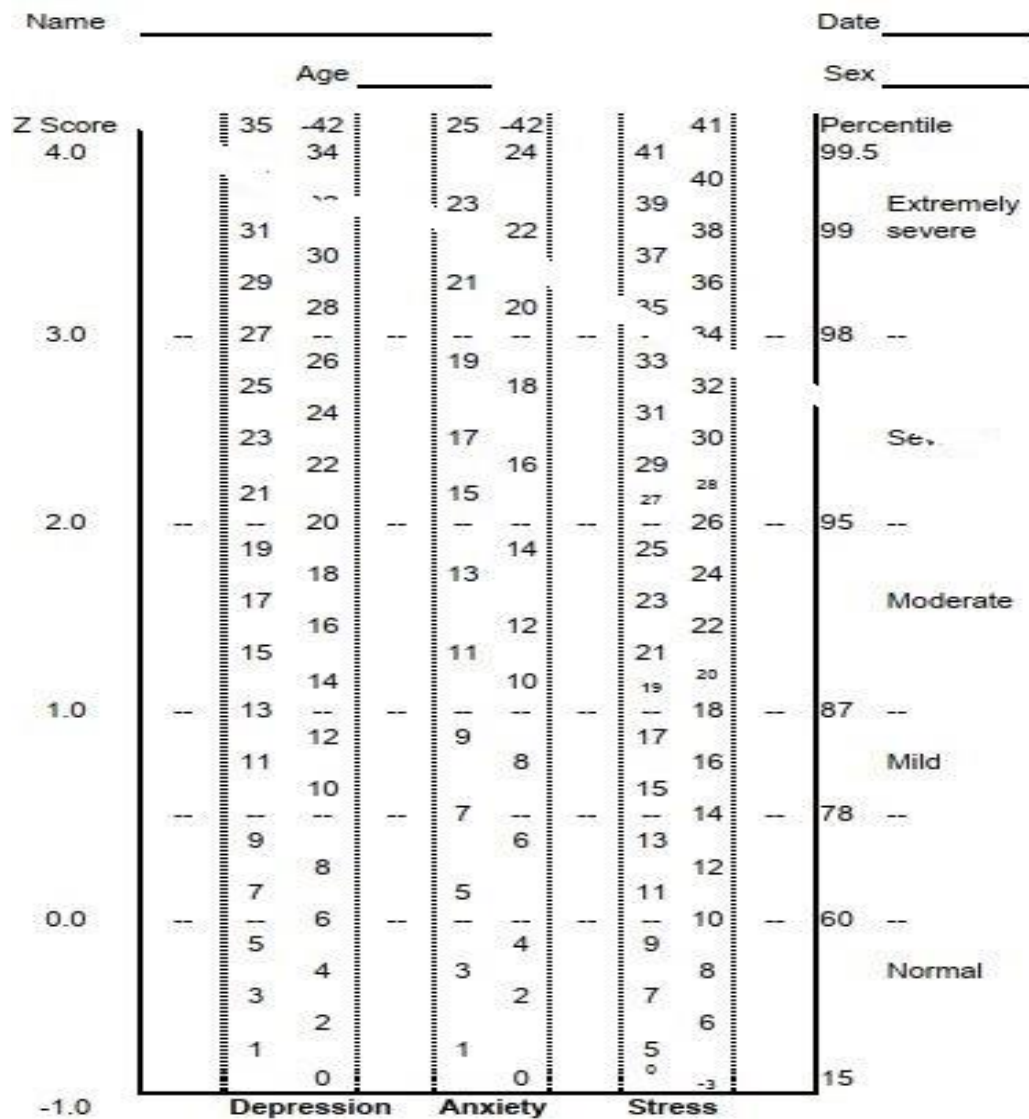


شكل (13) يوضح صفحة نفسية (بروفيل) لقائمة إدواردز للتفضيلات الشخصية (EPPS) ل(Edwards, 1959).

حيث تعتمد مفردات قائمة إدواردز للتفضيلات الشخصية Edwards Personality Preference Schedule ل Allen Louis Edwards (1914-1994). على نظرية موراي Murray في الاحتياجات. تتكون الاختبار نفسه من 125 زوجاً من العبارات، يختار الممتحن أي العبارات تنطبق عليه بشكل أفضل من خلال وضع دائرة حول العنصر "أ" أو "ب" في كل زوج، من خلال الجمع بين الدرجات، يعتقد "إدواردز"

أنه يمكنه الإدلاء ببيانات حول شخصية المتقدم للاختبار في مجالات تم ترجمتها من مفاهيم خاصة بنظرية "موراي" ومن هذه الحاجات الحاجة إلى: الإنجاز؛ والإذعان أو الانقياد؛ والنظام؛ والاستعراض؛ والاستقلالية؛ والانتماء؛ والشعور بالحالة الداخلية؛ وللمساعدة والإغاثة؛ والهيمنة أو السيطرة؛ وتجنب جلد الذات؛ والتغذية؛ والتغيير؛ والتحمل؛ وللجنس؛ والعدوان. (Allen, 1958, pp591-597)

DASS Profile Sheet



الشكل (14) بروفييل أو صفحة نفسية لمقياس الاكتئاب والقلق والضغط النفسي

(DASS) (Lovibond, 1995)

يتضح من الشكل (14) والذي يوضح بروفيل أو صفحة نفسية لمقياس الاكتئاب والقلق والضغط النفسي (DASS) أنه خصص لكل مقياس فرعي عموداً به الدرجات المتوقع الحصول عليه، حيث يمكننا رسم خطأ أفقياً من خلال النتيجة التي يتم الحصول عليها لكل مقياس فرعي، ثم توصيل الخطوط المنقطه لتشكيل رسم بياني شريطي. تكون ارتفاعات الأشرطة بوحدات درجة زائفة (Z) ويمكن مقارنتها ببعضها البعض ومع ملاحظة أن التحويل إلى نسب مئوية أو المئينيات على المحور الأيمن تقريبي فقط. كما يمكن أن نحدد كذلك من خلال هذا البروفايل مستويات الاضطرابات الانفعالية السلبية (الاكتئاب؛ القلق؛ الضغط النفسي) شديدة جداً؛ شديدة؛ معتدلة؛ خفيفة؛ عادية.

ولا تقتصر الصفحات النفسية والتربوية على نوع واحد وإنما تتعدد أنواعها وطرق تمثيلها. فالنوع الذي تم تقديمه يمثل درجات بطارية اختبارات تطبق على فرد واحد في وقت واحد لإعطاء صورة إستاتيكية مستعرضة لدرجاته، في حين أنه يوجد نوع آخر يسمى الصفحات النفسية الارتقائية أو الطولية Longitudinal Profiles حيث تمثل درجات الاختبار نفسه أو عدة اختبارات في أزمنة مختلفة لتوضيح نمو الفرد في سمة أو سمات متعددة. فمثلاً يمكن تمثيل درجات تلميذ في اختبارات القراءة والحساب والمعلومات العامة من الصف الثالث إلى الصف السادس في بروفايل تربوي Educational Profile واحد للإفادة منه في التعرف على التقدم الدراسي للتلميذ وتوجيهه توجيهاً تربوياً سليماً. وعادة يوضح مؤلف بطارية الاختبارات أمثلة لصفحات نفسية لبعض الحالات وكيفية استخدامها في الإرشاد والتوجيه التربوي أو النفسي لهذه الحالات للاسترشاد بها في استخلاص المعلومات من صفحة نفسية أخرى. (علام، 2000، ص260)

ثانياً: الخطأ المعياري للمقياس وتفسير الفروق في الدرجات:

غير أنه يجب مراعاة أن الفرق بين درجتين في اختبارين مختلفين يتأثر بأخطاء القياس المتعلقة بكل منهما، ولذلك فإن ثبات هذا الفرق يكون أقل من ثبات درجات كل اختبار على حدة. ونظراً لأن الصفحات النفسية أو التربوية تعكس الفروق بين درجات اختبارات فرعية مختلفة، فإنه ينبغي الحيلة قبل أن تعزى هذه الفروق إلى الاختلافات بين درجات الاختبارات الفرعية التي تمثلها الصفحة النفسية أو التربوية وبخاصة إذا كان عدد مفردات هذه الاختبارات قليل مما يؤدي إلى انخفاض ثبات درجاتها. (علام، 2000، ص260)

كما يجب مراعاة بعض العوامل الأخرى لكي لا تقدم الصفحة النفسية صورة غير صادقة عن سمات الفرد، إذ ينبغي أن تستخدم الجماعة المرجعية نفسها في إعداد معايير الاختبارات الفرعية، أما إذا استخدمت جماعات مرجعية مختلفة، فإننا لا نستطيع المقارنة المباشرة بين هذه الدرجات. كما ينبغي أن تمثل درجات مجموعة الاختبارات في الرسم البياني بميزان أو تدرج واحد، وأن يعبر عن الدرجات بالرتب المئينية أو أي نوع من أنواع الدرجات المحولة المقابلة لها. كما يفضل استخدام معيار لا يضحك الفروق الضئيلة أو يصغر الفروق الكبيرة. ويمكن التغلب على ذلك إلى حد ما بأن تتضمن الصفحات النفسية مؤشراً يبين مقدار أخطاء القياس لكي لا يساء تفسير الدرجات.

من المهم بخاصة مراعاة الاختبار وأخطاء القياس عند تقييم الفروق بين درجتين والتفكير من حيث المدى الذي ربما تتذبذب في نطاقه كل درجة يفيد في التحقق من المغالاة في تأكيد الفروق الضئيلة بين الدرجات وهذه الحيطة مرجوة عند مقارنة درجات الأفراد المختلفين، وكذلك عند مقارنة درجات نفس الفرد في قدرات مختلفة وبالمثل، فإن التغيرات في الدرجات بعد عملية التعليم أو أي متغيرات تجريبية أخرى تحتاج إلى أن تفسر في ضوء أخطاء القياس. (أناستازي، 2015، ص143-146)

والسؤال الشائع عن درجات الاختبارات يتعلق بالموقع النسبي للفرد في مجالات مختلفة فهل "أحمد" قدرته اللغوية أكبر من قدرته العددية؟ وهل "إلياس" لديه استعداد لأنشطة ميكانيكية أكبر من استعداده لأنشطة لفظية؟ وإذا حصل "أحمد" على درجات في الاختبار الفرعي اللفظي من بطارية استعدادات، أكبر من درجته في الاختبار الفرعي العددي، وحصل "إلياس" على درجة في الاستعداد الميكانيكي أكبر من درجته في الاستعداد اللفظي، فكيف يمكننا التأكد من أنهما سوف يظان هكذا عند إعادة الاختبار بصيغة أخرى للبطارية؟ وبعبارة أخرى هل كان من الممكن أن تكون الفروق بين الدرجات قد نتجت من مجرد الانتقال العشوائي لمفردات معينة في الاختبارات اللفظية، والعددية، والميكانيكية المستخدمة؟ وهذه الأسئلة موائمة بخاصة للتفسير المناسب للدرجات في البطاريات متعددة الدرجات في كل من القدرات وسمات الشخصية (Anastasi, 1985a).

والمهم أن نتذكر أن الخطأ المعياري للقياس للفرق بين درجتين مثلاً الفرق بين درجات اختبار الاستدلال المجرد والاستدلال الميكانيكي أكبر من خطأ القياس لأي من الدرجتين، وهذا يتأتى من حقيقة أن هذا الفرق يتأثر بأخطاء الصدفة الموجودة في كل من الدرجتين ويمكن إيجاد الخطأ المعياري للفرق بين درجتين من الأخطاء المعيارية للقياس لكل من الدرجتين بواسطة الصيغة التالية:

$$\sqrt{(SEM_1)^2 + (SEM_2)^2} = \text{الخطأ المعياري للفرق}$$

حيث أن: (SEM_1) و (SEM_2) هما الأخطاء المعيارية للمقياس لكل من الدرجتين، وبالتعويض في المعادلة $SE = \sigma\sqrt{1 - r_{11}}$ للخطأ المعياري (SEM_1) ، و $SE = \sigma\sqrt{1 - r_{22}}$ للخطأ المعياري (SEM_2) ، فإنه يمكننا إعادة كتابة الصيغة مباشرة باستخدام معاملات الثبات كالتالي:

$$SE = \sigma\sqrt{2 - r_{11} - r_{22}} = \text{الخطأ المعياري للفرق بين درجتين}$$

وقد استخدم في هذا التعويض نفس الانحراف المعياري (σ) لكل من الاختبارين الاستدلاليين المجرد1 والاستدلال الميكانيكي2، نظراً لأن درجاتهما سوف يعبر عنها في ضوء نفس الميزان أو متصل السمة أو القدرة قبل إمكانية المقارنة بين الاختبارين. (أناستازي، 2015، ص143-146)

فمثلاً يمكن توضيح الأجراء السابق من خلال نسبي الذكاء اللفظي والذكاء الأدائي المستمدتين من مقياس ويكسلر لذكاء الراشدين- الطبعة الرابعة (WAIS-R) فقيمتا ثبات كل من هذه الدرجات بطريقة التجزئة النصفية هما القيمة 0.97 والقيمة 0.93 على التوالي، ويعبر عن نسب الذكاء الانحرافية لهذا المقياس على ميزان متوسطه 100، وانحرافه المعياري 15. ولذلك فإن الخطأ المعياري للفرق بين هاتين الدرجتين (نسبي الذكاء) يمكن إيجاد قيمته كما يأتي:

$$SE = 15\sqrt{2 - 0.97 - 0.93} = 4.74 \text{ الخطأ المعياري للفرق يساوي}$$

ولتحديد مدى كبر حجم درجة الفرق التي يمكن الحصول عليها عن طريق الصدفة عند مستوى دلالة (0.05)، فإننا نضرب الخطأ المعياري للفرق في (1.96%)، والنتيجة تساوي 9.29 أو 10 نقاط تقريباً، ولذلك، فإن الفرق بين نسبي الذكاء اللفظي والذكاء الأدائي لفرد معين في مقياس ويكسلر لذكاء الراشدين- الطبعة الرابعة (WAIS-R) ينبغي ألا يقل عن 10 نقاط لكي يعد فرقاً جوهرياً عند مستوى دلالة (0.05). (أناستازي، 2015، ص143-146)

ومن هنا يتضح أن المعايير Norms تعد أساساً لتفسير درجات الأفراد في الاختبارات والمقاييس مرجعية الجماعة أو المعيار، فتحويل الدرجة الخام التي يحصل عليها الفرد إلى نوع من أنواع المعايير التي تم التطرق إليها سابقاً والتي تمكنا من تفسير هذه الدرجة في ضوء المركز النسبي للفرد بين أقرانه المماثلين له فيما يقيسه الاختبار.

وبالتالي المعايير ليست مطلقة أو مستقرة بل هي معايير نسبية تعتمد اعتماداً أساسياً على جماعات مرجعية معينة، وتتأثر هذه المعايير تأثراً كبيراً بتغير خصائص هذه الجماعات التي يقارن بها الفرد في سمة معينة أو مجموعة من السمات التي يقيسها الاختبار. ونظراً للخصائص النسبية للمعايير وعدم استقرارها مما يحد من عمومية استخدامها في تفسير درجات الاختبارات والمقاييس النفسية والتربوية، فقد حاول علماء القياس التوصل إلى أسلوب جديد يهدف لجعل خصائص أدوات القياس مستقلة عن خصائص جماعات مرجعية معينة في إطار نظرية معاصرة يطلق عليها نظرية الاستجابة للمفردة الاختبارية (IRT) Item Response Theory. (علام، 2000، ص261)

المحاضرة السادسة

الأساليب الإسقاطية

Projective Techinques

تمهيد:

ظهر لأول مرة لفظ "إسقاط" في علم النفس عند (فرويد) وذلك في مقالة له عن عصاب القلق سنة 1894، حيث أوضح فيها أن عصاب القلق يظهر عندما تشعر الذات بعجزها عن السيطرة على المثيرات الجنسية، وفي هذه الحالة تسلك النفس وكأنها تسقط هذه المثيرات على العالم الخارجي. (الفاعوري، ص05)

أولاً: الاختبارات الإسقاطية:

استخدم "فرويد" الإسقاط Projection في مواضع مختلفة ليشير إلى أحد ميكانزمات الأنا الدفاعية حيث تعزى من خلاله الرغبات اللاشعورية وما يرتبط بها من قلق إلى موضوعات أو مصادر خارجية بدلا من ربطها بأسبابها الحقيقية، وذلك في محاولة من الأنا لضبط القلق المرتبط بتلك الرغبات عند فشله في السيطرة عليها بطرق أكثر سواء. واعتمادا على هذا المفهوم بدأ علماء التحليل النفسي استخدام مفهوم الإسقاط في القياس للإشارة إلى الوسائل غير المباشرة أو الغامضة التي يمكن استخدامها لكشف رغبات ومشكلات الفرد وسماته الشخصية في علاقتها الديناميكية دون أن يلتفت إلى ذلك كنتيجة لتحرر الخبرات اللاشعورية من رقابة الأنا (Freud, 1931; Freud, 1946; , 1992). (الغامدي، 1424هـ، ص02)

وكنتيجة للتطور في مجال القياس النفسي تنوعت الاختبارات الإسقاطية مما حدا بالعلماء إلى تصنيفها بطرق مختلفة، ومن ذلك على سبيل المثال تصنيفها تبعا لطبيعة المثير إلى اختبارات تعتمد على اللغة أو الصور أو الرسم و تبعا لطبيعة الاستجابة إلى اختبارات ارتباطية كاختبار الرورشاخ، واختبارات بنائية كاختبار تفهم الموضوع حيث يعتمد المفحوص فيما إلى إنتاج قصة، واختبارات الإكمال ومنها اختبارات تكملة الجمل الناقصة أو تكملة القصص أو حتى تكملة الرسوم، واختبارات إعادة التركيب أو البناء حيث يطلب من المفحوص اختيار صور أو كلمات و إعادة بناءها في صورة ما، و اختبارات التعبير حيث لا يكون هناك مثيرا خارجيا ومنها اختبارات إنتاج الرسم كرسم الشخص الإنساني (Gregory, 1992). (الغامدي، 1424هـ، ص02)

تسعى المقاييس الإسقاطية إلى الحصول على عينات من السلوك التي ينظر إليها على أنها تزودنا في الوقت الواحد بمعلومات عن حشد من الجوانب المختلفة لبناء الشخصية والعوامل الديناميكية الخاصة بها، والصفة المشتركة بينها تتمثل في كونها تقدم للمفحوص مجموعة من المواد المثيرة تتميز بالغموض وعدم البنية. وتطلب منه أن يضيف عليها بنيانه الخاص، ويكون الناتج في حالتها نظاماً مركباً يتم تحليله فيما بعد من خلال الاسترشاد بمجموعة من التعليمات التي تكون في العادة أكثر تفصيلاً من مادة القياس، ويتم تفسير هذا النظام عادة بالنسبة للعديد من أوجه الشخصية.

والأدوات الإسقاطية تقوم على أرضية نظرية تختلف بدرجة ما عن الأرضية النظرية التي تقوم عليها الاختبارات الموضوعية، فالأدوات الإسقاطية تركز على نظرية التحليل النفسي وتبنى مفهوم الإسقاط بوصفه المفهوم المركزي لها، كما يمكن أن نلاحظ فيها تأثيرها بالنظرة الغشتالتية للشخصية التي تتعامل مع الشخصية بوصفها كلاً، أو وحدة واحدة متكاملة، وليست مجرد مجموعة من السمات، أو كما لا محدوداً من المثيرات والاستجابات. (الفاعوري، ص06)

ثانياً: تصنيف الأدوات الإسقاطية:

جرت عديد المحاولات لتصنيف الأدوات الإسقاطية ولعل من أبرزها تصنيف ليندزي (Lindzey, 1961-1959) الذي اعتمد نموذج استجابة المفحوص أساساً له، ووضعها ضمن الفئات الخمسة التالية:

- 1- أساليب التداوي: وتتطلب من المفحوص أن يستجيب بأقصى سرعة للمثيرات بإعطاء الكلمات أو الصور أو المدركات التي تخطر له.
- 2- الأساليب البنائية: وفيها يطلب من المفحوص تكوين قصة أو رسم صورة دون أي قيد على استجابته.
- 3- أساليب التكملة: وفيها يطلب من المفحوص أن يكمل مهمة ناقصة بأي طريقة يراها مناسبة.
- 4- أساليب الاختيار أو الترتيب: وفيها يطلب من المفحوص إعادة ترتيب الصور، أو اختيار صورة أو أكثر من مجموعة من الصور، وفقاً لمحك معين، مثل المواءمة أو الفرق أو الاستحسان.
- 5- أساليب تعبيرية: وفيها يطلب من المفحوص القيام بدور نشط كالرسم أو الطلاء أو أنشطة اللعب كوسيلة للتعبير عن شخصيته، وهنا يكون التركيز على النتاج، وأسلوب أو نمط التعبير. (علام، 2000، ص625-626)

كم أنه من التصنيفات الشهيرة التصنيف الثنائي البسيط والذي يقسم إلى قسمين رئيسيين هما:

1- أسلوب يركز على المحتوى **Content - Centered**: يهتم هذا الأسلوب بتحليل استجابات الفرد في ضوء المشاعر والأحاسيس التي يعبر عنها، حيث يفترض أن يكون هذا المحتوى نتاج اهتمامات الفرد ويعبر عن تطلعاته، ومن أمثلة هذا الأسلوب اختبار تفهم الموضوع (TAT) حيث يتم تفسير الاستجابات التي يقدمها الفرد في ضوء القصص التي ينسجها على الصور التي تقدم إليه.

2- أسلوب يركز على الشكل **Form - Centered**: يعتمد هذا الأسلوب على تقديم الشكل، والذي غالباً ما يكون مهماً وغامضاً ويتطلب من المرء الاستجابة إليه من خلال ما يمثل لديه من معنى، وخير مثال على هذا النوع من الاختبارات اختبار الرورشاخ Rorschach، ويركز هذا الأسلوب على تقصي إجابات المفحوص من حيث الزاوية والشكل واللون الذي أثاره والذي شده أو دفعه للاهتمام به. (علام، 2000، ص625-626)

ثالثاً: الفروق الأساسية بين الأدوات الإسقاطية والمقاييس الموضوعية:

1- تتضمن المقاييس الموضوعية للشخصية مثيرات لفظية واضحة ومحددة، بخلاف الأدوات الإسقاطية والتي تتضمن مثيرات غير لفظية تتيح ظهور استجابات تنطوي على قدر ما من الغموض وعدم التحديد.

2- تتميز الأدوات الإسقاطية بقدرتها على التغلغل إلى أعماق المفحوص والتصدي الجوانب اللاشعورية، بينما تتناول الاختبارات الموضوعية للشخصية الجوانب الشعورية والسطحية فقط.

3- يضعف لجوء المفحوص إلى تزيف إجاباته في الأدوات الإسقاطية باعتبار هذه الأخير تسبر أعماق لا شعوره، على عكس الاختبارات الموضوعية للشخصية التي ترتفع فيها نسبة التزيف في الإجابة.

4- تقوم الأدوات الإسقاطية على أساس مفهوم الإسقاط بوصفه المفهوم المركزي لها كما يمكن أن نلاحظ فيها تأثيرها بالنظرة الجشثالتية للشخصية التي تتعامل الشخصية بوصفها كلاً، أو وحدة واحدة متكاملة، وليست مجرد مجموعة مع من السمات، أو كما لا محدوداً من المثيرات والاستجابات.

5- تتفوق الاختبارات الموضوعية للشخصية من حيث إمكانية تطبيقها على عينات كبيرة نسبياً، مما يتيح استخراج معايير إحصائية لها على درجة لا بأس بها من الدقة ويجعلها تتفوق على الأدوات الإسقاطية من حيث هذه الناحية. (الفاعوري، ص07-08)

رابعاً: الاختبارات الإسقاطية الأكثر شيوعاً:

وسنتطرق الآن لبعض الاختبارات الإسقاطية الأكثر شيوعاً في هذا المجال، كالآتي:

1- اختبار بقع الحبر لرورشاخ (Ink blot Test Rorschach, 1924):

يعد هذا الاختبار من أكثر الأدوات الإسقاطية شيوعاً، وقد وضع هذا الاختبار من قبل الطبيب النفسي السويسري (هيرمان رورشاخ) بعد تجارب طويلة بدأها في عام 1911. وامتدت لأكثر من عشر أعوام واستخدم فيها بقع الحبر للكشف عن خصائص الشخصية. وقد نشر (رورشاخ) دراساته الأولى حول هذا الاختبار عام 1921م، وطرحت الطريقة التي اقترحها في دراسة الشخصية عام 1924م، ثم تبع ذلك ظهور عدد من الإضافات المتلاحقة التي تركزت حول طريقة التطبيق وأساليب تفسير الإجابات وغيرها. (الفاعوري، ص 07-08)

يحتوي اختبار الرورشاخ على عشر بطاقات بيضاء ترتسم على كل منها بقعة كبيرة من الحبر تظهر على شكل نصفين متناظرين. وتظهر بقع الحبر في خمس من هذه البطاقات باللون الأسود وتدرج الرمادية وهي البطاقات ذوات الأرقام (1، 4، 5، 6، 7)، بينما تظهر في بطاقتين أخريين، وهما البطاقتان رقم (2، 3) بلمسات إضافية من اللون الأحمر الفاتح، في حين تجمع البطاقات الثلاث الباقية (وهي البطاقات ذوات الرقم (8، 9، 10) بين عدة ظلال ملونة. (فيصل، 2001، ص 196)

يقوم اختبار الرورشاخ على أساس افتراض العلاقة بين الإدراك والشخصية، حيث يعكس إدراك الفرد لبقع الحبر طبيعة ووظائفه السيكولوجية، وذلك من خلال استثارة البقع بغموضها لاستجابات مرتبطة بحاجات الفرد وخبراته السابقة وأساليبه المعتادة للاستجابة للمثيرات المختلفة، ذلك البقع ليست موضوعات مقننة اجتماعياً تستوجب استجابات محددة أو مقبولة ثقافياً. وعلى هذا الأساس فالاختبار يكشف عن سمات الشخص الوجدانية والسلوكية غير المتعلمة. وبالرغم من أن ذلك يؤدي إلى ضعف الارتباط بين السلوك المتوقع كاستجابات على الرورشاخ والسلوك الفعلي في المواقف الحياتية، فإنه لا يقلل من قيمة الاختبار الإكلينيكية كونه لا يهدف إلى التنبؤ بالسلوك بطريقة جزئية بل إلى وصف شخصية الفرد بشكل كلي له قيمته الإكلينيكية لفهم السلوك الملاحظ لما يقدمه للمعالج من معلومات ضرورية تمس الشخصية الأكثر عمقاً، مما يعني مساعدة الفرد على تحقيق تكيف أفضل وأكثر صحية. (الغامدي، 1424هـ، ص 09)

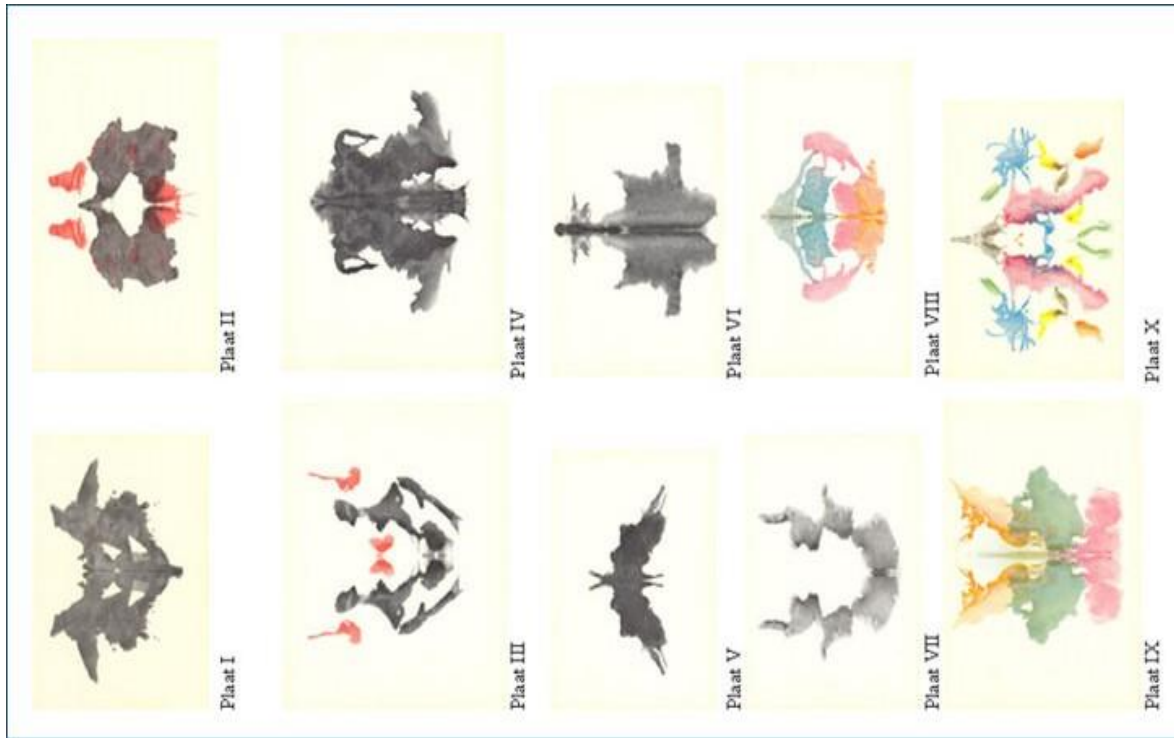
1-1- تصحيح الاختبار وتفسير نتائجه:

ظهرت أنظمة عديدة لتصحيح الاختبار وتفسير استجابات المفحوصين نحوها، لا بد في البداية أن نشير إلى أن تفسير نتائج هذا الاختبار يتطلب فاحصاً مدرباً تدريباً إكلينيكياً عالياً، نظراً لتنوع الانفعالات والاستجابات التي تصدر عن المفحوص، فعليه أن يتمتع بدقة عالية في ملاحظة حركات المفحوص أثناء الاستجابة ونوعية الانفعالات التي يصدرها والتغيرات الفيزيائية التي تصدر على وجهه. وترتكز الأساليب الأكثر شيوعاً بين هذه الأساليب على أربع فئات تمثل الأسس المعتمدة في تقدير استجابات المفحوصين وهي: التحديد المكاني أو الموضع Location، العوامل المحددة أو المحددات Déterminants، والمحتوى Content، بالإضافة إلى مؤشر الشيعوع. ثم يقدر الفاحص عدد الاستجابات، زمن الرجوع للاستجابة الأولى في كل بطاقة، ومن ثم نمط الاستجابات وتتابعها. (علام، 2000، ص 629)

فالموضع Location يشير إلى الجزء من بقعة الحبر الذي استند إليه المفحوص استناداً أساسياً في استجابته، إذ يمكن أن يستخدم المفحوص البقعة ككل، أو تفاصيل عامة أو تفاصيل بسيطة، أو الظلال البيضاء، ويمكن للفاحص التحقق من ذلك أثناء المرحلة الثانية. وموضع الاستجابات يعتقد بأنه مؤشر للمدخل العقلي لحياة الفرد من حيث اعتماده على تعميمات مترامية، أو يراعى التفاصيل بانتظام، أو تضيع لديه المعالم في تركيزه الشديد على قضايا جانبية غير ذات أهمية.

والمحددات Déterminants تشير إلى مظاهر البقعة التي تُحدد الاستجابة، مثل الشكل أم الصيغة، والظلال أم التركيب، واللون، والحركة. وتقدير درجات المحددات يُعد عملية معقدة وأكثر ذاتية من تقدير الموقع أو المحتوى، فلكل من هذه المظاهر تقديرات معينة ومؤشرات مختلفة.

والمحتوى Content يشير إلى القسم الذي ينتهي إليه المدرك. ويتم تقدير محتوى الاستجابات في ثلاثة أقسام تتعلق بالإنسان أو الحيوان أو المجردات كالعدوان أو العاطفة، بحسب مادة المثير الذي يراه المفحوص. وكذلك يمكن تقدير ذلك شيعوع محتوى إدراك الفرد أو أصالته فتكرار عدد معين من الاستجابات لكل بقعة حبر الذي يحدد شيعوعها أو أصالتها. ويختلف تفسير شخصية الفرد تبعاً للمحتوى الذي يراه في معظم البطاقات فمثلاً يمكن المقارنة بين تفسيرين ربما يتم التوصل إليهما من مجموعتين من الاستجابات؛ إحداهما رأى فيها المفحوص حيوانات تتصارع، ورأى في الأخرى أناساً يشاركون مشاركة كاملة في عمل ما. (أناستازي، 2015، ص 630) (ملحم، 2012، ص 315)



شكل (15) يمثل اللوحات العشر المكونة لاختبار هرمان رورشاخ (Rorschach, 1924).

2-1- مثال لتفسير الاستجابة البطاقة العاشرة:



البطاقة رقم (10)

الشكل (16) البطاقة العاشرة من اختبار رورشاخ الاسقاطي.

تبدو البقعة في هذه البطاقة كلوحة فنان مليئة بالألوان الموزعة على أجزاء متعددة منفصلة، ولهذا يجد غالبية المفحوصين صعوبة التعامل مع البقعة كوحدة واحدة فيما عد تلك لاستجابات مثل

"لوحة لفنان أو منظر تحت الماء". تساعد البطاقة على تقديم استجابات عن الحيوان في حالة حركة كما أنها نادرا ما تستثير استجابات ترتبط بالصور البشرية فيما عد المساحات القرنفلية الكبيرة في الجانبين. من الاستجابات الشائعة "ثعبان اخضر (أو دودة خضراء) للمساحة الخضراء المائلة للاستطالة في الأسفل أو سرطانات للبقع الزرقاء في الجانبين، أو رأس أرنب للجزء الصغير بين الثعابين". (الغامدي، 1424هـ، ص08)

ملاحظة: هناك بعض الاستجابات التي تكشف الاضطرابات: إما بسبب اللون أو بسبب تشتت البقع في أطراف متعددة من البطاقة، وهي عادة تشير إلى صدمة انفعالية قاسية، أو إلى حالة فصامية غير متكيفة مع الحياة الاجتماعية أو العمل.

2- اختبار تفهم الموضوع (TAT):

أعد اختبار تفهم الموضوع (TAT) The Thematic Apperception Test (TAT) كريستينا مورجان C. Morgan ، وهنرى موراي Murray عام 1935 بالعيادة النفسية بجامعة هارفارد كأسلوب للكشف عن الأفكار اللاشعورية والخيالات Fantacies. ويختلف هذا الاختبار عن اختبار الرورشاخ في أن مثيراته أكثر انتظاما وأقل غموضاً واستجاباته تكون أكثر تحديداً فهذا الاختبار يتطلب من الفرد أن يحكى قصصاً تتعلق بصور معينة بحيث يمكن للأخصائين من ذوي الخبرة تفسير الدوافع الكامنة لدى الفرد، وانفعالاته، وعواطفه وصراعاته وحاجاته النفسية التي افترضها موراي في نظريته للشخصية. وهذا الاختبار يلي اختبار رورشاخ في شيوع استخدامه، وفي عدد الدراسات والبحوث التي استثارها وقد استمد الاختبار أهميته أولاً في العيادات النفسية، ولكنه أصبح تدريجياً أداة بحث في علم النفس الارتقائي، وعلم النفس الاجتماعي، والشخصية، والدراسات الأنثروبولوجية عبر الثقافات، كما يستخدم في تقييم الشخصية في مجالات الإرشاد وعلم النفس الصناعي. (علام، 2000، ص633)

ويتكون الاختبار من 30 صورة باللونين الأبيض والأسود مرسومة على بطاقات عبارة عن 4 مجموعات متداخلة تتكون كل منها من 19 بطاقة إضافة إلى بطاقة بيضاء، وتختلف كل صورة من حيث محتواها ودرجة غموضها. وتتباين الصور في درجة انتظامها وبنيتها، إذ تتراوح بين الانتظام التام أو البنية المحددة بدرجة كبيرة إلى البنية غير المحددة على الإطلاق كما في البطاقة البيضاء. ويمكن تطبيق الاختبار على الراشدين والمراهقين وتعرض البطاقات على الفرد واحدة تلو الأخرى مصحوبة بتعليمات بسيطة، ويُطبق الاختبار عادة على مرتين أو أكثر، وإجراءات التطبيق المعتاد هو أن يُطلب من المفحوص أن يحكى قصة كاملة تتعلق بكل من 19 بطاقة يتم اختيارها بحيث تناسب عمره

ونوعه، وكذلك البطاقة البيضاء، ويطلب منه أن يستغرق خمس دقائق في كل قصة، ويحكي ما يحدث في الصورة الآن والأحداث التي أدت إليها. وما سوف تنتهي إليه، ويجب أيضاً أن تتضمن القصة تفاصيل عن الشخصيات التي في الصورة، وأنشطتهم وتفكيرهم ومشاعرهم. (علام، 2000، ص 633) وسنعرض مثال لأحد بطاقات اختبار تفهم الموضوع (TAT)، كالآتي:

البطاقة الثانية عشر (B): طفل صغير يجلس على عتبة منزل خشبي، تشير هذه الصورة إلى العزلة، وتستثير انفعالات حول الحرمان والوحدة وشعوراً بالعجز، كما أنها تثير ذكريات حول الماضي، وإهمال الوالدين وتخليهما عنه.



الشكل (17) البطاقة الثانية عشر (B) من اختبار تفهم الموضوع (TAT)،

2-1- ثبات وصدق اختبار تفهم الموضوع:

إن الخصائص السيكومترية لهذا الاختبار لا تختلف كثيراً عن اختبار رورشاخ وذلك نظراً لأنه لا توجد إجراءات مقننة لتطبيق الاختبار وتصحيحه وتفسير نتائجه. فإحدى طرق تقدير الثبات في هذا الاختبار تعتمد على اتفاق المحكمين في تحليل كل منهم لبروتوكول الاختبار ولكي تكون هناك درجة عالية من الاتفاق ينبغي أن يكون المحكمون على درجة عالية من الخبرة والتمرس فيما يتعلق بنظام تقييم الاستجابات للحصول على قيم لمعامل الثبات تقترب من 0.90 على الأقل لكي تكون هناك ثقة في درجة الاتفاق (Murstein, 1963). أما الثبات بطريقة إعادة تطبيق الاختبار فإن قيمه تتذبذب بحسب المدة الزمنية الفاصلة بين مرتي التطبيق. (علام، 2000، ص 635)

وقد وجد أن قيمة معامل الاستقرار تقل كثيراً بمرور مدة زمنية تبلغ عدة أشهر؛ لذلك يرى كثير من مستخدمي الاختبار أن معامل الاستقرار لا يناسب هذا الاختبار كما هو الحال في اختبار رورشاخ، وذلك لأننا نتوقع اختلافاً في موضوع القصة ومحتواها الذي يقدمه الفرد من وقت إلى آخر، وكذلك الحالة النفسية للفرد ربما تتغير بمرور الزمن، واختلاف صيغ القصص ينبغي أن يعكس هذا التغير دون أن يكون هذا دليلاً على عدم الثبات. غير أنه ينبغي التركيز على اتساق أقسام الاستجابات أو عدم اتساقها بدلاً من التركيز على المقارنة بين القصص في مرتي التطبيق. أما طريقة التجزئة النصفية في تقدير الثبات فإنها لا تناسب هذا الاختبار كما هو الحال في اختبار رورشاخ. فقد بينت الدراسات أن قيمة معامل الثبات باستخدام هذه الطريقة كانت منخفضة. (علام، 2000، ص 635)

ويُعد تقدير صدق هذا الاختبار من الأمور المعقدة أيضاً، فالسؤال الذي يتبادر إلى الذهن هو: هل القصص تعكس حقيقة شخصية الفرد المختبر أم أنها مجرد ردود فعل نمطية للموقف أي للصور؟ وحتى إذا افترضنا أن القصص تعكس شخصية الفرد فهل افتراض أن الفرد يتطابق مع البطل، وأن مشكلات البطل والضغوط المؤثرة عليه تعكس حقيقة جوانب حياة المختبر يعد افتراضاً صحيحاً؟ ومما يزيد من حدة المشكلة هو عدم الاتفاق على طريقة واحدة لتقييم الاستجابات ونظام التفسير. ونظراً لأن الاختبار يستخدم في الوصف العام للشخصية فإن مشكلة تحديد المحك عند تقدير الصدق المرتبط بمحك تُعد من المشكلات المعقدة. غير أن هناك اتفاقاً بين معظم الاختصاصيين في هذا المجال على أن هذا الاختبار يتميز بصدق محتواه في تقييم الشخصية. (علام، 2000، ص 636)

ومن الجدير بالذكر أن هناك اختبارات أخرى تعد تعديلات لاختبار تفهم الموضوع صممت بحيث تناسب صغار السن مثل اختبار القصة المصور لسيموندز Symonds, Picture Story Test، واختبار تفهم الموضوع للأطفال (CAT)، واختبار الصور لبلاكي Blacky Pictures، والاختبار السمعي لتفهم الموضوع Auditory Apperception Test (AAT)، واختبار أيوا لتفسير الصور Iowa Picture Interpretation Test (IPTT). كما أن هناك تعديلات أخرى للاختبار لقياس الاتجاهات نحو المدرسة، والسلطة، والعمل، وغير ذلك تستخدم أيضاً في مجالات الإرشاد المهني والمشروعات البحثية المختلفة.

وكذلك أعد ريتزل ومعاونوه (Ritz et al. 1980) اختباراً بديلاً لاختبار تفهم الموضوع، وذلك بانتقاء صور تسمح بتوازن القصص الإيجابية والسلبية، وكذلك بأنشطة متنوعة ومستويات مختلفة من الطاقة لبطل القصة، وذلك بهدف تطوير منهجية اختبار تفهم الموضوع بحيث يمكن تقييم الاستجابات تقيماً كمياً. (علام، 2000، ص 636)

3- مقارنة بين اختبار تفهم الموضوع واختبار رورشاخ:

يعد كل من اختبار تفهم الموضوع واختبار رورشاخ أداة إكلينيكية تتطلب أخصائياً متمرساً لاستخدامها ولتقييم استجابات المفحوصين. ويفيد اختبار رورشاخ في الكشف عن الاضطرابات في عمليات التفكير وبخاصة عند المرضى الذين يتعرضون لاضطرابات ذهانية، بينما يهتم اختبار تفهم الموضوع بالتوافق الاجتماعي. وعلى الرغم من اختلاف الاختبارين فيما يقيسه كل منهما، إلا أن لكل منهما ميزاته وعيوبه فاختبار تفهم الموضوع يهتم اهتماماً رئيساً بمحتوى القصص وخصائص البطل الذي يتطابق معه المفحوص، ويختلف تفسير الاستجابات باختلاف هذه الخصائص. أما محتوى الاستجابات فليس له نفس القدر من الأهمية في اختبار رورشاخ، فالجزء من بقعة الحبر الذي تشير إليه استجابات المفحوص لا يكون مهماً بقدر أهمية محددات المدرك (صيغة أو لون منطقة معينة من البقعة). (علام، 2000، ص 636)

ونظراً لأن هناك العديد من الأدلة التي أوضحت افتقار كل من الاختبارين إلى الخصائص السيكومترية مثل الثبات والصدق، وبخاصة اختبار رورشاخ، فإن كثيراً من علماء النفس الإكلينيكي يُفضّلون استخدام الاختبارين معاً في التشخيص والتوصل إلى فهم متكامل للشخصية. (علام، 2000، ص 637)

خامساً: صدق وثبات الاختبارات الإسقاطية:

ولا شك في أن للاختبارات الإسقاطية قيمتها الإكلينيكية التي لا يمكن تجاهلها وذلك على الرغم من بعض المآخذ المرتبطة بصدقها وثباتها من الناحية الإحصائية، ولعل غموض المثيرات فيها من أهم ميزات ذلك أنها تعمل على تحويل الإثارة من الخارج إلى الداخل فتصبح خبرات الفرد المكتبوتة مصدراً أساسياً لاستجاباته في ظل تحرر هذه الخبرات من سيطرة الأنا. (الغامدي، 1424هـ، ص 02)

إن المطلع على كتاب القياس العقلي السنوي Buross Mental Measures Yearbook يلاحظ وجهات النظر المتناقضة حول الخصائص السيكومترية لاختبار الرورشاخ ولعل هذا الأمر يرجع إلى أنه لا يوجد اتفاق بين المختصين على أسلوب التصحيح أو التفسير، والأمر الوحيد المتفق عليه بين المختصين هو مجرد عرض بقع الحبر على المفحوص وسؤاله ما الذي يراه في البقعة.

ولعل من طرق الثبات المستخدمة مع اختبار الرورشاخ هو مدى الاتفاق بين مجموعة من المحكمين في تقدير الدرجات، كما أنه من بين الطرق التي تستخدم لحساب الثبات طريقة الأشكال المتعادلة: حيث استخدمت مجموعة من البطاقات التي أعدها هاردر و شتاينر بافتراض أنها تمثل

الشكل المعادل للاختبار وكانت النتائج التي تم الوصول إليها مرضية إلى درجة ما. كما حسب الثبات أيضاً بطريقة التصنيف، وتوصلت دراسة واحدة فقط إلى معاملات ثبات مرتفعة نسبياً حيث تراوحت بين 0.60 و 0.95، أما طريقة الإعادة فقد قدمت نتائج متضاربة.

أما بالنسبة لصدق الاختبار فإن ذلك يعتمد على الإستراتيجية التي يتخذها الأخصائي النفسي في تفسير الاستجابات على الاختبار، وبشكل عام أظهرت الدراسات انخفاض مؤشرات الصدق بالنسبة لهذا الاختبار. (أناستازي، 2015، ص 631)

لقد استخدمت وسائل عديدة من اجل الحصول على الصدق والثبات منها.

1- قام "رابابورت" و"جيل" و"شاتر" بتحقيق صحة بعض الاختبارات التي تستخدم لتشخيص الأمراض العقلية عن طريق انتقاء مجموعات تجريبية تتكون من فئات إكلينيكية مختلفة من الذهانين كحالات الاكتئاب ثم مجموعة ضابطة للمقارنة تتكون من (50) رجلاً من رجال المرور باعتبارهم أسوأ بمقارنة استجابات المجموعات أمكن اختبار صدق العلامات المختلفة التي يستدل بها على المرض العقلي وهنا يعد الصدق صدقاً تلازمياً.

2- نسأل شخصاً يعرف المفحوص حق المعرفة فإذا كانت التأويلات التي حصلنا عليها من الاختبار تصدق فعلاً على هذا الشخص.

3- طريقة المضاهاة بإيجاد علاقة كمية بين نواحي نوعية في الشخصية كأن نضاهي بين تقريرين لباحثين مختلفين عن شخص واحد.

4- لقد تعرضت الاختبارات الإسقاطية إلى نقد كبير لها باعتبارها ذاتية وليست موضوعية وان ثباتها وصدقها غير موثوق به وان قدرتها على التمييز بين الحالات السوية وغير السوية ضعيفة للغاية ومشكوك فيها. (سوسن، 2005، ص 360)

سادساً: تقييم الأساليب الإسقاطية:

اتضح من عرضنا السابق تنوع الأساليب الإسقاطية واختلاف أدواتها وما تشتمل عليه من مثيرات، وما تتطلبه من استجابات وما تتوصل إليه من تفسيرات فالمثيرات التي تستدعي الاستجابات الإسقاطية يمكن أن تكون كلمات أو جملاً أو قصصاً ناقصة، أو بقع حبر، أو صوراً محددة أو غير محددة البنية، أو عرائس ولعب، أو رسوماً أو لوحات ملونة، أو غير ذلك. وقد صنفها ليندزي Lindzcy كما رأينا في خمسة أقسام عامة. ونظراً لأن الأساليب المختلفة تكشف عن أوجه متباينة إلى حد ما

للشخصية على الرغم من تداخل هذه الأوجه، فإنه يفضل استخدام أكثر من أسلوب في عملية التقييم. (علام، 2000، ص 637)

سابعاً: ميزات الأساليب الإسقاطية:

تتميز الأساليب الإسقاطية بأن مثيراتها وتعليماتها غير محددة البنية Unstructured أو غامضة مما يسمح بحرية الاستجابة وتنوعها، وتتباين الأساليب في درجة غموض مثيراتها. فبعضها مثل اختبار رورشاخ تكون مثيراته أكثر غموضاً من اختبار تفهم الموضوع أو إكمال الجمل مما يؤدي إلى تباين ملحوظ في مدى الاستجابات. ومعظم هذه الأساليب تكون غامضة الهدف أيضاً، مما يجعل من الصعب على الفرد المختبر تزييف استجاباته، وذلك لأن طرق تقييم الاستجابات أو تقدير الدرجات وتفسير الاستجابات يكون غير مألوف له وكذلك تسمح هذه الأساليب بحرية الأخصائي الإكلينيكي في سبر وتوضيح ما يقصده المختبر. كما أن معظمها لا يحتاج إلى مهارة في القراءة، ويمكن أن تُطبق على الأميين والأطفال الصغار، وتتميز أيضاً بعدم وجود استجابة صحيحة وأخرى خطأ أو استجابات محددة مسبقاً، حيث يتم تفسير الاستجابات كيفياً أو تصنيفياً بطرق مختلفة. فبعضها مثل اختبار رورشاخ يركز التفسير على الخصائص الشكلية، وفي غيره مثل اختبار تفهم الموضوع أو إكمال الجمل يكون التركيز على المحتوى وتستند معظم الأساليب على فرضية أن إضفاء الفرد بنية على المثيرات الغامضة أو ردود فعله لهذه المثيرات يعكس نزعاته وتصورات الشخصية وبخاصة نزعاته اللاشعورية وميكانيزماته غير اللفظية. (علام، 2000، ص 637)

ثامناً: عيوب الأساليب الإسقاطية:

من أهم عيوب هذه الأساليب افتقارها إلى التقنين Standardization فالاختبارات الإسقاطية ليست اختبارات بالمعنى السيكمومتري الدقيق، وإنما أدوات أو أساليب يستخدمها أخصائي مدرب تدريباً عالياً في تقييم الشخصية والكشف عن جوانبها. وعلى الرغم من أن هناك جهوداً لتقنين عرض المثيرات، إلا أنها تظل هناك فروق في المداخل التي يستخدمها الفاحصون المختلفون كما أن نتائج التقييم ووصف شخصية المختبر تعتمد على الأحكام المستبصرة للفاحص. (علام، 2000، ص 638)

ومن بين أوجه قصور الأساليب الإسقاطية افتقارها أيضاً إلى المعايير Norms، أو استنادها إلى معايير مقتضبة أو مستندة إلى جماعات مرجعية غير محددة تحديداً دقيقاً. وعدم توافر هذه المعايير يؤدي إلى تفسير الاستجابات في ضوء الخبرة الإكلينيكية، مما يسمح بتدخل كثير من الجوانب الذاتية

والتحيز في عملية التفسير. وقد أسهمت هذه العوامل مجتمعة في القيم المنخفضة لمعامل الثبات وعدم وجود أدلة كافية تتعلق بالجوانب المختلفة لصدق هذه الأساليب وجدوى استخدامها. وعلاوة على ذلك فإن جميع الأساليب الإسقاطية تتطلب التطبيق الفردي مما يجعل عملية التطبيق وتقييم الاستجابات وتفسيرها مكلفاً. كما يجب أن يقوم بتطبيق هذه الأساليب وبخاصة اختبار رورشاخ واختبار تفهم الموضوع، والدراما النفسية. والاجتماعية واللعب، واختبارات الرسم، أخصائيون تلقوا تدريباً خاصاً على استخدامها وتفسير نتائجها. وعلى الرغم مما يشوب هذه الأساليب من عيوب إلا أنها تزود الأخصائيين ببيانات تتعلق بديناميات الشخصية يصعب الحصول عليها باستخدام الاستبيانات أو أي أساليب أخرى غير أنه ينبغي عدم الاستناد إلى نتائجها في اتخاذ قرارات مهمة تتعلق بالأفراد في ضوء الطرق التفسيرية الحالية. كما لا ينبغي استخدامها في الإرشاد المهني أو انتقاء الأفراد، ولكن يمكن الاستفادة من هذه الأساليب في الأغراض البحثية من أجل التوصل إلى إجراءات كمية لتقييم الاستجابات لكي تصبح أكثر موضوعية، وكذلك لبحث صدق متغيراتها في علاقتها باختبارات أخرى تم التحقق من صدقها، والتحقق من صدقها التجريبي في ضوء مجموعات محكية مختلفة مثل مرضى البارانويا، والفصام. فإذا أجريت مثل هذه البحوث، فإن الأساليب الإسقاطية يمكن أن تصبح واحدة من أفضل الاختبارات السيكولوجية (Kline, 1976). (علام، 2000، ص 639)

المحاضرة السابعة:

ترجمة وتكييف الاختبارات النفسية والتربوية

تمهيد:

شهد مجال ترجمة وتكييف الاختبارات النفسية والتربوية تطوراً سريعاً على مدى الخمسة والعشرين السنة الماضية، والذي اتضح في العديد من الدراسات والكتب المنشورة ونماذج لأعمال الإستثنائية في عملية تكييف الاختبارات مثل أعمال كل من، (van de Vijver & Leung, 1997, 2000; Hambleton, Merenda, & Spielberger, 2005 ; Grégoire & Hambleton, 2009 ; Rios & Sireci, 2014) ويرجع سبب تزايد الاهتمام في هذا المجال هو تطور دراسات علم النفس بين الثقافات، ودراسات المقارنة الدولية واسعة النطاق حول النظم التعليمية، مثل دراسة الاتجاهات الدولية في الرياضيات والعلوم (TIMSS) وبرنامج التقييم الدولي للطلاب (PISA) وهي مجموعة من الدراسات تشرف عليها منظمة التعاون والتنمية الاقتصادية (OECD)، وكذلك برامج اختبارات الاعتماد الدولية المعتمد في مجال تكنولوجيا المعلومات كشركة (Microsoft)، واختبارات مستوى اللغة الذي يسمح بدخول الجامعة.

احرز التقدم التقني في المناهج الكمية والكيفية لقياس البنيات النفسية، والأساليب والطرق المستخدمة في تكييف الاختبارات والمقاييس لاستخدامها في السياق الثقافي، كإستخدام الإجراءات الإحصائية المعقدة مثل نظرية الاستجابة للمفردة الاختبارية (IRT)، والنمذجة بالمعادلة البنائية (SEM)، ونظرية إمكانية التعميم. (ITC, 2017, pp.01-68)

ورغم هذه التطورات في الإستخدامات، إلا أن لها عواقب كثيرة في ما يتعلق بتطوير وإدارة أدوات القياس، والتحقق من صدقها وثباتها، ومن بين النتائج المحتملة الحاجة إلى تكييف مفردات اختبار موجود من أجل زيادة فهمه من طرف المشاركين لغة الاختبار ليست لغتهم الأم.

وذلك بهدف الاستفادة من التطورات التقنية والخبرات المكتسبة من قبل الباحثين في هذا المجال، وشملت هذه التطورات التالي:

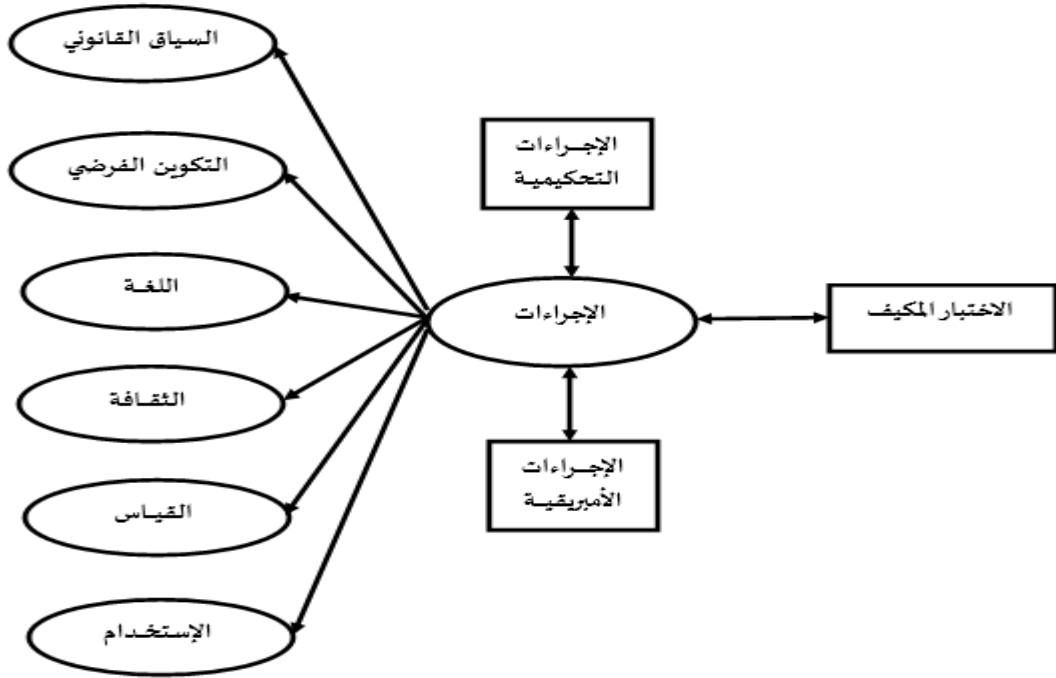
- تطور منهجية النمذجة بالعدالات البنائية (SEM) لتقييم تكافؤ مفردات الاختبار عبر مجموعات لغوية مختلفة.
- تطور طريقة تحليل الأداء التفاضلي للمفردة (DIF) واسعة الانتشار في المقياس المتعددة الاستجابات عبر مجموعات لغوية مختلفة. (ITC, 2017, pp.01-68)

• تطوير نماذج جديدة لتكييف الاختبارات بواسطة مشاريع التقييم الدولية (TIMSS, PISA) التي تشرف عليها منظمة التعاون والتنمية الاقتصادية (OECD).

وللتوضيح بشأن التمييز بين ترجمة الاختبار وتكييفه، فقد تكون الترجمة هي المصطلح الأكثر استخداماً، ولكن تكييف الاختبار هو مصطلح أشمل يشير إلى عملية نقل الاختبار من لغة وثقافة إلى ثقافة أخرى. من ناحية أخرى، فإن ترجمة الاختبار لها معنى أكثر تقييداً، ويقتصر على اختيار اجتياز الاختبار من لغة وثقافة أخرى مع الحرص على الحفاظ على المعنى اللغوي. إن عملية ترجمة الاختبار ليست سوى جزء من عملية التكييف، والتي تعتبر بمثابة الطريقة البسيطة لنقل الاختبار من لغة إلى أخرى دون النظر إلى التكافؤ الأكاديمي أو النفسي. (ITC, 2017, pp.01-68)

أولاً: الدليل الإرشادي لترجمة وتكييف الاختبارات:

يقدم الدليل الإرشادي لتقييم الاختبارات الذي أصدرته اللجنة الدولية للاختبارات (ITC) في طبعته الثانية (2017) مجموعة من التوصيات أو إرشادات أو معايير إجرائية تخص إجراءات تقييم وتكييف الاختبارات أو التطوير المتزامن للاختبارات النفسية والتعليمية بغية استخدامها مع عينات مختلفة. قدمت النسخة الثانية 18 توصية أو معيار إجرائي اندرجت تحت ستة أقسام رئيسية، وهي: الشروط الأولية (Pre-Condition) ثلاثة معايير، تطوير أو بناء الاختبار (Test Development) خمسة معايير، التأكد والتحقق من صدق التحليلات التجريبية (Confirmation [Empirical Analyses]) أربعة معايير، وإدارة الاختبار (Administration) معيارين، والتقييم والتفسير (Score Scales and Interpretation) معيارين، وتوثيق الاختبار (Documentation) معيارين كذلك. وكل توصية من هذه التوصيات (18) قدمت عنها شروحات مفصلة ومقترحات للممارسة في هذه النسخة من الدليل والتي سنتطرق إليها بعد أن نعرض شكل توضيحي لمكونات عملية تكييف الاختبار وفق دليل النسخة الثانية للهيئة الدولية لترجمة وتكييف الاختبار.



شكل رقم (18) يوضح إجراءات عملية تكييف الاختبارات

(Muniz and Hambleton, 2016, p.294).

القسم الأول: الشروط الأولية (التمهيدية):

1. الحصول إلزاماً على إذن من صاحب حقوق الملكية الفكرية الخاصة بالاختبار قبل البدء في تكييفه.

يتعلق هذا المعيار بحقوق الملكية الفكرية التي يتمتع بها الأفراد على إبداعاتهم أو اختراعاتهم أو منتجاتهم. ووفقاً للمنظمة العالمية للملكية الفكرية (www.wipo.int)، " يعد قانون المؤلف جزءاً من القطاع القانوني وعلى نطاق واسع يخص الملكية الفكرية، التي تهدف عموماً إلى حماية المصنفات وحماية حقوق الملكية الفكرية ومصالح المبتكرين والمبدعين من خلال منحهم حقوقاً حول أعمالهم". يتمتع مؤلفو الاختبارات النفسية والتربوية بحقوق ملكية فكرية خاصة محمية لابتكاراتهم مثل، منع بعض التشويهات عند نسخها أو تكييفها. يمكن ممارسة حقوق أخرى مثل عمل نسخ من قبل أشخاص آخرين على سبيل المثال، الناشر ممن حصلوا على ترخيص من المؤلف أو صاحب حقوق الطبع والنشر. بالنسبة للعديد من الاختبارات، كما هو الحال بالنسبة للمصنفات المكتوبة الأخرى، قد يتنازل المؤلف عن حقوق النشر للناشر أو الموزع. لا تتعلق حقوق الطبع والنشر في الغالب بالمحتوى الخاص بالمفردات، بل التنظيم الأصلي للاختبار كبناء الاختبار، شكله، ونظامه، أدوات التصحيح وما إلى ذلك. وبالتالي، فإن أي تقليد لإختبار أصلي أو الحفاظ على بنائه الأصلي ونظام التصنيف الخاص به

أثناء كتابة مفردات جديدة، يشكل إنتهاكاً لحقوق الملكية الفكرية للاختبار الأصلي. عندما يأذن المؤلف إجراء التكييف لمطور الاختبار، يجب عليه إحترام خصائصه الأصلية عند تعديل الاختبار (البنية، الأدوات، الشكل، التصحيح...)، إلا في حالة تحصله على موافقة من صاحب الاختبار تسمح له بتعديل تلك الخصائص.

2. التقرب من العينة المستهدفة لتقييم درجة التوافق بين التعريف ومحتوى البنية المقاسة بواسطة الاختبار الأصلي وأن يكون كل بند كافياً للاستخدام المقصود (أو الاستخدامات المقصودة) لنتائج الاختبار.

تتطلب هذه التوصية فهم الموضوع الذي يتم تقييمه بنفس الطريقة إلى جانب المجموعات المختلفة التي تشكل أساساً لمقارنات ثقافية ولغوية بين الثقافات. في هذه المرحلة من الإجراء، الاختبار أو أداة القياس ليس مكيف بعد. وعليه من المستحسن جمع أدلة تجريبية من خلال توثيق اختبارات مماثلة، وتقييم مدى اتفاق ملاءمة بنية المفردة للمجموعات اللغوية المستهدفة في الدراسة، أي التأكد من ثبات أو اللاتغاير البنية بالنسبة للإصدارات اللغوية المختلفة للاختبار. (ITC, 2017, pp.01-68)

لهذا يجب توظيف خبراء في ميدان البنية المقاسة والمجموعات الثقافية المستهدفة من أجل تقييم ملاءمة بنية الاختبار للمجموعات المختلفة لغوياً. ومن الأسئلة التي يجب أن يطرحها الخبراء هنا هو: هل للبناء معنى مختلف في كل ثقافة؟، والكثير من الخبراء يحكمون على أن البناء المقاس للاختبار لا معنى له أو أنه فقد معناه عند نقله إلى ثقافة أخرى، مثل بنية متغير نوعية الحياة، الاكتئاب، الذكاء... لهذا يجب استخدام طرق مختلفة لجمع مزيداً من الأدلة حول بنية الاختبار مثل مجموعات المناقشة والمقابلات والتحقيقات للحصول على معلومات أكثر حول درجة توافق البناء بين الثقافات.

3. الحد بشكل كبير من تأثير الاختلافات الثقافية واللغوية غير المرغوب فيها والغير الضرورية في استخدام قصدي للاختبار في العينات المستهدفة.

يجب تحديد الخصائص الثقافية واللغوية التي لا تتعلق بالمتغيرات التي من المفترض أن الاختبار وضع لقياسها. فقد تكون لها علاقة بشكل المفردات والمعدات المستخدمة كاستخدام الكمبيوتر أو الصور أو الصور التخطيطية، وما إلى ذلك، أو تكون لها علاقة بمدة تطبيق الاختبار وإدارته. والمنهج المستخدم لتحقيق ذلك هو تقييم "المسافة اللغوية الثقافية" بين اللغة المصدر واللغة المستهدفة للاختبار، وقد يشمل هذا المنهج اعتبارات تتعلق بالاختلافات لغوية، بناء الأسرة، والدين، والقيم، ونمط الحياة (van de Vijver & Leung, 1997).

تعتمد هذه التوصية بشكل أساسي على المناهج النوعية (الكيفية)، وتركز على المتخصصين الذين لهم علاقة بالبحث في الاختلافات الثقافية واللغوية. فهي تركز بشكل خاص على اختيار مترجمي

الاختبار الذين يجب أن تكون بين لغتهم الأم (الأصلية) واللغة المستهدفة، لأن مجرد معرفة اللغة المستهدفة ليس كافياً لتكون قادراً على تحديد مصادر تحيز البنية المحتمل. مثلاً، لوحظت مشاكل في شكل وطول الاختبار في دراسة مقارنة صينية أمريكية التي أجراها كل من Hambleton, Yu, Slater (1999) حول معرفة الرياضيات لدى طلاب الصف الرابع، بالإضافة إلى مجموعة من الخصائص الثقافية المرتبطة بهذا الاختبار. (ITC, 2017, pp.01-68)

القسم الثاني: تطوير أو بناء الاختبار:

1. التأكد من أن إجراءات الترجمة والتكييف والأخذ بعين الاعتبار الاختلافات اللغوية والنفسية والثقافية للعينات المستهدفة من خلال اختيار الخبراء ذوي الخبرة اللازمة.

تعتبر هذه التوصية إحدى التوصيات المهمة جداً، حيث أشارت العديد من الأدلة التجريبية التي تم جمعها من خلال الأبحاث التي أجرتها وكالات التقييم والمترجمين المؤهلين الذين تتجاوز قدرتهم مجرد معرفة اللغتين للمشاركين في تكييف الاختبار. لهذا أصبحت معرفة الثقافات والمعارف العامة لبنية الاختبارات جزءاً من معايير اختيار المترجمين، حيث لعبت هذه التوصية دوراً مهماً في تشجيع منظمات الترجمة وتكييف الاختبارات على استخدام مترجمين على الأقل وفقاً للنموذج الذي يتم اختياره مثل، نموذج الترجمة الثنائي الاتجاه أو العكسي. وهي عكس الممارسات القديمة المتمثلة في الاعتماد على مترجم واحد يتخذ جميع القرارات، بغض النظر عن مؤهلاته. فالمعرفة والخبرة في الثقافة المستهدفة بالنسبة للمترجم الذي يتقن لغته الأصلية ويعيش في المجتمع المستهدف بالتكييف، فهو لن ينتج ترجمة دقيقة فحسب، بل سينتج أيضاً ترجمة سهلة القراءة تتوافق والسياق أو البيئة المحلية. لذلك، فإن الخبير هو شخص أو فريق لديه معرفة مشتركة وكافية بما يلي:

- باللغات المعنية؛
- بالثقافات المعنية؛
- بمحتوى الاختبار المعني؛
- بالمبادئ العامة للاختبار. (ITC, 2017, pp.01-68)

والهدف من كل هذا هو إنتاج ترجمة أو تكييف ذا جودة عالية. ومن المستحسن كذلك في الممارسة العملية، دعوة فرق ذات مؤهلات مختلفة مثلاً، مترجمون ذوي خبرة أو دونها في المجال المعين للاختبارات، وخبير في الاختبارات، وغيرها، وذلك من أجل تحديد الجوانب المحتمل إهمالها في أي مرحلة من مراحل تكييف الاختبار. ومن الاعتبارات التي يجب أن يراعيها كذلك المترجم مثلاً، في ما يتعلق بمفردات الاختبار من متعدد، والتي لا يجب أن تكون الإجابة الصحيحة أطول أو أقصر من بقية

الاختيارات. كما ينبغي أن لا تشير الدلائل النحوية إلى الإجابة الصحيحة، بالنسبة للمفردات التي الإجابة عليها بصح/خطأ، لا يجب أن تكون الجملة الصحيحة أطول بشكل أو بآخر من الجملة الخاطئة. في واقع الممارسة العملية، من الصعب إيجاد مترجمين لديهم معرفة بمبادئ بناء وتطوير الاختبارات، لذلك من الضروري تدريب المترجمين على المبادئ الأساسية وكتابة وشكل المفردات التي سيتعاملون بها.

2. استخدم تصميمات وإجراءات ترجمة مناسبة لزيادة ملائمة تكييف الاختبار مع المجموعات المستهدفة.

تفرض هذه التوصية أن القرارات التي يتخذها المترجمون تزيد إلى حد أقصى الملاءمة والنسخة التي تم تكييفها للعينات المستهدفة. هذا يعني أن الكلمات المستخدمة يجب أن تكون طبيعية ومقبولة، والأخذ بالاعتبار معادلة التكافؤ اللغوي الوظيفي بدلاً من التكافؤ الحرفي. ومن النماذج الأكثر شعبية لتحقيق هذه الأهداف هي الترجمات أحادية الاتجاه والترجمات ثنائية الاتجاه أو العكسية والتي قدم فيها كل من (Brislin (1986), Hambleton & Patsula (1999) تحليلاً شاملاً للنموذجين، بما في ذلك تعريفهما ونقاط قوتها وضعفها.

فالاختلافات في البنية اللغوية يمكن أن تسبب مشاكل في ترجمة الاختبارات، فعلى سبيل المثال، في مقياس مشهور تم تطويره باللغة الإنجليزية من قبل كل مفرداته عبارة عن جمل يجب إكمالها: "أحب..."; "أتأسف..."; "لا أستطيع..."; وغيرها، فقد تبين أن هذا الشكل للجملة غير مناسب للغة التركيبية والتي يجب أن يسبق موضوع الجملة الفعل والفاعل. فاستخدام الجمل لإكمالها، كما هو الحال في النسخة الإنجليزية، من شأنه أن يغير تماماً الإجابة على المقياس. (ITC, 2017, pp.01-68)

3. تقديم أدلة بأن تعليمات الاختبار ومحتوى البنود لها نفس المعنى لجميع المجموعات المستهدفة. يمكن جمع الأدلة المطلوبة لهذه التوصية من خلال مجموعة متنوعة من الاستراتيجيات، وهي:

● الاعتماد على خبراء من المجتمع المستهدف؛

● استخدام عينات من المستجيبين متحكمين في اللغتين؛

● استخدام تحقيقات محلية لتقييم الاختبار؛

● تطبيق الاختبار غير المقنن في مواقف وطرق مختلفة من أجل زيادة ثباته وصدقه.

كما يفضل إجراء دراسة أولية أو تجريبية للنسخة المعدلة للاختبار. في هذه الدراسة لا تتضمن تطبيق الاختبار وتحليل بياناته فقط، ولكن الأهم هو إجراء مقابلات مع كل من مطبقي الاختبار والمشاركين في الدراسة بغرض جمع أحكامهم واقتراحاتهم بشأن معادلة نسخي الاختبار سواء فيما يخص التعليمات أو مفردات الاختبار. وهناك طرق أخرى ممكنة مثل دعوة مختصين من لغات

مختلفة أو مختصين في اللغتين لتقييم محتوى الاختبار، فعلى سبيل المثال يمكن أن يطلب من المختصين في اللغتين تقييم التشابه والصعوبة التي يتسبب فيها شكل المفردة ومحتوى كل نسخة من الاختبار، كما أن المقابلة المعرفية هي من الطرق الواعدة في الموضوع (Levin et coll, 2009).

4. تقديم دليل على أن شكل المفردات، ومقاييس التقدير (ليكرت)، وفئات التصحيح، وموثائق الاختبار وطرق تطبيقه وغيرها من الإجراءات المناسبة لجميع فئات المجتمع المستهدف.

يمكن أن يكون شكل المفردات مثل مقاييس التقدير (ليكرت) خماسية التدرج أو شكل المفردات الجديدة مثل طريقة "السحب والإفلات" أو "الإجابة على كل ما هو صحيح" أو حتى "الإجابة على خيار إجابة واحدة فقط" مربكة للمستجيبين الذين لم يروا من قبل المفردات بهذا الشكل. كذلك تصميم المفردات أو استخدام الرسومات أو صيغ المفردات المحوسبة التي تتطور بسرعة يمكن أن تكون مربكة للمرشحين. لهذا يجب أن تكون أشكال المفردات الجديدة مألوفة لدى المجيبين، لأنها تستطيع أن تكون مصدرا للتحيز يشوه نتائج الاختبار. ومن بين المقترحات العملية جمع المزيد من الأدلة التي تستند للبيانات الكيفية والكمية، ومرعاة خصائص الاختبار المكيف التي يجب التحقق منها:

- التحقق من أن التمارين أو بنود التدريب كافية لتحقيق المستجيبين للمستوى المطلوب، ما يسمح لهم بتقديم إجابات صادقة أو إجابات تعكس مستوى إتقانهم لمحتوى الاختبار.
- التحقق من مدى وضوح موثائق الاختبار للمستجيبين كوضع الرسوم التوضيحية و كيفية تدوين العلامات في ورقة الإجابة. (ITC, 2017, pp.01-68)

5. جمع بيانات تجريبية (استطلاعية) عن الاختبار المكيف لإجراء تحليل المفردات وتقييم الثبات ودراسات الصدق على نطاق ضيق بحيث تمكن من إجراء أي مراجعات ضرورية للاختبار المكيف. قبل البدء في أي دراسات واسعة النطاق للتحقق من ثبات وصدق درجات الاختبار ودراسات المعايير التي قد تستغرق وقتاً طويلاً وتكون مكلفة، لجمع أدلة مؤكدة للجودة السيكمومترية للاختبار المكيف. هناك العديد من تحليلات الخصائص السيكمومترية التي يمكن إجراؤها لتقديم أدلة أولية على ثبات وصدق درجاته. فعلى سبيل المثال، في مرحلة تطوير الاختبار، يمكن لتحليل المفردات باستخدام حجم عينة أقل من (100 فرد) مثلاً أن يوفر بيانات نحن في أمس الحاجة إليها بشأن أداء مفردات اختبار معين. وفيها يمكن تعديل المفردات السهلة جداً أو الصعبة بالمقارنة مع المفردات الأخرى أو المفردات التي تظهر تمييز منخفض أو سلبي. في حالة مفردات الخيارات المتعددة، سيكون من المناسب تقييم فعالية مشتتات المفردات. ويمكن كذلك رصد المشاكل والتنقيحات اللازمة. أيضاً، مع نفس البيانات التي تم تجميعها لتحليل مفردات الاختبار، يوفر معامل ألفا أو معامل أوميغا (McDonald,)

(1999) لمطور الاختبار معلومات قيمة للترجمة والتكيف (الطبعة الثانية) والتي يمكن استخدامها لدعم القرارات حول الطول المناسب لنسخ اللغة الأصلية والمستهدفة للاختبار. (ITC, 2017, pp.01-68) في بعض الحالات، تطرح مجموعة من التساؤلات حول جوانب معينة من التكيف، مثل: هل يمكن فهم تعليمات الاختبار بشكل صحيح؟ هل يجب أن تكون التعليمات مختلفة لتوجيه المتقدمين للاختبار بشكل فعال في اللغة الأصلية والثقافة الجديدة؟ هل سيأثر الاختبار المطبق بالكمبيوتر سلباً على المستجيبين مثل (الوضع الاجتماعي والاقتصادي المنخفض) في المجتمع المعني بالاختبار المكيف؟ هل أن طول الاختبار مناسب مقارنة بالوقت المتاح؟ ويمكن الإجابة على جميع هذه الأسئلة وغيرها من خلال نتائج صدق الدراسة الاستطلاعية. و الهدف هو تجميع ما يكفي من البيانات التي يمكن اتخاذ قرار بشأنها والتي من خلالها يقرر ما إذا كان ينبغي المضي قدماً في الاختبار المكيف أم لا. وإذا كان القرار للمضي قدماً، يمكن إجراء دراسات أخرى أكثر طموحاً يمكن التخطيط لها وتنفيذها مثل دراسات الأداء التفاضلي للمفردات (DIF)، ودراسات تقييم بنية وعوامل الاختبار. مقترحات للممارسة: (ITC, 2017, pp.01-68)

- إجراء دراسة التحليل الكلاسيكي للمفردات للحصول على معلومات حول متوسطات المفردات ومؤشرات التمييز، وفي حالة مفردات الخيار من متعدد يجب التحقق من فعالية المشتتات.
- إجراء تحليلات الثبات، مثل معامل كيوذر-ريتشارسون KR-20 مع المفردات ثنائية التدرج، أو معامل ألفا كرونباخ أو معامل أوميغا مع مفردات الاختيار من متعدد.
- إجراء دراسة استطلاعية أو اثنتين حسب الضرورة للحصول على رؤية واضحة حول صدق الاختبار المكيف. على سبيل المثال، افترض أن الاختبار المكيف يتم تطبيقه عبر جهاز كمبيوتر. في هذه الحالة من المستحسن إجراء دراسة لتقييم طريقة تطبيق الاختبار (أي طريقة الورق والقلم مقابل طريقة الكمبيوتر). لنفترض أن التعليمات كذلك تتطلب من المجيبين الإجابة على جميع الأسئلة. قد يكون من الضروري القيام ببعض البحوث لتحديد أفضل التعليمات لتحقيق هذا الهدف المحدد وتفادي وقوع المستجيبين في مشكلة التخمين. (ITC, 2017, pp.01-68)

القسم الثالث: التأكد والتحقق من صدق التحليلات التجريبية:

1. اختر عينة خصائصها مناسبة للاستخدام المستهدف للاختبار وحجمها كافي للتحليلات التجريبية. يشير تصميم جمع البيانات إلى الطريقة التي يتم بها جمع البيانات لوضع معايير إذا لزم الأمر والتكافؤ بين نسخ اللغة للاختبار، وإجراء دراسات الصدق والثبات والأداء التفاضلي للمفردات (DIF)، والشرط الأول فيما يتعلق بجمع البيانات هو أن تكون العينات كبيرة بما يكفي لجمع المعلومات الإحصائية الثابتة. ورغم أن هذا الشرط ينطبق على أي نوع من البحوث، فإنه وثيق الصلة بوجه

خاص في سياق دراسة التحقق من صدق تكييف الاختبار لأن التقنيات الإحصائية اللازمة للاختبار وتكافؤ المفردات مثل، التحليل العاملي التوكيدي (CFA)، ونماذج نظرية الاستجابة للمفردة (IRT) لتحديد تحيز محتمل لمفردات الاختبار، والتي تتطلب عند تطبيقها عينات كبيرة نوعاً ما لتكون أكثر جدوى بما يكفي لتقدير معالم النموذج بشكل ثابت، ويعتمد حجم العينة الموصى به على تعقيد النموذج المختار و جودة البيانات. تتضمن مقارنة بنية الاختبار عادة قيم التباينات المشتركة، ويظهر هذا في دالة توزيع الدرجات. فباستخدام عينات متطابقة، يتم مطابقة أي دور قد يلعبه توزيع الدرجات في نتائج العينتين الأصلية والمستهدفة، وبالتالي يمكن استبعاد دور توزيع الدرجات كتفسير لأي اختلاف في النتائج. (ITC, 2017, pp.01-68)

وربما يوضح مثال آخر تفسير مشكلة توزيع الدرجات المختلفة في مجموعة اللغة الأصلية واللغة المستهدفة. لنفترض أن ثبات درجة الاختبار 0.80 في مجموعة اللغة الأصلية، ولكن 0.60 في مجموعة اللغة المستهدفة. قد يبدو الفرق مثيراً للقلق ويثير تساؤلات حول مدى ملاءمة نسخة اللغة المستهدفة للاختبار. ومع ذلك، غالباً ما يتم تجاهل أن الثبات لأنه خاصية مشتركة بين الاختبار والمجتمع (McDonald, 1999) لأنه يعتمد على كل من تباين الدرجات الحقيقية (خصائص المجتمع) وتباين الخطأ (خصائص الاختبار). لذلك، يمكن أن يؤدي نفس تباين الخطأ إلى ثبات أعلى بسبب تباين الدرجات الحقيقية الكبير في مجموعة اللغة الأصلية. ويرى McDonald (1999) أن الخطأ المعياري للقياس والذي هو الجذر التربيعي لتباين الخطأ هو في الواقع أكثر ملاءمة كميّاً للمقارنة بين العينات، وليس الثبات. وثمة بديل آخر يستخدم معاملات الثبات وهو أخذ عينة متطابقة من المستجيبين من مجموعة اللغة الأصلية وإعادة حساب ثبات درجات الاختبار. (طباع و بلقيدوم، 2018، ص 60-74)

تسمح الطرق الحديثة لاختبار تباين القياس باستخدام تحليل العاملي التوكيدي المتعدد المجموعات (CFA) بتقييم العينات ذات التوزيعات المختلفة للسمات الكامنة. وفي مثل هذه النماذج، يفترض أن معالم القياس مثل تشعبات العوامل والقواطع متساوية عبر المجموعات، يسمح للمتوسطات والتباينات والتغايرات في السمات الكامنة بالتباين عبر المجموعات. وهذا يسمح باستخدام عينات كاملة، والسيناريو الأكثر واقعية هو ظهور توزيعات مختلفة من السمة المقاسة عبر المجتمع. اقتراحات للممارسة: في جميع الأبحاث تقريباً، هناك اقتراحان يتم تقديمهما عند وصف العينة:

- استخدام عينة كبيرة نوعاً ما ولضمان عدم تحيز محتمل لمفردات الاختبار يجب أن لا يقل عن حجم العينة عن 200 فرد لكل نسخة من الاختبار، لإجراء تحليلات نظرية الاستجابة للمفردة (IRT) والتحقق من مدى ملاءمة النماذج والتي تتطلب عينة من 500 فرد على الأقل، في حين أن الدراسات للتحقق من البنية العاملية للاختبار تتطلب أحجام عينة كبيرة إلى حد ما ربما 300 أو أكثر من

المستجيبين. ومن الواضح أنه يمكن إجراء تحليلات على عينات صغيرة أيضاً، ولكن باستخدام طريقة توليد عينات كبيرة كلما أمكن ذلك.

• اختيار عينات ممثلة للمجتمع كلما أمكن ذلك. لكي يمكننا تعميم النتائج المستخلصة من هذه العينات. لإزالة الاختلافات في النتائج بسبب عوامل منهجية مثل التباين في توزيعات الدرجات يجب سحب عينة من مجموعة اللغة الأصلية لمطابقة مجموعة اللغة المستهدفة، وقد تكون المقارنات بين أخطاء القياس المعياري أفضل من المقارنات على أساس معاملات الثبات.

2. تقديم الأدلة الإحصائية التي تتعلق بتكافؤ بنية الاختبار، وطرق التكافؤ، وتكافؤ المفردة لمختلف مجموعات المجتمع المستهدف.

إنشاء تكافؤ لبنية الاختبار في نسخته اللغة الأصلية و المستهدفة، ولكن ليس هو التحليل التجريبي المهم الوحيد الذي يجب إجرائه. كما تناولت المبادئ التوجيهية بصورة مختصرة الطريقة المتبعة في مجال تكافؤ البنية وطرق التكافؤ.

يحتاج الباحثون إلى معالجة التكافؤ على مستوى المفردة أيضاً. تتم دراسة تكافؤ المفردة بما يسمى "تحليل الأداء التفاضلي للمفردة (DIF)". ويظهر الأداء التفاضلي للمفردة عادة إذا كان اثنان من المتقدمين للاختبار من مجموعتين مختلفتين لغوياً وثقافياً، ولهما نفس مستوى السمة أو القدرة المقاسة ولكن يكون لديهما احتمال الاستجابة الصحيحة على المفردة مختلف. يمكن أن يحدث اختلاف كلي في الأداء على الاختبار عبر المجموعات، ولكن هذا لا يمثل مشكلة بمفردها. في حين أنه عندما تتم مطابقة أفراد المجتمع في البنية المقاسة بالاختبار عادة ما تكون الدرجات الكلية للاختبار، أو مجموع درجات الاختبار مطروح منها درجة المفردة التي تتم دراستها، وعندما نجد فروقات في الأداء على المفردة عبر المجموعات، بالتالي يمكن الحكم على أن المفردة تبدئ أداءً تفاضلياً (DIF).

يتم تنفيذ هذا النوع من التحليل لكل مفردة في الاختبار. وبعدها يتم إجراء محاولة لفهم أسباب وجود الأداء التفاضلي للمفردات (DIF)، ووفقاً لهذه المراجعة التحكيمية، قد يتم تحديد بعض المفردات السيئة، والتي يجب تعديلها أو حذفها تماماً من الاختبار. (ITC, 2017, pp.01-68)

وثمة هناك مصدران محتملان مهمان لتقييمهما وهما مشاكل الترجمة والاختلافات الثقافية. وبشكل أكثر تحديداً، قد يكون سبب وجود الأداء التفاضلي للمفردة (DIF) هو:

- (1) عدم تكافؤ الترجمة التي تتم من اللغة المصدر إلى اللغة المستهدفة للاختبار مثل الألفة مع المفردات المستخدمة، والتغيير في صعوبة المفردات، والتغيير في تكافؤ المعنى، وما إلى ذلك.
- (2) الاختلافات في السياقات الثقافية.

أثناء الترجمة، هناك إمكانية استخدام مفردات أقل شيوعاً في اللغة المستهدفة. يمكن أن تكون المعاني هي نفسها في النسخ المترجمة، ولكن حتى في الثقافة الواحدة يمكن أن تكون الكلمة أكثر شيوعاً مقارنة بكلمة أخرى. من الممكن أيضاً يتغير مستوى صعوبة المفردة نتيجة للترجمة بسبب طول الجملة وتعقيدها واستخدام المفردات السهلة أو الصعبة أيضاً. قد يتغير المعنى أيضاً في اللغة المستهدفة مع حذف بعض أجزاء من الجمل، والترجمات غير الدقيقة، ووجود أكثر من معنى واحد في المفردات المستخدمة في اللغة المستهدفة، والانطباعات غير المتكافئة لمعنى بعض الكلمات عبر الثقافات، وما إلى ذلك. (ITC, 2017, pp.01-68)

هناك ما لا يقل عن أربع مجموعات من طرق التحقق من الأداء التفاضلي للمفردة (DIF) عبر المجموعات اللغوية والثقافية. وهذه الإجراءات، هي:

(أ) الطرق القائمة على نظرية الاستجابة للمفردة (IRT)؛

(ب) طريقة مانتل - هاينزيل (MH) Mantel-Haenszel؛

(ج) طريقة الانحدار اللوجستي (LR)؛

(د) طريقة تحليل العوامل المقيدة (RFA) (Oort & Berberoglu, 1992).

بالنسبة للطريقة التي تستند إلى نظرية الاستجابة للمفردة (IRT)، يتم مطابقة المشاركين في الاختبار في نسخته استناداً إلى درجة السمة الكامنة. وفي طريقي طريقة مانتل - هاينزيل (MH) وطريقة الانحدار اللوجستي (LR)، تستخدم الدرجة الملاحظة للاختبار كمييار للمطابقة قبل مقارنة أداء المستجيبين في المجموعتين على المفردات. ومن معايير المطابقة الأكثر شعبية إضافة إلى هذه الطرق، يمكن كذلك استخدام درجات تقديرية أخرى، مثل التحليل العاملي والذي يستخدم الدرجات في التنقية عن طريق حذف المفردات السيئة. بشرط أن يكون محك المطابقة صادقاً وثابتاً بما يكفي لتقييم الأداء التفاضلي (DIF) بشكل صحيح. أم طريقة تحليل العوامل المقيدة (RFA)، وهي طريقة قوية يتم فيها التحكم في معادلات الخطأ المضخمة من الدرجة الأولى، وهي طريقة جيدة لاختبار الأداء التفاضلي غير المنتظم، والذي يتم فيها الفتح تحديد انحدار كل مفردة على متغير التقسيم (الفئوي) إلى جانب السمة الكامنة (Kolbe, 2019, pp.138-151).

يتم تعيين تشبع المفردة الحر و تقييم ملاءمتها للنموذج بناءً إلى النموذج الصفري حيث يتم تشبع أي مفردة على متغير التقسيم أي لا يوجد أداء تفاضلي (DIF) في النموذج. وإذا كان النموذج دال إحصائياً فإن هذا دلالة على أن المفردة تظهر أداءً تفاضلي (DIF).

عندما يكون الاختبار معقداً في مستوى الأبعاد، فإن العثور على معيار المطابقة المناسب هو مشكلة (Clauser, Nungester, Mazor, & Ripley, 1996). استخدام معايير مطابقة متعددة المتغيرات،

مثل درجات العامل التي تم الحصول عليها نتيجة للتحليل العاملي، والتي قد تغير في تفسيرات الأداء التفاضلي (DIF) على مستوى المفردة كذلك. وبناء على ذلك، يشير المبدأ التوجيهي إلى أنه إذا كان الاختبار متعدد الأبعاد، فقد يستخدم الباحثون معايير مختلفة لتقييم الأداء التفاضلي للمفردة (DIF)، وتقييم المفردات التي يتم الإشارة إليها بأنها تظهر أداءً تفاضلياً فيما بمختلف معايير المطابقة. يمكن أن تقلل المطابقة المتعددة المتغيرات من عدد المفردات التي تظهر أداءً تفاضلياً عبر المجموعات اللغوية والثقافية. (ITC, 2017, pp.01-68)

وقد تتطلب هذه المنهجيات أحجام عينات مختلفة. أما طرق MH، LR، و RFA هي نماذج يمكن أن تعمل بشكل ثابت وصادق في عينات صغيرة نسبياً مقارنة بالتقنيات المستندة إلى نماذج نظرية الاستجابة للمفردة (IRT)، والتي تتطلب عينات كبيرة لتقديرات معلمات صادقة. وثمة اعتبار آخر هو نمط بيانات الاستجابة على المفردة. يمكن تطبيق طرق MH و LR و RFA على البيانات الثنائية الاستجابة. وهناك حاجة إلى طريقة أخرى، مثل طريقة مانتل-هانزيل (MH) المعمم، والتي يمكن استخدامها مع بيانات الاستجابة المتعددة البدائل.

يتطلب هذا المبدأ من الباحثين تحديد المصادر المحتملة لتحيز الطريقة في الاختبار المكيف. وتشمل مصادر التحيز الطرق التالية:

- 1- المستويات المختلفة لدافعية المشاركين في الاختبار؛
 - 2- الخبرات المتفاوتة من جانب المجيبين على الاختبارات النفسية؛
 - 3- اختلاف في سرعة الاختبار بين مجموعة وأخرى؛
 - 4- الألفة المختلفة لشكل الاستجابة بالنسبة لمجموعتي الاختبار؛ (ITC, 2017, pp.01-68)
 - 5- عدم تجانس أسلوب الاستجابة، الخ. فالتحيز في الاستجابات، يشكل مصدر قلق كبير في تفسير نتائج برنامج التقييم الدولي للطلاب (PISA) مثلاً، والتي أصبحت محل اهتمام بحثي كبير.
- وأخيراً، والأهم من كل هذا، فإن هذه التوصية تحث الباحثين على معالجة البنية المتكافئة. وهناك أربعة طرق إحصائية على الأقل لتقييم تكافؤ البنية عبر نسخة اللغة الأصلية واللغة المستهدفة للاختبار، والطرق، هي: التحليل العاملي الاستكشافي (EFA)، التحليل العاملي التوكيدي (CFA)، والتدرج متعدد الأبعاد (MDS)، ومقارنة الشبكات المنطقية (Sireci, Patsula, & Hambleton, 2005). كما أن هناك العديد من الدراسات التي تم فيها استخدام التحليل العاملي التوكيدي (CFA) لتقييم ما إذا كان البنية العاملية للنسخة الأصلية للاختبار متسقة عبر إصداراته المكيفة (Byrne & van de Vijver, 2014) فهذه الطريقة جذابة لتقييم التكافؤ البنيوي عبر الاختبارات المكيفة لأنه يمكن التعامل مع مجموعات متعددة في وقت واحد، والاختبارات الإحصائية لتقدير مدى ملاءمة

النموذج متاحة، والتي يتم فيها كذلك توفير مؤشرات وصفية لملاءمة النموذج (Sireci et al, 2005). فالقدرة على التعامل مع مجموعات متعددة أصبح شيء مهم خاصة مع شيوع عملية تكييف الاختبارات لعدة لغات فمثلاً، وفي الاختبارات الدولية المقارنة كـ TIMSS و OECD/PISA، تم تكييف هذه الاختبارات إلى أكثر من 30 لغة. وبما أن الشرط الصارم المتمثل في عدم وجود تشعب تقاطعي أثناء استخدام التحليل العاملي التوكيدي، إلا أنه غالباً ما لا تكون هناك ملاءمة جيدة مع البيانات المتعلقة بأدوات القياس المعقدة والمتعددة الأبعاد، فإن الحل الذي أصبح له شعبية هو نمذجة المعادلة البنائية الاستكشافية (ESEM)، خاصة مع بيانات الشخصية أو المتغيرات الأكثر تعقيداً وتداخلاً (Asparouhov & Muthen, 2009). (ITC, 2017, pp.01-68)

فالتدرج المتعدد الأبعاد المرجح (WMDS) هو طريقة جذابة أخرى لتقييم تكافؤ البنية عبر نسخ مختلفة من لغات التقييم. وعلى غرار التحليل العاملي الاستكشافي (EFA)، لا يتطلب تحليل المتعدد الأبعاد المرجح تحديداً مسبقاً لبنية الاختبار، وهو، مثل التحليل العاملي التوكيدي (CFA)، يسمح بتحليل مجموعات متعددة (Sireci et al, 2003, pp.148-166).

واقترح (Van de Vijver and Tanzer 1997) أن يدرس الباحثون عبر الثقافات الثبات كل نسخة ثقافية للاختبار محل الاهتمام وأن يبحثوا عن أدلة الصدق التمييزي والتقاربي في كل مجموعة ثقافية. قد تكون هذه الدراسات في كثير من الأحيان أكثر عملية من دراسات بنية الاختبار التي تتطلب أحجام عينة كبيرة جداً.

تتطلب اختبارات مثل تلك المستخدمة في الدراسات الدولية للتقييم المقارن PISA و TIMSS أدلة على التداخل الكبير في المحتوى لأن النتائج تستخدم لمقارنة إنجاز الطلاب في العديد من البلدان. إن استخدام قائمة بيك للاكتئاب المترجمة من اللغة الإنجليزية إلى اللغة الصينية بالنسبة للباحثين الذين يدرسون الاكتئاب أو للمستشارين لتقييم الاكتئاب لن يتطلب تداخلاً كبيراً في المحتوى. ومع ذلك تبقى الحاجة قائمة إلى أدلة الصدق لدعم قائمة الاكتئاب في البيئة الصينية.

ويمكن أيضاً معالجة هذا الموضوع بطرق إحصائية أخرى بعد تكييف الاختبار. فعلى سبيل المثال، إذا كان يعتقد أن المجموعات الثقافية تختلف في متغيرات مهمة لا علاقة لها ببنية المفهوم المقاس، يمكن استخدام تصاميم شاملة وتحليلات إحصائية لضبط على هذه المتغيرات "الدخيلة". ويمكن استخدام تحليل التباين المشترك (التغاير)، وتصاميم الكتل العشوائية، وغيرها من التقنيات الإحصائية كتحليل الانحدار، والارتباط الجزئي، وما إلى ذلك، للسيطرة على تأثير التباين بين المجموعات. (ITC, 2017, pp.01-68)

اقتراحات للممارسة:

تعتبر هذه التوصية مهمة للغاية، لأن هناك العديد من التحليلات التي يمكن إجراؤها. بالنسبة لتحليلات التكافؤ، نقدم الاقتراحات التالية للممارسة:

- إذا كانت أحجام العينات كافية، علينا بالقيام بإجراء دراسة مقارنة لتكافؤ البنية بين النسخة الأصلية والمستهدفة للاختبار فهناك الكثير من الحزم والبرامج الإحصائية لتسهيل هذه التحليلات.
- إجراء التحليل العاملي الاستكشافي والذي يقوم على تدوير البنية المستهدفة والذي يسمى "الدوران المستهدف" أو التحليل العاملي التوكيدي، أو تحليل التدرج المتعدد الأبعاد المرجح، لتحديد مدى الاتفاق على بنية الاختبار محل الاهتمام عبر اللغة أو المجموعات الثقافية. إن شرط حجم عينات كبيرة (10 أفراد لكل متغير) يجعل من الصعب إجراء هذه الدراسات في العديد من الدراسات المشتركة بين الثقافات. (ITC, 2017, pp.01-68)

• البحث عن أدلة على الصدق التمييزي والتقاربي من الضروري إيجاد أدلة عن الارتباط بين مجموعة من البنى والتحقق من استقرار هذه الارتباطات عبر اللغات أو المجموعات الثقافية.

3. تقديم أدلة تدعم المعايير وصدق وثبات النسخة المكيفة من الاختبار في المجموعات المستهدفة. لا تنطبق المعايير وأدلة الصدق والثبات الخاصة باختبار ما في نسخته الأصلية تلقائياً على التعديلات المحتملة الأخرى للاختبار على ثقافات ولغات مختلفة. ولذلك، يجب أيضاً تقديم أدلة على صدق وثبات أي نسخ جديدة يتم تطويرها. وينبغي إدراج جميع أنواع الأدلة التجريبية التي تدعم الاستدلالات المستخلصة من الاختبار في دليل الاختبار. وينبغي إيلاء اهتمام خاص للمصادر الخمسة لأدلة الصدق استناداً إلى: محتوى الاختبار، وعمليات الاستجابة، والبنية الداخلية للاختبار، والعلاقات مع المتغيرات الأخرى، ونتائج الاختبار (AERA, APA, NCME, 2014). ومن بين أهم التقنيات التي يمكن استخدامها للحصول على بيانات حول أدلة الصدق وتحليلها استناداً إلى البنية الداخلية للاختبار. التحليل العاملي الاستكشافي والتوكيدي، والنمذجة بالمعادلات البنائية، والتحليلات المتعددة السمات- المتعددة الطرائق.

اقتراحات للممارسة: والاقتراحات هي نفسها المطلوبة لأي اختبار نريد استخدامه:

- اقترح تطوير المعايير التي وضعت للنسخة الأصلية من الاختبار مع النسخة المكيفة، ينبغي تقديم أدلة على أن هذا الاستخدام مناسب وعادل من الناحية الإحصائية. وإذا لم يكن بالإمكان تقديم أي دليل على استخدام معايير النسخة الأصلية، فينبغي وضع معايير محددة للصيغة المكيفة وفقاً لعملية التقنين وتطوير المعايير. (طباع و بلقيدوم، 2018، ص 60-74)

• تجميع أدلة كافية حول الثبات لتبرير استخدام إصدار اللغة المستهدفة من الاختبار. وقد تتضمن الأدلة عادة تقديراً للاتساق الداخلي مثل معامل كيوذر-ريتشارسون KR-20، أو معاملات ألفا كرونباخ أو أوميغا.

• تجميع أدلة حول الصدق بقدر كافي لتحديد ما إذا كان ينبغي استخدام نسخة اللغة المستهدفة من الاختبار. ويتوقف نوع الأدلة التي يتم تجميعها حول الاستخدام المقصود لدرجات الاختبار مثل، صدق المحتوى لاختبارات الإنجاز الدراسي أو التحصيل، والصدق التنبؤي لاختبارات الاستعدادات وما إلى ذلك. (طباع وبلقيدوم، 2018، ص 60-74)

4. استخدم إجراءات مناسبة في تصميم المعادلة وتحليل البيانات عند ربط أو معادلة درجات المقاييس في الإصدارات اللغوية المختلفة للاختبار.

عند ربط أو معادلة نسختين لغويتين للاختبار ليصبح اختبار واحد، يمكن توفير العديد من الامكانيات. إذا تم استخدام مجموعة مشتركة من المفردات، يجب تقييم أداء هذه المفردات الشائعة عبر مجموعتي اللغة، وإذا لوحظ الأداء التفاضلي، ينبغي النظر في إزالتها من اختبارات الترجمة والتكيف. ليس لدى أنواع المفردات نفس إمكانية الربط أو المعادلة بين النسخ اللغوية. ويمكن رسم تقديرات لبارامترات الصعوبة والتميز المستمدة من نظرية الاستجابة للمفردة (IRT) في تحديد المفردات المشتركة الملاءمة والغير الملاءمة (Hambleton et al, 1991).

ولكن ربط أو معادلة الدرجات عبر نسختين لغويتين من الاختبار يتطلب افتراضات قوية حول البيانات. في بعض الأحيان، نكون أمام إشكالية افتراض أن النسختين المختلفتين للاختبار متكافئة، وبالتالي يتم استخدام درجات نسختي الاختبار بالتبادل. يمكن أن هذا الافتراض صحيح مع اختبارات الرياضيات لأن الترجمة أو التكيف عادة ما تكون مباشرة في هذا النوع من الاختبارات. خاصة إذا تم إنشاء نسختين من الاختبار بعناية، وبالتالي يمكن افتراض أن النسخة اللغوية الأصلية للاختبار تعمل في المجتمع الأصلي بطريقة مكافئة التي تعمل بها النسخة اللغوية من الاختبار للمجتمع المستهدفة. قد يكون لهذا الافتراض تبرير إذا كانت جميع الأدلة تشير إلى أن النسختين اللغويتين للاختبار متساويتان ولا توجد تحيزات في الطريقة تؤثر على النتائج في النسخة اللغوية المستهدفة للاختبار. (ITC, 2017, pp.01-68)

وهناك حلان آخران، ولكن لا يوجد حل مثالي أولاً، وهو إمكانية الربط بين مجموعة فرعية من المفردات التي تعتبر متكافئة أساساً في نسختي الاختبار. مثلاً، قد تكون المفردات هي التي كانت سهلة الترجمة والتكيف. من حيث المبدأ، يمكن أن يكون هذا حل مناسب ولكن يتطلب أن تكون معادلة المفردات وباقي مفردات الاختبار لقياس نفس البنية. الحل الثاني ينطوي على المعادلة أو الربط من

خلال عينة من المتقدمين للاختبار الذين يتحدثون اللغتين. مع أخذ العينة نسختي الاختبار، سيكون من الممكن إنشاء جدول تحويل للدرجات (المعايير). ولا يجب أن تكون العينة صغيرة جداً، ويجب ترتيب عرض شكلي الاختبار. وتتوقف هذه الطريقة على افتراض أن المستجيبين على يتقنون اللغتين.
اقتراحات للممارسة:

ربط أو معادلة درجات نسختي الاختبار المكيف، لا يخلو من مشكلات لأن جميع تصاميم التكافؤ لديها مشكلة واحدة على الأقل. ربما أفضل استراتيجية هي معالجة جميع الخطوات لإنشاء تكافؤ للدرجات كلي. إذا كانت الأدلة التي تتناولها الأسئلة الأربعة التالية ثابتة، بالنسبة لدرجات الاختبار في نسخته يمكن بالتالي التعامل معهما بالتبادل: (ITC, 2017, pp.01-68)

- هل هناك دليل على أن نسخة اللغة الأصلية واللغة المستهدفة للاختبار تقيسان نفس البنية؟ هل يكون لبنية المفهوم نفس العلاقة مع المتغيرات الخارجية الأخرى في الثقافة الجديدة؟
- هل هناك أدلة قوية على أنه تم ضبط مصادر التحيز مثلاً، لا توجد مشكلة في وقت الاختبار، وأشكال الاختبار مألوفة بالنسبة للمستجيبين، ليس هناك غموض حول التعليمات، لا يوجد اختلاف منهجي بين مجموعة وأخرى، تعليمات الاختبار مقننة (موحدة)، وعدم وجود أنماط الاستجابة الشاذة كالتقييمات المتطرفة، والاداء التفاضلي للمفردة؟
- هل مفردات الاختبار خالية من التحيز المحتمل؟ في هذه الحالة يمكن التحقق من مفردات نسختي الاختبار باستخدام الشكل التخطيطي لقيم مستوى الدلالة (p) أو قيم دللتا. وينبغي دراسة الدرجات التي لا تقع على طول خط المعدلة أو الربط الخطي لتحديد ما إذا كانت قيم معاملات الارتباط بنفس القيمة في كلتا اللغتين. وتوفر تحليلات الأداء التفاضلي للمفردة (DIF) أدلة أقوى على تكافؤ المفردات عبر المجموعات اللغوية والثقافية.
- إذا تمت محاولة ربط أو معادلة الدرجات، فيجب اختيار تصميم ربط مناسب وتنفيذه. وينبغي تقديم أدلة على صدق التصميم. (طباع و بلقيدوم، 2018، ص 60-74)

القسم الرابع: تطبيق أو إدارة الاختبار:

1. تحضير الأدوات وتعليمات تطبيق الاختبار بغرض التقليل من أي مشكلة لها علاقة بالثقافة واللغة التي قد تنتج عن إجراءات تطبيق الاختبار وأنماط الاستجابة التي قد تؤثر على صدق التفسيرات المستمدة من الدرجات.

وينبغي أن يبدأ تنفيذ المبادئ التوجيهية لإدارة الاختبار بتحليل جميع العوامل التي يمكن أن تهدد صدق درجات الاختبار في سياق ثقافي ولغوي محدد. وقد تكون الخبرة في إدارة أداة ما في سياق أحادي اللغة أو أحادي الثقافة مفيدة بالفعل في توقع المشاكل التي يمكن توقعها في سياق متعدد اللغات أو

متعدد الثقافات. فمثلاً، غالباً ما يعرف مطبقوا الاختبار ذوي الخبرة جوانب التعليم التي قد تكون صعبة بالنسبة للمستجيبين. والتي ربما تبقى صعبة بعد الترجمة أو التكيف. إن تطبيق أدوات القياس في سياق لغوي أو ثقافي جديد يمكن أن يطرح مشاكل لم تكن موجودة من قبل في التطبيقات الأحادية الثقافية. (ITC, 2017, pp.01-68)

اقتراحات للممارسة:

من المهم في هذا المبدأ التوجيهي توقع العوامل المحتملة التي قد تخلق مشاكل في إدارة الاختبار. وفي ما يلي بعض العوامل التي يجب مراعاتها لضمان النزاهة أو العدالة في إدارة الاختبار، وهي:

• وضوح تعليمات الاختبار بما في ذلك ترجمة تلك التعليمات، وآلية الإجابة على ورقة الإختبار، و الوقت المسموح به وهو أحد المصادر الشائعة للخطأ هو عدم إعطاء الوقت الكافي للإجابة على الاختبار، دافعية المشاركين في إكمال الاختبار، ومعرفة الهدف من الاختبار، وكيفية إعطاء الدرجة (التصحيح).

2. تحديد شروط الاختبار التي يجب متابعتها عن كثب لدى أفراد المجتمع محل الاهتمام.

الهدف من هذا المبدأ هو تشجيع مطوري الاختبارات على وضع تعليمات الاختبار والإجراءات المتعلقة بها مثل، ظروف تطبيق الاختبار والحدود الزمنية وما إلى ذلك التي يمكن متابعتها عن قرب عند جميع شرائح المجتمع المستهدف. يهدف هذا المبدأ في المقام الأول إلى تشجيع مطوري الاختبار على الالتزام بتوحيد أو تقنين تعليمات الاختبار. وفي نفس الوقت يمكن تقديم بعض التوضيحات للأفراد في المجموعات الفرعية داخل كل مجتمع، مثل إعطاء الوقت الإضافي، والطباعة بالحجم الكبير، وتوفير ظروف أكثر هدوءاً عند تطبيق الاختبار، وما إلى ذلك. والتي تعرف في مجال الاختبارات اليوم، باسم "تسهيلات الاختبار". والهدف من هذه الترتيبات هو خلق بيئة ملائمة للمشاركين للتعبير عما يشعرون به أو يعرفونه. (ITC, 2017, pp.01-68)

اقتراحات للممارسة:

هنا يتم التركيز على على أهمية المشاركين الذين يطبق عليهم الاختبار في ظل ظروف متشابهة قدر الإمكان. وهذا أمر ضروري خاصة إذا كانت الدرجات من نسختي الاختبار سوف تستخدم بالتبادل. وفيما يلي بعض الاقتراحات:

• ينبغي تكييف تعليمات الاختبار والإجراءات الأخرى وإعادة كتابتها بطريقة مقننة (موحدة)، لتكون مناسبة للغة والثقافة الجديدتين.

• وإذا تم تغيير تعليمات الاختبار والإجراءات الأخرى إلى الثقافات الجديدة، ينبغي تدريب مطبقي الاختبار على الإجراءات الجديدة والتأكيد على احترامها. (طباع وبلقيدوم، 2018، ص 60-74)

القسم الخامس: لتسجيل الدرجات والتفسير:

1. تفسير أي اختلافات في الدرجات بين المجموعات مع الإشارة إلى جميع المعلومات المتاحة والتي لها علاقة بالموضوع.

حتى لو تم تكييف الاختبار بإتباع إجراءات سليمة تقنياً، وتم تحديد صدق درجات الاختبار إلى حد ما، يجب أن نضع في اعتبارنا أن معنى الاختلافات بين المجموعات يمكن أن يفسر بطرق عديدة بسبب الاختلافات الثقافية. قدم Sireci (2005) طريقة لتقييم معادلة نسختين لغويتين مختلفتين من الاختبار عن طريق تطبيق نسختي اللغتين للاختبار لمجموعة من المشاركين الذين يتقنون اللغتين والذين ينتمون إلى نفس المجموعة الثقافية أو اللغوية المستهدفة. ولخص بعض خيارات تصميم البحوث لدراسات التكافؤ باستخدام مزدوجي اللغة، ونها إلى المتغيرات التي يجب السيطرة عليها، وقدم بعض الاقتراحات القيمة لتفسير النتائج. (ITC, 2017, pp.01-68)

اقتراحات للممارسة:

يمكن النظر في عدد من التفسيرات الممكنة، حسب موضوع البحث أو السياق الذي أجريت من أجله مقارنات بين المجموعات، مثلاً، من المهم استبعاد الدافع التفاضلي للأداء على الاختبار قبل الحكم على أفضلية أداء مجموعة على مجموعة أخرى. قد يؤثر السياق الثقافي بشكل كبير على الأداء على الاختبار. فعلى سبيل المثال، قد تكون مجموعة ما جزءاً من نظام تعليمي أقل فعالية من النظام التعليمي في المجموعة الأخرى. (طباع و بلقيدوم، 2018، ص 60-74)

2. لا تتم مقارنة الدرجات بين المجموعات إلا عندما يتم تحديد مستوى التباين (اللاتغاير) على المقياس الذي يتم تسجيل الدرجات عليه.

عندما تكون الدراسات المقارنة بين المجموعات اللغوية والثقافية هي المحور الرئيسي في تصميم الترجمة والتكييف، يجب وضع النسخ المتعددة اللغات للاختبار على مقياس مشترك، ويتم ذلك من خلال عملية تسمى "الربط" أو "المعادلة". والتي تتطلب حجم عينة كبير نوعاً ما، وأدلة تبين أن النسخة المكيفة من الاختبار لا تحتوي على تحيز للبنية، أو في طريقة التطبيق، أو تحيز في المفردات.

وقد حدد كل من Van de Vijver and Poortinga (2005) عدة مستويات لمعادلة الاختبارات بين المجموعات اللغوية والثقافية، وفائدتها في الفهم الجيد للمفهوم في الثقافتين. على سبيل المثال، أشاروا إلى أن تكافؤ القياس يتطلب أن يكون للمقاييس في كل مجموعة نفس وحدة القياس، وبالتالي ضمان أن الاختلافات بين الأشخاص داخل المجموعات لها نفس المعنى. على سبيل المثال، يمكن مقارنة الاختلافات بين الذكور والإناث في عينة صينية بعينة فرنسية. ومع ذلك، لا يمكن إجراء مقارنات درجات مباشر بشكل صحيح إلا عندما تظهر الدرجات أعلى مستوى من التكافؤ، والذي يسمى

بالتكافؤ التدريجي أو معادلة الدرجات الكلية، مما يتطلب أن يكون للمقاييس في كل مجموعة نفس وحدة القياس ونفس الأصل عبر المجموعات.

اقترحت عدة طرق سواء في إطار النظرية الكلاسيكية للاختبارات أو نظرية الاستجابة للمفردة لربط أو معادلة الدرجات اختبارين للترجمة والتكيف. واقترح كل من Cook and Schmitt-Cascallar (2005) أساساً لفهم الأساليب الإحصائية المتوفرة لمعادلة الاختبارات التربوية والنفسية وتوسيع نطاق استخدامها. ووصف وانتقد المؤلفون إجراءات ربط أو معادلة المقياس المستخدم في دراسات تكيف الاختبارات حالياً، كما وضحو بعض مشكلات الربط المختارة من خلال وصف وانتقاد نتائج ثلاثة دراسات أجريت على مدى السنوات العشرين الماضية لمعادلة درجات اختبار التقييم المدرسي والمسمى بـ "Prueba de Aptitud Académica".

اقتراحات للممارسة: النقطة المهمة هنا هي أنه لا ينبغي المبالغة في تفسير درجات الاختبار: تفسير النتائج استناداً إلى مستوى دليل صدق الاختبار المتاح. على سبيل المثال، لا تدلي ببيانات مقارنة حول مستويات أداء المجيبين في مجموعتين لغويتين ما لم يتم تحديد مدى ثبات تباين القياس بالنسبة لدرجات الاختبار الذي يتم مقارنته. (طباع و بلقيدوم، 2018، ص 60-74)

القسم السادس: توثيق الاختبار:

1. تقديم الوثائق التقنية لأي تغييرات، بما في ذلك الأدلة التي تم الحصول عليها لدعم التكافؤ، عند تكيف اختبار للاستخدام في مجتمع آخر.

أدرك العديد من الباحثين أهمية هذا المبدأ وأكدوا عليه. وقد حققت دراسة الاتجاهات الدولية في الرياضيات والعلوم (TIMSS) وبرنامج التقييم الدولي للطلاب (PISA) نجاحاً كبيراً في مراعاة هذا المبدأ من خلال توثيق التغييرات بدقة طوال عملية التكيف. وينبغي أن تتضمن الوثائق التقنية أيضاً تفاصيل كافية عن الطريقة التي يجب أن يتبعها الباحثين مستقبلاً لتكرار نفس الإجراءات المستخدمة مع نفس المجتمع أو على مجتمعات أخرى. كما ينبغي أن تتضمن معلومات كافية عن أدلة تكافؤ البنية البناء وتكافؤ التدرج إذا تم تدرج الاختبار باستخدام نظرية الاستجابة للمفردة لدعم استخدام الأداة في المجتمع المستهدف. ولكي يمكننا إجراء مقارنة بين المجتمعات، يجب أن تقدم الوثائق عن الأدلة المستخدمة لتحديد معادلة درجات الاختبار بين المجتمعات. (ITC, 2017, pp.01-68)

اقتراحات للممارسة. يجب أن تحتوي الاختبارات المكيفة على دليل تقني يوثق جميع الأدلة النوعية والكمية المرتبطة بعملية التكيف. من المفيد بشكل خاص توثيق أي تغييرات تم إجراؤها لاستيعاب الاختبار بلغة وثقافة أخرى. وفيما يلي مجموعة من الأسئلة التي تحتاج إلى إجابة:

- ما هي الأدلة المتوفرة لدعم فائدة البناء والاختبار المكيف في المجتمع المستهدف؟

- ما هي البيانات التي تم جمعها من المفردات ومن أي عينات تم جمعها؟
- ما هي البيانات الأخرى التي تم الحصول عليها لتقييم صدق المحتوى، والصدق المرتبط بالمحك والصدق البنائي؟
- كيف تم تحليل البيانات المختلفة؟
- ما هي النتائج؟ (طباع و بلقيدوم، 2018، ص60-74)
- 2. توفير الوثائق لمستخدمي الاختبار التي من شأنها دعم التطبيق الجيد في استخدام الاختبار المكيف مع أفراد المجتمع الجديد.
- يجب كتابة الوثائق لأولئك الذين سيستخدمون الاختبار في إطار التقييم العملي أو التطبيقي. وينبغي أن يكون متسقا مع الممارسة الجيدة التي حددتها المبادئ التوجيهية للجنة الاختبارات الدولية بشأن استخدام الاختبارات أنظر (www.InTestCom.org).
- اقتراحات للممارسة: وينبغي أن يقدم مطور الاختبار معلومات دقيقة حول الطرق تأثير السياقات الاجتماعية والثقافية والبيئية على الأداء في الاختبار. يجب أن يحتوي دليل المستخدم على:
- وصف للبنية التي يتم قياسها بواسطة الاختبار ووصف ملخص لإجراءات التكييف.
- تلخيص الأدلة الداعمة للتكييف، بما في ذلك الأدلة على الملاءمة الثقافية لمحتوى المفردات، وتعليمات الاختبار، وطريقة الاجابة، وما إلى ذلك.
- تحديد مدى ملاءمة استخدام الاختبار مع مجموعات فرعية مختلفة من المجتمع وأي قيود أخرى على الاستخدام.
- توضيح القضايا التي تحتاج إلى تفسير فيما يتعلق بالممارسات الجيدة في تطبيق الاختبار.
- توضيح ما إذا كان يمكن إجراء مقارنات بين المجتمعات وكيفية ذلك.
- توفير المعلومات اللازمة للتصحيح والمعايرة مثل جداول المعايير أو وصف كيفية وصول المستخدمين إلى إجراءات التصحيح فمثلاً هناك إجراءات تستخدم الحاسوب في تصحيح الاختبارات.
- تقديم مبادئ توجيهية لتفسير النتائج، بما في ذلك معلومات عن الآثار المترتبة على بيانات الصدق والثبات بشأن الاستدلالات التي يمكن استخلاصها من درجات الاختبار. (ITC, 2017, pp.01-68)

ملحق(01):الدليل الإرشادي لترجمة وتكييف الاختبارات والمقاييس الذي أصدرته اللجنة الدولية للاختبارات (ITC) في طبعته الثانية (2017).

ITC Guidelines for Translating and Adapting Tests Checklist

Here is a checklist to remind you of the eighteen ITC Guidelines. We recommend that you check those that you feel you have dealt with satisfactorily in your test translation/adaptation project, and then attend to those that remain unaddressed.

Pre-Condition Guidelines

- PC-1 (1) Obtain the necessary permissions from the holder of the intellectual property rights relating to the test before carrying out any adaptation.
- PC-2 (2) Evaluate that the amount of overlap in the definition and content of the construct measured by the test in the populations of interest is sufficient for the intended use (or uses) of the scores.
- PC-3 (3) Minimize the influence of any cultural and linguistic differences that are irrelevant to the intended uses of the test in the populations of interest.

Test Development Guidelines

- TD-1 (4) Ensure that the adaptation process considers linguistic, psychological, and cultural differences in the intended populations through the choice of experts with relevant expertise.
- TD-2 (5) Use appropriate translation designs and procedures to maximize the suitability of the test adaptation in the intended populations.
- TD-3 (6) Provide evidence that the test instructions and item content have similar meaning for all intended populations.
- TD-4 (7) Provide evidence that the item formats, rating scales, scoring categories, test conventions, modes of administration, and other procedures are suitable for all intended populations.
- TD-5 (8) Collect pilot data on the adapted test to enable item analysis, reliability assessment and other small-scale validity studies, so that any necessary revisions to the adapted test can be made.

Confirmation Guidelines

- C-1 (9) Select sample with characteristics that are relevant for the intended use of the test and of sufficient size and relevance for the empirical analyses.
- C-2 (10) Provide relevant statistical evidence about the construct equivalence, method equivalence, and item equivalence for all intended populations.
- C-3 (11) Provide evidence supporting the norms, reliability and validity of the adapted version of the test in the intended populations.
- C-4 (12) Use an appropriate equating design and data analysis procedures when linking score scales from different language versions of a test.

Administration Guidelines

- A-1 (13) Prepare administration materials and instructions to minimize any culture- and language-related problems that are caused by administration procedures and response modes that can affect the validity of the inferences drawn from the scores.
- A-2 (14) Specify testing conditions that should be followed closely in all populations of interest.

Score Scales and Interpretation Guidelines

- SSI-1 (15) Interpret any group score differences with reference to all relevant available information
- SSI-2 (16) Only compare scores across populations when the level of invariance has been established on the scale on which scores are reported.

Documentation Guidelines

- Doc-1 (17) Provide technical documentation of any changes, including an account of the evidence obtained to support equivalence, when a test is adapted for use in another population.
 - Doc-2 (18) Provide documentation for test users that will support good practice in the use of an adapted test with people in the context of the new population.
-

قائمة المصادر والمراجع:

- أناستازي وسوسن، أرينا.(2015). القياس النفسي (ترجمة: صلاح الدين محمود علام). الناشر: دار الفكر، عمان، الأردن.
- أبو ناهية، صلاح الدين.(1994). القياس التربوي. مصر: مكتبة الأنجلو مصرية.
- أبو هاشم، السيد.(2006). الخصائص السيكمترية لأدوات القياس في البحوث النفسية والتربوية باستخدام SPSS، كلية التربية، جامعة الملك سعود. المملكة العربية السعودية.
- ليون، أ. تايلر.(1998). الاختبارات والمقاييس. الناشر: دار الشروق، القاهرة، مصر.
- دويدار، عبد الفتاح (1995). أسس علم النفس التجريبي، الناشر: دار النهضة العربية للطباعة والنشر بيروت- لبنان.
- الزامل، علي حسين (2017). بناء وتقنين المقاييس النفسية. الناشر: دار الكتب والوثائق بغداد- العراق.
- سوسن، شاكرا مجيد (2005). أساسيات بناء الاختبارات والمقاييس النفسية والتربوية. ط3. دار النشر: مؤسسة علاء الدين للطباعة والتوزيع. دمشق. سوريا.
- سوسن، شاكرا مجيد (2014). أسس بناء الاختبارات والمقاييس النفسية والتربوية. ط3. دار النشر: مركز ديونو لتعليم التفكير.
- سامية، براهيمي وآخرون.(2018). كتاب جماعي بناء وتكييف أدوات القياس النفسي والتربوية من منظور النظرية الكلاسيكية والنظرية المعاصرة. الناشر: ناصري للطباعة والنشر، المسيلة، الجزائر.
- صفوت، فرج.(2007). القياس النفسي، ط6، الناشر: مكتبة الأنجلو مصرية، القاهرة، مصر.
- طباع، فاروق وبلقيدوم، بلقاسم.(2018). كتاب جماعي بناء وتكييف أدوات القياس النفسي والتربوية من منظور النظرية الكلاسيكية والنظرية المعاصرة. الناشر: ناصري للطباعة والنشر، المسيلة، الجزائر.
- علام، محمود صلاح الدين.(2000). القياس والتقويم النفسي والتربوي، الناشر: دار الفكر العربي، القاهرة، مصر.
- عزيزان، خوخة و بحاش، عبد الحق.(2018). كتاب جماعي بناء وتكييف أدوات القياس النفسي والتربوية من منظور النظرية الكلاسيكية والنظرية المعاصرة. الناشر: ناصري للطباعة والنشر، المسيلة، الجزائر.

الغامدي، حسين عبد الفتاح.(1424هـ). تكنيك رورشاخ، سلسلة البحوث التربوية والنفسية، جامعة أم القرى، المملكة العربية السعودية.

الفاعوري، أيهم علي. (ب.ت). الاختبارات الإسقاطية. <https://oclademy.uk/pdf-booksL> فيصل، عباس. (2001). الاختبارات الإسقاطية نظرياتها تقنياتها إجراءاتها، الناشر: دار المنهل اللبناني للطباعة والنشر.

فيصل، عباس.(1996). الاختبارات النفسية تقنياتها وإجراءاتها، الناشر: دار الفكر، بيروت، لبنان. قدوري، رايح (2018). بناء وتكييف أدوات القياس النفسي والتربوي من منظور النظرية الكلاسيكية والنظرية المعاصرة (كتاب جماعي). الناشر: نواصري للطباعة والنشر-المسيلة.

محمود وآخرون.(2009). القياس النفسي والتربوي، الناشر: دار المسيرة للنشر والتوزيع، عمان، الأردن. ملحم، سامي محمد (2012). القياس والتقويم في التربية وعلم النفس. ط6. الناشر: دار المسيرة للنشر والتوزيع. عمان. الأردن.

الناشف، سلى (2001). دليلك في تصميم الاختبارات. الناشر: دار البشير عمان الأردن. النيهان، موسى.(2013). أساسيات القياس في العلوم السلوكية. ط2، الناشر: دار الشروق للنشر والتوزيع، عمان، الأردن.

نهى الزهراني، رغدة الشريف (ب.ت). اختبارات الاستعدادات. <https://www.emaze.com/@aoofqirf/>

Anastasi, A. (1985). Psychological testing: Basic concepts and common misconceptions. In Annual Meeting of the American Psychological Association, 1984, Toronto, ON, Canada; This lecture was presented at the aforementioned meeting.. American Psychological Association.

Allen, Robert M. "An Analysis of Edwards Personal Preference Schedule Intercorrelations for a Local College Population." The Journal of Educational Research, 1958, vol. 51, no. 8, pp. 591–597. https://americanhistory.si.edu/collections/search/object/nmah_692350

APA Dictionary of Psychology, December 27, 2022.

<https://dictionary.apa.org/test-profile>

Bernaude, J. L. (2014). Méthodes de tests et questionnaires en psychologie. Dunod.

Harsh, Katara ; Dheeraj, Vaidya .(2022). Percentile Rank Formula.

[/https://www.wallstreetmojo.com/percentile-rank-formula](https://www.wallstreetmojo.com/percentile-rank-formula)

International Test Commission. (2017). The ITC Guidelines for Translating and Adapting Tests (Second edition). <https://www.intestcom.org/>.

Irwing, P., & Hughes, D. J. (2018). Test Development. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), The Wiley Handbook of Psychometric Testing: A

- Multidisciplinary Reference on Survey, Scale and Test Development, V.I (pp. 4-47). Hoboken, NJ: Wiley.
- Kolbe, L., & Jorgensen, T. D. (2019). Using restricted factor analysis to select anchor items and detect differential item functioning. *Behavior Research Methods*, 51(1), 138-151.
- Muñiz, J., Elosua, P., Padilla, J. L., & Hambleton, R. K. (2016). Test adaptation standards for cross-lingual assessment.
- Price, L. R. (2016). *Psychometric methods: Theory into practice*. Guilford Publications.
- Reynolds, C. R., & Livingston, R. A. (2021). *Mastering modern psychological testing* (p. 573). Springer International Publishing.
- Sireci, S. G., & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language testing*, 20(2), 148-166.
- Van de Vijver, F., & Leung, K. (1997). Methods and data analysis of comparative research. *Handbook of cross-cultural psychology*, 1, 257-300.