محاضرات فى

(القياس والتقويم)

إعداد
قسم علم النفس

الفرقة الثالثة عام علم النفس

رؤيـة الـكـلـيـة :

تسعى الكلية إلى مساعدة الجامعة فى تحقيق أهدافها الاستراتيجية من خلال أن تكون واحدة من الكليات المتميزة والمنافسة داخلياً وخارجياً فى التعليم وخدمة المجتمع والبحث العلمي من خلال تحقيق مستوى رفيع من الأداء وتقديم خريج متميز يقابل الاحتياجات المتعددة لسوق العمل المحلى والخارجى.

رسـالـة الـكـلـيـة :

تهدف كلية التربية بالغردقة إلى التميز من خلال :

- ❖ إعداد المربين والمعلمين المتخصصين والقادة فى مختلف التخصصات التربوية .
- ❖ تنمية القدرات المهنية والعلمية للعاملين فى ميدان التربية والتعليم بتعريفهم بالاتجاهات التربوية الحديثة .
- ❖ إجراء البحوث والدراسات فى التخصصات التربوية المختلفة بالكلية.
- ❖ نشر الفكر التربوى الحديث وإسهاماته لحل مشكلات البيئة والمجتمع .
- ❖ تبادل الخبرات والمعلومات مع الهيئات والمؤسسات التعليمية والثقافية .
- ❖ تنمية جوانب شخصية الطلاب ورعاية الموهوبين والمبدعين .

مفهوم القياس وأهميته:

يعد القياس السلوك الإنساني مدخلاً اساسياً لفهم الفرد كإنسان متمايز وكعضو فى مجتمع ، وبذلك يجعل من الممكن التقييم الموضوعي لهذا السلوك ، وعن طريق القياس نتوصل إلى معلومات الإستناد إليها فى إصدار أحكام عن الوضع الراهن للأفراد والجماعات وتقدير إمكانات الأفراد المستقبلية فى مختلف مجالات السلوك الإنساني .

ومفهوم القياس Measurement له تعريفات متعددة ، لعل أبسطها هو أنه وصف البيانات المتعلقة بخصائص الأشياء باستخدام الأعداد أو الجوانب الكمية فى وصف سمات أو خصائص الأفراد (Guiford. 1954) ويرى سميث ، وادمز (Smith & Adams, 1972 أن القياس بمعناه الواسع هو الجمع المنظم للمعلومات بترتيب معين ، وهذا يتضمن عملية جمع وتنظيم المعلومات ، وناتج هذه العملية .

ويعرف " نانالي (Nunnally 1972)" القياس بأنه يشتمل على قواعد تعيين أعداد للأشياء بحيث تمثل مقادير سمات هذه الأشياء ، ويتفق هذا التعريف مع ما سبق أن قدمه " ستيفنز (Stevens, 1951) " من أن القياس هو تعيين أعداد أو رموز رقمية للأشياء أو الأحداث وفقاً لقواعد ، فالقياس إذن يتطلب إجراءات أو عمليات معينة ، استناداً إلى قواعد محددة تستخدم فى مقارنة بين الأشياء أو الأحداث وفقاً لمعيار Standard أو ميزان Scale معرف تعريفاً دقيقاً.

وعلى الرغم من أن هذه القواعد تكون واضحة فى بعض الحالات فى القياس الفيزيائي مما لا يتطلب صياغات تفصيلية لها ، كما فى حالة قياس الطول، إلا أن هناك حالات كثيرة ربما تتطلب إجراءات معقدة ، مثل قياس المقادير المختلفة لمكونات المركبات الكيميائية.

وقواعد قياس معظم السمات أو الخصائص النفسية ليست بسيطة أو مباشرة، مثل: ذكاء الأطفال، ومقدار الاسترجاع فى تعلم الترابط الثنائي، ومستوى الدافع، والاتجاه نحو جماعة معينة غير أن القياس النفسي والتربوي يتبع الخطوات المتبعة فى القياس الفيزيائي فنحن نُعرف السمة أو البعد المراد قياسه ، ونحدد عمليات القياس والقواعد التى

نستند إليها فى القياس، وكذلك ميزان معين Scale لوحدات القياس يُعبر عن القياس الناتج.

ويمكن الإفادة من القياس الكمي فى تحقيق ما يلي:

(أ) **مرونة الوصف:** فتطبيق إجراءات القياس تُمكنًا من التمييز بين الأفراد، ووصف الفروق الفردية.

(ب) تيسير عملية التفسير: فالقياس لا يقتصر على تحديد ووصف الفروق الفردية، وإنما يُمكننا من ترتيب الأفراد بالنسبة للمتغير موضع القياس، وبذلك يمكن اشتقاق تفسيرات استناداً إلى الوضع أو الأداء النسبي لهؤلاء الأفراد فى مقابل معيار معين.

(ج) تحديد أنماط السلوك: فمرونة الوصف تُمكننا من تحديد وقياس الفروق الفردية فى أنماط سلوكية مختلفة ، حيث يمكن الربط بين هذه الأنماط بطرق مفيدة.

(د) اختزال البيانات: فتحديد ميزان القياس ووحداته يسمح باستخدام أساليب رياضية وإحصائية تلخص كميات كبيرة من البيانات، كما سيتضح فى الفصل الثاني.

منطق السمات الإنسانية فى القياس:

تبين مما سبق أننا لا نقيس الفرد وإنما نقيس خاصة أو سمة معينة لديه ، فنحن نقيس طول المنضدة أو درجة حرارة الجسم، ونقيس ذكاء الطفل، والنضج الاجتماعي للمراهق، وعندما نتعامل مع أبسط الخصائص الفيزيائية مثل الطول، فإننا لا نتساءل عن معنى أو تعريف هذه الخاصة، حيث إن تعريف الطول محدد ومتفق عليه منذ زمن بعيد، ولكن هناك خواص فيزيائية كثيرة أخري تتطلب الاتفاق على المقصود منها، مثل: المقاومة، والتحمل، والإجهاد، والمتانة، وعدم الاتفاق على معاني أو تعريفات هذه الخواص يجعل هناك اختلافاً فى الإجراءات المناسبة لقياسها.

وتزداد حدة هذه المشكلة فى علم النفس والتربية، حيث تفتقر كثيرا من السمات والخصائص الإنسانية إلى تعريفات محددة متفق عليها ، وتختلف دقة هذه التعريفات باختلاف مدى تعقد السمة المعينة، ويؤثر ذلك بالطبع فى مدى دقة قياس هذه السمات، فالقياس يتعلق بمظهر معين من مظاهر الأشياء المراد قياسها، حيث إننا لا نقيس الشئ

، وإنما نقيس بعض خصائصه وكذلك لا نقيس الفرد ، وإنما نقيس سمة معينة لديه مثل : الذكاء ، والقلق ، والاتزان الانفعالي ، والقيادة ،٠٠٠ إلخ ، ونعني بالسمة مجموعة مترابطة من السلوك الذى يحتمل حدوثه معاً ، أي أن السمة ليست صفة منفردة ، وإنما تعد تكويناً فرضياً Construct ، او مفهوماً مجرداً ، وليست شيئا ملموساً ولتوضيح ذلك يفترض أن طفلاً استطاع أن يحل مجموعة من التمارين الحسابية المتعلقة بالاستدلال العددى ، والاستدلال اللفظي ، والمتشابهات فجميع هذه المظاهر السلوكية يمكن أن تندرج تحت التفكير المجرد ، أو حل المشكلات ، ولهذا فإننا ربما نستنتج أن هناك سمة عقلية تنظم أو تعكس هذه المجموعة المترابطة من السلوك الذى يمكن ملاحظته وتسجيله فى مواقف اختبارية مقننة ، ويمكن أن نطلق عليها سمة الاستدلال المجرد (Abstract Reasoning) لذلك فإن السمات الإنسانية يمكن تعريفها فى ضوء السلوك الذى يمكن ملاحظته ملاحظة مباشرة وتفسيره فى إطارها والاختبار الذى يصمم لقياس سمة معينة ينبغي أن يتضمن مجموعة من المواقف التى تستدعي السلوك المناسب ، ونستدل على موقع الفرد على متصل هذه السمة من استجابته لهذه المواقف .

وهذا يعني أن السمات الإنسانية لا تقاس قياسًا مباشراً ، وإنما يمكن الاستدلال عليها من أنماط سلوكية معينة ، تنظمها هذه السمات ، أي أن قياسها يكون غير مباشر ، وهذا هو الحال دائماً فى القياس النفسي والتربوي ، فالسمات ينعكس أثرها فى مجموعة من الأنماط السلوكية المترابطة التى يمكن ملاحظتها مباشرة فى مواقف اختبارية مبتكرة ، لذلك فإن تحديد هذه السمات ، وبناء اختبارات لقياسها أو تقييمها ليست بالأمر الهين ، فهو يعد عملاً فنياً وتقنياً متخصصاً ، ويتطلب جهداً كبيراً .

إن الأخذ بمفهوم السمة الذى ذكرناه يجعلنا نستنتج ما يلي :

يمكن القول بأن القياس النفسي والتربوي نسبي وليس مطلقاً، وهذا يعني أن هناك متصل Continuum من السلوك نحاول عن طريق عملية القياس أن تحدد موضع الفرد بالنسبة لأقرانه على هذا المتصل، فقياس الذكاء أو الاستعدادات العقلية مثلاً يعتمد على تطبيق اختبار مناسب ثم إعداده بطريقة معينة بحيث يشتمل على مهارات محددة

يؤديها الفرد، مثل : حل مشكلات استدلال ، تعرف أشكال، عمليات حسابية ورياضية، معاني كلمات .... إلخ ، ومقارنة الدرجة التى يحصل عليها بمتوسط درجات اقرانه ، بحيث نستطيع تحديد المركز النسبي للفرد فى الذكاء أو الاستعدادات لذلك تجري بعض العمليات الاحصائية على الدرجات الخام لتحويلها إلى نوع من الدرجات يسير هذه المقارنة ، كما سيتضح فى الفصل الثالث .

فالقياس النفسي والتربوي إذن يهدف لتكميم السمات والخصائص الإنسانية، أي تعبير أعداد أو قيم عددية تناظر سلوك أو أداء الفرد، وهذا يعني ما ذكرنا أن لدينا متصلاً نقيس عليه سمة معينة، وتهدف عملية القياس لتحديد موضع الفرد على هذا المتصل طبقاً لقواعد محددة ومن الجدير بالذكر أن بعض أساليب وأدوات التقويم التربوي لا تؤدي إلى قيم عددية ، أي أن نتائجها ليست كمية، ولذلك ربما لا يكون مناسباً اعتبار هذه النتائج قياسات .

هناك مواقف يمكن أن يقارن فيها أداء الفرد بمحك أداء مطلق Criterion يتفق عليه أو يعد أساسياً ، لكي يؤدي الفرد سلوكاً معيناً بطريقة سليمة، مثل : مهارة العزف على آلة موسيقية ، أو مهارة الطبيب فى إجراء عملية جراحية، أو مهارة إصلاح عطل معين فى السيارة ، أو غير ذلك . ويمكن أيضاً فى مثل هذه الحالات مقارنة أداء الفرد بمتوسط أداء جميع أفراد المهنة المعنية ، ولعل هذا المدخل يناسب بدرجة أفضل القياس التربوي والمهني ، بينما يناسب المدخل الأول القياس النفسي.

على الرغم من تشابه عمليات القياس الفيزيائي والقياس النفسي والتربوي، إلا أن القياس الفيزيائي يتعامل بوحدات معيارية Measurement Units متفق عليها، مثل: الكيلو متر، والكيلو جرام، والفولت ، أما القياس النفسي والتربوي فإنه يفتقر إلى مثل هذه الوحدات، فليس هناك وحدة للتحصيل فى الحساب، أو وحدة للقدرة المنطقية تتميز بوجود صفر مطلق، ووحدات متساوية على متصل ميزان القياس، كما سيتضح بعد قليل .

غير أن هناك مداخل متعددة للتوصل إلى وحدات قياس متسقة للسمات الإنسانية مع تأكيد أن هذا الاتساق يعد أمراً نسبياً ، حيث أنه يوجد عدد قليل من وحدات القياس المتسقة المطلقة فى مجالات العلوم النفسية والتربوية.

فالقياس النفسي بخاصة يستند إلى الأداء الملاحظ لمجموعة معيارية (نمطية) من الأفراد، حيث يقارن أداء الفرد بأداء هذه المجموعة المعيارية ، أي بمعيار الجماعة Group Norm كما سيتضح فى الفصل الثالث.

على الرغم من أن تعريف ستيفنز Stevens للقياس الذى سبق أن اشرنا إليه يؤكد قواعد التكميم Quantification Rules ويعتبرها أساساً لعملية القياس ، إلا أنه لم يحدد طبيعة هذه القواعد أو حدودها، وربما يرجع ذلك إلى أنه ينبغي التمييز بين كيفية القياس المستمدة من عملية القياس ، وكيفية التحقق من صدق أدوات القياس ، أو تحديد منفعتها، فهذه القواعد تعد افتراضات ينبغي التحقق من صحتها بالأساليب الإمبريقية فى ضوء محاكات خارجية عن عملية القياس ذاتها ، حيث يتم تحديد مدى وجود علاقات وظيفية بين المقياس وبعض المتغيرات المهمة ، وسوف أوضح ذلك بمزيد من التفصيل فى الفصل الرابع .

ومن الجدير بالذكر أنه ينبغي تحديد مجموعة واضحة من القواعد استناداً إلى نموذج استدلالي معين، أو الخبرة السابقة ، أو الحكم الصائب لمصمم الاختبار أو المقياس، ومحك اختبار هذه القواعد هو مدى فائدة المقياس فى تفسير الظاهرة المستهدفة.

ومن هذا يتبين أن علم القياس النفسي والتربوي يواجه كثيراً من المشكلات والصعوبات كما هو الحال فى بقية العلوم ، غير أنه ربما يبدو أن هذه الصعوبات تزداد فى العلوم النفسية والتربوية، وأحد الأسباب الرئيسية لذلك أن الإنسان الذى هو محو القياس فى هذه العلوم دائم التغير ، وأحياناً يصعب اخضاعه للضبط فكما سبق أن أشرنا يستطيع عالم الكيمياء مثلاً التعامل مع معظم عينات دراسته من المواد الكيميائية بالطريقة التى يراها، وتظل هذه العينات فى كثير من الأحيان ثابتة ومتسقة، ويمكنه

تقسيمها أو خلطها ، ولكن عينات الأطفال التى يقوم عالم النفس بدراستها يصعب بدرجة كبيرة أن تحقق هذا الاتساق وضبط الظروف ومع هذا حدث تقدم ملحوظ فى طرق تصميم المعالجات التجريبية للتغلب على ذلك، حيث أصبح من الممكن قياس مختلف السمات الإنسانية بأساليب متنوعة.

وينبغي تأكيد أن القياس مجرد أداة ووسيلة لغاية، وأهميته تكمن فى درجة الإفادة من نتائجه فى تحسين أداء الأفراد فى مختلف مجالات الأعمال، وإثراء تعلم الطلاب فى المؤسسات التعليمية، غير أنه تظل هناك حاجة إلى مزيد من الجهود المبتكرة لتطوير أساليب وأدوات القياس المتوفرة فى وقتنا الحاضر ، ومعاونة المشتغلين بالقياس ، والمربين وغيرهم فى استخدام الأدوات المتوافرة استخداماً أكثر فاعلية ، ولعل البحوث السيكومترية والإديومترية المعاصرة ، والتطورات التى أصبحت تميز علم القياس النفسي والتربوي ، أسهمت اسهاماً ملحوظاً فى هذا الشأن ، ويتمثل ذلك فى النظرية السيكومترية المعاصرة التى يُطلق عليها نظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT)، والقياس والتقويم البدلي Alternative Assessment ، حيث أفردت لكل منها مرجعاً مستقلاً سبق نشره .

خصائص موازين القياس:

تختلف السمات والخصائص الإنسانية فى مستويات قياسها ، فخاصية نوع الفرد ، أو مستوى تعليمه، أو حالته الاجتماعية ، تختلف فى مستوى قياسها عن خاصية ترتيب الطفل بين إخوته ، أو ترتيبه فى فصله ، كما تختلف عن خاصية الذكاء، او التحصيل أو الميول، فالخصائص فى الحالة الأولي تعد من المتغيرات الكيفية التصنيفية ، وفى الحالة الثانية تعد من المتغيرات الرتبية، وفى الحالة الثالثة تعد من المتغيرات الكمية القياسية ، وقد سبق أن ذكرنا أن القياس هو تعيين أعداد للسمات أو الخصائص طبقاً لقواعد معينة ، ولكي تمثل هذه الأعداد السمة أو الخاصية المطلوبة يجب أن يكون هناك تناظراً إحادياً بين بعض خصائص النظام العددى ، والعلاقات بين مختلف مقادير الخاصية المراد قياسها، فإذا قمنا بتعيين أعداد تعبر عن السمات أو الخصائص المختلفة

، فإننا نستطيع أن نتعامل مع هذه الأعداد بطرق رياضية معينة ، ونتوصل من ذلك غلى استنتاجات يمكن أن نعيد تطبيقها على الظاهرة المقاسة فالخصائص أو السمات الإنسانية يمكن وصفها عن طريق الأعداد بشرط أن يكون هناك تناظراً أحادياً بين هذه الخصائص أو السمات والنظام العددى المستخدم ، فهناك خصائص معينة للأعداد ينبغي أن يوجد ما يناظرها فى هذه الخصائص أو السمات المراد قياسها وهى: خاصية التفرد أو التميز ، وخاصية الترتيب، وخاصية قابلية الجمع ، وليس من الضروري أن يكون للسمة الإنسانية التى نُعين لها الأعداد الخصائص الثلاث السابقة لكي نتمكن من قياسها ، إلا أن الاستفادة من استخدام الأعداد فى القياس تعتمد على مدى توافر هذه الخصائص فى السمة المراد قياسها.

وتعتمد مستويات أو موازين القياس Measurement Seales على ثلاث خصائص أساسية يوضحها الشكل التخطيطي (١– ١) التالي :

خصائص مستويات (موازين) القياس		
المقدار	تساوي المسافات	الصفر المطلق

شكل (١–١) يوضح مستويات (موازين) القياس

فالميزان Scale يحقق هذه الخاصية إذا كانت إحدى حالات سمة معينة يمكن الحكم بأنها أكبر أو أصغر أو تساوي حالة أخري للسمة نفسها ، فإذا استطعنا على ميزان الطول مثلاً أن نقول أن فرداً معيناً أطول من فرد أخر ، فإن هذا الميزان يحقق خاصية المقدار ، أما إذا عينت أعداد لفرق كرة القدم للتعرف بكل فريق ، فإن الأعداد فى هذه الحالة لا تحقق هذه الخاصية ولكن إذا كانت هذه الأعداد تُعبر عن ترتيب كل فريق بحسب عدد المباريات التى فاز فيها ، فإن نظام الاعداد الجديد يحقق خاصية " المقدار"

وهذه الخاصية تعني أن مقدار السمة الممثلة بوحدة قياس على الميزان تكون متساوية بغض النظر عن موقع هذه الوحدة على الميزان وبعبارة أخري تتحقق هذه الخاصية إذا كان الفرق بين نقطتين فى أي موقع على الميزان تحمل نفس معني الفرق

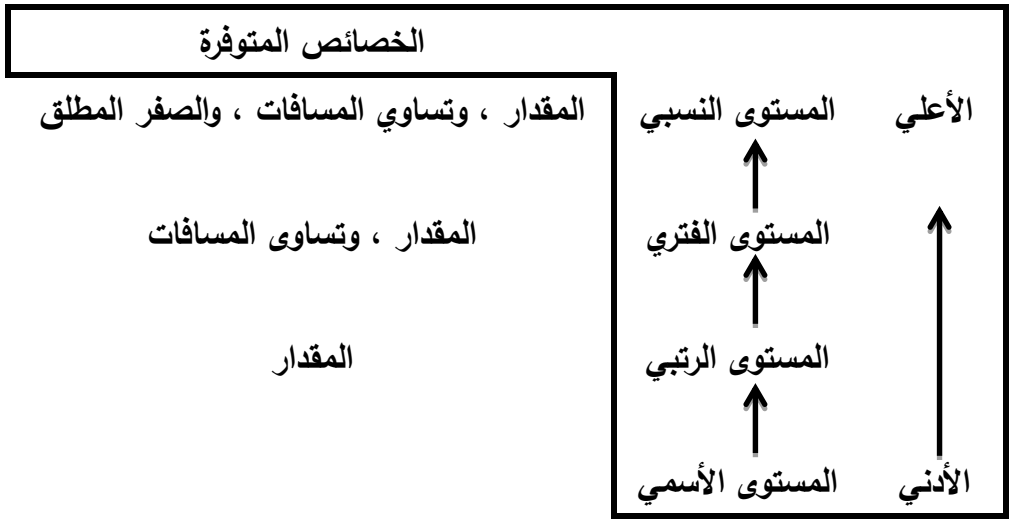
بين نقطتين أخريتين تختلفان بعدد مساوى من وحدات الميزان ، فالفرق بين ٢سم ، ٦ سم على المسطرة لا يختلف معناه عن الفرق بين ٨ سم ، ١٢ سم ، حيث إن هذا الفرق ٤سم فى الحالتين ، والسنتيمتر يعكس قدراً معيناً من الطول بغض النظر عن موقعه على المسطرة .

وعلى الرغم من أن خاصية تساوي المسافات تبدو بسيطة ، إلا أنه يندر أن نجد دليلاً على أن الاختبار النفسي أو التربوي يحقق هذه الخاصية ، فالفرق بين نسبتي الذكاء ٦٠ ، ٦٥ ليس له نفس معنى الفرق بين نسبتي الذكاء ١١٥، ١٢٠ ، على الرغم من أن الفرق فى كل من الحالتين ٥ نقاط ، وذلك لاختلاف مستوى الذكاء فى كل منهما .

تتحقق هذه الخاصية إذا انعدمت السمة موضع القياس ، فإذا كان عدد نبضات القلب صفراً ، فإن هذا يعني أن القلب توقف عن العمل ، أما إذا اعتبرنا أن أقل تقدير لسمة العدوانية (١) فإن هذا لا يعني انعدام السمة ، وحتى إذا كان الصفر نقطة من نقاط ميزان قياس العدوانية ، فإنه لا يعني بالضرورة انعدام العدوانية انعداماً كلياً ، لذلك فإن هذا الميزان لا يحقق خاصية الصفر المطلب ، والحقيقية أنه يصعب بدرجة كبيرة تعريف نقطة صفر مطلق لكثير من السمات النفسية .

يمكن التمييز بين أربعة مستويات أو موازين القياس استناداً غلى الخصائص الثلاث السابقة ، وتختلف هذه المستويات فى تحقيقها لهذه الخصائص لذلك فإن هذه المستويات هرمية بمعني أن المستويات العليا تحقق خصائص جميع المستويات الأدني منها بالإضافة إلى خصائص تميز المستوى المعين ، وكل منها يمثل مستوى من مستويات التقدير الكمي للسمة المراد قياسها ، كما يسمح بعمليات حسابية مختلفة .

ويوضح شكل (١ – ٢) التالي هذه المستويات والخصائص التى تتحقق فيها:



شكل (١-٢) مستويات القياس وخصائصها

وفيما يلي توضيحاً لكل من هذه المستويات الأربع :

المستوى الأسمى:

ويعد أبسط مستويات القياس ، حيث يقتصر على تصنيف الأفراد أو الأشياء فى أقسام متنافية مثني مثني وفقاً للسمة المقاسة ، ونعني بذلك أن الفرد لا يجوز أن ينتمي إلى أكثر من قسم واحد فى الوقت نفسه بالنسبة لسمة معينة ، ولكن يمكن بالطبع تصنيفه فى أكثر من سمة ، كأن يكون ذكراً ، ومستوى تعليمه جامعي ، ومتزوج مثلاً .

أي أن القياس فى هذه الحالة يتناول فروقاً كيفية أو نوعية ، وعادة تشير الأعداد الناتجة إلى تكرارات للأفراد الذين ينتمون إلى كل من هذه الأقسام ، وبالتالي لا نستطيع إجراء أي من العمليات الحسابية الأربع على هذه الأعداد لأنها تُعبر عن تكرارات لأقسام مختلفة النوع ، أي منفصلة ، وعملية تحديد التساوي هى العملية الأساسية فى هذا المستوى ، حيث يمكن فقط تحديد أن جميع الأفراد الذين ينتمون إلى قسم معين يمتلكون خاصة مشتركة لا يمتلكها من لا ينتمون إلى هذا القسم ، لذلك فإن الميزان الأسمي فى القياس محدود الفائدة فى علم النفس والتربية ، نظراً لأنه يسمح فقط بتصنيفات عريضة دون تحديد أي مقدار كمي وهو ما يهمنا فى القياس ، حيث إنه يعتمد على عملية العد

Counting ، أي تحديد عدد الأفراد الذين ينتمون إلى قسم معين ، وبذلك تتضح الفروق الكيفية بين الأفراد ، وليست الفروق الكمية .

المستوى الرتبي Ordinal Scale : يعتمد هذه المستوى على ترتيب الأفراد فى سمة أو خاصية معينة من الأكبر إلى الأصغر ، وتدل الأعداد الناتجة على الوضع النسبي للأفراد فى هذه السمة ، فهذا الطالب ترتيبه الأول فى اللغة العربية بالنسبة لطلاب فصله ، وهذا ترتيبه الثاني ، وهكذا ، وبالطبع لا تتساوى بالضروري الفروق بين درجات الطلاب من ذوي الرتب المتتالية ، وكثير من السمات أو الخصائص الإنسانية الكيفية يمكن ترتيبها بهذا الشكل ، لذلك فإن هذه الأعداد الرتبية لا تدل على مقدار مطلق للسمة لدى الفرد فى المجموعة ، ونظراً لصعوبة تحقق تساوي المسافات وتوافر صفر مطلق فى كثير من الظواهر النفسية والتربوية فإن المستوى الرتبي شائع الاستخدام فى قياس هذه الظواهر ، لأنه يحقق فقط خاصية " المقدار " .

المستوى الفتري Interval Seale : يتميز هذه المستوى بأنه يحقق خاصيتي المقدار ، وتساوي المسافات على ميزان قياس سمة أو خاصة معينة ، فالترمومتر مقسم إلى وحدات متساوية ، والفرق بين درجتي الحرارة ٣٥ ، ٣٧ يساوي الفرق بين الدرجتين ٣٩ ، ٤١ ، أو بين أي درجتين متجاورتين آخريتين.

ومما هو جدير بالذكر أن معظم الاختبارات المستخدمة فى قياس الذكاء ، أو التحصيل ، أو الاستعدادات ، أو الميول ، أو الشخصية ، تعتمد على هذا المستوى ، إذ يمكن اعتبار الفرق بين نسبتي الذكاء ٨٠ ، ١٠٠ فى اختبار للذكاء يكافئ الفرق بين نسبتي الذكاء ١٠٠ ، ١٢٠ أو ١٢٠ ، ١٤٠ بشروط معينة ، وهكذا ..

غير أنه لا يجوز أن نقول أن نسبة الذكاء ٨٠ نصف نسبة الذكاء ، ١٦٠ ، أو أن تحصيل الطالب الذى حصل فى اختبار تحصيلي مقنن فى الحساب على الدرجة ٩٠ ضعف تحصيل طالب أخر حصل على الاختبار نفسه على الدرجة ٤٥ ، وذلك لعدم وجود " صفر " مطلق على تدريج المقياس ، إذ لا يوجد فرد ذكاؤه صفراً ، أو تحصيله صفراً ، لذلك فإن نقطة الصفر على هذا التدريج تكون اعتبارية كما هو الحال فى نقطة الصفر على التدريج العددى للترمومتر ، إذ أن هذا الصفر لا يعني انعدام الحرارة وإنما

هو مستوى حرارة اتفق علماء الفيزياء على استخدامه كنقطة بدء قياس هذه الظاهرة على هذا التدريج ، ويمكن أن يناظر الصفر الاعتبارى فى الاختبارات والمقاييس النفسية والتربوية متوسط توزيع درجاتها، ويهتم علماء القياس بالحصول على ميزان فتري فى الاختبارات النفسية والتربوية لأنه يمكننا من تحويل الدرجات تحويلاً خطياً Linear Transformation في هذه الحالة ، أي تتحول الدرجات من ميزان إلى ميزان أخر وحداته مختلفة ، كما هو الحال عند تحويل درجات الحرارة المئوية تحويلاً خطياً إلى درجات فهرنهيتية أو العكس ، وهذا يعني أن درجات الاختبارات أو المقياس المختلفة يمكن وضعها على ميزان مشترك Common Scale ، وبالتالي يمكن مباشرة مقارنة أداء الأفراد على مقاييس مختلفة ، كما يمكن استخدام معظم الأساليب الاحصائية التى تفترض أن يكون ميزان القياس فترياً كما سيتضح فى الفصل الثاني .

المستوى النسبى Ration Seale : يعد أدق مستويات القياس ، حيث إن هذا المستوي يحقق الخصائص الثلاث السابق ذكرها ، وهى ، المقدار ، وتساوي المسافات ، والصفر المطلق الذى يمثل انعدام السمة المقاسة ، ولعل المسطرة العادية تعد مثالاً بسيطاً للميزان النسبي ، حيث إن الفروق بين التداريج المدونة عليها (السنتميترات أو البوصات) متساوية فى الطول ، ١٥ سم مثلاً نصف الطول ٣٠سم ، لذلك فإن هذا المستوى يسمح بإجراء العمليات الحسابية الأربع على الأعداد الناتجة عن القياس .

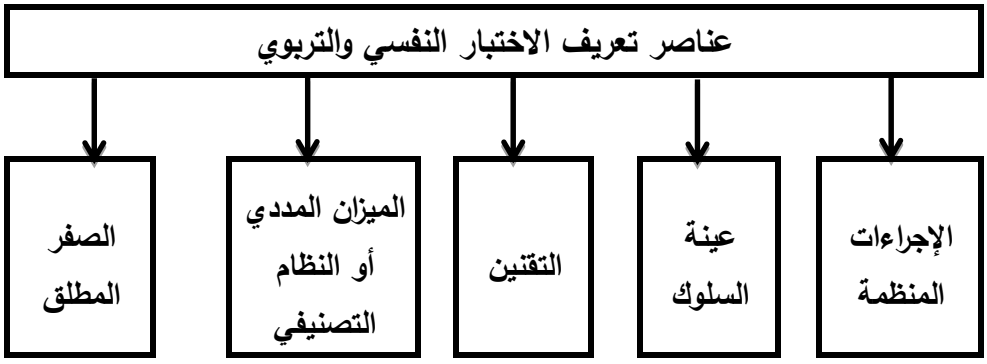
وعلى الرغم من أن معظم الخصائص الفيزيائية للأفراد مستوى قياسها نسبي ، إلا أنه يندر استخدام هذا المستوى فى القياس النفسي والتربوي لعدم توافر صفر مطلق يناظر حقيقة انعدام الظاهرة أو السمة المقاسة ، لذلك فإن مستوى قياس السمات والخصائص النفسية والتربوية يكون عادة فترياً أو رتبياً .

ولعل عرضنا لمستويات القياس الأربعة التي توضح الطرق المختلفة لتطبيق الموازين في تعرف سمات الأفراد والفروق بينهم يجعل أخصائي القياس ، ومستخدم الاختبارات والمقاييس مدركاً أن المستوى النسبي الذى يتميز بتساوي الفترات ، وتوفر الصفر المطلق ليس هو المستوى الوحيد للقياس ، حيث إن كثيراً من السمات النفسية

والتربوية لا تحقق متطلبات هذا المستوى ، كما أن تصنيف هذه المستويات فى مجموعات أربع يسمح بوضع الاختبارات والمقاييس فى ترتيب هرمي من حيث درجة تعقد القياس ودقته ، وبذلك يمكن أن يتخذ قراراً فيما يتعلق بالأدوات المتنوعة ، واختيار الأداة الأكثر ملائمة للغرض من القياس .

مفهوم الاختبار كأداة قياس:

تعد الاختبارات النفسية والتربوية أدوات قياس للسمات الإنسانية شأنها شأن أدوات قياس الظواهر الطبيعية ، مثل الترمومتر، غير أن كثيراً من الناس يعتقدون أن الاختبارات مرادفة للامتحانات المدرسية ، وهذه تعد نظرة ضيقة ومحدودة لمفهوم الاختبار Test الذى يُعد مفهوماً متسعاً، ويتضمن معظم المفاهيم والمبادئ التي سبق توضيحها ، وتشير أدبيات القياس إلى تعريفات متعددة لهذا المفهوم ، إذ ترى انستازي (Anastasi, 1976) أن الاختبار هو أداة قياس موضوعية ومقننة لعينة من السلوك ، ويرى براون (Brown,1970) أن الاختبار هو إجراء منظم لقياس عينة من سلوك الفرد.. بينما يرى شيس (Chase,1978) أن الاختبار هو أسلوب منظم لمقارنة أداء الفرد بمعيار أو مستوى أداء محدد ، أما كرونباخ (Cronbach,1970) فيرى أن الاختبار النفسي أو التربوي هو أسلوب منظم لملاحظة ووصف خاصية أو أكثر من خصائص الفرد استناداً إلى ميزان عددي أو نظام تصنيفي ، وعلى الرغم من الاختلاف الظاهري لهذه التعريفات إلا أنه توجد بينها عناصر مشتركة يوضحها الشكل التخطيطي (١-٣) التالي :



شكل (١-٣) يوضح عناصر تعريف الاختبار النفسي والتربوي

وفيما يلي توضيحاً موجزاً لكل من هذه العناصر :

الإجراءات المنظمة Systematic Proeedures: فالاختبار ينبغي أن يستند فى بنائه وتطبيقه وتصحيحه أو تقدير درجاته إلى قواعد واضحة ومحددة ، فمفردات الاختبار أو المهام التى يشتمل عليها لا يتم انتقاء محتواها من النطاق السلوكي المراد قياسه بأسلوب منظم فحسب ، بل تُطبق أيضاً جميع مفرداته أو مهامه على جميع الأفراد المختبرين فى ظروف متكافئة.

عينة السلوك Behavioral Sample : فالمفردات أو المهام التى يشتمل عليها الاختبار ينبغي أن تعكس السلوك الذى يمكن الاستدلال منه على السمة المراد قياسها ، وهذا يتطلب أن تكون عينة مفرداته ممثلة لجميع مكونات هذه السمة .

التقنين Standardization : فالاختبار المقنن هو ذلك الاختبار الذى تتوحد فيه وتتحدد بدقة طريقة تطبيقه ، ومواده ، وتعليمات وزمن إجابته ، وطريقة تصحيحه أو تسجيل درجاته ، بحيث يصبح الموقف الاختباري موحداً بقدر الإمكان لجميع الأفراد ، وهذا يجعل من الممكن المقارنة بين درجات الأفراد الذين طُبق عليهم الاختبار .

الميزان العددي أو النظام التصنيفي Seale or Classification System: فبعض الاختبارات مثل اختبار الذكاء وغيره من اختبارات الجوانب المعرفية يؤدي إلى قياسات تحقق المتطلبات التى تضمنها تعريف القياس الذى سبق أن أوضحناه ، بحيث يمكن وصف السلوك المراد قياسه كمياً ، غير أن هناك أساليب أو أدوات قياس أخري ، مثل المقابلة Interview يستخدمها الأخصائيون النفسيون فى تصنيف سلوك الفرد وفقاً لنمط استجاباته ، فاختبارات الجوانب المعرفية تستخدم موازين عددية تحدد مكانة الفرد بين أقرانه فى السمة المقاسة ، وذلك بتحديد نسبة عدد الأفراد الذين يتفوق عليهم الفرد فى هذه السمة ، في حين أن أسلوب المقابلة غير المقننة مثلاً يستخدم نظاماً تصنيفياً استناداً إلى مفاهيم نفسية تحليلية تصف استجابات كل فرد أجريت له المقابلة ، كما أن هناك اختبارات تربوية، مثل : بعض الاختبارات التحصيلية ، واختبارات الأداء التى تقيس مهارات إجرائية محددة لا يقارن فيها أداء الفرد بأداء أقرانه ، وإنما بمستوى أداء ، أو

محك كفاءة مناسب لمعرفة مدى تمكن الفرد من هذه المهارات ، وتعرف المهارات التى لم يستطع تحقيقها ، وهذا يعني أن هناك أطراً مرجعية مختلفة لتفسير درجات الاختبار ، والغرض من الاختبار أو المقياس هو الذى يحدد الأساس الذى فى ضوئه يُختار الإطار المرجعي المناسب .

الموضوعية Objectivity : فتطبيق الاختبار ، وتصحيحه ، وتفسير درجاته يكون موضوعياً إذا لم يتأثر ذلك بالأحكام الذاتية للمختبر ، ولعل هذا يرتبط بعملية تقنين الاختبار التى سبق توضيحها ، فالفرد الذى يُطبق عليه الاختبار أو المقياس ينبغي ألا تختلف الدرجة التى يحصل عليها باختلاف الشخص القائم بعملية التصحيح ، وهذه الموضوعية التامة تعد هدفاً يسعي إليه القائمون بإعداد الاختبارات والمقاييس التربوية والنفسية على الرغم من صعوبة تحقيق ذلك فى الممارسات الفعلية ، ومع هذا فإن كثيراً من الاختبارات والمقاييس المتوافرة تحقق ذلك بدرجة جيدة (Anastasi,1970) وخلاصة ذلك أن الاختبار الذى تُستخدم فى بنائه إجراءات موضوعية منظمة، ويحدد الميزان العددي أو النظام التصنيفي الذى يستند غليه تفسير درجاته، يؤدي إلى تقليل المتغيرات الموقفية والشخصية إلى أقل حد ممكن ، ويسمح بمقارنة الفرد بأقرانه أو يمحك أداء محدد .

أنواع الاختبارات والمقاييس النفسية والتربوية:

تختلف أدوات القياس النفسي والتربوي وفقاً لكثير من الجوانب أو الأبعاد مثل: المحتوى ، والغرض ، وطريقة التصميم والبناء ، وطريقة التطبيق ، ونمط الأداء ، وكيفية الاستجابة ، وطريقة التصحيح ، وكيفية تفسير الدرجات ، وفيما يلي توضيحاً لكل من هذه الجوانب .

المحتوى : تختلف الاختبارات والمقاييس فى محتواها أو المهام التى تشتمل عليها : فبعضها يكون محتواه لغوياً ، مثل : المفردات والجمل ، والبعض الآخر يكون محتواه غير لغوي ، مثل: المسائل الحسابية والجبرية ، والأشكال والخرائط والصور ، والرسوم ، وغير ذلك ، وتختلف أيضاً الاختبارات التحصيلية فى محتواها ، فبعضها يكون محتواه

محدوداً حيث يقتصر مثلاً على مصطلحات جغرافية ، أو صيغ رياضية ، أو رموز كيميائية ، والبعض الآخر يكون محتواه متسعاً حيث يشتمل على تعبيرات لغوية ، ومسائل ، ونصوص ، وغير ذلك ، كما توجد اختبارات تشتمل على مواد عيانية ، مثل : المكعبات ، والمتاهات ، والأجهزة وغيرها ، وتتطلب أن يتعامل الفرد معها يدوياً ، وتسمي اختبارات الأداء Performance Tests ، وبعض هذه الاختبارات تناسب الأطفال الصغار ، والأميين .

الغرض : يمكن التمييز بين أدوات القياس من حيث طريقة الفرد فى الأداء ، ففي الاختبارات التى تقيس الجوانب المعرفية لدى الفرد كالتحصيل ، والذكاء ، والاستعدادات تُعطي تعليمات للمختبرين بأن يبذلوا أقصي جهدهم للحصول على أعلي درجة ممكنة ، لذلك يسمي هذا الأداء بالأداء الأقصي Maximum Performance ، أما الاختبارات والمقاييس التى تقيس الجوانب غير المعرفية ، أي الوجدانية ، مثل : الميول والاتجاهات والشخصية ، فإنها تهتم بقياس أداء الفرد الذى يقوم به فى نشاطه اليومي المعتاد ، أي السلوك النمطي الذى يميز هذا الفرد، مثل : هل يحب تنسيق الزهور ؟ أو هل يحبذ الاختلاط فى الجامعة " أو هل هو قلق باستمرار ؟ والأداء الذى تتطلبه هذه الاختبارات والمقاييس يسمى الأداء المميز Typical Performance ، وتشتمل هذه الاختبارات عادة على عبارات أو فقرات تتطلب أن يجيب عليها الفرد بنعم أو لا ، أو موافق أو غير موافق ، ولا يوجد لاي عبارة منها إجابة صحيحة واحدة ، بل تعتمد على الرأي الشخصي للفرد الذى يعكس سلوكه المميز فى مواقف معينة .

طريقة التصميم والبناء: بعض الاختبارات والمقاييس تكون مقننة Standardized Tests والبعض الآخر يكون غير مقنن ، وقد سبق أن أوضحنا مفهوم التقنين ، فالاختبار أو المقياس المقنن هو ذلك الاختبار الذى تكون طريقة تطبيقه ومواده وتعليمات إجابته وطريقة تصحيحه موحدة قدر الامكان لجميع المختبرين ، ويمر بناء الاختبارات المقننة بعامة بمراحل متعددة وتتطلب خبرة عدد من الاختصاصيين فى القياس والتقويم التربوي والنفسي ، وتجريب الاختبار على عينة عشوائية ممثلة لمجتمع الأفراد الذى أعد

من أجله ، والتوصل إلى ما يسمي معايير الاختبار Test Norms التى تفسر درجة الفرد بالنسبة لأقرانه استناداً إليها ، وسوف أوضح هذه المعايير فى الفصل الرابع .

ويلاحظ أن معظم اختبارات الذكاء والاستعدادات تكون مقننة ، وكذلك توجد اختبارات تحصيلية مقننة تتعلق بمجالات دراسية متعددة ، وتتميز هذه الاختبارات بأن محتواها متسع ، بمعني أنها تشمل معظم الأهداف العريضة المتعلقة بهذه المجالات ، وتقيس عادة معلومات أتفق بوجه عام على أنها ضرورية لمتوسطي الأداء في فرقة دراسية معينة ، أما الاختبارات غير المقننة فتشمل بعض أدوات القياس النفسي ، مثل ، المقابلة الشخصية ، والمقابلة الكلينيكية والمقابلة لدراسة تاريخ الحالة ، وغيرها ، كما تشمل الاختبارات التحصيلية الصفية التى يكتبها المعلم لتلاميذه وترتبط ارتباطاً مباشراً بما ينتهي المعلم من تدريسه لهم ، ولا تتطلب إجراء دراسات استطلاعية على عينات ممثلة من طلاب المدارس المختلفة فى الفرقة المستهدفة ، كما لا تتطلب تكوين جداول معايير ، وإنما تتطلب عناية بتحديد الإجرائية للوحدات للوحدات الدراسية التى يهدف المعلم لتحقيقها لدى تلاميذه ، ووضع مواصفات دقيقة للاختبار بحيث تقيس مفرداته الأهداف المحددة تحديداً مباشراً .

طريقة التطبيق: تُطبق بعض الاختبارات والمقاييس على فرد واحد فى وقت واحد ، مثل : اختبارات الذكاء الفردية Individualized Tests كاتبار بينية Binet ، ويقوم بتطبيقها عادة شخص مدرب على تطبيق الاختبار ، وتصحيحه وتفسير نتائجه.

وبعض المقاييس النفسية الفردية ، مثل: اختبار بقع الحبر لرورشاخ Rorshach الذي يعد من الاختبارات الإسقاطية المستخدمة بكثرة فى قياس الشخصية الإنسانية ، واختبار ويكسلر للذكاء يحتاج تطبيقها إلى أخصائي نفسي ، حيث تسمح له بأقصي قدر من التفاعل بينه وبين الفرد المختبر ، وتُقدم فرصاً ثرية للملاحظات الكلينيكية التى تفيد فى تشخيص الاضطرابات النفسية ، والتخلف العقلي أو الدراسي ، وغير ذلك ، فالأخصائي يمكنه ملاحظة المدخل الذى يستخدمه المختبر وأدائه فى المهام الاختبارية ويتدخل الأخصائي أحياناً بتساؤلات تفيد فى متابعة استجابات الفرد وتوضيحها وفهمها

بدرجة أفضل ، غير أن هذا النوع من الاختبارات يتطلب ظروفاً معينة ، مثل توافر حجرة يُجرى فيها الاختبار ، كما يتطلب وقتاً وكلفة لتقييم كل فرد مختبر .

أما النوع الآخر فهو الاختبارات الجماعية Group Tests التى تعتمد عادة على الورقة والقلم ، وتُطبق على مجموعة كبيرة من الأفراد فى وقت واحد ، وهى شائعة الاستخدام فى الصفوف المدرسية العادية لأغراض التوجيه التعليمي أو المهني ، وكذلك فى المؤسسات العسكرية لانتقاء الجنود للمهام المختلفة ، وتتميز هذه الاختبارات بأنها أكثر فعالية من الاختبارات الفردية فى تطبيقها ، وأقل كلفة ، كما أن تطبيقها لا يتطلب شخصاً مدرباً ، غير أن تفسير الدرجات المستمدة من كل من نوعي الاختبارات يتطلب توافر معارف ومهارات إحصائية معينة لدى القائمين بذلك .

نمط الأداء : نقصد بنمط الأداء نوع النشاط الذى يؤديه الفرد فى الاختبار أو المقياس فبعض الاختبارات تتطلب الكتابة ، أي الورقة والقلم ، مثل : الاختبارات التحصيلية ، واختبارات الذكاء والاستعدادات ، وبخاصة الجماعية منها ، ومقاييس الجوانب الوجدانية .

ونقسم اختبارات الورقة والقلم إلى اختبارات تعتمد على التعرف Recognition وأخري تعتمد على الاسترجاع Recall .

ففي اختبارات التعرف يقدم للفرد سؤالاً أو عبارة بالاضافة إلى مجموعة من الإجابات أو الخيارات التى يختار منها الإجابة الصحيحة ، أو أفضل الإجابات ، أو ما ينطبق عليه بدرجة أفضل ، أما فى اختبارات الاسترجاع فيُقدم السؤال أو العبارة دون خيارات معينة ، وعلى الفرد أن يستجيب بأي أجابة يراها مناسبة فى ضوء المحكات المعطاة .

وبعض آخر من الاختبارات يتطلب الأداء العملي ، مثل : بعض اختبارات التربية الفنية والموسيقية ، وبعض اختبارات الذكاء الفردية وبخاصة التى تطبق على الأطفال ، والمقاييس الوجدانية .

كيفية الاستجابة: يعتمد بعض الاختبارات النفسية والتربوية على السرعة فى الاستجابة Speed Tests والبعض الآخر يعتمد على القوة Power Tests ، فأختبار

السرعة يكون له زمن محدد تحديداً دقيقاً ينبغي أن يلتزم به الفرد أثناء إجابته على الاختبار ، وتتميز مفردات هذه النوع من الاختبارات بالسهولة ، بحيث إننا نتوقع عدداً قليلاً من الأفراد يخطئون فى بعض المفردات ، ولكن لا يستطيع الانتهاء من إجابة جميع مفردات الاختبار إلا بعض الأفراد مرتفعي القدرة ، ومثال ذلك : بعض اختبارات الذكاء والاستعدادات الكتابية ، والاختبارات التحصيلية التى تقيس معارف ومهارات أساسية ، وتعد درجة الفرد فى هذا النوع من الاختبارات مؤشراً لسرع إجابته .

أما اختبارات القوة فإنه لا يكون لها زمن محدد للإجابة ، ويُعطي الفرد فرصة كافية للإجابة قدر الإمكان على جميع المفردات ، وتميل مفردات هذا النوع من الاختبارات إلى الصعوبة وتتباين فى مستوى صعوبتها ، ومثال ذلك : بعض اختبارات الذكاء والاستعدادات واختبارات الكفايات ، ويمكن اعتبار الاختبارات التى تعتمد بدرجة أساسية على السرعة أو القوة ممثلة بطرفي متصل لتباين زمن الإجابة ، كما هو موضح بالشكل (١-٤) التالي :

	اختبارات القوة		اختبارات السرعة
الزمن	غير محدد		محدد
صعوبة المفردات	صعبة نسبياً	←	سهلة نسبياً

شكل (١-٤) يوضح متصل اختبارات السرعة والقوة

ويتضح من شكل (١-٤) أنه على الرغم من اعتماد بعض الاختبارات اعتماداً أساسياً على سرعة أو القوة ، إلا أن هناك اختبارات تعتمد على السرعة اعتماداً جزئياً Partially Speeded Tests ، أي تتطلب السرعة والدقة معاً.

طريقة التصحيح: تختلف الاختبارات والمقاييس التربوية والنفسية فى طريقة تصحيحها فبعضها يعتمد على طريقة موضوعية فى التصحيح Objective ، وبعض أخر يعتمد على طريقة ذاتية Subjective ، ونقصد بموضوعية التصحيح عدم تأثر درجات الفرد بالقائم بعملية التصحيح بغض النظر عن نوع المفردة أو محتواها ، فإذا كان محك التصحيح واضحاً ومتفقاً عليه ، فإنه يمكن التقليل من الأحكام الذاتية قدر الأمكان،

ولعل مفردات الاختيار من متعدد ، والصواب أو الخطأ ، والمزاوجة ، تعد من المفردات الموضوعية حيث إنه يجري تصحيحها عادة باستخدام مفتاح تصحيح Scoring Key ، أما أسئلة المقال فإن تصحيحها ربما يختلف باختلاف القائمين بتقدير الدرجات ، مما يجعل التصحيح متأثراً بالذاتية ، غير أنه يمكن العناية باعداد محكات تفصيلية محددة يسترشد بها فى عملية التصحيح ، مما يسهم فى زيادة موضوعية تقدير درجات أسئلة المقال.

كيفية تفسير الدرجات : يمكن تفسير درجات الاختبارات والمقاييس إما عن طريق موازنة أداء كل منهم بأداء اقرانه فى السمة التى يقيسها الاختبار بهدف ترتيبهم بالنسبة لبعضهم البعض الآخر فى هذه السمة ، أو عن طريق موازنة أداء كل منهم بمحك أو مستوى أداء محدد بغض النظر عن المركز النسبي لكل منهم فى السمة موضع القياس .

وعلى الرغم من أنه يمكننا تصميم وبناء اختبار أو مقياس تُفسّر الدرجات المستمدة منه بالطريقتين معاً ، إلا أنه يحسن التمييز بين كل من التفسيرين ، فإذا كان الهدف من بناء الاختبار تقدير أداء الفرد بالنسبة لأداء أقرانه فى قدرة أو سمة معينة ، فإن الاختبار يسمي اختبار مرجعي المعيار Norm – Referenced Test (NRT) ، أما إذا كان الهدف هو تقدير أداء الفرد بالنسبة لمحك أو مستوى أداء معين (أي نسبة معينة من الأهداف أو المستويات التعليمية Standards المحددة لمقرر دراسي معين) فإن الاختبار يسمي اختبار مرجعي المحك Criterion – Refemced Test (CRT) .

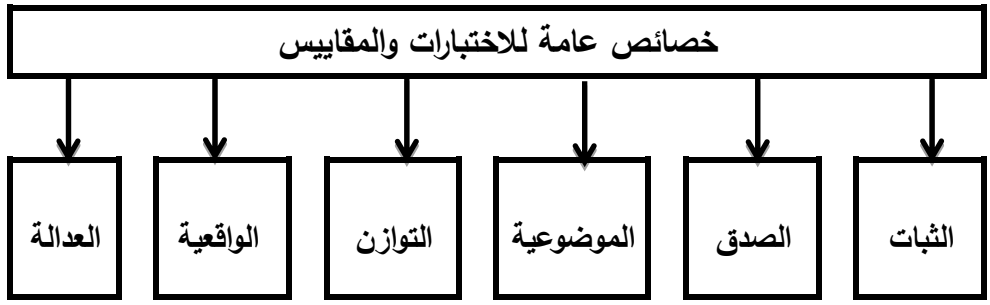
ويمدنا النوع الأول من الاختبارات بتفسير لدرجة الفرد فى ضوء معيار الجماعة Group Norm ، أي فى ضوء متوسط أداء الجماعة ، أو فى ضوء نسبة مئوية من الأقران الذين ينتمي إليهم فرد معين يتفوق عليهم ، وتصبح الدرجة لا معني لها ولا دلالة ويصعب تفسيرها إذا أغفلنا ردها إلى هذا المعيار ، وتعد الاختبارات التحصيلية المقننة ، واختبارات الذكاء والاستعدادات ، وبعض مقاييس الجوانب الوجدانية من هذا النوع .

أما النوع الثاني فيمدنا بمعلومات تفصيلية محددة عن تحصيل الطالب ، وكفأياته ، وما انجزه من الأهداف أو المستويات المتعلقة بمقرر دراسي معين ، لذلك فإن الخطوة

الأساسية فى بناء هذا النوع من الاختبارات هى تحديد الأهداف والمستويات أو التوقعات السلوكية التى يقيسها الاختبار ، وتصبح هذه الأهداف أو التوقعات بمثابة المحك الذي يُنسب إليه أداء الفرد ، وتفسر درجته فى ضوئها لمعرفة المستوى التحصيلي الذى حققه ، والأهداف التى لم يستطع تحقيقها ، لذلك فإن هذه الاختبارات تتميز بأن لها قيمة تشخيصية فى تحديد جوانب القوة والضعف فى تحصيل كل طالب بالنسبة لمجموعة محددة من المعارف والمهارات ، كما أنها تساعد فى تحديد المستوى المدخلي للطالب Entry Behavior الذى يُمكنه من البدء فى دراسة وحدة تعليمية معينة ، أو تعلم مهارة مرجوة ، وكذلك فى تحديد الصف المناسب للطالب المنقول ، أو الذى انقطع عن الدراسة لظروف معينة.

خصائص عامة للاختبارات والمقاييس النفسية والتربوية :

إن الاختبارات والمقاييس التربوية والنفسية هى أدوات قياس عينة مقننة من السلوك بطريقة منظمة كما أوضحنا فيما سبق ، ولكي يفي الاختبار أو المقياس بهذه الشروط ينبغي أن تتوافر فيه مجموعة من الخصائص يوضحها الشكل التخطيطي (١-٥) التالي :



شكل (١-٥) يوضح خصائص عامة للاختبارات والمقاييس

وفيما يلي توضيحاً لكل من هذه الخصائص :

(١) **الثبات Reliability** : ويعني اتساق الدرجات التى يحصل عليها الفرد فى الاختبار أو المقياس ، واتساق القياس يمكن الحكم عليه من حيث : مفردات الاختبار ، وزمن تطبيق الاختبار ، والقائمين بالعملية الاختبارية ، والمختبرين ، وربما تتضح خاصية

الثبات عندما نستخدم المسطرة فى قياس طول خط مستقيم ، فلو وجدنا أن طوله ٦.٥سم ، ثم أعدنا قياسه مرة أخري بالمسطرة نفسها ووجدنا أن الطول أصبح ٦.٧سم ، فإننا بلا شك سوف نفكر فى سبب وجود هذا الفرق فى القياس ، فربما نقول أن المسطرة بها عيب معين ، أو أن اليد قد اهتزت أثناء عملية القياس ، أو غير ذلك من الأسباب فعندئذ نستنتج أن العدد المناظر للطول الذى حصلنا عليه غير ثابت ، أو غير متسق ، وهذا يقلل من الثقة فى القياس ، ويجعل الاعتماد عليه مضللاً ، كذلك الحال فى القياس النفسي والتربوي ، إذ لا بد أن تكون البيانات المستمدة من أدوات القياس ثابتة حتى يمكن الاعتماد عليها فى اتخاذ قرارات صائبة تتعلق بالأفراد ، وهذا يتطلب التحقق من خاصية الثبات قبل استخدام أداة قياس معينة .

(٢) **الصدق Vatidity** : فأدوات القياس تُستخدم فى قياس سمة معينة لدى الفرد وفقاً لتعريف محدد لهذه السمة ، فإذا كانت هذه الأداة تقيس هذه السمة بحق ، فإن البيانات المستمدة منها تفيد فى اتخاذ قرارات مناسبة تتعلق بغرض معين ، وتشتمل أدوات القياس على عينة من الأسئلة أو المفردات أو العبارات المتعلقة بمكونات سمة معينة ، ونحاول من هذه العينة الاستدلال على أداء الفرد فى مواقف معينة خارج حدود الموقف الاختباري ، فنحن لا نقيس الذكاء باختبار ذكاء معين لمجرد القياس ، وإنما نود الاستدلال على سلوك الفرد وأدائه فى مواقف تعليمية أو مهنية خارجة عن نطاق المواقف التى طُبق فيها الاختبار ، أذ ربما نود التنبؤ بنجاحه فى الدراسة فى إحدى الكليات مثلاً استناداً إلى درجة ذكائه التى حصل عليها فى الاختبار . فصدق الاختبار إذن يعني مدى استفادتنا من الدرجات التى يحصل عليها الأفراد فى اختبار أو مقياس نفس أو تربوي معين فى اتخاذ قرارات مناسبة تتعلق بموقف أو مواقف معينة بقدر كبير من الثقة .

(٣) **الموضوعية Objectivity**: فدرجات الاختبارات ينبغي ألا تتأثر قدر الإمكان بعوامل تختلف عن السلوك المحدد المراد قياسه ، مثل : غموض الأسئلة أو الاستجابات ، وأخطاء التصحيح البسيطة ، ومزاج القائم بالتصحيح ، وكذلك ربما

تتأثر الدرجات بوجود انطباعات مسبقة فيما يتعلق بمستوى أداء الأفراد المختبرين ، وأسلوب الكتابة ، وملائمة الوقت لعملية التصحيح ، لذلك ينبغي العناية بوضع خطة تصحيح الاختبار وتقدير درجاته بغض النظر عن صيغة الاختبار أو محتواه لتحقيق درجة عالية من الموضوعية .

(٤) **التوازن Balance**: فلكي يكون الاختبار صادقاً ، فإنه يجب أن يقيس السلوك المتعلق بمجال المحتوى المراد قياسه ، لذلك فإن التوازن يعني مدى كفاية نسبة عدد المفردات التى تقيس سلوكاً أو ناتجاً معيناً ، ويمكن مراعاة ذلك فى الاختبارات التحصيلية استناداً إلى جدول مواصفات الاختبار الذى يحدد الأهمية النسبية لكل من المحتوى والأهداف التعليمية المراد قياسها .

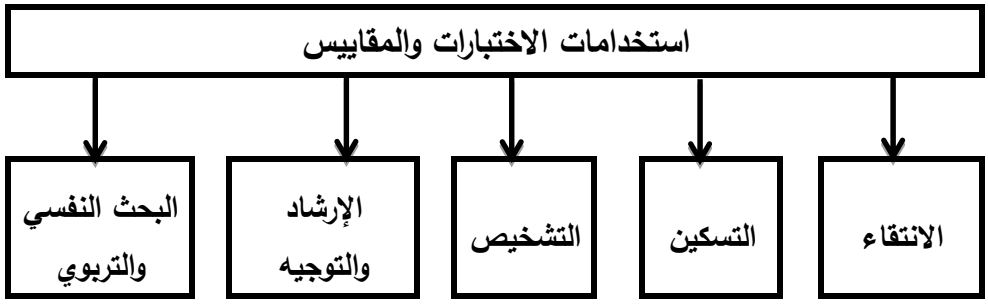
(٥) **الواقعية Practicality** : فالاختبارات والمقاييس يجب أن تكون مفيدة لغرض معين ، وهذا يتطلب أن يكون الاختبار واقعياً ، من حيث : الظروف ، والإمكانات المتاحة لبنائه وتطبيقه ، وتصحيحه ، وتقرير نتائجه ، كما يجب أن يتميز بالفعالية Efficiency من حيث عدد الاستجابات التي يمكن أن نحصل عليها من زمن معين ، فمفردات الاختبار من متعدد ، أو الصواب أو الخطأ مثلاً يمكن الاجابة على عدد كبير منها فى زمن مناسب كما يمكن تصحيحها ألياً ، لذلك فإن الاختبار الذى يشتمل على مثل المفردات يكون أكثر فعالية لو اشتمل على أسئلة تستغرق إجابتها وقتاً طويلاً ، مثل أسئلة المقال ، غير إنه يجب مراعاة أن كلاً من نوعي الأسئلة له استخداماته المناسبة ، ولا يمكن إحلال أي من النوعين محل الآخر .

(٦) **العدالة Fairness** : فالاختبارات وبخاصة التى تقيس التحصيل الدراسي ينبغي أن تؤكد المعارف والمهارات والقدرات التى اهتم المعلم بتنميتها لدى طلابه أثناء دراستهم مقرر معين ، ولتحقيق العدالة ينبغي الابتعاد عن الأسئلة أو المفردات الاختبارية الخادعة أو الغامضة ، كأن تكون صياغتها غير واضحة أو غير مناسبة للمرحلة العمرية للطلاب أو تحتوي على الفاظ غير مألوفة لهم ، أو توحي بإجابات أو تفسيرات غير مطلوبة ، فمثل هذه الأسئلة أو المفردات ربما تعوق الطالب المستذكر

عن الاجابة الصحيحة ، وبذلك تقل درجته التى يحصل عليها عن درجته التى تُعبر عن تحصيله الفعلي ، وهو ما لا يجب أن يكون .

استخدامات الاختبارات والمقاييس النفسية والتربوية :

تُستخدم الاختبارات والمقاييس كما أوضحنا فى إتخاذ قرارات متعددة تتعلق بالأفراد بأقل درجة من المخاطرة ، فلكل اختبار أو مقياس استخدامات تصلح لأغراض معينة ولا تصلح لأغراض أخري ، لذلك فإن انتقاء أداة القياس المناسبة للغرض المرجو يزيد احتمال اتخاذ قرار أفضل ، ولا يعتمد ذلك فقط على جودة الأداة (الاختبار أو المقياس) ، وإنما على اعتبارات أخري مثل : مدى كفاية تحديد مكونات السمة المراد قياسها ، ومدى مواءمتها فى قياس التباين الفعلي بين الأفراد ، وكيفية استخدام البيانات أو الدرجات المستمدة منها ، ويوضح الشكل التخطيطي (١-٦) التالي بعض أهم استخدامات الاختبارات والمقاييس التربوية والنفسية للمعاونة في اتخاذ القرارات المختلفة.



شكل (١-٦) يوضح بعض أهم استخدامات الاختبارات والمقاييس التربوية والنفسية

وفيما يلي توضيحاً لكل هذه الاستخدامات :

الانتقاء Selection: فالاختبارات والمقاييس يمكن أن تُستخدم فى التنبؤ بالنجاح أو الفشل فى الأداء الأكاديمي أو المهني ، ففي مواقف الانتقاء يتقدم عادة عدد من الأفراد يزيد عما هو مطلوب ، لذلك ينبغي أتخاذ قرار بقبول أصلح الأفراد للالتحاق بدراسة أو مهنة معينة ، ويكون الغرض من الاختبار المعاونة فى اتخاذ قرار الانتقاء الذين يكون احتمال نجاحهم كبيراً .

وعندما تُستخدم الاختبارات والمقاييس فى الانتقاء ، ينبغي أن تكون هناك علاقة بين درجاتها والنجاح فى البرنامج التعليمي أو العمل ، فإذا لم تبين الدراسات أن الاختبارات تميز بين الأفراد المحتمل نجاحهم ، والذين لا يحتمل نجاحهم ، فإن هذه الاختبارات تكون فى حاجة إلى تعديل أو استبعاد ، غير أن قرارات الانتقاء لا تكون عادة صادقة تماماً ، لذلك يكون هناك دائماً احتمال قبول أفراد يتبين فيما بعد إخفاقهم ، أو عدم قبول أفراد كان من الممكن أن ينجحوا إذا تم قبولهم ، وينبغي التقليل من هذين الخطأين قدر الأمكان ، ويمكن تحسين قرارات الانتقاء بجعل المهام ، أو المهارات ، أو القدرات التي يشتمل عليها الاختبار مناظرة لمتطلبات النجاح فى الدراسة أو المهنة المعينة .

التسكين Placcment: بعد انتقاء الأفراد لدراسة أو مهنة معينة ينبغي تسكينهم فى أحد البرامج أو الأقسام المناسبة ، والغرض من ذلك إحداث مزاوجة أمثل بين سمات الفرد التى تم قياسها ، ومتطلبات الدراسة أو العمل .

والاختبار الذى يصلح للانتقاء ربما لا يصلح للتسكين ، فاختبار الاستعداد العام الأكاديمي مثلاً ربما يفيد التنبؤ بالنجاح فى الدراسة بالكلية ، ولكن ليس بالضرورة يكون ذا فائدة فى التنبؤ بمن يحتمل نجاحهم فى برنامج دراسي معين.

التشخيص Diagnosis : يهدف لمقارنة جوانب القوة والضعف لدى الفرد فى مجالات متعددة من أجل تحديد أسباب الصعوبات التى يواجهها أثناء أدائه مهاماً معينة ، بحيث يمكن توفير الاساليب المناسبة للعلاج ، فمثلاً الاختبار التشخيصي فى القراءة يمكن أن يزودنا بدرجات فى مجالات مختلفة ، مثل : المفردات اللغوية ، ومعاني الكلمات والجمل ، ومعدل القراءة وذلك بهدف التحديد التدقيق لجوانب الضعف والقوة فى القراءة لدى التلميذ ، ومقياس الشخصية ربما يشتمل على مجالات فرعية تتناول قضايا متعلقة بالعلاقات الأسرية والاجتماعية ، والاستقرار الانفعالي ، وذلك لتحديد أو تشخيص المجال الذى يعاني فيه الفرد من صراعات ويساعد هذا التشخيص فى تقديم برنامج إرشادي أو علاجي .

والاختبارات التحصيلية العادية تعد اختبارات مسحية Survey Tests ، حيث تُستخدم في تقدير المستوى العام لأداء الطالب فى مادة دراسية معينة ، لذلك فإن محتوى الاختبار يكون متسعاً ، مما يجعل من الصعب استخدامه فى إعطاء صورة صادقة عن جوانب القوة والضعف فى تحصيل الطالب ، أما الاختبار التشخيصي Diagnostic Test فإنه يقيس التحصيل بدرجة أكثر عمقاً وتفصيلاً نظراً لتركيزه على قياس مكونات المادة الدراسية أو المهارة المعينة ، لذلك ينقسم الاختبار التشخيصي عادة إلى عدد من الاختبارات الفرعية ، وتقدر درجة لكل طالب في كل اختبار فرعي يجعل الاختبار أطول نسبياً من الاختبار المسحي .

غير إن بناء المقياس التشخيصية يحتاج إلى منهجيات وتقنيات متطورة لا تزال فى حاجة إلى مزيد من البحث والتطوير .

الإرشاد والتوجيه Counseling & Guidance : تُستخدم الاختبارات والمقاييس فى كثير من الأحيان فى معاونة الأفراد فى اختيار البرامج الدراسية ، والمهن ، والوظائف الأكثر ملاءمة لهم ٠

فمعرفة درجات الفرد فى عدد من اختبارات الاستعدادات ، وبطاريات اختيارات التحصيل ومقاييس الميول ، واستبيانات الشخصية ، وسجلات التحصيل الدراسي السابق ، والمعلومات التى يتم جمعها من خلال المقابلات ، تُمكن المرشد التربوي أو المهني من تلخيص جميع هذه البيانات والاستناد إليها فى معاونة الفرد فى عمل تنبؤات تتعلق بمستقبله التعليمي أو المهني وعادة لا يستخدم اختبار أو مقياس واحد فى هذا الشأن ، وإنما تطبق سلسلة من الاختبارات كما أشرنا ، وتعد قرارات الإرشاد والتوجيه قرارات فردية ، على العكس من القرارات المؤسسية التى تتخذها المدارس والجامعات والهيئات الاجتماعية، والمصانع، وغيرها؛ فقرارات الإرشاد والتوجيه تتعلق بفرد معين ، وربما لا تتناسب فرداً آخر .

البحث النفسي والتربوي Research: تُستخدم الاختبارات والمقاييس فى كثير من الاحيان فى التحقق من صحة الفروض ؛ فعلماء النفس والتربية يستخدمون الاختبارات

فى قياس متغيرات تتضمنها فروضهم البحثية من أجل التوصل إلى نظريات تفسر السلوك الإنساني ، وبذلك تعد الاختبارات والمقاييس أدوات علمية، ربما لا يكون لها تطبيقات عملية مباشرة فى حياة الأفراد .

وبالطبع ليست هذه هي الاستخدامات الوحيدة للاختبارات والمقاييس التربوية والنفسية ، إذ أن هناك العديد من الاستخدامات التى تناسب مختلف الأغراض، ولعل استخدام الاختبارات فى المجالات التربوية يعد أكثر الاستخدامات انتشاراً فى الدول المختلفة ، فالاختبارات تُستخدم فى تحسين وأثراء تعلم الطلاب ، وتقويم تحصيلهم ، ومتابعة تقدمهم الدراسي ، وتقديم تغذية راجعة للمعلم والطالب، كما تعد الاختبارات مصدراً لزيادة دافعية الطلاب ، وتطوير البرامج والمناهج التعليمية، وتقويم مختلف مكونات المنظومة التربوية

مفهوم التقويم ، وعلاقته بالاختبارات والمقاييس Exaluation:

يعد مفهوم التقويم أكثر اتساعاً من كل من مفهومي القياس والاختبارات ، فنحن نحاول فى القياس تحري الموضوعية قدر الإمكان ، وتقليل دور الأحكام إلى أقل حد ممكن ، ولكن التقويم يعد عملية تؤكد دور الأحكام استناداً إلى أنواع مختلفة من البيانات الكمية والكيفية التى يتم جمعها وتحليلها ، وذلك للتوصل إلى أدلة تفيد فى تعريف القيمة، أو الجودة ، أو الفاعلية أو اتخاذ قرارات معينة ، أو معاونة فرد آخر فى اتخاذ القرار.

فالمرشد التربوي يعاون الطالب فى جمع بيانات متنوعة مناسبة عن تحصيله وقدراته وميوله ، والتوصل منها إلى معلومات يُتخذ فى ضوئها قراراً معيناً يتعلق بالطالب ففي هذه الحالة تُوجِّه عملية التقويم نحو الاختبار الصائب من بدائل متعددة .

ومن هذا يتضح أن مفهوم التقويم يرتبط ارتباطاً وثيقاً بمفهومي القياس والاختبار، فعلي الرغم من أنه ربما يبدو أن هذه المفاهيم تشير إلى عمليات منفصلة، إلا أنها متداخلة ، فعملية التقويم شأنها شأن أي اختبار ، تكون مفيدة إذا أدت إلى بيانات دقيقة ومتسقة وصادقة بحيث يمكن الاعتماد عليها فى إصدار أحكام صائبة، كما أنه من الضروري أن نتحري في عملية التقويم موضوعية تفسير البيانات، وعدم تأثر الأحكام بالتحيزات الشخصية قدر الإمكان .

وينبغي أن تعتمد عملية التقويم على أنواع مختلفة من أدوات القياس ، والربط بين البيانات المستمدة منها لكي يتم فهم السلوك المراد تقويمه فهماً أفضل ، ومعاونة الأفراد ، وهذا يعني أن عملية التقويم لا تقتصر على استخدام الاختبارات فقط ، وأنما يمكن أن تستخدم قوائم الملاحظة ، وأساليب المقابلة ، واستبيانات الميول والاتجاهات والشخصية ، والطرق الإسقاطية وغيرها.

ومن الجدير بالذكر أن التقويم ليس عملية واحدة ، وإنما تتضمن سلسلة من الأنشطة أو عدد من الخطوات، فاتخاذ القرارات التربوية مثلاً يستند إلى الأحكام، التي تستند بدورها إلى المعلومات، والترابط بين مفاهيم القرارات، والأحكام، والمعلومات يعرِّف مفهوم التقويم، لذلك فإن عملية التقويم تمر بمراحل متعددة نوجزها فيما يلي :

(١) **مرحلة الاعداد:** وتتضمن تحديد نوع الأحكام والقرارات المراد اتخاذها ، ووصف المعلومات المطلوبة ، وكيفية الحصول عليها .

(٢) **مرحلة جمع البيانات:** وتتضمن الحصول بالفعل على المعلومات ، وتحليلها وتدوينها .

(٣) **مرحلة الحكم واتخاذ القرار:** وتتضمن الموازنة بين المعلومات التي تم التوصل إليها، ومحك معين لإصدار الأحكام ، ومن ثم اتخاذ قرار ، أي الاختبار من بين بدائل أفعال فنظام اتخاذ القرارات ينبغي أن يلقي الضوء على أوجه التماثل والاختلاف بين مختلف البدائل ، بحيث يمكن اختبار البديل المناسب بسهولة ، وينبغي بعد ذلك تلخيص النتائج وكتابة ، تقرير التقويم الذي يُوجه لفئة معينة ، مثل : الآباء أو المعلمين ، أو الإدريين ، أو غيرهم من المعنيين بنتائج التقويم .

أخلاقيات استخدام الاختبارات:

على الرغم من أن الاختبارات والمقاييس لها فوائد واستخدامات جيدة متعددة أوضحناها فيما سبق ، إلا أن إساءة استخدام هذه الأدوات يكون له عواقب سلبية كثيرة ، وهذا يتطلب الإجابة على تساؤلات عديدة (Mehrens & Lehman,1969) مثل:

١- ما مستوي تأهيل الشخص الذى يوكل إليه تطبيق الاختبارات وتفسير نتائجها؟

٢- ما نوع الاختبارات التى ينبغي استخدامها ؟
٣- كيف تُستخدم نتائج الاختبارات ؟
٤- من يسمح له بالاطلاع على نتائج الاختبارات ؟
٥- هل الاختبارات تتميز بالعدالة ؟
٦- هل عدد الاختبارات التى تطبق على الطلاب كثيرة أم قليلة ؟

وهذه التسؤلات مترابطة ، حيث إن أنواع الاختبارات التى ينبغي تطبيقها على مستوى تأهيل مستخدميها ، وأغراض تطبيق هذه الاختبارات .

وقد حرصت الجمعية الامريكية لعلم النفس (APA) على إصدار دليل متجدد بعنوان " معايير للاختبارات التربوية والنفسية ، وكذلك المعايير الأخلاقية للمشتغلين بعلم النفس " ، ويشتمل الدليلان على أقسام متعددة تقدم خطوطاً إرشادية لتفسير درجات الاختبارات وخصوصياتها ، وسريتها ، وكذلك معايير تتعلق باستخدام الاختبارات ، ومبادئ التحقق من صدقها واستخدامها فى انتقاء الأفراد ، وغير ذلك من الجوانب الأخلاقية والاجتماعية للاختبارات واستخداماتها ، ويمكن للقارئ ، أو الباحث المهتم بذلك الرجوع إلى هذين الدليلين .

غير أننى سوف اقتصر على إيجاز بعض أوجه النقد والجدل حول الاختبارات والمقاييس النفسية والتربوية ، وإساءة استخدام وتفسير نتائج الاختبارات المقننة ، وبعض الإرشادات للتغلب على ذلك ، فقد أشارت أدبيات القياس والتقويم إلى أوجه النقد التالية :
١- الاختبارات تكون متحيزة ضد بعض الأفراد فى مواقف معينة ، وبخاصة الاقليات والمحرومين .
٢- التنبؤ باستخدام الاختبارات يكون منقوصاً ، ويتطلب الاستعانة بأنواع من المعلومات ، كما أن المحكات المستخدمة فى التحقق من صدق الاختبار ليس مثالية .
٣- الاختبارات تؤدي إلى تصنيف الأفراد تصنيفات قاطعة ، مما يعني عدم الاعتراف بأن الأفراد دائمي التغير ، وأن نتائج الاختبارات تعد بمثابة قياسات تقريبية لسمات الأفراد وخصائصهم .

٤- الاختبارات تفترض قياس سمات أو خصائص موروثة .
٥- نتائج الاختبارات تسهم بدرجة أساسية فى تغيير نظرة الآخرين عن الفرد وفقاً لما يتوقعونه منهم .
٦- الاختبارات تؤثر تأثيراً غير مرغوب فيه فى نمط تفكير الفرد وفقاً لنوع مفرداتها.
٧- الاختبارات تسهم عن طريق إساءة استخدامها فى تشكيل المناهج التعليمية ، وتعوق التطوير التربوي .
٨- الاختبارات تؤثر تأثيراً سلبياً فى مفهوم الذات لدى الفرد ، ومستوى طموحه ، وبخاصة إذا قام بتطبيقها وتفسير نتائجها أخصائي نفسي غير متمرس .
٩- الاختبارات تتدخل فى خصوصيات الأفراد ، وبخاصة الاختبارات والمقاييس المتعلقة بالجوانب الوجدانية والشخصية ، وهذا يؤثر فى عملية انتقاء الأفراد لوظائف معينة ، إذ ربما يؤدي ذلك إلى استبعاد أفراد أكفاء من أعمال يصلحون لها استناداً إلى نتائج هذه الاختبارات التى يري النقاد أنها تستكشف عقول المتقدمين للوظائف .

وقد حاول بعض علماء النفس وخبراء القياس تفنيد هذه المزاعم ، ورأوا أن كثيراً منها لا يرجع الاختبارات والمقاييس فى حد ذاتها ، وإنما إلى إساءة استخدامها بما لا يتفق مع الأهداف المرجوة منها ، فنوعية الاستخدام فى مختلف مجالات الخدمات النفسية والتربوية ، وكفاءة القائمين بذلك ، يزيد من فاعلية أدوات القياس التى يعتني بتصميمها وأعدادها .

لذلك حددت الجمعية الأمريكية لعلم النفس ١٩ مبدءاً أخلاقياً ينبغي الاستناد إليه فى استخدام الاختبارات والمقاييس ، وتتناول هذه المبادئ : الكفايات اللازم توافرها فيمن يستخدم الاختبارات ، وأدوات القياس وإجراءاتها ، وحماية خصوصيات الأفراد ، وسرية الاختبارات والبيانات المستمدة منها ، وتوصيل النتائج للمختبرين ، وحقوق الاقليات والمحرومين ، وسوف أوضح فيما يلي بعض هذه المبادئ بإيجاز (Ailen,1976, : Anastasi,1976 ; Sax,1974)

الخصوصية Confidentiality : فدرجات الاختبارات والتواصل اللفظي بين المرشد التربوى أو الأخصائي النفسي والفرد المختبر تعد معلومات لها خصوصيتها ، ولا يجوز نقل هذه المعلومات إلى الآخرين دون موافقة الفرد ، إلا إذا كان هناك ضرر واضح يمكن أن يعود على الآخرين أو على المجتمع ، وعلى الرغم من أن الدرجات التى يتم الحصول عليها فى البرامج الاختبارية التى تجريها مؤسسات معينة تكون أقل سرية منها فى حالات الإرشاد الفردي ، إلا أنه ينبغي فى الحالتين تحديد الجهة التى تحتاج إلى معرفة نتائج الاختبارات والمقاييس ، وأنسب طرق توصيل المعلومات المتعلقة بها.

السرية Test Security: فالاختبارات تعد أدوات قياس يستخدمها المشتغلون بعلم النفس ، والأخصائيون ، لذلك فإن توزيع هذه الاختبارات ينبغي أن يقتصر على أولئك الذين يتمتعون بكفاءات فنية تمكنهم من استخدام هذه الأدوات استخداماً مناسباً ، فمعظم الاختبارات المقننة ينبغي ألا يسمح بتداولها الا للمتخصصين ، وهذا يتطلب تنظيم عملية نشرها وتوزيعها تجنباً للأضرار التى تنجم عن إساءة استخدامها .

تطبيق الاختبارات Test Administration : عندما تُطبق الاختبارات والمقاييس التربوية والنفسية ينبغي توضيح الغرض من الاختبار لكل فرد مختبر ، وفيما سوف تُستخدم نتائجه ، كما ينبغي إخبار الأفراد بحقوقهم المتعلقة بالخصوصية ، وانه يمكنكم رفض الإجابة على الأسئلة التى يرون أنها غير ملائمة أو تتدخل مباشرة فى خصوصياتهم .

كفايات مستخدمي الاختبارات User Qualifications : يمكن التقليل من إساءة استخدام الاختبارات كبيرة إذا كان القائمون بتطبيق الاختبارات وتفسير نتائجها لديهم الكفايات المناسبة لذلك ، وتختلف هذه الكفأيات باختلاف نوع الاختبار أو المقياس ، فالاستخدام المناسب للاختبارات الفردية فى الذكاء ، ومعظم مقاييس الشخصية ، يحتاج إلى تدريب مكثف لمدة طويلة بإشراف متخصصين ، أما الاختبارات التربوية والمهنية ، فإنها تحتاج إلى تدريب متخصص أقل ، بحيث يمكن اختيار الاختبارات المناسبة للغرض من القياس وطبيعة الأفراد المختبرين ، وكذلك تقييم الخصائص الفنية لهذه الاختبارات ،

مثل : المعايير ، والصدق ، والثبات ، ويكون المختبر مدركاً للظروف التى تطبق فيها الاختبارات والتي ربما تؤثر فى أداء الأفراد ، وتفسير نتائج الاختبارات باستخدام الدرجات الناتجة عنها فى ضوء معلومات أخري تتعلق بالفرد ، مع مراعاة دقة الاستنتاجات فى تفسير الدرجات .

الدرجات المحولة ومعايير الاختبارات

- مقدمة .
- أنواع معايير الاختبارات .
- الدرجات المعيارية Standard Scores
- الدرجات التائية (ت) T scores
- المئينيات Percentiles ، والإعشاريات Deciles
- التساعيات المعيارية Stanines :
- نسبة الذكاء The Intelligent Quotient (IQ)
- نسبة الذكاء الانحرافية Deviation IQ (DIQ)
- معايير الفرق الدراسية Grade Norms
- الصفحات النفسية (البروفيلات) Profiles

❖ مقدمة :

أوضحنا فيما سبق أن الاختبارات التربوية والنفسية هى أدوات قياس مقننة لعينة ممثلة من السلوك بطريقة منتظمة وفقاً لقواعد محددة، وتُستخدم نتائج القياس فى المقارنة بين الأفراد وفقاً لمعاير أو ميزان معرف تعريفاً واضحاً، مما يساعد فى تفسير درجات الاختبارات ، فالدرجات التى يحصل عليها الأفراد فى اختبار معين تعبر عن عدد المفردات أو الأسئلة التى أجاب عليها كل منهم إجابة صحيحة ، وتسمي الدرجات الخام Raw Scores وهذه الدرجات يصعب تفسيرها ما لم يتفق على إطار مرجعى Refrencing Scheme يتم على أساسه تفسير الدرجات ، فالدرجة (١١٥) التى حصل عليها فرد فى اختبار الذكاء لا تعني شيئاً ما لم تُنسب إلى معيار أو محك معين، إذ ربما يتبادر إلى الأذهان أن هذا الفرد أجاب عن (١١٥) سؤالاً إجابة صحيحة ، أو ربما نقول أنه حصل على (١٣٧) نقطة من المجموع الكلي من النقط وهو (٢٠٠) مثلاً، ولكن هذه الدرجة لا تنقل لنا صورة واضحة عن مركز الفرد بالنسبة لأقرانه فى الذكاء، فمثلاً لو أن الدرجة (١١٥) التى حصل عليها الفرد تفوق ٦٥% من درجات أقرانه الذين يتفقون معه فى العمر ، والمستوى التعليمي ، والمستوى الاجتماعي والاقتصادي وغير ذلك ، فإننا نكون قد فسرنا هذه الدرجة فى أطار الجماعة المرجعية Refrence Group التى ينتمي إليها هذا الفرد .

وهذا يدل على أن تفسير درجات الاختبارات والمقاييس المقننة لا يكون تفسيراً مطلقاً كما هو الحال فى القياس الفيزيائي، وإنما يكون بإرجاء الدرجة إلى معيار الجماعة Group Norm، أي أنه تفسير نسبي ، وتوجد مواقف قليلة فى القياس النفسي والتربوي يمكن تفسيرها استناداً إلى محكات أو معايير مطلقة Absolute Standards ؛ فإن قلنا أن فرداً ما استطاع أن يجيب عن ٧٥% من أسئلة اختبار تقيس مهارة حسابية معينة، وتم تحديد النطاق الشامل للأسئلة التى يمكن أن ينتقي منها أسئلة الاختبار بطريقة عشوائية بحيث تمثل هذا النطاق تمثيلاً جيداً ، فإنه يمكن فى هذه الحالة مقارنة درجة الفرد فى الاختبار بمحك أو مستوى أداء مطلق يمثله النطاق الشامل للمهارة المقاسة،

ولكننا لا نستطيع تحديد ما إذا كانت النسبة ٧٥% جيدة أو متوسطة أو ضعيفة بالنسبة لمن هم فى عمر هذا الفرد، وفى مستوى تعليمه وتدربه على هذه المهارة ، فهذه المقارنة تتطلب معرفة أداء مجتمع هؤلاء الأفراد فى المهارة الحسابية ، لكي نحدد مركز الفرد فى هذه المهارة بالنسبة لهذه الجماعة المعيارية .

لذلك فإن المعيار Norms يعد الأساس الذى يُسند إليه فى تفسير درجات الاختبارات والمقاييس المقننة ، حيث يمكن فى ضوئها مقارنة أداء الفرد فى الاختبار بأداء اقرانه ، أي الجماعة المرجعية المعيارية Standard Group ، ونظراً لوجود جماعات مرجعية مختلفة يمكن استخدامها فى الحصول على معايير الاختبارات، فإن تحديد الجماعة المرجعية المناسبة يعد أمراً ضرورياً، فإذا كان الاختبار قد صُمِّم لقياس الاستعداد الدراسي للطلاب الذين يودون الالتحاق بكلية معينة ، فإن الجماعة المرجعية يجب أن تشتمل على طلاب الثانوية العامة الذين سوف يلتحقون بهذه الكلية ، وإذا كان الهدف من الاختبار قياس سمات شخصية معينة لدى المراهقين من البنين مثلاً فإن الجماعة المرجعية يجب أن تشتمل على عينة مستعرضة من المراهقين البنين، وهكذا ، أما إذا كان للاختبار أكثر من غرض ، أو صُمِّم لكي يستخدم مع جماعات متعددة ، فإن هذا سوف يتطلب أكثر من جماعة مرجعية واحدة ، فمدي ملائمة جماعة مرجعية معينة يعتمد على التشابه بين المجتمع الذى تتم معاينته والمجموعة التى يود مستخدم الاختبار مقارنة درجته بها .

❖ أنواع معايير الاختبارات:

يتضح مما سبق أن الدرجة الخام للفرد فى الاختبار لا تفيد شيئاً فى تفسير هذه الدرجة لذلك ينبغي إجراء نوع من التحويل على هذه الدرجة لكي تصبح قابلة للمقارنة ، وتحدد مكانة الفرد أو مركزه بالنسبة لأقرانه فى الجماعة المرجعية ، ومن هنا تأتي أهمية معايير الاختبارات Tests Norms ، فدرجات اختبارات الذكاء، والاستعدادات، والشخصية ، والاختبارات التحصيلية المقننة ، وغيرها تُفسر درجاتها عادة أستناداً إلى جداول المعايير Norms Tables التى تكن مدونة فى أدلة الاختبارات ، ويمكن

باستخدام هذه الجداول معرفة الدرجة المحولة المناظرة لدرجة خام معينة ، وهذه الدرجات المحولة أو المعايير يتم أعدادها إمبريقياً ، وذلك بتحديد أداء عينة من الأفراد ممثلة لمجتمع معين فى الاختبار، أي الجماعة المرجعية ، فإذا اشتملت هذه الجماعة على عينة قومية مستعرضة مستمدة من جميع مناطق بلد معين بحيث تمثل هذه المناطق ، فإن المعايير عندئذ تسمي المعايير القومية National Norms ، وهذه المعايير شائعة الاستخدام فى كثير من الاختبارات المقننة لأنها تتميز بقابلية التعميم .

أما إذا اشتملت الجماعة المرجعية على عينة من منطقة أو محافظة معينة ، فإن المعايير المستمدة منها تسمي عندئذ المعايير المحلية Local Norms ، وتفسير درجات الاختبارات استناداً إلى هذه المعايير يكون محدوداً بهذه العينة المحلية ، ولكنها تكون أكثر فائدة من المعايير القومية أذا أردنا مقارنة أداء الفرد بأداء أفراد آخرين مماثلين له فى الخبرة ، كما أن بعض الاختبارات يكون لها معايير منفصلة للذكور والإناث ، أو الصفوف المدرسية ، أو العمر ، كما أن بعض مقاييس الميول تستند إلى معايير منفصلة للمجموعات المهنية المختلفة ، وبعض اختبارات القبول بالدراسات العليا بالجامعات تستند إلى معايير منفصلة للتخصصات الاكاديمية المختلفة .

فالمعايير أذن هى مجموعة من الدرجات التى تشتق بطرق إحصائية معينة من الدرجات الخام ، بحيث تأخذ فى الاعتبار توزيع الدرجات الخام التى نحصل عليها نتيجة تطبيق الاختبار على عينة عشوائية ممثلة لمجتمع معين من الأفراد ، واحياناً تسمي عينة التقنين Standardization Sample ، وتصمم هذه المعايير لتحقيق هدفين : أحدهما تحديد مكانة أو مركز الفرد بالنسبة لعينة التقنين ، وبذلك يمكن تقييم أدائه بالنسبة لأقرانه ، والآخر تقديم قياسات متقاربة ، أي تسمح بالمقارنة المباشرة بين درجات الفرد فى اختبارات مختلفة ، فمثلاً إذا حصل الفرد على الدرجة (٢٥) فى اختبار فى الاستدلال الحسابي ، والدرجة (١٤) فى اختبار فهم معاني الكلمات ، فإنه يصعب باستخدام هاتين الدرجتين معرفة المركز النسبي للفرد فى كل من الاختبارين ، أي لا نستطيع معرفة ما إذا كان أفضل فى الاستدلال الحسابي أم فى فهم معاني الكلمات ، ويرجع ذلك إلى أن مستوى صعوبة مفردات الاختبار تؤثر فى إمكانية المقارنة بين الدرجتين الخام ، أم

الدرجات المحولة فإنه يعبر عنها بنفس الوحدات ، وتستند إلى نفس عينة تقنين كل من الاختبارين ، مما يمكننا من مقارنة أداء الفرد فى اختبارات مختلفة .

وتوجد طرق متعددة لإجراء تحويل للدرجات الخام تؤدي إلى أنواع مختلفة من المعايير التى تُستخدم فى تفسير درجات الاختبارات التربوية والنفسية أهمها ما يلي:

❖ الدرجات المعيارية Standard Scores:

تعد الدرجات المعيارية من التحويلات الخطية للدرجات الخام التى تعتمد على المتوسط والانحراف المعياري لهذه الدرجات، وتؤدي إلى توزيع الدرجات المحولة مطابق لتوزيع الدرجات الخام ، وهذا يعني أنه إذا مثَّلنا توزيع الدرجات الخام الأصلية على المحور الأفقي، والدرجات المعيارية على المحور الرأسي فى شكل بياني، فإن النقط المناظرة لنوعي الدرجات سوف تقع على خط مستقيم.

والدرجات المعيارية هي عدد الانحرافات المعيارية لدرجات الاختبار عن متوسط هذه الدرجات لمجموعة معينة من الأفراد، فإذا حصل فرد على الدرجة (١٢٠) فى اختبار متوسط درجاته (١٠٠) ، وانحرافه المعياري (١٠) فإن الدرجة المعيارية المناظرة لهذه الدرجة تساوي

$$\frac{١٢٠ - ١٠٠}{١٠} = +٢$$

فالدرجة الخام (١٢٠) تنحرف عن المتوسط (١٠٠) بقدر (٢٠) درجة ، ولكن (٢٠ درجة تعني ٢٠ ÷ ١٠ = +٢ انحراف معياري ، أي أن درجة هذا الفرد تبعد بقدر انحرافين معياريين موجبين عن متوسط الجماعة المرجعية أو عينة التقنين ، أن درجة الفرد تفوق متوسط درجات أقرانه بقدر انحرافين معيارين قيمة كل منهما (١٠) درجات ٠

وعندما نحوِّل جميع درجات اختبار إلى درجات معيارية، فإن التوزيع الناتج يكون متوسطه صفراً ، وانحرافه المعياري مساوياً الواحد الصحيح ، وذلك لأن الدرجة المعيارية صفر تدل على أن الدرجة الخام لا تنحرف عن المتوسط أي تساوي المتوسط ، والدرجة المعيارية (+١) تدل على أن الدرجة الخام تزيد عن المتوسط بقدر انحراف معياري واحد

، والدرجة المعيارية (-١) تدل على أن الدرجة الخام تقل عن المتوسط بقدر وحدة انحراف معياري .

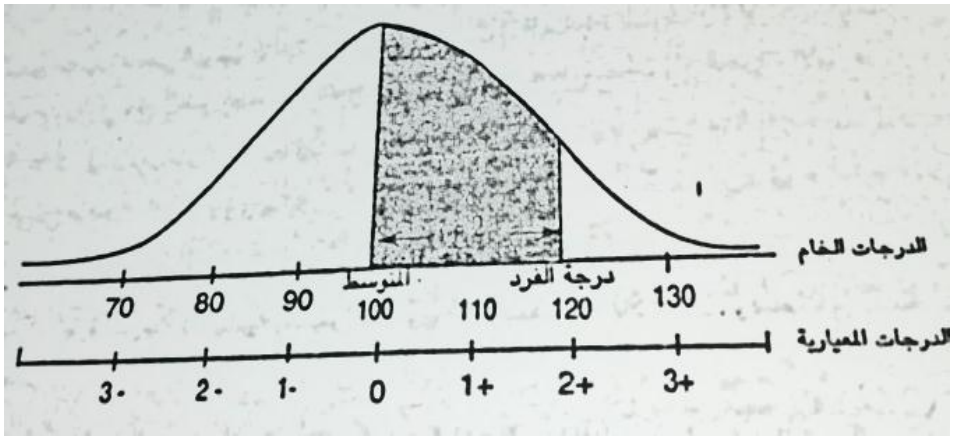
ومدى الدرجات المعيارية لاختبار معين ينحصر عادة بين (+٣) ، (-٣) انحراف معياري .

والصيغة المستخدمة فى حساب الدرجات المعيارية كالتالي :

$$\text{الدرجة المعيارية} = \frac{\text{الدرجة الخام – المتوسط}}{\text{الانحراف المعياري}}$$

$$\text{أي أن : د} = \frac{\text{س – سَ}}{\text{ع}} \quad \cdots\cdots\cdots (٣ - ١)$$

ويمكن تمثيل الدرجة الخام (١٢٠) المذكورة فى المثال السابق على المنحني الاعتدالي كما هو مبين بشكل (٣-١) التالي :



شكل (٣-١) يوضح توزيع الدرجات المعيارية على المنحني الاعتدالي

ومن هذا يتضح أن الدرجات المعيارية تدل على بُعد الدرجة الخام عن متوسط توزيع الدرجات ، وأن الدرجات التى تفوق المتوسط تكون أشارتها موجبة، والتى تقل عن المتوسط تكون إشارتها سالبة ، ونظراً لأن الدرجات المعيارية يُعبر عنها على ميزان

فتري، فإنه يمكن إجراء العمليات الحسابية عليها ، وكذلك فإن نظراً لأن تحويل الدرجات الخام إلى درجات معيارية يُعد تحويلاً خطياً Linear Transformation ، فإن شكل توزيع الدرجات المعيارية سوف يكون مماثلاً لتوزيع الدرجات الخام كما سبق أن أوضحنا ، فإذا كان توزيع الدرجات الخام اعتدالياً ، فإن توزيع الدرجات المعيارية سيكون اعتدالياً كذلك ، وإذا اقترب توزيع الدرجات الخام من التوزيع الاعتدالي ، فإن مدى الدرجات المعيارية سوف يتراوح تقريباً بين (−٣) ، (+٣) .

❖ الدرجات التائية (ت) T scores:

على الرغم من مميزات الدرجات المعيارية، إلا أنه يُعاب عليها أنها تشتمل على كسور وإشارات سالبة ، وللتغلب على ذلك يجري عادة تحويل ميزان الدرجات المعيارية إلى ميزان أخر مناسب ، وهذا يماثل تحويل درجات الحرارة المئوية إلى درجات فهرنهيتية أو العكس، وقد اقترح مكول (McCall, 1922) أن نضرب الدرجة المعيارية فى (١٠) للتخلص من الكسور ، وإضافة (+٥٠) إلى الناتج للتخلص من الإشارات الجبرية السالبة، وبذلك نحصل على درجة محولة جديدة تسمي الدرجة التائية (ت)، وقد اختار مكول (McCall,) الحرف (t) أو(ت) تيمَّناً بكل من رائدي علم النفس ثورنديك (Thorndike) وتيرمان (Terman) حيث إنه الحرف الأول من أسميهما.

أي أن : ت = ١٠ د + ٥٠ ..... (٣−٢)

فالدرجة التائية المناظرة للدرجة المعيارية (−١.٣) + ٥٠ = ٠.٣٧ والدرجة التائية المناظرة للدرجة المعيارية (صفر) تساوي ١٠ (صفر) +٥٠ = ٥٠ . وهكذا فى بقية الدرجات .

وعندما تُستخدم الدرجات المعيارية التى توزيعها اعتدالي ، فإننا نحصل على توزيع الدرجة التائية التى تتخذ شكل التوزيع الاعتدالي أيضاً ، ويكون متوسطها ٥٠ وانحرافها المعياري ١٠ ، ومن الجدير بالذكر أنه يمكن أجراء تحويلات خطية أخري على الدرجات المعيارية للحصول على ميزان جديد للدرجات يناسب أغراضاً معينة ، فمثلاً اختبار القبول بالدراسات العليا بالجامعات الامريكية Graduate Record

Examination (GRE) ، وكذلك اختبار القبول بالجامعات College Entrance Examination Board (CEEB) متوسط درجات كل منهما ٥٠٠ وانحرافه المعياري ١٠٠ .

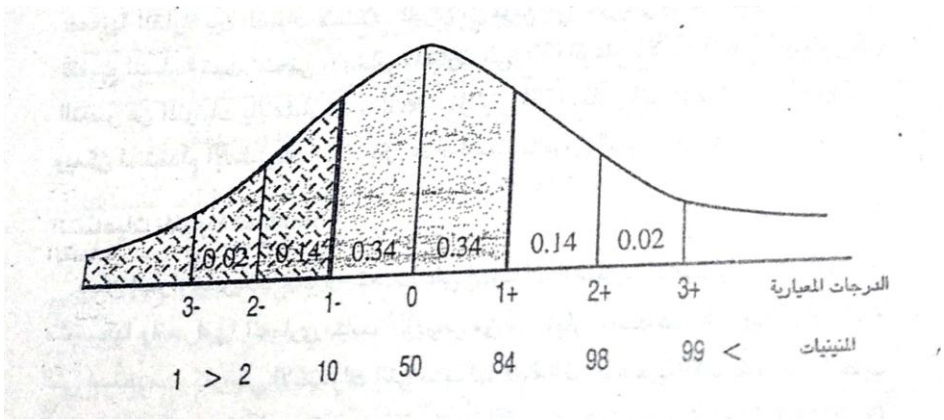
أي أن : الدرجة المحولة = ١٠٠ د + ٥٠٠ .....(٣-٣)

❖ المئينيات Percentiles، والإعشاريات Deciles

تعتمد كثير من الاختبارات المقننة فى تفسير درجاتها على المئينيات ، أو الرتب المئينية Percentile Ranks ، فالمئيني هو نقطة على توزيع الدرجات تقابلها أو تقل عنها نسبة معينة من الأفراد ، فالمئيني ٥٠ مثلاً هو النقطة التى تقسم التوزيع إلى نصفين متساويين ، وهذه النقطة تناظر وسيط الدرجات ، وإذا كانت الدرجة الخام التى حصل عليها فرد فى الاختبار تناظر المئيني (٧٢) ، فإن هذا يعني أنه يفوق ٧٢% من أفراد الجماعة المرجعية المعيارية أو عينة التقنين ، ويقل عن ٢٨% من هؤلاء الأفراد ، لذلك فإن المئينيات تحدد بطريقة مباشرة المركز النسبي للفرد فى مجموعته .

ويمكن إيجاد المئينيات المناظرة للدرجات الخام فى توزيع معين بطريقة بيانية بإيجاد التكرار المتجمع الصاعد المئوي لكل فئة من فئات التوزيع ، ورسم المنحني التكراري المتجمع المئوي كما سبق أن أوضحنا فى الفصل الثاني ، ورسم خط رأسي من الدرجة التى تقع على المحور الأفقي حتى يلاقي المنحني فى نقطة أخري نسقط منها عموداً على المحور الرأسي ونقطة الالتقاء بهذا المحور تناظر المئيني المطلوب .

ويمكن توضيح العلاقة بين الدرجات المعيارية والمئينيات على المنحني الاعتدالي المعياري بالشكل (٣-٣) التالي :



شكل (٣-٣) يوضح العلاقة بين الدرجات المعيارية والمئينيات

ويتضح من شكل (٣-٣) أن الدرجة المعيارية (+١) تفوق المتوسط بقدر انحراف معياري واحد ، وتناظر المئيني ٨٤ ، والدرجة المعيارية (-١) تقل عن المتوسط بقدر انحراف معياري واحد ، وتناظر المئيني ١٤ ، وقد حصلنا على المئيني ٨٤ بإضافة ٥٠% من الحالات التى تقل عن المتوسط إلى ٣٤% من الحالات التى تنحصر بين الدرجتين المعياريتين صفر ، +١ .

وعلى الرغم من أن الدرجات المعيارية فى التوزيع الاعتدالي يكون مستوى قياسها فترياً ، وذلك لأن الفرق بين كل درجتين معياريتين متساوي ، إلا أن هذا ينطبق على المئينيات ، فالفرق بين المئيني ٥٠ والمئيني ٨٤ يناظر فرقاً قدره انحراف معياري واحد (د = صفر ، د = +١) كما هو مبين بالشكل ، ويحد نسبة من المساحة تحت المنحنى الاعتدالي قدرها ٣٤% .

بينما نجد أن الفرق بين المئيني ٨٤ والمئيني ٩٨ يناظر أيضاً فرقا قدره انحراف معياري واحد (د= +١ ، د = +٢) ولكن يحد نسبة من المساحة تحت المنحني الاعتدالي قدرها ١٤% فقط .

فالمئينيات مستوى قياسها يكون رتبياً ، ولذلك لا نستطيع إجراء العمليات الحسابية الأربع عليها ، وهذا يعد أمراً أساسياً عند تفسير درجات الاختبارات ، حيث إن الفرق بين

مئيني وآخر بالقرب من وسط التوزيع يكون أقل بكثير من الفرق نفسه بالقرب من طرفي التوزيع ، فالطالب الذى تحسن أداؤه بحيث تغير المئيني المناظر لدرجته فى الاختبار من المئيني (١) إلى المئيني (٥) مثلاً يكافي تحسن طالب آخر من المئيني (٥٠) إلى المئيني (٧٠) ، ولعل هذا يؤدي إلى صعوبة المقارنة بين المئينيات لاختلاف الميزان من موقع إلى آخر على متصل الدرجات ، ويمكن تقسيم المساحة تحت المنحني الاعتدالي المعياري إلى (١٠) أقسام بدلاً من (١٠٠) ، وبذلك يمكن التعبير عن المئينيات بالإعشاريات Deciles ، فالمئيني (٧٠) مثلاً يناظر الإعشاري (٧) ، وهكذا .. ويمكن استخدام الإعشاريات إذا لم يكن هناك اهتمام بالفروق الدقيقة بين الأفراد .

❖ التساعيات المعيارية Stanines:

يمكن إجراء تحويلات خطية متعددة على الدرجات المعيارية الاعتدالية بحيث يصبح متوسطها وانحرافها المعياري مناسباً للغرض من الاختبار ، وأحد هذه التحويلات الشائعة التى استُخدمت بكثرة فى الاختبارات التى صممتها هيئة القوات الجوية الامريكية أثناء الحرب العالمية الثانية ، وفى أغراض تقدير درجات الطلاب فى مؤسسات تربوية معينة ، أطلق عليها " التساعيات المعيارية Stanines " وهو أسم مكون من ضم كلمتي Nine, Standard ، وميزان التساعيات Stanine Scale هو درجة معيارية اعتدالية متوسطها (٥) وانحرافها المعياري (٢) تقريباً ، استناداً إلى أن الدرجات تتراوح من (١) إلى (٩) ، لذلك فإن التساعيات المعيارية يعبر عنها بأعداد ذات رقم واحد ، ويمكن تحويل الدرجات الخام إلى تساعيات معيارية ، وذلك بترتيب هذه الدرجات ترتيباً تصاعدياً أو تنازلياً ، وتعيين التساعيات وفقاً للنسب المئوية من المساحات تحت المنحني الاعتدالي ، وتوجد تسع نقاط درجات أو تساعيات على ميزان القياس ولكن ميزان التساعيات المعيارية لا يعد ميزان درجات معيارية حقيقي ، وذلك لأن كلاً من التساعي الأول والتاسع مفتوح النهاية .

ويتضح أن الدرجات التى تقع فى حدود كل منطقتين يُعين لها أحد التساعيات ، وكل منطقة اتساعها نصف انحراف معياري ، فيما عدا كل من التساعي الأول والتاسع ، وتحدد النسب المئوية من الحالات فى المنحي الاعتدالي داخل كل من منطقة ، وكذلك مدى المئينيات المناظرة لكل منها : فجميع الأفراد الذين تقع درجاتهم فى مدى معين يُعين لهم التساعي المعياري المناظر لهذا المدى، فمثلاً الأفراد الذين تقع درجاتهم الخام فى المدى المئيني (٦٠-٧٦) يُعين لهم التساعي المعياري (٦) وهكذا ، حيث يُعين لنسبة ١٧% من أفراد عينة التقنين التساعي المعياري (٦) .

وتوصي بعض المؤسسات المعينة بنشر الاختبارات مرجعية المعيار – Norm Referenced Tests مثل : الاختبارات التحصيلية المقننة ، واختبارات الاستعدادات فى كثير من الأحيان باستخدام المعيارية فى تفسير درجات هذه الاختبارات ، وربما يرجع ذلك إلى أن التساعيات المعيارية تتعامل برقم واحد ، كما أنها تتعامل مع مدى المئينيات بدلاً من نقط المئينيات ، وذلك لأن درجات الاختبارات يشوبها عادة قدراً من الخطأ ، كما أنه نظراً لأن التساعيات المعيارية تعد درجات معيارية اعتدالية ، فإنه يمكن إجراء معالجات رياضية عليها، فهي على العكس من المئينيات يمكن إيجاد متوسطها دون الحاجة إلى تحويلها إلى موازين أخري .

غير أن بعض خبراء القياس يري أن تفسير درجات الاختبارات باستخدام التساعيات المعيارية يعد أكثر صعوبة من تفسيرها باستخدام المئينيات ، وبخاصة إذا كانت الدرجات تتميز بالثبات ، حيث أن التساعيات المعيارية لا تتعامل مع كل درجة على حدة ، وإنما مع مجموعة من الدرجات فى فئة معينة كما يتضح من شكل (٣-٤) ، لذلك يري هؤلاء الخبراء أنه يُفضل استخدام عبارات لفظية لوصف كل تساعي معياري ، فمثلاً التساعي المعياري الأول يعني مستوى أدني (٤%)، والتساعي المعياري الثاني يعني مستوى ضعيف (٧%)، والتساعي المعياري الثالث يعني مستوى أقل من المتوسط (١٢%) ، وهكذا ، غير أن هذه العبارات اللفظية تنطبق فقط إذا كانت الدرجات تتوزع اعتداليا ، وذلك لأنه إذا كان الاختبار غاية فى السهولة أو الصعوبة ، فإن توزيع الدرجات فى

هاتين الحالتين ربما يكون ملتوياً التواءً ملحوظاً ، مما يتطلب الحيطة فى تفسير التساعيات المعيارية ، وبخاصة فى طرفي التوزيع .

❖ نسبة الذكاء (IQ) The Intelligent Quotient :

تعد نسبة الذكاء من معايير الاختبارات شائعة الاستخدام ، كما أنها أكثر المعايير غموضاً فى فهمها وتفسيرها ، وقد اقترح المفهوم الأصلي لنسبة الذكاء عالم النفس الألماني " ويلهلم شترن Wilhelm Stern " فى مستهل القرن الماضي ، ويرجع الفضل إلى تيرمان Terman عالم النفس الأمريكي فى انتشار هذا المفهوم فى الطبعة الأولي لاختبار ستانفورد بينية للذكاء عام ١٩١٦، ونسبة الذكاء (IQ) من هذا المنظور يُعبر عنها بالنسبة بين العمر العقلي للفرد مقسوماً على عمره الزمني ، وضرب الناتج فى (١٠٠) للتخلص من الكسور ، أي أن

$$\text{نسبة الذكاء (IQ)} = \frac{\text{العمر العقلي}}{\text{العمر الزمني}} \times ١٠٠ \quad .....(٣-٣)$$

وتعد نسبة الذكاء بمثابة مؤشر لمعدل النمو العقلي العام ، غير أن العمر العقلي Mental Age لا يعطي صورة مكتملة عن النمو العقلي ، حيث أن العمر العقلي (١٢) مثلاً يكون تضمينات مختلفة إذا كان الطفل عمره الزمني (٨) أو (١٤) عاماً ، لذلك اقترح ضرورة قياس كل من معدل النمو العقلي ، ومستوى النمو (أي العمر العقلي) ، فالطفل الذى يكون ذكاؤه مناظراً لمتوسط ذكاء أطفال هذا العمر تكون نسبة ذكاؤه (IQ)تساوي ١٠٠ ، وذلك لأن عمره العقلي يكون مساوياً لعمره الزمني ، أما الأطفال الذين يكون معدل نموهم العقلي أسرع من المتوسط سوف تكون نسبة الذكاء لديهم أكبر من ١٠٠ ، بينما تكون أقل من ١٠٠ لمن كان معدل نموهم العقلي أبطأ من المتوسط .

والعمر العقلي الذى يُستند إليه فى حساب نسبة الذكاء هو درجة خام محولة.

غير أن نسبة الذكاء يشوبها كثيراً من أوجه القصور من بينها ما يلي :

١- الانحرافات المعيارية لنسب الذكاء ليست ثابتة للأعمار المختلفة .

٢- القيمة القصوى للمقام فى نسبة الذكاء (أي العمر الزمني) لا تزال مثار خلاف بين علماء النفس وخبراء القياس ، وهذا يتعلق بتحديد العمر الذى يتوقف عنده النمو العقلي للفرد ، لذلك يرى كثير من هؤلاء أن نسبة الذكاء قليلة الفائدة فى قياس ذكاء الراشدين .

٣- تقلص وحدات العمر العقلي بتقدم العمر ، وعلاقة ذلك بالانحراف المعياري ، فمثلاً إذا نظرنا لنسب الذكاء الآتية :

(¾ × ١٠٠ = ١٣٣) ، (٨/٦ × ١٠٠ = ١٣٣)، (١٢/٩ × ١٠٠ = ١٣٣)

نلاحظ أن الطفل الذى عمره العقلي (٤) يزيد عن عمره الزمني (٣) بقدر عام واحد ، ونسبة ذكاؤه ١٣٣ ، بينما الطفل الذى عمره العقلي (٨) يزيد عن عمره الزمني (٦) بقدر عامين ، ومع هذا تظل نسبة الذكاء (١٣٣) ، والطفل الذى عمره يفوق عمره العقلي بثلاثة أعوام تكون نسبة ذكاؤه ١٣٣ أيضاً ، فإذا كان نسبة الذكاء (١٣٣) تعني الشئ نفسه فى جميع الأعمار ، فإن مقدار الانحراف المعياري ينبغي أن يزداد فى كل عام تال ، غير أن هذا ليس متحققاً فى الواقع ، فالنمو العقلي لا يزداد خطياً بزيادة العمر ، لذلك فإن نسبة الذكاء لم تعد تُستخدم فى كثير من اختبارات الذكاء الرئيسية ، وإنما استعيض عنها بمعيار يعتمد على إيجاد الدرجات المعيارية المستمدة من عينة ممثلة من أفراد مجتمع لكل مستوى عمري معين ، وتسمي هذه الدرجات " نسبة الذكاء الإنحرافية " .

❖ نسبة الذكاء الانحرافية Deviation IQ (DIQ) :

تعد نسبة الذكاء الانحرافية نوعاً من الدرجات المعيارية المحولة مثل الدرجات التائية وليست نسبة بالمفهوم الرياضي ، وعادة يكون متوسطها ١٠٠ وانحرافها المعياري (١٥) أو (١٦) ، على الرغم من أن بعض اختبارات الذكاء يكون الانحراف المعياري لدرجتها (١٢) أو (٢٠) ويتم حساب نسبة الذكاء الانحرافية لكل مجموعة عمرية فى عينة التقنين.

فنسبة الذكاء الانحرافية (١٣٠) تضع الفرد دائماً فى موقع يزيد انحرافين معياريين عن المتوسط ، وهذا يناظر المئيني (٩٨) ، وقد استُخدمت نسبة الذكاء (IQ) فى اختبار

بينية Binet إلى أن أجريت مراجعة للاختبار عام ١٩٦٠، واستُخدمت فيه عندئذ نسبة الذكاء الانحرافية (DIQ) ، كذلك استخدمها ويكسلر Wechsler فى اختباراته المعروفة التي تقيس ذكاء الراشدين ، والأطفال ، وأطفال ما قبل الحضانة ، والتى سوف أوضحها فى الفصل السادس .

ومن الجدير بالذكر أن نسبة الذكاء ، ونسبة الذكاء الانحرافية ، يعدان من المعايير الارتقائية أو النمائية Developmental Scales .

❖ معايير الفرق الدراسية Grade Norms :

تعد معايير الفرق الدراسية أكثر المعايير استخداماً ، وذلك لسهولة فهمها بواسطة مستخدمي الاختبارات ، فمعيار الفرقة الدراسية Grade Norm أو معيار مكافئ الفرقة Grade Equivalent هو الدرجة المتوسطة التى يحصل عليها التلاميذ فى فرقة دراسية معينة فى اختبار معين ، فإذا استطاع تلاميذ الفرقة الثالثة الابتدائية فى الجماعة المرجعية أو عينة التقنين الإجابة الصحيحة على عدد متوسط من الأسئلة فى اختبار حساب ، وليكن (١٨) ، فإن الدرجة الخام (١٨) تناظر الفرقة الرابعة ، وتُستخدم معايير الفرق الدراسية عادة فى الاختبارات التحصيلية المقننة للمرحلة الابتدائية ، ودرجة الفرقة المكافئة تتحدد بكسر عشري مثل (٢.٩) أو (٤.٤) حيث يشير الرقم الصحيح إلى الفرقة الدراسية ، والجزء الكسري العشري إلى أحد شهور العام الدراسي لهذه الفرقة .

لذلك فإن الدرجة المكافئة للفرقة ، وهى (٢.٩) تشير إلى الفرقة الثانية والشهر التاسع ، وهذا يعني أنه إذا كانت الدرجة (٢٤) فى اختبار تحصيلي معين تناظر الفرقة المكافئة (٦.٢) فإن هذا يمكن تفسيره بأن المتوسط لجميع التلاميذ الذين طبق عليهم هذا الاختبار خلال الشهر الثاني فى الصف السادس كان مساوياً (٢٤).

ومكافئ الفرق الفعلي يمكن الحصول عليه للنقط الزمنية التى يطبق فيها الاختبار ، أما مكافئ الفرق لنقط زمنية أخري ، فإنه يتم الحصول عليه بواسطة عملية استكمال Exterpolation ، أو استيفاء Interpolation ، وذلك بتحديد وسيط مكافئ الفرق وقت

الاختبار لكل درجة من درجات الاختبار ، وتوصيل هذه النقط بخط يستخدم فى الاستكمال أو الاستيفاء ، وبذلك يمكن تقدير القيم التى تقع بين مكافئات الفرق .

وعلى الرغم من شيوع استخدام هذا النوع من المعايير ، إلا أنه يشوبه أوجه قصور متعددة من أهمها ما يلي :

١- نمو الخصائص النفسية والتربوية لدى التلاميذ لا يكون منتظماً على المدى الكلي للفرق الدراسية ، ولهذا فإن فرق شهرين من النمو فى التحصيل فى الفرقة الخامسة أي من (٥.٣) مثلاً ، لا يدل على نفس درجة التغيير فى التحصيل في الفرقة السابعة ، أي من (٧.٣) إلى (٧.٥) مثلاً ، والحقيقة أن وحدة الفرقة تصبح أصغر تدريجياً كلما اتجهنا نحو الفرق الدراسية الأعلي ، أي أن وحدة القياس فى معايير الفرق الدراسية غير متساوي كما هو الحال فى معيار نسبة الذكاء التى سبق توضيحها .

٢- معايير الفرق الدراسية لا تقدم معلومات كافية فيما يتعلق بالمئيني المناظر لدرجة الفرد بالنسبة لصفة الدراسي.

٣- اختلاف تحصيل التلاميذ الذين يحصلون على مكافئات فرق متساوية ، فتلميذ الفرقة الرابعة الذى يكون مكافئ فرقته (٧.٠) لا يمتلك نفس معارف أو مهارات تلميذ الفرقة التاسعة الذى يكون مكافئ فرقته (٧.٥) أيضاً .

٤- معايير مكافئ الفرق الدراسية غير متقارنة عبر مجموعات فرعية من الاختبارات التي تشمل عليها بطارية اختبارات ، وذلك للتلاميذ الذين يفوقون متوسط فرقتهم أو يقلون عنه .

❖ الصفحات النفسية (البروفيلات) Profiles :

عندما نود عرض وتفسير درجتين أو أكثر للفرد نفسه أو لمجموع من الأفراد فى بطارية من الاختبارات ، فإن هذا يتطلب استخدام ما يسمي بالصفحة النفسية أو البروفيل Profile ، فالمقارنة بين هذه الدرجات لا يكون لها معنى إلا إذا أجرينا تحويلاً للدرجات الخام إلى نوع واحد من الدرجات المحولَّة استناداً إلى المجموعة المرجعية ، أو عينة

التقنين ذاتها ، وعندما يشتمل الاختبار على عدة اختبارات فرعية مثل : اختبارات الاستعدادات متعدد العوامل ، وبطاريات الاختبارات التحصيلية المقننة ، وكثير من مقاييس الشخصية والميول ، فإنه ينبغي تفسير درجاتها استناداً إلى معايير مستمدة من نفس العينة ، وبذلك يمكن تكوين صفحات نفسية ذات معنى للأفراد المختبرين وكذلك يمكن تحديد مقدار الخطأ فى القياس .

وينبغي أن تشتمل الصفحات النفسية على معلومات كافية عن الاختبار ، مثل: عنوانه وصيغته ، ومستواه ، وأسم الفرد المختبر ، وتاريخ إجراء الاختبار ، والدرجات الخام التى حصل عليها الفرد فى كل اختبار فرعي والتي اشتقت منها الدرجات المحولة (المعايير) ، وعادة تُقدم تقارير وصفية إلى جانب هذه المعايير أو الدرجات المحولة ، والصفحات النفسية ، وتكون هذه التقارير مطبوعة باستخدام الحاسوب كما هو الحال فى اختبارات أيوا للمهارات الاساسية Iowa Tests of Basic Skills، وهي اختبارات تقيس المهارات الأساسية اللغوية والرياضية لطلاب المدارس فى الولايات المتحدة الأمريكية ، وهذه التقارير تقدم معلومات بطريقة بسيطة يسهل فهمها بواسطة الطلاب وأولياء أمورهم.

والخلاصة أن هناك معايير متعددة ومتنوعة يُستند إليه فى تفسير درجات الاختبارات والمقاييس التربوية والنفسية ولكل منها مزاياه وعيوبه وحدود استخدامه، وكل منهما يناسب غرضاً معيناً ، ومعظم مؤسسات نشر الاختبارات المقننة يستخدم جميع هذه المعايير ، وبوجه عام ، فإن معايير المئينيات والتساعيات المعيارية أكثر ملاءمة للمقارنة بين درجات اختبارات فرعية Subtests ، أما الدرجات المعيارية فإنها تستخدم بدرجة أكبر فى المقارنة بين درجات مستويات وصيغ مختلفة لنفس الاختبار الفرعي ، ويُفضل استخدام معايير الفرقة الدراسية لتقرير متوسطات مجموعات من التلاميذ إضافة إلى المعايير المئينية ، وعموماً فإن تفسير الدرجات الخام للاختبارات المقننة بطرق مختلفة يثرى هذا التفسير ويزيد من المعلومات المستمدة من الاختبار .

الخصائص الأساسية
للاختبارات والمقاييس

- مقدمة .

(أولاً) مفهوم الثبات وأهميته.

- مصادر أخطاء القياس التى يؤثر فى الثبات .
- أخطاء تتعلق بأداة القياس .
- أخطاء تتعلق بإجراء الاختبار، وظروف تطبيقه.
- أخطاء تتعلق بالأفراد المختبرين .
- طرق تقدير ثبات درجات الاختبارات
- بعض العوامل التى تؤثر فى تفسير قيم معامل الثبات

(ثانياً) مفهوم الصدق وأهميته

- أنواع صدق الاختبار .
- صدق المحتوي Content Validity :
- الصدق المرتبط بمحك Criterion – related Validity :
- صدق التكوين الفرضي Construct Validity :
- تعقيب عام على أنواع صدق الاختبارات والمقاييس .

(ثالثاً) تحليل مفردات الاختبار وأهميته .

- معامل الصعوبة المفردة .
- معامل تمييز المفردة .
- تفسير تمييز المفردة .

➢ مقدمة :

أوضحنا فى الفصل الأول أن أي أداة قياس ينبغي أن تتوافر فيها بعض الخصائص الفنية الأساسية ، لعل من أهمها الثبات والصدق ، وهما مفهومان مترابطان ، فالاختبارات والمقاييس التربوية والنفسية ينبغي أن تُستخدم استخداماً ذكياً ، حيث إن بعض الاختبارات ربما لا تكون قد أعدت إعداداً جيداً ، وهنا ينبغي الحيطة أو الحذر من استخدام مثل هذه الاختبارات ، وهذا الأمر يقتصر على المقاييس التربوية والنفسية ، وإنما يمتد ليشمل أيضاً المقاييس فى العلوم الفيزيائية ، فما بالك بميزان أو أميتر أو مايكرومتر تعطي قراءات مختلفة إذا تكرر استخدامها فى قياس الشيء نفسه فى أوقات مختلفة ، أليس هذا ممكنا ؟ بل ويحدث كثيراً فى حياتنا اليومية .

فإذا أعدت قياس وزنك مثلاً فى فترات متقاربة جداً بميزان معين ، وأعطاك قياساً مختلفاً ، فما معنى ذلك ؟

ربما تستنتج أن الميزان به خلل معين ، أو أن السطح الذى يرتكز عليه الميزان ليس ثابتاً أو مستوياً ، أو غير ذلك، فهذه العوامل تجعل القياس مشوباً بالخطأ ، وكلما زادت هذه العوامل ازداد تأثيرها ، وقلت دقة القياس ، والعكس صحيح.

فالثبات أذن يعد خاصية من خصائص القياس الناتج ، وليس خاصية من خصائص أداة القياس ذاتها .

فأدوات القياس ليست ثابتة أو غير ثابتة ، وإنما القياس عنها هو الذي يتميز بدرجة معينة من الثبات .

غير أن المشكلة تكون إلحاحاً وتعقيداً فى العلوم السلوكية ، فنتائج الاختبار ينبغي أن تكون ثابتة أو منسقة ولا تتغير جوهرياً من تطبيق إلى آخر على الفرد نفسه ما دامت الظروف لم تتغير تغيراً ملحوظاً إذا لنا أن نستخدم نتائج الاختبار فى اتخاذ قرارات صائبة تتعلق بالفرد أو الجماعة ، فمفهوم الثبات Reliability إذن يعد من المفاهيم الأساسية فى القياس النفسي والتربوي .

كذلك ينبغي أن تتميز نتائج الاختبار بالصدق Validity ، أي ترتبط نتائجه بشئ خارج نطاق الاختبار ذاته ، فالاختبار الذي ترتبط درجاته بالسلوك الذكي يعد مقياساً للذكاء وليس مقياساً لزمن الرجع مثلاً ، والحقيقة أن مفهوم الصدق يعد من المفاهيم التى اهتم علماء القياس بدراستها دراسة مستفيضة نظراً لأهميته فى اتخاذ القرارات المتعلقة بالأفراد والجماعات .

لذلك سوف أتناول فى هذا الفصل كلاً من مفهومي الثبات والصدق، وأنواعهما ، والطرق المختلفة لتقدير كل من هذه الأنواع والعوامل التى تؤثر فيها، وكذلك اتناول كيفية تحسين نوعية الاختبارات عن طريق تحليل مفرداتها لتحديد درجة صعوبة كل مفردة، وقدرتها على التمييز بين الأفراد فى السمة المراد قياسها.

أولاً : مفهوم الثبات وأهميته:

يشير مفهوم الثبات إلى اتساق درجات الاختبارات والمقاييس لمجموعة معينة من الأفراد أي الاتساق عبر الزمن ، أو اتساق صيغ مختلفة من نفس الاختبار ، أو اتساق مفردات الاختبار ذاته ، أو الاتساق عبر أفراد مختبرين ومصححين مختلفين ، فدرجات الاختبارات ربما لا تكون متسقة لأسباب متعددة ، إذا ربما يكون سلوك الأفراد غير مستقر ، أو أن يكون هناك تباين فى التنبؤ بهذا السلوك من وقت إلى آخر ، أو اختلاف فى عينة مفردات الاختبار ، أو محتوى أساليب الملاحظة ، أو عدم اتساق الأفراد أو إجراءات تصحيح الاختبار أو تقدير درجاته ، وغير ذلك ، لذلك ينبغي عند تفسير درجات الاختبار معرفة المصادر أو العوامل التى تسبب عدم اتساق هذه الدرجات ، فالقياس الذى يتميز بالاتساق يعد شرطاً ضرورياً للاختبار أو المقياس الجيد (Nitko,1983) فإذا كانت درجات الاختبار غير متسقة وتتذبذب تذبذباً كبيراً من عينة من المفردات إلى عينة أخري ، فإنه لا يكون هناك أساساً يستند إليه فى اتخاذ قرار فيما يتعلق بمكانة الفرد أو تحقيقه لما يشتمل عليه النطاق السلوكي لهذه المفردات ، لذلك فإن ثبات درجات الاختبارات يؤثر فى نوعية القرارات المتعلقة بالأفراد أو الجماعات .

➢ مصادر أخطاء القياس التى تؤثر فى الثبات :

تتميز درجات الاختبارات بقدر مرتفع من الثبات إذا كانت تعكس الجوانب الحقيقة للسمة أو الخاصة أو القدرة المراد قياسها ، ولا تعكس عوامل الخطأ والعشوائية ، ويزداد ثبات الدرجات إذا قلت هذه العوامل ، وتوجد مصادر متعددة تسهم فى أخطاء القياس التي تؤدي إلى انخفاض ثبات درجات الاختبارات والمقاييس التربوية والنفسية ، بعضها يتعلق بأداة القياس ذاتها ، وبعضها يتعلق بإجراءات وظروف تطبيقها ، وبعض آخر يتعلق بالأفراد المختبرين.

➢ أخطاء تتعلق بأداة القياس :

سبق أن أوضحنا فى الفصل الأول أن هناك خصائص عامة ينبغي توافرها فى أدوات القياس لكي تكون دراجاتها متسقة بدرجة كبيرة ، فعدم ملاءمة محتوى الاختبار للأفراد المختبرين ، ومستوى صعوبة مفرداته ، أو غموضها وعدم دقتها ، وكذلك غموض تعليمات الإجابة عليها ، وطول الاختبار ، والزمن اللازم للإجابة عليه ، واشتماله على عينة ممثلة لمكونات السمة المراد قياسها ، وغير ذلك تعد من مصادر الخطأ المتعلقة بأداة القياس ، وتؤدي إلى عدم اتساق الدرجات ، كما تؤثر أيضاً فى صدق الاختبار ، حيث إن المعلومات والقرارات التي تستند إليها تكون غير موثوق بها ، أي غير صادقة .

➢ أخطاء تتعلق بإجراء الاختبار ، وظروف تطبيقه :

يتأثر ثبات درجات الاختبارات ليس فقط بتصميم وبناء أداة القياس التي تستخدم فى جمع المعلومات ، وإنما أيضاً بكيفية استخدام الأداة ، فتطبيق الاختبارات يتطلب أن تكون تعليمات الإجابة واضحة وكافية ومناسبة للمختبرين ، كما أن ظروف تطبيق الاختبار ينبغي أن تكون جيدة ومقننة ، فالاختبار الذى يطلق على مجموعة كبيرة من الأفراد فى غرفة ضيقة وغير جيدة التهوية ، وقليلة الإضاءة ، وتحيط بها الضوضاء ، تؤثر تأثيراً كبيراً فى درجات الأفراد ، مما يؤدي إلى انخفاض ثبات هذه الدرجات وعدم صدق المعلومات المستمدة منها ، وكذلك تعد عملية وإجراءات تصحيح الاختبار أو تقدير درجاته ، وتدوين نتائجه من مصادر الأخطاء الشائعة فمن السهل الوقوع فى

الأخطاء الحسابية عند جمع الدرجات أو إغفال تصحيح بعض الأسئلة ، أو عدم الدقة فى تصحيح الأسئلة التى تتطلب استجابات مفتوحة ، وغير ذلك .

- **أخطاء تتعلق بالأفراد المختبرين :**

توجد أخطاء متعددة تتعلق بالأفراد المختبرين ، وتسهم فى انخفاض ثبات درجات الاختبارات ، ومن أمثلة ذلك :

انخفاض الدافعية لدى المختبرين وقد تطبيق الاختبار لأسباب ربما ترجع إلى عدم معرفتهم بالغرض من الاختبار أو اتجاههم السلبي نحوه ، أو قلة أنتباههم أو شعورهم بالإجهاد أو التعب أو الملل ، أو انخفاض قدرتهم القرائية ، أو حالتهم الصحية والنفسية ، وأيضاً مستواهم الاجتماعي الاقتصادي ، أو تفاوت تدربهم على نوع معين من الاختبارات ، كما أن عدم استقرار بعض السمات الإنسانية وتذبذبها عبر الزمن يجعل من الضروري الحصول على معلومات تتعلق بهذه السمات فى الوقت الذى سوف تستخدم فيه هذه المعلومات فى اتخاذ قرارات بشأن الأفراد المختبرين .

لذلك ينبغي معرفة مصادر الخطأ الممكنة التي يمكن أن تؤثر فى استجابات الأفراد للاختبارات والمقاييس التربوية والنفسية ، وتؤدي إلى عدم اتساق الدرجات، مما يسهم فى تحسين طرق القياس ، واستخدام نتائج القياس استخداما مستبصراً ، ومعرفة حدود الدقة فى تفسير هذه النتائج .

- **طرق تقدير ثبات درجات الاختبارات :**

لكي تكون هناك ثقة فى استخدام نتائج الاختبارات والافادة منها ينبغي ضبط مصادر الأخطاء المتعددة التي أوضحناها ، فتقنين الاختبار ، أي ضبط جميع العوامل المغتربة بقدر الامكان يؤدي إلى زيادة ثبات الدرجات المستمدة منه.

والدرجات التى يحصل عليها الفرد فى الاختبار تسمي الدرجات الملاحظة Obtained Scores ، وسواء كانت هذه الدرجات تقديرات أو درجات خام ، أو درجات محولة ، فإنه يشوبها جميعاً قدراً من الخطأ ، فإذا استطعنا تقدير هذا الخطأ فى

الدرجات الملاحظة فإنه يكون لدينا " درجة الخطأ Error Score "، ويمكن الحصول على " الدرجة الحقيقة True Score " للفرد بطرح درجة الخطأ من درجة الملاحظة .

أي أن الدرجة الحقيقة = الدرجة الملاحظة – درجة الخطأ .

وهذا يعني أن الدرجة الملاحظة تكون من الدرجة الحقيقة ، ودرجة الخطأ، ويطلق على درجة الخطأ " خطأ القياس Error of Measurement " ، ودرجة الخطأ يمكن أن تكون موجبة أو سالبة أو صفر ، فإذا كانت درجة الخطأ موجبة تزيد الدرجة الملاحظة عن الدرجة الحقيقة ، والعكس إذا كانت درجة الخطأ سالبة.

ومعامل الثبات Reliability Coefficient يزودنا بتقدير كمي للعلاقة بين الفروق الفردية فى الدرجات الحقيقية ، والفروق الفردية فى الدرجات الملاحظة ، والدرجات الملاحظة الخالية من الأخطاء تنتج عن اختبار تتسم درجاته بالثبات التام ، وهذا يجعل الفروق الفردية فى الدرجات الملاحظة مساوياً للفروق الفردية فى الدرجات الحقيقية ، وتكون قيمة معامل الثبات فى هذه الحالة مساوياً الواحد الصحيح ، أما الاختبار الذى تتسم درجاته بعدم الثبات على الإطلاق فسوف تكون الدرجات الملاحظة المستمدة منه ليست إلا أخطاء قياس ، وعندئذ تكون قيمة معامل الثبات صفراً .

لذلك فإن قيمة معامل الثبات تتراوح بين (صفر ، ١) ، وهو بذلك يعكس درجة اتساق درجات الاختبار ، ويمكن تفسيره بانه نسبة من الفروق الملاحظة بين الأفراد التى تعزي إلى الفروق الحقيقية بينهم .

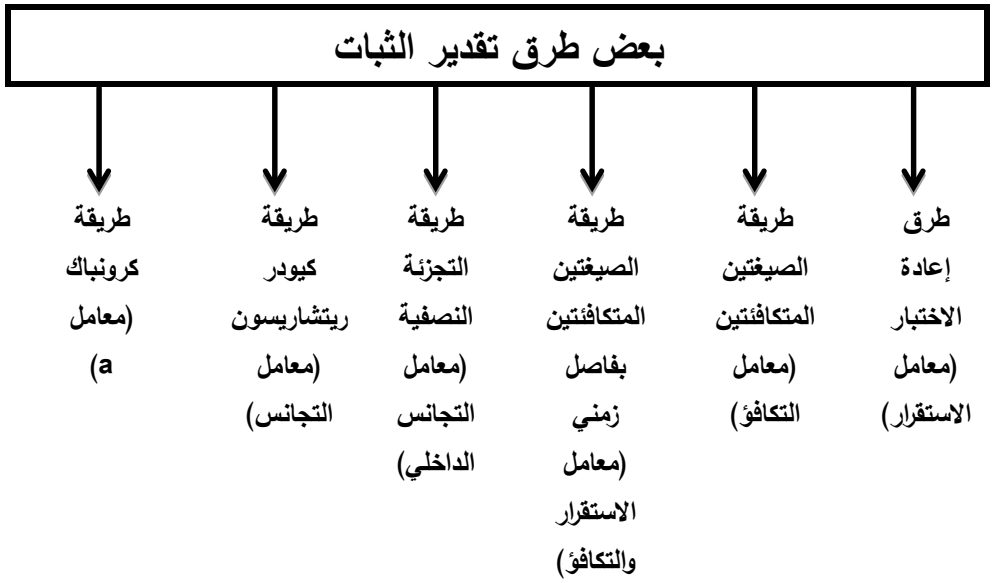
ويمكن تكميم الفروق فى الدرجات الحقيقية بين الأفراد باستخدام تباين هذه الدرجات وكذلك الفروق فى الدرجات الملاحظة .

لذلك يمكن تعريف معامل الثبات بأنه النسبة بين تباين الدرجات الحقيقية وتباين الدرجات الملاحظة .

$$\textbf{أي أن معامل الثبات} = \frac{\textbf{تباين الدرجات الحقيقية}}{\textbf{تباين الدرجات الملاحظة}}$$

وهذا يعني أن معامل الثبات هو مدى ما تعكسه الفروق فى الدرجات على الفروق فى الدرجات الحقيقية ، أو مدى إعزاء تباين الدرجات الملاحظة إلى تباين الدرجات الحقيقية ، ونظراً لأن تباين الدرجات الحقيقية يكون غير معلوم ، فإننا نود تقدير الدرجة الحقيقية للفرد بأستخدام درجته الملاحظة ، وهذا يتطلب تقدير درجة الخطأ أو تباين الخطأ ، وهو ما يهدف إليه مفهوم الثبات .

وتختلف طرق تقدير ثبات درجات الاختبارات باختلاف مصادر الخطأ الذى نود التحكم فيها بحيث لا ينعكس أثرها فى الدرجة الملاحظة ، ويوضح الشكل التخطيطي التالي (٤-١) أهم هذه الطرق شائعة الاستخدام فى الاختبارات المقننة المنشورة :



شكل (٤-١) يوضح بعض طرق تقدير الثبات

وفيما يلي توضيحاً لكل من هذه الطرق والاستخدام المناسب لها :

(١) طريقة إعادة الاختبار (معامل الاستقرار) (Test- retest (SAtability:

للتحقق من استقرار درجات الاختبار ، يطبق الاختبار على عينة ممثلة من الأفراد ، ثم يعاد تطبيقه بعد مدة زمنية معينة، وتتوقف هذه المدة الزمنية الفاصلة على

طبيعة السمة التي يقيسها الاختبار، فإذا كان يقيس سمة سريعة التذبذب والتغير ، مثل : ذكاء الأطفال الصغار ، أو السمات المزاجية ، فإن هذه الطريقة لا تكون مناسبة ، ويحسن استخدام طريقة أخري من الطرق التى سنذكرها ، وإذا كانت المدة الفاصلة قصيرة ، فإن الذاكرة قصيرة الأمد سوف تؤثر فى نتائج التطبيق فى المرة الثانية بحيث تجعلها أكثر اتساقاً مع نتائج التطبيق الأول ، وإذا كانت هذه المدة طويلة، فإن النضج ، والعوامل البيئية ، سوف تؤثر أيضاً فى نتائج التطبيق الثاني، وعندئذ يصبح معامل الثبات أو ما يسمي فى هذه الحالة "معامل الاستقرار " Stability Coefficient " محل تساؤل ، لذلك يجب العناية باختيار المدة الزمنية الفاصلة اعتماداً على لطبيعة السمة المراد قياسها ، وخصائص المجتمع المستهدف.

ويمكن توضيح هذه الطريقة بالشكل (٤-٢) التالي :

مدة زمنية فاصلة

تطبيق الاختبار ⟵ إعادة تطبيق الاختبار

شكل (٤-٢) يوضح طريقة إعادة الاختبار

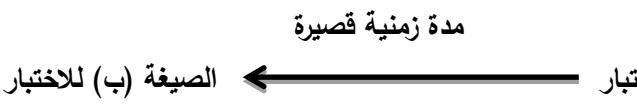
ويمكن عندئذ إيجاد معامل ارتباط بيرسون الذى سبق توضيحه فى الفصل الثالث بين درجات مجموعة الأفراد فى الاختبار نفسه فى مرتي التطبيق ، فإذا كانت استجابة الأفراد متسقة فى المرتين، يكون الارتباط مرتفعاً ، وبالتالي يكون ثبات الدرجات مرتفعاً ، ويحسن ذكر المدة الزمنية الفاصلة إلى جانب مرتفعاً ، ويحسن ذكر المدة الزمنية الفاصلة إلى جانب قيمة معامل الثبات التى نحصل عليها.

ومن الجدير بالذكر أن التغيير الذي يحدث فى الدرجات فى المرتين يُنظر إليه على أنه نتيجة أخطاء القياس ، بافتراض أن السمة ليست سريعة التغير، أي مستقرة نسبياً .

(٢) طريقة تطبيق صيغتين متكافئتين للاختبار (معامل التكافؤ Equivalence Coefficient) :

يصعب أحياناً تطبيق الاختبار على نفس المجموعة من الأفراد مرتين ، أو أن السمة المراد قياسها تتميز بعدم الاستقرار ، فعندئذ يمكن استخدام صيغتين متكافئتين

للاختبار نفسه ، وتطبيقها على مجموعة الأفراد نفسها الواحدة تلو الأخري ، كما يوضحه الشكل (٤-٣) التالي :



شكل (٤-٢) يوضح طريقة إعادة الاختبار

وتستند هذه الطريقة على فرضية أن كلاً من صيغتي الاختبار تكونان فى الحقيقة متكافئتين من حيث المحتوى ، ونوع المفردات ، وعددها ، وصعوبتها، والمتوسط الانحراف المعياري لدرجات كل منهما .

ويمكن إيجاد معامل الارتباط بين درجات كل من الصيغتين للحصول على معامل الثبات الذى يسمي فى هذه الحالة " معامل التكافؤ Equivalence Coefficient " .

وترجع أخطاء القياس فى هذه الحالة إلى اختلاف مفردات أو أسئلة كل من الصيغتين فكلما زاد الاختلاف بينهما انخفض الثبات ، أي أن أخطاء القياس فى هذه الحالة ترجع إلى اختلاف عينات الأسئلة فى كل من الصيغتين ، وليس إلى التغيرات التى تحدث للأفراد كما هو الحال فى معامل الاستقرار .

وكثير من الاختبارات المقننة يكون لها صيغتين متكافئتين على الأقل للاستخدام فى حالة تغيب بعض الأفراد عن الاختبار ، أو لأي أسباب أخري تتطلب عدم تذكر الإجابات فى الصيغة الأولي ، أو غير ذلك ، وتدوَّن تقديرات معامل التكافؤ فى أدلة هذه الاختبارات .

(٣) طريقة تطبيق صيغتين متكافئتين بفاصل زمني Stability & Equivalence :

أحياناً نهتم بالتنبؤ بعيد المدى ، والاستدلال على أداء الفرد فى نطاق من المعارق أو المهارات ، فعندئذ نحتاج إلى معامل يشمل كل من معاملي الاستقرار والتكافؤ ، فإذا كان لدينا صيغتين للاختبار ، فإننا نطبق الصيغة الألي ، وبعد انقضاء مدة زمنية نطبق الصيغة الثانية كما هو موضح بالشكل (٤-٤) التالي:

مدة زمنية فاصلة

الصيغة (١) للاختبار ⟵ الصيغة (ب) للاختبار

شكل (٤-٤) يوضح طريقة تطبيق صيغتين متكافئتين بفاصل زمني

ويمكن إيجاد معامل الارتباط بين مجموعتي الدرجات فى كل من صيغتي الاختبار.

ويلاحظ أن جميع مصادر الخطأ التي تؤثر فى الطريقتين الأولي والثانية والتى تؤدي إلى عدم اتساق درجات الاختبار تعد مصادر أخطاء فى هذه الطريقة، لذلك يتوقع انخفاض قيمة معامل الثبات فى هذه الحالة عنها فى الطريقتين السابقتين ، ويسمي المعامل الناتج " معامل الاستقرار والتكافؤ Stability & Equivalence Coefficient " .

غير أن هذا المعامل يكون ذا أهمية فى قياس التحسن فى الأداء ، أو عندما نود التعميم على المواقف أو الظروف ، وعينات مفردات الاختبار ، ونعني بذلك مدى دقة الاختبار فى قياس سمة معينة بغض النظر عن صيغة الاختبار المستخدمة ، أو الظروف التى طبق فيها الاختبار .

ويحدث ذلك فى كثير من الأحيان ، فإذا أردنا تقدير قدرة فرد معين على التدريس مثلاً ، فإننا ربما نود التركيز على اتساق التقديرات التى يتم الحصول عليها فى ظروف ومواقف مختلفة بواسطة أكثر من شخص قائم بالتقدير ، لذلك ينبغي عند استخدام هذا المعامل فى تقدير ثبات درجات الاختبارات تحديد الظروف والمواقف بدقة لكي يستند إليها تفسير هذه الدرجات .

(٤) طريقة التجزئة النصفية (معامل الاتساق الداخلي Internal Consistency Coefficient :

الطرق الثلاث السابقة لتقدير الثبات تناسب بدرجة كبيرة الاختبارات المقننة Standard – I zed غير أنه فى كثير من المواقف يصعب معامل الاستقرار، أو أي من معاملي التكافؤ السابقة ، فالمعلم فى الصف المدرسي ربما لا يود تطبيق للاختبار إذا توافرتا ، أو يعيد تطبيق الاختبار نفسه فى مرتين متتاليتين ، لذلك توجد عدة طرق لتقدير

الثبات تعتمد على تطبيق الاختبار مرة واحدة ، وتعد طريقة التجزئة النصفية Split-Half Method أكثر هذه الطرق استخداماً ، وتتطلب هذه الطريقة تقسيم الاختبار إلى نصفين متكافئين إحصائياً ، وهذان النصفان يكونان مستقلين ومتكافئين فى المحتوى ، والصعوبة ، ويتساوي كل من المتوسط والانحراف المعياري لدرجاتهما ، وليس من الضروري تقسيم الاختبار بحسب الأسئلة أو المفردات الفردية والزوجية إلا إذا حقق هذا التقسيم الشروط المذكورة .

وبعد تطبيق الاختبار كوحدة وتصحيح نتائجه ، ترصد درجات أسئلة أو مفردات النصف الأول على حدة ، ودرجات أو أسئلة النصف الثاني على حدة ، ثم نوجد معامل الارتباط بين درجات كل من النصفين .

ونظراً لأن معامل الارتباط الذى نحصل عليه يكون مقياساً لثبات درجات نصف الاختبار ، فإنه يجب استخدام القيمة الناتجة فى تقدير ثبات درجات الاختبار ككل ، حيث إن قيمة معامل ثبات درجات نصف الاختبار تكون أقل من ثبات درجات الاختبار التى كان من الممكن الحصول عليها من الاختبار بطوله الأصلي .

وقد اقترح سيبرمان وبراون Spearman & Brown صيغة رياضية يمكن استخدامها فى تقدير معامل ثبات درجات الاختبار ككل ، وذلك من قيم معامل الارتباط بين درجات كم من نصفي الاختبار ، وهذه الصيغة كالتالي :

$$\text{معامل ثبات درجات الاختبار} = \frac{\text{٢ × الارتباط بين نصفي الاختبار}}{\text{١ + الارتباط بين نصفي الاختبار}}$$

$$ر_{١١} = \frac{٢\, ر_{٢١}}{١ + ر_{٢١}} \quad \ldots\ldots\ldots (٤-١)$$

حيث ترمز (ر١١) إلى معامل ثبات درجات الاختبار الذي يتم تقديره ، (ر٢١) إلى معامل الثبات الناتج عن التجزئة النصفية .

فإذا كان معامل الارتباط بين نصفي الاختبار ($ر_{٢١}$ = ٠.٧٥) .

فإن تقدير معامل ثبات الاختبار ككل = $\frac{٢ \times ٠.٧٥}{١ + ٠.٧٥}$ = $\frac{١.٥٠}{١.٧٥}$ = ٠.٨٠ **تقريبا**

ويلاحظ أن قيمة ثبات درجات الاختبار ككل قد زادت من ٠.٧٥ إلى ٠.٨٦ ، وهذا يدل على أن ثبات درجات الاختبار يزداد بازدياد طول الاختبار بشرط أن تقيس أسئلته أو مفرداته التى أدت إلى زيادة الطول السمة نفسها التى تقيسها بقية الأسئلة أو المفردات .

ويمكن تعميم الصيغة (٤-١) لتقدير ثبات درجات الاختبار إذا زاد نقص طوله (ن) من المرات لتصبح كالتالي :

$$ر_{نن} = \frac{ن\ ر_{كك}}{١ + (ن-١)\ ر_{كك}} \quad .....(٤-٢)$$

حيث ترمز ($ر_{نن}$) إلى القيمة التقديرية لمعامل الثبات إذا زاد طوله (ن) من المرات (ن) تساوي عدد مفردات الاختبار بعد الإطالة مقسوماً على عدد مفرداته الأصلية .

، ($ر_{كك}$) إلى ثبات درجات الاختبار الأصلي قبل الإطالة

ويلاحظ أننا إذا وضعنا ن =٢ فى الصيغة (٤-٢) فإننا نحصل على الصيغة (٤-١) السابقة التى تعد حالة خاصة منها ، ويتطلب استخدام الصيغة العامة (٤-٢) تحقق الشروط التى سبق ذكرها .

فإن كان معامل ثبات درجات الاختبار الأصلي ٠.٧٥ ، وزاد طوله ثلاث مرات ، فإن القيمة التقديرية لثبات الدرجات بعد إطالة الاختبار تكون كالتالي :

$$ر = \frac{٣ \times ٠.٧٥}{١ + (٣-١)\ ٠.٧٥} = \frac{٢.٢٥}{٢.٥٠} = ٠.٩٠$$

وينبغي التحقق من تساوي تباين درجات كل من نصفي الاختبار ، أما إذا اختلف التباين ، فإنه يمكن استخدام الصيغة التى اقترحها جاتمان Guttman,1945) وهي كالتالي :

$$ر_{ن\,ن} = ٢ \left[١ - \frac{ع_{٢\,أ} \div ع_{٢\,ب}}{ع_{٢\,ن}} \right] \quad \ldots\ldots\ldots (٣-٤)$$

حيث ترمز (ع أ) ، (ع ب) إلى الانحراف المعياري لدرجات كل من نصفي الاختبار ، (ع ن) إلى الانحراف المعيارى لدرجات الاختبار ككل .

وتطبيق هذه الصيغة لا يتطلب إيجاد الارتباط بين درجات نصفي الاختبار ، وإنما يكتفي بإيجاد الانحرافات المعيارية للدرجات .

ويلاحظ أن أي من هذه الصيغ يتطلب تطبيق الاختبار مرة واحدة ، أي تطبيق كل من نصفي الاختبار فى الوقت نفسه ، ومع هذا يعاب على طريقة التجزئة النصفية أن المصدر الرئيس للخطأ يتعلق بالمفردات ذاتها ، وهذا يعني أن عدم الثبات ينتج عن الفروق فى عينة المفردات التى يشتمل عليها كل من نصفي الاختبار ، لذلك فإن هذه الطريقة تؤدي إلى قيم مرتفعة نسبياً لمعامل الثبات ، وينبغي عدم استخدام هذه الطريقة فى تقدير ثبات درجات الاختبارات التى تعتمد على السرعة ، وذلك لأنها تؤدي إلى قيم زائفة لمعامل الثبات ، وإنما يقتصر استخدامها على الاختبارات التى تعتمد على القوة .

(٥) طريقة كيودر – ريتشاردسون (معامل التجانس) Homogeneity Coefficient :

تعتمد هذه الطريقة أيضاً على تطبيق الاختبار فى مرة واحدة ، ويمكن استخدامها التحقق من تجانس جميع مفردات الاختبار Test Homogeneity ، وهذا التجانس يتأثر بكل من معاينة المحتوى Contentt Sampling كما فى طريقة التجزئة النصفية ، وعدم اتساق النطاق السلوكي الذي تستمد منه عينة مفردات الاختبار ، فكلما زاد هذا الاتساق دل ذلك على تجانس مفردات الاختبار ، فالاختبار الذى يشتمل على مفردات تقيس عملية الجمع البسيط فقط ، يكون أكثر اتساقاً أو تجانساً من الاختبار الذى يقيس

عمليتي الجمع والطرح ، وتتغلب هذه الطريقة على مشكلة تجزئة الاختبار إلى نصفين متكافئين ، حيث إنها تعتمد على إيجاد قيمة تقديرية لمتوسط قيم معاملات الارتباط بين كل من نصفي الاختبار إذا تم تجزئته بجميع الطرق الممكنة دون أن نلجأ إلى هذه التجزئة فعلياً .

وقد اقترح كل من كيودر وريتشاردسون Kuder & Richardson صيغتين رياضيتين أطلقا عليهما الصيغة (٢٠) والصيغة (٢١) ، ويرمز لهما بالرمزين (K–R20) ، (k–R21) غير أنني سأقتصر على ذكر الصيغة (٢٠) لأنها أكثر شيوعاً واستخداماً فى الاختبارات والمقاييس وأكثر دقة من الصيغة (٢١).

$$\text{معامل التجانس} = \frac{\text{ن}}{\text{ن}-1} \times \left(\frac{ع2 - \text{مجـ س ص}}{ع2}\right) \quad \ldots\ldots (٤-٤)$$

$$= \frac{\text{ن}}{\text{ن}-1} \times \left(1 - \frac{\text{مجـ س ص}}{ع2}\right) \quad \ldots\ldots (٤-٥)$$

حيث ترمز (ن) إلى العدد الكلي لمفردات الاختبار .

، (س) إلى نسبة عدد الأفراد الذين أجابوا على أي مفردة فى الاختبار إجابة صحيحة

، (ص) إلى نسبة عدد الأفراد الذين أجابوا على أي مفردة إجابة خطأ (ص=١–س).

، (مجـ س ص) إلى مجموع تباين درجات مفردات الاختبار .

، (ع٢) إلى تباين الدرجات الكلية فى الاختبار (مربع الانحراف المعياري) .

ويتطلب تطبيق هذه الصيغة إيجاد قيمة الانحراف المعياري لدرجات الاختبار كما أوضحنا فى الفصل الثالث ، ونسبة عدد الإجابات الصحيحة والإجابات الخطأ لكل مفردة من مفردات الاختبار ، وإيجاد حاصل ضرب النسبتين لكل مفردة وجمع الناتج .

فإذا كانت ع٢ = ٢.٥ ، مجـ س ص = ١.٠٩ ، ن = ٥

فإن معامل التجانس $= \frac{5}{1-5} \times \left(\frac{1.09-2.5}{2.5}\right) = \frac{5}{4} \times = \frac{1.41}{2.5} =$ ٠.٧١

وتفترض هذه الطريقة أن الاختبار من نوع القوة كما فى طريقة التجزئة النصفية ، لذلك فهي لا تناسب الاختبارات التى تعتمد على السرعة ، كما تفترض أن مفردات الاختبار تقيس سمة واحدة بُعد واحد ، وأن المفردات ثنائية الدرجة كما فى مفردات الاختبار من متعدد ، أو الصواب أو الخطأ ، وتعد طريقة النصفية حالة خاصة من هذه الطريقة .

(٦) معامل (a) لكرونباك Coefficient (a) :

نظراً لأن معامل التجانس السابق يقتصر استخدامه على المفردات ثنائية الدرجة ، فقد اقترح كرونباك (Cronback,1951) طريقة تناسب اختبارات القوة التى تشتمل على مفردات متعددة الدرجات ، مثل : أسئلة المقال ، أو فقرات الاستبيانات التى تتطلب إجابتها الاختيار من بين بدائل متعددة ، وتكون درجاتها متدرجة من ١ إلى ٥ مثلاً .

والصيغة الرياضية لمعامل (a) ، ويقرأ (معامل ألفا) كالتالي :

$$\text{معامل (a)} = \frac{\text{ن}}{\text{ن}-1} \times \left(\frac{\text{مجـ ع}^2\text{م}}{\text{ع}^2\text{ك}}\right) \quad \ldots\ldots \quad (٤-٦)$$

حيث ترمز (ن) إلى العدد الكلي لمفردات الاختبار .

، **(مجـ ع٢م) إلى مجموع تباين درجات كل مفردة (هـ) .**

، **(ع٢ ك) إلى تباين الدرجات الكلية فى الاختبار .**

والصيغة (٤٠٦) تماثل الصيغة (٤-٥) السابقة فيما عدا الرمز (مجـ ع٢م) الذى يمثل تباين درجات جميع الأفراد فى إحدى المفردات التي يقوم درجاتها محكم واحد ، ومن ثم نجمع تباين درجات جميع المفردات .

فإذ ت كانت ن = ٤ مفردات .

، تباين درجات المفردات الربع (ع٢م)= ١.٨١+١.٠٠+٢.٠٠+١.٥٨= ٦.٣٩.

، وتباين الدرجات الكلية فى الاختبار يتطلب جمع درجات كل فرد فى المفردات الأربع وحساب الانحراف المعياري لهذه الدرجات الكلية كما سبق أن أوضحنا فى الفصل الثالث ، وتربيع القيمة الناتجة للحصول على التباين وهو يساوي ١٦.٨٩.

ولإيجاد قيمة معامل (a) نستخدم الصيغة (٤-٦) كالتالي :

$$\text{معامل (a)} = \frac{4}{1-4} \times \left(\frac{6.39}{16.89}\right) = \frac{4}{3} \left(١ - ٠.٣٨\right)$$

$$= \left(١.٣٣\right) \left(٠.٦٢\right) = ٠.٨٣$$

ومن الجدير بالذكر أن صيغتي كيوردر – ريتشاردسون ، وكرونباك تستندان إلى أن اتساق استجابات الأفراد عبر مفردات الاختبار يمكن الاعتماد عليه فى تقدير معامل الثبات للدرجات الكلية فى الاختبار ، فإذا كانت مفردات الاختبار متجانسة ، أي تقيس سمة أو بُعد واحد ، فإن كلاً من الصيغتين تؤدي من النتيجة إلي نحصل عليها من تطبيق طريقة التجزئة النصفية ، أما إذا كانت المفردات غير متجانسة فإن قيمة معامل الثبات الناتجة عن تطبيق كل من الصيغتين سوف تكون أقل من القيمة الناتجة عن التجزئة ، والحقيقة أن قيم ل من معامل التجانس (الصيغة ٢٠) ، أو معامل (a) تُعد بمثابة الحد الأدني Lower Bound للقيم التقديرية لمعامل الثبات.

الخطأ المعياري للقياس :

نظراً لأن القياس النفسي والتربوي يتأثر بمصادر أخطاء متعددة كما أوضحنا ، فإنه لا يوجد اختبار يقيس سمة معينة لدى الفرد قياساً متسقاً تاماً ، لذلك ينبغي تقدير قيمة هذه الاخطاء ، وأخذها بعين الاعتبار عند تفسير درجات الاختبار ، فإذا استطعنا تقدير الدرجات الحقيقة للأفراد فى الاختبار ، فإن الانحراف المعياري لتوزيع هذه الدرجات يُطلق عليه " الخطأ المعياري للقياس Standard Error of Measurement (SEM) وهو مؤشر لمدى اختلاف الدرجات الملاحظة للأفراد فى الاختبار عن درجاتهم الحقيقة ، ويفيد هذه المؤشر فى تقدير قيمة الخطأ المعياري للقياس لكل فرد فى الاختبار دون الحاجة إلى تكرار تطبيق الاختبار عليه عدة مرات لتعرف مدى اتساق درجاته .

والخطأ المعياري للقياس هو الانحراف المعياري للخطأ ، ويزودنا بقيمة مطلقة وليست نسبية لمدى انطباق الدرجات الملاحظة والدرجات الحقيقية ، فمعاملات الثبات تعتمد على تباين درجات مجموعة الأفراد ، بينما الخطأ المعياري للقياس لا يتأثر بهذا التباين ، وتقدير الخطأ المعياري للقياس لا يتطلب معرفة الدرجات الحقيقة.

وقد سبق أن أوضحنا فى مستهل هذا الفصل أن معامل الثبات (ر١١) هو النسبة بين تباين الدرجات الحقيقية (ع٢ح) ، وتباين الدرجات الملاحظة (ع٢م).

أي أن : $ر_{11} = 1 - \frac{ع^2_{ح}}{ع^2_{م}}$ ..... (٤-٧)

ولكن تباين الدرجات الحقيقية = تباين الدرجات الملاحظة – تباين درجات الخطأ.

أي أن : $ع^2_{ح} = ع^2_{م} - ع^2_{ح}$ ..... (٤ -٨)

وبالتعويض من (٤-٨) في (٤-٧) نجد أن :

$$ر_{11} = 1 - \frac{ع^2_{م} + ع^2_{ح}}{ع^2_{م}} = \frac{ع^2_{م}}{ع^2_{م}} = \frac{ع^2_{ح}}{ع^2_{م}} = 1 - \frac{ع^2_{ح}}{ع^2_{م}}$$

أي أن : $ر_{11} = 1 - \frac{ع^2_{ح}}{ع^2_{م}}$

$$1 - ر_{11} = \frac{ع^2_{ح}}{ع^2_{م}}$$

$ع^2_{ح} = ع^2_{م} (1 - ر_{11})$

$ع_{ح} = ع_{م} \sqrt{1 - ر_{11}}$ ..... (٤ – ٩)

وهذه هي الصيغة المستخدمة فى إيجاد القيمة التقديرية للخطأ المعياري للقياس ، أو الانحراف المعياري للدرجة للفرد فى الاختبار (بالتطبيق المتكرر للاختبار نفسه) حول درجة الحقيقة فى الاختبار .

فإذا افترضنا أن أخطاء القياس للفرد فى التطبيقات المتكررة للاختبار تُعد أخطاء عشوائية ، فإن متوسط الدرجات الملاحظة فى هذه المرات يكون تقديراً لدرجته الحقيقية ، وذلك لأن متوسط الأخطاء العشوائية يساوي صفراً ، لذلك فإن التوزيع الافتراضي للدرجات الملاحظة يكون اعتدالياً متوسطه هو الدرجة الحقيقية للفرد فى الاختبار .

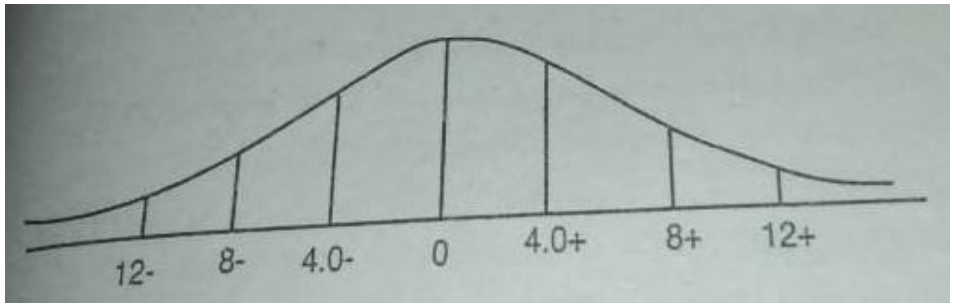
ويمكن استخدام خصائص المنحنى الاعتدالي المعياري التي سبق توضيحها فى الفصل الثالث فى تفسير القيمة التقديرية للخطأ المعياري للقياس التى تنتج عن تطبيق الصيغة (٤-٩)

ولتوضيح ذلك نفترض أن الانحراف المعياري للدرجات الملاحظة = ١٠ ، ومعامل ثبات درجات الاختبار = ٠.٨٤ ، فإن الخطأ المعياري للقياس .

$$ع_{ح} = ع_{م} \sqrt{١ - ر_{١١}}$$

$$= ١٠ \sqrt{١ - ٠.٨٤} = ١٠ \sqrt{٠.١٦} = ١٠ (٠.٤) = ٤.٠$$

ولتفسير هذه القيمة التقديرية لأخطاء القياس نرسم المنحني الاعتدالي ، ونحدد عليها هذه القيمية باعتبارها تمثل الانحراف المعياري للأخطاء كالتالي :



شكل (٤-٥) يوضح التوزيع الاعتدالي لدرجات الخطأ التى متوسطها صفر ، وانحرافها المعياري = ٤.٠

ويتضح من شكل (٤-٥) أننا نتوقع أن تنحصر درجة الفرد الحقيقية بين (الدرجة الملاحظة + الخطأ المعياري للقياس) ، أي بين (الدرجة الملاحظة +٤) ، ٦٨% من المرات .

فإذا كانت درجته الملاحظة ٢٠ ، فإن درجته الحقيقية يحتمل أن تنحصر بين (٢٠-٤ ، ٢٠ +٤) أ بين (١٦ ، ٢٤) عدد من المرات نسبتها ٩٨% تقريباً ، وبذلك يمكن معرفة حدود الخطأ فى الدرجة التي يحصل عليه الفرد من الاختبار ، وتقرير ذلك عند تفسير هذه الدرجة .

➢ بعض العوامل التى تؤثر فى تفسير قيم معامل الثبات :

عند تفسير القيم التقديرية لمعامل الثبات التى تحصل عليها باستخدام أي من الطرق السابقة ، وبخاصة عند مقارنة معاملات ثبات درجات اختبارات مختلفة، ينبغي مراعاة بعض العوامل ، نذكر منها ما يلي :

(١) الدرجات المستمدة من اختبار عدد مفرداته أو أسئلته كبيراً نسبياً تكون أكثر ثباتاً مما لو كان عدد المفردات قليلاً ، لعل هذا يتضح إذا رجعنا إلى الصيغة العامة لسبيرمان وبراون (٤-٢) المذكورة سابقاً .

(٢) تتذبذب القيم العددية لمعامل الثبات من عينة إلى أخري من الأفراد .

(٣) كلما كان مدى الدرجات ضيقاً قلت قيمة معامل الثبات ، وهذا يعني أنه كلما كانت قدرة الأفراد المختبرين متقاربة ، فإن قيمة معامل الثبات تميل إلى الانخفاض ، وذلك لأنها تعتمد على معامل الارتباط الذى يتأثر قيمته بتشتت الدرجات كما ذكرنا فى الفصل الثاني .

(٤) كلما زادت المدة الزمنية الفاصلة بين مرتي تطبيق الاختبار ، انخفضت قيمة كل من معامل الاستقرار ، ومعامل التكافؤ .

(٥) الموضوعية فى تصحيح الاختبار أو تقدير درجاته تزيد من ثبات درجات الاختبار ، أما الذاتية التى تعني عدم اتساق التصحيح ، فإنها تسهم فى الأخطاء العشوائية التى تؤدي إلى انخفاض قيمة معامل الثبات .

(٦) زيادة صعوبة الاختبار ، وصعوبة مفرداته ، يجعل الأفراد يلجأون إلى التخمين العشوائي فى إجابتهم ، مما يزيد الأخطاء العشوائية ، ويقلل من ثبات درجات الاختبار .

ثانياً : مفهوم صدق وأهميته :

تؤدي معظم الاختبارات التربوية والنفسية إلى نتائج كيمة يُستند إليها فى وصف سمات معينة لدى الأفراد ، غير أننا لا نهتم فى معظم الأحيان بالدرجات ذاتها ، وإنما ما يمكن أن تمثله أو تفسره هذه الدرجات ، وبالطبع نود أن يكون هذا التفسير دقيقاً وموثوقاً به ثقة كاملة ، وهو ما لا يتحقق فى الواقع الفعلي : فمثلاً إذا طبقنا اختباراً على أحد الأفراد ، وأعدنا تطبيقه عليه بعد مدة زمنية معينة ، فإننا لا نتوقع أن يحصل هذا الفرد على الدرجة نفسها فى المرتين .

وكذلك إذا أردنا التنبؤ بأداء طلب الثانوية العامة فى الكلية التى سوف يلتحق بها ، فإنه يكون لدينا نتائج اختبار استعداد دراسي معلوم علاقته بالتقديرات فى العام الدراسي الأول بالكلية ، واستناداً إلى درجة الطالب فى الاختبار نتنبأ بأنه سيحصل على تقدير معين ، غير أنه ربما يحصل بالفعل على تقدير مختلف إلى حد ما عن التقدير المتنبأ به ، ففي كل من هذين المثالين نلاحظ أن هناك أخطاء فى القياس ، وأخطاء فى التنبؤ على الترتيب ، وقد أوضحنا فيما سبق أن هناك مصادر متعددة لهذه الأخطاء بعضها أخطاء غير منتظمة Unsystematic Errors، والبعض الآخر أخطاء منتظمة Systematic Errors ، فالأخطاء غير المنتظمة لا يكون لها نمطاً متسقاً ، وتتذبذب من موقف اختباري إلى آخر ، فتشتت الانتباه ، والتخمين ، وانخفاض الدافعية تعد من مصادر الخطأ غير المنتظم، بينما الأخطاء المنتظمة يكون تأثيرها متكافئ على الفرد فى مرتي تطبيق الاختبار ، أو على جميع الأفراد الذين يطبق عليهم الاختبار فى وقت معين ، فالتدريب ، والنسيان ، والتعب ، والنمو ، يمكن اعتبارها مصادر خطأ منتظم ، حيث أنها تعمل أحياناً على رفع الدرجات ، وأحياناً أخري تؤدي إلى أنخفاضها ، غير أن

مصادر الخطأ المنتظم وغير المنتظم تكون متداخلة فملائمة الغرفة التى يطبق قيها الاختبار ربما تسهم فى الأخطاء المنتظمة فى بعض الأحيان إذا ما اختبر جميع الأفراد فى غرفة سيئة التهوية أو الإضاءة مثلاً ، أو تسهم فى أحيان أخرى فى الأخطاء غير المنتظمة إذا ما طبق الاختبار فى غرفة متسعة ثم أعيد تطبيقه فى غرفة ضيقة ، ومفهوم ثبات درجات الاختبار Reliability الذى أوضحناه بتعلق بمشكلة ضبط وتقييم تأثير الأخطاء غير المنتظمة فى الدرجات ، بينما يتعلق مفهوم الصدق Validity بالإجراءات التجريبية المختلفة التى تستخدم فى تحديد مدى تأثر الاختبار بالأخطاء المتنظمة ، وهذا يتطلب توفر محك أو محكات لتقييم تواجد أو عدم تواجد هذه الأخطاء ، لذلك فإنه ينبغي تحديد هذه المحكات ، وإعدادها ، وجمع بيانات حولها ، لكي نتمكن من الحكم على الصدق، وتختلف طبيعة هذه المحكات التى تستخدم فى تقييم صدق الاختبار باختلاف أغراض الاختبار الذى يتم بناؤه ، واختلاف استخداماته .

➢ أنواع صدق الاختبار :

حددت الجمعية الأمريكية لعلم النفس ثلاثة أغراض رئيسية للاختبارات واستخداماتها كالتالي :

١- تحديد كيفية أداء الفرد فى نطاق شامل Universe من المهام أو المعارف أو المهارات التى يفترض أن الاختبار يشتمل على عينة ممثلة له، وهذا النوع من الصدق يسمي "صدق المحتوى Content Validity"

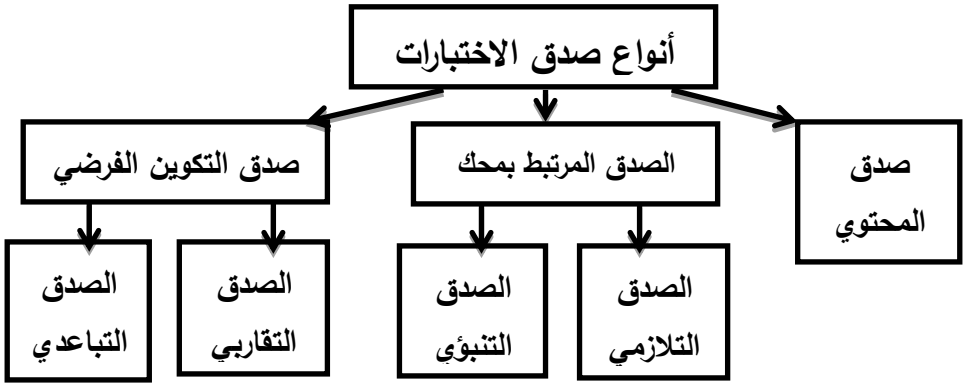
٢- التنبؤ بأداء مستقبلي ، أو تقدير الوضع الراهن للفرد فى تفسير معين يختلف عما يقدسه الاختبار ، ومن أمثلة ذلك ، التنبؤ بتقديرات الطلاب فى الكلية من درجاتهم فى أختبار الاستعداد الدراسي ، أو التنبؤ بالحالة الانفعالية للفرد من بروفيل درجات الاختبارات الفرعية لاختبار ويكسلر لذكاءة الراشدين ، وهذا النوع من الصدق يسمي " الصدق المرتبط بمحك Criterion-related Validity " .

٣- الاستدلال على درجة امتلاك الفرد لسمة أو خاصة أو تكوين فرضي ينعكس فى الاختبار ، فمثلاً ربما نود معرفة ما إذا كان الفرد يمتلك درجة مرتفعة من سمة

مجردة معينة مثل الذكاء ، أو الابتكارية ، ولا يمكن ملاحظة أي منهما ملاحظة مباشرة ويمكن التوصل إلى معلومات تتعلق بهاتين السمتين لدى الفرد ، أو تتعلق بدراسة الاختبار ذاته ، أو علاقته بغيره من الاختبارات ، أو ببناء نظرية سيكولوجية ، فنحن هنا نهتم بما يسمي " صدق التكوين الفرضي Construct Validity" للاختبار .

ونوع المعلومات المتعلقة بالصدق التى يتم الحصول عليها سوف تعتمد على الغرض من الاختبار وفيم تستخدم نتائجه ، فالاختبار ربما يكون صادقاً لغرض أو أغراض معينة ولا يكون كذلك لأغراض أخري ، لذلك فإن تحديد أنواع التفسيرات لدرجات الاختبار تجديداً واضحاً ، أو أنواع القرارات المراد اتخاذها استناداً إلى هذه الدرجات يعد أمراً أساسياً قبل التحقق من صدق الاختبار .

ويلخص الشكل التخطيطي (٤-٦) التالي هذه الأنواع الثلاثة من الصدق وفرعياتها :



شكل (٤-٦) يوضح أنواع صدق الاختبارات

ويعتمد بعض هذه الأنواع الثلاثة من الصدق على التحليل المنطقي للاختبار ومفرداته ومحتواها ، والبعض الآخر يعتمد على أدلة إمبريقية وإحصائية ، وسوف أوضح فيما يلي كل من هذه الأنواع بشئ من التفصيل :

➢ صدق المحتوي Content Validity :

يتعلق صدق محتوى الاختبار بالحكم على مدى كفاية مفرداته كعينة ممثلة لنطاق محتوى أو أهداف يُفترض أن الاختبار يقيسها ، وعلى الرغم من أن صدق المحتوى ربما ينطبق على جميع أنواع الاختبارات المعرفية ومقاييس الشخصية ، إلا أنه يناسب بدرجة أكبر الاختبارات التحصيلية المقننة ، أو التى يكتبها المعلم لطلابه ، والحقيقة أنه لا يوجد مؤشر كمي أو عددي لصدق المحتوى ، وإنما يتم تقييم هذا النوع من الصدق عن طريق فحص محتوى مفردات الاختبارات فحصاً منطقياً فى ضوء ما يقيسه الاختبار التحصيلي من أهداف أو مستويات تعليمية Standards ، وذلك لتعرف مدى كفاية مفرداته فى قياس هذه الأهداف أو المستويات فى ضوء محتوى مجال دراسي معين ، وهذا يتطلب تحديد النطاق الشامل Universe للمفردات المراد معاينتها ، أي يتم انتقاء عينة عشوائية أو عشوائية طبقية منها، وتضمينها فى الاختبار .

وفى حالة الاختبارات التحصيلية المقننة يحدد دليل الاختبار كيفية إجراء ذلك ، أما فى حالة الاختبارات الصفية التى يكتبها المعلم لطلابه ، فإنه يستند إلى جدول مواصفات الاختبار الذى يقوم بإعداده كما سيتضح فى الفصل الخامس .

➢ الصدق المرتبط بمحك Criterion – related Validity :

تُستخدم كثير من الاختبارات بخلاف الاختبارات التحصيلية فى التنبؤ بسلوك مستقبلي معين للأفراد ، إذ ربما يود باحث تربوي أن يتنبأ بالأداء المدرسي للطلاب بمعلومية درجاتهم فى اختبار للذكاء ، أو يود باحث فى علم النفس الصناعي أن يتنبأ بآداء الأفراد فى عمل معين بمعلومية آدائهم فى بطارية من اختبارات الاستعدادات ، أو يود بأحدث فى علم النفس الكلينيكي أن يتنبأ بقابلية علاج مرضي معينين باستخدام المعلومات التى يتم جمعها باستخدام أسلوب المقابلة الكلينيكية ، أو مقياس مناسب للشخصية ، وهكذا.. ففي مثل هذه الحالات تكون وظيفة الاختبار أو المقياس التنبؤ بالأداء فى متغير يتضمن أداء يختلف كيفياً عن الأداء الذى يتطلبه الاختبار ، وهذا الأداء المتنبأ به يكون له أهمية عملية أو اجتماعية والمتغير المتنبأ به باستخدام درجات

الاختبار أو المقياس يسمي المحك Criterion ، لذلك يُطلق على هذا النوع من صدق الاختبارات والمقاييس " الصدق المرتبط بمحك " .

ونظراً لأن تقدير الصدق المرتبط بمحك دائماً جمع بيانات إمبريقية عن العلاقة بين درجات الاختبار ودرجات المقياس المحك Crierion Measure ، فإنه يشار أحياناً إلى هذا النوع من الصدق على أن الصدق الإمبريقي Empirical Validity والاهتمام الرئيس هنا ينصب على أداء الأفراد على المقياس المحك ، حيث إن درجات الاختبار المراد التحقق من صدقها ليست هى موضع التركيز ، وإنما تُستخدم فى التنبؤ بسلوك محك ذى أهمية ، لذلك فإن محتوى الاختبار لا يعد أمراً مهماً ، كما أن مفرداته لا ينبغي بالضرورة أن تكون لها علاقة واضحة بالمقياس المحك ، فما يهمنا هو أن يكون للاختبار صدق إمبريقي ، أي أن درجاته تتنبأ بالفعل بالأداء فى المحك.

ويمك التمييز بين نوعين من هذا الصدق : الصدق التلازمي Concurrent Validity والصدق التنبؤي Predictive Validity .

والفرق الرئيس بينهما يتعلق بالمدة الزمنية الفاصلة بين تطبيق الاختبار المراد التحقق من صدقه ، وتطبيق المحك ، فالصدق التلازمي يتطلب تطبيق كل من الاختبار والمقياس المحك فى نفس الوقت بعد انقضاء مدة زمنية مناسبة على تطبيق الاختبار ، فاختبار الشخصية متعددة الأوجه (MMPI) مثلاً يفيد فى تصنيف الاضطرابات العقلية ، لأنه قد تبين أن الأفراد الذين يتم تشخيص حالة الاضطراب العقلي لديهم بواسطة الطبيب النفسي (المحك) يميلون إلى نمط معين من الدرجات فى مجموعات معينة من مفردات أو فقرات هذا الاختبار ، ونظراً لأن العلاقة بين الاختبار والمحك مرتفعة ، فإن يمكن القول بأن اختبار الشخصية متعدد الأوجه يتميز بصدق تلازمي مرتفع.

أما اختبارات الذكاء والاستعدادات (الاختبارات المنبأة) ، فإنها غالباً ترتبط ارتباطاً مرتفعاً بعلامات الطلاب ، وتقديرات المعلمين ، ودرجات الاختبارات التحصيلية ، لذلك فإن يمكن تقدير الصدق التنبؤي لمثل هذه الاختبارات بإيجاد قيمة معامل الارتباط بين درجاتها (المحك) ، ودرجات الطلبة أو علاماتهم فى وقت لاحق ، غير أن قيمة هذا

المعامل لا تزيد فى كثير من الأحيان عن ٠.٦٠ ، مما يعني أن ٣٦% من تباين درجات أو علامات الطلاب يمكن التنبؤ بها من درجات الاختبار المحك .

لذلك ينبغي الحيطة عند استخدام الاختبارات والمقاييس فى التنبؤ بالسلوك المستقبلي ، كما ينبغي مراعاة أن التنبؤ يتعلق بمجموعة من الأفراد وليس بأداء فرد معين ، فمعامل الارتباط يعتمد على درجات اختبارين على مجموعة من الأفراد ، وبالتالي يشير الصدق التنبؤي إلى صدق مجموعة من الدرجات ، وليس لدرجة فرد معين ، وعلى الرغم من إمكانية التوصل إلى بعض الاستدلالات عن درجة الفرد ، إلا أن طريقة جمع البيانات المتعلقة بالصدق ، وأخطاء القياس التى ربما تؤثر فى كل درجة تقلل من الثقة فى الاستدلالات الفردية ، ولكنها يمكن أن تتميز بدرجة معقولة من الثقة إذا تعلقت بمجموعة الأفراد .

ويعتمد اختيار أي من نوعي الصدق المرتبط بمحك على ما إذا كنا نود التنبؤ بسلوك مستقبلي ، أو تقييم الوضع الراهن للسلوك ، وتستخدم العديد من المتغيرات كمقاييس محكة فى تقييم الصدق المرتبط بمحك كما أشرنا فى الأمثلة السابقة ، واختيار المقياس المحك يعتمد على القرار المراد اتخاذه ، ومجال السلوك المستقبلي موضع الاهتمام ، فالمقياس المحك فى المجال الصناعي أو مجالات الأعمال يختلف عنه فى المجال التربوي والكلينيكي .

غير إنه فى جميع الحالات ينبغي العناية باختيار المحك ، وأن تتوافر فيه خصائص معينة فالتحقيق من الصدق المرتبط بمحك يتطلب أن تكون العلاقة بين المحك والاختبار كبيرة ، لكي يتسنى اتخاذ قرار صائب ، فإذا لم يعكس المقياس المحك السلوك أو النتائج المرجوة ، فإن القرار الذى يستند إلى هذا النوع من الصدق يكون أقل فاعلية .

لذلك فإنه ينبغي التحقق من أن المقياس المحك يكون موائماً لطبيعة القرار المراد اتخاذه والسلوك المراد التنبؤ به ، وأن يؤدي إلى قياسات متسقة عبر الزمن واختلاف الظروف ويكون خالياً بقدر الإمكان من عوامل التحيز والذاتية ، بحيث لا تؤثر هذه العوامل فى درجات الأفراد .

➢ صدق التكوين الفرضي Construct Validity :

يتعلق صدق المحتوى ، وصدق التكوين الفرضي باستخدامات عملية مفيدة لنتائج الاختبارات والمقاييس ، فصدق المحتوي يفيد فى تحديد مدى تمثيل درجات الاختبار لتحصيل الطلاب فى مجال دراسي معين وما يتضمنه من نواتج تعليمية ، والصدق المرتبط بمحك يفيد فى تقدير أو التنبؤ بأداء حالي أو مستقبلي فى مجال أكاديمي أو مهني معين ، غير أنه في حالات أخري ربما نود تفسير درجات الاختبارات فى ضوء تكوينات فرضية Constructs ، أو سمات نفسيه Traits ، أو خصائص عامة Qualities مستمدة من نظرية سيكولوجية معينة مثل : القلق والاتجاه العلمي ، والذكاء ، والابتكارية ، والدافعية ، والتفكير الابداعي ، وغير ذلك ، فبدلاً من الاقتصار على معرفة درجة الفرد فى اختبار حسابي معين ، أو الاستناد إلى هذه الدرجة فى التنبؤ بنجاحه فى الرياضيات ، ربما نود الاستدلال من درجته على أنه يمتلك درجة معينة من القدرة الاستدلالية مثلاً ، وهذا يمكننا من وصف أداء الفرد وصفاً أكثر اتساعاً ، ويساعد فى العديد من الاستخدامات ، فالتكوين الفرضي إذن هو صفة أو خاصية نفسية نفترض وجودها ، وذلك لتفسير جوانب معينة من سلوك الأفراد ، ونوع الصدق المناسب فى هذه الحالة يسمي " صدق التكوين الفرضي " .

ويتطلب تقدير هذا النوع من الصدق دراسات متعددة وأساليب منطقية وإحصائية وتجريبية .

وقد أقترح كرونباك ، ومهيل (Cronback & Meehl, 1955) خمسة أنواع من الأدلة ينبغي جميعها لتأييد صدق التكوين الفرضي للاختبارات والمقاييس كالتالي :

١) الفروق بين مجموعات من الأفراد : فعينات من الأفراد الذين يفترض اختلافهم فى المتغير موضع البحث يمكن التنبؤ بتباين أدائهم ، والبيانات التى يتم جمعها تكون مستعرضة بدرجة أساسية .
٢) التغير فى الأداء : يمكن إجراء دراسات طولية لتعرف التغيرات عبر الزمن أو الناجمة عن اختلاف الظروف لنفس المجموعة من الأفراد .

٣) الارتباط : ينبغي أن ترتبط درجات اختبار معين بدرجات اختبار آخر يقيس المتغير نفسه ، بينما لا ترتبط درجاته بدرجات اختبارات أخري لا تقيس هذا المتغير ، ففي الحالة الأولي يمكن التوصل إلى الصدق التقاربى Convergent Valitdity ، وفى الحالة الثانية يمكن التوصل إلى الصدق التمايزي Discriminant Valitdity للاختبار ، وقد اقترح كامبل وفيسك (Campell & Fiske,1959) هذين النوعين من الصدق ، حيث يقدمان أدلة مهمة عن صدق التكوين الفرضي للاختبار ، ولعل هذا يتطلب استخدام أسلوب التحليل العاملي Factor Analysis لمصفوفة معاملات الارتباط بين درجات الاختبار ، ودرجات اختبارات ومتغيرات أخري موائمة .

٤) الاتساق الداخلي : ويقصد به الارتباط بين درجات مفردات الاختبار، أي درجة قياس المفردات للسمة نفسها.

٥) دراسات عملية لطريقة تناول الفرد للاختبار : ويقصد بذلك ماذا يفعل الفرد عندما يطبق عليه الاختبار ؟ وما العمليات العقلية التى تنطوي عليها استجابات الفرد لمفردات الاختيار ؟ وهل تختلف استجابات الفرد لصيغ متكافئة من الاختبار ؟ فهذه الأسئلة وغيرها ينبغي دراستها بأساليب تجريبية وإحصائية .

- **تعقيب عام على أنواع صدق الاختبارات والمقاييس :**

من هذا يتضح أن أنواع الصدق الثلاثة ليست مستقلة ؛ فالتحقق من صدق المحتوى ينبغي أن يسبق عادة الصدق المرتبط بمحك ، والتواصل إلى أدلة عن صدق التكوين الفرضي يتطلب جميع بيانات مستمدة من تلك المستخدمة فى التحقق من صدق المحتوي ، والصدق المرتبط بمحك ، كما أوضح " كرونباك (Cronbach, 1971,) " أيضاً أنه يمكن التحقق من صدق المحتوي إمبريقياً بإيجاد الارتباط بين درجات صيغتين من صيغ الاختبار نفسه إذا استند بناؤهما إلى نفس نطاق المحتوى Content Domain .

وكذلك يمكن ضم اجراءات تقدير صدق التكوين الفرضي ، وتقدير الثبات معاً للتوصل إلى أدلة عن صدق محتوى الاختبار .

وخلاصة القول أنه لا توجد طريقة واحدة للتحقق من صدق الاختبارات والمقاييس التربوية والنفسية ، وذلك لأن نوع الأدلة المرجوة يعتمد على كيفية استخدام هذه الاختبارات والمقاييس ، فالاختبار لا يكون صادقاً أو غير صادق بعامة ، وإنما يمكن تحديد ذلك فى ضوء أغراض بناء الاختبار واستخدامه ، ومن المعلوم أن الاختبار ينبغي أن تتميز درجاته بالثبات قبل الحكم على صدقه ، والصدق ليس خاصة من خصائص الاختبار على وجه الخصوص ، وإنما يتعلق بالاستدلال الذى يمكن التوصل إليه استناداً إلى درجات الاختبار ، لذلك فإن الاهتمام ينبغي أن يُوجه بدرجة أساسية إلى صدق التفسير المستمد من هذه الدرجات .

ونظراً لأن البيانات المستمدة من دراسات الصدق تتعلق بمواقف معينة ، فإن نتائج هذه الدراسات سوف تعتمد على خصائص هذه المواقف ، ويترتب على ذلك أن يكون للاختبار أكثر من صدق واحد بقدر المواقف ، التى يتم دراسة الصدق فى ضوئها ، وتعميم نتائج صدق الاختبارات والمقاييس على عينات مختلفة من الأفراد مستعدة من مجتمع معين يتطلب إجراء دراسات مستعرضة للصدقCross– Validation وذلك للتحقق من تماثل النتائج.

➢ ثالثاً : تحليل مفردات الاختبار وأهميته :

نظراً لأن نوعية الاختبار وجودته تعمد اعتماداً كبيراً على نوعية وجود المفردات التى يشتمل عليها ، فإن تحليل المفردات Item Analysis يعد أمراً ضرورياً لتحسين الاختبارات وبخاصة الاختبارات التحصيلية ، سواء المقننة أو الصفية التى يكتبها المعلم لطلابه ، وتحليل المفردات هو أسلوب منظم يصمم للحصول على معلومات محددة تتعلق بكل مفردة من مفردات الاختبار ، وهذه المعلومات يمكن الاستفادة منها فى تحديد المفردات الغامضة أو المريكة أو غير الفعالة من أجل مراجعتها أو استبعادها ، وانتقاء أفضل المفردات المتوافرة لتضمينها فى الصيغة النهائية للاختبار ، وكذلك يفيد تحليل إجابات الطلاب على كل مفردة من مفردات الاختبار فى تعرف صعوبات التعلم لديهم لمعالجتها .

ويتضمن تحليل المفردات أساليب احصائية لتقدير مستوى صعوبة المفردة ، ودرجة تمييزها بين الطلاب ، ونظراً لأن معظم هذه الأساليب تقتصر على المفردات ثنائية الدرجة مثل : مفردات الاختبار من متعدد ، والصواب أو الخطأ ، والمزاوجة ، فإنه يمكن أيضاً باستخدام أساليب تحليل المفردات تقييم فاعلية المشتتات Distractors ، أو بدائل الاختبارات التى تكون جزءاً من بعض هذه المفردات .

وعلى الرغم من أن هناك أساليب إحصائية معقدة تستخدم فى تحليل مفردات الاختبارات المقننة التى يقوم بأعدادها خبراء القياس والتقويم ، وتقويم مؤسسات متخصصة بنشر هذه الاختبارات إلا أنه توجد أساليب إحصائية بسيطة يمكن استخدامها فى تحليل مفردات الاختبارات التحصيلية الصفية التي يكتبها المعلم لطلابه ، وهو ما سوف أوضحه فى الجزء التالي ، ويمكن للقارئ المهتم بالاساليب المتقدمة ، وبخاصة التى تستند إلى نظرية الاستجابة للمفردة Item Response Theory (IRT) الرجوع إلى كتاب : القياس والتقويم التربوي والنفسي : أساسياته وتطبيقاته وتوجهاته المعاصرة ، وكتاب نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد ، وتطبيقاتها فى القياس النفسي والتربوي للمؤلف نفسه .

- **معامل صعوب المفردة :**

إن تحديد مستوى صعوبة كل مفردة من المفردات الاختبار يعد ضرورياً نظراً لأنه يبين للمعلم كيفية أداء الطالب فى المهمة التى تقيسها المفردة ، وكذلك المستوى العام لأداء طلبة صف معين فى كل مفردة من مفردات الاختبار ، وذلك يستطيع المعلم تحديد مدى تحقق الأهداف التعليمية التى تقيسها هذه المفردات ، كما أن معرفة مقدار معامل صعوبة المفردة يساعد فى تعرف المفردات التى تكون غاية فى الصعوبة أو السهولة بالنسبة لطلاب صف معين .

ويمكن التعبير عن صعوبة المفردة Item Difficulty بنسبة عدد الطلاب الذين أجابوا على المفردة إجابة صحيحة ، ومقدار هذه النسبة يسمي " معامل الصعوبة

Difficulty Index " فكلما زاد هذا المقدار دل ذلك على سهولة المفردة ، وكلما قل اعتبرت المفردة صعبة .

$$\text{أي أن معامل الصعوبة المفردة} = \frac{\text{عدد الطلاب الذين أجابوا على المفردة إجابة صحيحة}}{\text{العدد الكلي للطلاب الذين طبق عليهم الاختبار}}$$

..... (٤-٩)

فإذا كان عدد الطلاب الذين أجابوا على أحدى المفردات إجابة صحيحة (٢٤) مثلاً ، والعدد الكلي للطلاب (٤٠) ، فإن معامل صعوبة هذه المفردة يساوي $0.06 = \frac{24}{40}$ ، أي أن هذه المفردة تعد متوسط الصعوبة إلى حد ما بالنسبة لمجموعة الطلاب ومقدار معامل الصعوبة يتراوح بين (صفر -١).

ومن الجدير بالذكر أنه لا يوجد أتفاق بين خبراء القياس والتقويم التربوي على التوزيع الأمثل لمعاملات صعوبة اختبار معين ، إذا أن يعتمد جزئياً على نوع الاختبار المستخدم ، أي ما إذا كان الاختبار مرجعي المعيار NRT أم مرجعي المحك CRT فالاختبار من النوع الأول يهدف عادة للتمييز الأقصى بين المختبرين فيما يقيسه ، لذلك ربما يكون من المناسب أن تتراوح صعوبة مفرداته بين (٠.١٥ – ٠.٨٥) أي أن تتوزع قيم معامل الصعوبة توزيعاً متعادلاً على مدى السمة أو القدرة المقاسة ، ويرى بعض خبراء القياس أن الاختبار يمكن أن يميز تمييزاً أقصى بين الطلاب إذا كان متوسط مقدار معامل صعوبة مفرداته (٠.٥٠) ، أما إذا كان الاختبار من النوع الثاني الذى لا يهتم بالتمييز بين الأفراد ، وإنما تحديد مدى تحقيق كل طالب للأهداف التعليمية المتعلقة بمجال دراسي معين ، ففي هذه الحالة تقيس مفردات الاختبار هذه الأهداف قياساً مباشراً ، وتتحدد صعوبة المفردة وفقاً للمستوى المعرفي للهدف .

فإذا كان الهدف التعليمي من مستوى التذكر مثلاً ، فإن المفردة التى تقيس هذا الهدف تكون بالطبع أسهل من المفردة التي تقيس هدفاً تعليمياً من مستوى التحليل أو التقويم ، وفى هذه الحالة لا يكون ضرورياً الالتزام بأن يكون متوسط صعوبة المفردات (٠.٥٠) كما فى الحالة الأولي .

ويمكن القول بعامة أن مستوى صعوبة مفردة اختبار يعتمد على الهدف من استخدام نتائج الاختبار ، وطبيعة مجموعة الأفراد التى يطبق عليها الاختبار ، فمقدار معامل الصعوبة يعتمد على هذه المجموعة ، ويؤدي إلى تغير هذا المقدار نظراً لتغير قدرة أفراد المجموعة المراد اختبارها .

- **معامل تمييز المفردة :**

إن المفردات الاختبارية التي تتساوى فى مقدار صعوبتها لا تتميز بالضرورة بنس القدر بين المستويات التحصيلية المختلفة للطلاب فيما يقيسه الاختبار ، لذلك ينبغي إيجاد درجة تمييز كل مفردة للتحقق من فاعليتها فى التمييز بين الطلاب أقوياء التحصيل ، والطلاب ضعاف التحصيل .

فإذا وجد المعلم مثلاً أن نسبة الطلاب الذين أجابوا على مفردة معينة إجابة صحيحة فى مجموعة الطلاب من ذوي الدرجات الكلية المنخفضة فى الاختبار تساوي نسبة من أجابوا على المفردة نفسها إجابة صحيحة فى مجموعة الطلاب من ذوي الدرجات الكلية المرتفعة ، فإن مثل هذه المفردة لا تكون قد ميزت بين المجموعتين الطرفيتين وتحتاج بلا شك إلى مراجعة .

ويمكن الحصول على مقدار معامل تمييز المفردة Discrimination بإيجاد الفرق بين نسبة عدد طلاب المجموعة العليا (ذوي الدرجات الكلية المرتفعة فى الاختبار) الذين أجابوا على المفردة إجابة صحيحة ونسبة عدد طلاب المجموعة الدنيا (ذوي الدرجات الكلية المنخفضة فى الاختبار) الذين أجابوا على المفردة على المفردة إجابة صحيحة .

ولكن كيف يمكن تحديد المجموعتين الطرفيتين العليا والدنيا؟ لقد أوضح " كيلي Kelley " أن اختيار نسبة عدد الأفراد فى كل من هاتين المجموعتين بحيث تساوي (١٠%) مثلاً ربما يقلل من ثبات الدرجات ، وذلك لأن عدد أفراد كل من المجموعتين سوف يكون قليلاً ، وقد وجد أنه فى التوزيع الاعتدالي تكون النسبة الفاصلة المثلي لكل من المجموعتين هى (٢٧%) ، أما إذا كان عدد طلاب قليلاً ، فإن أخطاء المعاينات فى

معامل التمييز سوف تكون كبيرة نسبياً ، لذلك ينبغي الحيطة عند اختيار نسبة عدد الأفراد المجموعتين الطرفيتين ، وتري " أنستازي Anastasi " أن أي نسبة تتراوح بين (٢٥% و ٣٣%) سوف تكون مرضية .

ولإيجاد مقدار معامل تمييز مفردة ينبغي أولاً ترتيب أوراق الإجابة بحسب الدرجات الكلية ترتيباً تنازلياً ، ثم نفصل (٢٧%) من الأوراق التي حازت على أعلى الدرجات (المجموعة العليا) ، (٢٧%) التي حازت على أقل الدرجات (المجموعة الدنيا) ثم نتبق الخطوات التالية :

(١) إيجاد عدد طلاب كل من المجموعتين العليا والدنيا الذين أجابوا على كل بديل من بدائل مفردة الاختبار من متعدد ، ويمكن تكوين جدول تحليل المفردة كالتالي مثلاً :

الاجابة الصحيحة	رقم المفردة	المجموع	البـــدائل				
			أ	ب	جـ*	د	هـ
جـ	١	العليا	١	–	٨	–	١
		الدنيا	١	–	٦	٢	١

جدول (٤-١) يوضح نتائج تحليل مفردة اختبارية من نوع الاختبار من متعدد

ويلاحظ من الجدول أن :

* عدد طلاب المجموعة العليا (٢٧%) =١٠ .
* وعدد طلاب المجموعة الدنيا (٢٧%) = ١٠
* وعدد طلاب المجموعة العليا الذين أجابوا على المفردة إجابة صحيحة = ٨
* وعدد طلاب المجموعة الدنيا الذين أجابوا على المفردة إجابة بصحيحة = ٦

(٢) نوجد الفرق بين نسبة عدد طلاب المجموعة العليا الذين أجابوا على المفردة إجابة صحيحة ، وعدد طلاب المجموعة الدنيا الذين أجابوا على المفردة صحيحة للحصول على مقدار معامل التمييز .

ويتضح من الجدول أن :

معامل التمييز = $\frac{8}{10} - \frac{6}{10} = \frac{2}{10}$ = + 0.20

وعلى الرغم من أن المفردة تميز تمييزاً موجباً (+ 0.20) إلا أن مقدار معامل تمييزها منخفضاً ، مما يستدعي مراجعة البدائل المختلفة للإجابات فى المفردة .

- **تفسير معامل التميز :**

تقدم قيمة معامل التمييز للمعلم مؤشرات عن فاعلية كل مفردة فى الاختبار، وما تتضمنه من بدائل أو اختيارات ، وتتراوح هذه القيم بين (−1) ، (+1) .

فإذا كانت قيمة هذا المعامل (0.40) أو أكبر ، فإن هذا يكون دليلاً على أن المفردة تميز بين المجموعتين بدرجة جيدة .

أما إذا كانت هذه القيمة تتراوح بين (0.20 , 0.04) ، فإن التمييز يكون لا بأس به ، وإذا قلت عن (0.20) يكون التمييز ضعيفاً .

وإذا كانت قيمة التمييز صفراً ، فإن هذا يعني أن المفردة غاية فى السهولة أو الصعوبة ، ولا تميز بين مرتفعي ومنخفضي التحصيل ، فإذا كان الاختبار مرجعي المعايير (NRT) ويهدف للتمييز بين الطلاب فيما يقيسه ، فإن هذه المفردة يجب استبعادها من الاختبار ، أما إذا كان الاختبار مرجعي المحك (CRT) ويقيس تمكن الطلاب من مهارة أساسية ، فإن المفردة السهلة التى تمييزها صفراً لا ينبغي استبعادها ، لأنها تكون مناسبة للمستوى المعرفي للهدف الذى تقيسه ، واستطاع جميع الطلاب الإجابة عليها .

ومما لا شك فيه أن قيمة معامل التمييز تعتمد على درجة صعوبة المفردة الاختبارية ، فكلما ابتعدت قيمة معامل الصعوبة عن (0.50) تقل القيمة القصوى لمعامل التمييز ، لذلك فإنه إذا استخدم المعلم مفردات غاية فى السهولة أو الصعوبة بالنسبة لمجموعة معينة من الطلاب فإنه لا يجب أن يتوقع الحصول على قيم مرتفعة لمعاملات التمييز .

ولكن يجب أن نؤكد أنه إذا كان الهدف من الاختبار قياس إتقان الطلاب لمحتوى دراسي أو مهارة معينة ، فإن المعلم ربما يجد أن مقدار معامل الصعوبة (0.80) مثلاً ،

ومعامل التمييز (صفراً) ، ولا مانع من ذلك ما دام المعلم ليس مهتماً بإبراز الفروق الفردية بين الطلاب ، وإنما يهدف لقياس درجة تمكنهم من المحتوى أو المهارة المرجوة .

أدوات القياس التربوي والنفسي

* **الفصل الخامس : الاختبارات التحصيلية .**
* **الفصل السادس : اختبارات الذكاء العام .**
* **الفصل السابع : اختبارات الاستعدادات الخاصة .**
* **الفصل الثامن : اختبارات ومقاييس الشخصية .**
* **الفصل التاسع : اختيار وتقييم الاختبارات والمقاييس .**

الاختبارات التحصيلية

- **مقدمة**
- **مفهوم الاختبار التحصيلي**
- **الفروض الأساسية التى يرتكن إليها قياس التحصيل .**
- **تصنيف الاختبارات التحصيلية .**
- **استخدامات الاختبارات التحصيلية .**
- **الاختبارات التحصيلية الصفية التى يكتبها المعلم لطلابه**
- **مراحل بناء الاختبارات الصفية .**
- **الاختبارات التحصيلية المقننة .**
- **خصائص الاختبارات التحصيلية المقننة .**
- **بناء الاختبارات التحصيلية المقننة .**
- **الفروق بين الاختبارات التحصيلية المقننة، والاختبارات الصفية.**

➢ مقدمة :

عرضنا فى الباب الأول أهم الأسس المنهجية والإحصائية التى يستند إليها القياس النفسي والتربوي بعامة ، والاختبارات والمقاييس بخاصة ، وكذلك أخلاقيات استخدام الاختبارات ، وسوف أتناول فى هذا الباب أدوات القياس التربوي والنفسي المتعلقة بالتحصيل الدراسي ، والذكاء العام ، والاستعدادات الخاصة ، والشخصية الإنسانية ، وعلى الرغم من أن هذه الجوانب المعرفية وغير المعرفية أو الوجدانية تعد جوانب متكاملة للسلوك الإنساني ، إلا إنني سأتناول كل منها فى فصل مستقل لمزيد من التوضيح لهذه الجوانب أو السمات الإنسانية الرئيسية ، وما يتعلق بها من اختبارات ومقاييس متنوعة فى ضوء المفاهيم والمبادئ التى تناولتها بالتفصيل فى الفصول الأربعة السابقة ، وينتهي هذا الباب بفصل مستقل عن كيفية اختيار وتقييم الاختبارات والمقاييس التربوية والنفسية ، وسوف أستهل هذا الباب بالاختبارات التحصيلية لأنها أكثر شيوعاً واستخداماً فى المؤسسات التربوية والتدريبية .

ويتناول هذا الفصل مفهوم الاختبار التحصيلي ، والفروض الأساسية التى يتسند إليها قياس التحصيل ، وتصنيف الاختبارات التحصيلية واستخداماتها المتعددة ، ومراحل بناء الاختبارات التحصيلية الصفية والاختبارات التحصيلية المقننة ، والفروق بين نوعي الاختبارات .

➢ مفهوم الاختبار التحصيلي Achievement Test:

ويُستخدم مفهوم التحصيل Achievement للإشارة إلى درجة أو مستوى النجاح الذى يحذر الطالب فى مجال دراسي عام أو متخصص ، فهو يمثل اكتساب المعارف والمهارات والقدرة على استخدامها فى مواقف حالية أو مستقبلية ، ويعد التحصيل هو الناتج النهائي للتعلم ، ويتأثر مستوى التحصيل والأداء بعوامل متعددة توجد وقت التعلم ، كما يكون لها تأثير وسيط ما بين التعلم واستخدامات نواتجه ، ويستخدم المعلمون مفهوم التحصيل للإشارة إلى قدرة الطلاب على تحقيق الأهداف التعليمية للمواد الدراسية ، على الرغم من أن هذا يعد استخداماً محدوداً للمفهوم ، حيث إنه يمكن أن ينطبق على مختلف

مجالات السلوك الإنساني الفاعل (Bertrand & Cebula, 1980, Ward & Ward,1999) ، وهناك أدلة متعددة مستمدة من البحوث النفسية والتربوية تؤكد العلاقة الموجبة بين الذكاء والتحصيل ، كما أن الدافعية ومفهوم الذات ، والضبط الذاتي ، ومستوى طموح الطالب تؤثر فى مستوى وجودة تحصيله ، وكذلك تؤثر البيئة المنزلية والمدرسية، واتجاهات الآباء نحو إمكانات أبنائهم ، وتوقعات المعلمين عن طلابهم فى مستوى التحصيل لديهم ، وقد بينت دراسة " جوغ ,Gough, 1953 " وتيهان Teahan,1963 وإينز Innes,1972 وجنسن وفيورستاين Jensen & Feuerstein,1986 " أن السمات الشخصية المختلفة تؤثر تأثيراً واضحاً فى التحصيل الدراسي للطلاب فى مختلف مراحل التعليم ، كما أن طلاب المدارس الثانوية مرتفعي التحصيل يتسمون بالكفاءة العقلية ، والضبط الذاتي ، والثراء المعرفي ، والاستقلالية ، والاتساق ، والمثابرة ، ويؤثر المناخ النفسي الذى يوفره المعلمون فى سلوك الطلبة ، الذي يؤثر بدوره فى دافعيتهم ، ومفهوم الذات لديهم ، واتجاههم نحو المدرسة ، فإذا كان هذا المناخ إيجابياً وبناءً ، فإن يساعد فى تيسير تعلم الطلبة ورفع مستوى تحصيلهم ، ولكنه يعوق ويحد من التحصيل إذا كان سلبياً .

وربما يتسأل البعض عن كيفية اختلاف الاختبارات التحصيلية عن غيرها من الاختبارات النفسية والتربوية إذ ربما يعتقد البعض أن الاختبارات التحصيلية تقيس ما تعلمه الطالب فى المواد الدراسية ، وعلى الرغم من أن هذا أن هذا يبدو معقولاً ، إلا أن معظم الاختبارات التربوية والنفسية تقيس أيضاً ما تعلمه الفرد ، فأختبار قيادة السيارة مثلاً ، يعد اختباراً تحصيلياً على الرغم من أنه لا يرتبط بمادة دراسية معينة ، وربما يري البعض أن الاختبارات التحصيلية تتميز بأنها تتطلب عمليات معرفية للإجابة على مفرداتها ، مثل تذكر الحقائق والاستدلال ، وحل المشكلات وغير ذلك ، غير أنه إذا فحصنا اختبارين من الاختبارات التى تقيس الجوانب العقلية أو المعرفية أحدهما يقيس التحصيل والآخر يقيس الاستعداد الدراسي ربما يصعب التمييز بينهما تمييزاً واضحاً .

لذلك يرى " براون Brown,1970 " أن الاختبار التحصيلي يقيس التعلم الذى يحدث نتيجة للخبرة فى موقف تعلم يتم التحكم فيه ، مثل : صف مدرسي ، أو برنامج

تدريبي ، وأن الإطار المرجعي الذي نهتم به هو الحاضر أو الماضي ، أي الوضع الراهن لما تعلمه الطلاب .

➢ الفروض الأساسية التى يرتكن إليها قياس التحصيل :

فى ضوء تعريف مفهوم التحصيل الذى أوضحناه يتبين أن الاختبارات التحصيلية تصمم لقياس المعارف والمهارات التى اكتسبها الطلاب فى مجال محدود نسبياً ، وهذا المجال قد يكون محدوداً جداً ، مثل : تحصيل الطلاب لموضوع دراسي يتعلق بجغرافية الدول العربية ، أو أحدى العمليات الحسابية الأربع، أو يكون عريضاً ومتسعاً بحيث يشمل كل ما يدرس خلال العام الدارسي ، مثل : رياضيات المدرسة الثانوية ، إلا أننا فى الحالتين نحاول قياس ما يعرفه الطالب ، أو ما يمكنه أداؤه فى نهاية مدة دراسية معينة ، ونهتم بما تعلمه نتيجة لسلسلة معينة من الخبرات التى اكتسبها من دراسة منهج دراسي معين ، ونظراً لأنه لا يمكن أن نقيس جميع المهارات الممكنة ، فإن مفردات الاختبار تكون بمثابة عينة ممثلة لجميع المفردات المتعلقة بالمجال الدراسي المعين .

فإذا أخذنا بهذه النظرة عن مفهوم التحصيل ، فإننا فى الحقيقة نفترض بعض الافتراضات المتعلقة بطريقة القياس كالتالي (Anstasi,1976, Brown, 1970, Elliot, 1984,) :

(١) يمكن تحديد محتوى أو نطاق المعارف والمهارات التي يشتمل عليها الاختبار فى عبارات سلوكية ، وعلى الرغم من أن بعض المربين يرون أن هناك بعض الأهداف التربوية ربما يصعب تحديدها بطريقة سلوكية ، مثل :الاتجاه الإيجابي نحو التعلم ، أو التذوق الفني ، إلا أن الأهداف التربوية ينبغي أن تحدد التوقعات أو المستويات Standards المرجوة للمادة الدراسية لكي نتمكن من قياسها .

(٢) يقيس الاختبار الأهداف أو التوقعات أو المستويات المهمة ، ولا يقيس أهدافاً غير ضرورية ، وهذا الإفتراض يتعلق بصدق محتوى الاختبار Content Validity الذى أوضحناه فى الفصل السابق ، فمثلاً لكي يحل الطالب مسألة جبرية ، فإنه يحتاج إلى عدد من المهارات ، مثل : قراءة المسألة وفهمها ، وترجمتها إلى صيغة جبرية ،

والقيام بالعمليات الحسابية المطلوبة ، فإذا كان الاختبار قد صُمم لقياس قدرة الطالب على حل المسائل الجبرية الكلامية ، فعندئذ يكون الاختبار قد صُمم لقياس الغرض المطلوب .

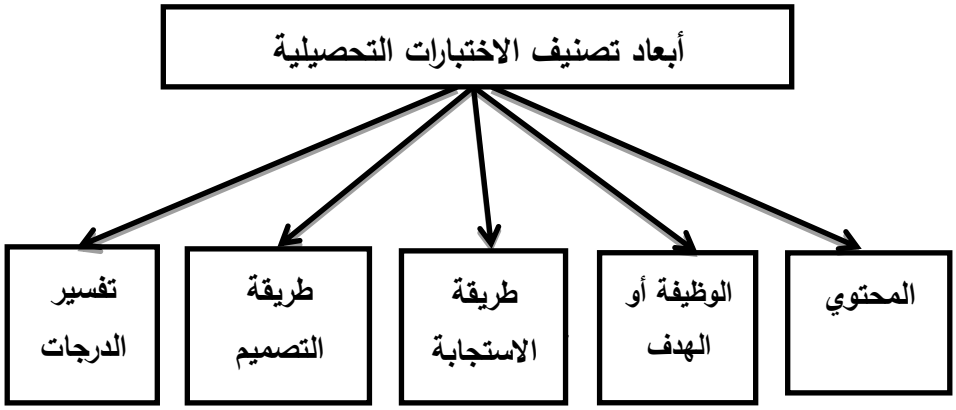
أما إذا كنا نهتم فقط بقدرة الطالب على استخدام الصيغ الجبرية ، فإن الاختبار السابق يكون مقياساً لقدرة الطالب على القراءة ، ويصبح الاختبار لا يقيس الغرض الذي صُمم من أجله .

(٣) توفر فرصة كافية للطلاب الذين سيطبق عليهم الاختبار لتعلم المادة التي سوف يشتمل عليها الاختبار ويقيسها ، وبتعبير آخر ، إذا أردنا المقارنة بين الطلاب أو التحقق من اتقانهم للأهداف أو المستويات التعليمية المرجوة فإنه يجب أن توفر لهم الفرص المتكافئة للتعلم ، وهذه الفرص تتحقق بدرجة معقولة فى الاختبارات الصفية التى يكتبها المعلم لطلابه ، لأن جميع الطلاب صف مدرس معين يدرسون المنهج نفسه ، ويحدد لهم المعلم محتوى وأهداف الاختبار .

إلا أن تكافؤ الخبرات التعليمية تعد مشكلة فى حالة الاختبارات التحصيلية المقننة التي سأوضحها فى هذا الفصل ، حيث يقيس الاختبار محتوى تعليمي ربما يختلف من مدرسة إلى أخرى ، وتتباين في كل منها خبرات الطلاب ، أى تكون خبراتهم غير متكافئة مع خبرات طلاب المرجعية Reference Group التي تم تقنين الاختبار عليها.

- **تصنيف الاختبارات التحصيلية :**

يمكن تصنيف الاختبارات التحصيلية وفقاً لعدد كبير من الأبعاد ، لعل أهمها ما يوضحه الشكل التخطيطي (٥-١) التالي



شكل (٥-١) يوضح الأبعاد المختلفة لتصنيف الاختبارات التحصيلية

وفيما يلي توضيحاً لكل من هذه الابعاد :

*** المحتوى Content :**

تختلف الاختبارات التحصيلية فى محتواها : فبعض الاختبارات – كما سبق أن ذكرنا – يكون محدود المحتوى ؛ مثل : قياس حالة الطقس ، أو تكون متسعة وعريضة فى محتواها ، مثل : مهارات القراءة، أو حل المشكلات ، وبعض الاختبارات التحصيلية يكون على شكل بطاريات Batteries of Tests تقيس مواداً مختلفة بدلاً من مادة واحدة ، مثل اختبار أيوا للمهارات الأساسية (ITBS) ، واختبارات كاليفورنيا التحصيلية (CAT) والاختبارات المتتابعة للتقدم التربوي (STEP) وغيرها، فهذه الاختبارات تركز على مهارات تربوية مثل : القراءة ، والحساب ، واستخدام اللغة ، ومهارات الاستذكار المتعلقة بقراءة الخرائط والرسوم التوضيحية والأشكال البيانية ، واستخدام المصادر المتنوعة ، وغير ذلك .

وهنا يحسن تصنيف الاختبار على أساس الفرقة الدراسية ، أي بطارية اختبار تحصيلي للمرحلة الأولي ، أو المرحلة الإعدادية أو الثانوية .

*** وظيفة أو هدف الاختبار Purpose :**

يمكن تصنيف الاختبارات وفقاً للأغراض التى يُستخدم فيها وهي : الاختبارات المسحية Survey Tests ، والاختبارات التشخيصية Diagnostic Tests ، واختبارات الاستعداد أو المقدرة Readiness Tests ، فالاختبار التحصيلي يعد اختباراً مسحياً إذا استُخدم فى تقدير المستوى العام لأداء الفرد فى مادة دراسية أو محتوى دراسي معين ، لذلك يمكن استخدامه في إعطاء صورة تقريبية عن مكانة الفرد بالنسبة لأقرانه ، فيما يقيسه الاختبار ، أي مقارنة أداء الطالب بأداء أقرانه ، ويمكن أن يكون الاختبار المسحي

على شكل بطارية من الاختبارات المسحية الفرعية التى يتعلق كل منها بمادة دراسية مختلفة تناسب فرقة معينة Achievement Batteries ، ومن أمثلة ذلك اختبار ستانفورد التحصيلي Stanford Achievement Test,(SAT) الذى أجريت عليه تغييرات كثيرة فى طبعاته المتتالية ، وهذا الاختبار يتناسب صفوف المرحلتين الابتدائية والمتوسطة ، ويشتمل على (٦) بطاريات يحتاج تطبيق كل منها حوالي (٥) ساعات يمكن تقسيمها إلى جلسات متعددة وتتكون الاختبارات الفرعية للبطارية من صيغتين متكافئتين ، ويوجد دليل إرشادي للتدرب على الاختبارات التى تشمل عليها قبل التطبيق الفعلي لأي منها ، فمثلاً بطارية المرحلة المتوسطة تشتمل علىل (١١) اختباراً فرعياً تتعلق بالمفردات اللغوية ، وفهم المقروء ، ومهارات التعامل مع الكلمات ، ومفاهيم رياضية ، وتطبيقات رياضية، وتهجي كلمات ، واستخدام القواعد اللغوية والمهارات المستخدمة فى العلوم الاجتماعية ، مثل (مهارات الاستدلال المتعلقة بالجغرافيا ، والتاريخ والاقتصاد، والعلوم السياسية ، وعلم الأجناس البشرية ، وعلم الاجتماع) ، ومفاهيم علمية ، وفهم الموضوعات التى تقدم شفوياً ، وجميع مفردات هذه الاختبارات من نوع الاختبار من متعدد ، كما يمكن تفسير الدرجات استناداً إلى الأهداف التعليمية ، التى تقيسها كل مفردة أو مجموعة من المفردات ، وكذلك يمكن تفسيرها أستناداً إلى معايير مئينية وتساعيات معيارية تم أعدادها لكل اختبار ، وقد سبق توضيح هذه المعايير فى الفصل الثالث وبذلك تسمح هذه البطاريات من الاختبارات بالتفسير مرجعي المحك (CRT) ، والتفسير مرجعي المعيار (NRT).

وقد تم تقنين هذه البطاريات فيما بعد على عينة قومية فى مدارس الولايات المتحدة الأمريكية قوامها ٢٧٥.٠٠٠ طالب ممثلة لكل شرائح المجتمع ، وأصبح لهذه البطاريات معايير قومية National Norms .

ونظراً لأن الاختبار المسحي يكون محتواه متسعاً ، فإنه يصعب بأستخدامه إعطاء صورة صادقة عن جوانب القوة والضعف فى تحصيل الطلاب.

وعلى العكس فإن الاختبار التشخيصي يقيس التحصيل بدرجة أكثر تفصيلاً وعمقاً نظراً لتركيزه على قياس مكونات المادة الدراسية ، ويُستخدم فى تحديد جوانب القوة والضعف فى تحصيل الطلاب ، وينقسم عادة إلى عدد من الاختبارات الفرعية ، وتقدر درجة لكل طالب فى كل اختبار فرعي ، ولذلك فهو يكون أطول من الاختبار المسحي ، فمثلاً الاختبار التشخيصي فى القراءة ربما يشتمل على اختبارات فرعية أحدها يقيس تعرف الكلمات ، وآخر يقيس فهم معاني الكلمات ، وثالث يقيس سرعة القراءة ، ورابع يقيس فهم الموضوع المقروء ، وغيرها ، وهذا يتطلب تحليل مكونات مهارة القراءة إلى عناصرها الرئيسية والفرعية التى يتم قياسها ويُستخدم الاختبار التشخيصي عادة بعد ما يتضح من تطبيق الاختبار المسحي ضعف مستوى أداء الطلاب ، وبذلك نستطيع عن طريق استخدام الاختبار التشخيصي التحديد الدقيق لجوانب الضعف ، وتقدم النوع المناسب من التعلم العلاجي .

ويوجد العديد من الاختبارات التشخيصية المنشورة باللغة الانجليزية مثل : اختبار ستانفورد التشخيصي فى القراءة (SDRT) ، وكذلك فى الرياضيات (SDMT) ، واختبار كاليفورنيا التشخيصي فى الحساب (CDAT) ، وغيرها، وهذه الاختبارات تناسب الطلاب فى مختلف مراحل التعليم العام ، وقد أعد مؤلف هذا الكتاب مجموعة من الاختبارات التشخيصية التى تقيس مهارات أساسية مختلفة لدى طلاب بعض الكليات (علام ، ٢٠٠٢ ، ٢٠٠٤) .

أما اختبار المقدرة فهو يبين ما إذا كان الطالب لديه المهارة المطلوبة لتعلم المادة التعليمية اللاحقة ، وبتركيز اختبار المقدرة على المهارات المهمة اللازمة للتعلم اللاحق ، فإن اختبار المقدرة يشبه فى ذلك الاختبار التشخيصي ، وباستخدامه للتنبؤ بالأداء اللاحق ، فإنه يشبه اختبار الاستعداد الدراسي .

وأفضل مثال لهذا النوع من الاختبارات هو اختبار الاستعداد للقراءة Reading Readiness الذى يطبق على أطفال الحضانة الذين يزعمون الالتحاق بالصف الأول

الابتدائي ويُستخدم هذا الاختبار فى تحديد ما إذا كان الطفل قد تعلم المهارات الضرورية للبدء فى تعلم القراءة .

*** طريقة الاستجابة Response Type:**

يمكن أيضاً تصنيف الاختبارات التحصيلية وفقاً لطريقة الاستجابة لمفرداتها أو المهام التى تشتمل عليها ، فعلي الرغم من أن كثيراً من الاختبارات التحصيلية تعد من نوع الورقة والقلم Paper– and – Pencil Test ، إلا أنه توجد اختبارات تتطلب مهارات أدائية مختلفة Performance Tests مثل : اختبارات الأداء الموسيقي ، والاختبارات العملية ، واختبارات مهارات تشغيل أجهزة معينة ، وغير ذلك ، ولعل هذا النوع من الاختبارات يناسب تقويم الأداء Performance Assessments الذى اصبح ينادي به خبراء القياس والتقويم فى وقتنا الحاضر ، فى إطار ما يسمي بالتقويم البديل أو التقويم الواقعي Authentic Assessments ، ويمكن للقارئ الرجوع غلى كتاب " التقويم التربوي البديل : أسسه النظرية والمنهجية وتطبيقاته الميدانية للمؤلف ، وذلك للاطلاع على تفاصيل هذا المجال المستحدث من مجالات القياس والتقويم التربوي المعاصر .

وتختلف اختبارات الورقة والقلم باختلاف أنواع مفرداتها ، إذ يمكن أن تشتمل على مفردات تتطلب التعرف من بين بدائل اجابات معطاة ، مثل : مفردات الاختبار من متعدد ، والصواب أو الخطأ ، والمزاوجة ، وكذلك يمكن أن تشتمل على مفردات أو أسئلة تتطلب إجابات مفتوحة دون تقديم بدائل إجابات معينة ، حيث يجيب الطالب بما يراه مناسباً للسؤال ، مثل أسئلة المقال بأنواعها المختلفة.

*** طريقة تصميم وبناء الاختبار Test Design:**

يمكن أيضاً تصنيف الاختبارات التحصيلية إلى نوعين رئيسين : الاختبارات المقننة Standardized Tests ، والاختبارات الصفية التي يكتبها المعلم Teacher –Made Tests ، فالاختبار التحصيلي المقنن يعد بواسطة خبراء القياس ومؤسسات نشر الاختبارات ، ويُستخدم فى مختلف المدارس ، ويختلف عن الاختبار الذى يكتبه المعلم

فى أن محتواه يكون متسعاً بحيث يتفق مع الأهداف العريضة للمادة الدراسية ، وبذلك يقيس معلومات اتفق خبراء التربية على أنها ضرورية لمتوسط الأداء فى مادة دراسية معينة أو فرقة معينة ، ويُطبق الاختبار المقنن على عينة كبيرة نسبياً من الطلاب خصائصها معلومة ويتم انتقاؤها بطريقة عشوائية ، وذلك لإيجاد معايير الاختبار استناداً إلى متوسط أداء هذه العينة ، وتصبح هذه المعايير الوسيلة التى يقارن فى ضوئها أداء كل طالب بمتوسط أداء هذه العينة ، أما الاختبارات التى يكتبها المعلم لطلابه فلا تتطلب هذه المعايير ، نظراً لأنها تتعلق بالطلاب الذين يقوم المعلم بالتدريس لهم ، وتقيس الأهداف التعليمية المباشرة لما تم تعلمه ، لذلك فإن محتواها يكون أكثر تحديداً ، وسوف اتناول كل من هذين النوعين الرئيسيين من الاختبارات التحصيلية بعد قليل .

*** تفسير درجات الاختبار Score Interpretation:**

يمكن تصنيف الاختبارات التحصيلية فى ضوء مرجعية تفسير درجاتها ، فإذا كان الاهتمام منصباً على تمييز الفروق الفردية بين الطلاب فى مجال دراسي معين استناداً إلى درجات الاختبار ، فإن الاختبار يسمي عندئذ الاختبار مرجعي المعيار Norm– Referenced Test (NRT) ، أما إذا أنصب الاهتمام على معرفة مدى أتقان الطالب لمهارات معينة محددة فى إطار محتوى دراسي معين ، فإننا نقارن درجة الفرد بمحك أو مستوى أداء سبق تحديده بدلاً من مقارنته بأقرانه، ويسمى الاختبار فى هذه الحالة " الاختبار المرجعي المحك Crierion –Referenced Test (CRT) ، وقد سبق أن أشرنا إلى هذين النوعين من الاختبارات فى بعض المواضع فى الباب الأول .

فالاختبار مرجعي المعيار هو ذلك الاختبار الذى يُستخدم فى تقييم أداء الفرد بالنسبة لأداء الأفراد الآخرين فى القدرة أو السمة التي يقيسها الاختبار ، والهدف من ذلك ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم فى الاختبار ، ويكون للدرجة تفسيراً فقط فى ضوء معيار الجماعة Test Norm أي متوسط أداء الجماعة فى الاختبار .

وتصبح الدرجة لا معني لها ويصعب تفسيرها إذا أغفلنا رد هذه الدرجة إلى هذا المعيار وتعد الاختبارات المقننة أو المسحية التى سبق ذكرها من هذا النوع.

أما الاختبار مرجعي المحك فيعرُف بأنه ذلك الاختبار الذى يُستخدم فى تقييم أداء الفرد بالنسبة إلى محك أو مستوى أداء مطلق دون الحاجة إلى مقارنة بأداء الأفراد الآخرين .

ويمدنا هذا النوع من الاختبارات بمعلومات محددة وتفصيلية عن تحصيل كل طالب فى موضوع دراسي معين أو مهارة معينة ، ودرجة الفرد فى الاختبار مرجعي المحك تعبر عن مدى تحقيقه للأهداف التعليمية المرجوة ، لذلك فإن هذا النوع من الاختبارات له قيمة تشخيصية لتحديد جوانب القوة والضعف فى تحصيل كل طالب لمجموعة محددة من المعارف والمهارات وبذلك يكون أقرب إلى الاختبارات التى يكتبها المعلم لطلابه .

- **استخدامات الاختبارات التحصيلية :**

تُستخدم الاختبارات التحصيلية استخدامات متعددة لعل أهمها أنها تحدد ما يعرفه الطالب وما أكتسبه من مهارات نتيجة تعلمه لموضوع أو مجال دراسي معين، غير أنه يمكن الإفادة من نتائج هذه الاختبارات فى انتقاء الطلاب للدراسة فى برامج معينة مناسبة لهم، وإخبارهم بإنجازاتهم ، وزيادة دافعيتهم ، وإثراء تعلمهم ، وكذلك فى تزويد المعلمين والمسئولين على بناء المناهج بمعلومات يمكن الاستناد إليها فى تطوير هذه المناهج أو تعديلها ، وتقييم البرامج التعليمية ، كما تُستخدم الاختبارات التحصيلية فى التقويم القومي لإجراء مسوح للمعارف والمهارات التى اكتسبها الطلاب من البرامج التعليمية المدرسية فى مختلف المجالات الدراسية ، وذلك بغرض مراقبة التقدم التربوي على مستوى الوطن ، وكذلك تُستخدم هذه الاختبارات لأغراض المساءلة التربوية Accountability للمناطق التعليمية والمعلمين عما تعلمه الطلاب ، وفى المنافسات الدولية فى الرياضيات والعلوم ، وغير ذلك ، وسوف أوضح فيما يلي كلاً من هذه الاستخدامات بشئ من التفصيل:

- **إثراء تعلم الطلاب :**

تلعب الاختبارات التحصيلية وبخاصة الاختبارات الصفية التي يكتبها المعلم لطلابه ، دوراً رئيساً فى عمليتي التعليم والتعلم ، فبعض المعلمين يظن أن هذه الاختبارات تطبق عقب تدريس مقرر معين بغرض معين تقدير درجات كلية لطلاب فى هذا المقرر ، وعلى الرغم من أهمية بل وضرورة ذلك ، إلا أن هذا يعد أحد أدوار الاختبارات التحصيلية ، فالهدف الرئيس لهذه الاختبارات هو تحسين عملية التعليم ، وإثراء عملية التعلم ، من خلال التغذية الراجعة التى تقدمها للمعلمين والطلاب فيما يتعلق بفاعلية عملية التعلم ، وجودة عملية التعلم ، إذ يمكن باستخدام نتائج الاختبار تحديد مدى واقعية وملاءمة الأهداف التعليمية للمواد الدراسية ، ومدى مناسبة طرائق التدريس وتقنياته لمستوى الطلاب ، فإذا بينت النتائج أن أداء الطلاب فى الاختبار كان ضعيفاً ، فإن هذا ربما يدل على أن طريقة التدريس غير مناسبة لهم ، أو أن الاختبار لا يعكس الأهداف التعليمية المرجوة ، أو لا يشتمل على عينة من المفردات ممثلة لهذه الأهداف أو أن أنواع المفردات المستخدمة لا تناسب طبيعة النواتج المحددة .

ولكي تكون التغذية الراجعة فاعلة ، ينبغي أن تكون فورية ومستمرة أثناء عمليتي التعليم والتعلم ، وأن تفيد الطالب فى معرفة اجاباته الصحيحة ، وإجاباته الخطأ ، لكي يتمكن من تصحيح أخطائه وإثراء تعلمه .

- **استثارة دافعية التعلم :**

فالاختبارات التحصيلية تقيس النواتج التعليمية التى ينبغي أن تكون واضحة للطلاب فى مستهل عملية التعلم لكي يبذلوا أقصي جهد لتحقيقها ، وعندما تُستخدم هذه الاختبارات فى إخبار الطلاب عما حققوه وما لم يحققوه أثناء تعلمهم ، فإنه يزيد دافعيتهم للعمل والانجاز ، كما أن هذه الاختبارات إذا ما أعتُني بتصميمها وبناؤها بحيث تقيس مهارات معرفية عليا كالتطبيق والتحليل والإبداع ، واهتم المعلم بتنمية هذه المهارات المهمة لدى طلابه ، فإن الطلاب سوف يبذلون قصاري جهدهم لتعلمها لكي يحصلوا على درجات مرتفعة فى الاختبار ، ومناقشة نتائج الاختبار مع الطلاب تعد بمثابة تغذية

راجعة توضح لهم طبيعة هذه المهارات وأهميتها ، وكيفية تناول الأسئلة المتعلقة بها ، لذلك فإن الاختبارات التحصيلية التى يُتني ببنائها يمكن أن تستثير دافعية الطلاب للتحصيل الجيد ، وتحقيق الأهداف التعليمية بمختلف مستوياتها ، وهذا يساعد فى استبقاء المعلومات وتوظيفها وانتقال أثرها تعلمها إلى مواقف خارج حدود المنهج المدرسي ، كما يساعد فى تأكيد أهمية التقويم الذاتي الذى يعد مكونة أساسية من مكونات التعلم المستقل للطالب وزيادة دافعيته ، وتقديره لذاته ، وبذلك يعتمد الطالب على الضبط والتحكم غير المباشر ، والدافعية الداخلية فى تحقيق الأهداف التعليمية المرجوة.

- **انتقاء الطلاب :**

تُستخدم الاختبارات التحصيلية فى انتقاء الطلاب للدراسة فى برامج تعليمية معينة مناسبة ، كما أن الطلاب ينتقلون من فرقة دراسية إلى أخري ، ومن مرحلة تعليمية إلى مرحلة أعلى بناء على نتائجها إضافة إلى نتائج اختبارات استعدادات ، لذلك ينبغى أن تتميز هذه الاختبارات بالصدق ، وتتميز درجاتها بالثبات .

وكذلك تُستخدم هذه الاختبارات فى تسكين الطلاب المختلفة بالكليات أو فى المستويات المختلفة لمقررات معينة ، وتُستخدم أيضاً فى تحديد ما إذا كان الطالب قد حقق مستوى الأداء المطلوب للقيام بأنشطة معينة .

- **تقويم المناهج والبرامج التعليمية :**

تُستخدم الاختبارات التحصيلية فى تقويم ملاءمة المناهج الدراسية ، وفاعلية البرامج التعليمية ، والمقارنة بين تحصيل الطلاب فى المدارس المختلفة ، غير أن هذه المقارنة تتطلب الحيطة فى تفسير نتائج الاختبارات التحصيلية وبخاصة التى تطبق على نطاق واسع .

- **المساءلة التربوية :**

تُستخدم درجات الطلاب فى الاختبارات التحصيلية أيضاً فى تقويم أداء المعلم ؛ فالمعلم يكون مسئولاً عن مدى نجاحه فى تعليم طلابه ، إذ يمكن تطبيق الاختبارات

التحصيلية نفسها أو صيغ مكافئة لها قبل قيامه بتدريس مقرر معين ، وعقب الانتهاء منه ، وذلك لتحديد مكتسبات الطلاب ، غير أن هذا الإجراء يشوبه بعض العيوب من الوجهة الإحصائية ، حيث إن مقدار ثبات درجات التحسن فى أداء الطلاب يكون عادة أقل من ثبات كل من درجات الاختبار القبلي والبعدي.

- **التقويم القومي :**

تُستخدم الاختبارات التحصيلية وغيرها من أدوات تقويم الأداء فى إجراء مسوح قومية لمنجزات الطلاب فى مختلف المناطق التعليمية ، والهدف من ذلك معرفة ما حققه طلاب الوطن من معارف ومهارات واتجاهات فى مواد دراسية متعددة مثل : العلوم ، والكتابة ، والمواطنة ، والقراءة ، والدراسات الاجتماعية ، والرياضيات ، والمواد التي تنمي المعارف المهنية ، والفنون ، وتمال ذلك التقويم القومي للتقدم التربوي فى الولايات المتحدة الأمريكية الذى يهدف للحصول على معلومات مهمة عن مدى تحقيق النظام التعليمي فى مراحل التعليم العام المختلفة للمستويات التربوية Education Sandards المتوقعة ، وتتبع التغيرات التى تحدث فى هذه المستويات أو النواتج التعليمية عبر الزمن على المستوى القومي ، والاستفادة من هذه المعلومات فى تحسين صنع القرارات التربوية ، وتطوير العملية التعلمية فى واقع الصفوف الدراسية .

➢ الاختبارات التحصيلية الصفية التى يكتبها المعلم لطلابه – Teache Made Classrom Tests :

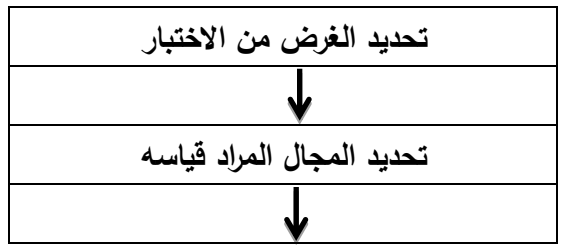
يهتم المعلمون بتقويم طلابهم بصفة دورية متكررة ، ويركزون عادة على تقويم أدائهم وتحصيلهم ، وذلك لأغراض تعليمية متنوعة كما أوضحنا فيما سبق ، ولعل أهم هذه الأغراض من الناحية التربوية هو معرفة مدى تحقيق كل طالب للأهداف أو المستويات المتوقعة Standards للمواد الدراسية المختلفة بغرض تبين مواطن القوة الضعف فى تحصيل الطلاب ، لكي يتسنى معالجة الصعوبات التى يواجهها معظمهم فى تحقيق هذه الأهداف أو المستويات ، ولذلك فإن الاختبارات الصفية التى يكتبها المعلم لطلابه تكون عادة مرجعية المحك Criterion Referenced، وتستند إلى المنهج

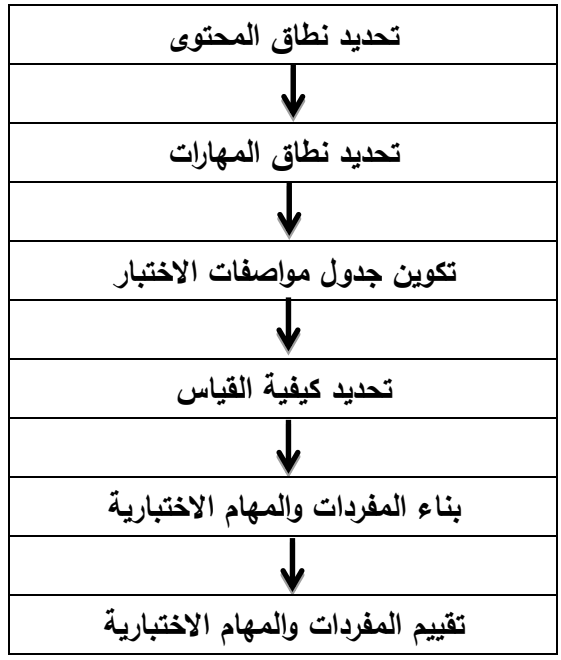
الدراسي Curriculum– Based Tests ، تناسب حاجات الطلاب الفعلية فى صف مدرسي معين ، وفى هذه الحالة لا ينبغي المعلم مقارنن أداء الطالب بأداء أقرانه فى الصف ، وإنما تعد الأهداف أو المستويات المتوقعة هى المحك Crierion الذى يتم تقييم الطالب فى ضوئه لتحديد درجة تحقيقه هذا المحك ، لذلك ينبغي أن يكون المحك واضحاً ومعرفاً تعريفاً دقيقاً ، ومحدداً قبل بدء التعليم ، وأن يكون الطلاب على دراية تامة بمتطلبات هذا المحك ، لكي يبذل كل منهم جهداً كافياً لتحقيقه .

- **مراحل بناء الاختبارات الصفية :**

يتطلب بناء الاختبارات التحصيلية الصفية إعداداً جيداً استناداً إلى خطة تفصيلية مدروسة تراعي الغرض من العملية الاختبارية ، وخصائص الاختبارات مرجعية المحك ، وطبيعة محتوى المادة الدراسية والمهارات المعرفية المتضمنة فى تعلم هذا المحتوى ، وأنسب المفردات والمهام الاختبارية التى تقيس هذه المهارات ، وهذا يحتاج من المعلم أن يكون متعمقاً فى مجاله الدراسي ، ومتمكناً من تفاصيله الدقيقة ، وأن يكون لديه معرفة كافية وتدرباً جيداً فيما يتعلق بكيفية بناء مختلف أنواع المفردات والمهام التى تقيس الأهداف المعرفية العليا ، صياغة الأفكار المتعلقة بالمثيرات التى يمكن أن تشتمل عليها هذه المفردات والمهام صياغة واضحة وموجزة ودقيقة ، أي ترجمة هذه المثيرات إلى موقف اختبارية مناسبة فى ضوء الهدف التعليمي والمحتوى الدراسي ، فبناء الاختبارات التحصيلية الصفية يعد علماً وفناً يستطيع معظم المعلمين اتقانه بدرجة معقولة من خلال التدريب المستمر ، والخبرة المتراكمة ، والممارسة الفعلية ٠

ويوضح الشكل التخطيطي (٥–٢) التالية المراحل الرئيسية فى إعداد هذه الاختبارات :





شكل (٥-٢) يوضح خطوات بناء الاختبارات التحصيلية الصفية

وسوف أوضح كل من هذه الخطوات بشئ من التفصيل فيما يلي :

أولاً – تحديد الغرض من الاختبار :

قبل البدء في تصميم وبناء الاختبار التحصيلي الصفي ينبغي على المعلم أن يحدد الغرض الرئيس من أجله يعد الاختبار ، وذلك هذا الغرض هو الذى يوجه جميع الخطوات التالية : فقد سبق أن أوضحنا أن الاختبارات التحصيلية الصفية تتعدد أغراضها وتتنوع استخداماتها بحسب أهداف عملية التقويم ، أي ما إذا كان التقويم قبلياً تمهيدياً لمعرفة المستوى المدخلي ، أو تشخيصياً لتعرف جوانب القوة والضعف فى أداء الطلاب أثناء عمليتي التعليم والتعلم ، أو ختامياً لوضع تقديرات للطلاب عقب الانتهاء من دراسة وحدة تعليمية أو مقرر بأكمله ، لذلك ينبغي على المعلم أن يحدد فى ضوء ذلك ما يلي :

(أ) الغرض من اختبار طلابه ومتى سيختبرهم :

(ب) قيم تُستخدم نتائج الاختبار ، ونوع القرارات التى ستتخذ فى ضوء هذه النتائج.

(ج) المرحلة التعليمية ، والصف الدراسي ، والمستوى المعمري للطلاب .

(د) المادة أو الوحدة الدراسية ، أو الموضوعات التى يود قياس التحصيل فيها.

(هـ) الزمن المتاح لتطبيق الاختبار ، وما تتطلبه الإجابة من أدوات أو مصادر.

فمما لا شك فيه أن تحديد هذه الأمور تحديداً واضحاً يُمكن المعلم من تصميم اختباره تصميماً مناسباً يحقق الغرض الرئيس منه ، ويفيد نتائجه الفائدة المرجوة .

ثانياً – تحديد المجال أو النطاق المراد قياسه :

إن تحديد المجال أو النطاق المراد قياسه يعد خطوة أساسية فى بناء الاختبار التحصيلي الصفي مهما اختلف الغرض منه ، فبعد أن يحدد المعلم الغرض الرئيس من الاختبار ، أي لماذا يختبر طلابه ، وفيم تُستخدم نتائج الاختبار ، ينبغي معرفة ماذا يقيس الاختبار .

فقياس التحصيل الدراسي شأنه شأن قياس الظواهر المختلفة يتطلب تحديد المتغير المراد قياسه ، نظراً لأن الاختبار باعتباره أداة قياس ما هو إلا ترجمة إجرائية لهذا المتغير ، وكلما كان التحديد دقيقاً لا يحتمل الجدل أو التأويل كان محتوى الاختبار أكثر مواءمة ووضوحاً وتعريف هذا المتغير يكون فى ضوء نطاق سلوكي Behavioral Domain معرفاً تعريفاً إجرائياً ، ويختلف خبراء القياس والتقويم فى تعريفهم لمفهوم النطاق Domain ؛ فمنهم من يرى أنه يحتوى الأهداف الضمنية أو الصريحة أو الموضوعات المراد قياسها ، أو أنه مخطط للمعارف والقدرات التي يشتمل عليها الاختبار ، وعدد المفردات التي تقيس أنماط المعارف المختلفة، وكيفية انتقاء النوع المناسب من المفردات والمهام الاختبارية، على أن يتضمن المخطط ما ينبغي إجراؤه على المعارف ، مثل : فهم المصطلحات والحقائق والتعميمات ، التنبؤ بالنواتج تقديم تفسيرات ، حل مشكلات، وهكذا ، ومنهم من يرى البدء بجدول مواصفات للاختبار كوسيلة تنظيمية لتصميم خطة بناء المفردات والمهام بحيث تضمن توازن المحتوى ، ومستويات السلوك أو العمليات المعرفية التى تتطلبها الإجابة على المفردات والمهام الاختبارية ، كما يرى بعض آخر أن تحديد النطاق السلوكي يتطلب إعداد جدول المواصفات ، إضافة إلى توصيف المفردات والمهام .

ويمكن القول بعامة أن النطاق Domain هو مصطلح مستمد من العلوم الفيزيائية ، ويقصد به فى هذه العلوم مجموع المعلومات التى يتوقع أن تفسيرها نظرية معينة ، وقد استخدم خبراء القياس هذا المصطلح للإشارة إلى النطاق الشامل للمحتوى Content Universe الذى يفترض أن الاختبار يشتمل على مفردات أو مهام تمثل عينة مستمدة من هذا النطاق الشامل ، وتتطلب معاينة النطاق Domain Sampling تحديد مكوناته وعناصره ، وهذا ما يجعل تحديد النطاق خطوة أساسية فى تصميم خطة الاختبار .

وعادة يحدد النطاق السلوكي للاختبارات التحصيلية الصفية بأهداف سلوكية أو نواتج متوقعة للمادة الدراسية ، وهذا يتطلب معرفة ما يجب تعلمه فى مادة دراسية معينة ، وكيف يتم تعليمها وتعلمها قبل البدء فى بناء الاختبار ، لذلك لا يجب أن تكون الأهداف التعليمية مجرد قوائم لمهارات غير منظمة ، ومهام بسيطة محدودة الفائدة ، والطريقة الأكثر فاعلية فى هذا الشأن هي البدء بالمفاهيم أو المهام الرئيسية للطلاب ، ومن ثم تحديد المحتوى والمهارات المتعلقة بهذه المفاهيم أو المهام (MSEB, 1991) .

ثالثاً – تحديد نطاق المحتوى :

عند تحليل نطاق المحتوى Content Domain يفضل الاسترشاد بأحد الأطر المرجعية التي حاولت تصنيف محتويات المواد الدراسية ونواتجها التعليمية ، مثل : تصنيف بلوم المعدل Bloom`s Taxonomy ، أو تصنيف مواري ، ووارد Murray & Ward Taxonomy ، أو غيرهما من التصنيفات ، ويمكن تلخيص ذلك فى جدول (٥-١) التالي (Ward & Ward, 1999).

تصنيف بلوم	تصنيف مواري ووارد
	العناصر
(١) النوعيات	(أ) المصطلحات
المصطلحات	(ب) الحقائق
الحقائق	(ج) الاتفاقات
	(د) الرموز
	(ه) المكونات
(٢) الطرق والأساليب	الطرق والأساليب (التفاعلات)
الاتفاقات	(و) النزعات
التصنيفات والأقسام	(ز) نتائج البحوث
النزعات والتتابعات	(ح) القواعد والقوانين

المنهجية	(ط) الأساليب
المحكات	(ك) المحكات
	العموميات
(٣) العموميات والتجريدات	(ل) المبادئ
المبادئ والتعميمات	(م) النظريات
النظريات والبنى	(ن) المعايير والاخلاقيات

جدول (٢-١) يوضح تصنيف المحتوى عند كل من بلوم ، وموارى ووارد.

وتحديد نطاق المحتوى ينبغي تصميم خطة المحتوى Content Outline التي تنظم المحتوى المراد قياسه في وحدات منطقية ، وتحديد اوزان او نسب مئوية لكل منها . ويمكن اتباع الخطوتين الرئيستين التاليتين في هذا الشأن :

(أ) تحديد المتغير المراد قياسه : ونقصد بالتغير مفهوم ، او المبدأ ، او الصيغة، او الخاصة المراد قياسها. ويمكن ان يكون المتغير متسعا، مثل: "المبادئ والقوانين العلمية "، أو محدودا مثل: "الاعداد المحصورة بين صفر، ١٠". وتعريف المتغير يعد بمثابة الاطار المرجعي الذى يستند اليه في تحديد نطاق المحتوي الذي سيتناوله الاختبار. وهذا يعتمد علي الهدف من استخدام نتائج الاختبار؛ فاذا كان الهدف توجيه مسار عملية التعليم، اي التقويم البنائي، فانه يفضل ان يكون المتغير محدودا، اما اذا كان الهدف هو التقويم الختامي ، فان هذا ربما يتطلب ان يكون المتغير متسعا. ويمكن الاعتماد في تحديد المتغير المراد قياسه علي مفردات المناهج ، او محتويات الكتب المدرسية ، او اي مصادر معلومات اخرى مناسبه .

(ب) تحديد المحتوي وتنظيمه : والهدف من ذلك التحقق من عدم وجود تداخل في محتوى ، حيث يمكن ان يتضمن موضوعا معينا جميع الموضوعات الاخرى. لذلك يفضل تحديد الموضوعات الرئيسة وما يندرج تحتها من موضوعات فرعيه مثل:

جغرافية أقطار الوطن العربي فى إفريقيا :

- جغرافية ليبيا والسودان .
- جغرافية أقطار المغرب العربي .
- جغرافية جمهوريتي موريتانيا الإسلامية والصومال .

كما يمكن الاسترشاد بالتصنيف الموضح بجدول (٥-١) السابق فى تحديد مكونات محتوى وحدة دراسية معينة ، مثل : العناصر وفرعياتها ، الطرق والأساليب وفرعياتها ، العموميات وفرعياتها ، وذلك فى ضوء الموضوعات التى تشتمل عليها هذه الوحدة ، ويمكن تنظيم المحتوى ومكوناته فى جدول مقسم بحسب عدد الموضوعات المعينة .

(ج) تحديد أوزان المحتوى : بعد تحديد الموضوعات وتنظيمها ، ينبغي تقدير أوزان لكل من الموضوعات الرئيسة والفرعية ، من حيث أهميتها النسبية فى الوحدة أو المادة الدراسية ، وهذا يساعد المعلم فى إمعان التفكير فى مكونات مادته الدراسية وعناصرها والمطلوب تعلمه ، والزمن الذى يستغرقه تدريس الموضوعات المختلفة ، والجهد اللازم لذلك ، ويمكن التعبير عن هذه الأوزان بنسب مئوية لكل من الموضوعات الرئيسة والفرعية .

رابعاً – تحديد نطاق المهارات :

بعد تحديد نطاق المحتوى تكون الخطوة التالية تحديد نطاق المهارات Skills Domain التى سوف يقيسها الاختبار فى ضوء هذا المحتوى ، وكذلك أوزانها النسبية .

وتحديد نطاق المهارات يتطلب معرفة العمليات العقلية التي يمكن أن يستخدمها الطالب في إجابة مفردة أو مهمة اختبارية معينة تتعلق بأحد المحتويات ، وتتباين الموضوعات فى عدد المهارات التى تتطلبها ومستوياتها المعرفية ، فبعضها ربما يتطلب مهارة واحدة من مستوى معرفي مرتفع ، والبعض الآخر ربما يتطلب عدة مهارات تختلف فى مستواها المعرفي ، وفى أي الحالات ينبغي صياغة المهارة صياغة واضحة ومحددة ، بحيث يمكن ملاحظة السلوك المتعلق بها ، وقياسه بمفردات أو مهام اختبارية .

ويمكن الاسترشاد أيضاً فى تحديد نطاق المهارات بالتصنيفات التى أشرنا إليها فيما سبق ، ويوضح جدول (٥-٢) التالي تصنيفات المهارات عند كل من بلوم Bloom ، ومواري ووارد) (Ward & Ward1999) .

تصنيف بلوم	تصنيف مواري ووارد
	العناصر
(١) المعرفة (التذكر	(١) التعرف
(٢) الفهم	(٢) الاستدلال
- الترجمة	- الترجمة (التحويل – التصنيف)
- التفسير	- التفسير (التلخيص – التعميم)
- الاستكمال	- الاستكمال (العلية – التنبؤ – التضمينات)
(٣) التطبيق	(٣) التحليل (المتشابهات ، المصفوفات ، التتابع ، المخططات ، الحقائق ، الافتراضات ، الفروض ، الآراء ، الاتساق الداخلي ، وجهات النظر)
(٤) التحليل	
(٥) التركيب	
(٦) التقويم	(٤) التقويم
	- المقارنة بمعايير محددة
	- المقارنة بمعايير غير محددة
	- الإسهام في غرض أو تأثير معين
	- الموازنة / المغايرة
	(٥) حل المشكلات
	- تحديد المشكلات
	- اختيار المبدأ المناسب
	- استخدام المبدأ الصحيح

جدول (٥-٢) يوضح تصنيف المهارات عند كل من بلوم ، ومواري ووارد.

ويعد تحديد نطاق المهارات أمراً مهماً لأنه يُستند إليه فى تحديد المستوى المعرفي أو المهارى للمفردات أو المهام الاختبارية ، لذلك ينبغي أن يكون السلوك المراد ملاحظته وقياسه لدى الطالب واضحاً لكي يتمكن المعلم من التحقق من أداء الطالب للمهارة المرجوة ، وهذا يتطلب تعريف المتغيرات أو المهارات فى ضوء الإجراءات التى سوف تُستخدم فى قياسها ولعل أفضل الطرق لتعريف المهارات تعريفاً إجرائياً هو إعداد مواصفات للمفردات والمهام التى سوف يستند إليها فى قياس هذه المهارات فى ضوء نطاق المحتوى الذى تم تحديده ، وهذا يعني تحديد المثيرات التى سوف تشتمل عليها المفردات أو المهام الاختبارية ، وخصائص الإجابة على كل منها .

ويمكن اتباع الخطوتين التاليتين لتحديد نطاق المهارات :

أ- **تحديد المهارات المتوقع ارتباطها بكل موضوع من موضوعات المحتوى** ، وتتطلب منه الخطوة وضوح نطاق كل من المحتوى والأهداف ، والاسترشاد فى ذلك بجدولي (٥-١) ، (٥-٢) السابقين وعندئذ يمكن تحديد أنواع المهارات ومستوياتها المناسبة لكل

موضوع فرعي ، حيث إن الموضوعات المختلفة ربما تتطلب مستويات مهارية متباينة ، فمهارة التحليل تتطلب مستويات أخري من المهارات يمكن قياسها ؛ فتحليل قطعة مقروءة ربما يتطلب معرفة الكلمات ، وفهم معناها ، والاستدلال ، والربط وإعادة الصياغة ، وغير ذلك .

ب- اختبار المهارات المناسبة : يمكن قياس مهارة واحدة فى موضوع معين ، كما يمكن قياس عدة مهارات فى الموضوع نفسه ، وهذا يعتمد على الهدف من الاختبار ، وتوقعات المعلم عن مستوى المهارات التي يرى ضرورة اكتسابها ، وعدد مفردات الاختبار ، والزمن المتاح للإجابة عليه ، وأنواع المفردات والمهام المناسبة للمحتوى والمهارات المراد قياسها .

خامساً – تكوين جدول مواصفات الاختبار :

بعد تحديد نطاق كل من المحتوى ، والمهارات ، والأوزان ، التى تدل على الأهمية النسبية لكل من الموضوعات الرئيسة والفرعية ، والمهارات ذات المستويات المختلفة ، فإن المعلم يحتاج إلى وسيلة تنظيمية تساعده فى ربط المهارات بالمحتوى بحيث يمكن تقدير أوزان تتناسب مع أهمية المهارة والمحتوى فى أن واحد لكي يضمن توازن جميع مستويات المهارات فى إطار محتوى المادة الدراسية، وتمثيلها فى الاختبار ، وعلى أساس هذه الأوزان يمكنه تحديد عدد المفردات والمهام الاختبارية المطلوب اعدادها ، وهذه الوسيلة التنظيمية تسمي " جدول مواصفات الاختبار Test Specification Table .

وهذا الجدول ثنائي البعد ، حيث يمثل بُعده الأفقي مستويات بُعده الرأسي أقسام المحتوى (الموضوعات الرئيسة والفرعية) ، ومن ثم تحديد فى خلايا هذا الجدول الأوزان المتعلقة بالأهمية النسبية لكل من المهارة والمحتوى باستخدام أوزان المحتوى ، وأوزان المهارة التى سبق تحديدها فى الخطوتين الثالثة والرابعة ، وباستخدام هذه الأوزان يمكن تحديد عدد المفردات والمهام التى يجب إعدادها لكل مهارة فى إطار كل موضوع من الموضوعات المحددة ، ويمكن إجراء ذلك بأن نضرب الوزن النسبي (النسبة المئوية) للعمود الذى يقع فيه خانة معينة فى الجدول × الوزن النسبي للصف الذى تقع فيه

الخانة نفسها ، ثم نضرب الناتج × عدد المفردات والمهام الكلية التى سوف يشتمل عليها الاختبار .

سادساً – تحديد كيفية القياس :

يتطلب قياس المهارات التى تم تحديدها فى ضوء نطاق المحتوى أنواعاً مختلفة من المفردات والمهام الاختبارية ، وكثير من المهارات تتطلب مستويات معرفية عليا ، وأداءات ، وعمليات ونتاجات ، لذلك فإنه ينبغي أن يكون المعلم على دراية بفنيات كتابة مختلف أنواع المفردات الاختبارية التى تتطلب الورقة والقلم ، وكذلك المهام الاختيارية الأخرى التى تتعلق بأساليب ملاحظة السلوك ، وموازين التقدير ، وملفات الأعمال (البورتفوليو) ، وتقييم النتاجات ، وغير ذلك. وقد سبق أن أوضحنا فى مستهل هذا الفصل تصنيف الاختبارات التحصيلية وفقاً لأنواع مفرداتها ومتطلبات إجابتها .

ويمكن القاء الضوء على بعض من هذه المفردات والمهام ، حيث إن كثيراً من المهارات العليا المهمة يصعب قياسها بمفردات اختيار من متعدد ، أو صواب أو خطأ ، أو مزاوجة ، أو إكمال فمثلاً مهارة كتابة تقرير علمي استناداً إلى بيانات احصائية يحتاج قياسها إلى سؤال مقال مفتوح وفق محكات معينة ، ومهارة الإلقاء أو التعبير الشفوي تتطلب اختباراً شفوياً ، ومهارة إجراء تجربة مختبرية تتطلب أداء عملياً ، ومهارة تصميم وتنفيذ مشروع فردي أو جماعي معين ويجب تحديد عدد هذه المجموعات قبل عملية التقدير للدرجات .

وإذا لم نستطيع تحديد المجموعة التى تنتمي إليها أحد أوراق الإجابة ، فإننا يجب إعادة قراءتها بعناية حتى يتسني لنا تصنيفها فى إحدى المجموعات ويجب التأكد من أن أفقر الإجابات فى المجموعة الممتازة تفوق أفضل الإجابات في المجموعة غير الجيدة .

ومن الأفضل بعد الانتهاء من عملية التصنيف المبدئي أن نعيد تصنيف أوراق الإجابة فى كل مجموعة للتأكد من تجانسها ، والهدف من إعادة التصنيف أن نجعل الفروق بين المجموعات أكبر ما يمكن ، والفروق داخل كل مجموعة أقل ما يمكن ومن ثم نضع تقديراً لكل مجموعة من هذه المجموعات مثل : ممتاز ، جيدجدا ، جيد ، مقبول

، راسب ، وهكذا .. على حسب النقط أو طريقة التصنيف ، فإنه يمكن زيادة موضوعية التصحيح إذا استبعدنا أسماء الطلاب من على أوراق الإجابة ، وقررنا كيفية تقدير الإجابات الصحيحة ولكن غير المتعلقة بموضوع السؤال وأيضاً الأخطاء اللغوية ، كما أنه من المفيد أن نكتب بعض التعليقات على الإجابات حتى يتسني للطلاب معرفة أخطائهم ، مما ييسر لهم عملية التعلم.

سابعاً – تقييم المفردات والمهام الاختبارية :

يتطلب الاختبار أو الأداء المستخدمة فى قياس تحصيل الطالب ونتائجه مراجعة وتدقيقاً بعد تطبيقه ، إذ يمكن أن يتبين من النتائج قصور بعض المفردات، حيث تُــترك بعض المفردات دون إجابة من ذوي التحصيل المرتفع ، فى حين يجيب عليها ذوي التحصيل المنخفض إجابة صحيحة ، لذلك ينبغي إجراء تحليل المفردات Item Analysis الذى سبق أن أوضحناه فى الفصل الرابع ، للتعرف على مثل هذه المفردات وغيرها التى تكون غامضة ، أو مربكة ، أو حدث خطأ فى تحديد إجابتها الصحيحة فى مفتاح التصحيح ، وغير ذلك ، وهذا يفيد المعلم فى إجراء التعديلات المناسبة على هذه المفردات ، وتجميعها فى قوائم تشتمل على المفردات ، وقيمة معامل صعوبة كل منها ومعامل تمييزها ، وتوجد برامج جاهزة متعددة يمكن استخدامها فى إجراء هذا التحليل بسرعة ودقة ، مثل برنامج MieroCat ، وبرنامج Iteman ، أعدتهما مؤسسة نظم التقييم (ASC) الأمريكية.

وبذلك يمكن أن يستخدم المعلم هذه المفردات مرات متعددة فى أعوام متتالية مما يوفر جهده ووقعته وتكريسهما لعملية تعليم طلابه .

- **الاختبارات التحصيلية المقننة**

: Standardized Achievement Tests

تُستخدم الاختبارات التحصيلية المقننة فى قياس درجة تعلم الطالب فى مجالات مشتركة للمناهج المدرسية ، فالاختبارات التي يكتبها المعلم لطلابه تقتصر على ما يقوم

بتدريسه وما يناسب حاجات طلابه ، أما الاختبارات التحصيلية المقننة فإنه يمكن تطبيقها على طلاب فرقة دراسية معينة فى مختلف المدارس فى ظروف مقننة منتظمة من حيث التطبيق ، والتصحيح ، وهذا يجعل من الممكن مقارنة مستوى تحصيل طلاب مجموعة معينة بمعيار الجماعة Reference Group ، فكما سبق أن ذكرنا فى الفصل الرابع أن التقنين Standardization هو الاسلوب المنتظم الذى نتبعه عند تطبيق الاختبار، وتصحيحه لكي تكون المقارنة عادلة، حيث تتوحد الشروط والظروف لجميع الطلاب الذين يطبق عليهم الاختبار، وهذا المطلب يعد شرطاً أساسياً لجميع المشاهدات العلمية .

لذلك يتضمن الاختبار تعليمات تفصيلية واضحه تُراعى عند تطبيق الاختبار ، لضمان انتظام الإجراءات والظروف التى يطبق فيها الاختبار ، وتعد صياغة هذه التعليمات جزءاً من عملية تقنين هذا النوع من الاختبارات .

وكذلك يجب أن تتضمن عملية التقنين جميع المواد الاختبارية المستخدمة ، وزمن الإجابة على الاختبار ، وأمثلة توضيحية ، وغير ذلك من تفاصيل الموقف الاختباري الذي يمكن أن يؤثر فى أداء الطلاب فى الاختبار ، وهذا التقنين يوفر للاختبار نوعاً من الموضوعية ، كما يضمن عدم تأثر الدرجات باختلاف المصححين .

ونحن نحتاج فى مجال التربية إلى معلومات نحصل عليها من عينة متسعة مقننة من أداء الطلاب ، ومثال ذلك : قياس النمو التعليمي من عام إلى آخر ، ومقارنة النتائج أو طرائق التدريس عبر المدراس ، ومدى انتقال أثر التعلم ، وتصنيف الطلاب ويصعب الحصول علي معلومات كافية في كل من هذه المواقف بإستخدام الاختبارات الصفية التى يكتبها المعلم لطلابه ، ومن هنا تبرز الحاجة إلى بناء الاختبارات التحصيلية المقننة التى يقوم بإعدادها خبراء متميزين فى القياس النفسي والتربوي ، وتقوم بنشر هذه الاختبارات مؤسسات نشر تجارية Test Publishers ، وتشارك فى إعدادها .

➢ خصائص الاختبارات التحصيلية المقننة:

(١) تشتمل هذه الاختبارات على مهام يشترك جميع الطلاب فى تعلمها فى جميع المدارس ، وتكون تعليماتها محددة لكي يبدأ الطلاب الإجابة عليها بفهم مشترك

فجميع الطلاب يجيبون على مهام تتطلب تذكر الحقائق ، وحل المشكلات ، والاستدلال مثلاً . وهذه المهام المشتركة تعد بمثابة الميزان الذي يجعل المقارنة بين المجموعات المختلفة من الطلاب ممكنة.

(٢) تشتمل هذه الاختبارات على معايير Norms لتفسير الدرجات المستمدة منها، وذلك لتحديد مركز الطالب بالنسبة لأقرانه فيما يقيسه الاختبار. وعادة يتم تحويل الدرجات الخام إلى نوع من الموازين المعيارية ، مثل : الرتب المئينية، والتساعيات المعيارية ، كما أوضحنا فى الفصل الرابع ، وهذا يسمح بترتيب الطالب بين أفراد مجموعة كبيرة من الطلاب الآخرين ، مما ييسر التعميم فيما يتعلق بالدرجات الناتجة .

(٣) يشتمل معظم هذه الاختبارات على صيغتين متكافئتين أو أكثر ، وتحتوي كل من هذه الصيغ على مجموعات متمايزة من المفردات ، ولكنها تعد بمثابة عينة ممثلة لنفس النطاق السلوكي ، ويكون متوسط درجات مجموعة من الطلاب فى إحدى هذه الصيغ مساوياً المتوسط فى صيغة أخري ، وكذلك يتساوى الانحراف المعياري لدرجاتهم في الصيغتين ، وهذا يساعد فى اختبار الطلاب قبل عملية التعليم وعقب الانتهاء منها دون أن تتأثر درجاتهم فى المرتين بعوامل الذاكرة ، وذلك لتعرف التغير الذي يحدث فى هذه الدرجات .

(٤) يُرفق مع كل من هذه الاختبارات دليل للاختبار Test Manual يُسترشد به فى كيفية تطبيق الاختبار وشروط ذلك ، وكيفية تفسير درجاته ، وهذا يساعد فى توحيد الإجراءات والتفسير .

(٥) تعد هذه الاختبارات من نوع الاختبارات مرجعية المعيار (NRTs) التى تهتم بمقارنة أداء الطالب بمحك أداء مطلق .

غير أن بعض الاختبارات التحصيلية المقننة يتم تصميمها وبناؤها بحيث تسمح بتفسير الدرجات استنادا إلى معايير ، وكذلك إلى محك نجاح أو مستوى أداء محدد مسبقاً فى ضوء الأهداف التعليمية المرجوة ، وهناك بعض الاختبارات الأجنبية المتوافرة التى تفي بذلك فى مجالات القراءة والحساب .

➢ بناء الاختبارات التحصيلية المقننة :

يقوم بتصميم وبناء الاختبارات التحصيلية المقننة مؤسسات متخصصة، وعادة تكون مؤسسات نشر الاختبارات التربوية والنفسية في الولايات المتحدة الامريكية، وتكون عملية بناء هذه الاختبارات تحت إشراف وتوجيه مجموعة استشاريين من خبراء القياس والتقويم واختصاصي المناهج، وهؤلاء يضعون الخطوط العريضة للاختبار فى ضوء الغرض من استخدامه ، أي ما إذا كان مسحياً أو تشخيصاً ، وكذلك يقومون بتحديد محتواه والمهارات المزمع قياسها، وأوزانها النسبية فى ضوء المحتوى ، والمستوى العمري الملائم لمن سيتطبق عليهم الاختبار ، وطول الاختبار ، والعلاقة بينه وبين غيره من الاختبارات التى ستتكون منها البطارية ، أو الاختبارات التى سبق إعدادها ونشرها ، ونوع المفردات التى ستُستخدم ، وعدد الاختبارات الفرعية التى سوف تشتمل عليها البطارية ، وحيث سيكون لكل اختبار فرعي درجة مستقلة .

وعلى الرغم من أن خطوات بناء الاختبار المقنن ربما تختلف قليلاً بحسب الغرض منه وطبيعته ، إلا أن هناك تسلسلاً معيناً ينبغي مراعاته يوضحه الشكل التخطيطي (٥-٥) التالي :

شكل (٥-٥) يوضح خطوات بناء الاختبارات التحصيلية المقننة

وفيما يلي توضيحاً لكل من هذه الخطوات :

إعداد خطة الاختبار :

الفرق الرئيس بين الاختبارات التحصيلية المقننة والاختبارات التى يكتبها المعلم لطلابه هو أن الاختبارات المقننة يقوم بإعداد خطتها مجموعة من الخبراء، وتعكس هذه الخطة آراء ووجهات نظر هذه المجموعة ، مما يساعد شفى تقليل عوامل التحيز ، ويكون عادة محتوى الاختبار عرضاَ وأكثر شمولاً ، كما يكون عادة جزءاً من بطارية اختبارات تناسب فرقة دراسية معينة ، ويكون الاختبار مناسباً للبطارية التى هو جزء منها .

وفيما عدا ذلك فإن خطة بناء الاختبار التحصيلي المقنن تماثل خطة بناء الاختبار الذى يكتبه المعلم ، ويبدأ إعداد الاختبار بالفحص الدقيق للمادة الدراسية، والكتب المقررة لتحديد الموضوعات التى سيتناولها الاختبار ، وربما يتطلب ذلك استشارة مجموعة من اختصاصي المادة الدراسية والاستنارة بمقترحاتهم وتوصياتهم فى هذا الشأن ، وتُحدد المهارات الرئيسة والفرعية المتعلقة بهذا المحتوى، ويتم إعداد جداول المواصفات التى يُعين فيها المحتوى والمهارات المطلوبة ، والأوزان النسبية لكل منها .

كتابة المفردات :

يقوم اختصاصيو المواد الدراسية – بالاستعانة بخبراء القياس – بكتابة مفردات الاختبار المناسبة طبقاً للخطة الموضوعة ، وتُراجع بواسطة مجموعة أخرى من الخبراء ، ومن ثم تُعاد صياغة المفردات ومراجعتها ، وتتكرر هذه العملية حتى يتم الاتفاق على صلاحية هذه المفردات وعادة يُبني عدد من المفردات أكبر مما هو مطلوب ، وذلك لأن بعض المفردات ربما يُستبعد لعدم صلاحيته ، ويعتمد عدد المفردات التي يتم كتابتها لكل موضوع على الأوزان المحددة فى جدول مواصفات الاختبار الذى تم إعداده ، وعلى عدد الصيغ المتكافئة للاختبار .

وعادة يُستخدم فى هذا النوع من الاختبارات مفردات اختبار من متعدد، وذلك لأنها أكثر تمثيلاً للمحتوى والمهارات المراد قياسها ، كما أن تصحيحها يكون أكثر سهولة وموضوعية غير أن استخدام مفردات أخري مختلفة عن المفردات الموضوعية يمثل مشكلة تتعلق بتصحيح هذه المفردات ، وعلى الرغم من البحوث التى تحاول جعل تصحيح أسئلة المقال موضوعياً، بل واستخدام الحاسوب في تقدير درجات هذا النوع من الاسئلة ، إلا أن هذه المحاولات لا تزال مبكرة ، لذلك اتجه خبراء القياس والتقويم فى الآونة الأخيرة إلى استخدام أساليب التقويم البديل التى تختلف عن أساليب بناء الاختبارات المقننة ، وذلك لقياس المهارات التى لا نستطيع المفردات الموضوعية قياسها .

التطبيق المبدئي للاختبار ، وتحليل مفرداته :

بعد وضع خطة الاختبار وكتابة مفردات وتدقيقها يجب أن يطبق الاختبار تطبيقاً مبدئياً على عينة من عينات من الطلاب تماثل المجموعة الفعلية المستهدفة التي سيُطبق عليها الاختبار فى صورته النهائية ، وهذه الخطوة تعد من الخطوات الضرورية ، لأنه سيتم بناءً على نتائج هذا التطبيق تحليل المفردات Item Analysis ومراجعتها ، وحذف المفردات غير المميزة ، أو الغامضة ، أو التي مستوى صعوبتها مرتفع ، وغير ذلك ، ومراجعة وتعديل هذه المفردات بما يضمن اشتمال الاختبار على أفضل المفردات .

تجميع مفردات الاختبار :

عند تجميع مفردات الاختبار فى صورته النهائية يجب أن تمثل المفردات التى يشتمل عليها الاختبار جميع أقسام المحتوى والمهارات طبقاً للخطة الموضوعة ، كما يجب أن يكون عدد المفردات مناسباً للزمن المحدد للإجابة عليها، وأن يكون مستوى صعوبتها مناسباً ، ومثل هذا التوازن يعد من العمليات الدقيقة التي جيب أن تعتمد على البيانات الإحصائية ، والأحكام المستنيرة.

تحديد الخصائص السيكومترية للاختبار :

إن تحديد الخصائص السيكومترية للاختبار المقنن يعد عملية مستمرة ، فعلى الرغم من أن بعض خصائص مفردات الاختبار ، مثل : تمييز المفردة ، وصعوبتها ، وزمن الإجابة عليها ، يمكن تقديره من خلال نتائج التطبيق المبدئي للاختبار ، إلا أن إجابات الطلاب عند التطبيق النهائي للاختبار ربما تختلف عن إجابة الطلاب الذين طبق عليهم الاختبار فى التجربة الاستطلاعية المبدئية ، لذلك فإنه يجب تحليل الصورة الكاملة للاختبار لمعرفة مدى اتساق الدرجات ، وصدق الاختبار ، وبخاصة صدق محتواه ن وينبغي اختيار معامل الثبات المناسب لطبيعة الاختبار ، والغرض منه ، والمجموعة المستهدفة ، ونظراً لأن تفسير درجات الاختبار ينبغي أن يستند إلى إطار مرجعي معين ، فإن هذا يتطلب إعداد معايير الاختبار Test Norms التى أوضحناها فى الفصل الرابع ، ويكون التفسير فى هذه الحالة مرجعي المعيار Norm– Referenced – أي يُقارن أداء الطالب فى الاختبار بأداء الجماعة أو الجماعات المرجعية التى تمثل الفرق الدراسية المختلفة ، والمناطق الجغرافية ، وحجم المدارس التى سيتم انتقاء عينات منها .

غير أن هذا التفسير لا يقدم دليلاً على درجة كفاءة الطالب فى المادة الدراسية ، مما قد يستلزم تحديد محكات أو مستويات يُستند إليها فى تحديد مقدار كفاءة أو مهارة الطالب فى المادة الدراسية ، وهذا يسمح بأن يكون التفسير مرجعي المحك – Crierion Referenced .

إعداد دليل الاختبار :

لا ينتهي إعداد الاختبار التحصيلي المقنن ببنائه واعداد جداول معاييره ، وإنما ينبغي إعداد دليل أو أدلة الاختبار Test Manuals والمواد المساعدة ، لمن يستخدمون الاختبار ، بحيث تزودهم بمعلومات كافية لتحديد ما إذا كان الاختبار مناسباً لأغراضهم ، وتُمكنهم من تفسير الدرجات تفسيراً صحيحاً ، ويتبع معظم ناشري الاختبارات في إعداد هذه الأدلة العناصر والمستويات التى حددتها الجمعية الأمريكية لعلم النفس (APA) في مطبوعها بعنوان " مستويات الاختبارات النفسية والتربوية وأدلتها ".

وهذه المجموعة من المستويات ينبغي أن تكون مألوفة لدى من يود استخدام الاختبارات المقننة ، وسوف أوضح ذلك فى الفصل التاسع .

فجميع هذه الاختبارات ينبغي أن يصاحبها أدلة الاختبار التي تتضمن معلومات تتعلق بتفسير درجات الاختبار ، وثباتها ، وصدق الاختبار ، وتعليمات عن تطبيق الاختبار وتصحيحه ، والمعايير المناسبة للاختبار فى صورة جداول تشتمل على الدرجات الخام الممكنة فى الاختبار ، والدرجات المحولة المناظرة لها التى تكون عادة مئينيات ، أو تساعيات معيارية ، وينبغي أن تكون هذه المعايير حديثة ، وجميع المعلومات مكتملة .

وعندما يود ناشر الاختبار إعداد دليلين أو أكثر للاختبار نفسه ، أحدها يناسب المعلمين والموجهين وغيرهم من المربين ، فإن الدليل يمكن أن يتضمن معلومات يمكن فهمها بسهولة دون التركيز على الجوانب الإحصائية المعقدة التى تحتاج إلى أخصائيين مدربين على تفسير البيانات السيكومترية .

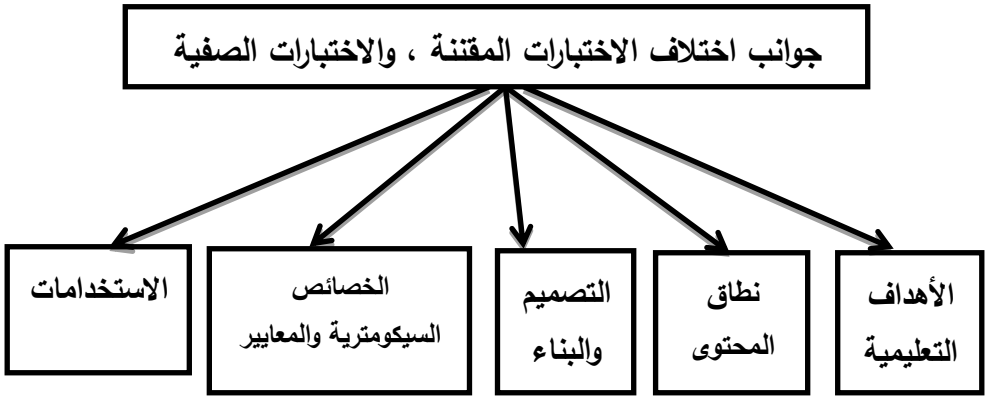
أما الدليل الآخر فيكون دليلاً فنياً يتضمن بيانات ومعلومات إحصائية تتعلق بالاختبار ومعاييره ، وخصائصه السيكومترية ، وذلك للمتخصصين الذى يودون فحص الاختبار فحصاً متعمقاً ، وبذلك يكون الاختبار التحصيلي المقنن صالحاً للنشر .

ومن هذا يتضح أن بناء هذا النوع من الاختبارات يعد عملاً فنياً متخصصاً، ويتطلب قدراً كبيراً من الجهد والوقت والكلفة ، ولذلك يوكل هذا الأمر إلى مؤسسات متخصصة لديها الخبرات الفنية المتميزة فى القياس والتقويم ، والموارد المالية اللازمة لذلك .

- **الفروق بين الاختبارات التحصيلية المقننة، والاختبارات الصفية :**

يمكن استخدام الاختبارات التحصيلية المقننة إلى جانب الاختبارات الصفية فى تقويم الطلاب ، حيث إن كلاً منها يُزوَّد المعلم بمعلومات مختلفة عن تحصيل طلابه ،

على الرغم من أن كليهما يقيس معارفهم ومهاراتهم ، غير أن هناك فروقاً أساسية بين نوعي الاختبارات فى عدد من الجوانب الموضحة بالشكل التخطيطي (٦-٥) التالي :



شكل (٦-٥) يوضح جوانب اختلاف الاختبارات التحصيلية المقننة ، والاختبارات الصفية

الأهداف التعليمية : يستند المعلم فى بنائه للاختبارات الصفية التى يكتبها لطلابه إلى أهداف تعليمية لوحدات دراسية يقوم بتدريسها فى الصف المدرسي ، وهذه الأهداف ربما تختلف من صف إلى آخر بما يناسب طلاب الصف المعين ، وهذه تُعد من مميزات هذه الاختبارات في تحديد الأهداف التى حققها معظم الطلاب ، والأهداف التى تتطلب مزيداً من الجهد لتحقيقها ، أما الاختبارات التحصيلية المقننة فتُصمم لتقييم أهداف تربوية عريضة ومتسعة تناسب فرقة دراسية معينة .

نطاق المحتوى : تتعلق الاختبارات الصفية بمحتوى وحدة دراسية معينة أو موضوعات محددة ، أما الاختبارات التحصيلية المقننة فإنها تتناول مجالات محتوى أكثر أتساعاً ، لأنها ربما تُستخدم فى تقييم تحصيل الطلاب فى أكثر من عام دراسي واحد ، وكذلك إذا كانت تشتمل على عدة اختبارات فرعية ضمن بطارية من الاختبارات ، لذلك ينبغي معرفة الغرض من الاختبار ، ومدى ملاءمته لقياس تحقُق أهداف تعليمية معينة .

التصميم والبناء : يقوم بإعداد الاختبارات التحصيلية المقننة نخبة من خبراء القياس والتقويم واختصاصي المواد الدراسية المختلفة ، ويستغرق بناؤها وقتاً طويلاً ، وجهداً وافراً

وكلفه عالية ، ويمر بناء مفرداتها بمراحل متعددة للتحقق من توافر الشروط الفنية فى كتابتها وصياغتها وتدقيقها ، وكذلك تجريبها ، وتحليل مفرداتها باستخدام الأساليب الإحصائية لتحديد معامل صعوبتها وتمييزها بين الطلاب ومراجعتها أو تعديلها فى ضوء ذلك ، أما الاختبارات الصفُية فإن المعلم يقوم بكتابة مفرداتها على الرغم من أنه ربما لا يكون لديه خبرة كافية أو تدرباً فى فنيات بناء مختلف أنواع المفردات ، ولا يقوم فى كثير من الأحيان بتجريب المفردات للتحقق من جودتها باستخدام أساليب تحليل المفردات .

الخصائص السيكومترية والمعايير : تتميز الاختبارات التحصيلية المقننة بتحديد خصائصها السيكومترية ، مثل الثبات ، والصدق ، ويزداد ثبات درجاتها فى كثير من الأحيان عن (٠.٩٠) ، مما يجعل هناك ثقة في اتساق الدرجات .

أما الاختبارات الصفُية فإن المعلم نادراً ما يقوم بالتحقق من ثبات درجاتها ، وإذا تم ذاك فإن معامل الثبات يكون عادة منخفضاً ، وربما يرجع ذلك إلى تذبذب درجات الاختبار ، وقلة عدد مفرداته .

وتتميز الاختبارات المقننة كذلك بأنها تشتمل على معايير من نوع معين، فبعضها يشتمل على معايير قومية فى ضوئها مقارنة أداء طالب ، أو صف ، أو فرقة ، أو مدرسة معينة بالتقدم الأكاديمي للطلاب على المستوى القومي .

وبعض أخر يشتمل على معايير محلية أو على مستوى مناطق معينة ، لذلك فإن هذه الاختبارات تكون مرجعية المعيار ، أما الاختبارات الصفُية فلا يتوافر لها هذه المعايير ، حيث إنها تكون عادة مرجعية المحك.

الاستخدامات : على الرغم من أن الفرض من الاختبارات المقننة ، والاختبارات الصُفية هو تقييم معارف الطلاب ومهاراتهم في وقت معين ، إلا أنه نظراً لاختلافهما فى المنظور والمحتوى ، والبيانات المتعلقة بالصدق والثبات والمعايير، فإنهما يختلفان فى الاستخدامات ، فالاختبارات الصُفية تُستخدم فى تقييم مدى تحقق أهداف تعليمية محددة بدرجة أفضل من الاختبارات التحصيلية المقننة ، أما الاختبارات التحصيلية المقننة فإنه يمكن الاستناد إليها فى اتخاذ قرارات تتعلق بالتوجيه والإرشاد التربوي والمهني للطلاب ،

وانتقائهم وتسكينهم ورفعهم ، وتقويم المناهج وأداء المعلم ، وإجراء البحوث ، وكذلك المقارنة بين الطلاب فى المدارس المختلفة ، وغير ذلك .

ويمكن تلخيص الفروق بين الاختبارات التحصيلية المقننة ، والاختبارات الصفُية فى جدول (٥-٣) التالي :

جدول (٥-٣) تلخيص الفروق الأساسية بين الاختبارات التحصيلية المقننة ، والاختبارات الصفية

	الاختبارات الصفَية	الاختبارات المقننة
نواتج التعلم ومحتوى الاختبار	تناسب حاجات الطلاب فى صف مدرسي معين ، ومرونة النواتج والمحتوى تسمح بتكييف الاختبار وفقاً لأهمية كل وحدة من الوحدات التعليمية.	تعتمد على أهداف عريضة لفرقة دراسية معينة تشترك فيها معظم المدارس ، وتتم معاينة المحتوى بطريقة منتظمة
نوعية المفردات	يقوم معلم واحد ببناء المفردات وعادة تكون نوعيتها محدودة لقلة خبرة المعلم بفنيات كتابة الأنواع المختلفة من المفردات وضيق وقته.	تُبنى المفردات بواسطة مجموعة من الخبراء، وتخضع لعمليات تدقيق ومراجعة وتحرير، وتجريب ميداني، وتحليل إحصائي لكي تكون المفردات ذات نوعية عالية الجودة .
طرق التطبيق والتصحيح	يحددها المعلم ، وتكون موحدة فى الصف المدرسي الواحد، ولكن يمكن تعديلها بما يتفق وحاجات الطلاب.	يحددها ناشري الاختبارات ويجب اتباع تعليمات الاختبار فى هذا الشأن بمنتهي الدقة.
الثبات	ثبات الدرجات لا يكون عادة معلوماً ، ويمكن أن يكون مرتفعاً إذا اعتنى المعلم ببناء الاختبار وفق خطة مدروسة.	ثبات الدرجات يكون مرتفعاً ، ويتراوح عادة بين(٠.٨ ، ٠.٩٥)، وربما يزيد عن ذلك.
المعايير	لا تُبني معايير Norms لهذه الاختبارات ، ولكن يمكن للمعلم حساب المعايير للصف الذى يقوم بالتدريس فيه.	تُبني معايير لهذه الاختبارات لمجموعات مختلفة من الطلاب، وتُفسر الدرجات الخام فى ضوئها ، ويستخدمها جميع المعلمين لمقارنة أداء طلاب صف معين بمختلف المجموعات العمرية والصفوف المدرسية على المستوى القومي أو المحلي .

	الاختبارات الصفِية	الاختبارات المقننة
الأغراض والاستخدامات	تناسب قياس أهداف تعليمية يحددها المعلم ، ويقارن فى ضوئها أداء الطالب بأداء اقرانه فى الصف المدرسي ، أو تحديد ما اكتسبه الطالب من معارف ومهارات تتعلق بمحتوى دراسي معين	تناسب بدرجة أفضل قياس الأهداف العريضة لمنهج دراسي معين ، وتُستخدم هذه الاختبارات فى المقارنة بين الصفوف والمدارس ، واتخاذ قرارات تتعلق بالطلاب والمعلمين والمناهج المدرسية .

اختبارات الذكاء العام

- مقدمة
- مفهوم الذكاء وتطوراته التاريخية .
- التعريفات المتعددة للذكاء .
- نظريات بنية الذكاء .
- تصنيف اختبارات الذكاء العام .
- اختبارات الذكاء العام الفردية .
- اختبارات الذكاء العام الجماعية .
- تعقيب على اختبارات الذكاء العام .

❖ مقدمة :

تناولنا فى الفصل السابق الاختبارات التحصيلية التى تُستخدم فى قياس نواتج التعلم المتعلقة بالمناهج والمواد الدراسية ، ومقارنة أداء الطالب بأداء أقرانه ، غير أنه من الأهمية أيضاً مقارنة أداء الطالب بقدرته على التعلم لكي نتحقق من توظيفه لإمكاناته العقلية إلى أقصي حد ممكن ، فالمعلمون يدركون أن جميع الطلاب يمكنهم إحراز تقدم دراسي والوصول إلى مستوى واحد ، ولكنهم يعملون على توفير فرصة كافية لكل منهم لتحقيق أعلي مستوى مناسب لإمكاناته ، وهذا يتطلب معرفة مستوى الذكاء العام للطالب واستعداداته الخاصة ، وكيفية استخدام الاختبارات المناسبة لقياس إمكاناته العقلية استخداماً مستنيراً ، ولتحقيق ذلك ينبغي على العلم معرفة المقصود بمفهوم الذكاء والأطر النظرية والمنهجية التى تستند إليها اختبارات الذكاء العام ، وأنواع هذه الاختبارات ومحتواها ، وكيفية تفسير درجاتها ، واستخدام نتائجها استخداماً مناسباً .

لذلك يتناول هذا الفصل أفضل التعريفات المتعددة لمفهوم الذكاء فى ضوء التطور التاريخي لحركة القياس العقلي ، ونظريات بنية الذكاء ، وتصنيف اختبارات الذكاء العام ، وأمثلة لأهم اختبارات الذكاء العام الفردية والجماعية ، وينتهي الفصل بتعقيب عام على هذه الاختبارات ، وسوف أفرد الفصل التالي لاختبارات الاستعدادات الخاصة .

❖ مفهوم الذكاء وتطوراته التاريخية :

يُعد مفهوم الذكاء من المفاهيم السيكولوجية التى نالت كثيراً من الخلاف والجدل بين علماء النفس وخبراء القياس منذ بداية القرن الماضي ، ولعل معرفة التطور التاريخي لحركة القياس العقلي تلقي الضوء على هذه المفهوم بأعتباره من المفاهيم التي تتميز بالتغير المستمر ، فقد نشأ الاهتمام بضعاف العقول ، والكشف عن الدرجات المختلفة للضعف العقلي فى القرن التاسع عشر ، وذلك بهدف وضع أسس لتصنيف الأفراد غير الأسوياء ، والاهتمام بالتمييز بين مرضي العقول ، وضعاف العقول ، ويعد " أسكرول Esquirol " ، وهو طبيب فرنسي ، أول من حاول تصنيف ضعاف العقول عام ١٩٣٨ فى مستويات مختلفة تبدأ من الأسوياء إلى أدني مستويات الضعف العقلي ، ولكن المعايير التى اعتمد عليها فى هذا التصنيف لم تكن موضوعية ، حيث اعتمد على شكل جمجمة الإنسان ، وحجمها، ونسبتها إلى الجسم ، وهذا تعد من المظاهر الجسمية ، وبخاصة فى تشخيص المستويات الدنيا للضعف العقلي ، غير إنه اعتمد على بعض طرق قياس السلوك اللفظي فى تشخيص المستويات العليا للضعف العقلي ، وحاول بعض العلماء فى ذلك الوقت وضع بعض الاختبارات التى تقيس المهارات الحسية الحركية والعملية لتدريب ضعاف العقول ، وبالطبع لم تجدي هذه الطرق غير الموضوعية فى الكشف عن ضعاف العقول أو تدريبهم (Hunt, 1961).

وقد شهد منتصف القرن التاسع عشر تحولاً ملحوظاً فى تطور حركة القياس العقلي ، حيث تحوَّل الاهتمام من التأملات الشخصية إلى التجارب .

فقد أسُس " فندت Wundt عام 1879 " أول مختبر لعلم النفس التجريبي فى ليبتزج بألمانيا وانصب اهتمام فندت وتلاميذه على بحث الظواهر النفسجسمية، مثل : الظواهر الحسية ، وزمن رد الفعل ، وفى أواخر القرن التاسع عشر حدث تقدم كبير فى علم البيولوجي نتيجة لإسهام شارلز دارون بنظريته فى التطور والنشوء والارتقاء .

ولعل العالم البيولوجي الإنجليزي " فرانسيس جالتون Galton "، وهو من أتباع دارون البارزين هو الذى وضع المعالم الرئيسة للقياس العقلي، فقد اهتم "جالتون " بمشكلة الوراثة، وأنشأ معمل دراسة الوراثة فى لندن، وأجري دراسات وثائقية من سجلات طلاب

بعض المدارس الإنجليزية تتعلق بالمظاهر الجسمية لهؤلاء الطلاب واستخلص أن الطلاب الذين يتميزون بقدرة عالية فى التمييز الحسي يُعدون أذكى الطلاب بعامة .

وقد أدت جهود جالتون والأساليب الإحصائية التى استخدمها فى ذلك الوقت إلى حفز تلاميذه من أمثال " بيرسون Pearsoon وكاتل Cattell " ، إلى الاهتمام بعلم الإحصاء ، كوسيلة للبحث النفسي ، وقد عاد كاتل إلى الولايات المتحدة الأمريكية ، حيث أشرف على معامل علم النفس التجريبي ونشر أراءه عن القياس العقلي ، ولكنه اعترض على وجهة نظر جالتون ، وهى أن قياس الوظائف العقلية يمكن الاستدلال عليه عن طريق قياس الوظائف النفسية البسيطة ، مثل: التمييز الحسي ، وقياس زمن رد الفعل ، وغير ذلك من الوظائف الدنيا ، لذلك تحولَّ اهتمامه إلى دراسة الفروق الفردية فى نواحي أخري مثل الذكاء ، وقد قام ببناء مجموعة من الاختبارات العقلية التى اشتملت على مقايس لسرعة حركة الذراع ، أو التمييز بين الأوزان وغير ذلك ، إلا أنه تبين بعد ذلك أنه لا توجد علاقة بين درجات هذه الاختبارات ودرجات التحصيل الدراسي ، مما أدي إلى دحض افتراض وجود ارتباط بين القدرة الحسية والذكاء ، ومع هذا فإن هذه الفكرة أسهمت فى اقتراح مقاييس جديدة للذكاء .

وقد انتقد " بينيه Binet " وهو عالم نفس فرنسي ، الاختبارات المستخدمة فى ذلك الوقت لقياس الذكاء ، واقترح هو وزميله " سيمون Simon " بناء اختبار جديد لقياس القدرة العقلية Mental Ability على أساس الوظائف العقلية العليا ، ورأي بينيه أن الذكاء قدرة عقلية عامة تتكون من مجموعة من القدرات المنفصلة وهى الاستدلال ، والحكم والفهم الجيد ، واتخاذ توجه مناسب للتفكير والحفاظ على هذا التوجه ، وتعديل التفكير بما يحقق هدفاً معيناً ، والنقد الذاتي .

وقد استغرق بناء الاختبار حوالي عشر سنوات ، وظهر الاختبار أو المقياس الجديد عام ١٩٠٥ ، حيث نال اهتماماً كبيراً من علماء النفس فى العالم .

وقام " تيرمان Terman " وهو عالم نفس أمريكي ، بمراجعة الاختبار وترجمته إلى اللغة الإنجليزية فى جامعة ستانفورد بالولايات المتحدة الأمريكية ، وأطلق عليه أسم اختبار ستانفورد – بينيه للذكاء .

ويرجع الفضل فى تقرير أول وحدة فى القياس العقلي ، وهى " العمر العقلي Mental Age (MA) " إلى بينيه .

بينما يرجع الفضل إلى تيرمان فى إدخال مصطلح جديد ، وهو مصطلح "نسبة الذكاء Intelligence Quotient (IQ) " وعلى الرغم من انتشار استخدام اختبار بينيه فى قياس الذكاء العام ، إلا أنه تبين فيما بعد أنه مشوب بأوجه قصور متعددة ، لذلك قام " دافيد ويكسار Wechsler عام ١٩٣٩" ، ببناء اختبار ذكاء فردي يختلف عن اختبار بينيه فى أنه يقيس بدرجة اساسية ذكاء الراشدين ، والتعرف على الأنماط المنحرفة للاستجابات التى تميز مختلف أنواع الأمراض النفسية ، غير أن نتائج الدراسات النفسية لم تستطع التوصل إلى أنماط معينة من الاستجابات تميز الأفراد الذين يعانون من مشكلات نفسيه مختلفة ، لذلك أصبح يقتصر استخدام اختبار ويكسلر على قياس الذكاء ، وليس تشخيص الأمراض النفسية .

ومن التطورات الأخرى التى حدثت فى حركة القياس العقلي هى ظهور الاختبارات الجماعية Group Tests لقياس الذكاء ، فمن المعلوم أن اختبار بينيه اختبار فردي Individual Test ، أي يطبق على فرد واحد فى وقت معين ، ويحتاج إلى شخص على درجة عالية من التدرب ليشرف على تطبيقه وتفسير نتائجه .

ولكن عندما دخلت الولايات المتحدة الأمريكية الحرب العالمية الأولي عام ١٩١٧ تطلب الأمر تصنيف الأفراد وفقاً لذكائهم بغرض توجيههم نحو الوظائف العسكرية المناسبة لقدراتهم .

لذلك شرع أحد الخبراء النفسيين فى الجيش الأمريكي وهو أرثر أوتيس Arther Otis " مع مجموعة من زملائه فى بناء مجموعة من الاختبارات الجماعية لقياس القدرة العقلية للمجندين فى الجيش الأمريكي ، أي تطبق على مجموعة كبيرة من الأفراد فى

الوقت نفسه ، ومن أشهر هذه الاختبارات : اختبار ألفا Alpha Test ، وهو يقيس القدرة العقلية للأفراد غير الناطقين باللغة الانجليزية أو الأميين .

وبعد انتهاء الحرب العالمية الأولي انتشرت هذه الاختبارات ، وترتب على ذلك تطور سريع لحركة القياس العقلي فى الولايات المتحدة الأمريكية ، وفي كثير من دول العالم ، ومن أهم هذه التطورات بناء اختبارات لقياس مختلف مظاهر السلوك ، واختبارات جماعية لقياس الذكاء فى مختلف الأعمار بما فى ذلك الأطفال والراشدين ، وبطاريات لقياس الاستعدادات الخاصة ، وتبع ذلك بناء مقاييس للسمات غير المعرفية ، مثل : استبيانات الشخصية ، والاختبارات الإسقاطية ، وغيرها من طرق قياس الشخصية التى ستناولها فى الفصل الثامن.

❖ التعريفات المتعددة للذكاء :

تعدد وجهات نظر علماء النفس فيما يتعلق بتعريف مفهوم الذكاء Intelligence " لذلك لا يوجد اتفاق على تعريف موحد يمكن الاستناد إليه فى قياس الذكاء ، ولعل التطور التاريخي لحركة القياس العقلي الذى أوجزناه يوضح ذلك ، فمحاولات تعريف الذكاء تواجه مشكلة اعتبار أن الذكاء له كينونة حقيقة بحيث يمكن تحديدها وقياسها ، فالذكاء هو مجرد مصطلح يدل على مجموعة من السلوك المهم فى حياة الأفراد ، لذلك لا ينبغي النظر إلى مصطلح أو كلمة الذكاء على أنها بديلاً للشئ الذى تمثله ، أي سلوك ، ومن ثم تحدي الشئ بدلاً من تحديد السلوك ، فمصطلح الذكاء مثل مصطلح " الطقس " هو كلمة تصف ظاهرة تسلك بطريقة معينة ، فنحن لا نقيس الذكاء ، ولا نقيس الطقس ، وإنما نقيس السلوك فى كل من الحالتين ، وليس المفهوم أو المصطلح ذاته ، غير أننا نستدل على المفهوم من نتائج قياس السلوك ، ونحكم على الفرد بأن ذكاءة مرتفع أو منخفض أو أن قدرته اللفظية أفضل من قدرته غير اللفظية.

لذلك تحاول كثير من التعريفات وصف الذكاء ، وكيف يعمل ولعل هذا يتضح فى تعريف بينيه للذكاء ، حيث أشار إلى كيفية إحداثه لأثر فى السلوك ، وخصائصه الأساسية .

وكذلك اقتراح " ثورنديك Thorndike " تعريفاً للذكاء بأنه قدرة الفرد على القيام بمهام معينة مقارنة بأقرانه ، وأداء هذه المهام ونتائجها هو الذى يعكس الذكاء .

وقد اقتراح " بيرت Burt " التمييز بين التعريف الأسمي Nominal Definition ، والتعريف الفعلي Real Defintion للذكاء ، فالتعريف الأسمي يشير إلى كيفية استخدام مفهوم الذكاء ، بينما يشير التعريف الفعلي إلى تفسير طبيعة الذكاء ، ولهذا تبدو أهمية تعريف المفهوم تعريفاً أسمياً قبل محاولةل قياس الذكاء ، أما التعريف الفعلي للذكاء فإنه يتطلب مزيداً من البحث والدراسة .

غير أن المتتبع للتعريفات الحالية للذكاء يلاحظ تنوع السلوك الذى يشير إليه التعريف ، مما ينعكس على المهام الاختبارية ، فاختبار بينيه يتضمن مهاماً فرعية متعددة ، ولكنها تضم جميعاً فى درجة واحدة تُعبر عن القدرة العقلية العامة ، بينما تشتمل اختبارات ويكسلر على موازين فرعية Subscales متعددة تؤدي إلى نسب ذكاء لفظية وأخري غير لفظية ، وكذلك نسب ذكاء كلية ، أما " جليفورد Guiford,1983 " فقد اقترح بنية تنظيمية للذكاء ثلاثية الأبعاد ، هي : المحتوى، والعمليات ، والنواتج ، فى حين أن " هوارد جاردنر Gardner,1983 " اقترح نظرية الذكاءات المتعددة Multiple Intelligences ، وقد انتقد " روبرت سترنبرج Sternberg,1988 " نظرية جاردنر ، واقترح نظرية أخري تعتمد على النظرية الثلاثية Triarchic Theory ، حيث تميز بين العمليات الداخلية ، الخبرة ، والعلاقة بينها وبين متطلبات العالم الخارجي ، وقد أكدت هذه النظرية أهمية مفهوم الضرورة Necessity إلى جانب مفهوم الذكاء ، فالقدرات التى يحتاجها الفرد من أجل الحياة تعد جزءاً من الذكاء ، أما القدرات الأخري فيمكن اعتبارها قدرات خاصة .

وقد بينت نتائج كثير من البحوث أنه على الرغم من تعدد القدرات أو المهارات العقلية الخاصة ، إلا أنها مترابطة ، مما يعزز بدرجة معينة ما سبق تسميته العامل العام "g" Factor أو ما يُعرف بالقدرة العقلية العامة ، غير أن بعض الأفراد يتميزون فى إحدى القدرات المتعلقة بمفهوم الذكاء ، وربما يختلف ذلك عن وجهات النظر المبكرة التى

كانت تري أن الذكاء قدرة غير متغيرة ، فهناك الآن أدوات تستند إلى فكرة تفاعل السمات والمعالجات ، أو التفاعل القدرات والمعالجات ، مثل : استبيانات الانماط المعرفية ، واختبار السلوك التكيفي الذى يقيس قدرة الفرد على أداء مهام تعلم فى الحياة اليومية .

والخلاصة أن هذه التعريفات تؤكد أن الذكاء ليس قدرة فطرية كامنة Capacity ، وإنما سمة سلوكية تعتمد إلى حد ما على التعلم السابق والسلوك المكتسب ، وينبغي أن يقيس اختبار الذكاء سلوكاً مكتسباً يتفق علماء النفس على تسميته ذكاءً ، وعندئذ يصبح الاختبار بمثابة تعريف إجرائي للذكاء ، وعلى الرغم من اختلاف اختبارات الذكاء فى أنماط السلوك الذى تقيسه ، إلا أن الارتباط بين درجات هذه الاختبارات مرتفع نسبياً .

❖ نظريات بنية الذكاء :

اتضح من عرضنا للتطور التاريخي لحركة القياس العقلي أن بينيه Binet وسيمون Simon قاما بتطوير أول اختبار فردي للذكاء لقياس مستوى الذكاء العام لدى الفرد ، وعلى الرغم من أن كثيراً من علماء النفس الآخرين كانت تصوراتهم الفكرية عن الذكاء أنه خاصة أو قدرة عقلية عامة ، إلا أن هناك نظريات أخري معارضة لهذه التصورات .

ويُعد " سبيرمان Spearman,1927 " من العلماء الذين اعتبروا الذكاء مكون من عاملين ، ولذلك أطلق على نظريته فى الذكاء " نظرية العاملين Two Factor Theoy " فقد اقترح أن الذكاء يشتمل على عامل عام General Factor (g) ، وعوامل نوعية Specific Factors .

أما " ثيرستون Thurstone, 1944 " فقد قام بتطوير نظرية العوامل المتعددة Multiple Factors استند إليها فى بناء اختبار القدرات العقلية الأولية الذي يقيس سبعة عوامل ، هى : القدرة على الطلاقة اللفظية ، القدرة على الفهم اللغوي ، القدرة العددية ، القدرة المكانية ، السرعة الادراكية ، القدرة على التذكر ، القدرة على الاستدلال الاستقرائي .

أما " فيرنون Vernon,1961 " فقد اقترح بينة هرمية للقدرات العقلية تبدأ بالعامل العام الذى يتفرع إلى عاملين طائفتين رئيسين هما : عامل لفظي – تربوي، وعامل حركي

– ميكانيكي ، وهذين العاملين الطائفيين أقل عمومية من العامل العام (g) الذى اقترحه سبيرمان ، ولكنهما أكثر عمومية من العوامل التى اقترحها ثيرستون .

واقترح " كاتل Cattell,1963 " نظرية توضح أن الذكاء يشتمل على مكونتين إحداهما الذكاء السائل Fluid Intelligence ، والأخرى الذكاء المتبلور Crystallized Intelligence ، ويُعد الذكاء السائل أكثر عمومية ، وارتفاع مقداره يدل على أن الفرد يستطيع القيام بمهام كثيرة ومتنوعة ، لذلك فإن الذكاء السائل هو كفاءة عقلية تتميز بالتجريد وغير لفظية بدرجة أساسية ومتحررة من الثقافة ، أما الذكاء المتبلور فهو أكثر صلة بالثقافة أو البيئة ، ويمثل قدرة الفرد على إنجاز مهام أكثر نوعية وارتباطاً بالثقافة ، وبذلك فإن هذه القدرة تماثل التحصيل .

واقترح " جيلفورد Guilford,1959 " نموذجاً لبنية العقل Structure of Intellect (SI) حيث اعتبر أن الذكاء يتكون من عوامل متعددة قام بتصنيفها فى أبعاد ثلاثة رئيسة هي : العملية التى يجربها الفرد ، ونوع المادة أو المحتوى ، ونوع النتاج ، وكل بعد من هذه الأبعاد أعيد تقسيمه إلى خمس عمليات ، وأربعة أنواع من المحتوى ، وستة أنواع من النواتج ، ويمكن تمثيل هذه الأبعاد بمكعب يحتوي على ٤ × ٥ × ٦ =١٢٠ خلية ، يمثل كل منها مظهراً معيناً من مظاهر النشاط العقلي ، واستطاع جيلفورد أن يبين وجود (٨٢) عاملاً من هذه العوامل ، واقتراح قياس كل منها على حده ، وقد نال هذا النموذج اهتماماً كبيراً من جانب علماء النفس بإعتباره نموذجاً نظرياً منطقياً أكثر من كونه نموذجاً عملياً تطبيقياً ، حيث إنه ربما يصعب بناء اختبارات تناظر كل خلية من خلايا المكعب ، وحتى إذا أمكن ذلك فإن هذه الاختبارات لا يكون لها فائدة حقيقية من حيث قيمتها التنبؤية ، نظراً لأنها تقيس عوامل نوعية بدرجة كبيرة ومن الجدير بالذكر أن جيلفورد اعتمد فى هذا النموذج على أسلوب التحليل العاملي Factor Analysis ، وهو من الأساليب الإحصائية متعددة المتغيرات الذى يهدف لاختزال عدد كبير من المتغيرات وتجميعها فى عدد قليل من العوامل التى تفسر التباين المتسق فى هذه المجموعة من المتغيرات ، وهذه العوامل Factors تعد تكوينات فرضية Constructs تنطوي تحتها مجموعة المتغيرات وتفسرها .

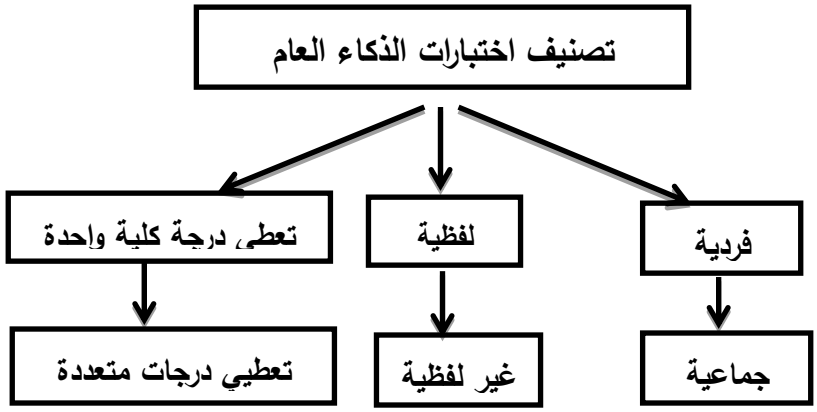
أما " جاردنر Gardner,1989 " وهو من علماء النفس المعرفي بجامعة هارفرد فقد حاول خلال العقدين الماضيين تطوير نظرية الذكاءات المتعددة Multiple Intelligences وكيفية الاستفادة منها فى التعليم ، وأوضح عدداً من هذه القدرات والمهارات المتمايزة التي يمتلك كل فرد درجات معينة من كل منها ، وهى : القدرة البصرية – المكانية – القدرة الجسمية – الحركية ، القدرة الموسيقية – الإيقاعية ، القدرة على التفاعل بين الأفراد ، القدرة على معرفة مشاعر الفرد الذاتية ، القدرة المنطقية – الرياضية ، القدرة اللفظية – اللغوية .

وأوضح " جاردنر " أن هذا يخالف الاعتقاد بأن الذكاء الإنساني سمة ثابتة واحدة ، وإنما توجد ذكاءات متعددة يمكن تنميتها بأساليب تعليم مناسبة ، وتتشكل بتفاعل الفرد مع بيئته تفاعلاً ايجابياً مثمراً .

ومن هذا يتضح عدم اتفاق علماء النفس حول بنيه الذكاء ، لذلك تعددت الاختبارات التى يُطلق عليها اختبارات الذكاء ، وتنوعات تنوعاً كبيراً ، وربما لا تقيس جميعها الشئ نفسه ، لذلك ينبغي عند انتقاء أحد هذه الاختبارات وتفسير درجاتها مراعاة تعريف الذكاء الذي استند إليه بناء الاختبار ، ولمزيد من تفاصيل نظريات بنيه الذكاء يمكن الرجوع إلى كتاب " القياس والتقويم " التربوي والنفسي " للمؤلف .

❖ تصنيف اختبارات الذكاء العام :

سبق أن أوضحنا فى الفصل الأول أنواع الاختبارات التربوية والنفسية وفقاً لتصنيف يستند إلى عدد من الأبعاد ، وينطبق هذا التصنيف فى كثير من أبعاده على اختبارات الذكاء ، حيث إن هذه الاختبارات يمكن أيضاً تصنيفها بطرق متعددة ، ولعل التصنيف الذي يوضحه الشكل التخطيطي (٦-١) التالي يفيد فى التعريف على أنواع اختبرات الذكاء المتوافرة وانتقاء الاختبار المناسب لغرض معين :



شكل (٦-١) يوضح تصنيف اختبارات الذكاء العام

❖ الاختبارات الفردية والجماعية

تنقسم اختبارات الذكاء إلى إختبارات فردية ، واختبارات جماعية ، والاختبار الفردي Individual Test يقوم بتطبيقه أخصائي نفسي على فرد واحد ، ويقوم الأخصائي بتقديم المهام التي يشتمل عليها الاختبار للفرد المختبر شفوياً وجهاً لوجه ، ويستجيب الفرد بالإشارة إلى مثير معين ، أو تقديم إجابة شفوية ، أو التعامل يدوياً مع مهمة معينة ، ويتطلب تطبيق الاختبار أخصائي نفسي متمرس على هذا النوع من الاختبارات الفردية ، حيث أن الإخصائي يعد جزءاً من الإجراءات المقننة للاختبار ، إذ ينبغي أن يقنن تعامله مع المختبر وفقاً لتعليمات الاختبار المحددة ، وكثير من مفردات الاختبارات الفردية ، لا يتم تقدير درجاتها على أنها صحيحة أو خطأ ، وأنما ربما تكون لها إجابات متعددة مقبولة ، لذلك تقدر عادة درجات مختلفة بحسب درجة صحة الإجابة ، ويحدد فى دليل الاختبار الإجابات التي يمكن اعتبارها صحيحة ، والدرجة الجزئية التي تقدر لكل منها ، وينبغي أن يلتزم الأخصائي بهذه التعليمات ولا يجيد عنها ، لكي لا تتأثر الدرجات بأحكامه الذاتية عن ذكاء المختبر .

لذلك يستغرق تطبيق وتصحيح هذا النوع من الاختبارات وقتاً طويلاً نسبياً ، ويصلح بدرجة أفضل للأطفال الصغار ، وكذلك فى المواقف الكلينيكية ، حيث تسمح للأخصائي

بالملاحظة الدقيقة للمختبر ، وتتميز هذه الاختبارات بأن تقديراتها للذكاء تكون أكثر صدقاً لأنها تطبق فى ظروف مثلي ، فالمختبر تكون دافعيته للإجابة وتعاونه أفضل ، ويتأكد الأخصائي بأن المختبر فى صحة جيدة ، وغير متوتر ، وانتباهه مركزاً عل المهام المعطاة ، ومستوعباً استيعاباً كاملاً للتعليمات ، كما أن قدرته على القراءة لا يكون لها تأثير على أدائه أو درجته فى الاختبار .

ومن أهم أمثلة الاختبارات الفردية : اختبار بينيه ، واختبارات ويكسلر للذكاء التى سأتناولها بالتوضيح بعد قليل .

أما الاختبارات الجماعية Group Tests فقد برزت كما سبق أن أوضحنا نتيجة لحاجة ملحة ، وهى قياس القدرة العقلية للجنود فى الحرب العالمية الأولي ، حيث يصعب بدرجة كبيرة تطبيق اختبار فردي مثل اختبار بينيه على أعداد كبيرة من هؤلاء الجنود .

وتشتمل الاختبارات الجماعية عادة على مفردات تتطلب إجابتها الورقة والقلم ، لذلك يمكن تطبيقها على مجموعة كبيرة من الأفراد فى وقت واحد ، مما يوفر الوقت والجهد والكلفة ، ولا تتطلب هذه الاختبارات من القائم بتطبيقها تدريباً متخصصاً ، حيث إنه يقرأ تعليمات الاختبار ويسمح ببعض الوقت لكي يألف المختبرون كيفية الإجابة على مفردات الاختبار ، ويعطي لهم الإذن ببدء الإجابة ، ويجمع أوراق الإجابة عند انتهاء الزمن المخصص للاختبار .

ويتميز هذا النوع من الاختبارات بسهولة تطبيقه ، والانخفاض النسبي لكلفة تصحيح الاختبار ، وكذلك سرعة التصحيح ، غير أن في مقابل ذلك تكون ظروف تطبيق الاختبار غير مقننة بدرجة كاملة ، وربما تتأثر الدرجات بعوامل مختلفة ، مثل : سرعة القراءة ، والخلفية التعليمية للمختبرين ، لذلك فإن درجات اختبارات الذكاء الجماعية تقلل عادة من تقديرات ذكاء المختبرين .

ومع هذا ، فإن هذه الاختبارات تناسب الأفراد الراشدين العاديين فى المواقف الجماعية مثل : الصفوف المدرسية ، والمؤسسات الصناعية والعسكرية.

ومن أفضل أمثلة الاختبارات الجماعية ، اختبار أوتيس للقدرة العقلية (اختبار ألفا) الذى أشرنا إليه من قبل ، واختبار ثيرستون لليقظة العقلية الذى يزودنا بدرجتين فرعيتين للذكاء إحداهما درجة للمفردات الكمية ، والأخري للمفردات اللغوية ، وكذلك درجة كلية ، وهناك اختبار التصنيف العام (AGCT) الذى تم تقنينه على مجتمع من المجندين بالجيش الامريكي أثناء الحرب العالمية الثانية ، واصبح يُستخدم فى قياس ذكاء الأفراد المدنيين ، ويركز على الاستدلال العددى ، والاستدلال اللفظي ، والادراك المكاني (القدرة على تصور الأشكال فى الفراغ) .

وزمن تطبيق هذا الاختبار ٤٠ دقيقة ويؤدى إلى درجة محولة ، ويعد من اختبارات الذكاء العام الجماعية التي تتميز بالصدق ، ويُستخدم فى إنتقاء الأفراد، ولكنه أقل فائدة في التنبؤ بنجاح الطلبة فى الكليات ، وذلك لأن نوع الذكاء الملائم للنجاح فى الكليات يتطلب قدرات خاصة ، لذلك فإن الاختبار الشائع الاستخدام فى هذا المجال هو اختبار القبول بالجامعات (CEEB) الذى أعده مركز الخدمات الاختبارية بالولايات المتحدة الأمريكية (ETS) ، يشتمل على نوعين من الاختبارات أحدهما اختبار الاستعداد العام ، والثاني اختبارات تحصيلية في مختلف المواد الدراسية ، ويفيد هذا الاختبار فى التنبؤ بنجاح الطلاب فى المرحلة الجامعية الأولي .

❖ الاختبارات اللفظية وغير اللفظية :

تُستخدم معظم اختبارات الذكاء الجماعية ، وبخاصة التى تناسب الراشدين أنواعاً معينة من المفردات مثل ، مفردات لغوية ، إكمال جمل ، متشابها لفظية ، حساب ، متضادات . تسلسل لفظي ، تمييز الفروق ، تفسير أمثال ترتيب جمل ، استدلال ، اتباع تعليمات ، استنتاج مترتبات علاقات بين أشكال ، متشابهات شكلية ، متسلسلات عددية ، تعامل مع مساحات ، وغير ذلك٠ لذلك تعتمد هذه الاختبارات على القراءة والكتابة والتعبير الشفوي ، ويطلق عليها الاختبارات اللفظية Verbal Tests ، أما اختبارات

الذكاء غير اللفظية Nonverbal Tests فإن اللغة لا تلعب دوراً فى الإجابة : فهذه الاختبارات تُستخدم الصور والأشكال أو اي مواد زمزية ، لذلك فإنها تناسب بدرجة أفضل الأطفال الصغار والأميين لقياس الجوانب غير اللفظية للذكاء ، مثل الأدراك المكاني ، أما الاختبارات غير اللغوية Non language فهي لا تعتمد على أي شكل من أشكال التواصل الكتابي أو الشفوي أو القرائي ، وإنما يقدم الفاحص التعليمات من خلال الإيماءات والإشارات، ويستجيب الفرد بالإشارة إلى الأشياء أو التعامل معها يدوياً مثل : الكتل الخشبية، والمتاهات ، والصور ، وغير ذلك ، وتطبق هذه الاختبارات على الأطفال الجسم، أو الذين لا يمكنهم التواصل بأي شكل من أشكال اللغة العادية .

غير أن بعض اختبارات الذكاء تشتمل على أجزاء لفظية وأجزاء غير لفظية ميزات هذا النوع من الاختبارات أنه يمكن باستخدامه تحديد قدرة الفرد على التعلم فى مستويين مختلفين ، فدرجته فى الاختبار اللفظي تفيد فى التنبؤ بنجاحه الدراسي ، وذلك لأن القدرة اللفظية تلعب دوراً مهماً فى تعلم المهام الدراسية ، أما درجته فى الاختبار غير اللفظي فتقدم تقديراً أفضل لقدرته الكامنة على التعلم التى تبدو منخفضة ، فمثلاً الطالب الذى قدرته العقلية تفوق المتوسط ولكن قدرته على القراءة منخفضة يُتوقع أن يحصل على درجة منخفضة فى الجزء اللفظي من الاختبار ، بينما ربما يحصل على درجة مرتفعة فى الجزء اللفظي ، وعلى الرغم من أن احتمال نجاحه الحالي فى الدراسة يكون منخفضاً نظراً لانخفاض قدرته على القراءة ، إلا أنه يُحتمل نجاحه المستقبلي إذا قدم له تعليماً علاجياً مناسباً .

غير أن الفروق بين درجات الجزء اللفظي ، والجزء غير اللفظي ينبغي أن تكون كبيرة نسبياً قبل استخدامها فى أغراض التشخيص ، وذلك لارتباط كل منها بالآخر إرتباطاً مرتفعاً إلى حد ما ، لاشتراكهما فى عامل القدرة العقلية العامة .

❖ الاختبارات التى تعطي درجة كلية واحدة أو تعطي درجات متعددة :

بعض الاختبارات الذكاء التى ينتج عنها درجة واحدة كلية تُصمم بحيث تقيس القدرة العقلية العامة لذلك فإن الجوانب النوعية لهذه القدرة ، مثل : الجوانب اللفظية ، والعددية

، والاستدلال المجرد تُضم معاً فى قياس واحد للقدرة العقلية ، ويتم ترتيب مفردات هذا النوع من الاختبارات الى ينتج عنها درجة واحدة كلية فى نمط حلزوني Spiral – Omnibus Pattern أي خلط الأنواع المختلفة من المفردات ، وترتيبها ترتيبا تصاعدياً من حيث صعوبتها ، وبذلك لا يكون هناك اختبارات فرعية ، مما يجعل تعليمات الاختبار مجموعة واحدة ، ويكون زمن الاختبار قصيراً نسبياً ، ومن أمثلة هذا النوع من اختبارات الذكاء ، اختبارات القدرة العقلية لهلمون ، ونيلسون Helmon – Nelson Tests of Mental Ability للصفوف من الحضانة إلى نهاية المرحلة الثانوية ، واختبار القدرة المدرسية لأوتيس وليتون Otis – Lennon School Ability Test للصفوف من الأول الابتدائي إلى نهاية المرحلة الثانوية .

وتستخدم عادة الاختبارات التي تعطي درجة واحدة كلية فى قياس الاستعداد للتعلم المدرسي ، والتنبؤ بالنجاح فى التعليم المدرسي فى المستقبل ، ومفردات هذه الاختبارات تكون لفظية مما يزيد من قيمتها التنبؤية ، وذلك لأن طبيعة التعلم المدرسي يكون لفظياً .

أما اختبارات الذكاء التى تعطي درجات متعددة فإنها تقيس عدداً من القدرات العقلية ومثال ذلك اختبار القدرات المعرفية Cognitive Abilities التى تناسب التلاميذ من الحضانة إلى نهاية المرحلة الثانوية ، وتُستخدم هذه الاختبارات فى التنبؤ التمايزي ، وفى بطاريات الاختبارات التى ينتج عنها درجات متعددة يكون هناك تداخل بين الاختبارات الفرعية يعكس القدرة العقلية العامة ، وكذلك القدرات النوعية التى تقيسها هذه الاختبارات .

❖ اختبارات الذكاء العام الفردية :

- **مقياس ستانفورد – بينيه للذكاء العام :**

يعد هذا المقياس أكثر مقاييس الذكاء العام انتشاراً واستخداماً ، وقد مر هذا القياس بمراجعات وتعديلات متتالية منذ الطبعة الأولي لمقياس بينيه – سيمون الأصلي عام ١٩١٦. وقد أجريت هذه المراجعات فى الولايات المتحدة الأمريكية ، ولعل أحدثها طبعة عام ١٩٨٤ . ويُستخدم هذا القياس فى قياس الذكاء العام للأعمار من عامين إلى الرشد

، على الرغم من أنه يُستخدم بدرجة أكبر مع الأطفال من الأطفال من عمر عامين إلى ثمانية أعوام ، ويعتمد المقياس على معايير عمرية للأداء ، أي ترتب المجموعات المنفصلة من المفردات بحيث يكون مستوى صعوبتها مناسباً للمستوى العمري المصممة له ، فمجموعة المهام التي تقدم للأعمار من (٢) غلى (٥) أعوام يفصل بينها ستة أشهر عمرية ، أما بالنسبة للأعمار من (٥) إلى (١٤) عاماً فإنه يفصل كل منها عن الأخر عام واحد .

ويوجد اختبار للراشد المتوسط وثلاثة اختبارات للراشد المتفوق ، ويستغرق تطبيق القياس عادة ما بين ٥٠ ، ٧٥ دقيقة .

ويقيس هذا المقياس عدة مهارات يمكن تصنيفها فى سبعة أقسام من المحتويات كالتالي :

* اللغة : تسمية الأشياء فى صور ، تعريف كلمات تسمية كلمات مقفاة .
* الاستدلال : رسم توجهات معرفة سبب سخافة عبارات لفظية .
* الذاكرة : تذكر جمل ، تذكر أرقام .
* التفكير : شرح أمثال سائرة ، توضيح أساس تماثل أشياء .
* الذكاء الاجتماعي : فهم علاقات اجتماعية ،البحث عن سخافات فى صور.
* الاستدلال العددي : إجراء تغييرات ، حل مسائل رياضية حلولاً ذكية .
* النشاط البصري – الحركي تكوين شكل معين على لوحة أشكال ، رسم شكل معين عن طريق النقل ، والمثيرات المستخدمة تشتمل على : كلمات ، وأشياء ، وصور ، والاستجابات تتخذ أنماطاً متعددة ، مثل : التحدث ، الرسم ، إجراء حسابات الكتابة ، وأنشطة حركية .

غير أنه يلاحظ أن هذه المهام السبعة تعتمد اعتماداً أساسياً على الاستدلال المجرد ، ولا تعد القدرة على القراءة أمراً ضرورياً ، ولكن يبدو أن المهارات اللفظية تلعب دوراً كبيراً فى الإجابة على مفردات المقياس .

وهذا يضفي أهمية على مقياس ستانفورد – بينيه لأن نسب الذكاء الناتجة عن المقياس تعد منبئاً جيداً بالنجاح فى العمل المدرسي المعتاد .

ويتضمن المقياس مدى متسعاً من المستويات ، حيث يشتمل على ستة اختبارات فرعية لكل من الأعمال المختلفة ، ويبدأ الفاحص بتقديم المهام التي تقل مباشرة عن العمر الزمني للمتغير والتي يستطيع الإجابة عليها جميعاً إجابة صحيحة ثم ينقله غلى مهام أعلي إلى أن يصل على المستوى الذى عنده يخفق فى الإجابة على جميع الاختبارات الفرعية ، وبذلك يمكن تحديد العمر العقلي القاعدي Basal Age ، والعمر الأقصي Maximal Age .

وتتحدد درجة الإجابة الصحيحة لكل مهمة بمقارنتها بقائمة مطولة من إجابات متنوعة في دليل الاختبار ، ومن ثم تحول الدرجة إلى ميزان عمري Age Scale وذلك بإضافة عدد معين من الشهور لكل إجابة صحيحة ، وعندما تُجمع هذه الشهور على العمر القاعدي نحصل على العمر العقلي للمختبر .

ونظراً لصعوبة مساواة توزيع نسب الذكاء (IQ`s) فى مختلف المستويات العمرية ، وعدم استقامة منحنيات النمو العقلي وافتقادها للمعني بالنسبة لذكاء الراشدين ، حيث أن هناك خلافاً على العمر الزمني الذي يمكن استخدامه فى مقام نسبة الذكاء ، فقد استُخدمت نسبة الذكاء الإنحرافية (DIQ) التي سبق إيضاحها فى الفصل الرابع بدلاً من نسبة الذكاء المعتادة التى نحصل عليها بقسمة العمر العقلي على العمر الزمني ، وضرب الناتج فى (١٠٠) ، وذلك فى الطبعة المعدلة للمقياس عام ١٩٦٠ ، ونسب الذكاء الإنحرافية كما أوضحنا تعد درجات معيارية محولة متوسطها (١٠٠) وانحرافها المعياري (١٦) نقطة ، ويمكن تفسير درجات الاختبار إما باستخدام نسبة الذكاء الإنحرافية (DIQ) التى تدل على المعدل النسبي للنمو العقلي للفرد مقارنة بأقرانه فى

المستوى العمري نفسه ، أو بإستخدام العمر العقلي الذى يدل على المستوى الحالي للنمو العقلي للفرد ، وعلى الرغم من أن كلاً من نسبة الذكاء الإنحرافية والعمر العقلي تعد درجات محولة معيارية لأنها تعتمد على أداء جماعة مرجعية أو عينة تقنين المقياس ، إلا أن العمر العقلي يمكن أيضاً تفسيره من حيث نوع وصعوبة المهمة التي يستطيع الفرد أداءها ، أي يمكن تفسيره على أنه درجة متعلقة بمحتوى معين .

ويُعد مقياس ستانفورد – بينيه من المقاييس التي تستند إلى أسس فنية ، حيث اعُتني ببنائه وتقدير خصائصه السيكومترية ، فمعظم الدراسات التي تناولته بالبحث توصلت إلى أن قيمة معاملات ثبات درجاته تزيد عن (٠.٩٠) ، غير أن هذه القيمة تقل قليلاً بالنسبة لدرجات الأطفال الصغار ، ويفيد هذا القياس كغيره من مقاييس الذكاء فى التنبؤ بالنجاح الدراسي المستقبلي ، حيث إن معاملات الصدق التنبؤى والتلازمي باستخدام محكات، مثل: تقديرات الطلاب فى المدارس، ودرجات الاختبارات التحصيلية تتراوح قيمها بين (٠.٤٠ – ٠.٧٠) وذلك بحسب المستوى العمري للأفراد .

❖ مقاييس ويكسلر للذكاء العام :

تعد مقاييس ويكسلر للذكاء Wechsler Intelligence من المقاييس الفردية أيضاً التى تتميز بأنها حاولت تلافي عيوب مقياس ستانفورد – بينيه السابق ، فمفردات مقياس بينيه Binet صُممت للأطفال ، ولا تناسب الراشدين فى معظم الاحيان ، واستخدام مفهوم العمر العقلي والمعايير العمرية تبين محدودية فائدته للراشدين فى معظم الأحيان ، واستخدام مفهوم العمر العقلي والمعايير العمرية تبين محدودية فائدته للراشدين ، كما أن معايير المقياس استندت إلى بيانات مستمدة من الأطفال وليس الراشدين ، لذلك قام ويكسلر بتطوير مقياس خاص بذكاء الراشدين نُشر فى صورته الأولي الذى أطلق عليها مقياس ويكسلر – بلفيو للذكاء – Wechsler Bellevue Scales عام ١٩٣٩ نسبة غلى مستشفي بلفيو للأمراض النفسية بولاية نيويورك الامريكية الذي كان يعمل فيها خبيراً سيكولوجياً رئيساً ، وكان الهدف من المقياس تجنب استخدام مفهوم العمر العقلي ، وتقليل أثر عامل السرعة فى الاستجابة الذى يعوق الراشدين ، وتصميم مهام أكثر تشويقاً

وملاءمة لهم ، وكذلك زيادة الفائدة التشخيصية للمقياس باشتماله على مقاييس لفظية وادائية .

وقد أجريت مراجعة لهذا المقياس عام ١٩٥٥ وأعيد تقنينه ، وأصبح يسمي مقياس ويكسلر لذكاء الراشدين (WALS) ويناسب الأفراد الذين يبلغون من العمر (١٦) عاماً أو أكبر وتم ترتيب مفردات الاختبار فى اختبارات فرعية منفصلة بدلاً من ترتيبها وفقاً للأعمار الزمنية وأجرى تحويل للدرجات الخام إلى نسب ذكاء انحرافية Deviation IQ (DIQ`s) ، ودُونت التحويلات فى جداول المعايير.

ويستند المقياس إلى تعريف ويكسلر بأن الذكاء هو القدرة العامة للفرد للتصرف والسلوك الهادف ، والتفكير المنطقي ، والتعامل بفاعلية مع بيئته ، وبذلك تبنُي ويكسلر مفهوم الذكاء العام ، وأن المقاييس المختلفة للذكاء ليست مقاييس لأنواع مختلفة من الذكاء ، وأكد أن الذكاء ليس مجرد جمع قدرات تقيسها مفردات الاختبارات الفرعية للمقياس ، وإنما هو تكوين نسبي للقدرات بالإضافة إلى العوامل الدافعية والشخصية التي تؤدي إلى سلوك ذكي .

ويشتمل مقياس ويكسلر لذكاء الراشدين (WAIS) على (١١٠) اختباراً فرعياً ، تقيس ستة لكل من القسمين .

- **المقياس اللفظي Verbal Scale :**

١) المعلومات (تذكر معارف): ويشتمل على أسئلة تتعلق بمعلومات عامة فى موضوعات متعددة تناسب معظم الراشدين فى المجتمع .

٢) الفهم (فهم المعلومات) : ويقيس الحكم الواقعي المعلومات ، مثل : لماذا يدفع الناس الضرائب ؟

٣) الحساب (القدرة العددية) : مسائل كلامية تتطلب أقل قدر ممكن من العمليات الحسابية .

٤) التشابه (الاستدلال) : ويشتمل على (١٣) زوجاً من الكلمات ، ويُطلب من الفرد توضيح أوجه التشابه بين كل زوج منها .

٥) المفردات (القدرة اللفظية) : تتطلب تعريف مفردات لغوية عامة .
٦) وسع الذاكرة للأرقام (الذاكرة) : تتطلب إعادة ذكر ارقام بترتيبها الأصلي ، أو بالترتيب العكسي .

- **مقياس الأداء Performance Scale :**

١) رموز الأرقام (الترميز) : تتطلب مزاوجة الأعداد برموز مجردة .
٢) تكملة الصور (تحليل الأجزاء من الكل): تتطلب تسمية الجزء الناقص من صورة معطاة .
٣) تصميم المجسمات (تحليل مجسم معقد) : يتطلب ترتيب أربعة مجسمات مختلفة الألوان لكي تكون تصميماً معيناً .
٤) ترتيب الصور (الاستدلال غير اللفظي) : يتطلب ترتيب مجموعة من الصور بحيث تحكي قصة ذات معنى.
٥) تجميع الأشياء (توليف الكل من الأجزاء) : يتطلب تجميع أجزاء متاهات تتباين فى درجة تعقدها لكي تُكون شيئاً معيناً ، وذلك بأسرع ما يمكن.

ويستغرق تطبيق مقياس ويكسلر لذكاء الراشدين ساعة واحدة تقريباً ، وتطبق الاختبارات الفرعية واحداً تلو الآخر ، حيث تُقدم مفردات كل من هذه الاختبارات وفقاً لترتيب صعوبتها ، ويتوقف التطبيق إذا أخفق الفرد فى الإجابة عدداً محدداً من المرات ، وبعض المفردات لها زمن محدد للإجابة ، والبعض الآخر غير موقوت .

والقائم بتطبيق هذا المقياس ينبغي أن يكون على درجة عالية من التدرب والخبرة كما هو الحال فى مقياس ستانفورد – بينيه ، ويطبق المقياس بطريقة فردية، ويتم تقدير درجات المختبر بمقارنة إجاباته بقائمة تشتمل على بدائل الإجابات المقبولة ، والمدونة فى دليل المقياس وتعتمد الدرجة على نوعية الإجابة، وتُحول الدرجات الخام التي يحصل عليها الفرد فى كل اختبار فرعي إلى درجات معيارية متوسطها (١٠) وانحرافها المعياري (٣) ومن ثم تجمع الدرجات المعيارية للاختبارات الفرعية ، وتحول إلى ثلاث نسب ذكاء انحرافية ، الأولي للمقياس اللفظي الثانية لمقياس الأداء ، والثالثة للمقياس ككل ، وقد

استخدم ويكسلر نسبة الذكاء الانحرافية (DIQ) التى متوسطها (١٠٠) وانحرافها المعياري (١٥) بدلاً من (١٦) المستخدمة فى مقياس بينيه .

ونظراً لأن الأداء فى المقياس يتدهور تدريجياً مع تقدم العمر ، فإن نسبة الذكاء الانحرافية تم حسابها لكل مجموعة عمرية على حده ، وبذلك تدل هذه النسبة على مركز الفرد بالنسبة للجماعة المرجعية التى تشتمل على عينة من الأفراد فى نفس المجموعة العمرية للفرد .

وقد أوضحت البيانات التى جمعها ويكسلر فيما يتعلق بالمقياس أن درجات المقياسين الفرعيين ، والمقياس ككل تتميز بقدر مرتفع من الثبات يساوي أو يزيد عن (٠.٩٣)، كما أن درجات المقياس ككل ترتبط بدرجات مقياس ستانفورد – بينيه ارتباطاً تبلغ قيمته حوالي (٠.٨٥)، كما تتراوح قيمة الارتباط بين المقياس اللفظي ومقياس الأداء بين (٠.٧٧ – ٠.٨١) في عينات ثلاث مختلفة من الأفراد (Wechsler,1955).

ومن الجدير بالذكر أن ويكسلر أعد مقياسين آخرين أحدهما لذكاء الأطفال Wechsler Scale for Children (Wisc) عام ١٩٤٩ للأعمار من (٥) أعوام إلى (١٥) عاماً ، وأجريت مراجعه على الاختبار ، وأصبح يرمز له (WISC–R) ويناسب الأعمار من ٦ أعوام ، ٦ شهور إلى (١٦) عاماً ، والآخر لذكاء أطفال ما قبل المدرسة (Washier Preschool and Primary Scale of Intelligence (WPPSI) عام ١٩٦٧ ، ويناسب الأطفال الذين تتراوح أعمارهم بين (٤) أعوام ، وستة أعوام ونصف .

واستند ويكسلر فى بناء هذين المقياسين إلى الأسس التى استند إليها فى بناء مقياس الذكاء للراشدين ، حيث اشتمل كل من المقياسين المذكورين على مقياسين فرعيين أحدهما لفظي ، ويحتوى على ستة اختبارات فرعية ، والآخر أداء ، ويحتوى على خمسة اختبارات فرعية

❖ اختبارات الذكاء العام الجماعية :

سبق أن أوضحنا أن الاختبارات الجماعية Group Test تتميز على الاختبارات الفردية فى فعالية العملية الاختبارية وخفض كلفة تطبيق الاختبارات ، حيث إنها تطبق

على عدد كبير من الأفراد فى وقت واحد ، وتوجد أنواع متعددة من اختبارات الذكاء العام الجماعية ، لعل أكثرها استخداماً الاختبارات التالية .

١) **اختبار أوتيس– لينون للقدرة العقلية (OLMAT):** ويناسب المراحل من الحضانة إلى نهاية التعليم الثانوي، ويشتمل على ستة مستويات تناظر الصفوف الدراسية المختلفة ، ويستغرق تطبيق اختبار المستوى الابتدائي حوالي (٣٠) دقيقة، أما اختبارات المستويات الأخري فيستغرق تطبيق كل منها حوالي ساعة واحدة، ويشتمل الاختبار فى صورته الأخيرة على أنواع مختلفة من المفردات التى تُرتب حلزونياً أو دائرياً فى الاختبار ، وتتعلق بالجوانب اللفظية والتعليمية للعامل العام (g) فى نظرية سبيرمان لبنية الذكاء، وينتج عن الاختبار درجة خام كلية واحدة يتم تحويلها إلى نسبة ذكاء انحرافيه (DIQ)، ورتب مئينية للصفوف المدرسية ، وتساعيات معيارية ، واقتصرت عينة تقنين الاختبار على مجتمع طلاب المدارس فقط وتتراوح قيم معامل ثبات درجات الاختبار بطريقة الصيغ المتكافئة بين (٠.٨٣) إلى ما يزيد عن (٠.٩٠)، وتزداد قيمة معامل الثبات بزيادة العمر الزمني للأفراد ، وهناك بيانات للتحقق من صدق محتوى ، وصدق التكوين الفرضي للاختبار.

٢) **اختبار لورج– ثورنديك للذكاء (LTIT):** وهو من الاختبارات متعددة المستويات أيضاً من الحضانة إلى نهاية المرحلة الثانوية ، ويستغرق تطبيقه حوالي ساعتين ، وينتج عنه درجات للجانبين اللفظي ، وغير اللفظي ، وكذلك درجات كلية ، ويتميز كلية ، ويتميز هذا الاختبار بأن ثبات درجات الجانب اللفظي يتراوح بين (٠.٨٣ – ٠.٩١) ، وثبات درجات الجانب غير اللفظي يتراوح بين (٠.٨٠ – ٠.٨٨) باستخدام طريقة الصيغ المتكافئة ، ويتم تحويل الدرجات الخام إلى نسبة ذكاء انحرافيه لكل مستوى على حدة ، ويتراوح مقدار الارتباط بين درجات هذه الاختبارات ، ودرجات الاختبارات التحصيلية بين (٠.٦٠ ، ٠.٧٠).

٣) اختبار القدرات المعرفية (CAT) : ويعد أيضاً من الاختبارات متعددة المستويات من (٣) أعوام إلى نهاية المرحلة الثانوية، وتوجد طبعة منه تناسب الأعمار الأقل من (٣) أعوام، وتشتمل على ثلاث بطاريات منفصلة يستغرق تطبيق كل فيها (٣٥)

دقيقة بالإضافة إلى زمن التعليمات الذى يبلغ حوالي (١٥) دقيقة ، وتشتمل هذه البطاريات على جوانب لفظية (مفردات لغوية ، إكمال جمل ، تصنيف لفظي ، متشابهات لفظية) ، وجوانب كمية (سلاسل أعداد ،علاقات كمية ، تكوين معادلات) ، وجوانب غير لفظية (تصنيف أشكال متشابهات شكلية ، تكوين أشكال)

أما الطبعة الخاصة بالأعمار الأقل من (٣) أعوام فإنها تستخدم مواداً بصرية، وتعليمات شفوية ، وتتضمن اختبارات فرعية تشتمل على مفردات لغوية شفوية، ومفاهيم علاقية ، ومفاهيم كمية ، ويطبق هذا الاختبار مفرده تلو الأخري لضمان عدم تأثره بسرعة الإجابة .

وقد أعدت معايير مئينية ، تساعيات معيارية ، وغيرها من المعايير المستمدة من عينة قوامها ٢٠٠.٠٠٠ طالب من مختلف الصفوف المدرسية ، ويتخطي معامل ثبات درجاتها القيمة (٠.٩٠).

غير أن الارتباط بين درجات اختباراته الفرعية تتراوح بين (٠.٤٠ ، ٠.٧٧) ، ولم يوضح دليل الاختبار بيانات كافية للتحقق من صدقه .

٤) اختبار القدرات العقلية الأولية (PMA): أعد ثيرستون هذا الاختبار استناداً إلى نظريته التي سبق ذكرها فى مستهل هذا الفصل والتى تُعرف بنظرية العوامل المتعددة للذكاء ، ويعد هذا الاختبار بمثابة عدة بطاريات تناسب المراحل : الحضانة والصف الأول الابتدائي والصفوف من الثاني إلى الرابع، ومن الرابع إلى السادس ، ومن السادس إلى التاسع ، ومن التاسع غلى نهاية المرحلة الثانوية ، وتقيس خمس قدرات هي :

معاني الكلمات : التعامل مع الأعداد بيسر وسهولة ، العلاقات المكانية ، السرعة الإدراكية ، الاستدلال أو التفكير المنطقي ، ويقتصر الاختبار الأخير على الصفوف من الرابع إلى نهاية المرحلة الثانوية ، ويستغرق تطبيق الاختبار حوالي ساعة ونصف وتُقرر الدرجات الخام لكل اختبار فرعي وللبطارية ككل ، ثم تُحوَّل إلى نسبة ذكاء انحرافيه ، ورتب مئينية.

٥) اختبار كاتل للذكاء غير المتحيز ثقافياً : ويهدف هذا الاختبار لقياس الذكاء السائل Fluid Intelligence ، أي العامل العام (g) الذى سبق إيضاحه ، ويتمثل فى السلوك العقلي المتكيف فى مواقف غير مألوفة ، بحيث لا يسترشد الفرد فى استجابته لها بمهارات تعلم سابقة ، ومن أمثلة ذلك : مادة لا معنى لها ، ومواد غير مألوفة ، ومواد شائعة الاستخدام ، وتُعرض هذه المثيرات فى أشكال تتطلب تسلسل معين ، أو تصنيف أو مصفوفات ، أو تحديد مواضع مناسبة لأشكال معطاة .

❖ تعقيب على اختبارات الذكاء العام :

يتضح مما سبق أنه توجد كثير من الاختبارات الجيدة التى تقيس القدرة العقلية العامة تناسب متطلبات المواقف التطبيقية المختلفة ، وبعض هذه الاختبارات يطبق بطريقة فردية مثل : مقياس ستانفورد – بينيه ، ومقاييس ويسكلر ، والبعض الآخر يطبق بطريقة جماعية ، ويوجد عدد كبير من هذه الاختبارات يتراوح بين تلك التي تناسب أطفال ما قبل المدرسة ، وتلك التي تناسب الراشد المتفوق ، وبعض هذه الاختبارات يكون على شكل بطاريات تشتمل على اختبارات فرعية تقيس كل من الجوانب اللفظية ، والجوانب غير اللفظية ، وينتج عنها درجات لكل من هذه الجوانب ، وكذلك درجة كلية للذكاء العام ، أما البعض الآخر فإنه لا يعتمد على اللغة ، وإنما على الصور والأشكال والمصفوفات ، وبعضها أيضاً يكون متحرراً من الثقافة بدرجة ما .

ويعتمد تفسير معظم هذه الاختبارات على الدرجات المعيارية المحولة مثل : نسب الذكاء الانحرافية ، والتساعيات المعيارية ، وكذلك على المئينيات ، وغيرها من الدرجات المحوَّلة ، ومن الجدير بالذكر أن بعض هذه الاختبارات تم تعريبه وتقنينه على بيئات عربية .

اختبارات الاستعدادات الخاصة

- مقدمة
- أنواع اختبارات الاستعدادات الخاصة
- استخدامات اختبارات الاستعدادات الخاصة

أولاً : بطاريات الاستعدادات الخاصة المتعددة

- اختبارات الاستعدادات الفارقة
- بطارية اختبارات الاستعداد العام

ثانياً : اختبارات استعدادات وقدرة التوعية

- اختبارات الاستعدادات الميكانيكية

❖ **اختبارات الاستعداد الكتابي**

❖ **اختبارات الاستعدادات الموسيقية والفنية**

❖ مقدمة

أوضحنا فى الفصلين السابقين كيفية قياس التحصيل والذكاء ، وذكرنا أن اختبارات التحصيل تهدف لقياس الأداء الحالي للفرد أو نواتج تعلمه فى مدة دراسية معينة ، بينما تهدف اختبارات الذكاء لقياس القدرة العقلية العامة للفرد ، والتنبؤ بتحصيله الأكاديمي باستخدام درجة كلية واحدة فى الاختبار ، غير إن بعض اختبارات الذكاء كما رأينا فى الفصل السابق ينتج عنه أكثر من درجة واحدة للفرد اعتماداً على نظريات الذكاء متعدد العوامل .

فاختبارات الذكاء العام تقيس مجموعة متنوعة من القدرات ، لذلك يتبين أن فائدتها فى التنبؤ بمختلف المحكات ليست كبيرة ، ويري "كرونباك Cronbach, 1970) " أن الاختبار الأقل اتساعاً فى محتواه يمكن الإفادة منه بدرجة أفضل فى التنبؤ بمحكات محدودة .

لذلك اقترح بعض علماء القياس أن اختبارات الذكاء العام ينبغي أن تطبق إلى جانب اختبارات الاستعدادات أو القدرات الكامنة ، والتنبؤ بمدى استفادة الفرد من تدريب فى مجال مهاري معين ، ولعل هذا أيضاً هو اختلاف اختبارات الذكاء العام ، حيث إن

الهدف من اختبارات الاستعدادات التنبؤ بالنجاح فى مجال أكثر محدودية مما يهدف إليه التنبؤ من درجات اختبارات الذكاء .

فاختبارات الاستعدادات الخاصة تقيس مهارات نوعية أو معارف تتعلق بمجال معين ، وهذه المهارات والمعارف تعد ضرورية للأداء المستقبلي فى هذا المجال ، لذلك فإن اختبار الاستعداد الميكانيكي مثلاً ربما يشتمل على مهام تجميع أشياء مفككة كالجرس الكهربائي ، أو جزء من أجزاء جهاز أو أله ، أو على مهام تتطلب التمييز بين مثيرات مسجلة تتباين فى درجة الصوت وعلوه ، وسرعة النغمة ، والإيقاع ، وغير ذلك ، وتعتمد هذه المهارات على تحليل العمل المطلوب Job Analysis لتحديد نوعيات الأداء الناجح فى مهنة أو مجال معين .

ومن ثم يتم بناء اختبار الاستعدادات لقياس هذه المهارات أو القدرات المحددة بموضوعية ودقة .

ويتناول هذا الفصل أنواع اختبارات الاستعدادات الخاصة ، واستخداماتها ، وبعض بطاريات اختبرات الاستعدادات المتعددة ، وأمثلة لبعض اختبارات استعدادات وقدرات نوعية تتضمن اختبارات الاستعدادات الميكانيكية ، واختبارات الاستعداد الكتابي ، واختبارات الاستعدادات الموسيقية والفنية .

❖ أنواع اختبارات الاستعدادات الخاصة :

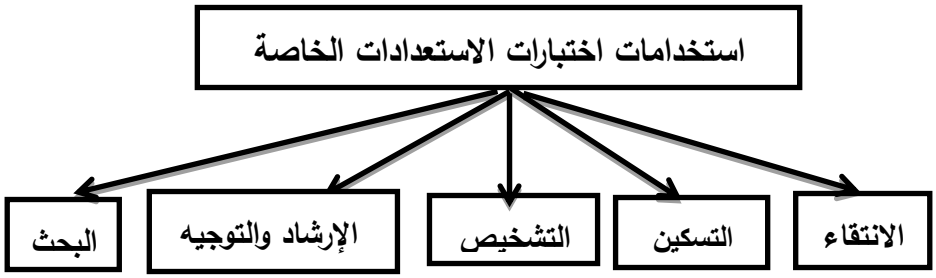
تختلف اختبارات الاستعدادات الخاصة فى درجات خصوصية أو نوعية القدرات أو المهارات التى يقيسها وتتطلبها مهن أو أعمال أو أنشطة معينة ، وبعض هذه الاختبارات يُستخدم فى التنبؤ بالنجاح أو الفشل فى دراسات معينة ، أو فى تعلم مجال دراسي تخصصي ، كما يُستخدم فى التنبؤ بالأداء فى مهنة أو عمل معين ، ففي الحالة الأولي نحن نهتم بالاستعداد الدراسي Scholastic Aptitude ، وفي الحالة الثانية نهتم بالاستعداد المهني Vocational Aptitude ، وبعض اختبارات الاستعدادات يمكن أن يحقق هدفين معاً ، حيث يقيس الذكاء العام ، وكذلك الاستعداد الدراسي ، ويوجد عدد كبير من اختبارات الاستعداد الدراسي في مجالات متنوعة ، مثل الفنون الحرة ، والطب ،

والتمريض ، وغير ذلك ، كما يوجد أيضاً عدد كبير من اختبارات الاستعداد المهني الميكانيكي ، واليدوي ، والكتابي أو التجاري ، والموسيقي والفني ، وغيرها .

وتوجد أيضاً بطاريات واختبارات استعداد متعددة Multiple Aptitude Tests Batteries تُستخدم فى أغرض الإرشاد التربوي والتوجيه المهني ، حيث إن كلاً منها يشتمل على عدة اختبارات تقيس استعدادات خاصة مختلفة ، وبعض هذه الاختبارات يُضم معاً فى بطاريات تزودنا بمعلومات تتعلق بكل من الاستعدادات الدراسية والمهنية .

❖ استخدامات اختبارات الاستعدادات الخاصة

تُستخدم اختبارات الاستعدادات الخاصة بكثرة في المؤسسات التعليمية والمهنية لأغراض متعددة يوضحها الشكل التخطيطي (٧-١) التالي :



شكل (٧-١) يوضح أغراض استخدامات اختبارات الاستعدادات الخاصة

وفيما يلي توضيحاً لكل من هذه الاستعدادات :

أغراض الانتقاء : وتُعد من أهم استخدامات اختبارات الاستعدادات غير إن استخدام نتائج هذه الاختبارات فى أغراض الانتقاء Selection لا يضمن اتخاذ قرارات صائبة فيما يتعلق بقبول الأفراد أو رفضهم فى جميع الحالات ، بل نتوقع بعض الخطأ فى هذه القرارات .

لذلك يهتم القائمون ببناء اختبار استعداد معين التحقق من الصدق المرتبط بمحك للاختبار بحيث تكون درجاته بمحك معين ارتباطاً مرتفعاً .

أغراض التسكين : وهنا يُفترض أن الأفراد قد تم انتقاؤهم فى مؤسسة تربوية أو مهنية معينة ، ويكون مطلوباً تسكينهم فى مجال دراسي ، أو عمل معين مناسباً لهم ومتفقاً مع ميولهم وكثير من المشكلات المتعلقة باتخاذ قرارات الانتقاء تتواجد أيضاً عند استخدام اختبارات الاستعدادات فى أغراض التسكين Placement.

فالبعض يرى أن اتخاذ قرارات التسكين تكون لصالح المؤسسة ، لذلك ربما يتم تسكين الأفراد الذين يتبين أن لديهم الاستعدادات الجيدة فى عمل معين ، مما يجعل هؤلاء الأفراد يضطرون لقبول هذا العمل على الرغم من أنهم يرون أنهم لا يميلون إليه ، أو أنه غير مناسب لهم ، وهذه إحدى مشكلات استخدام اختبارات الاستعدادات في أغراض التسكين .

لذلك ينبغي أن تكون قرارات الانتقاء مسئولية مشتركة بين المؤسسة والأفراد المتقدمين للدراسة أو العمل .

أغراض التشخيص : تفيد اختبارات الاستعدادات فائدة كبيرة فى تشخيص مشكلات الطلاب إذ تطبق اختبارات التهيؤ Readiness Tests فى المرحلة الابتدائية للتعرف على التلاميذ الذين يُحتمل مواجهتهم بعض الصعوبات أثناء تعلم مواد دراسية معينة ، مثل : القراءة أو الحساب ، كما سبق أن أوضحنا فى الفصل الخامس .

وكذلك فى المدرسة الثانوية ربما تُطبق اختبارات استعدادات فى اللغات الأجنبية لتحديد الطلاب الذين لديهم المهارات اللازمة لتعلم لغات أجنبية معينة ، وتحليل أخطائهم فى هذه الاختبارات يساعد فى تصميم الإجراءات العلاجية المناسبة .

أغراض الإرشاد والتوجيه : تفيد أيضا اختبارات الاستعدادات فى الإرشاد والتوجيه التربوى ، والمهنى ، والشخصى وتستخدم عادة بطاريات اختبارات الاستعدادات متعددة العوامل Multifactor Aptitude Tests فى الإرشاد والتربوى Educational Counseling لتعريف الطلاب بقدراتهم المتمايزة . وعلى الرغم من أن قياس هذه القدرات المتمايزة لا يؤدى بالضرورة إلى تحسين التنبؤ ، إلا أنه يجعل الطالب مدركا لجوانب أو مجالات تميزه ، ولكن ينبغى الحيطة من إساءة استخدام نتائج بطاريات

اختبارات الاستعدادات فى هذا الغرض ، إذ لا يجوز للمرشد التربوى تفسير درجات منخفضة فى اختبار معين بطريقة تؤثر سلبيا فى مفهوم الذات لدى الطالب ، وإنما ينبغى توظيف الدرجات المرتفعة فى الاختبارات الأخرى فى حفز دافعية الطالب وإشعاره بقدرته كما ينبغى التأكد على أن درجات اختبارات الاستعدادات بمفردها لتعد كافية للإرشاد والتوجيه التربوى والمهنى ، وإنما يجب أيضا الأخذ بعين الاعتبار أداء الفرد السابق ، وميوله والعوامل الموقفية المختلفة .

أغراض البحث : يمكن استخدام اختبارات الاستعدادات فى كثير من البحوث النفسية والتربوية فدرجات هذه الاختبارات تستخدم عادة كمتغيرات مستقلة فى تصميم البحوث على الرغم من إمكانية استخدام هذه الدرجات كمتغيرات تابعة، كما هو الحال فى دراسة تأثيرات البيئة فى الذكاء كما أن البحوث تجرى حول طبيعة الذكاء والاستعدادات الخاصة تؤدى إلى تطوير الاختبارات ذاتها ، وتقدم بيانات مهمة تفيد فى إعادة النظر فى بعض نظريات التكوين العقلى ، وفهم النمو المعرفى .

(أولا) بطاريات الاستعدادات الخاصة المتعددة :

تعد بطاريات الاستعدادات الخاصة المتعددة Multiaptitude Batteries من المقاييس التى استندت فى بنائها على أسلوب التحليل العاملى الذى سبق أن أشرنا إليه فى مستهل الفصل السابق من خلال تناولنا نظريات بنية الذكاء ، فقد اعتمد ثيرستون ، وجيلفورد ، وغيرهما على هذا الأسلوب فى التوصل إلى الأبعاد أو العوامل التى أطلق عليها القدرات العقلية . فمحاولات عزل وتحديد عوامل نوعية للذكاء العام أثرت فى جهود علماء القياس المتعلقة بتحديد الاستعدادات الخاصة وقياسها .

فقد حاول هؤلاء العلماء بناء بطاريات من الاختبارات لقياس قدرات أو سمات متعددة تتطلبها أنواع مختلفة من الأعمال أو الأنشطة ، فإذا ارتبطت درجات الاختبارات التى تشتمل عليها البطارية بمهن معينة أو مجموعات من المهن ، فإنه يمكن تقدير أوزان لهذه الدرجات وفقا لأهميتها فى هذه المهن (Alken,1976 & Noll , 1972) .

لذلك تسمى هذه البطاريات أحيانا بطاريات استعدادات فارقة ،أو بطاريات استعدادات عاملية .

والاختبارات التى تشتمل عليها هذه البطاريات تقيس عادة عدة قدرات متسقة ومعاييرها تسمح بإجراء مقارنات بين قدرات الفرد نفسه ، لذلك تعتمد معايير اختبارات البطارية جميعها على نفس عينة التقنين .

وتفيد هذه البطاريات فى الإرشاد والتوجيه والمهنى ، وانتقاء الأفراد وتصنيفهم، حيث يكون هناك اهتمام بقياس مهارات الفرد ومعارفه فى مجالات متعددة، وتعتمد معظم هذه البطاريات عادة على مفردات اختبارية تتطلب الورقة والقلم، لذلك يمكن تطبيقها على مجموعة كبيرة من الأفراد فى وقت واحد ، والمعلومات المستمدة من هذه البطاريات تفيد الطالب فى معرفة جوانب قوته وضعفه ، ويمكن أن يستند إليها فى اتخاذ قرارات تربوية ومهنية تتعلق بمستقبله غير إنه لا ينصح عادة باستخدام بطاريات الاستعدادات المتعددة قبل وصول الطالب إلى المرحلة التعليمية المتوسطة نظرا لأن قدرات الأطفال تميل إلى العمومية فى المرحلة الابتدائية .

ويوجد العديد من بطاريات الاستعدادات المتعددة ، غير أن أقدمها وأكثرها شيوعا واستخداما البطاريتين التاليتين :

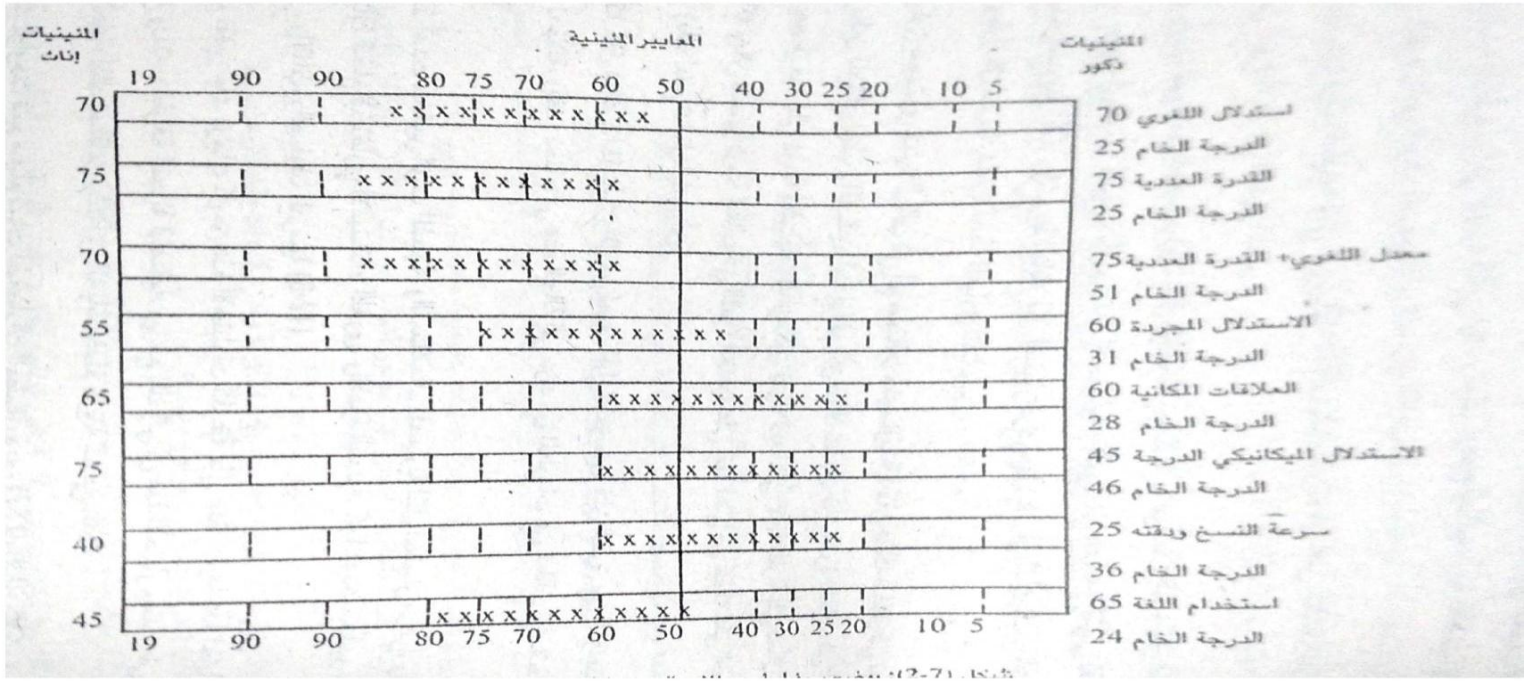
(١) اختبارات الاستعدادات الفارقة(DAT) Differential Aptitude Tests :

أعد هذه البطارية من الاختبارات بينيت ، وسيشور ، وويزمان , Bennett) (Seashore & Wesman , 1974 وصدرت طبعتها الخامسة عام 1973، وأجريت عليها بعض التعديلات عام 1982. والغرض الأساسي من هذه البطارية تقديم معلومات تتعلق ببروفيل قدرات متعددة لاستخدامه فى التوجيه والإرشاد التربوى للطلاب فى المرحلتين المتوسطة والثانوية لكى يتخذ كل منهم فى ضوء ذلك قرارات تتعلق باختياراته المهنية المستقبلية.

وتشتمل البطارية على ثمانية اختبارات فرعية يطبق كل منها على حدة ، وهى كالتالي :

(١) الاستدلال اللغوي (تكملة جمل ناقصة باستخدام كلمات معطاة).
(٢) القدرة العددية (حساب وعمليات حسابية).
(٣) الاستدلال المجرد (إكمال متسلسلات شكلية).
(٤) العلاقات المكانية (تكوين شكل معين من أنماط شكلية معطاة).
(٥) الاستدلال الميكانيكي (إجابة أسئلة تتعلق بمبادئ ميكانيكية ممثلة بصور معطاة) .
(٦) سرعة النسخ ودقته (التعرف على مجموعات معينة مركبة من حروف وأرقام).
(٧) استخدام اللغة (التهجى ، والنحو ، وعلامات الترقيم) .

وإجابة بعض مفردات هذه الاختبارات تتطلب الاختيار من متعدد ، ومجموع درجات اختبار الاستدلال اللغوي ، واختبار القدرة العددية يعطى مؤشرا للاستعداد الدراسي ، ويستغرق تطبيق هذه البطارية ثلاث ساعات تقريبا ، وهى تعتمد على القوة ، وتقرر النتائج للطلاب فى شكل بروفيل كالتالي :



شكل (٧-٢) يوضح بوفيل أحد طلاب الصف الأول الثانوي فى بطارية اختبارات الاستعدادات الفارقة

ويتضح من شكل (٧-٢) أن الدرجات الخام لكل اختبار فرعى مدونة فى العمود الأول ، والرتب المئينية المناظرة لها فى العمود الثانى ، وهذه الرتب توضح للطالب رتبه بالنسبة لأقرانه من نفس النوع ونفس الصف المدرسى على المستوى القومي .

بينما دونت الرتب المئينية التى تقارن أداء الطالب بأداء أقرانه فى نفس الصف المدرسى ولكن من النوع الآخر فى العمود الأخير.

وقد رسم شريط من العلامات (X) يمثل الرتبة المئينية للطلاب من نفس النوع لكل اختبار فرعى، بحيث أن الرتبة المئينية لكل من هذه الاختبارات المدونة فى العمود الثانى تقع فى منتصف الشريط (X)، والشريط المناظر لكل اختبار فرعى يدل على أن ثبات الدرجات ليس تاما، وطول كل شريط يعكس الخطأ المعيارى للقياس .

ويكون مستوى قدرة الطالب فى أى محالين متساويا إذا كان هناك تداخل فى نهايات الشريطين الممثلين لهذين المجالين، أما إذا لم يكن هناك تداخل ، فإن قدرة الطالب فى المجال الذى حصل فيه على أعلى يحتمل أن تكون قدرته فيه أعلى، فإذا نظرنا إلى شكل (٧-٢) نجد أنه يمكن تفسير درجة الطالب فى القدرة العددية على أنها أعلى من قدرته على استخدام اللغة، وذلك لعدم تداخل الشريطين.

وإذا كان هناك تداخل ولكن أقل من نصف طول الشريطين، فإن الفرق بين الدرجتين ربما يكون له دلالة، ولكننا نحتاج فى هذه الحالة إلى مزيد من المعلومات عن الطالب قبل اتخاذ قرار بشأنه.

وقد أدت معايير الرتب المئينية، والتساعيات المعيارية لكل اختبار فرعى، ولمجموع الدرجات الخام لاختبار الاستعداد اللغوي، والقدرة العددية، كما أعدت معايير منفصلة للذكور ، والإناث ، والصفوف المدرسية (8-12) .

وتم تقدير ثبات درجات كل من هذه الاختبارات الفرعية على حدة – فيما عدا اختبار سرعة النسخ ودقته – بطريقة التجزئة النصفية. ويرجع ذلك إلى أن هذا الاختبار يعتمد على السرعة فى الإجابة ، لذلك تم تقدير ثبات درجاته باستخدام طريقة الصيغ المتكافئة.

وقد تراوحت قيم ثبات اختبارات البطارية بين (0.87 /0.97). للذكور، وبين (0.96/0.83) للإناث، وقد أجرى العديد من دراسات الصدق التنبؤي لهذه البطارية من الاختبارات ، وذلك بإيجاد الارتباطات بين درجات الاختبارات الفرعية وتقديرات الطلاب فى مختلف المواد الدراسية ، ودرجاتهم فى اختبارات الذكاء ، وكذلك درجاتهم فى بطاريات أخرى للاستعدادات ، واختبارات تحصيل موضوعية ، وبيانات مهنية، وقد بينت النتائج بعامة أنه توجد علاقة تمايزية بين هذه الاختبارات والتحصيل فى مختلف المواد الدراسية ، ومستويات القدرة العقلية ، والتصنيفات المهنية .

(٢) بطارية اختبارات الاستعداد العام: General Aptitude Test Battery (GATB)

أعد هذه البطارية مكتب خدمات التوظيف بالولايات المتحدة الأمريكية، واعتمد بناؤها على تحليل مختلف الأعمال، ونتائج التحليل العاملى لعدد كبير من الاختبارات المختلفة المتعلقة بأعمال معينة، وبلغ عددها (59) اختبارا .

واختير فقط (10)عوامل قدرات مختلفة لكى تقيسها البطارية وتشتمل البطارية فى طبعتها الحديثة على (12) اختبارا فرعيا، ينتج عنها درجات لتسعة عوامل مختلفة ، واحد هذه العوامل هو القدرة العقلية العامة (G) الذى ينتج من درجات ثلاثة اختبارات فرعية هى: المفردات اللغوية، والاستدلال الحسابى، والتصور المكانى فى ثلاثة أبعاد .

وفيما يلى توضيحا للاختبارات الفرعية التى تشتمل عليها البطارية :

(١) الذكاء :وهو مركب من الاختبارات الفرعية الثلاثة المشار إليها.

(٢) القدرة اللفظية : مفردات لغوية ، استخدام المترادفات ، والمناقضات فى المعنى .

(٣) القدرة العددية : إجراء عمليات حسابية ، واستدلال حسابي .

(٤) القدرة المكانية : القدرة على تصور أشكال فى الفراغ ثلاثى الأبعاد ، ولكنها عرضت فى بعدين .

(٥) إدراك صيغ وأنماط بصرية : مزاوجة رسوم لأدوات وأشكال هندسية .

(٦) الإدراك الكتابى : مزاوجة أسماء .

(٧) التآزر الحركي : القدرة على وضع علامات معينة داخل مجموعة من المربعات.

(٨) مهارة الأصابع : القدرة على تجميع وفك مسامير البرشام، والحلقات المعدنية.

(٩) المهارة اليدوية : القدرة على النقل والتركيب فى لوحة مثقبة .

وبمقارنة هذه الاختبارات بالاختبارات الفرعية التى اشتملت عليها بطارية الاستعدادات الفارقة نلاحظ أن البطارية الأخيرة تحتوى على اختبارات للفهم الميكانيكى واللغة وهو مالم تشتمل عليه بطارية اختبارات الاستعداد العام فى حين أن هذه البطارية اشتملت على اختبارات حركية لم تشتمل عليها بطارية الاستعدادات الفارقة لذلك فإن بطارية اختبارات الاستعداد العام موجهة للعمل بدرجة أكبر من توجهها للأداء الأكاديمي ، وقد وجد أن الاختبارات الحركية الثلاثة التي تشتمل عليها هذه البطارية ترتبط فيما بينها ارتباطا مرتفعا نسبيا، ولكنها في الواقع لا ترتبط بالاختبارات الفرعية الأخرى .

كما وجدت ارتباطات مرتفعة إلى حد ما بين العوامل المتناظرة التى تقيسها كل من البطاريتين ، والفرق الرئيس بينهما هو أن اختبارات البطارية الأولى

(DAT) تعد فى معظمها اختبارات قوة Power Tests ، أما اختبارات البطارية الثانية (GATB) فتعد من الاختبارات التى تعتمد إلى حد كبير على السرعة Speed Tests ، لذلك تفيد هذه البطارية فى التنبؤ بمحكات تتطلب التناسق الحركى البصرى .

والدرجات الخام للعوامل التسعة المستمدة من الاختبارات الفرعية تحول إلى رتب مئينية أو درجات معيارية متوسطها (100) ، وانحرافها المعيارى (20) ، ويمكن مقارنة الدرجات المعيارية للفرد بتلك المستمدة من أنماط من الاستعدادات المهنية الناتجة عن تحليل درجات الأفراد فى هذه البطارية الذين ينتمون إلى أكثر من (٨٠٠) مهنة أو عمل ، وحُددت درجات قطع Cutting Scores لكل اختبار ينبغي أن يصل إليها الفرد للأداء الفاعل فى المهنة المعنية .

ويستغرق تطبيق هذه البطارية قرابة الساعتين ونصف ، وتناسب طلاب نهاية المرحلة الثانوية ، وكذلك الراشدين ، وقد تراوحت قيم معاملات ثبات درجات الاختبارات الفرعية باستخدام طريقتي التجزئة النصفية ، والصيغ المتكافئة بين (٠.٨٠ ، ٠.٩٠) ، والخطأ العياري للقياس (٧) نقاط بالنسبة للدرجات المعيارية، كما أن صدق الدرجات التسع ، وأنماط القدرات المهنية التى عددها (٣٦) نمطاً فى مقابل محكات مهنية وأكاديمية تراوح بين (صفر – ٠.٩٠) .

وفى ضوء هذه المعلومات يمكن اعتبار هذه البطارية (GATB) من أفضل البطاريات التى تُستخدم فى الإرشاد المهني ، وانتقاء الأفراد فى حين أن بطارية الاستعدادات الفارقة (DAT) يبدو أنها تناسب بدرجة أفضل أغراض الإرشاد التربوي .

ومن الجدير بالذكر أن هناك بطاريات أخري مثل : بطارية فلاناجان لتصنيف الاستعدادات Flanagan Aptitude Classificatian (FACT) ،

وبطارية فلاناجان الصناعية Flanagan Industrial Tests (FIT). وبطارية جليلفورد – زيمرمان لمستح الاستعدادات Guilford Zimmerman Aptitude Survey، وجميعها تقيس جوانب القوة والضعف فى الاستعدادات المختلفة لدى الأفراد .

ثانياً : اختبارات استعدادات وقدرة التوعية:

تقيس بطاريات الاستعدادات المتعددة التي أوضحنا مجموعة من القدرات في بطارية اختبارات واحدة متكاملة ، غير إنه يمكن أيضاً بناء اختبارات منفصلة لقياس استعدادات أو قدرات نوعية Specific Altitudes لذلك سوف أوضح فيما يلي بعض أهم هذه الاختبارات التي يُستخدم بعضها فى المواقف الأكاديمية، وبعض آخر ى المواقف المهنية وفى أي الحالات يهتم مصمم الاختبار بقياس قدرة نوعية محددة قياساً أكثر عمقاً بدلاً من اهتمامه بتقديم بروفيل لعدة قدرات.

وهذا يجعل الاختبار أكثر صدقاً لتعلقه بغرض أو موقف معين ، غير إنه نظراً لأن اختبارات الاستعدادات النوعية المختلفة تستند معاييرها إلى مجتمعات مختلفة من الأفراد ، فإن البيانات التى يُستخدم فى حساب المعايير لا تكون متقاربة ، مما يؤدى إلى صعوبات فى تفسير الدرجات وبخاصة إذا أُجريت مقارنات فيما بينها .

(١) اختبارات الاستعدادات الميكانيكية Mechanical Aptitude Tests :

يرى بعض خبراء القياس أن الاستعداد الميكانيكي يُعد سمة أحادية البُعد يختلف مقدراها من فرد إلى أخر ، بينما يري بعض أخر أنه سمة متعددة الأبعاد ، أي تشتمل على عدد من القدرات المستقلة نسبياً ، ولعل هذا المدخل الذى يمايز بين مكونات الاستعداد الميكانيكي هو الذى حظي بتأييد أكبر ، وهذا يعني أن

قياس هذا الاستعداد بدرجة افضل يتطلب تطبيق عدة اختبارات أو اختبارات فرعية وليس اختباراً واحداً .

وبعض هذه الاختبارات يقيس آداءات أو مهارات يدوية أو نفسية حركية معينة، وبعض أخر يقيس معلومات ميكانيكية تعتمد على الورقة والقلم ، وفيما يلي بعض اختبارات الاستعدادات الميكانيكية :

اختبار التجمع الميكانيكي Mechanical Assembly Test :

تتطلب هذه الاختبارات تجميع الأجزاء ، وتعتمد درجة الفرد على عدد الأجزاء التى يمكنه تجميعها فى زمن محدد ، ولعل اختبار مينسوتا للتجميع Minnesota Assembly Test يعد أحد أهم هذه الاختبارات ، ويشتمل على (٣٣) شيئاً أو شكلاً ميكانيكياً مفككاً ينبغي على الفرد تجميعها فى زمن محدد ، غير أن اختبارات التجميع الميكانيكي كلفتها عالية وصدقها منخفض حيث لا ترتبط درجات اختبار مينسوتا بتقديرات الطلاب .

ونظراً لأن معايير هذا الاختبار مستمدة من طلاب المدارس الثانوية ، فإن المعلومات المتعلقة بقدرة الاختبار على التنبؤ بالنجاح فى مهنة معينة غير متوافرة.

اختبار المعلومات الميكانيكية Mechanical Information Test :

يقيس هذا الاختبار معلومات الفرد أو خبرته الميكانيكية، حيث يمكن باستخدامه انتقاء القائمين بتدريب الأفراد، ويُفترض أن هؤلاء المدربين لديهم الاستعداد الميكانيكي.

اختبار أوروك للاستعداد الميكانيكي O`Rourk Mechanical Aptiliude Test :

ويشتمل على مفردات تتطلب المزاوجة بين الصور والأدوات، وهذه المفردات تمثل الجزء الأول من الاختبار، أما الجزء الثاني فيشتمل على مفردات اختبار من متعدد تقيس المعلومات المتعلقة بالأدوات واستخداماتها ، ومفردات كل من الجزئيين تحتوي على أسئلة متنوعة ومتعددة تقيس المعلومات فى مجالات النجارة ، والكهرباء ، والطلاء ، والطباعة ، وغيرها من المهن ، غير أن البيانات المتوافرة حول صدق الاختبار وثبات درجاته محدودة .

اختبار الاستدلال الميكانيكي Mechanical Reasoning Test :

تُعد اختبار بينيت للفهم الميكانيكي Bennett Mechanical Comprehension Test (BMCT) واحداً من أفضل اختبارات الاستدلال الميكانيكي ، ويشتمل على أسئلة تحتوى على صور لأشياء مألوفة بينها علاقة ميكانيكية معينة ، وعلى الفرد أن يتوصل إلى استدلالات صحيحة عن هذه العلاقات ، ومن أمثلة ذلك ، صورة لغرفتين تحتويان على عدد مختلف من قطع الأثاث ، ويُطلب من الفرد الاستدلال على الغرفة التى يكون صدى الصوت فيها أكبر .

وقد أُعدت لهذا الاختبار معايير ، وكذلك صيغة أخري للاختبار تناسب الإناث .

(٢) اختبارات الاستعداد الكتابي Clerical Aptitude Tests :

يُعد الاستعداد الكتابي أيضاً من الاستعدادات متعددة الأبعاد ، وعلى الرغم من أن المهارة اليدوية وسرعة إدراك التماثل والاختلاف بين الأشياء يُعد ضرورياً للعمل الكتابي ، إلا أن القدرات اللغوية والرياضية التى تعد من مظاهر الذكاء العام، تكون مهمة أيضاً فى العمل الكتابي ، لذلك فإن كثيراً من اختبارات

الاستعداد الكتابي تشتمل على مفردات متشابهة لتلك التى تحتويها اختبارات الذكاء العام ، إضافة إلى مفردات تقيس السرعة الإدراكية ، والدقة وكثير من اختبارات الاستعداد الكتابي الأكثر شمولية تحاول قياس استعدادات متنوعة مثل : المهارات اللفظية ، وملأ فراغات بحروف ، وقواعد لغوية ، وتهجُى ، وجميعها مهارات مهمة للنجاح فى العمل الكتابي ، ومن الملاحظ أن بطارية اختبارات الاستعدادات الفارقة (DAT) التي أوضحناها تحتوي على اختبارات فرعية تتعلق بالاستعداد الكتابي .

اختبار مينسوتا للاستعداد الكتابي : Minnesota Clerical Test :

يُعد هذا الاختبار من اختبارات الاستعداد الكتابي العام ، وأعده أندرو Andrew ، وبترسون Paterson ، ويُستخدم بكثرة فى أنتقاء العاملين فى المجالات الكتابية ، والمدققين ، وغيرهم ممن يلتحقون بوظائف تتطلب السرعة الإدراكية ، والتعامل مع الرموز بدقة ، ويشتمل الاختبار على جزئيين هما : المقارنة بين الأعداد بين الأعداد ، والمقارنة بين الأسماء ، حيث يتحقق المختبر من الأزواج المتطابقة فى قائمة تشتمل على (٢٠٠) زوج من الأعداد ، (٢٠٠) زوج من الاسماء على الترتيب ، وكل من الجزئين له زمن محدد ، ويتم تصحيح الدرجات من أثر التخمين فى الإجابة .

وتشتمل معايير الاختبار على المئينيات لكل من الذكور والإناث ، والفرق الدراسية (٧-١٢) ولمجموعات من العاملين فى المجالات الكتابية ، والمتقدمين لشغل هذه الوظائف .

ومعامل ثبات درجات الاختبار بطريقة التجزئة النصفية تتراوح بين (٠.٧٠ – ٠.٨٠) وترتبط هذه الدرجات بتقديرات المعلمين والمشرفين على الأعمال الكتابية ارتباطاً متوسطاً .

وتوجد اختبارات أخري تقيس الاستعداد الكتابي العام وتتناول مفرداتها ، التهُجى ، الترتيب الأبجدي ، الجمل غير المرتبة ، العمليات الحسابية البسيطة.

وتوجد أيضاً اختبارات استعدادات تتعلق بمجالات خاصة ، مثل الاختزال ، وبرمجة الحاسوب ، وغير ذلك من المجالات الكتابية النوعية .

(٣) اختبارات الاستعدادات الموسيقية والفنية Musical and ARTISTIC Tests :

توجد اختبارات استعدادات موسيقية وفنية متعددة ، غير أنها محدودة الاستخدام بوجه عام ، ولعل اختبارات الاستعدادات الموسيقية شيوعاً اختبار سيشور للموهبة الموسيقية Seashore Measures of Mental Tests .

وتُقدَّم مفردات هذه الاختبارات مُسجلة على شرائط وتشتمل على ستة اختبارات هي : تتميز درجة النغم ، وتمييز علو الصوت ، وتمييز الزمن ، والحكم على الزمن ، والحكم على الإيقاع ، وذاكرة النغم ، ويعتبر سيشور وزملاؤه أن هذه القدرات أساسية فى تنمية الكفاءة في الموسيقي ، وهذه الاختبارات تناسب الطلاب من الصف الرابع إلى الرشد ويستغرق تطبيقها ساعة واحدة تقريباً بمعدل (١٠) دقائق لكل اختبار .

وقد وُجد أن معامل ثبات درجاتها بطريقة التجزئة النصفية يتراوح بين (٠.٥٥ ، ٠.٩٠) ، وتم حساب المعايير المئينية للصفوف (٤-٥) ، (٦-٨)، (٩-١٦)، وبينت نتائج البحوث أن درجات هذه الاختبارات ترتبط بمحكات التدريب الموسيقي ارتباطاً تتراوح قيمته بين (٠.٣٠ ، ٠.٤٠) ولكن البيانات المتعلقة بصدق الاختبارات غير كافية ، ويُعاب على هذه الاختبارات أنها جزيئية ، أي أن كلاً منها يقيس جانباً محدوداً من الموهبة الموسيقية .

أما اختبارات الاستعدادات الفنية فإنها تنقسم إلى نوعين هما :

الحكم الفني Aesthetic Judgment ، والانتاج الفني Aesthetic Production ، ويُعد الحكم الفني ضرورياً للنقد الفني ، أما الإنتاج الفني فإنها

يتعلق بالعمل الفني ، ويتطلب الإنتاج الفني أيضاً حكماً فنياً ، ولكن العكس لا يكون ضرورياً .

اختبار ماير للحكم الفني Meier Art Judgment Test :

يُعد أكثر اختبارات الاستعدادات الفنية شيوعاً ، ويشتمل على (١٠٠) زوج من اللوحات الفنية الشهيرة ، وإحدى لوحات كل زوج تُعد أصلية ، أما الأخري فإنها تُعد نسخة معدلة من الأولي ، ويطلب من الفرد تحديد أوجه الاختلاف بينهما ، وأيهما يفضَّل ، ويمكن باستخدام هذا الاختبار التمييز بين مجموعات الأفراد الذين يختلفون اختلافات محددة فى خبرتهم الفنية .

قائمة هورن للاستعدادات الفني Horn Art Aptitude Inventory :

تُعد مقياساً للإنتاج الفني ، تشتمل على جزئين : أحدهما يتطلب رسم (٢٠) شيئاً مختلفاً ولكنه مألوف ، مثل : شجرة ، أو كتاب ، أو ملعقة ، فى زمن يتراوح بين (١٠.٣) ثواني ، أما الآخر فيتطلب التخيل ، حيث يشتمل على (١٢) مستطيلاً يتضمن كل منها خطوطاً متعددة يستند إليها الفرد فى رسم صورة مُمثلة ، وتتميز هذه القائمة بصدق مرتفع ، كما أن قيمة معامل ارتباط درجاتها بمتوسط تقديرات معلمي التربية الفنية بلغت (٠.٥٣) .

ومن هذا يتضح تنوع الاستعدادات الخاصة ، وتعدد الاختبارات التي تقيسها والتى تختلف فى مدى اتساع محتواها وأغراض استخداماتها والاختبارات التى تناولها هذا الفصل تُعد بمثابة عينة من كثير من اختبارات الاستعدادات التي يعرضها الكتاب السنوي للقياس العقلي Buros` Mental Measurement Year book .

غير إن هذه العينة اشتملت على بعض الاختبارات الأكثر شيوعاً واستخداماً فى التنبؤ بالأداء الأكاديمي ، والكفاءة فى مجالات مهنية متنوعة

اختبارات ومقاييس الشخصية

- ❖ مقدمة
- ❖ مفهوم الشخصية
- ❖ أهداف قياس الشخصية
- ❖ بعض مشكلات قياس الشخصية
- ❖ استراتيجيات بناء استبيانات الشخصية
- ❖ أنواع اختبار ومقاييس الشخصية

(أولاً) استبيانات تستند إلى الاستراتيجية المنطقية

(ثانياً) استبيانات تستند إلى الاستراتيجية الإمريقية

(ثالثاً) استبيانات تستند إلى استراتيجية التجانس

- ❖ تعقيب عام على استبيانات الشخصية
- ❖ المقاييس الإسقاطية للشخصية
- ❖ تعقيب عام على المقاييس والاسقاطية للشخصية

❖ مقدمة

عرضنا فى الفصول الثلاثة السابقة بعض أهم الاختبارات والمقاييس التى تُستخدم فى تقييم الخصائص أو السمات المعرفية لدى الأفراد ، غير أن الشخصية الإنسانية تُعد تركيباً منفرداً من الخصائص المعرفية وغير المعرفية ، أي الوجدانية، التي تمثل نمط السلوك المتسق والمميز للفرد ، لذلك فإن قياس وتقييم الشخصية يشتمل على نطاق واسع من المتغيرات المعرفية وغير المعرفية ، مثل التحصيل ، والذكاء العام ، والاستعدادات الخاصة التى أوضحناها ، وكذلك الميول، والاتجاهات ، والقيم ، كما أن هناك خصائص أخري مميزة للفرد تتعلق بالجوانب الانفعالية والمزاجية التى يُشار إليها عادة بأنها متغيرات شخصية Personality Variables وهذه الجوانب لها أهمية كبيرة فى فهم سلوك الأفراد والتنبؤ به ، وبالطبع يصعب الفصل بين الخصائص المعرفية وغير المعرفية فى الشخصية الإنسانية ، إذ أنها تعد خصائص متكاملة وتؤثر فى سلوك الأفراد ، وتناول كل منها على حدة لا يعني أنها تؤثر تأثيراً منفرداً فى السلوك ، وإنما لتيسير دراستها وفهمها .

لذلك سوف أفرد هذا الفصل لتوضيح الأساليب المختلفة التى تَستخدم فى قياس الشخصية الإنسانية من جوانبها الانفعالية والمزاجية .

ويتضمن ذلك التعريفات المتعددة لمفهوم الشخصية وأهداف قياس الشخصية، وبعض مشكلاته ، والاستراتيجيات المختلفة لبناء مقاييس الشخصية ، وأنواع هذه المقاييس ، وأمثلة لكل نوع منها متضمنة مكونات المقياس ، وكيفية تفسير درجاته، واستخداماته ، وميزاته وعيوبه .

❖ مفهوم الشخصية Personality :

يُعد مفهوم الشخصية من المفاهيم المعقدة التي تتميز بالغموض ، إذ لا يوجد اتفاق بين علماء النفس أنفسهم حول طبيعة الشخصية الإنسانية ، فهذا المفهوم مشتق من المصطلح اللاتيني " Persona " إلى تعني " القناع " ، وقد أضيفت

دلالات أخري لهذا المفهوم تدريجياً تعبُر عنها الكلمة الإنجليزية "Personage" ، أي " أشخاص ذوى أهمية " ولعل هذا يتضح من تعدد تعريفات هذا المفهوم الذى أشارت إليه الأدبيات ، فأحد المراجع التي تناولت نظريات الشخصية عرضت ما يزيد عن عشر نظريات رئيسية ، ويندرج تحتها نظريات فرعية (Hall & Lindzey, 1957) ، ويمكن وضع معظم نظريات الشخصية على مُتحصل يتراوح فيه التأكيد التام بين السلوك الملاحظ والعمليات الداخلية التي تنعكس فى السلوك ، ولعل نظريات الشخصية ، التي تنبأها علماء النفس فى أمريكا تميل إلى تأكيد المنظور السلوكي ، حيث يمكن تعريف الشخصية بأنها مجموع السلوك الكلي للفرد ، أو أنها النزعة المميزة للفرد إلى أن يسلك بطريقة متشابهة فى المواقف المتشابهة ، أما النظريات التي تبنُّاها علماء النفس فى أوربا ، فإنها تميل إلى تأكيد الطرف الآخر للمتصل ، أي العمليات الداخلية .

ويتمثل ذلك فى نظريات التحليل النفسي ، وبخاصة نظرية فرويد Freud ، وعلى الرغم من تأكيد علماء الأمريكيين على السلوك بدرجة أساسية ، إلا أن المعاصرين منهم اعترفوا بوجود دوافع كامنة ، أو حالات عضوية داخلية تحفز الفرد للسلوك ، أو حاجات تؤدي إلى تنشيط السلوك وتوجُّهه ، وبالطبع يصعب فى هذا الكتاب المدخلي مناقشة النظريات الرئيسة للشخصية ، وإنما يمكن الإشارة إلى بعض التعريفات الشائعة للشخصية كمفهوم سيكولوجي ، وربما تعتمد هذه التعريفات على ما إذا كانت تنظر إلى الشخصية على أنها كينونة Entity متمايزة عن الكينونات الأخري ، ولها خصائص معينة ، أم تنظر إليها على أنها وصف ليكفية اختلاف الأفراد فى من درجة امتلاكهم مجموعة من السمات ، والواقع أن كلاً من النظريتين تكمل إحداهما الأخري ، فقد عرفً البورت Allport الشخصية بأنها التنظيم الدينامي داخل الفرد للأنظمة السيكولوجية التى تحدد تكيفه المتفرد لبيئته ، ومع هذا فإنه فى مراجعته للأدبيات التي تناولت معاني مفهوم الشخصية ، استطاع ألبورت التمييز بين ٥٠ معنى مختلفاً ، وربما هناك المزيد من المعاني .

ويمكن تصنيف تعريفات الشخصية بعامة فى خمسة أقسام منفصلة كالتالي :

(١) التعريفات الشاملة Inchusive : تنظر هذه التعريفات إلى الشخصية على أنها مجموع سمات أو وحدات سلوكية متنوعة مستقلة ، لذلك فإنها تميل إلى النظر إلى الشخصية نظرة جزئية ، وأن الشخصية محصلة وليست مظهراً متكاملاً نسبياً ينبثق من عملية النمو ولعل برينس Prince يعد أحد التعريفات، حيث يري أن الشخصية هى المجموع الكلي لموروثات فطرية، ونزعات ، وبواعث ، وغرائز الفرد ، وكذلك نزعاته المكتسبة نتيجة الخبرة .

(٢) التعريفات التكاملية Integrative: تركز هذه التعريفات على الجوانب التنظيمية للعوامل التي يتكون منها سلوك الفرد ، وتنظر إلى الحالات والعمليات الداخلية فى ضوء العلاقات القائمة بينها ، ومن هذه التعريفات تعريف وارن ، وكارما يكل Warren & Carmichael ، وهو أن الشخصية هى التنظيم العقلي الكامل للفرد فى أي مرحلة من مراجل نموه ، وهذا يتضمن : العقل والمزاج ، والمهارة ، والخُلق ، وجميع الاتجاهات التي تتكون خلال حياة الفرد .

(٣) التعريفات الهرمية Hierarchical : تؤكد هذه التعريفات الجوانب التنظيمية للشخصية كما فى التعريفات التكاملية ، ولكن مدخلها فى التنظيم رأسي وليس أفقياً ، ومن أمثلة ذلك التعرف المبكر لماكد وجل McDougal الذى يرى أن هناك مرحلتين هرميتين للعواطف Sentiments ، إحداهما بنائية ، والأخرى تتعلق بتكوين عواطف أساسية فى كل متناغم .

(٤) التعريفات الخُلقية Idiosyncratic : تُركز هذه التعريفات على تفُرد الشخص فى علاقته بالآخرين ، وتحاول فى معظم الأحيان التمييز بين الشخص وثقافته ، إذ يرى سشون Schoen أنه إذا تصرف ، وفكر ، وشعر جميع أفراد مجتمع معين بنفس الطريقة ، فلن يكون هناك مجال للحديث عن الشخصية ،

لذلك يُعرف الشخصية بأنها النظام المرتب ، والوظيفية الكلية للعادات والنزعات ، والعواطف التى تميز الفرد عن غيره من أفراد جماعته .

(٥) التعريفات التكثيفية Adaptive : تؤكد هذه التعريفات قدرة الفرد على التعامل مع بيئته بفاعلية ، إذ يرى بأودن Bowden أن الشخصية هى نزاعات التكيف المستقرة لدى الفرد التى تحكم سلوكه فى بيئته .

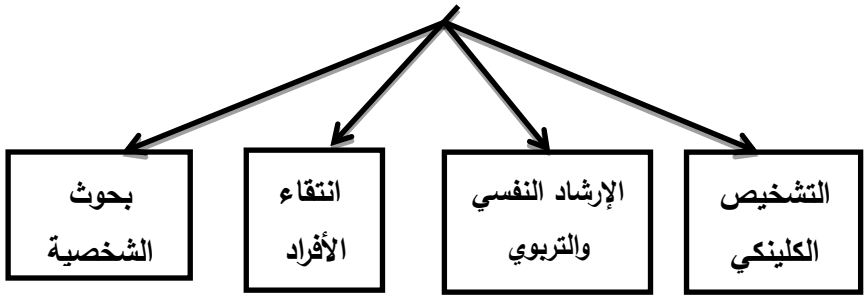
ومن هذا يتضح تباين وجهات النظر التى بينتها التعريفات الخمسة السابقة، وربما يرجع ذلك إلى تباين الأطر النظرية التى تبنأها علماء النفس فى مجال الشخصية ، ومما لا شك فيه أن هذا ينعكس على أساليب وطرق قياس الشخصية ، حيث إن كل اختبار أو مقياس يستند إلى نظرية معينة ، ويُمثَّل مفاهيمها ومبادئها ، لذلك يصعب الحصول على اختبار أو مقياس للشخصية لا يستند فى بنائه استناداً صريحاً أو ضمنياً إلى إطار نظري معين ، حيث يُعد هذا الإطار بمثابة مرجعية الاختبار ، والمحك المحدد لانتقاء المفردات أو المهام التى يشتمل عليها .

لذلك ينبغي على من يستخدم اختبارات ومقاييس الشخصية ، أو يقوم بتفسير درجاتها ، أن تكون لديه معرفة كافية بالنظرية التى استند إليها بناء الاختبار أو المقياس ، وكذلك ينبغي أن يكون مدركاً للإطار النظري الذى تبناه مصمم الاختبار يقيس بعض السمات دون الأخرى .

أهداف قياس الشخصية :

كثيراً من المشتغلين بعلم النفس فى مختلف المجالات التطبيقية ، أو مجالات البحث يهتمون اهتماماً مباشراً بقياس الشخصية الإنسانية ، ويوضح الشكل التخطيطي (٨-١) التالي هذه المجالات :

أهداف قياس الشخصية



شكل (٨-١) يوضح أهداف قياس الشخصية

التشخيص الكلينكي :

نحتاج إلى قياس الشخصية فى التشخيص الكلينيكي فى مستشفيات الأمراض العقلية ، والعيادات النفسية ، وذلك لكي يقرر الأخصائي مدى وطبيعة الاضطراب فى الشخصية الذى يعاني منه المريض ، ويحدد كيفية العلاج المناسب .

الإرشاد النفسي والتربوي :

يسهم الارشاد والعلاج النفسي فى علاج اضطرابات الشخصية بأساليب سيكولوجية ، إذ يستخدم كل منها التفاعل اللفظي بين المريض والأخصائي النفسي بغرض مساعدة المريض فى تغيير سلوكه لكي تتيسر له عملية التكيف ، واستخدام مقاييس الشخصية قبل العلاج ، وفى أثنائه ، وبعد انتهائه ، يصاحب عادة التشخيص الكلينيكي ، والغرض من هذا القياس هو تقييم مدى التغير الذى ينتج عن أسلوب العلاج .

إنتقـــاء الافراد :

تُستخدم اختبارات ومقاييس الشخصية فى انتقاء الأفراد ، أو ترقيتهم فى وظائف أو أعمال معينة ، أو قبولهم فى برامج تدريبية متخصصة ، مثل : تدريب الممرضات ، والأطباء ، ورجال البوليس ، وغيرهم لزيادة فاعليتهم فى أدائهم لأعمالهم .

والمشكلة الرئيسة فى هذه الحالة هى ما إذا كان المتقدمون يتميزون بسمات شخصية معينة يتطلبها العمل المرشحون له ، وهو ما يمكن تحديده باستخدام مقاييس الشخصية .

بحوث الشخصية :

تُستخدم بحوث الشخصية فى المواقف العملية التى تتطلب مثلاً تقدير أثر برنامج تدريبي معين على الشخصية ، كما تُستخدم فى البحوث المختبرية ، حيث تكون عينة البحث مجموعة من الأفراد المتطوعين للاشتراك فى التجربة ، فمثلاً يمكن إجراء تجربة عن أثر الحرمان من الطعام لفترة طويلة على شخصيات مجموعة من الأفراد .

كما تُستخدم بحوث قياس الشخصية فى التحقق من كفاية طريقة القياس ذاتها، وهذا النوع من البحوث يصمم للكشف عما إذا كانت أداة قياس معينة تقيس سمات شخصية بطريقة متسقة ، وما إذا كانت السمات التي تقيسها هي ما يود الباحث قياسه بالفعل ، إذا ربما تدعو حاجة مؤسسة صناعية معينة ، أو مستشفي معين إلى بناء مقياس جديد للشخصية لأغراض معينة ، فهنا يقوم فريق البحث بدراسة مدى تحقيق المقياس للغرض الذى أعد من أجله .

وكذلك تُستخدم بغرض الإسهام فى زيادة المعرفة وتطوير النظريات المتعلقة بالوظائف النفسية ، أى تطوير نظام منطقي يمكن فيه قياس التكوينات الفرضية Constructs التي يمكن الاستدلال منها على العلاقات القائمة بين الأنماط السلوكية المختلفة .

ومن هذا يتضح أن هناك هدفين رئيسين لقياس الشخصية أحدهما نظري تطويري ، والآخر عملي تطبيقي ، فالهدف النظري التطويري يتعلق بالإسهام فى التعرف على متغيرات ومفاهيم نظرية للشخصية ، ومن خلال القياس يمكن ربط

هذه المفاهيم بأنماط السلوك الملاحظ ، أما الهدف العملي التطبيقي فإنه يتعلق باتخاذ قرارات وتنبؤات عن الأفراد فى مواقف عملية .

بعض مشكلات قياس الشخصية :

يختلف قياس الشخصية عن قياس الذكاء أو غيره من الجوانب المعرفية فى أن قياس الشخصية وغيرها من الجوانب غير المعرفية يكون أكثر تعقيداً ، فالفرد الذي يُطبق عليه اختبار أو مقياس الشخصية لا يكون واضحاً لديه السمات أو الخصائص التى يجري قياسها ، أو كيفية تفسير الدرجات ، كما أن الاستجابات تكون عرضة لتحيزات مختلفة ، وكذلك ربما تؤثر عملية القياس أو الموقف الاختباري فى سلوك الفرد المختبر مما قد يؤدى إلى تغيير استجابته فى الاختبار، وبالتالي يجعل صدق الاختبار موضع شك .

غير أن تحيز استجابات الفرد ، أو اتجاهه العقلي Response Sets ، تعد من المشكلات التى لا توجد ، أو على الأقل توجد بدرجة محدودة فى قياس الجوانب المعرفية أو الأداء الأقصى ، والمقصود بالاتجاه العقلي فى الاستجابة ميل أو نزعة الفرد للاستجابة فى اتجاه معين بغض النظر عن محتوى المفردة أو الفقرة الاختبارية ، وكذلك نود عند قياس الانبساط مثلاً Extraversion معرفة ما إذا كان الفرد الذي يُطبق عليه المقياس يسلك بطريقة تُعبر عن الانبساط وليس مجرد الموافقة أو عدم الموافقة على عبارات معينة يتضمنها هذا المقياس .

فبعض الأفراد يسلك سلوك الانبساط فى مواقف وأوقات معينة دون سواها ، ولذلك إذا طُلب من أحد الأفراد أن يسلك سلوكاً يدل على الانبساط فإنه يستطيع ذلك ، ولكننا لا نريد قياس هذا السلوك ، وإنما نود تحديد كيف يسلك هذا الفرد عادة ، لذلك فإن الاختبارات ومقاييس الشخصية يجب أن تُبني بحيث تجعلنا نتأكد أن استجابات الأفراد تدل على سلوكهم النمطي أو أدائهم المميز .

وتوجد ثلاثة أنواع رئيسة من الاتجاهات العقلية هى :

١) **الإذعان أو الموافقة Acquiescence** : اهتمت كثير من الأبحاث بدراسة هذا النوع من تحيز الاستجابات ويُقصد بالإذعان أو الموافقة اتجاه الفرد أو نزعته للاستجابة بـ " نعم " أو بـ " موافق " فى اختبارات الشخصية ، والدرجات التي يحصل عليها مثل هؤلاء الأفراد تكون غير صادقة ، لأنه يُفترض أنها لا تقيس الاستجابات لمحتوى فقرات الاختبار ، بقدر ما تقيس هذه النزعة شبه الشعورية للموافقة على فقرات أو عبارات الاختبار .

ويوجد تعريف آخر للإذعان أو الموافقة يختلط أحياناً بالتعريف السابق ، وهو أنه " يُعد سمة شخصية عند الأفراد الذين يستجيبون فى معظم الأحيان لفقرات أو عبارات استبيانات الشخصية ، ويختلف علماء النفس فيما إذا كانت هذه الاستبيانات يجب أن تقيس هذه السمة .

٢) **مراعاة القبول الإجتماعي Social Desirability** : كثير من الأفراد سوف يوافقون بدرجة كبيرة على عبارة مثل : " أنا شخصي ذكي ومبدع " أما عبارة مثل " أشعر بصداع شديد إذا سارت الأمور على غير توقعاتي " فإنها تعد بالنسبة لهم غير مرغوبة اجتماعياً ، وبالتالي سوف لا يوافقون عليها ، وهذا يرجع إلى أن الفرد يحاول بطريقة شبه شعورية أن يكون مقبولاً اجتماعياً ، وقبوله العبارة الأولي لا يعد نوعاً من الكذب لأن الاستجابة تكون بطريقة شبه شعورية ، فى حين أن الكذب يكون متعمداً ، ويعتمد الفرد فى استجابته على محتوى العبارات أو الفقرات ، ويكون مدركاً للعلاقة بين محتوى العبارات وطريقة تصحيحها ، ويمكن الاستناد إلى أحكام مجموعة من الخبراء فى تقدير قيمة وزنية للقبول الاجتماعي لكل عبارة من عبارات استبيان الشخصية.

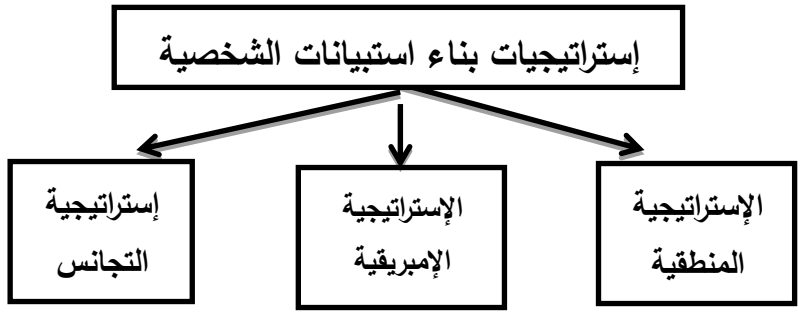
٣) **الاستجابات الشاذة Deviant Responses**: فهناك نزعة لدى بعض الأفراد أن يستجيبوا استجابات شاذة Atypical على فقرات أو عبارات مقاييس

الشخصية ، فنحن نتوقع من الأفراد غير الأسوياء ، أن يستجيبوا لمحتوى المقياس استجابات غير نمطية .

ويفترض " بيرج Berg " أن الاستجابة الشاذة للأفراد غير الأسوياء لا تعتمد على محتوى فقرات أو عبارات مقياس الشخصية ، ولذلك فإنه يُفترض أنه يمكن ببساطة اكتشاف هذا الشذوذ عن طريق معرفة عدد الاستجابات الشاذة التي يختارها الفرد بغض النظر عن محتوى الفقرات أو العبارات ومعظم البحوث التى تهتم بدراسة الاستجابات الشاذة تتطلب من الفرد أن يختار استجابته على بُعد يتراوح بين " أوافق جداً ، " لا أوافق على الإطلاق " ، وهنا يلجأ الفرد غير السوي إلى اختيار استجابة تختلف عن استجابات معظم الأفراد الآخرين ، ويرى بيرج أن هذه الاستجابة تتكرر فى جميع المقاييس الأخري للشخصية .

❖ إستراتيجيات بناء استبيانات الشخصية :

يعتمد كثير من مقاييس الشخصية على أساليب التقرير الذاتي Self Report التى تتطلب أن يستجيب الفرد لعبارة واضحة ومكتوبة إما (بنعم) أو (لا) ، ويستند بناء هذه المقاييس إلى ثلاث استراتيجيات رئيسية يوضحها الشكل التخطيطي (٨ –٢) التالي :



شكل (٨–٢) يوضح إستراتيجيات بناء استبيانات الشخصية

(١) **الإستراتيجية المنطقية Logical Strategy :** استُخدمت هذه الإستراتيجية فى بناء المقاييس المبكرة للشخصية ، فقد بُنيت على أساس نظري أو منطقي أو حدَّسي ، وهذا يعنى أنه بمجرد تحديد السمة أو السمات المراد قياسها تُصاغ فقرات أو عبارات تبدو أنها تقيس هذه السمة أو السمات ، فمثلاً يمكن أن يشتمل المقياس الذى يقيس الانطواء على فقرات أو عبارات مثل :

أخجل بسهولة :

" أفضل المشاركة فى الحفلات الصغيرة مع الأصدقاء على المشاركة فى الحفلات الكبيرة التى تجمع عدداً كبيراً من الناس " .

" أعتبر القراءة من هواياتي المفضلة " .

ويتم تصحيح الفقرات على أساس منطقي ، وعلى الرغم من أنه ربما لا يكون هناك اتفاق عام على تقدير درجة فقرة معينة ، إلا أنه عند بناء المقياس يجب الاهتمام بانتقاء المفردات التى يسهل تصحيحها أو تقدير درجاتها ، وبناء المقياس استناداً إلى هذه الإستراتيجية يُعد خطوة أوَّلية ، وذلك لأن الفقرات التى تبدو أنها تقيس سمة معينة من حيث محتواها ، ربما لا تقيس هذه السمة فى الحقيقة ، وهذا يحد من استخدامها .

(٢) **الإستراتيجية الإمبريقية Empirical Strategy :** تُنتقي الفقرات أو العبارات فى هذه الحالة على أساس علاقتها التجريبية بمحك خارجي ، والخطوة الأولي هى اختيار مجموعة محكة Criterion Group تتميز بالسمة المراد قياسها.

فإذا افترضنا مثلاً أننا نود قياس سمة " القيادة Leadership " – إذا كانت القيادة هى في الحقيقة سمة نفسية = فإننا نحدد تعريف هذه السمة ، ونختار مجموعة من الأفراد الذين ينطبق عليها هذا التعريف ، فالقائد ربما يمكن اعتباره

الشخص الذى تنتخبه الجماعة لقيادة هيئة أو جمعية معينة ، وهنا يبرز التساؤل : ما الخصائص التي تميز أعضاء مجموعة القادة عن غيرهم من الأفراد .

وإجابة هذا التساؤل تحتاج إلى تصميم عينة من المثيرات (الفقرات) ومجموعة مقارنة تتكون من أشخاص لا يتميزون بسمة القيادة .

وربما يتطلب تصميم وبناء عينة الفقرات الاسترشاد بالدراسات المتعلقة بالقيادة، ومراجع على النفس الاجتماعي ، وغير ذلك ، كما يمكن أن نطلب من مجموعة من الخبراء كتابة قائمة بالصفات التى يرون أنها تميز القائد ، وكذلك الاطلاع على المقاييس الأخري التى تقيس القيادة والسمات المتعلقة بها .

وبعد كتابة عينة ممثلة من الفقرات ، نقوم بتطبيقها على مجموعة القادة ، ومجموعة أخري لا تتصف بالقيادة، ولكي نحدد أي الفقرات تميز بين المجموعتين، يمكن إيجاد نسبة عدد الأفراد الذين استجابوا لكل مفردة بطريقة معينة فى كل من المجموعتين ، ومن ثم نختار الفقرات التى تميز بدرجة كبيرة بينهما .

ومن الجدير بالذكر أن مقياس الشخصية الذى يُبني استناداً إلى هذه الاستراتيجية يمكن أن يشتمل على فقرات أو عبارات ربما لا يبدو أن محتواها له علاقة منطقية بالسمة أو السمات المراد قياسها .

غير إنه ينبغي مراعاة أن انتقاء الفقرات على أساس نتائج مجموعة محكة معينة ربما تعكس طبيعة وخصائص هذه المجموعة على المقياس ، فاختيار مجموعة القادة فى هذا المثال وفقاً لتعريف معين ربما لا يكون كافياً ، وإذا اختلف التعريف ، أو لم تحقق المجموعة المحكة صفات القيادة المذكورة فى التعريف الذى نختاره ، فإن المقياس لا يكون مناسباً ، كما ينبغي مراعاة أن الاستناد إلى هذه الاستراتيجية فى بناء مقاييس الشخصية ربما يضمن أن تتميز هذه المقاييس بالصدق التلازمي ، ولكن يبرز التساؤل حول مدى الاستفادة من المقياس فى

التنبؤ بمحك يختلف عن مجرد قياس انتماء الفرد لعضوية المجموعة المحكة الأصلية ؛ فهذا التساؤل يتعلق بالأهمية النظرية للمقياس .

(٣) استراتيجية التجانس Homogeneity Strategy :

يرى بعض علماء القياس أن المقاييس التى تتميز بتجانس مفرداتها هي التى يمكن استخدامها فى قياس متغيرات سيكولوجية ذات معني ، أي أن الأساس فى المقياس أن يكون متجانساً ، وهذا يعني أن الفقرات التي لا ترتبط بغيرها من فقرات المقياس يجب حدفها ، لأنه يمكن اعتباراها فى هذه الحالة أنها تقيس سمة أو تكويناً فرضياً مختلفاً .

ويتطلب بناء مقياس الشخصية استناداً إلى هذه الاستراتيجية البدء بعدد كبير من الفقرات ، وتطبيقها على عينة تقنين مناسبة من الأفراد ، وتُحسب معاملات الارتباط بين المفردات ، ويُستخدم أسلوب التحليل العاملي Factor Analysis – وهو من الأساليب الإحصائية متعددة المتغيرات سبقت الإشارة إليه فى الفصل الرابع – لتجميع الفقرات فى أقل عدد من العوامل المتعامدة التى تفسر التباين فى في مصفوفة معاملات الارتباط .

وبذلك يكون قد صنَّفنا الفقرات في مجموعات متجانسة ، تقيس كل مجموعة منها سمة معينة يشتمل عليها المقياس ، ويُعطي اسماً للمقياس فى ضوء محتوى الفقرات التي تشتمل عليها كل مجموعة .

ومقاييس الشخصية التي تستند فى بنائها إلى هذه الاستراتيجية تكون احادية البُعد One–dimensional ، وتتميز بقدر من صدق التكوين الفرضي Construct Validity الذى أوضحناه فى الفصل الرابع .

ومن الجدير بالذكر أن هذه الاستراتيجيات الثلاث فى بناء مقاييس الشخصية التي تعتمد على التقرير الذاتي باستخدام الاستبيانات ، تعد مداخل رئيسه فى هذا

الشأن ، وعلى الرغم من أن الطريقة المنطقية أقلها استخداماً ، إلا أنه يمكن الاسترشاد بها فى انتقاء المفردات فى المراحل الأولية لبناء المقياس ، لكي يكون للمفردات أساس منطقي ، وربما يُفضل الجمع بين الاستراتيجية الإمبريقية ، واستراتيجية الإمبريقية ، واستراتيجية التجانس فى بناء مقاييس الشخصية حيث تحاول بناء مقياس فقراته متجانسة ، ثم تُقَّدر صدقه عن طريق محك تجريبي ، ويستبعد الفقرات التى لا تحقق هذا المحك ، وبذلك يصبح للمقياس أهمية نظرية وفائدة عملية .

❖ أنواع اختبارات ومقاييس الشخصية :

يمكن قياس الشخصية بأنواع مختلفة من الاختبارات والمقاييس ، فقياس وتقييم الشخصية يتضمن الملاحظة الموضوعية المنظمة لسلوك الفرد تحت شروط محددة فى علاقتها بمثيرات معينة ، وللحصول على بيانات تتعلق بفرد ما ، يجب محاولة فهم أثر مختلف مكونات الموقف المثير على سلوكه .

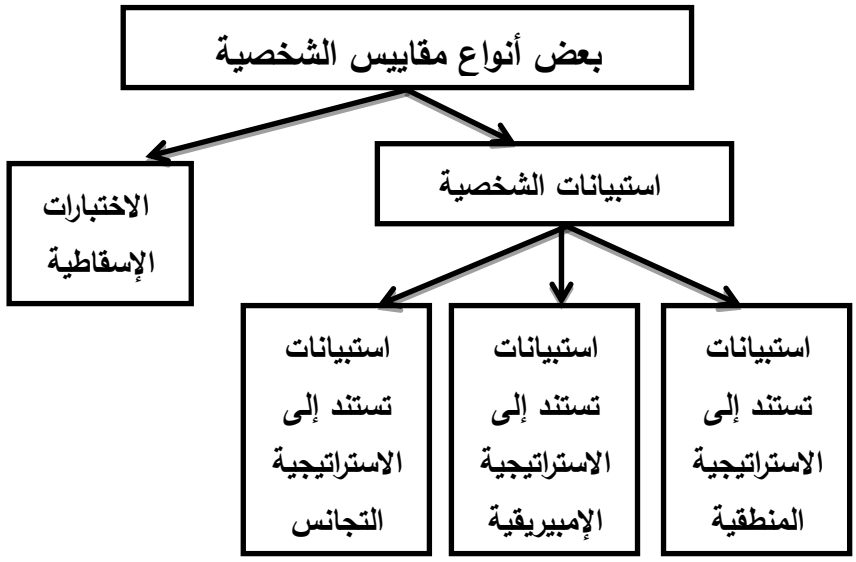
ويتحدد الموقف المثير بخصائصه الفيزيائية ، وهو ما يجد الفرد نفسه فى مواجهته وبالتعليمات التي تُعطي لفرد ، وبالمطلب المحدد (المثيرات) التى يواجهها ، فنحن نستطيع أن نُكون فكرة واضحة عن طبيعة الشخصية ، وتجمع بيانات تستند إلى نظرية معينة من خلال ملاحظة استجابات الأفراد لمثل هذه المواقف المحددة تحديداً واضحاً .

ونظراً لاختلاف المتغيرات المرتبطة بعملية جمع البيانات المتعلقة بشخصية الفرد ، فإن هناك تبايناً واضحاً فى طرق وأساليب جمع هذه البيانات من حيث ، طبيعة الموقف الاختباري ، وطبيعة المثيرات ، وطبيعة التعليمات المعطاة ، وطبيعة الاستجابة المطلوبة ، وكيفية تقدير الدرجات وتحليلها ، وتفسيرها .

واختلاف نظريات الشخصية يؤدى إلى اختلاف تعريفات مفهوم الشخصية كما سبق أن أوضحنا ، وبالتالي يؤدي إلى اختلاف طرق وأساليب القياس والتقدير، كما تؤدي إلى طرق مختلفة لملاحظة سلوك الأفراد ، وهذا يتطلب بالضرورة فهم العلاقات القائمة بين نظريات الشخصية وأساليب القياس والتقدير، فنظرية الشخصية التي لا تسهم فى تحديد وتيسير وسائل جمع معلومات عن الأفراد تُعد نظرية غير ذات جدوى من الناحية العملية ، كما أن البيانات التي لا تتعلق بنظرية معينة تكون عديمة المعني ، وقليلة الفائدة .

وينبغي أن تكون الطريقة المستخدمة في قياس الشخصية أكثر حساسية للسمة المراد قياسها ، ويجب ألا تتأثر بأي متغيرات أخري لا ترتبط بهذه السمة ، وهذا يتطلب تحديد الهدف من القياس ، وماذا نقيس ، وفيم تستخدم النتائج .

ويمكن تصنيف مقاييس الشخصية التي سأتناولها فى هذا الفصل فى ثلاثة أقسام رئيسه وفقاً للاستراتيجية المستخدمة فى بنائها ، كما هو موضح بالشكل التخطيطي (٨ – ٣) التالي :



شكل (٨–٣) يوضح أنواع مقاييس الشخصية

وسوف أقدم فيما يلي بعض الأمثلة لكل نوع من أنواع اختبارات ومقاييس الشخصية الموضحة بهذا الشكل .

أولاً : استبيانات تستند إلى الاستراتيجية المنطقية :

قائمة مراجعة المشكلات لموني Mooney Problem Check– List :

تتضمن هذه القائمة أربعة مستويات :

المستوى الأول : للصفوف من السابع إلى التاسع ، وينتج عن القائمة درجات تتعلق بالصحة والنمو الجسمي ، المدرسة ، المنزل والعائلة، المال، العمل ، المستقبل ، العلاقات الاجتماعية، أمور تتعلق بذات الفرد .

المستوى الثاني : للصفوف من التاسع إلى نهاية المرحلة الثانوية ، ويتناول مشكلات تتعلق بالصحة والنمو الجسمي، النواحي المالية، ظروف المعيشة ، التوظيف ، الأنشطة الاجتماعية والترويحية، العلاقات الاجتماعية السيكولوجية ، العلاقات الشخصية السيكولوجية ، التودد ، الزواج ، المنزل والعائلة ، الأخلاقيات والتدين ، التكيف فى البيئة المدرسية ، المستقبل الدراسي والمهني .

المستوى الثالث : المرحلة الجامعية ، ويتناول نفس مشكلات المستوى الثاني .

المستوى الرابع : للراشدين، ويتناول مشكلات تتعلق بالصحة، والأمن الاقتصادي، التطوير الذاتي، الشخصية، المنزل والأسرة ، التودد ، التدين، المهنة .

وقد اعتمد بناء فقرات هذه القائمة على تحليل دقيق لعدد كبير من العبارات المختلفة بمشكلات الطلاب والراشدين صاغتها عينة بلغ عددها (٤٠٠٠) فرد

تمثل هذه الشرائح الأربع ، وكذلك على الأدبيات ، وخبرات معد القائمة فى مجال الإرشاد ، ومن بين هذه المشكلات ما يلي :

- أود أن أتعلم كيف أقرأ بطريقة أفضل .
- لا أعيش مع والدي .
- أعاني من مشكلات فى أسناني .

ويمكن فى كل مستوى من المستويات الأربعة أن يوضح الفرد ما إذا كان يود مناقشة مشكلاته مع شخص معين يحدده بنفسه ، وهذه القائمة ليست اختباراً بالمفهوم السيكومتري ، حيث إن الفرد يستجيب للعبارة التي يشعر أنها تُعبر عن مشكلة لديه ، لذلك فإن القائمة تفتقر إلى بيانات تتعلق بالثبات والصدق ، على الرغم من أن دليل القائمة أشار إلى بعض الدراسات التى أُجريت فى هذا الشأن ، والهدف من هذه القائمة هو تقديم معلومات مساعدة أو إضافية للطالب أو الراشد، وليس لعباراتهما إجابات صحيحة أو خطأ ، وإنما تحدد مشكلات ، أو مجالات اهتمام فقط .

ومن الجدير بالذكر أن قائمة التفضيل الشخصي التى أعدها أدواردز Edwards Personal Preference Schedule (EPPS) تعد أحد استبيانات الشخصية شائعة الاستخدام ، وأستندت في بنائها إلى الاستراتيجية المنطقية أيضاً ، ولا تعد اختباراً شانها شأن قائمة موني ، لأن عباراتها ليس لها إجابة صحيحة أو خطأ .

وتستند القائمة إلى الإطار النظري لنظام الحاجات عند موراي Murray الذى أثر تأثيراً كبيراً فى بناء اختبارات الشخصية .

ثانياً : استبيانات تستند إلى الاستراتيجية الإمبريقية :

استبيان مينسوتا للشخصية متعددة الأوجه

Minnesota Multiphasic Personality Inventory (MMPI) :

يعد هذا الاستبيان من أكثر استبيانات الشخصية التي تستند إلى التقرير الذاتي ، والاستراتيجية الإمبريقية ، شيوعاً واستخداماً ، حيث أُجريت حوله عدة آلاف من الدراسات والبحوث ، وقد أعد هذا الاستبيان كل ما هاثاوأي Hathaway ، وماكنلي Mckinley بجامعة مينسوتا الأمريكية عام ١٩٤٢ ، بهدف التشخيص الكلينيكي لمرضى العيادات النفسية ، ويشتمل الاستبيان على (٥٥٦) عبارة من نوع الصواب أو الخطأ ، وتشير جميعها إلى الفرد Self– referenced مثل : أحب الطعام الجيد ، لا أعاني صعوبة عندما أريد النوم ، أصرخ بسهولة ، أشعر بالخوف عندما أنظر إلى أسفل من مكان مرتفع ، وقد تم تصنيف عبارات الاستبيان فى عشرة مقاييس فرعية كلينيكية ، واربعة مقاييس لضبط مدى صدق إجابات الفرد على العبارات .

وهذه المقاييس الكلينيكية كالتالي :

- توهم المرض Hypochondriasis .
- الانقباض Depression.
- الانحراف السيكوباتي Psychopathic Deviate.
- الذكورة – الأنوثة Masculinity – Feminity.
- البارانويا Paranoia .
- الوهن النفسي Psychasthenia.
- الفصام Schizophrenia .
- الهوس الخفيف Hypomania .
- الانطواء الاجتماعي Social Introversion .

ولبناء الاستبيان بدأ العالمان باختبار حوالي (٢٠٠٠) عبارة من بطاقات الفحوص الكلينيكية ، ومراجع الطب النفسي ، وأساليب الفحص العصبي ، ومن مقاييس الاتجاهات الشخصية والاجتماعية التى سبق بناؤها ، ثم قاما بتطبيق هذه العبارات على حوالي (٢٠٠) من المرضي المصابين بأمراض عصبية نفسية فى مستشفيات جامعة مينسوتا ، كما طبقت على حوالي (١٠٠٠) من الأفراد الأسوياء الذين تم اختيارهم من بين الذين كانوا يزورون أقاربهم المرضي فى هذه المستشفيات ، وطبقت العبارات أيضاً على عدة مئات من الطلاب المتقدمين للالتحاق بجامعة مينسوتا ، وكذلك على عينة من سكان إحدى المدن التابعة لهذه الولاية الأمريكية ، وقد تم اختيار العبارات التي اشتمل عليها الاستبيان في صورته النهائية على أساس تمييزها بين هؤلاء المرضى والأسوياء .

وتغطي هذه العبارات مدى واسعاً من الموضوعات التى تتناول جوانب الشخصية المختلفة ، مثل : الجوانب الصحية ، والاتجاهات الجنسية والاجتماعية، والدينية ، والسياسية ، والأسرية ، والمهنية ، والمخاوف المرضية ، والحالات الانفعالية المختلفة ، والروح المعنوية ، وما يتصل بالذكورة والأنوثة ، واتجاه الفرد نحو الاختبار .

ويُستخدم هذا الاستبيان فى صيغته العربية فى التشخيص الكلينيكي لمن هم فى مستوى المرحلة الثانوية وما بعدها ، ويمكن استخدامه كاختبار فردي ، أو جماعي .

ولتفسير الدرجات الخام فى جميع المقاييس الفرعية، ومقاييس ضبط نسبة الصدق ينبغي تحويلها إلى درجات تائية متوسطها(٥٠)، وانحرافها المعياري(١٠).

وقد وُجِّه بعض النقد إلى هذا الاستبيان لانخفاض ثبات بعض المقاييس التي يتكون منها ، وقد برَّر المؤلفان ذلك بأن الطرق السيكومترية التعارف عليها فى تقدير الثبات لا تصلح لاختبارات الشخصية، وحجتهما فى ذلك أن سمات

الشخصية سريعة التغير ، وأن البيانات المستمدة من إعادة تطبيق الاستبيان تعد مقياساً لتباين السمات ، وليست تقديراً لثبات الدرجات ، حيث يتوقع حدوث هذا التباين فى السمات لدى من يعانون من أمراض نفسية .

ومن المقاييس الأخري المعروفة التي استندت فى بنائها إلى الاستراتيجية الامبريقية استبيان كاليفورنيا السيكلوجي California Psychological Inventory (CPI) .

ويشتمل هذا الاستبيان على (٤٨٠) عبارة يشبه أكثر من ثلثها عبارات استبيان مينسوتا متعدد الأوجه ، وجميعها من نوع الصواب والخطأ ، ومن ميزات هذا الاستبيان أنه يطبق على الأفراد العاديين ، وأُجريت حوله كثير من الدراسات التطويرية لإثراء استخداماته ، وزيادة صدقه المتعلق بالتكوين الفرضي .

ثالثاً : استبيانات تستند إلى استراتيجية التجانس :

استبيان عوامل الشخصية الستة عشر لكاتل Cattell 16- PF Questionanire

استخدم " كاتل Cattell " أسلوب التحليل العاملي فى بناء هذا الاستبيان ، وقد سبق أن ذكرنا أن هذا الأسلوب يعد طريقة إحصائية لإرجاع العدد الكبير من معاملات الارتباط الملاحظة بين المتغيرات أو الاستجابات المختلفة لفقرات أو عبارات الاختبارات أو الاستبيانات إلى عدد قليل من العوامل غير المترابطة ، والفقرات التي تضم فى تجمع معين ترتبط ببعضها ارتباطاً عالياً ، وترتبط بغيرها من الفقرات التي فى التجمعات الأخري ارتباطاً منخفضاً ، أو لا ترتبط بها على الاطلاق ، ويعد تحديد الفقرات التي تُكَّون التجمعات أو العوامل المختلفة ، يمكن استخدامها كميزان Scale لقياس سمات الشخصية ، ويقوم معد الاختبار أو الاستبيان بإعطاء اسم للميزان فى ضوء محتوى الخاصة أو السمة التي يتضح أن الفقرات تشترك فيها ، ونظراً لأن هذه التسمية تعتمد على الأحكام الشخصية ، فإنه

من المحتمل أن يصل باحثون مختلفون إلى تسميات مختلفة للميزان نفسه ، غير أن الفقرات وليست التسمية فى ذاتها هي التي تُستخدم كتعريف للميزان .

ولقد أجرى " كاتل " تحليلاً عاملياً لاستبيان يشتمل على (٨٠) متغيراً طبقه على عينة تشتمل على (٣٧٠) فرداً، وتمكن من استخلاص (٢٠) عاملاً ، ولكنه استبعد عدداً من هذه العوامل لعدم وضوحها عندما أعاد النظر فى بناء المقياس ، وتبقي لديه (١٥) عاملاً من عوامل الشخصية، وعامل يدل على الذكاء العام ، وقد أضاف " كاتل " بعض الفقرات الآخري على أساس ارتباطها بهذه العوامل الستة عشر، وبذلك أصبح عدد فقرات الاستبيان (١٨٦) فقرة، ويتكون من صيغتين متكافئتين ، وكل عامل منها يقاس بعدد يتراوح بين (١٠-١٣) فقرة .

وآجري " كاتل " مرة أخري تحليلاً عاملياً على الاستبيان بعد إضافة عدد آخر من الفقرات لكل من صيغتيه المتكافئتين ، ومن أمثلة بعض عباراته " أحب مشاهدة الألعاب الرياضية الجماعية " ، ويستجيب الفرد للعبارات إما بكلمة "نعم" أو " لا " ويُنصح الفرد بالاستجابة بأول انطباع يأتي لديه ، بحيث تكون استجابته للموقف الذى تصفه العبارة هي استجابته لمعظم المواقف التي تشبه ذلك الموقف، كما يُنصح بعد الاستجابة بكلمة " أحياناً " إلا إذا لم يستطع الاختيار بين البديلين الآخرين .

ويمكن اعتبار هذا الاستبيان من نوع المقابلة المقننة المنتظمة غير الشخصية وهو محدد البنية ، لأن كل فقرة أو عبارة لها ثلاث استجابات ذاتية أو حرة Voluntary Responses ، كما أنه مباشر وواضح الهدف ، بمعني أن المختبر يعلم أن الاستبيان هو مقياس للشخصية ، ويمكن تقدير درجاته بطريقة موضوعية يدوياً ، أو باستخدام الحاسوب .

ويحصل الفرد على درجة لكل عامل من العوامل الستة عشر ، حيث يفترض أن هذه تعطي صورة عن شخصيته بوجه عام ، ويمكن الرجوع إلى دليل الاختبار

Test Manual لمعرفة المعني السيكلوجي لكل من هذه العوامل حيث تُفسر على أنها سمات الشخصية ، ومن بين هذه السمات : التحفظ فى مقابل الانطلاق ، الخضوع فى مقابل تأكيد الذات ، الخجل فى مقابل الجسارة والإقدام ، ويمكن رسم صفحة نفسيه (بروفيل) لدرجات الفرد تساعد فى التفسير والتشخيص والتنبؤ .

وقد بينت بعض الدراسات أن ثبات درجات عوامل أي من صيغتي هذا الاستبيان منخفض بعامة بسبب قلة عدد فقرات كل من الصيغتين ، كما تراوحت معاملات ثبات المقاييس الفرعية بطريقة التجزئة النصفية بين (٠.٩٣ ، ٠.٥٤) بينما معامل الثبات الناتج من إعادة تطبيق الاستبيان بعد حوالي أسبوع كان أقل من(٠.٨٠) أما البيانات الاحصائية المتعلقة بصدق المقياس فلم تكم كافية للاستفادة منها فى معرفة ما إذا كانت درجات الاستبيان تتنبأ بالفعل بالسلوك المطلوب قياسه، مما يستدعي مزيداً من الدراسات والبحوث للتوصل إلى أفضل طرق استخدام هذا الاستبيان فى المجالات الصناعية والتعليمية والكلينيكية.

تعقيب عام على استبيانات الشخصية :

يتضح من عرضنا لاستبيانات الشخصية أن هناك من الأدلة الإمبريقية ما يبين أن اتساق درجات هذه الاستبيانات ، سواء معاملات الاستقرار، أو التكافؤ أو الاتساق الداخلي ، أقل بوجه عام من غيرها من مقاييس الجوانب العقلية المعرفية، ولكن هل هذا الانخفاض يرجع إلى طبيعة الظاهرة المقاسة ، أم إلى عملية القياس ذاتها ؟ إن هناك من الأدلة أيضاً ما يؤكد أن الاستبيان الذى يُعتني ببنائه استناداً إلى أي من الاستراتيجيات الثلاث السابقة يمكن أن يحقق قدراً معقولاً من الاتساق الداخلي والاستقرار في درجاته .

أما فيما يتعلق بالصدق فإنه يبدو أن الصدق التكوين الفرضي أكثر ملاءمة فى معظم الأحيان لتقدير صدق استبيانات الشخصية ، إلا أنه لا يمكن التوصل

إلى هذا الصدق فى وقت قصير ، كما لا يمكن التعبير عن هذا النوع من الصدق بمؤشر واحد .

والسؤال الأخر هو : هل أمكن التوصل إلى ما يرتبط تجريبياً بدرجات استبيانات الشخصية ؟ والحقيقة أن معظم البيانات التي جمعتها الدراسات المتعددة فى هذا الشأن ارتبط كل منها بموقف معين ، واستُخلصت من عينات مختلفة فى أوقات مختلفة ولأغراض مختلفة ، بدلاً من كونها جزءاً من محاولات منظمة ومتكاملة لفهم معنى الدرجات وتفسيرها ، لذلك يصبح من الصعب دراستها واستيعاب دلالاتها بشكل متكامل .

أما المشكلة المهمة الأخري فتتعلق بتزوير الاستجابات ، أو ما يعرف بتحيز الاستجابات Response Bias ، والذى يبدو تأثيرها واضحاً فى استبيانات الشخصية اكثر من غيرها من المقاييس ، فعلي الرغم من الأبحاث المستفيضة التي اهتمت بضبط تحيز الاستجابات وبخاصة الاتجاهات العقلية فى الاستجابات ، إلا أنها لم تستطع استبعاد هذه المتغيرات من أي استبيان الشخصية ، وعلى الرغم من أن تحديد اتجاه ومقدار تحيز الاستجابات فى بعض الاستبيانات – كما فى مقياس ضبط مدى صدق الاستجابات – واستخدام هذه البيانات فى عملية تفسير الدرجات يعد خطوة مطلوبة نحو التغلب على هذه المشكلة ، إلا أن الحل الأفضل ربما يكون عن طريق التوصل غلى نمط معين من أنماط الاختبار بحيث يسهم فى الاستبعاد أثر تحيز الاستجابات .

والآن ما أكثر الاستراتيجيات فاعلية فى بناء استبيانات التقرير الذاتي للشخصية ، وإجابة هذا التساؤل تتطلب معرفة الغرض الذى يُبني من آجله الاستبيان ليكون أكثر فاعلية فالاستبيانات التي يستند بناؤها إلى الاستراتيجية الإمبريقية (مثل استبيان الشخصية متعدد الأوجه) تكون أكثر استخداماً ، وتهتم كثير من البحوث السيكولوجية بدراستها ، وهذه النتيجة يمكن أن تنطبق على

الاستبيانات التي تستند غلى استراتيجية التجانس (مثل استبيان كاتل للعوامل الستة عشر) ، حيث يجب أن تتميز هذه الاستبيانات بمزايا الاستبيانات التى تستند إلى الاستراتيجية الإمبريقية بالإضافة إلى أنها يمكن أن تمدنا بدرجات يسهل تفسيرها .

وبالطبع فإن أداة القياس المثالية يجب أن تشتمل على فقرات أو عبارات يتم انتقاؤها باستخدام الاستراتيجيتين معاً .

المقاييس الإسقاطية للشخصية :

إن بناء استبيانات الشخصية التي تعتمد على التقرير الذاتي لا يختلف كثيراً عن بناء مقاييس الجوانب المعرفية ، مثل : التحصيل ، والذكاء والاستعدادات الخاصة ، حيث تهتم أيضاً بالموضوعية والتقنين ، وربما يكون الاختلاف فقط فى محتوى المقياس ومحكات تقدير صدق درجاته ، غير أنه وجه كثير من النقد لهذه الاستبيانات الموضوعية فيما يتعلق بتركيزها على كيفية قياس السمات الشخصية أكثر من تركيزها على ما الذي تقيسه هذه الاستبيانات لذلك يرى بعض علماء النفس أنه ينبغي أستناد مقاييس الشخصية بدرجة أكبر على النظريات السيكولوجية، ولعل المقاييس الإسقاطية للشخصية حاولت ذلك من أجل التوصل إلى بيانات متكاملة يمكن أن تعطي صورة ديناميكية لشخصية الفرد بدلاً من التركيز على أبعاد مستقلة للشخصية كما في استبيانات الشخصية ، لذلك تعتمد المقاييس الإسقاطية على منهجية مختلفة ، وتمثل توجها مغايراً لاستبيانات الشخصية.

ويمكن أن تتح العلاقة بين الطرق الإسقاطية Projective Techniques والنظريات السيكوديناميكية Psychodynamics فى أسس التي بنيت على أساسها المقاييس الإسقاطية للشخصية ، فالنظريات السيكوديناميكية تؤكد الجوانب المزاجية الذاتية للفرد ، وتعتبر أن السلوك يحدث نتيجة تفاعل قوي متعددة ،مثل :

الدوافع ، والحوافز ، والحاجات ، والصراعات ، وأن تكون الشخصية يتضمن تنظيمات مختلفة ، مثل الشعر واللاشعور، وأن الشخصية هي العملية التي يفرض فيها الفرد تنظيماً وتكويناً على المثيرات الخارجية فى بيئته المحيطة به.

ويمكن أن نرى العلاقة بين الطرق الإسقاطية في تقييم الشخصية ، والنظريات السيكوديناميكية فى الشخصية فى خصائص هذه الطرق التى تتمثل فى :

(أ) تأكيد إعطاء حرية كافية للفرد لاختيار استجابته من بين عدد لا نهائي من البدائل ، وذلك لأن الاختبار يكون غير محدد البنية وذاتي .

(ب) تعليمات الإجابة ، والمثيرات التي يشتمل عليها الاختبار تعطي مؤشرات قليلة للاستجابة .

(ج) الغرض من الاختبار يكون غير واضح بالنسبة للفرد المختبر .

(د) يهدف الأخصائي الكلينيكي الذى يستخدم هذه الطرق للقيام بتفسيرات كلينيكية وشاملة من البيانات التى يجمعها .

وعندما يستخدم المصطلح " إسقاط Projection فى الطرق الإسقاطية للشخصية ، فإنه يعني بعامة العملية التي تؤثر بها خصائص الفرد ، أي بينة شخصية ، فى كيفية إدراكه ، وتنظيمه ، وتفسيره لبيئته وخبراته ، وهذه التأثيرات يمكن ملاحظتها بدرجة أفضل وقياسها عندما يستجيب الفرد لمثيرات أو مواقف تتميز بالغموض أو الحداثة .

وقد استخدم هنري موراي Murray عام ١٩٣٨ مصطلح الإسقاط Projection لأول مرة كطريقة لتقدير الشخصية ، وصمم اختباراً لهذا الغرض سنوضحه بعد قليل ، ولكن يرجع الفضل في تأكيد أهمية الطرق الإسقاطية إلى فرانك Frank عام ١٩٣٩ ؛ فقد عارض استخدام الاختبارات الموضوعية للشخصية بحجة أنها تُستخدم فى تصنيف الأفراد ، ولا تعطينا معلومات مفيدة عن الفرد كإنسان متمايز ، وحث على استخدام اختبارات تساعد في فهم أكثر وضوحاً

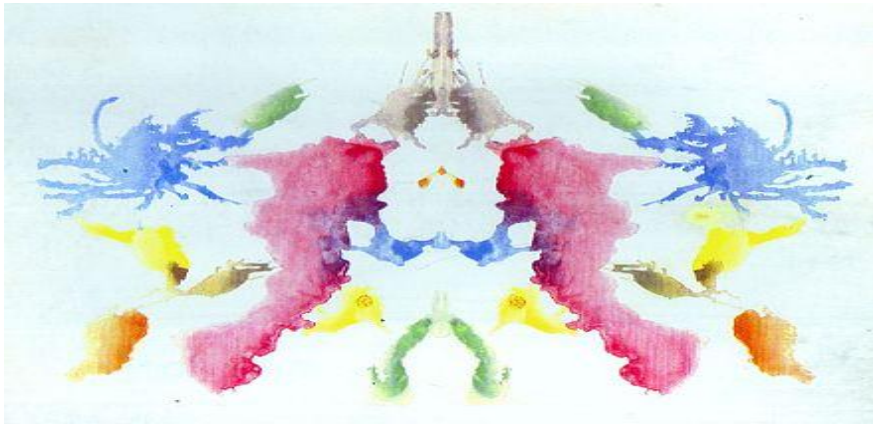
للمشاعر والأفكار الخاصة عند الأفراد ومثل هذه الاختبارات تسمح للفرد أن يفرض تكويناً وتنظيماً خاصاً به على المثيرات ، وبهذه تُعبر عن تصور ديناميكي للشخصية ، ويمكن من استجاباته التوصل إلى استدلالات فيما يتعلق ببنية شخصية .

غير أنه لا يوجد اتفاق كامل بين علماء النفس حول المظاهر الرئيسة للطرق الإسقاطية ، فقد قام ليندزي (Lindzey,1961) مثلاً بإجراء مسح للتعريفات المختلفة للطرق الإسقاطية ، وتوصل إلى أن الخصائص المميزة لهذه الطرق هي: حساسيتها للعمليات اللاشعورية أو المظاهر الكامنة للشخصية ، وتعدد استجابات الأفراد وذلك لتعدد أبعاد الشخصية التى تقاس فى أن واحد ، وعدم معرفة الفرد بالغرض من الاختبار ، وثراء الاستجابات التي تنتج عن المثيرات وعدم وجود استجابة صحيحة وأخري خطأ .

وعلى الرغم من تعدد أنواع الاختبارات الإسقاطية ، إلا أنني سأقتصر على تقديم مثالين فقد لاختبارين يعدان أكثر الاختبارات الإسقاطية شيوعاً واستخداماً فى الأغراض الكلينيكية ، والتي أُجريت حولها عدة آلاف من الدراسات والبحوث أشار إلى بعض منها الكتاب السنوي الأمريكي للقياس العقلي Buros` Mental Measurment Yearbook .

(١) اختبار بقع الحبر" لرورشاخ Rorschach Inkblot Test ":

صمّم " هيرمان رورشاخ " – الطبيب النفسي السويسري – اختبار بقع الحبر الذى سُمَّي باسمه ، وقد صمم مثيرات عن طريق وضع بقع الحبر على ورقة وقام بثني الورقة بحيث ينتج عنها أشكالاً متماثلة ولكن غير محددة البنية ، ثم قام بتطبيق هذا المثيرات على مرضي المستشفيات العقلية ، واحتفظ بالبقع التي استدعت استجابات مختلفة من مجموعات مختلفة من المرضي ، واستبعد البقع الأخري ، ويوضح الشكل التالي (٨ – ٣) صورة مماثلة لإحدى هذه البقع :



شكل (٨-٣) يوضح صورة مماثلة لإحدى بقع الحبر فى اختبار رورشاخ

وقد أستخدم رورشاخ في محاولاته ألافا من بقع الحبر حتى استقر فى النهاية على عشر بقع فى بطاقات ، وصمم طريقة لتقدير الدرجات ، وتم نشر الاختبار، وطريقة تقدير درجاته ووصف طريقة بنائه عام ١٩٢١، ولا تزال هذه البقع تُستخدم حتى وقتنا الحاضر فى صورة بطاقات على كل منها بقعة حبر تتخذ شكلاً معيناً غير محدد البنيه ، وأعيد النظر فى تقدير درجات الاختبار ، ولكن الجهود التي بذلها رورشاخ فى البداية ظلت أساسية فى معظم نظم تقدير درجات الاختبارات المستخدمة حالياً .

ولقد كان رورشاخ متأثراً بنظرية فرويد ، وما تتضمنه من مفاهيم ، مثل: اللاشعور وديناميكية الشخصية ، ولكنه لم يصمم الاختبار على أساس هذه النظرية ، ومع هذا فإن طبيعة بناء الاختبار يبدو أنها قد تأثرت بها بلا شك .

ويحاول الفاحص عند تطبيق بطاقات بقع الحبر أن يجعل المفحوص فى حالة عدم توتر وراحة تامة ، وتكون التعليمات كافية فقط بالقدر الذي يمكنه من استكمال استجاباته ، ولهذا فإنه يُطلب من المفحوص أن ينظر إلى كل بطاقة ، ويذكر للفاحص ما يراه فيها .

وأهم ما تتميز به هذه الطريقة هو حرية الفرد فى اختبار ما يراه ، وأين يراه ، ومحددات إدراكاته ، ويقوم الفاحص أثناء التطبيق بتسجيل الاستجابات التي يقدمها المفحوص ، كما يسجل المواضع المختلفة فى البطاقة التى ركز عليها المفحو فى استجاباته ، وما الذي جعل البقعة تأخذ هذا الشكف ، وبذلك يمكن للفاحص أن يحدد الأساس الذى بني عليه المفحوص استجابته بدقة أكبر .

ويتركز الاهتمام عند تفسير الاستجابات بطريقة تكوين هذه الاستجابات أو المدركات Percelts ومحددات الاستجابة ، ومحتواها ، وفي جميع الحالات يُفترض أن هناك تناظراً بين طريقة الفرد فى تكوين مدركاته ، والطريقة التى يتعامل بها بعامة مع المثيرات فى البيئة المحيطة به ، ويُقصد بذلك كيف ينظم ويشُكل هذه المثيرات ،فالمدركات الجيدة التنظيم ، والتي تتميز بخصائص تكوينيه جيدة فى علاقتها بتكوين بقعة الحبر ، ربما تؤدى غلى اقتراح أن الوظائف النفسية موجهة توجيهاً واقعياً ، وتعمل فى مستوى جيد ، أما المدركات التي تكوينها غير جيد ولا تناسب تكوين بقعة الحبر ، ربما تؤدي إلى اقتراح أن المفحوص يسلك سلوكاً وهمياً ، أو غير واقعي .

ويختلف تفسير شخصية الفرد ، تبعاً للمحتوى الذى يراه فى معظم البطاقات ، مثل الحيوانات أو الإنسان ، أو المحتوى الذى يعبر عن عدوان أو عاطفة ، فمثلاً

يمكن المقارنة بين تفسيرين ربما يصل إليهما الفاحص من مجموعتين من الاستجابات إحداهما رأي فيها الشخص حيوانات تتصارع ، والأخري رأي فيها أناساً يشاركون مشاركة كاملة فى عمل معين .

وفى كثير من الحالات يكون محتوى الاستجابات للبقع على شكل قصص ، ويفسرها الفاحص على أنها تعبَّر عن خصائص شخصية المفحوص ، وأسلوب حياته .

كما يمكن تفسير المحتوى بطريقة رمزية فمثلاً : الأنفجار ربما يرمز إلى عنوان شديد ، والثعلب ربما يرمز إلى الميل إلى العدوان ، والخنزير إلى الشره والنهم ، والعنكبوت والأخطبوط والساحر والغول ربما يرمز كل منها إلى اتجاهات سلبية نحو الأب المسيطر ، والنعامة ربما ترمز غلى محاولة الابتعاد عن الصراعات .

والأقسام العامة للمحتوى ، مثل : الطعام ، والحيوانات التي يؤكل لحمها ، والكائنات العدوانية ، والحيوانات المفترسة ، والشخصيات التي تمثل القوة والسلطة يكن تفسيرها تفسيراً رمزياً ، وتحليلياً .

كما يجب أن نتذكر أن التفسير لا يتم على أساس استجابة واحدة ، وإنما على أساس علاقة الاستجابة بالتنظيم الكلي للاستجابات .

ولكن يمكن أن تفيد كل استجابة فى اقتراح فروض أو تفسيرات محتملة عن شخصية الفرد ، وهنا ينبغي على الفاحص التحقق من هذه القروض ومراجعتها فى ضوء التفسيرات التى تعتمد على الاستجابات الأخري ، والتفسيرات التي تعتمد على النمط العام للاستجابات ، وكذلك التفسيرات التي تعتمد على سلوك المفحوص أثناء استجابته لاختبار " رورشاخ " وبالنسبة لهذه النقطة الأخيرة ، فإن الفاحص يقوم بتسجيل ملاحظاته عن سلوك المفحوص أثناء استجابته ، ويستخدم هذه الملاحظات كمصدر إضافي للبيانات عند تفسيره للاختبار. فالمفحوص الذى

يطلب دائماً العون والإرشاد من الفاحص ، وربما يفسر سلوكه هذا بأنه سلوك شخص غير مستقل .

والمفحوص الذى يبدو متوتراً ، أو يسأل أسئلة مراوغه ، ثم ينظر غلى الوجه الآخر البطاقة ، ربما يفسر هذا السلوك بأنه سلوك شخص متشكك ، أو ربما يكون مصاباً بالبارانويا .

ويعتمد تفسير أي أستجابة على مجموعة من خصائص هذه الاستجابة وهى: المحددات المستخدمة ، وخصائص المدرك الحسي ، المحتوى ، وما يقترن بالسلوك أثناء تقرر المدرك ، كما يعتمد التفسير أو التقرير بوجه عام على تحليل كمي ، وكيفي للاستجابات .

ثبات وصدق اختبار رورشاخ :

إن تنوع وثراء استجابات اختبار رورشاخ التي تستخدم فى تقدير معظم أو جميع أبعاد شخصية الفرد يؤدي إلى مشكلات تتعلق بتقدير ثبات وصدق الاختبار.

فمعظم طرق تقدير ثبات الاختبارات السيكومترية ، مثل: اختبارات ومقاييس الجوانب المعرفية ، واستبيانات الشخصية تكون مناسبة لأن هذه الاختبارات والمقاييس تؤدي غلى نتائج كمية .

أما فى حالة الاختبارات الإسقاطية مثل اختبار رورشاخ وغيره ، فإننا لا نستطيع الرجوع إلى جداول معايير Norms Tables لمعرفة الدلالة المباشرة لأقسام الدرجات كما هو الحال فى الاختبارات المقننة ، ولذلك فإن إحدى طرق تقدير ثبات اختبار رورشاخ هو تحديد مدى اتفاق مجموعة من المحكمين فى تقدير درجات الاختبار ، ولكن هذه الطريقة غير شائعة الاستخدام بسبب اختلاف نظم تقدير الدرجات ، كما أنها لا تقدم دليلاً على ثبات تفسيرات الدرجات ويؤدى

اختلاف طبيعة نظام تقدير درجات الاختبار إلى عدة مشكلات فى تقدير ثبات الدرجات بطريقة أعادة الاختبار ، أو بطريقة التجزئة النصفية ، ومن بين هذه المشكلات صعوبة تحديد ما إذا كان عدم استقرار الدرجات يرجع إلى عدم استقرار طريقة القياس ، أو إلى التغير المؤقت الذى يحدث فى التنظيم الانفعالي للمفحوص ، وتختلف بطاقات اختبار رورشاخ عن مفردات الاختبار السيكومتري، مما يجعل من الصعب تجديد تكافؤها ، كما أن عدد استجابات الأفراد لبقع الحبر يختلف من بطاقة إلى أخري ، كل هذا يؤدي إلى صعوبات فى تقسيم البطاقات إلى نصفين متكافئين .

والخلاصة أن عدم وجود أدلة كافية تؤكد ثبات اختبار رورشاخ يجعل من الصعب معرفة ما إذا كانت الاستجابات غير مستقرة حقيقة – أي غير ثابته – أم أنه يصعب القيام بدراسات مناسبة لتقدير ثبات تفسير هذا الاختبار .

أما التحقق من صدق اختبار رورشاخ ، فإنه يعد أمراً بالغ التعقيد ، إذ تزخر الأدبيات السيكولوجية بدراسات متناقضة فى هذا الشأن ، كما لا يوجد اتفاق بينها حول أفضل طريقة لتقدير صدق هذا الاختبار .

وإحدى طرق تقدير صدق الاختبار هو التحقق من مدى مضاهاة التفسيرات الناتجة عن الاختبار بالبيانات الناتجة من مصادر أخري ، فقد بينت بعض الدراسات وجود اتفاق بين تفسيرات اختبار رورشاخ ، والتقرير أو الوصف الكلينيكي ، بينما وجد آخرون اتفاقاً ضئيلاً بين تفسيرات الاختبار والتفسيرات التي تعتمد على الاختبارات الإسقاطية الأخري (Lindzey, 1959) .

ويرى بعض الباحثين أن الطرق المعروفة لتقدير صدق الاختبارات لا يمكن تطبيقها على اختبار رورشاخ . إذ يصعب الاتفاق على المحك الذى يمكن استخدامه فى التحقق من تفسيرات هذا الاختبار للمستويات أى اللاشعورية للسلوك.

(٢) اختبار تفهم الموضوع The Thematic Apperception Test (TAT) :

تختلف الاختبارات الإسقاطية فى مدى تحديد وانتظام بنية كل منها ، فالمثيرات التى تكون اختبار الرورشاخ تعد اكثرها غموضا وعدم انتظاما ، وتسمح بتنوع وثراء الاستجابات .

أما المثيرات التى يشتمل عليها اختبار تفهم الموضوع (TAT) فهى أكثر انتظاما والاستجابات تكون أكثر حديدا من اختبار رورشاخ ، ويتكون هذا الاختبار الذى أعده مورجان Morgen ، وموارى Murray عام 1935 من (30) صورة كل منها مرسومة على بطاقة منفصلة ، وبطاقة بيضاء ، وتختلف كل صورة من حيث محتواها ودرجة غموضها .

وتتباين هذه الصور فى درجة انتظامها وبنيتها ، إذ تتراوح بين الانتظام التام أو البنية المحددة بدرجة كبيرة إلى البنية غير المحددة على الإطلاق كما فى البطاقة البيضاء .

ويوضح الشكل (٨-٤) التالي إحدى هذه الصور :

شكل (٨-٤) يوضح بطاقة صورة مماثلة للصور فى اختبار تفهم الموضوع

ويمكن تطبيق الاختبار على الراشدين والمراهقين ، وتُعرض البطاقات على الفرد واحدة تلو الآخرى مصحوبة بتعليمات بسيطة ، ويُطبق الاختبار عادة على مرتين أو أكثر ، وتُستخدم فقط (٢٠) بطاقة مع أي فرد ، ويُطلب من الفرد الذى يختبر أن يكون قصة عما حدث في الصورة ، وعما يحدث فيها ، وما سوف يمكن أن يحدث ، ويجب أيضاً أن تتضمن القصة تفاصيلاً عن الشخصيات التي فى الصورة ، وأنشطتهم ، وتفكيرهم ، ومشاعرهم .

ويعتمد تقدير الدرجات على النظام المستخدم ، إلا أن معظم أنظمة التقدير تتضمن التعرف على بطل أو بطلة القصة (أي البطل الذى يتطابق معه الفرد المختبر ، أو يُسقط عليه شخصيته) وتحديد حاجاته أو حاجاتها ، وتحليل الضغوط أو القوى الاجتماعية التي تؤثر عليه أو عليها ، وهذا النظام التقديري مستمد مباشرة من نظرية مواري فى الشخصية .

ومما هو جدير بالذكر أن هناك بطاقات خاصة يمكن استخدامها مع الأطفال، ويسمي اختبار تفهم الموضوع للأطفال Children Apperception Test (CAT) .

وتوجد مجموعات أخري من البطاقات لقياس بعض متغيرات الشخصية مثل: البطاقات المعدلة لاختبار تفهم الموضوع ، والتي استخدمها ماكليلاند McClelland,1953 لدراسة دافعية الإنجاز Achievement Motivation .

ويختلف اختبار تفهم الموضوع عن اختبار رورشاخ في أنه يُستخدم فى الكشف عن جوانب الشخصية ، مثل : الدوافع ، والحاجات ، والعواطف والصراعات والخيالات بينما يستخدم اختبرا رورشاخ في الكشف عن بناء وتنظيم الشخصية ، ولا يهتم على وجه الخصوص بالمحتوى ، فليس من المهم كثيراً في هذا الاختبار أن يشير المفحوص إلى جزء معين من بقعة الحبر على أنه ثعبان،

أو حبل ، ولكن المهم هو مدى معقولية المدرك ، وهل تحدد عن طريق شكل أو لون هذه المنطقة من البقعة .

ويعد تقدير صدق اختبار تفهم الموضوع من الأمور المعقدة أيضاً كما هو الحال فى اختبار رورشاخ ، فهل القصص تعكس حقيقة شخصية الفرد المختبر، أم أنها مجرد ردود فعل نمطية للموقف المصور ؟ وحتى إذا افترضنا أن القصص تعكس حقيقة شخصية الفرد ، فهل افتراض أن الفرد يتطابق مع البطل أو البطلة ، وأن مشكلات البطل أو البطلة ، والضغوط المؤثرة على أي منهما تعكس بالفعل جوانب حياة الفرد المختبر ، يعد افتراضاً صحيحاً ؟

ومما يزيد من تعقد المشكلة عدم الاتفاق على طريقة واحدة لتقدير الدرجات ونظام تفسيرها ، كما أن تقدير الصدق الخارجي للاختبار يعد من المشكلات التي لم تحسم بعد .

وعلى أية حال ، فإن اختبار رورشاخ ، واختبار تفهم الموضوع يعد كلاً منهما أداة كلينيكية تتطلب أخصائيا متمرساً لاستخدامها ، وتفسير استجابات المختبرين، ويفضل كثير من علماء النفس الكلينكي استخدام الاختبارين معاً للتوصل إلى فهم متكامل للشخصية .

وينبغي الإشارة إلى أن هناك اختبارات إسقاطية كثيرة أخرى تعتمد على طريقة التداعي مثل : اختبار تداعي الكلمات Word Association Test الذى أعده كنت ، ورسانوف Kent & Rosanoff ، وطريقة الترتيب Ordering ، وطريقة التعبير مثل : تمثيل الأدوار ، والسيكودراما ، واللعب بالدمي ، وطريقة الرسم ، مثل اختبار رسم الرجل Draw A–Person Test ، واختبار تذكر وإنتاج تصميمات هندسية معينة من الذاكرة بعد عرض تسعة أشكال هندسية واحداً تلو الآخر كل على بطاقة منفصلة ، ويسمي اختبار بندر ، وجستلت Bender & Gestalt Test ، وغيرها .

تعقيب عام على المقاييس الإسقاطية للشخصية :

يتضح مما سبق أن من أهم مميزات المقاييس الإسقاطية للشخصية حرية الاستجابة المسموح بها للفرد المختبر ، وحرية الأخصائي الكلينكي في سير وتوضيح ما يقصده الفرد ، ومعظم هذه المقاييس تكون غامضة الهدف ، وبذلك يصعب على الفرد المختبر تزوير استجاباته ، لأن طرق تقدير الدرجات ، وتفسير الاستجابات تكون غير مألوفة لديه ، كما أن معظمها لا يحتاج إلى مهارة فى القراءة ويمكن أن تطبق على الأميين والأطفال الصغار ، إلا أن هذه المقاييس يشوبها كثيراً من العيوب ، فهي تطبق بطريقة فردية ، ولذلك فهي مكلفة فى التطبيق وفى تقدير الدرجات والتفسير ، ومما يزيد الأمر تعقيداً عدم وجود أدلة كافية على جدوي استخدامها ، وثباتها ، وصدقها .

ويجب أن يقوم بتطبيق هذه المقاييس ، وبخاصة تلك التى تكون بنيتها غير محددة ، مثل اختبار رورشاخ ، واختبار تفهم الموضوع ، والسيكودراما ، واللعب واختيارات الرسم أخصائيون متمرسون على كيفية استخدام هذه الاختبارات وتفسير نتائجها .

اختبار وتقييم الاختبارات والمقاييس

- مقدمة
- اعتبارات عملية عند اختيار الاختبارات والمقاييس
- اعتبارات فنية عند اختيار الاختبارات والمقاييس
- دليل الاختبار
- مصادر الحصول على معلومات عن الاختبارات
- المقاييس المنشورة

❖ مقدمة :

تناولنا فى الفصول السابقة كيفية قياس السمات أو الخصائص الإنسانية المتعددة المتمثلة فى التحصيل ، والذكاء العام ، والاستعدادات والشخصية ، وأوضحنا طبيعة الاختبارات والمقاييس المناسبة لها ، وأنواعها المختلفة ، وخصائصها السيكومترية ، وكيفية تفسير درجاتها ولعل التساؤل المهم الذى يبرز الآن يتعلق بكيفية اختيار الاختبار أو المقياس المناسب لغرض أو أغراض معينة من بين العديد من الاختبارات المنشورة والمتوافرة ، بحيث يمكن الاستناد إلى نتائجها في اتخاذ القرارات المتعلقة بالأفراد ، أو الجماعات أو البرامج بدرجة عالية من الثقة وكذلك الأسس التي يمكن الإستناد إليها في تقييم هذه الاختبارات والمقاييس التي يتم اختيارها للغرض أو الأغراض المعينة .

لذلك سوف يتناول هذا الفصل بعض العوامل أو الاعتبارات العملية والفنية التي ينبغي مراعاتها عند اختيار الاختبارات والمقاييس لاغراض معينة والمعايير اللازم توافرها فى أدلة الاختبارات ، ومصادر الحصول على معلومات عن الاختبارات والمقاييس المنشورة .

❖ اعتبارات عملية عند اختيار الاختبارات والمقاييس :

عند اختيار اختبار أو مقياس لاستخدامه فى غرض معين ، ينبغي مراعاة بعض الاعتبارات العملية التي تؤثر فى هذا الاختبار ، لعل أهمها ما يلي :

(١) **السمة المراد قياسها، والغرض من القياس :** فقد سبق أن أوضحنا فى كثير من المواضع أن تحديد السمة المراد قياسها يعد خطوة أساسية توجه عملية اختيار أداة القياس المناسبة لهذه السمة ، فإذا كنا نود قياس الذكاء العام ، فإننا سوف نقتصر على الاختبارات والمقاييس العقلية الفردية أو الجماعية التي تقيس العامل العام (G) ؛ وإذا كنا نود قياس الشخصية ، فإننا ينبغي تحديد السمات التي نهتم بها أكثر من غيرها ، لأن ذلك يجعل اختيار استبيان الشخصية المناسب أكثر تحديداً ، وهكذا ، كما ينبغي مراعاة

الغرض من القياس سواء كان من أجل الانتقاء أو التسكين ، أو التصنيف، أو التشخيص مثلاً ، وهذا التحديد يعد أمراً ضرورياً لأنه يوجه عملية الاختبار الوجهة الصحيحة ، فالاختيار الذى يهدف للانتقاء يختلف فى محتواه وتفسير درجاته عن محتوى الاختبار الذي يُصمم من أجل التشخيص ، حيث يشتمل فى هذه الحالة على مفردات تقيس مختلف عناصر المحتوى أو مجالاته الفرعية ، ويمدنا بدرجات لكل من هذه المجالات .

(٢) **طبيعة المجموعة المستهدفة من الأفراد :** فقد لاحظنا فى الفصول الأربعة السابقة أن اختبارات الراشدين تختلف عن اختبارات الأطفال والأميين مثلاً؛ فالاختبارات تُصمم عادة بحيث تناسب فئة أو فئات معينة ، وأعمار مختلفة ، لذلك فإن تحديد طبيعة المجموعة المراد اختبارها يُيسر عملية اختيار الاختبار أو الاختبارات المناسبة لها ، وعندما يتباين أداء الأفراد بتباين خصائص معينة ، مثل : العمر ، والتعليم ، والمستوى الاجتماعي الاقتصادي ، وغير ذلك ، فإنه ينبغي مراعاة هذه الخصائص عند اختيار الاختبار أو المقياس .

(٣) **كلفة العملية الاختبارية :** فتطبق الاختبارات التحصيلية المقننة على نطاق واسع فى المدارس – كما هو الحال فى كثير من الدول – يعد عملية مكلفة ولا تقتصر الكلفة على شراء كراسة الاختبار ، وأوراق الإجابة ، وإنما أيضاً تتضمن تصحيح الاختبار ، وتقرير نتائجه ، وتزداد الكلفة عندما تُستخدم كراسة الاختبار مرة واحدة ، حيث تتطلب الإجابة ووضع علامات على كراسة الاختبار مباشرة ، كما أن اعتماد الاختبار على أجهزة معينة أو قاعات خاصة ربما يضيف مزيداً من كلفة تطبيق الاختبار ، غير أنه لا ينبغي التغاضي عن استخدام اختبار جيد بسبب كلفته العالية .

(٤) **زمن تطبيق الاختبار** : فقد سبق أن أوضحنا أن بعض الاختبارات وبخاصة اختبارات الذكاء الفردية ، وبعض استبيانات ومقاييس الشخصية يستغرق تطبيقها كثيراً من الوقت ، وهناك مواقف عملية تتطلب تطبيق اختبار معين في زمن لا يزيد عن ساعة واحدة مثلاً ، فعندئذ ينبغي مراعاة ذلك عند اختيار أو المقياس ، وعلى الرغم من أن الاختبار القصير يكون أقل كلفة، ويستغرق تطبيقه وتصحيحه وقتاً محدوداً ، إلا أن ثبات درجاته يكون منخفضاً لقلة عدد مفرداته ، لذلك ينبغي مراعاة التوازن بين زمن تطبيق الاختبار ، وثبات درجاته ، عند اختيار الاختبار المناسب .

(٥) **سهولة تطبيق الاختبار** : عند اختيار اختبار أو مقياس معين ينبغي مراعاة مدى سهولة التطبيق ، وكفاية التعليمات الموضحة بدليل الاختبار ، فكلما كانت تعليمات التطبيق كاملة ، أمكن جعل عملية التطبيق منتظمة ، فالتعليمات التفصيلية لكل من الفرد المختبر والقائم بتطبيق الاختبار ، تُيسر العملية الاختبارية ، وينبغي أن تشتمل هذه التعليمات على معلومات تتعلق بزمن الاختبار ، إذا كان من النوع الذى يتطلب السرعة فى الإجابة ، وذلك لكل قسم من أقسامه ، وتنظيم المقاعد ، وكيفية المراقبة أثناء أداء الاختبار ، ونوع القلم الذي يُستخدم فى وضع العلامات في ورقة الاجابة ، وطرق تصحيح أخطاء الإجابة سواء بالشطب أو بوضع علامة (×) على الإجابة المراد تصحيحها ، ومستوى الإضاءة ودرجة حرارة الغرفة ومتطلبات تهويتها، والإجراءات التى تتبع إذا خالف فرد شروط العملية الاختبارية ، وكذلك الاجراءات التي تُتبع فى حالات الطواري ، وطريقة توزيع أوراق أو كراسات إجابة الاختبار وجمعها ، وطرق تصحيح وتقدير درجات هذه الأوراق أو الكراسات ، ويجب على مستخدم الاختبار قراءة دليل الاختبار بعناية ويتخذ قراراً فيما يتعلق بكفاية تعليمات تطبيقه .

(٦) **ظروف تطبيق الاختبار** : فالظروف الفيزيائية للعملية الاختبارية تعد الأمور التي ينبغي مراعاتها ، مثل : تنظيم مقاعد غرفة الاختبار ، وإضاءة الغرفة، ونظام التهوية ، والحرارة ، وغير ذلك ، فهذه الظروف تؤثر بدرجة ما فى درجة الفرد فى الاختبار ، وينبغي أن يُختبر جميع الأفراد فى ظروف تماثل بدرجة كبيرة الظروف التى كانت متواجدة وقت تقنين الاختبار .

(٧) **الجوانب النفسية للمختبرين** : فالاتجاه الإيجابي للأفراد نحو الاختبار والعملية الاختبارية ، يعد من الاعتبارات المهمة ، فالاختبار الذى لا يراعي عند تطبيقه الجوانب النفسية للمختبرين يؤدي إلى نتائج غير موثوق بها ، فقلق الاختبار مثلاً يؤثر فى الأداء الأقصى للمختبرين ، وهذا يلقي دوراً على القائم بتطبيق الاختبار لتهيئة الأفراد للاختبار من الناحية الانفعالية ، وزيادة دافعيتهم لبذل أقصي جهدهم دون قلق ، فإذا أدرك المختبرون الفائدة التي تعود عليهم من النتائج الدقيقة للاختبار ، فإن ذلك سوف ينعكس إيجابياً علي تهيئتهم التهيئة الانفعالية المناسبة .

(٨) **صيغة الاختبار** : فالاختبار الذي تكون صيغته ، وطريقة تنظيمه متناسبة مع طبيعة المجموعة المستهدفة ، بحيث يراعي مستوى نضج أفراد المجموعة ، وعدم ازدحام الاختبار بالمفردات ، ووضوح الصور والأشكال ، وضم جميع التعليمات فى مكان واحد ، كل ذلك يزيد من دافعية المختبرين لبذل أقصي جهدهم فى الاختبارات المقننة ، فمثلاً دليل اختبار الاستعدادات الفارقة (DAT) الذى سبق أو أوضحناه فى الفصل الخامس يفصل بوضوح بين تعليمات تطبيق الاختبار ، وطريقة تصحيحه ، ومعاييره ، ويضع عناوين جانبية واضحة لكل من هذه العناصر .

(٩) **طريقة التصحيح** : توجد أنواع مختلفة من طرق تصحيح الاختبارات والمقاييس ؛ فالاختبارات التي تتطلب ورقة إجابة منفصلة يمكن تصحيحها يدوياً أو آلياً باستخدام الماسح الضوئي Optical Scanner ، وقد

أصبحت هذه الماسحات شائعة الاستخدام فى الآونة الأخيرة لتصحيح الأف من الأوراق إجابة الاختبارات التي تشتمل على مفردات الاختيار من متعدد ، أو الصواب أو الخطأ ، وذلك في وقت قصير ، كما أن بعض الاختبارات يتم تصحيحها ، وتقرير نتائجها آلياً باستخدام الحاسوب .

(١٠) **توافر صيغ متكافئة من الاختبار**: فهذه الصيغ المتكافئة Alternate Forms تساعد فى إعادة اختبار الأفراد فى أوقات مختلفة دون الحاجة إلى تطبيق الاختبار نفسه ، مما يجد من مشكلة سرية الاختبارات ، فاستخدام صيغة مكافئة يجعل من الصعب تذكر مفردات الصيغة الأولي التي سبق تطبيقها على الفرد نفسه ، لذلك ينبغي مراعاة الهدف من العملية الاختبارية، واتخاذ قرار عما إذا كانت تستلزم توافر صيغتين متكافئتين من الاختبار نفسه ، وذلك يوجه عملية اختيار الاختبار المناسب .

(١١) **سهولة تفسير درجات الاختبار** : ينبغي أن يتضمن دليل الاختبار معلومات تُيسِّر تفسير درجات الاختبار ، وهذا يتطلب تقديم معلومات تتعلق بمعايير الاختبار Test Norms ، ومزاياها ، وعيوبها ، وكذلك أمثلة لتفسيرات غير صحيحة من أجل العمل على تجنبها ، وأمثلة أخري لتفسيرات مناسبة ، وكثير من التفسيرات غير الصحيحة ربما ترجع إلى أمور فنية تتعلق بثبات الدرجات ، وصدق الاختبار ، كما سيتضح بعد قليل .

❖ اعتبارات فنية عند اختيار الاختبارات والمقاييس :

إن مراعاة الاعتبارات العملية السابقة عند اختيار الاختبارات والمقاييس ليست كافية لاتخاذ قرار فى هذا الشأن ، فإذا كان هناك اختباران أو أكثر يختلفان اختلافاً كبيراً فى هذه الاعتبارات العملية ينبغي التفكير جيداً فى المزايا النسبية لكل منها إذا استُخدم فى الموقف الاختباري المعين ، إذ ليس من الحكمة أحياناً اختيار اختبارا أو مقياس على أساس سهولة تطبيقه ، وتصحيحه مثلاً ، فى حين أنه

يفتقر إلى الصدق والثبات ، مما يجعل نتائج الاختبار غير موثوق بها ، وبذلك يكون مضيعة للوقت والجهد .

لذلك ينبغى عند اختيار الاختبارات والمقاييس مراعاة اعتبارات فنية أيضاً فيما يتعلق بالأدلة المقارنة لكل من الصدق والثبات ، وكفاية المعايير ، وفيما يلي توضيحاً لكل من هذه الخصائص :

(١) **صدق الاختبار** : يُعد الصدق Validity وكيفية تقديره من الاعتبارات الفنية المهمة التي ينبغي التحقق منها عند اختيار اختبار لغرض معين ، وقد سبق أن أوضحت مفهوم الصدق وطرق تقديره فى الفصل الرابع ، فالاختبارات المقننة والاستبيانات ينبغي أن تكون صادقة فيما تقيسه ، بما يؤدي إلى الثقة فى استخداماتها لأغراض معينة .

فالاختبارات التحصيلية مثلاً : ينبغي أن تتميز بصدق محتواها ، بحيث تقيس التحصيل فى مجال دراسي معين ، وتشتمل على عينة ممثلة لنطاق المهارات التي يتضمنها هذا المجال ، وهذا يعني وجود علاقة بين التعلم المراد تقييمه ، والمواد الاختبارية التي يُفترض أنها تقيس نواتج هذا التعلم ، لذلك فإن توافر بيانات تتعلق بصدق الاختبار يعد أمراً ضرورياً لتقديم أدلة عن أنواع الصدق المناسبة لاستخدامات الاختبار ، وهذه البيانات تكون مذكورة عادة فى دليل فني للاختبار Technical Manual ، لذلك ينبغي تقييم صدق الاختبار فى علاقته بالقرارات أو الأغراض التي سوف يُستخدم فيها الاختبار .

(٢) **ثبات درجات الاختبار** : يعد الثبات Reliability خاصية تلي الصدق فى أهميتها ، وينبغي أخذها بعين الاعتبار عند اختيار اختبار أو مقياس معين، فالغرض من استخدام نتائج الاختبار سوف تحدد مستوى الثبات المقبول ، فإذا كان الغرض من هذا الاستخدام تقييم الفرد ، ينبغي أن يكون معامل الثبات أعلى بكثير مما لو كان الغرض هو التقييم الجماعي ، ففي الحالة الأولي ربما

يُفضل أن تقترب قيمة معامل الثبات من (٠.٩٠) ، بينما يكفي فى الحالة الثانية أن تكون هذه القيمة (٠.٥٠) أو أكبر قليلاً (Crunbach,170; Seibel,1968) ، وبالطبع ينبغي أن يبين دليل الاختبار كيفية تقدير ثبات درجات الاختبار ، فمعامل الثبات يعد مهماً في المقارنة بين الاختبارات لأنه يكون مؤشراً لاتساق ما يقيسه الاختبار ، وكذلك يرتبط الخطأ المعياري للقياس Standard Error of Measurement بمفهوم الثبات ، وذلك لأنه يعد مقياساً إحصائياً مهماً إذا أردنا أن تُفسر درجات الاختبار بطريقة واحدة .

وكلما كانت قيمة معامل الثبات مرتفعة ، والخطأ المعياري للقياس منخفضاً ، فإن الاختبارات يكون لها ثبات أكبر .

لذلك ينبغي معرفة إجراءات تقدر ثبات درجات الاختبار ، وحجم العينة ، وقيمة معامل الثبات الذى يتحدد فى دليل الاختبار ، وكذلك تقدير الخطأ المعياري للقياس ، وتحديد مصادر أخطاء القياس .

(٣) **معايير الاختبار** : وتتعلق بتفسير درجات الاختبار ، وتعتمد هذه المعايير Norms على حجم عينة التقنين الذي ينبغي أن يكون كبيراً نسبياً ، لكي يزداد احتمال تمثيلها للمجتمع المستهدف ، وينبغي أن تراعي هذه العينة المناطق الجغرافية المختلفة والمستويات الاقتصادية والاجتماعية ، والنوع ، والعمر ، والفرق الدراسية ، والخلفية الثقافية ، وغير ذلك ، وهذا يساعد فى التوصل إلى استنتاجات صادقة من جداول المعايير عندما تُطبق على مجموعة أخري مماثلة لعينة التقنين ، وعادة يُوضح الدليل الفني للاختبار خصائص عينة التقنين ، وتاريخ جمع البيانات المتعلقة بالمعايير ، ونوع الدرجات المحولَّه المستخدمة ، أي : الدرجات التائية ، أو الرتب المئينية ، أو التساعيات المعيارية ، أو نسب الذكاء الانحرافية ، أو غير ذلك .

وينبغي عند اختيار اختبار أو مقياس معين مراعاة أن تكون المعايير حديثة بحيث تواكب التغيرات المتسارعة فى مجالات التعليم والمهن المختلفة ، وأن تكون

قابلة للمقارنة بين الاختبارات المختلفة ، وبخاصة عندما تُقارن الصفحات النفسية (البروفيلات) لأغراض التشخيص ، كما يجب وصف هذه المعايير وصفاً تفصيلياً يوضح طبيعة عينة التقنين ، وإجراءات الحصول على جداول الدرجات المحولَّة ، وينبغي أن يشتمل ذلك على أسلوب المعاينة ، وعدد أفراد العينة المستخدمة وتوزيعها ، وخصائصها الديموغرافية ، وتاريخ الحصول على هذه المعايير فمعرفة جميع هذه المعلومات يساعد فى الحكم على مدى ملاءمة معايير الاختبار للغرض أو الأغراض المعينة .

❖ دليل الاختبار The Test Mannal :

يُعد دليل الاختبار هو المصدر الرئيس للمعلومات المتعلقة بجودة ونوعية الاختبار أو المقياس المنشور ، وعادة يكون ملحقاً به كتيباً يتضمن البيانات والخصائص السيكومترية أو الفنية للاختبار .

وينبغي أن يكون دليل الاختبار واضحاً بدرجة كافية بحيث يمكن لمستخدمي الاختبار مهما اختلفت مؤهلاتهم فهم محتوياته ، كما ينبغي أن تكون المعلومات التي يشتمل عليها دقيقة بحيث تفي باحتياجات المتخصصين فى البحوث الاختبارات والمقاييس .

ونظراً لأهمية أدلة الاختبارات أعدت لجنة مشتركة من أعضاء الجمعية الأمريكية لعلم النفس (APA) ، والجمعية الأمريكية للبحث التربوي (AERA) ، والمجلس القومي الأمريكي للقياس التربوي (NCME) مجموعة من المعايير Standards يمكن الاسترشاد بها عند إعداد الاختبارات والمقاييس التربوية والنفسية ، وأدلتها ، وهذه المعايير تتجدد باستمرار نظراً للتطورات المتسارعة فى علم القياس النفسي والتربوي ، ولكن ليس من الضروري من الوجهة القانونية أن يلتزم ناشرو الاختبارات بهذه المعايير ، لذلك ربما نجد اختباراً أو مقياساً منشوراً بصحبة دليل غير جيد .

غير إنه إذا استوفي دليل الاختبار هذه المعايير ، فإنه مشتملاً على المعلومات الضرورية التي تساعد فى اختبار المناسب لاستخدام معين ، وقد صنُفَّت هذه المعايير فى العناصر التالية : (APA, 1999)

أ) نشر المعلومات :

(١) عند نشر اختبار للاستخدام العملي ، ينبغي أن يصاحبه دليل أو مجموعة من الأدلة التى توفر لمستخدمي المعلومات التي يحتاجونها لكي يتمكنوا من تفسير درجات الاختبار .

(٢) ينبغي تحديث الاختبار ودليله أو أدلته بصورة منتظمة ، نظراً لأن المعلومات المتعلقة بالاختبار تتقادم سريعاً ، وبخاصة مع التغير المستمر فى الظروف الاجتماعية ، ومواصفات الأعمال ، والضغوط التربوية ، وهذا يؤدي إلى تغير تفسير درجات بعض الاختبارات ، وربما ينبغي إيقاف استخدام الاختبار إذا لم تُجر مراجعة على دليل الاختبار فى غضون (١٥) عاماً منذ نشره .

ب) تفسير درجات الاختبار :

(١) ينبغي أن يساعد الاختبار، ودليله، ومواده المصاحبة مستخدمي الاختبار في تفسير درجاته تفسيراً صحيحاً له معنى .

(٢) ينبغى أن يحدد دليل الاختبار بوضوح أغراض استخدام الاختيار وتطبيقاته التي يُوصي بها .

(٣) ينبغي أن يحدد دليل الاختبار المؤهلات المطلوبة لتطبيق الاختبار، وتفسير درجاته تفسيراً مناسباً.

ج) صدق الاختبار :

(١) ينبغى أن يوضح دليل الاختبار صدق الاختبار لكل نوع من الاستدلالات التي يوُصي بها.

(٢) إذا صُمَّم الاختبار بدرجة أساسية لمعاونة مستخدمي الاختبار فى إصدار أحكام ذات طبيعة تنبؤية ، فانه ينبغى تقدير الصدق التنبؤى للاختبار بعناية للاختبار بعناية .

(٣) إذا صُمَّم الاختبار بدرجة أساسية لتقديم معلومات تتعلق بالتحصيل السابق للفرد ، فإنه يجب تقدير صدق محتوى الاختبار ، لأن هذا النوع من الصدق يكون أكثر ملاءمة فى هذه الحالة.

(٤) إذا تم تقدير الصدق التنبؤى أو التلازمي للاختبار ، فإنه ينبغى وصف المقاييس المحكمة Measures Criterion التى استُخدمت فى هذا الشأن بعناية ، وكذلك تقرير معلومات تصف العينة المستخدمة فى دراسة الصدق ، وظروف إجراء الاختبار، وهذا يساعد مستخدمي الاختبار فى الحكم على مدي ملاءمة البيانات المتعلقة بصدقة للأغراض المناسبة لهم .

(٥) ينبغي عرض نتائج التحليل الإحصائي الذى استُخدم فى تقدير الصدق المرتبط بمحك بطريقة واضحة ومفهومة .

(٦) إذا اقترح مؤلف الاختبار تفسير درجاته كمقياس لتغير نظري (قدرة أو سمة ، أو اتجاه معين) ، ينبغي تحديد التفسير المقترح تحديداً كاملاً ، فمثلاً إذا اقتُرض أن الاختبار يقيس تكويناً فرضياً Construct ، مثل القلق ، فإنه ينبغي أن يُميز دليل الاختبار بين تعريف القلق المُستخدم ، والتعريفات المحتملة الأخري .

د) ثبات درجات الاختبار :

(١) ينبغي أن يُقدم دليل الاختبار معلومات عن ثبات درجاته ، بحيث تسمح لمستخدمي الاختبار الحكم على ما إذا كانت الدرجات يمكن الاعتماد عليها بدرجة كافية للغرض الذي سوف يُستخدم الاختبار من أجله ، وإذا لم يتم جمع أي معلومات أو أدلة ضرورية ، فإنه ينبغي ذكر المعلومات الناقصة.

(٢) إذا قدَّم دليل الاختبار تقديرات الثبات ، أو تقديرات أخطاء القياس ، فإنه ينبغي تحديد الإجراءات التي استُخدمت فى الحصول على هذه التقديرات ، وكذلك وصف عينة الأفراد الذين طُبِّق عليهم الاختبار ، وخصائص العينة ، مثل : مستوى نُصح الأفراد ، وتباين قدرتهم ، واتجاههم نحو الاختبار ، وغير ذلك ، حيث إن هذه المتغيرات ربما تؤثر في هذه التقديرات .

(٣) إذا تم الحصول على معاملات الثبات من مجموعة غير متجانسة من الأفراد، وأراد مستخدم الاختبار تطبيقه على مجموعة متجانسة ، فإنه يجب ألا يتوقع الحصول على نفس قيمة معامل الثبات الذى حدده دليل الاختبار.

(٤) إذا كان للاختبار أكثر من صيغة واحدة ، فإنه ينبغي أن يُقدم دليل الاختبار بيانات فيما يتعلق بتكافؤ هذه الصيغ الأخري .

(٥) إذا صُمَّم الاختبار لقياس سمة عامة متجانسة ، فإنه ينبغي أن يُقَّدم دليل الاختبار ما يُوضح الاتساق الداخلي لدرجاته .

(٦) إذا استُخدم الاختبار في أغراض التنبؤ ، أو مقارنة أداء الفرد عبر الزمن ، فإن معامل الاستقرار بطريقة إعادة الاختبار يكون ضرورياً ، ويتطلب تقديراً لقيمته ، وتقرير ذلك فى دليل الاختبار .

هـ) تطبيق الاختبار ، وتصحيحه :

(١) ينبغي أن تكون تعليمات الاختبار واضحة بدرجة كافية فى دليل الاختبار بحيث يمكن اتباعها ، وتكرارها كما فى عينة التقنين .

(٢) ينبغي أن تكون إجراءات تصحيح الاختبار محددة وتفصيلية وواضحة فى دليل الاختبار للتقليل بقدر الإمكان من أخطاء التصحيح .

و) الموازين والمعايير :

(١) الموازين المستخدمة فى تقرير نتائج الاختبار Scales ينبغي أن توصف بعناية في دليل الاختبار لزيادة احتمال التفسير الصحيح للدرجات .

(٢) إذا أُجريت مراجعة لموازين الاختبار ، أو أُضيفت جديدة ، أو أي تعديلات أخري ، فإن دليل الاختبار المنقُح ينبغي أن يُقدم الجداول التي توضح التكافؤ بين الموازين الأصلية والموازين الجديدة .

(٣) معايير الاختبار Test Normsينبغي أن توصف بعناية ، بحيث يمكن لمستخدمي الاختبار تحديد ما إذا كانت عينة التقنيين مماثلة للمجموعة التي سوف يُستخدم معها الاختبار .

وهذا يتطلب أن يشتمل دليل الاختبار على وصف دقيق للمجتمع المستهدف الذى يُقارن به الأفراد المختبرون ، ومتوسطات درجات الاختبار ، وانحرافاتها المعيارية ، وكذلك الخصائص الديموغرافية لهذا المجتمع .

(٤) ينبغي أن يشتمل دليل الاختبار على جداول المعايير التي تحتوى على الدرجات الخام فى الاختبار ، والدرجات المحولة المناظرة لها ، مثل الدرجات المعيارية أو التائية ، أو التساعيات المعيارية ، أو الرتب المئينية التي تعكس توزيع الدرجات فى مجموعة أو مجموعات معيارية مناسبة من الأفراد .

(٥) ينبغي أن يشتمل دليل الاختبار على وصف متكامل لأي صفحة نفسية (بروفيل) يمكن تكوينها من بطارية تشتمل على عدد من الاختبارات الفرعية.

(٦) ينبغي نشر البيانات المتعلقة بمعايير الاختبار فى نفس وقت نشر الاختبار.

❖ مصادر الحصول على معلومات عن الاختبارات :

يعتمد اختيار الاختبار أو المقياس المناسب لغرض معين على معرفة مستخدمي الاختبارات مصادر الحصول على معلومات عن مختلف أنواع الاختبارات والمقاييس الأجنبية والعربية المنشورة ، فالفرد الذى يود استخدام اختبارات معينة يحتاج إلى معرفة الاختبارات المتوافرة التى تساعده فى الحصول على المعلومات التي تهمه ، وآراء الخبراء فى الاختبارات التي سوف يختارها ، والبحوث التي أُجريت حولها ، والبيانات السيكومترية المتوافرة المتعلقة بصدقها

وثبات درجاتها ، ومجموعات الأفراد التي يناسبها اختبار معين ، وهذه المعلومات تتطلب الرجوع إلى مصادر مختلفة متخصصة ، سوف أوضح فيما يلي بعضاً منها :

(١) الكتاب السنوي للقياس العقلي :

Buros` Mental Measurement Year book:

يُعد هذا الكتاب الدوري من أهم مصادر المعلومات فيما يتعلق بالاختبارات والمقاييس ، وهو يصدر كل عام منذ عام ١٩٣٨، وكان يحرره أوسكار بيروس Oscar Baros حتى وفاته منذ عدة أعوام ، ولا يزال يصدر سنوياً عن معهد بيروس للقياس العقلي بأمريكا، ويشتمل على مجموعة متجددة وشاملة من المصنفات ، ومراجعات الاختبارات ، والمراجع المتعلقة ، بالقياس والتقويم النفسي والتربوي، وقد صدر عن هذه المؤسسة حوالي تسع كتب سنوية حتى نهاية القرن الماضي، وتهدف هذه الكتب إلى تنمية الاتجاه الناقد نحو الاختبارات والعمليات الاختبارية من جانب ناشري الاختبارات ومستخدميها ، وتفعيل التواصل ، وتحسين جودة ونوعية الاختبارات المنشورة، وكذلك تقديم معلومات تتعلق بالاختبارات المنشورة التى تتناول بناء الاختبارات المقاييس واستخداماتها وصدقها ، والمراجع الجديدة المتعلقة بالعمليات الاختبارية ، ومقالات ناقدة ممثلة للمُراجعات التى تُنشر فى الدوريات المتخصصة حول هذه المراجع .

(٢) اختبارات فى مرحلة الطباعة Tests in Print :

وهى عبارة عن قائمة شاملة للاختبارات المستخدمة فى المجالات النفسية، والتربوية والصناعية ، ويقوم بنشرها أيضاً معهد بيروس للقياس العقلي، ويشتمل كل اختبار مذكور في هذه القائمة على معلومات تتعلق بعنوان الاختبار ، والفرقة الدراسية المناسبة، وتاريخ النشر، وتعليق موجز حول الاختبار، وعدد ونوع

الدرجات الناتجة، والمؤلف أو المؤلفين، والناشر، وإشارة إلى مراجعة الاختبار فى الكتاب السنوي للقياس العقلي.

(٣) منشورات مركز دراسات التقويم :

Center for the Study of Ecaluation (CSE)

يُصدر هذا المركز التابع لجامعة كاليفورنيا بأمريكا مجلدات تشتمل على دراسات منشورة تتعلق بتقييم اختبارات مختلفة ، وتستند محكات التقييم إلى الصدق والملاءمة ، والواقعية ، وجودة المعايير ، وتقييم عام للاختبار.

(٤) نشرة دورية تتعلق بالاختبارات News on Tests :

تصدر هذه النشرة الدورية عن مركز الخدمات الاختبارية Educational Testing Service (ETS) بولاية نيوجيرسي الأمريكية ، وكان يُطلق عليها فيما سبق " نشرة تجميع الاختبارات Test Collection Bulletin ويحتفظ هذا المركز بمجموعات كبيرة من الاختبارات المنشورة وغير المنشورة التي يمكن الاطلاع على مسمياتها على موقع الإنترنت ، أو على الأفلام المصغرة Microfiche عبر النظام ERIC .

كما تشتمل هذه النشرة على ملخصات موجزة للاضافات المستحدثة لمجموعة الاختبارات ولكن لا توجد مراجعة لهذه الاختبارات كما فى الكتاب السنوي للقياس العقلي .

(٥) قوائم إصدارات الاختبارات Bibliographies of Tests :

توجد العديد من قوائم إصدارات الاختبارات التى تصف الاختبارات المنشورة وغير المنشورة ، وهى قوائم لمجالات معينة تخصصيه مثل : قياسات تتعلق بنمو الطفل ، وتقييم مهارات القراءة ، وغير ذلك ، وكثير من هذه الإصدارات مذكورة فى مجلدي (Baches, 1977 , Aaronson, 1974,) .

(٦) دوريات القياس والتقويم Professional Journals :

توجد دوريات متعددة متخصصة في القياس والتقويم النفسي والتربوي ، وتُجري مراجعات للاختبارات التى يستخدمها المربون ، والمشتغلون بعلم النفس ، ويمكن الرجوع إلى الكتاب السنوي للقياس العقلي المذكور سابقاً للحصول على معلومات تتعلق بهذه الدوريات .

❖ مؤسسات نشر الاختبارات ،

يمكن الحصول على معلومات تتعلق بالاختبارات والمقاييس النفسية والتربوية من ناشري الاختبارات فى الولايات المتحدة الأمريكية بخاصة، حيث يمكن من خلال مواقعها على الإنترنت معرفة أنواع الاختبارات والمقاييس ووصفها، وكذلك من فهارس إصدارات هذه المؤسسات وتشتمل هذه الفهارس على الإصدارات الحديثة للاختبارات المختلفة ، والخدمات المتعلقة بتصحيحها ، وكلفتها ، وعينات من مفرداتها ، وأداتها وتقارير فنية متعلقة بها .

وكذلك يمكن معرفة أسماء وأنواع الاختبارات والمقاييس التي يتم تعريبها وتقنينها في الوطن العربي من دور النشر الكبري ، والمكتبات المتخصصة ، ومواقعها على الإنترنت .

❖ مراجع الكتب

- Adkins, D. (1974). Test Construction : Development and Intepretation of Achievement Tests (2nd ed,) Columbus: Charles Merill Publishing Company.
- Aiken, L. (1973). Readings in Psychological and Educational Testing, Boston: Allyn and Bacon, Inc.
- Alien, L. (1976). Psychological Testing and Assessment, (2nd ed.) Boston : Allyn and Bacon. Inc.
- Aiken, L. (1987). Assessment of Intellectual Functioning, Boston : Allyn and Bacom, Inc.
- Airasian,P. (1999) Classroom Assessment, N. Y. : McGraw, Hill.
- Allam, S. (1981) Minimum Measurement Competency Test: New Jersey : Educational Testing Service.
- Allam, S. (1990) Systems and Methods of Evaluation and Examination in the Arab States: Present Status and Future Trends, Jordan : UNESCO. IPDAS.
- Allpor. G. (1963). Patterns and Growth in Personality, N.J.: Holt Rinehart & Winston.
- American Psychological Association (1999), Standards for Educational and Psychological Testing, Washington, D.C.: Author.
- Anastasi, A. (1988) Psychological Testing (6th. Ed.) . N.Y: Macmillan,
- Baker, T. (1977) A Directory of Information of Tests, Princeton, N.J.: ERIC Clearinghouse on Tests.
- Bejar. I. (1974). Achievement Testing: Recent Advances, Calif: SAGE publications Inc,
- Belanof, P.& Dckson, M. (1991). Portfolios : Process and Product. Portsmouth, N.H: Boynton Cook Publishers.
- Berk, R. (1986) Performance Assessment Methods and Applications, Baltimore MD: John Hopkins University Press.

- Bertrand, A. & Cebula, J. (1980). Measurement, and Evaluation: A Devlopmental Approach, Mass: Addision Wesly Publishing Company.
- Birenbaum, M.& Dochy, F. (1996). Alternative Assessment of Achievement, Learning Processes and Prior Knowledge, Boston: Kluwer Academic Publishers.
- Brown, F. (1978). Principles of Educational and Psychological Testing (2nd ed.) ILL: Charles Thomas.
- Buros, O. (1977). Fifly years in Testing. Educational Researcher 6 (7) 9-15.
- Buros, O. (1978). The Eighth Mental Measurement Yearbook, Highland Park, N.J.: The Gyphon Press.
- Chase,C. (1978). Measurenent for Educational Evaluation (2nd ed.). Mass : Addison Wesley.
- Crocker, L. & Alginas, J. (1986). Introduction to Clasical and Modern Test Theory, New. York; Holt, Rinehart & Winston,
- Cronbach, L. (1990). Educational Psychology (5th ed.) New York: Harpex and Row,
- Daron, I. & Wolf, D. (Eds.) (1996). Performance- Based Student Assessment, Chicago: National Society for the Study of Education.
- Desforges C. (1989). Testing and Assessment. London,: Cassell.
- Donald L. & Emest, L. (1975). Measurement and Evaluation in the Schools. N.J.: Houghron Mifflin Company.
- Dubois, P. (1970). A. Hisstory of Psychological Testing. Boston: Allyn and Bacon.
- Dunnette, M. (1966). Personal Selection and Placement. Calif. : Brooks Colepublishing Company,

- Ebel, R. (1965). Measuring Educational Achievement, Englewood Cliffs, N.J.: Prentice Hall.
- Ebel. R. (1980). Practical Problems in Educational Measurement, Lexington, M.A.: D.C. Hearth.
- Edel. R. & Frisbie D. (1991). Essentials of Educational Measurement (5^{th} ed.) Englewood Cliffs. N.J.: Prentice Hall.
- Edwards, A. (1959). Edwards Personal Preference Rocord, San Antonio, Psychological Company.
- Elliot, C. (1990). Differential Ability Scales : Introductory and Technical Handbook. San Antonio, TX: Psychological Corporation,
- Embertson, S. (1985) Test D esign: Development in Psychology and Psychometrics. M.Y. : Academic Press.
- Embertson, S. & Reise (2000). Item Response Theory for Proychologists. N.J.: Lawrence Eribaum.
- Eysenck, H. (1973). The Measurement of Intelligence, Lancaster Medical and Technical Publishing Company.
- Eysenck, H. (1979). The Structure and Measurement of Intelligence, N.Y.: Springes – Verlagy.
- Farm R. & Tone B. (1994), Portfolios Performance Assessment, N.Y.: Harcourt.
- Feldi, L. & Forsyth, R. (1993) Lowa Tests of Educational Development . lowa City : University of lowa.
- Fredriksen N: Miselevy R. & Bejar I. (1993), Test Theory for a New Generation of Tests, S. N.J., Erlbaum.
- Gehoski, N. (1964). Psychological Testing: Theory. Interperation and Practices. III. : Charles Thomas Publishing.
- Ghiselli, E. (1960). Theory of Psychological Measurement, N.Y. : McGraw – Hill.

- Gipps. C. (1994). Beyond Testing : Toward a Theory of Educational Assessment, London : Falmes,
- Goldstein, H. & Lewis, T. (eds) (1996). Assessment : Problems, Development and Statistical Issues. Chichester, John Wiley.
- Gougy. H. (1953). What Determines the Achievement of High School Students? Journal of Educational Rescarch, 46, 321 -331,
- Green, D. (1974). The Aptitude Achievement Distinction., CA.: CTB McGraw Hill.
- Gronluned, N. (1968). Readings in Measurement and Evaluation, N.Y.: The Macmillan Company.
- Gronluned, N. (1985). Measurement and Evaluation in Teaching. New York: the Macmillan Company.
- Gruijter, D. & Kamp, L. (1976). Advances in Psychological and Educational Measurement. N.y.: Wiley & Sons.
- Gyiford, J. (1954). Psychometric Methods. N.Y.: McGraw Hill.
- Hambleton, R. & Eigner, D.(1978). Guidelines for Evaluating Criterion Referenced Tests and Test Manuals, Journal of Educational Measurement, 15, 321 – 327.
- Hambleton, R. & Swaminathan, H.(1985) Item Response Theory : Principles and Applications, Nowell, MA: Kluwer Academic Press.
- Hambleton, R. & Zaal, J. (1991). Advances in Educational and Psychological Testing : Theory and Applications. MA: Kluwer Academic Publishers.
- Haney, W. (1984). Testing Reasoning and Reasoning About Testing . Reweiw of Educational Reseach 54, 4. 597 – 654.
- Harris. D. & Bell, C. (1990). Evaluating and Assessing for Learning. London: Kogem Page.

- Holmen, M. & Doetor, R. (1972). Educational and Psychological Testing : A Study of the Industry and its Practices. N.Y.: Russell Sage.
- Hopkins. K., Stanley, J & Horbins, B. (1990). Educational and Psychological Measurement and Evaluation (7^{th} ed.)., Needram Heights, MA: Allyn Bacon.
- Horrocks, H. & Schoonover, T. (1968). Measurement for Teachers, Columbus: Charles Merill Publishing Company.
- Houts, P. (1977). The Myth of Measurability. N.Y.: Hort Publishing Company,
- Hunt, J. (1960). Intelligence and Experience N.Y. : The Ronald Press Company
- Ingenkamp, K.(1968). Developments in Educational Testing (Volume 1)N.Y.: Gordon and Breach.
- Innes, T. (1972). The Prediction of Achievement : Measurement of schools from Non – School Factors. Invied Address at Southern Invitational Conference on Testing. Athens. Ga.
- Jensen. A. (1980). Bias in Mental Testing. N.y.: Free Press.
- Jensen. M. & Feuerstein, R. (1986) Cultural Differences and Cultural Deprivation A Theoretical Framework for Differential Interventions. In R.M. Goupta & P. Cathead (eds.) Cultural Diversity And Learning Efficiency. London : The Macmillan Press.
- Joint Committee on Testing Practices (1988). Code of Fair Testing Practice in Education. Washington, D.c. American Psychological Association.
- Keats, J. (1971). An Introduction to Quantitative Psychology. N.Y.: John Wiley.
- Kelly. E. (1967). Assessment of Human Characteristics. Calif : Brooks Cole Publishing Company.

- Kenneth, S.; Melissa, S. Anderson, R. (2002). The Educational Guide to Ethical Standrads. Washington, D.C.: AERA Publications.
- Kohn, A. (2000). A case Against Standardized Testing. Portsmouth, NH.: Heineman.
- Koretz, D. (1997). What Happens to Test Scores and Why, Educational Leadership. 52 (2) 38 – 41.
- Lien, A.(1971). Measurement and Evaluation of Learning. (2nd ed.) lowa: brown Company Publishers.
- Lindzey. G.(1959). On the Classification of Projective Techniques. Psycholgical Bulletin, 59.2, 158 - 168.
- Linn. R. (1989). Educational Measurement, (3rd ed.) N.Y.: Macmillan.
- Linn, R. & Gronlund, N. (1995). Measurement and Evaluation in Teaching (7th ed.) N.Y.: Merill.
- Lohman, D. (1993). Teaching and Testing to Develop Fluid Abilities. Educational Researcher, 22 (7) 12- 23.
- Lubin, B. : Larsen, R.& Matarazzo. (1984). Patterns of Psychological Test Usage in the United STATES : 1935- 1982 American Psychologist, 39, 451- 454.
- Maeroff. G. (1991). Assessing Alternative Assessment Phi Delta Kappan, 72, 272- 281.
- Marshall, J. & Hales, L. (1971). Classroom Test Construction. Calif.: Addison – Wesley Publishing Company.
- Mathematical Sciences Educational Board (MSEB) (1991) A Summary of the Regional Presummit Forum. Washington, D.C.: National Research Council.
- McCallum, R. (1990) . Determinig the Factor Structure of The Stanford- Binet Fourth Edition. Journal of Psycholeducational Assessment. 8, 436 – 442.

- Mchaughin, M. (1991) Test – Based Accountability as Reform Strategy Kappan 73 (3) 248 – 251.
- Mcmillan, J. (1997). Classroom Assessment : Principles and Practice for Effective Instruction, Boston : Allyn and Bacon.
- Mehrens, W. & Ebel, R. (1967) Prinicples of Educational and Psychological Measurement III : Rand Mcnally & Company.
- Mehrens, /w. & Lehmann, I. (1969). STANDARDIZED Tests in Education N.y.: Holt, Rinehart and Winston, Inc.
- Miller, K. (1975). Psychological Testing in Personnel Assessment. : N.Y. John Wiley & Sons.
- Millman, I. : Bishop, C. & Ebel, R. (1965). An Analysis of Test Wiseness. Educational and Psychological Measurement 25, 707 -726.
- Nitko, J. (1983) : Educational Tests and Measurement An Introduction. N.Y.: Harcurt Brace Jovanovich, Inc,
- Nitko., J. (1996) Educational Assessment of Students. Englewood Cliffs: Prentice Hall.
- Noll. V.& Scannell, D. (1972). Introduction to Educational Measurement (3rd ed.) N.K.: Houghton Mifflin Company.
- Nunnally, J. (1972) Educational Measurement and Evaluation. N.Y.: MGraw – Hill.
- Oosterhof. A. (1990). Classroom Applications of Educational Measurement. Columbus : Merill.
- Payne, D. (1974). The Assessment of Learning : Cognitive and Affective. Mass.: D.C. Heath and Company.
- Perin, l. (1975) . Personality Theory Assessment and Research. N.Y.: John Wiley.

- Phye. G. (1996). Handbook of Classroom Assessment : Learning, Adjustment, and Achievement N.Y.: Academic press.
- Popham, W. (1978). Crieterion – Referenced Meferenced Measurement, N.J. Prentice Hall, Inc.
- Popham, w. (1981). Modern Educational Measurement Educational (Engelwood Cliffs). N.T.
- Popham, W. (1995). Classroom Assessment : What Teachers Need to Know. Boston : Allyn And Bacon,
- Psychological Corporation (1993). Metropolitan Achievement Tests. San Antonio Tx:, Author.
- Rabin, A. (1968). Projective Techniques in Personality Assessment : Modern Introduction. N.Y.: Springes Publishing Company, Inc.
- Reschly, D. & Wilson, . (1990) Cognitive Processing Versus Traditional Intelligance. School Psychology Review. 191. 443 – 458.
- Resnick, L. (1976). The Nature of Intelligence. N. J. Lawrence Eribaum.
- Roaed. G. (1989). Item Writing and Item Banking by Microcomputers An : Update. Educational Measurement : Issues and Practice. 8. (3) 38, 17 – 20.