الفصل السابع

الاختبارات المحكية المرجع والاختبارات المعيارية المرجع والاختبارات التشخيصية

- مقدمة
- الاختبارات محكية المرجع والاختبارت المعيارية المرجع
- أوجه الاختلاف بين الاختبارات معيارية المرجع ومحكية المرجع
- تصنيفات الاختبارات محكية المرجع Categories of CRT
- خطوات بناء الاختبارات محكية المرجع (الجانب المعرفي)
- بناء الاختبارات محكية المرجع لقياس الجانب الوجداني
- طرق تقدير صدق الاختبارات محكية المرجع
- طرق تقدير ثبات الاختبارات محكية المرجع
- مسميات وتعريف درجة القطع Definitions of cut off score
- طرق تحديد درجة القطع
- أسس تقويم طرق تحديد درجة القطع
- الاختبارات التشخيصية
- خطوات بناء الاختبارات التشخيصية
- أنماط التعليم ومستوياته

obeikan.com

obeikan.com

مقدمة

وجه العديد من علماء القياس التربوي انتقاداتهم إلى الاختبارات المعيارية المرجع لكون إن معظم هذه الاختبارات نشأت مرتبطة بالفلسفة التربوية التي كانت سائدة في الولايات المتحدة الأمريكية في أوائل القرن العشرين، وهي تصنيف الأفراد بحسب مركزهم النسبي بين أقرانهم في القدرات المختلفة، وبخاصة عندما دعت الحاجة إلى تصنيف الضباط والجنود قبل اشتعال الحرب العالمية الأولى، حيث ظهر في ذلك الوقت اختبار آرثر أوتيس A.Otis وهو أول اختبار جمعي، وقد أدى تحقيق هذا الاختبار للغرض الذي وضع من اجله إلى تشجيع حركة الاختبارات جماعية المرجع لتصنيف الأفراد تبعاً لقدراتهم وتحصيلهم وغير ذلك من الصفات. كما استخدمت في مجالات التوجيه المهني، والتشخيص الإكلينيكي، وانتقاء الأفراد للوظائف المختلفة وتحديد الفرق في تحقيق كثير من الأغراض، إلا انه تبين قصورها في تقويم فاعلية المنظومة التعليمية، وفي قياس المهارات المختلفة المطلوب تحققها لدى الدراسة المناسبة للطلاب، وقياس فاعلية البرامج التربوية. وعلى الرغم من نجاح هذه الاختبارات، كما بينت الدراسات المستفيضة حول الاختبارات المقننة المستخدمة في قياس الذكاء أن كثيرا منها يشتمل على مفردات تتحيز ضد مجموعة أقلية تعيش في نفس المجتمع الذي استمدت منه معايير هذه الاختبارات. كما أنها تؤدي في كثير من الأحيان إلى التفرقة العنصرية بين الأجناس. وقد ترتب على ذلك العديد من المشكلات الاجتماعية والتربوية مثل التشخيص غير الدقيق للأطفال المتخلفين عقلياً والاختيار غير الصائب لطلاب الدراسات العليا أو المتقدمين للحصول على المنح الدراسية.

كما أن كثيراً من المقاييس المقننة للشخصية المستخدمة في انتقاء أفضل المتقدمين لشغل وظائف معينة لا تتميز بالثبات والصدق اللازم توفرهما بقدر كبير من الثقة. وكثيراً من المفردات والعبارات التي تشتمل عليها وبخاصة تلك التي تتعلق

بالأمور الجنسية والسياسية والدينية تمس الحرية الفردية للأشخاص الذين تطبق عليهم هذه المقاييس.

وأن الاختبارات التحصيلية المقننة لا تقيس الأهداف التعليمية التي يسعى المعلم إلى تحقيقها في الفصل المدرسي. كما أنها لا تقدم الفائدة المرجوة في تشخيص مواطن الضعف والقوة في تحصيل الطلاب حتى يستطيع المعلم أن ييسر وسائل المعالجة ورفع كفاءة العملية التعليمية.

وقد استثارت أوجه النقد هذه المربين وعلماء القياس إلى أهمية استخدام مدخل آخر لقياس نواتج التعلم أطلق عليها ً القياس محكي المرجع Referenced Criterion Measurement. وفرق جليزر بين نوعين من القياس أحدهما القياس المعياري المرجع Norm Referenced Measurement، ويتركز الاهتمام فيه على ترتيب الفرد بالنسبة لاقرانه في القدرة التي يقيسها الاختبار، والآخر هو القياس المحكي المرجع الذي يهدف إلى موازنة أداء الفرد بمستوى أداء مطلق أو مستوى كفاية معين دون الحاجة إلى موازنة أدائه بأداء الأفراد الآخرين.

وقد أثار جليزر كثيرا من المناقشات والجدل بين علماء القياس بعامة والمتخصصين في تطبيقات تكنولوجية التعليم بخاصة، ولكن لم يحدث نشاط علمي ملحوظ نحو تحقيق ما نادى به جليزر الا حوالي عام ١٩٦٩.

ففي عام ١٩٦٩ بدأ جيمس بابام James Popham عالم النفس الأمريكي بجامعة كاليفورنيا يتزعم حركة القياس محكي المرجع. ودعا إلى عقد مؤتمر متخصص في أمريكا عام ١٩٧٠ لمناقشة القضايا والمشكلات السيكومترية المتعلقة بهذا المفهوم من القياس. ونشر له أول كتيب عن مقدمة في القياس المحكي المرجع عام ١٩٧١.وترتب على ذلك حركة بحثية كبيرة منذ ذلك الوقت ولحد الآن. فقد صدر كتاب لنورمان جرونلند N. Grounlun عن إعداد الاختبارات المحكية المرجع والكتاب موجه للمعلم المهتم ببناء هذا النوع من الاختبارات في المستويات المعرفية المختلفة.كما ويعتبر كتاب جيمس بابام Popham الذي صدر عام ١٩٧٨ أول كتاب متكامل عن القياس محكي المرجع وتلى ذلك صدور عدة كتب من أهمها كتاب بيرك Berk عام ١٩٨٠ وآخر في عام ١٩٨٣.

الاختبارات محكية المرجع والاختبارات المعيارية المرجع

تعتمد عمليتي القياس والتقويم التربوي على أدوات ومقاييس تقدم بيانات كمية تتيح للتربويين فهم الظاهرة التربوية بشكل دقيق ومحدد، وتعتبر الاختبارات والمقاييس التربوية جزء اً هاماً من أجزاء العملية التربوية فمن خلالها يستطيع المعلم الحكم على مدى تحقق أهداف البرامج التعليمية والتدريبية التي يقوم بتدريسها، كما أنها تفيد الطالب أيضاً في التعرف على مستوى تحصيله الدراسي من خلال أدائه في الاختبار. كما تتعدد استخدامات الاختبارات في المؤسسات التربوية فمنها ما يستخدم للقياس والتقويم أو التشخيص ...الخ.

وبنظرة سريعة على حركة بناء المقاييس التربوية المعاصرة يلاحظ أن هناك ثلاثة اتجاهات رئيسة يختلف كل اتجاه منها عن الآخر في الافتراضات التي يقوم عليها، وفي الإطار المرجعي الذي يعتمد عليه تفسير نتائج الطلاب في الاختبار، وفي إعداد أدوات القياس وهي القياس معياري المرجع Norm-Refernced-Measurement، والقياس محكي المرجع Referenced-Measurement Criterion، ونظرية السمات الكامنة.Latent Traits Theory وأشار كل من جيلزر وبابام وهيوزيك (Popham & Husek, ١٩٦٩; Glaser, ١٩٩٤) بأن الاتجاه الأول يعتمد على تفسير درجات الطالب في ضوء معايير محددة (Norms).أي ترد درجات الطالب في الاختبار إلى جداول معيارية ومن ثم يحدد مستوى أداء الطالب النسبي بناءاًعلى موقع درجة الطالب مقارنةً بمتوسط تحصيل المجموعة المعيارية في الاختبار، ويركز هذا الاتجاه على الفروق الفردية بين الطلاب بحيث يوضح مدى الفارق بين مستوى طالب وآخر، أي أنه يقيس مقدرة طالب معين مقارنةً بمقدرة طلبة آخرين. وتعرض الاتجاه معياري المرجع لبعض أوجه النقد أشير إليها في أدبيات القياس التربوي ومنها ماذكرها بابام وهيوزيك ١٩٦٩ Popham & Husek حيث تبين أن الاختبارات معيارية المرجع تعتمد على مقارنة أداء الطالب بأداء مجموعته الصفيّة، وبالتالي فإنه من المحتمل أن يختلف موقع الطالب باختلاف خصائص مجموعته المعيارية، بالإضافة إلى أنه ليس بالضرورة أن يمثل متوسط المجموعة

المعيارية الأداء المطلوب للنجاح، وتهتم هذه الاختبارات أيضاً بإبراز الفروق الفردية بين الطلاب دون النظر إلى مدى إتقان الطالب للمهارات والمعلومات المراد قياسها وذلك لتأهيله لبرامج تدريبية وتعليمية جديدة، وبالتالي فهي لا تساعد على تقويم وتشخيص كل من كفاية الطالب وبعض البرامج التدريبية (برامج التعليم الفردي - البرامج القائمة على الكفايات بالقدر المطلوب). أي أنها لا تزود المعلم بمعلومات دقيقة تساعد على اتخاذ القرارات التربوية المناسبة حول مستوى كفاية الطالب والمنهج الدراسي. كما أكد بابام (Popham ١٩٧٨) بأن هذه الانتقادات لعبت دوراً في ظهور تطورات جديدة في منهجية الاختبارات والمقاييس التربوية أدت إلى ظهور اتجاه آخر أطلق عليه القياس محكي المرجع Criterion-Referenced-Measurement، حيث تسجل أول مقالة لجليزر في عام ١٩٦٣م تناقش هذا الاتجاه.

كما أشار كل من هامبلتون وآخرون (Hambleton, et al ١٩٧٦) إلى أن هذا الاتجاه لا يعتمد على تحديد مستوى الطالب النسبي داخل مجموعته الصفية ولكنه يعتمد على تحديد ما يستطيع وما لا يستطيع الطالب أن ينجزه. أي ينسب أداء الطالب إلى محتوى الاختبار ذاته، ويعتمد هذا المقياس في بنائه على تحديد نطاق الأهداف السلوكية تحديداً دقيقاً، ومن ثم قياس مدى إتقان الطالب لهذه الأهداف من خلال مقارنة أداء الطالب بمستوى معين من الأداء (درجة القطع).

وأضاف ايبل وهامبلتون وايجنور ((Hambleton & Eignor, ١٩٧٩; Ebel, ١٩٧٩) إلى أن القياس محكي المرجع يهدف إلى تصنيف الطلاب إلى مجموعتين الأولى متقنة masrey والثانية غير متقنة Non-Mastery للمهارات والمعارف وذلك اعتماداً على (مستوى أداء – معيار –محك –درجة القطع –درجة النجاح) محددة تحديداً مسبقاً. ويتمثل هذا المحك (Criterion) في نسبة عدد الأسئلة التي يجب أن يقوم الطالب المتقن بالإجابة عليها بصورة صحيحة، وبالتالي إذا كانت درجة الطالب في الاختبار تزيد أو تساوي درجة القطع (Cut off Score) اعتبر الطالب متقناً للمهارات وإذا لم تكن كذلك اعتبر الطالب غير متقن. يتضح من ذلك أن عملية تصنيف الطلاب (Classification) إلى فئتين تعتمد على قيم درجات قطع الاختبارات، لذا يجب تحديد هذه الدرجة بعناية فائقة جداً لأنه يترتب على

تحديدها اتخاذ قرارات هامة منها: تقويم مستوى تحصيل الطالب في الاختبار، وتقويم البرامج والمناهج التعليمية والتدريبية ومن ثم تشخيصها بصورة مناسبة. ويعتبر تحديد درجة القطع (النجاح) أمراً بالغ في الصعوبة و الأهمية لذلك كثرت المداخلات بين علماء القياس التربوي حول كيفية تحديد درجة القطع المثلى التي يمكن على ضوئها اتخاذ القرارات المناسبة، نتج عن تلك المداخلات طرائق متعددة ومختلفة لتحديد درجة القطع في الاختبارات محكية المرجع، والتي ذكر بيرك (Berk, ١٩٨٦) أنها بلغت ثمان وثلاثين طريقة وتقسم طرق تحديد درجة القطع إلى ثلاثة أقسام:

١ـ يحتوي القسم الأول على الطرق التحكيمية Judgmental - Methods حيث تعتمد هذه الطرق على آراء) المحكمين) المختصين في المجال أو المحتوى المراد قياسه لتحديد درجة قطع الاختبار.

٢ ـ يحتوي القسم الثاني على طرق تعتمد جزئياً على التحكيم وتسترشد ببيانات تجريبية Judgmental ـ Empirical - Methods وهي تعتمد على آراء المحكمين مع الاعتماد على بيانات ترتبط بالأداء الفعلي للطلاب.

٣ ـ ويحتوي القسم الثالث على طرق تعتمد بشكل رئيس على البيانات التجريبية وجزئياً على التحكيم حيث تعتمد هذه الطرق على أداء الطلاب الفعلي في الاختبارات والتحليلات الإحصائية، وينحصر دور عملية التحكيم في انتقاء الطلاب المتقنين وغير المتقنين من الطلاب.و أشار أبو علام (١٩٩٥) إلى أن كثرة الانتقادات الموجهة إلى القياس معيار ي المرجع كانت سبباً في ظهور نظرية السمات الكامنة (LTT)، ومن تلك الانتقادات مايلي:

اعتماد بناء جميع خصائص الاختبارات السيكومترية على خصائص عينة الدراسة، وعلى صعوبة عينة مفردات الاختبار، يفترض خطية ميزان درجات الطالب في الاختبار غير أنه عادة ما يكون الميزان منحنياً لأن الفرق الثابت بين درجتين من درجات الاختبار يختلف تفسيره بناء على موقع الدرجات على متصل السمة أو القدرة المراد قياسها، كما يفترض أن العلاقة بين زيادة درجة الطالب في الاختبار وبين مقدار السمة أو القدرة علاقة خطية مطردة. وقد لا يكون هذا الافتراض صائباً، إذا لم يحصل بعض الطلاب من ذوي القدرات المتميزة على درجات مرتفعة

في الاختبار وفي المقابل يمكن أن يحقق بعض الطلاب من ذوي القدرات المتدنية على درجات مرتفعة في الاختبار، كما تتأثر مفردات الاختبار بالظروف المحيطة به حيث يمكن أن تتغير درجات الطلاب بحذف أو تغيير أي مفردة من مفردات الاختبار الأمر الذي يؤثر في تمثيل المفردات للنطاق السلوكي المراد قياسه. كما إن هذه النظرية وفرت نماذج تقدر السمات الكامنة (القدرة) في الفرد بصورة مستقلة عن أفراد العينة الدراسية.

وتوجد ثلاثة أبعاد توفر الأسس الموضوعية في نماذج السمات وهي: بعد السمة المقاسة الذي يعرف بواسطة مجموعة من البنود وهي إما أن تكون أحادية (صعوبة المفرد ة) أو ثنائية (صعوبة المفردة ومعامل التمييز) أو ثلاثية المعلم (صعوبة المفردة ومعامل التمييز وعامل تخمين الإجابة)، وبعد استقلالية القياس حيث لا يعتمد تقدير صعوبة البند على صعوبة البنود الأخرى، كما لا يعتمد أيضاً تقدير قدرة الأفراد على قدرة أي مجموعة أخرى. وأما توازي المنحنيات المميزة للبنود الذي يوفر الحد الذي تميز فيه المفردات بين الأفراد ذوي المستويات المختلفة في قدرة ما.. و أشار علام (١٩٩٥) إلى نماذج النظرية الثلاثية وهي نموذج (راش) ويعد نموذجاً أحادي المعلم، ونموذج (لورد) ويعد ثنائي المعلم ونموذج (بيرنبوم) ويعد ثلاثي المعلم.

ومن المعروف أن الاختبارات التحصيلية تهدف إلى:

- التأكد من توافر الحد الأدنى اللازم من مستويات الأداء بالنسبة لعمل أو مهنة معينة.
- التصنيف المهني والتعليمي والعسكري.
- مقارنة أداء الطالب بأداء أقرانه.
- تحفيز دافعية التلاميذ على التحصيل.

ولكي تحقق الاختبارات التحصيلية الأهداف المنشودة، يجب أن يكون الاختبار معداً بطريقة جيدة بحيث يغطي عينة ممثلة من النواتج التعليمية والمادة الدراسية، وأن يصاغ بطريقة تتناسب مع الغايات التي ستستخدم النتائج من أجلها وأن يتمتع بقدر كبير من الصدق والثبات، وأن يتوخى الحذر في تفسير نتائجه وذلك

بسبب أخطاء القياس، وأن يستخدم لتحسين مستوى تعلم الطلاب وأن يحتوى على أسئلة ذات نوعية ملائمة لقياس الناتج التعليمي المطلوب. وتوجد العديد من تصنيفات الاختبارات التحصيلية وذلك حسب نوع الاستجابة، وطريقة التطبيق، وصياغة الأسئلة، وتفسير الدرجات الناتجة، ولعل أهم تصنيف للاختبارات التحصيلية من حيث طريقة تفسير الدرجات الناتجة من تطبيق الاختبار كالتالي:

الاختبارات معيارية المرجع NRT:

ذكر بابام وهيوسك (Pohpam & Husek ١٩٦٩) بأن هذه الاختبارات تعتمد عند تفسيرها للدرجات المستحصلة على خصائص المجموعة التي طبقت عليها حيث يجب أن تكون هذه الخصائص ممثلة مع خصائص الطلاب من حيث السن والنوع والمنطقة الجغرافية والمرحلة الدراسية، وتهتم هذه الاختبارات بالكشف عن الفروق الفردية بين الطلاب وذلك من خلال مقارنة أداء الطالب بأداء أقرانه كجماعة معيارية. بمعنى أن هذه الاختبارات تركز على وضع الطالب النسبي في علاقته بالآخرين.

الاختبارات محكية المرجع CRT

أشار بابام وهيوسك (Pohpam & Husek ١٩٦٩) بأن هذه الاختبارات لا تعتمد عند تفسيرها لدرجة الطالب على خصائص مجموعته التي طبق عليها الاختبار، ولكنها تعتمد على مقارنة أداء الطالب بمحك أو مستويات أداء متوقعة ومحددة مسبقاً. ويتم تحديد هذه المستويات في ضوء الأهداف السلوكية (الإجرائية) المراد قياسها وتقيس هذه الاختبارات مدى تحقيق الطالب لهذه الأهداف السلوكية وبذلك يمكن التحقق من كفاءة كل طالب من اكتسابه للمهارات والمعارف المرجوة ومن ثم تشخيص نواحي الضعف والقوة.أوجه الاختلاف بين الاختبارات معيارية المرجع ومحكية المرجع، أشار كثير من علماء القياس التربوي مثل جليزر، وإيبل، وبابام وهيوسك، وهامبلتون،(Hambleton et al, Papham &Husek, ١٩٦٩, Ebel, ١٩٧٩, Glaser, ١٩٩٤) إلى وجود عدد من الاختلافات بين الاختبارات معيارية المرجع ومحكية المرجع وذلك في عدة نواحي منها مفهومها وبناؤها وأهدافها وخصائصها والتي يمكن أن تقدم على النحو التالي:

١ـ الهدف الرئيس للاختبار: يهدف الاختبار محكي المرجع إلى مقارنة أداء الطالب بمحك محدد مسبقاً وذلك لمعرفة الأهداف التي حققها والتي لم يحققها الطالب، أما الاختبار معياري المرجع فيهدف إلى مقارنة أداء الطالب بأداء مجموعته المعيارية من خلال تحديد عدد الأسئلة التي أجاب عليها الطالب بشكل صحيح.

٢ ـ استخدام الاختبار: يستخدم الاختبار محكي المرجع للتأكد من تحقيق الطالب للأ هداف السلوكية المحددة، أما الاختبار معياري المرجع فيعتمد على تزويدنا بمؤشرات عن درجة نجاح الطالب، وبالتالي إظهار الفروق الفردية بين الطلاب.

٣ ـ خصائص أسئلة الاختبار: تتجمع الأسئلة في الاختبار محكي المرجع حول عدد محدد من الأهداف، بينما تنتشر الأسئلة في الاختبار معياري المرجع بشكل واسع حول نطاق الأهداف.

٤ ـ تفسير الأداء: تفسر الدرجة في الاختبار محكي المرجع بناء على درجة القطع، بينما تفسر درجة الطالب في الاختبار معياري المرجع بناء على درجات معيارية تمثل موقع الطالب بين مجموعته الصفية والرتب المئينية.

٥ ـ بناء الفقرات:يعتمد بناء الفقرات في الاختبار محكي المرجع على مدى تحقيق المفردة للهدف السلوكي التعليمي، بينما يعتمد بناؤها في الاختبار معياري المرجع على تباين الدرجات ومعامل التمييز، ويتجنب اختيار المفردات الصعبة جداً والسهلة جدا.

٦ـ التقويم: يسعى الاختبار محكي المرجع إلى تحديد جوانب الضعف والقوة في كل من أداء الطلاب والبرامج التعليمية، بينما يسعى الاختبار معياري المرجع إلى تقويم أداء الطالب ومقارنته مع أداء الطلاب الآخرين.

تعريف الاختبارات محكية المرجعDefinition of a CRT

أشار عدد من علماء القياس التربوي مثل بيرك، وجايغر، وهامبلتون ونوفيك Hambleton & Novick, ١٩٧٣ Berk, ١٩٨٢ Jaeger, ١٩٨٩, إلى أن تعريف الاختبارات محكية المرجع يتعدد وفقاً لاختلاف آراء وخبرات علماء القياس التربوي. حيث توجد في أدبيات القياس التربوي أكثر من خمسين تعريفاً يصف الاختبارات محكية المرجع، وتتمحور تلك التعريفات حول النقاط التالية:

١ـ تعريف النطاق السلوكيcontent-domain الذي يقيسه الاختبار تعريفاً دقيقاً ذلك مما يساعد على اختيار مفردات تمثل هذا النطاق.

٢ـ استخدام نتائج الاختبارات محكية المرجع لاتخاذ القرارات الخاصة بتحديد مستوى كفاءة الطالب مما يساعد على تشخيص نواحي ضعفه وقوته.

٣ـ مقارنة أداء الطالب بمستويات أداء محددة مسبقاً (أهداف سلوكية محددة) يتخذ في ضوئها قرارات تربوية تتعلق بمدى إتقان الطالب أو عدم إتقانه للمهارات المراد قياسها.

طرق تحديد النطاقات السلوكية

تعتمد الطرق التقليدية في بناء الاختبارات المعيارية المرجع على وضع خطة تفصيلية تتضمن تحديدا للمحتوى المراد قياسه على البعد الأفقي لجدول مواصفات الاختبار، والأهداف المرتبطة بهذا المحتوى (غالبا ما تصاغ في عبارات عامة مثل معرفة المصطلحات، أو فهم المقروء، أو تطبيق المبادئ ...الخ على البعد الرأسي).

وتقدر أوزان تتناسب مع أهمية كل من المحتوى والهدف في كل خلية من خلايا الجدول. وتبنى مفردات الاختبار طبقا لهذه الأوزان. أما الاختبار محكي المرجع فيتطلب تحديدا أكثر دقة وتفصيلا للإطار السلوكي الذي يقيسه الاختبار قبل البدء الفعلي في كتابة المفردات. فالهدف من بناء هذا النوع من الاختبارات هو اتخاذ قرارات تعليمية تختلف عن القرارات التي تتخذ في حالة الاختبارات المعيارية المرجع.

فنحن هنا نهتم باتخاذ قرار عن مستوى أداء كل فرد بالنسبة إلى مجموعة من المهام أو المهارات أو المعارف التي تشكل النطاق السلوكي الذي يقيسه الاختبار. والحقيقة أن هذا النطاق يكون بمنزلة " المحك Criterion " الذي ينسب إليه أداء الفرد. فلكي نقول إن الفرد أجاب إجابة صحيحة عن ٩٥٪ من المفردات التي يشتمل عليها الاختبار، يجب أن يكون المجتمع الشامل للمفردات محدداً تحديدا واضحا. وتصبح المشكلة الأساسية في بناء الاختبارات محكية المرجع هي كيفية تحديد النطاقات السلوكية المراد قياسها والتي تنسب إليها درجة الفرد في الاختبار.وفيما يلي عرضا لبعض الطرق في تحديد وصياغة النطاقات السلوكية المراد قياسها:

أولا: طريقة بابام في إعداد مواصفات الاختبار

يرى بابام Popham أن تحديد الأهداف التي يقيسها الاختبار وصياغتها بطريقة إجرائية (سلوكية) لا يعد كافيا لبناء الاختبار محكي المرجع. إذ يجب أن نعد مواصفات أكثر تفصيلا للنطاق السلوكي الذي يتضمن هذه الأهداف.

ولكن يمكن أن نسأل ما هي درجة الوضوح والدقة والنوعية اللازم توفرها في النطاق السلوكي حتى نعتبره معرفاً تعريفاً جيداً ؟.

وفي الحقيقة لا توجد إجابات شافية لهذا السؤال. فالنطاقات السلوكية تختلف من موقف إلى آخر. فمثلا النطاق السلوكي في مجال الرياضيات ربما يشتمل على المهارات المختلفة اللازمة لحل مجموعة من المعادلات الآنية. والنطاق السلوكي في مجال الاتجاهات ربما يشتمل على مجموعة من الاستجابات التي تتطلب الموافقة أوعدم الموافقة على عبارات تتصل بموضوع معين. وهذه المهارات تحتاج إلى تحليل دقيق لمكوناتها حتى نستطيع تحديد النطاق السلوكي الذي يشتمل عليها. وفي أي من الأحوال يجب التأكد من أن صياغة النطاق السلوكي أصبحت من الدقة والوضوح بحيث يمكن استخدامها بسهولة في بناء الاختبار، ويجب ان يكون هناك اتفاق بين مجموعة من المحكمين على إمكانية كتابة مفردات متجانسة ومتسقة مع صياغة مكونات النطاق.

ونظرا لان الهدف من صياغة النطاق هو تحديد ما يشتمل عليه الاختبار أو ما يقيسه، فقد أطلق بابام على هذا المخطط اسم " مواصفات الاختبار Test Specification ولذلك تسمى هذه الطريقة بطريقة إعداد مواصفات الاختبار. وتشتمل هذه الطريقة على أربع مكونات تستخدم في تحديد ووصف السلوك الذي يمكننا من معرفة ما يستطيع الفرد أداءه، ومالا يستطيع أداءه في الاختبار.وينبغي أن نؤكد إن استخدام هذه الطريقة لا يمكننا من الحصول على نطاق سلوكي محدد تحديداً كاملاً.

ويمكن التغلب على هذه المشكلة إلى حد ما بأن يحاول مصمم الاختبار معرفة حدود النطاق السلوكي المطلوب قياسه. وأن يقرر ما إذا كان من الأفضل اختيار قطاع عريض من السلوك ويحاول صياغته بالتفصيل لكي يحدد النطاق السلوكي تحديدا كافياً أو اختيار قطاعات صغيرة من السلوك ويحدد مواصفات كل قطاع منها.

ويفضل بابام اختيار قطاعات صغيرة من السلوك لأنه يرى أنها أيسر في تحديدها وصياغتها. ولكنه يقترح مراعاة العوامل التالية عند اتخاذ قرار الاختيار بين نوعي القطاعات:

١ ـ الفترة التعليمية: أي الفترة التي تستغرقها عملية التعليم لكي يكتسب الفرد المهارة أو السلوك المطلوب قياسه.

٢ ـ طرق القياس الممكنة: اذ ان السلوك يمكن قياسه بطرق متعددة.

٣ ـ تجانس فقرات الاختبار: فالنطاق السلوكي الذي يقيسه الاختبار يجب ان يسمح ببناء فقرات اختبار متجانسة، أي متشابهة في الشكل والمحتوى، وهذا لا يتأتى إلا إذا كان النطاق محدودا. ولتحديد النطاق السلوكي الذي يقيسه الاختبار محكي المرجع اقترح بابام استخدام الخطوات أو المكونات الآتية:

أ ـ وصف عام لما يقيسه الاختبار: وهو عبارة عن وصف عام للسلوك الذي يتم تحديده بالتفصيل. ويمكن اعتبار هذا الوصف هو صياغة للأهداف الإجرائية (السلوكية) التي يقيسها الاختبار الا ان بابام يفضل تسميته " الوصف العام ". وينبغي ملاحظة أن الوصف العام لما يقيسه الاختبار يمكن تعديله وتطويره بعد إجراء الخطوات التي تلي ذلك.

ب ـ عينة من فقرات الاختبار: وهي توضح لمن يقوم ببناء الاختبار كيفية قياس الهدف قبل البدء في قراءة تفاصيل المواصفات التي سيأتي ذكرها في الخطوات التالية.

ج ـ تحديد عناصر المثيرات: والمقصود بذلك تحديد ووصف المثيرات التي تشتمل عليها مفردة الاختبار. فمفردة الاختبار تعد بمثابة مثير يقدم للطالب لكي يستجيب له بطريقة معينة.لذلك يجب تحديد ووصف مكونات هذا المثير وتحليله إلى عناصره.والحقيقة أن "عناصر المثيرات " تعد من أهم مكونات مواصفات الاختبار، اذ يعتمد عليها الفرد الذي يقوم ببناء الاختبار اعتماداً أساسياً، وتحديد هذه العناصر يحتاج إلى كثير من الجهد والوقت. فهي تتطلب معرفة جميع العناصر الرئيسة التي يمكن الاستعانة بها في تكوين وصياغة مفردات الاختبار.

ويرى بابام وكرونلاند انه إذا كان محتوى النطاق المراد قياسه غير محدود، يمكن اختيار عينة ممثلة من هذا المحتوى وكتابة عناصرها الأساسية: فمثلا إذا كان المحتوى هو القصة الأدبية، ربما يكون من الصعب تحديد جميع أشكال

وخصائص هذه القصص التي تشتمل عليها مثيرات مفردات الاختبار. لذلك من الأفضل تحديد عينة ممثلة لنطاق القصص الأدبية ولتكن عشرون قصة مثلا، أو أي عدد من القصص يمثل النطاق المطلوب. وهذا ييسر عملية تحديد العناصر التي تتضمنها المفردات.

د- تحديد عناصر الاستجابة: والمقصود بذلك تحديد ووصف طريقة استجابة الفرد لعناصر المثيرات التي تشتمل عليها مفردة الاختبار. ويوجد نوعان رئيسيان من الاستجابات هما:

الاستجابات المقيدة: مثل الاستجابات لمفردات الاختيار من متعدد أو الصواب والخطأ أو المزاوجة .

الاستجابات المفتوحة:مثل الاستجابات لمفردات المقال والمفردات التي تتطلب إجابة قصيرة. وتتكون عناصر الاستجابة من مجموعات من العبارات التي نحاول عن طريقها تحديد نوع وطريقة الاستجابة لمفردات الاختبار المستخدمة.فإذا كانت المفردة من نوع الاختيار من متعدد مثلاً، فإنه يجب تحديد طبيعة الإجابة الصحيحة، وكذلك قواعد تكوين " البدائل " أي الإجابات غير الصحيحة بحيث يمكن أن يسترشد بها الفرد الذي يقوم بكتابة المفردات وبالمثل في الأنواع الأخرى من المفردات التي تتطلب إجابة مقيدة.

أما إذا كانت المفردة من النوع الذي يتطلب إجابة مفتوحة، فإنه يجب وضع محكات تفصيلية واضحة لطريقة الإجابة، وطريقة تصحيح الإجابات المختلفة. وهذا في الحقيقة ليس بالأمر اليسير، إذ انه يتطلب إجراء عملية مسح للإجابات المتوقعة وتحديد عناصرها الرئيسة في عبارات واضحة. ويشترط في كتابة الهدف السلوكي أن يرتبط بالهدف العام بحيث يحقق الشروط الآتية:

- يجب أن يمثل الهدف السلوكي عينة من السلوك التي يحتويها الهدف العام.
- يجب أن يتفق الهدف السلوكي والهدف العام في المستوى المعرفي.
- يجب أن يشتمل الهدف السلوكي على المكونات الرئيسة الأربع وهي:

الدارس - الفعل السلوكي – شروط الأداء - مستويات الأداء.

- يجب أن لا تزيد عبارة الهدف الإجرائي عن (٢٠) كلمة.

وتعد الإجابة صحيحة إذا حقق الهدف المحكات الأربعة السابقة وتعد الإجابة خاطئة إذا لم يحقق الهدف جميع المحكات أو حقق بعضاً منها فقط.

هـ:ملحق الموصفات: يضاف أحياناً إلى المكونات السابقة ملحق للمواصفات تكتب فيه التفصيلات التي تؤدي إلى مزيد من التوضيح لمحتوى النطاق السلوكي أو محتوى مفردات الاختبار أو غير ذلك من عناصر المثير والاستجابة. وهذا يقلل من حجم التفاصيل التي تكتب في الخطوتين السابقتين وبذلك يمكن استخدام المواصفات بطريقة أكثر فاعلية.

ثانياً: صيغ المفردات Item Forms وهذه الصيغة تتضمن قواعد تكوين المفردات أو هيكل أو تصميم يستخدم في تحديد المحتوى المراد قياسه، وبخاصة إذا كان المحتوى له طابع تكويني من مثل موضوعات الحساب والفيزياء وغيرها.

وقد توصل هايفلي Hively وأوزبرن Osburn إلى فكرة هذه الصيغ. وفي هذه الصيغة يمكن أن نضمن تكوين مفردات متجانسة من حيث الشكل والمحتوى بحيث يمكن عن طريقها الحصول على النطاق الشامل للمفردات التي يتكون منها الاختبار.

ثالثاً: مخططات تعتمد على القواعد اللغوية Linguistic-Based schemes يرى بورموث ١٩٧٠ Borsht أنه إذا اعتمد تفسير درجات الاختبارات محكية المرجع اعتماداً كليا على خصائص التعليم ـ وهو بالطبع الهدف الذي يبني من اجله هذا النوع من الاختبارات ـ فأن هذا يتطلب افتراض ان كل مفردة من مفردات الاختبار تكون مرتبطة ارتباطاً تاماً بالبرنامج التعليمي. فالطرق التقليدية في بناء الاختبارات معيارية المرجع تسمح لكاتب الاختبار أن يغير من الصياغة والأسلوب اللغوي الذي استخدم في التعليم تبعا لما يراه. ومن الظواهر المؤكدة أن تغيير شكل المفردة أو صياغتها يكون له تأثير معين على مستوى صعوبة المفردة.

ونظراً لان كاتب المفردات يكون له حرية مطلقة في إجراء هذا التغيير، فإن صعوبة المفردات، وبالتالي الدرجات تتأثر بتحيز كاتب المفردات، وبالأسلوب اللغوي الذي أستخدمه في صياغتها. ولذلك فإن الاختبارات التي تبنى بالطرق التقليدية تقيس خصائص الشخص الذي يكتب المفردات إلى جانب قياسها لخصائص البرنامج التعليمي. ولذلك اقترح بورموث أساسا لغويا لتحديد النطاق السلوكي الذي

يستخدم في كتابة مفردات الاختبار أطلق عليها التعريفات الإجرائية Operational Definitions وهو يقصد بذلك " تحويلات المفردات Item Transformations "

وهو في هذا الصدد يذكر بورموث انه يجب أن نتمكن من تطبيق هذه الإجراءات Operations بطريقة منتظمة في أي برنامج تعليمي إذا أردنا تكوين جميع المفردات التي تشتق من هذه الإجراءات. وتحقق هذا الشرط لا يضمن فقط إمكانية تعريف النطاق الشامل للمفردات، وإنما يضمن أيضاً تحييد الشخص الذي يقوم بتكوين هذه المفردات. أي أن لا يكون له أي تأثير على الصياغة اللغوية للمفردات. فمفردة الاختبار ترتبط بالتعليم إذا أمكن اشتقاقها بطريقة موضوعية من جزء محدد من التعليم، وهذا لا يتأتى إلا إذا استخدمت مجموعة من الإجراءات التي يمكن تعميمها على البرامج التعليمية المختلفة. وهذا يعني انه يجب ان يكون هناك تناظر بين المفردة والجزء التعليمي المرتبط به، وكذلك بين الاستجابة الصحيحة للمفردة وذلك الجزء.

مثال: في تعليم اللغة العربية، ربما تشير الإجراءات إلى " التحويلات " التي تجري على " الخصائص اللغوية للجملة " من مثل:

" ركب الولد الحصان "، للحصول على المفردات:

من ركب الحصان ؟

هل ركب الحصان الولد ؟

ماذا ركب الولد ؟ ... وهكذا.

وبذلك يتضح من هذه الأمثلة انه يمكن بناء القواعد التي تكون على أساسها مفردات الاختبار، بحيث نحصل على نفس المفردات أو الأسئلة مهما اختلف الأفراد الذين يقومون ببناء الاختبار. إذ ان هذه القواعد تتضمن الإجراءات التي تشتق على أساسها المفردات.

تصنيفات الاختبارات محكية المرجع Categories of CRT

تصنف الاختبارات محكية المرجع على حسب نطاقها السلوكي إلى قسمين وذلك ما أشار به نيتكو Nitko (١٩٨٠) هما:

١ ـ اختبارات محكية المرجع تعتمد على نطاقات مرتبة: Ordered Domains

يعتمد هذا النوع على محكات تساعد على ترتيب النطاقات السلوكية وهذه المحكات عبارة عن:

أ ـ تحكيم المعايير الاجتماعية أو نوعية الأداء (Judged Social or esthetic quality of performance)

ب- مستوى صعوبة أو تعقد المقرر الدراسي Complexity or difficulty level of subject matter)

ج- درجة الكفاءة المطلوبة للمهارات المركبة (Degree of proficiency with which complex skills are performed)

د- متطلبات اكتساب المهارات العقلية والحركية Prerequisite sequence for acquiring intellectual and psychomotor skills

هـ ـ موقع السمات الكامنة المعرفة (Location on empirically defined latiat trait)

٢ ـ اختبارات محكية المرجع تعتمد على نطاقات غير مرتبة:

كما أشار نيتكو (Nitko ١٩٨٠) إلى أن هذا النوع يتطلب تعريف النطاقات السلوكية ويحتوي على أربعة أنواع من النطاقات يمكن إيجازها كالتالي:

أ - خصائص مثيرات النطاق وخطة اختيار مفردات الاختبار (Stimulus Domain and Sampling Plan of Test) Properties of

ب- نطاقات تهتم بالخصائص اللفظية للمثيرات والاستجابات (Verbal statements (of Stimuli and responses in domin

ج- نطاقات تعتمد على تشخيص تصنيفات الأداء (Diagnostic Categories of Performance),

د- نطاقات تعتمد على التجريد والسمات أو التكوينات الفرضية (Abstractions , raits,. (or Constructs

خطوات بناء الاختبارات محكية المرجع (الجانب المعرفي)

تستخدم الاختبارات محكية المرجع في تحديد مستوى أداء كل طالب بالنسبة إلى نطاق سلوكي من المعارف والمهارات الأساسية المعرفة تعريفا دقيقا دون موازنة هذا الأداء بأداء غيره من الطلاب. فالنطاق السلوكي يكون بمثابة المحك

obeikan.com

Criterion الذي ينسب إليه أداء الطالب كما سبق ان ذكرنا. وهذا بالطبع يجعل تحديد النطاق السلوكي الذي يقيسه الاختبار تحديدا دقيقا هو المشكلة الأساسية التي تواجه من يقوم ببناء اختبار محكي المرجع. ويمر بناء الاختبار محكي المرجع لقياس الجانب المعرفي بالخطوات الاتية:

اولاً: تحديد المحتوى المراد قياسه : يعتمد ذلك على طبيعة وحدود ذلك المحتوى، فإذا كان المحتوى المراد قياسه محددا (مثل وحدة دراسية معينة، أو موضوع دراسي واضح المعالم) فإنه يمكن الاكتفاء بمعرفة مكونات هذه الوحدة أو هذا المحتوى. أما إذا كان المحتوى المراد دراسته واسعا وكبيرا فإنه من المستحسن تقسيمه إلى موضوعات فرعية مرتبطة ببعضها البعض بحيث يمكن قياسها كوحدة. ويتطلب ذلك بناء عدة اختبارات كل منها يرتبط بأحد هذه الموضوعات الفرعية. ولا توجد قاعدة ثابتة لتقسيم المحتوى المراد قياسه، ولكن يجب أن يسمح هذا التقسيم ببناء اختبار أو اختبارات تكون مفرداتها بمثابة عينة ممثلة تمثيلا كافيا للموضوعات المختلفة التي يشتمل عليها المحتوى. كما يجب مراعاة طبيعة الطلاب المختبرين، والفترة الزمنية اللازمة لتعليم هذا المحتوى.

ثانيا: تحديد الأهداف العامة التي سيقيسها الاختبار: إن لكل محتوى أهداف مرتبطة به تهدف إلى قياس مدى تحققها لدى الطلاب. ومثال ذلك الهدف العام " يطبق الطالب الطريقة الاستنباطية في التفكير عند حل المسائل الرياضية "

وهذا الهدف العام يعبر عن نواتج متوقعة من عملية التعليم، ولكنها تحتاج إلى أعادة صياغة حتى تعبر عن أنشطة سلوكية تكون دليلا على أن الطالب قد طبق الطريقة الاستنباطية وما إلى ذلك.

ثالثا: تحليل الأهداف العامة إلى أهداف جزئية ونعني بذلك إعادة صياغة الأهداف العامة عن طريق وصف عينة من الأهداف السلوكية التي يمكن إن تتخذ دليلاً على تحقق كل هدف من الأهداف العامة. وعادة يواجه مصمم الاختبار بعض الصعوبات عند إجراء هذا التحليل.ويعتمد هذا على النطاق السلوكي المراد قياسه. فإذا كان هذا النطاق محدودا فإنه يمكن استخدام طريقة جاجني ١٩٦٨ Gagne في تحليل المهام Task Analysis، أو طريقة ميجار ١٩٧٢ Mager في تحليل الأهداف

Goal Analysis. وهاتان الطريقتان تفيدان في التوصل إلى مجموعة الأهداف أو المهام التي تسهم في تحقيق الهدف العام.

رابعا: تحديد مواصفات النطاق السلوكي الذي يقيسه الاختبار عند الحصول على الأهداف الإجرائية من التحليل السابق قد لا تكون واضحة بدرجة تسمح بتحديد نطاق مفردات الاختبار التي تقيس السلوك المطلوب. وبالطبع يصعب اختيار عينة ممثلة لمفردات الاختبار من النطاق إذا كان هذا النطاق الذي يقيس هدفا معينا غير واضح. وقد سبق وان ذكرنا أن أداء الفرد في عينة من مفردات الاختبار التي تقيس هدفا معينا يستخدم في تقدير درجة إتقانه للنطاق الشامل للمفردات التي تقيس هذا الهدف. لذلك فأنه يجب تعريف وتحديد هذا النطاق الشامل للمفردات تحديدا جيدا يمكن سحب عينات ممثلة من هذه المفردات.

فإذا كان النطاق الشامل يصعب تحديده تحديدا كاملا، فإنه يمكن اعداد مواصفات الاختبار باستخدام طريقة بابام التي تم عرضها في تحديد النطاقات السلوكية. أما إذا كان النطاق الشامل حدوده معرفاً تعريفا جيدا فأنه يمكن استخدام صيغ المفردات التي اقترحها هايفلي التي اشرنا لها سابقاً.

خامسا:تكوين مفردات الاختبار : وتشمل الخطوات الآتية:

١ ـ اختيار أنواع المفردات المناسبة لقياس الأهداف بعد تحديد مواصفات النطاق السلوكي يجب اختيار أفضل أنواع المفردات التي تقيس السلوك الذي تم تحديده في المواصفات الخاصة بكل هدف. فمن المعلوم ان بعض أنواع المفردات تصلح لقياس أهداف معينة بدرجة أفضل من غيرها. فمثلا تصلح مفردات الاختيار من متعدد، ومفردات المزاوجة، والصواب والخطأ، لقياس التذكر والفهم وأحيانا التطبيق، بينما تصلح مفردات المقال لقياس القدرة على تنظيم المعلومات، والاستنتاج، والتفسير، وإعادة صياغة الأفكار.... وهكذا.

٢ ـ تحديد العدد المناسب من المفردات في هذه الخطوة نحاول تحديد العدد المناسب من المفردات التي تقيس الأهداف الممثلة في النطاق السلوكي الذي سبق تحديده ويجب مراعاة العوامل الاتية:

ذكر كل من هامبلتون وإيقنور (١٩٨٠) Hambleton & Eignor)بأن هناك عاملين ينبغي وضعهما في الاعتبار عند تحديد أسئلة الاختبار وهما:

أ ـ العلاقة بين عدد أسئلة الاختبار والأهمية الخاصة بكفاية معينة:تتفاوت أهمية الكفايات المراد قياسها في الاختبار، لذا ينبغي أن يختار مصمم الاختبار عند قياسه كفايات متعددة أسئلة تتناسب مع أهمية النطاق المراد قياسه.

ب- العلاقة بين عدد أسئلة الاختبار ومستوى الحد الأدنى المقبول لدرجة ثبات الاختبار:

ففي حالة الاختبارات معيارية المرجع تستخدم المعادلة العامة لسبيرمان براون لتقدير معامل الثبات اذا زاد أو نقص طول الاختبار (ن) من المرات. اما في حالة الاختبارات محكية المرجع فقد اقترح هامبلتون Hambleton بتحديد قيمة الدقة المطلوبة ((Degree of Precision) في تقدير درجة الطالب في النطاق، ويتم ذلك وفق الصيغة التالية:

$$\text{طول الاختبار} = \frac{٠,٢٥}{(\text{درجة الدقة المطلوبة})^{٢}}$$

حيث ((٠,٢٥) مقدار ثابت.

وأشار كل من هامبلتون، وآخرون (١٩٧٨ Hambleton et al) بأن الهدف من تطبيق الاختبار محكي المرجع هو تصنيف الطلاب إلى فئتين متقنة وغير متقنة، لذا فإن تحديد طول الاختبار المناسب الذي يقلل من أخطاء التصنيف (Number of Classification Errors)يتطلب تطبيق اختبار طويل يقلل من أخطاء القياس.)

وأضاف كل من هالادين ورويد Haladyn & Roid بأن قيمة أخطاء التصنيف تقل بشكل عام عند زيادة طول الاختبار، لكن النقص في قيمة الخطأ يكون ملحوظاً عند الانتقال من اختبار طوله (١٠) فقرات إلى اختبار طوله (٢٠) فقرة، وأما النقص في قيمة الخطأ عند الانتقال من اختبار طوله (٢٠) فقرة إلى اختبار طوله ٣٠- ٤٠ فقرة يكون صغيراً نسبياً لدرجة قد لا تبرر استخدام فقرات تزيد عن (٢٠) فقرة.

٣ ـ العلاقة بين عدد المفردات وزمن تطبيق الاختبار: يعتمد عدد المفردات بلا شك على الزمن اللازم لتطبيق الاختبار، ويعتمد هذا بالتالي على نوع مفردات الاختبار المستخدمة، وعلى خصائص الأفراد المختبرين.

ح ـ كتابة مفردات الاختبار :وهذه الخطوة تحتاج إلى عناية كبيرة، فمفردات الاختبار محكي المرجع تبنى على أساس مواصفات النطاق السلوكي الذي سبق أعداده. لذلك يجب أن تقيس هذا النطاق بدرجة كبيرة من الدقة، وان يكون مستوى صعوبة كل مفردة مناسبا لمستوى صعوبة الهدف الذي تقيسه، ولمستواه المعرفي. وان تكون عينة المفردات ممثلة للنطاق السلوكي للأهداف. كما يجب ان تراعى الأصول الفنية في كتابة الأنواع المختلفة للمفردات.

بناء الاختبارات محكية المرجع لقياس الجانب الوجداني

يحتاج بناء المقاييس والاختبارات محكية المرجع في الجانب الوجداني إلى قدر كبير من المهارة. فتقويم الجانب الوجداني للسلوك يعد أكثر صعوبة من تقويم الجانب المعرفي لأسباب كثيرة نذكر منها:

١ ـ صعوبة تحديد وتعريف مكونات الجانب الوجداني للسلوك. وكذلك صعوبة ملاحظة هذه المكونات غير المنظورة يحتاج إلى تحليل سلوكي، وملاحظة سلوك الفرد ملاحظة مباشرة تعكس هذه المكونات.

وبالطبع يواجه مصمم الاختبار نفس المشكلة عند قياس مكونات الجانب المعرفي، أو ما يسمى بالسلوك أو الأداء الأقصى حيث يؤدي الفرد أفضل أداء ممكن ليحصل على درجة مرتفعة في الاختبار. ولكن المشكلة تزداد حدتها عند قياس الجانب الوجداني للسلوك أو ما يسمى بالسلوك أو الأداء المميز، حيث يستجيب الفرد للمواقف الاختبارية كما هو بالفعل في سلوكه اليومي المعتاد.

٢ ـ يتطلب تقويم الجانب الوجداني للسلوك ملاحظة الفرد ملاحظة متكررة في عينات مختلفة من المواقف التي يمكن ان تظهر فيها الاستجابات التي نهتم بملاحظتها وتقويمها. فلكي ندرس ميول فرد، فإننا نحتاج إلى ملاحظته في وقت فراغه مثلاً، ولكن ملاحظة سلوكه في موقف واحد لا يكفي للحكم على أدائه

المتميز. فالسلوك الوجداني يعتمد على مزاج الفرد، والألفة بالموقف، وتفاصيل البيئة المحيطة به، وعوامل أخرى كثيرة. فالفرد الذي يبدو عدوانياً في كثير من المواقف والظروف ربما نحكم عليه بأنه شخص عدواني، فقد يكون في حالة من الضيق والتوتر نتيجة قلقه المستمر على شيء ما، ولكنه ربما يبدو في حالة تكيف تام بعد انقضاء عدة أسابيع أو عدة شهور. ولذلك فإن ما نلاحظه من سلوك فرد معين في مواقف قليلة، ربما يكون مجرد انحراف وقتي، ولكن بلاشك إن هذا السلوك قد ميز الفرد في هذا الوقت.

ونظرا لصعوبة ملاحظة الفرد طوال اليوم أو خلال أطول فترة ممكنة، فإننا نعتمد على الاستجابات اللفظية التي نطلبها منه، عن طريق الأسئلة المباشرة أو عن طريق الاستبيانات. ولكن يجب تفسير البيانات المستمدة من هذه الاستبيانات بحذر شديد، والتأكد من صدق التفسير باستخدام ما نجمعه عن الفرد من ملاحظات أثناء قيامه بنشاطات مختلفة.

٣ ـ نظرا لصعوبة التحليل السلوكي الدقيق لكثير من المفاهيم الوجدانية من مثل مفهوم الذات أو مفهوم القيم، أو مفهوم الاتجاهات، فإننا لا نستطيع الاعتماد على أداة قياس واحدة تقيس أبعاد كل مفهوم من هذه المفاهيم، إذ من الصعب احتواء هذه الأبعاد جميعا وقياسها مباشرة بأداة قياس واحدة. لذلك ربما يكون من الأفضل إتباع إستراتيجية الالتفاف حول هذه الأبعاد، أي نستخدم عددا من المقاييس المختلفة التي تحتوي أهم هذه الأبعاد، وبذلك نحصل على صورة أفضل عن مكونات كل من هذه المفاهيم.

٤ ـ يحاول الأفراد في كثير من الأحيان تزوير استجاباتهم أو سلوكهم المطلوب قياسه إذا ما شعروا أنهم سوف يكونون عرضة للنقد، أو أن الدرجات التي ستقدر لهم في المواقف الاختبارية سوف تؤثر بشكل من الأشكال على التحاقهم بعمل معين أو على مستقبلهم الدراسي مثلا. ولذلك يجب أن نفرق بين استخدام البيانات المستمدة من استجابات الأفراد لهذه المواقف لغرض التقويم الجماعي، واستخدامها لغرض التقويم الفردي. ففي الحالة الأولى يمكن ان يستجيب الفرد لهذه المواقف (التي تكون عادة استجابات لفظية) دون الحاجة إلى ذكر أسمائهم على صحيفة

الاستبانة مثلاً.ومن هنا يمكن الاستفادة من البيانات في اتخاذ قرارات جماعية مثل تأثير موضوع أو قضية معينة في السلوك الوجداني لمعظم الأفراد. أما في الحالة الثانية فإنه يمكننا ان ننصح الأفراد بكتابة أسمائهم على صحيفة الاستبانة أو غيرها من التقارير الذاتية حتى يتمكن كل منهم من تقويم نفسه تقويماً ذاتياً بالنسبة للبعد الوجداني المطلوب قياسه.

ويفضل في مثل هذه الحالة استخدام الاستبيانات قهرية الاختيار للتغلب على عامل الرغبة الاجتماعية أو الميل إلى المعايير الاجتماعية، وهو نوع من تحيز الاستجابات، حيث يحاول الفرد بطريقة شبه شعورية أن يكون مقبولا اجتماعياً. لذلك فهو يوافق على عبارة مثل " أحاول قدر استطاعتي تجنب إثارة المشاكل " أو " أنا شخص ذكي " وقد أكد الين ادوردز E. Edwards وهو من علماء النفس الذين اهتموا ببحث هذا النوع من تحيز الاستجابات، إن موافقة الفرد على مثل هذه العبارات لا يعتبر نوعا من الكذب لأن الاستجابة تكون شبه شعورية في حين ان الكذب يكون متعمدا. ويعتمد الفرد في استجابته على محتوى العبارات، ويكون مدركاً للعلاقة بين محتوى العبارات وطريقة تصحيحها. وللتغلب إلى حد ما على هذه المشكلة اقترح ادوردز استخدام هذا النوع من الاستبيانات، وفيه تعرض بنود الاستبانة على الفرد بحيث يشتمل كل بند منها على عبارتين تبعد كل منهما عن المعايير الاجتماعية بنفس الدرجة، وعلى الفرد أن يحدد قبوله أو رفضه لاحداهما.

فمثلاً يكون البند مشتملاً على عبارتين كالتالي:

أ ـ أحب أن أتحدث عن نفسي أمام الناس.

ب ـ أحب أن اعمل لتحقيق الهدف الذي أصبوا إليه.

فالفرد الذي يختار الاستجابة الأولى مثلاً ربما يقدر له درجة واحدة على ميزان أو مقياس الاستعراضية Exhibition Scale، أما إذا اختار الاستجابة الثانية، فأنه ربما يقدر له درجة واحدة على ميزان أو مقياس التحصيل والإنجازAchievement Scale.

ويمكن باستخدام هذا النوع من الاستبيانات التغلب بدرجة معقولة على مشكلة الميل إلى المعايير الاجتماعية في استجابات الأفراد للاستبيانات والقوائم Inventories.

وقد اقترح كراندال Crandall وكاتكوفسكي ١٩٦٥ Katkovsky طريقة أخرى تفيد في التعرف على العبارات المرغوبة اجتماعياً. وتتمثل هذه الطريقة في انتقاء مجموعة من العبارات التي تؤدي إلى استجابات مرغوبة اجتماعياً وأضافتها إلى عبارات المقياس المطلوب. وتعتبر عندئذ هذه العبارات المنتقاة مقياساً للميل إلى المعايير الاجتماعية. فاستخدام هذا المقياس الأخير يمكننا من التعرف على عبارات المقياس الوجداني التي ترتبط ارتباطاً عالياً بدرجة الفرد في الميل إلى المعايير الاجتماعية. وبذلك يمكن حذف هذه العبارات من المقياس المطلوب.

تحليل الجوانب الوجدانية للسلوك: من هنا يتضح أن قياس الأبعاد المختلفة للسلوك الوجداني باستخدام الاختبارات والمقاييس محكية المرجع، يتطلب التفكير في العوامل التي تؤثر في الطريقة المميزة لاستجابات الفرد للمواقف المختلفة المرتبطة بهذه الأبعاد، أي في النزعة العامة لاستجاباته. وبالطبع لا نستطيع قياس هذه النزعة بدرجة تامة من التأكد، وإنما يمكن تقديرها. فمن المستحيل التنبؤ بدرجة تامة من الثقة بكيفية سلوك الفرد في ظروف أو مواقف معينة، وإنما يمكن معرفة احتمال ان يستجيب بطريقة معينة في بعض المواقف.

ويكمن الفرق الرئيسي بين قياس كل من المجالين المعرفي والنفسحركي من جهة، والمجال الوجداني من جهة أخرى، في الفرق بين ما يستطيع الفرد أن يفعله، وبين ما سيفعله في المستقبل. بمعنى أنه في الجانبين المعرفي والنفسحركي نفترض ان الفرد ربما لا يكون قادرا بالفعل على تحقيق هدف معين، بينما في الجانب الوجداني نفترض ان الفرد يمكنه تحقيق هدف وجداني، ولكنه ربما لا يفضل ان يسلك السلوك المرتبط بالهدف بمحض إرادته.

ولذلك فإن قياس الجوانب الوجدانية للسلوك يتطلب الحصول على دلائل عن الأفراد أما في صورة نواتج معينة، أو في صورة سلوك يمكن ملاحظته ملاحظة مباشرة في بعض المواقف. وهذه المواقف التي تعتبر بمنزلة المثيرات التي نلاحظ عن طريقها النتاج أو السلوك المعين يجب أن تسمح لنا بالاستدلال على مشاعر الأفراد في مواقف أو مثيرات مستقبلية مشابهة لها. فمثلاً ربما يحاول معلم قياس توافق طالب معين مع أقرانه عن طريق ملاحظة سلوكه في مواقف تتطلب علاقات مع الأقران،

أو عن طريق استخدام التقارير الذاتية على أمل أن يحصل المعلم بهذه الطرق على بعض الدلائل التي تساعده في التنبؤ بسلوك الطالب مع الآخرين في مستقبل حياته.

فعلى الرغم من اختلاف المواقف التي يمكن أن يستخدمها المعلم في هذا التقويم عن مواقف الحياة المستقبلية للطالب، إلا انه يجب أن يكون هناك نوع من التطابق أو التشابه بين هذه المواقف. فهذا التشابه هو الذي يسمح بالتنبؤ بقدر من الثقة بسلوك الطالب في المستقبل. وبالطبع يرتبط هذا ارتباطاً وثيقاً بطرق بناء أدوات قياس الجوانب الوجدانية. اذ أن هذه الأدوات تشتمل عادة على مواقف سلوكية ترتبط بالأبعاد الوجدانية المطلوب قياسها. وهذا يتطلب بالطبع تحليل هذه الأبعاد تحليلاً سلوكياً أي تحديد النطاق السلوكي تحديداً جيداً. ويجب ان يعتنى بتحديد هذا النطاق السلوكي بالقدر نفسه من الدقة الواجبة في الاختبارات محكية المرجع التي تقيس الجوانب المعرفية.

طرق تقدير صدق الاختبارات محكية المرجع:

١ـ الصدق الوصفي Descriptive - Validity ذكر بابام ١٩٧٨ Popham بأن الصدق الوصفي عبارة عن محاولة لتأكيد مدى ما يقيسه فعلاً الاختبار محكي المرجع، ويعتبر تحديد الصدق الوصفي شرطاً أساسياً لأنواع الصدق الأخرى وذلك لأنه أفضل وسيلة لتفسير أداء الطالب في الاختبار. ويمكن تحديد الصدق الوصفي باستخدام أساليب مختلفة، ومن ذلك الاستعانة بمجموعة من المتخصصين لتحديد مدى مطابقة Congruent مفردات الاختبار مع المجال السلوكي المراد قياسه.

٢ ـ الصدق الوظيفي Functional - Validityأشار بابام (١٩٧٨ Popham) بأن المقصود بالصدق الوظيفي هو دقة الاختبار محكي المرجع في تحقيق الغرض) الوظيفة(الذي صمم من أجله الاختبار، أي أنه لا يقتصر على وصف أداء الطالب في الاختبار، ويستخدم الصدق الوظيفي عندما يكون الهدف من تطبيق الاختبار وصف وتحديد الوظيفة المراد تحقيقها من الاختبار، لذا فإن الصدق الوصفي يعتبر شرطاً لازماً لحساب الصدق الوظيفي. ويقابل الصدق الوظيفي في القياس المعياري المرجع صدق المحك الخارجي criterion Related Validity والذي يدل على قدرة

الاختبار على التنبؤ بأداء طالب ما في موقف تالي من أداء الطالب على اختبار سبق التأكد من صدقه، وأطلق على هذا النوع من الصدق (بالصدق الوظيفي) بدلاً من صدق المحك الخارجي لأن بعض الوظائف المراد قياسها في الاختبار محكي المرجع تتطلب محكاً خارجياً والبعض الآخر لا يتطلب ذلك.

وأضاف هامبلتون بأنه يمكن تحديد صدق القرار بمقارنة أداء مجموعتين محكيتين أو أكثر بدرجة قطع محددة. كما يمكن تحديد صدق القرارات التعليمية المبنية على درجات الاختبارات محكية المرجع بمقارنة أداء مجموعتين (الأولى تلقت التعليم والثانية لم تتلق التعليم) بدرجة قطع الاختبار. ويمكن تقييم صدق القرار لكل هدف تعليمي بجمع النسب المئوية للطلاب المتقنين الذين اجتازوا درجة القطع والنسب المئوية للطلاب غير المتقنين الذين لم يجتازوا درجة القطع، وذلك بهدف تحديد الطلاب المتقنين فعلاً والطلاب غير المتقنين فعلاً. ويمكن تحديد صدق القرار أيضاً بحساب معامل الارتباط بين متغيرين ثنائيين الأول متغير الانتماء إلى المجموعة (Group-Membership) والثاني متغير قرار الإتقان. (Mastery Decision)

٣ ـ صدق انتقاء النطاق السلوكي Domain Selection Validity أشار بابام (١٩٧٨) Popham بأن صدق انتقاء النطاق يتعلق بمدى دقة باني الاختبار في اختيار نطاق سلوكي محدد يمكن أن يستخدم كمؤشر ((Indicator) لإيضاح مستوى الطالب بالنسبة إلى (نطاق عام) يعد أكثر عمومية. ويمكن تحديد صدق انتقاء النطاق السلوكي بأساليب متعددة منها مايلي:

أ- تدريس مجموعات صغيرة من الأفراد المهارات التي يشتمل عليها كل نطاق من النطاقات السلوكية حتى يصلوا إلى درجة الإتقان، يلي ذلك قياس إمكانية تعميم إتقان الطلاب لنطاق واحد تجاه النطاقات الأخرى.

ب- مقارنة أداء مجموعتين من الأفراد إحداهما (تلقت التعليم) والأخرى (لم تتلق التعليم)، وذلك بتطبيق عدد من الاختبارات، ثم يحدد النطاق السلوكي الذي يميز (Discriminates) بصورة أفضل بين المجموعتين.

طرق تقدير ثبات الاختبارات محكية المرجع: تختلف وتتعدد طرق تقدير ثبات الاختبارات محكية المرجع باختلاف استخداماتها، وصنف هامبلتون، وآخرون (Hambleton, et al ١٩٧٨) تلك الطرق إلى ثلاث مجموعات رئيسة، وسنتطرق لمجموعتين هما:

تقدير ثبات درجات الطلاب في النطاق السلوكي وتقدير ثبات تصنيف الطلاب في مجموعات على حسب درجة تمكنهم من النطاق السلوكي.

أولاً - تقدير ثبات درجات الطلاب في النطاق السلوكي:

طرق هذه المجموعة تهتم بتقليل خطأ التباين الناتج عن الفروق بين درجات الطلاب في النطاق السلوكي بالإضافة إلى درجات النطاق، ومن تلك الطرق مايلي:

معامل ليفنجستون Livingston-Index: ذكر كل من كروكر، والجاينا Algina & Crocker بأن هذا المعامل يهتم بإيجاد انحراف درجات الأفراد عن درجة القطع، ويهتم هذا المعامل أيضاً بمفهوم متوسط مجموع مربعات انحرافات الدرجات الملاحظة المتوقعة وذلك بحساب قيمة انحراف درجات كل طالب عن درجة القطع في الاختبار، ويتم ذلك وفق الصيغة الرياضية التالية:

$$K(X,T)= \frac{\sigma T+(Mt-NiC)}{\sigma X+(Mx-NiC)}$$

حيث أن:

K(X,T): ترمز إلى معامل ليفنجستون.

T: ترمز إلى متوسط مجموع مربعات انحرافات الدرجات الملاحظة حول درجة القطع (C)، X: ترمز إلى متوسط مجموع مربعات انحرافات الدرجات الحقيقية عن درجة القطع (C).

M: ترمز إلى متوسط درجات الطلاب في النطاق السلوكي Domin الذي يقيسه الاختبار، N i :ترمز إلى عدد الأسئلة ، C : ترمز إلى درجة القطع .

ولتقدير معامل ليفنجستون K(X,T) في حالة استخدام اختبارين لكل طالب تستخدم الصيغة

الرياضية التالية:

$$K\hat{}(X,T) = \frac{P\hat{}xx` \; б\hat{}x \; б\hat{}x` + (M\hat{}x - N\,i\,C)(M\hat{}x` - N\,i\,C)}{\sqrt{\{ б\hat{}x + (M\hat{}x - N\,i\,C)\}\{ б\hat{}x` + (M\hat{}x` - N\,i\,C)\}}}$$

حيث أن:

$P\hat{}xx`$: ترمز إلى معامل الارتباط بين درجات كلا الاختبارين.

$x\hat{}$: ترمز إلى متوسط مربعات انحرافات الدرجة الحقيقية عن درجة القطع.

$M\hat{}$: ترمز إلى متوسط الدرجات الخام

N i :ترمز إلى عدد الأسئلة

C : ترمز إلى درجة القطع

ثانياً ـ تقدير ثبات تصنيف الطلاب في مجموعات على حسب درجة تمكنهم من النطاق السلوكي: تهتم طرق هذه المجموعة بتحديد الأخطاء الناتجة عن عدم اتساق التصنيف عند تطبيق اختبارين متوازيين أو في حالة إعادة تطبيق الاختبار على أفراد العينة، ومن تلك الطرق مايلي:

١. طرق تتطلب تطبيق الاختبار مرة واحدة:

أ- **طريقة هاريس** Harris Method أشار هاريس (Harris ١٩٧٤) بأن هذه الطريقة لا تعتمد في تحديدها على متغير طول الاختبار (k) وإنما تقوم على الارتباط بين متغير يمثل مجموع درجة الطالب في الاختبار ومتغير ثنائي آخر يمثل تصنيف الطالب إلى متقن أو غير متقن على ضوء درجة القطع، لذا فإنه يتم تصنيف الطلاب إلى أربع مجموعات في (جدول رباعي) وهي طلاب متقنين فعلاً، و طلاب غير متقنين فعلاً. ومجموعة طلاب الخطأ الأول (α) ومجموعة طلاب الخطأ الثاني (β) ويتم تحديدهم بناءً على درجة قطع الاختبار.ويطلق على هذه الطريقة بمعامل (هاريس) ويرمز له بالرمز (Mc) ويتم تقديره بالصيغة التالية:

$$M_c = \frac{SS_b}{SS_b + SS_w}$$

حيث إن:

M_c = معامل ثبات هاريس.

SS_b ، SS_w = عبارة عن مجموع المربعات (داخل، بين) المجموعات.

وتمتد قيمة المعامل ما بين (صفر وواحد).

ب ـ معامل كابا (هاينا) Huynh Kappa Coefficient

يمكن تقدير معامل (هاينا) في حال تطبيق اختبار واحد أو اختبارين متوازيين، ونظراً لصعوبة تقدير هذا المعامل في حال تطبيق اختبارين متوازيين، لذا سيتم التطرق لكيفية تقدير معامل الثبات (هاينا) لاختبار واحد فقط، وفي هذا الصدد أشار سابكوفياك Subkoviak بأن تقدير هذا المعامل يستند إلى عدة فروض هي:

(Beta. Distribution)أن يتمثل توزيع درجات الطلاب الحقيقية في الاختبار شكل بيتا إذا طبق عدد من الاختبارات على طالب ما، فإنه يفترض أن يتخذ توزيع درجاته في الاختبار شكل التوزيع ذي الحدين، وفي هذه الحالة يجب أن يكون تقدير درجات أسئلة الاختبار إما صفراً أو واحداً، وأن تكون الإجابة على الأسئلة مستقلة إحصائياً بحيث لا تؤثر الإجابة على إحدى المفردات على إجابة المفردة الأخرى، وأن تتساوى تقريباً معاملات صعوبة جميع أسئلة الاختبار. ويتم تقدير معامل ثبات(هاينا) وفق إحدى الصيغتين التاليتين:

$$K = \frac{P\hat{}ZZ - P\hat{}Z}{P\hat{}Z - P\hat{}Z}$$

$$P\hat{}o = ١ - ٢ (P\hat{}Z __ - \ P\hat{}ZZ \ )$$

ويتطلب تحقيق الفروض السابقة إلى إجراء الخطوات التالية:

- تقدير متوسط درجات الطلاب $M\hat{}$ والانحراف المعياري $a\hat{}$ ومعامل ثبات كودر ريتشاردسون ٢١ ($a\hat{}_{٢١}$) وتحديد درجة القطع (C) وبارمتر توزيع الفا (a).

-تحويل القيم السابقة إلى بيانات معيارية normalize data وفق الصيغ الآتية :

$$M\hat{} = Sin \quad 1\sqrt{M \quad \hat{} \ / \ n}$$

$M\hat{}$ = ترمز إلى المتوسط

N = ترمز إلى عدد المفردات

$$6 = \sqrt{\quad (\alpha\hat{}٢١)\ б = \quad +١) + (\alpha\hat{} + \quad N\ )}$$

`^ б = ترمز إلى الانحراف المعياري

$$P\hat{} = \sqrt{\quad \alpha\hat{}٢١ \quad (n١) / (n + \alpha\hat{}٢١)}$$

^P = ترمز إلى معامل الثبات

$$C\hat{} = \sin^{-1}\sqrt{(c\text{-}0.5)n}$$

^C = ترمز إلى درجة القطع

تقدير الانحراف المعياري ((normal deviate بناءاً على درجة القطع (Z)

Z = (C` – M) / б`^

تقدير قيمة الاحتمال (P^Z) من الجداول الخاصة بذلك بحيث تكون هذه القيمة اقل من قيمة (Z).

ثم تحديد قيمة الاحتمال (P^ZZ) من جداول دالة التوزيع الشائي الاعتدالي بحيث تكون قيمة الاحتمال ايضاً اقل من قيمة (Z).

٢.طرق تتطلب تطبيق الاختبار مرتين:

أ- **معامل كارفر** Carver Method أشارت سابكوفياك Subkoviak إلى أن معامل كارفر يعد من الطرق الأولى المقترحة لتقدير ثبات الاختبارات محكية المرجع التي تتعلق باتساق قرار التصنيف. وتقوم هذه الطريقة على تطبيق اختبارين متوازيين على مجموعة واحدة من الأفراد، يليه مقارنة نسبة عدد الطلاب الذين تم تصنيفهم ضمن المتقنين في كلا الاختبارين، فإذا كانت النسبتين متماثلتين أو قريبتين من بعضهما فبذلك يعتبر الاختبارين ثابتين، ويعاب على هذه الطريقة بأنها غير حساسة لاتساق قرارات كل طالب في الاختبارين. **ويمكن إيجاد قيمة معامل (كارفر) وفق الجدول التالي:**

الاختبار (أ)	الاختبار (ب)		
		غير متقن	متقن
	متقن	ب	أ
	غير متقن	ج	د

ويحسب معامل كارفر باستخدام الصيغة التالية:

أ + ج

= ———————

ن

حيث إن:أ = مجموع الطلاب المتقنين بالفعل.

ج = مجموع الطلاب غير المتقنين بالفعل ن = أ+ ب+ ج+ د

ب ـ معامل (كابا) Swaminathan- Hambleton - Algina ذكر كل من هامبلتون، وآخرون ١٩٧٨ Hambleton et al بأن هذه الطريقة تعتبر امتداداً لطريقة) كارفر(السابقة حيث تعتمد على فرق النسبة بين الطلاب المتمكنين وغير المتمكنين في صورتي الاختبار المتوازيتين أو في مرتي تطبيق الاختبار، وتمتد قيمة معامل الثبات بين (+١، - ١).

ويمكن تقدير معامل الثبات وفق الصيغة التالية

$$Po\hat{} = \sum(P\hat{} KK)$$

حيث إن:

Po^ ترمز إلى معامل الثبات أي إتساق التصنيف.

KK ^P^ ترمز إلى نسبة الطلاب الذين تم تصنيفهم ضمن المتقنين في المستوى K في مرتي تطبيق الاختبار.

M : ترمز إلى عدد مستويات الإتقان. ونظراً لأن هذه الطريقة لا تأخذ بعين الاعتبار إتساق التصنيف الناتج عن عامل الصدفة، لذا أقترح كل من سواميناثان وهامبلتون وألجاينا استخدام معامل كابا(Kappa Coefficient) الذي ينسب إلى كوهن.

$$K = \frac{Po\hat{} - Pc\hat{}}{١ - pc\hat{}}$$

حيث أن:

K: معامل كابا

Po^: نسبة الاتفاق الملاحظ في التصنيفات وتحسب بالصيغة السابقة

$$Po\hat{}=\sum (PKK)$$

pc^ : ترمز إلى نسبة الاتفاق المتوقع في التصنيفات.وتحسب بالصيغة التالية:-

$Po\hat{} = \sum (PK.\ P.\ K)$

حيث إن: (K .P) (.PK) = نسب الطلاب المصنفين ضمن فئة المتقنين في المجموعة. (K) في مرتي تطبيق الاختبار على التوالي.

مسميات وتعريف درجة القطع.Definitions of cut off score

مسميات درجة القطع: تتعدد مسميات درجة قطع الاختبارات بتعدد آراء وخبرات علماء القياس والتقويم التربوي في مجال الاختبارات محكية المرجع، ومن تلك المسميات ما ذكرها هالبن، وآخرون ١٩٨٣ Halpin, et al وهي كما يلي:

- درجة القطع.Cut off Score
- الحد الأدنى من الكفاية.Minimum Competency
- درجة المحك.Criterion Score
- درجة الاجتياز Passing Score .
- مستوى الحد الأدنى من الكفايةLevel of Minimal Competency .

تعريف درجة القطع: نظراً لأهمية مفهوم درجة القطع في القياس محكي المرجع، الأمر الذي أدى إلى تعدد تعريفات علماء القياس التربوي والنفسي لمفهوم درجة القطع، ومن تلك التعريفات ما أشار إليه هامبلتون (Hambleton, ١٩٧٨) حيث عرف درجة القطع بأنها نقطة على متصل درجات الاختبار تستخدم لتصنيف الطلاب إلى فئتين تعكس مستويات الأداء المختلفة بالنسبة لهدف معين أو (الأهداف) المراد قياسها في الاختبار. كما عرّف بيرك Berk درجة القطع بأنها " تلك الدرجة التي تزيد من نسبة التصنيفات الصحيحة أو في المقابل تقلل من نسبة التصنيفات الخاطئة"كما عرّفها بابام Popham بأنها " مقياس لمدى ملائمة أداء الطالب لهدف محدد".

طرق تحديد درجة القطع:

تبنى خبراء القياس والتقويم التربوي العديد من الطرق والمحاولات لتحديد درجة القطع، واختلفت تلك الطرق والأساليب باختلاف خبرات ود رجات تأهيل

obeikan.com

وتخصصات هؤلاء الخبراء، فهنالك عدة طرق متعددة لتحديدها والتي ذكر بيرك Berk ١٩٨٦ أنها بلغت ثمان وثلاثين طريقة مختلفة من حيث أسلوب بنائها وتطبيقها وتفسير نتائجها، ويمكن تصنيفها إلى ثلاثة تصنيفات كالتالي:

أولاً: الطرق التحكيمية Judgmental - methods تستند الطرق التحكيمية على آراء المتخصصين في مجال القياس أو في محتوى الاختبار المراد قياسه، وذلك لتحديد درجة القطع المناسبة، حيث يسترشد بآرائهم في تحديد مستوى الطالب ذي الحد الأدنى من الكفاية (تحديد عدد الأسئلة التي يجب أن يجيب عليها الطالب بصورة صحيحة) واضعين في اعتبارهم المتغيرات التي تؤثر في تحديد هذه الكفاية.

مميزات وعيوب الطرق التحكيمية: لكل طريقة من الطرق التحكيمية مميزات وعيوب خاصة بها وهنالك عيوب ومميزات تكاد تكون مشتركة بينها وهي كالتالي:

مميزات الطرق التحكيمية: كما استرسل (بيرك Berk) بأن من مميزات الطرق التحكيمية مايلي:

١ـ أنها سهلة الاستخدام مقارنه بالطرق الأخرى.

٢ـ أنها سهلة التطبيق والفهم.

٣ـ أنها سهلة التفسير.

٤ـ لا تحتاج إلى جهد ووقت كبيرين.

٥ـ أنها سهلة الحساب ولا تعتمد على أساليب إحصائية معقدة.

عيوب الطرق التحكيمية:

أشار كل من بيرك، وهامبلتون، وايغنور(Berk, ١٩٨٠ Hambleton & Eignor, ١٩٨٦) إلى عيوب استخدام الطرق التحكيمية وهي:

١ـ اختيارية Arbitrary وسميت كذلك لأن هذه الطرق تعتمد كلياً على آراء المحكمين وتلعب كفاية هؤلاء المحكمين دوراً بارزاً في الوصول إلى المعيار المناسب لتصنيف الطلاب وتحديد بنود الاختبار التي يمكن لأقل الطلاب كفاءة الإجابة عليها بشكل صحيح ويعتقد بأن كون هذه الطرق اختيارية لا يعد سبباً

كافياً لرفض هذه الطرق لأن هذه العملية الاختيارية ليست مطلقة ولكنها تعتمد على مجموعة من الأسس في تحديد مستويات الأداء.

٢- صعوبة تحديد الطلاب الأقل كفاءة.

٣- لا تعتمد على بيانات تجريبية فعلية لكي يسترشد بها المحكمون في تقدير أحكامهم.

٤ـ الحصول على درجات قطع مختلفة عند تطبيقها على اختبار موحد وعينة دراسية واحدة.

٥ـ ذاتية المحكمين في اختيار الحد الأدنى من الكفاية وذلك بسبب اختلاف تصوراتهم حول عملية الإتقان.

٦- تفاوت عملية الاتساق الداخلي في عملية التحكيم.

٧- صعوبة التنبؤ بأداء الطلاب في الاختبار.

على الرغم من وجود بعض العيوب في استخدام الطرق التحكيمية إلا أنه توجد بعض المعايير المطروحة في أدبيات القياس التربوي التي تحد من تأثير هذه العيوب حيث ذكر بيرك (١٩٨٦) بعض المعايير نذكر منها ما يلي:

أ - استخدام طرق لتوحيد بيانات الأداء الفعلي.

ب - يجب أن تكون عملية تحكيم مفردات الاختبار متزامنة مع تطبيق الاختبار.

ج - الاهتمام بتدريب المحكمين.

د - تنظيم عملية تحكيم بنود الاختبار.

هـ - إيجاد الحلول المناسبة لتسوية الاختلافات في أحكام المحكمين.

و- تصميم استمارات تنظم عملية التحكيم.

ز- تزويد المحكمين ببيانات مسبقة حول أداء الاختبار.

وتناول كثير من الباحثين الطرق التحكيمية وأسهبوا في ذكر تفصيلاتها وتطبيقاتها ومميزاتها وعيوبها ومن ضمن هؤلاء الباحثين علام ١٩٩٥، وبيرك١٩٨٦، وجايقر ١٩٨٩ ومن هذه الطرق ما يلي:

١. **طريقة إيبل:** Ebel's method اقترح إيبل Ebel طريقة لتحديد درجة القطع تعتمد على تعليقات المحكمين على مفردات الاختبار وذلك من خلال اعتبار أن

مواصفات المفردات لها بعدين هما: بعد الصلة Relevance وبعد الصعوبة difficulty ويشتمل بعد الصعوبة على ثلاثة مستويات (سهلة، متوسطة، صعبة) ويشتمل بعد الصلة على أربعة مستويات (أساسية، مهمة، مقبولة، غير مهمة) وبذلك يستخدم المحكم أعداداً مختلفة من تصنيفات الصعوبة والصلة، ثم يطلب من كل محكم ما يلي:

أ - تحديد الخلية المناسبة لكل مفردة من المفردات بناءً على توافق محتواها مع تصنيفات البعدين.

ب – تحديد النسبة المئوية لعدد الطلاب ممن لديهم الحد الأدنى من الكفاية الذين يجب أن يجيبوا عن كل مفردة بصورة صحيحة وذلك لكل خلية من الخلايا، ثم يضرب عدد مفردات كل خلية في النسبة المئوية، ثم يقسم مجموع نواتج جميع الخلايا على عدد المفردات، وبالتالي نحصل على درجة القطع.

عيوبها: تناول بيرك ١٩٨٦ Berk عيوب هذه الطريقة حيث ذكر بأن أيبل لم يلتزم بالمواصفات التي تستخدم في كل من البعدين والتمييز بينهما، لكي يسترشد بها المحكمين في عملية التحكيم، لذلك ربما يعتمد كل محكم على مواصفات يحددها لنفسه مما يؤثر بلا شك في دقة التصنيف واتساقه، كما أن هذه الطريقة اعتمدت على تحكيم المفردة ككل ولم تأخذ بعين الاعتبار بدائل الإجابات التي تشتمل عليها مفردة الاختيار من متعدد. بالإضافة إلى أن طريقة إيبل تحتاج إلى تحكيم خاص يتمثل بنسب مئوية للمفردات وذلك لكل خلية عبر بعدي الصلة والصعوبة المستمرين، فإذا لم تكن هنالك معايير خارجية يبنى عليها هذا الحكم فإنه يبدو أنه سيكون حكماً اعتباطيا للحد من هذه العيوب قام كل من سكاكون وكلينك Skakun & Kling بإجراء تعديلات على طريقة إيبل بحيث أصبحت تعتمد هذه الطريقة على بعدين هما: معامل الصعوبة ويشتمل على)سهل ـ متوسط الصعوبة ـ صعب جد ا (ومعامل التصنيف ويشتمل على) واقعي ـ شامل ـ حل المشكلات) وتستخدم بيانات الأداء في البنود لتقدير تصنيفات الصعوبة ويحدد كلاً من باني الأسئلة ولجنة الاختبار الأسئلة طبقاً لمستويات التصنيف، بناءً على ذلك تضرب الأسئلة في الخلايا المكونة من المصفوفة ٣ × ٣ ثم يقوم المحكمين

بالإشارة إلى نسبة المفردات في كل خلية والتي يجب أن يجيب عليها الطالب المتقن إجابة صحيحة.

المميزات:

لها نفس مميزات طريقة إيبل ويحدد تصنيف معامل صعوبة البند بالاستعانة ببيانات الأداء، وتضرب في كلا البعدين مما يسهل عمل المحكم.

العيوب:

إن عيوب هذه الطريقة تكمن في عدم وضع عاملي الصلة والأهمية في الاعتبار ومن المحتمل أيضاً أن يصعب على بعد التصنيف تحديد مناطق المحتوى المحددة للحد من هذه العيوب.

٢ـ طريقة أنجوف: Angoffs method ١٩٧١يطلب من كل محكم في هذه الطريقة فحص كل فقرة من فقرات الاختبار، ثم يطلب منهم تصور مجموعة من الأفراد الذين حققوا الحد الأدنى للكفاية التي يقيسها الاختبار، ثم يقدر نسبة عدد الأفراد الذين يحتمل أن يجيبوا إجابة صحيحة عن كل مفردة من مفرداته، ويمثل متوسط هذه النسب الحد الأدنى لمستوى الاجتياز في الاختبار.

مميزاتها:

أ ـ سهلة التنفيذ. ب ـ سهلة الفهم.

ج ـ أنها تعتمد على تصور عدد من الأفراد الذي حققوا الحد الأدنى للكفاية المطلوبة بدلا من تصور فرد واحد.

عيوبها: يصعب على كثير من المحكمين تعريف الطلاب من ذوي الحد الأدنى للكفاية، ويصعب أيضاً تقدير القيم الاحتمالية لبعض المفردات المصاغة في عبارات منفية (Negatively) أو المتضمنة عمليات حسابية.(Computational)

للحد من هذه العيوب أشار كل من جايقر وكروس وآخرون ;Jaeger, ١٩٨٩ Cross et al.,١٩٨٤ بأن أنجوف Angoof أجرى تعديلات على طريقته وذلك لتسهيل عملية تقديرات قيم المحكمين الاحتمالية لأسئلة الاختبار. حيث يطلب من كل

محكم تقدير احتمال أن يجيب الطلاب ذوي الحد الأدنى للكفاية على كل سؤال دون أن يلجأوا إلى تخمين (Guessing) الإجابة، وتسجل تقديراتهم الاحتمالية في استمارة خاصة تحتوي على مقياس عشري الفترات Equal Interval scale

١= صفر – ٠,١٠ ٢=٠,١١- ٠,٢٠ ٣=٠,٢١- ٠,٣٠ ٤=٠,٣١- ٠,٤٠

٥=٠,٤١- ٠,٥٠ ٦=٠,٥١- ٠,٦٠ ٧=٠,٦١- ٠,٧٠ ٨=٠,٧١- ٠,٨٠

٩=٠,٨١- ٠,٩٠ ١٠=٠,٩١- ١,٠٠

ثم يطلب من كل محكم اختيار القيمة الاحتمالية التي تمثل تقديره الخاص بكل سؤال من أسئلة الاختبار، ثم تجمع القيمة الاحتمالية التي قدرها كل محكم لكل مفردة، ثم نوجد متوسطها لتمثل درجة القطع.

٢ـ طريقة ندلسكاي Nedlesky's Method: اقترح نيدلسكي Nedelsky ١٩٥٤هذه الطريقة لتحديد درجة القطع للاختبارات التي تشتمل على أسئلة ذات اختيار من متعدد حيث تقوم مجموعة من المحكمين المختصين بتحديد درجة القطع المناسبة للاختبار، وفي هذه الطر يقة يطلب من كل محكم فحص كل مفردة من مفردات الاختبار، ثم يحدد من بين الخيارات الخاصة لكل مفردة تلك التي يمكن أن يتجنب الطلاب ذوي الحد الأدنى للكفاية اختيارها لأنها لا تمثل في نظرهم الإجابة الصحيحة للمفردة، بعد ذلك يضع المحكم درجة للسؤال بناءً على ما تبقى من الخيارات، وذلك بأخذ مقلوب الخيارات المتبقية.

فمثلاً إذا كانت المفردة تشتمل على خمسة خيارات ورأى المحكم أن الطالب ذو الحد الأدنى للكفاية يمكن أن يتجنب اختيار ثلاثة منها عندها تكون درجة السؤال عبارة عن:

$$= \frac{١}{\text{٣-٥}} = \frac{١}{٢} = ٠,٥$$

بعد ذلك يقوم المحكم بجمع درجات الأسئلة، ويكون المجموع الكلي هو درجة القطع بالنسبة لذلك المحكم، وتجمع في النهاية الدرجات التي يتحصل عليها المحكمون، ثم يحسب المتوسط وبالتالي نحصل على درجة القطع للاختبار ككل، غير أن ندلسكاي وجد أنه من الأفضل إيجاد الانحراف المعياري (σ)

لدرجات القطع ثم ضربها في مقدار ثابت (K) يتفق عليه المحكمون وتعتمد قيمة هذا المقدار على النسبة المئوية التي يرى المحكمون أنها مطلوبة لنجاح الطلاب من ذوي الحد الأدنى للكفاية. ويتم تحديد درجة القطع في هذه الطريقة وفق المعادلة التالية:

$$MPL = \bar{M}FD + K\,\sigma FD$$

$$\bar{M}FD = \frac{\Sigma\ FD}{N}$$

$$\sigma FD = \sqrt{\Sigma\ \frac{(MFD - \bar{M}FD)}{N}}$$

حيث إن:

MPL: تمثل درجة القطع.

MFD : تمثل متوسط الحد الأدنى المحدد من المحكمين.

K: تمثل المقدار الثابت المحدد من المحكمين.

σ F D: تمثل الانحراف المعياري لدرجات القطع

N: تمثل عدد المحكمين.

وأشار بيرك ١٩٨٦ ,Berk **إلى بعض عيوب هذه الطريقة وهي كالتالي:** لا يوجد تبرير منطقي يؤيد افتراض (ندلسكاي) أن الطالب الذي لا يعرف الإجابة الصحيحة لسؤال معين يلجأ إلى التخمين العشوائي لأن مشتتات الأسئلة تصمم عادةً لجذب غير المتمكنين، كما أن هذه الطريقة تسمح فقط بقيم احتمالية منفصلة تعتمد على عدد خيارات الإجابات، لذلك يضطر المحكمون لتعيين القيمة (٠,٥) لمعظم المفردات، هذا بالإضافة إلى أن هذه الطريقة تحتاج لقدر كبير من التدريب لكي يستخدمها المحكمون استخداماً مناسباً.

٤ ـ طريقة انجوف ـ ندلسكاي المختلطة: ١٩٨٤ (Angoff Nedelsky ombination Reid), في هذه الطريقة يتم حساب المعيار على أساس حساب متوسط المعايير الناتجة عن استخدام طريقتي أنجوف وندلسكاي.

المميزات: لها نفس مميزات طريقة أنجوف.

العيوب: لها نفس عيوب طريقتي أنجوف وندلسكاي بالإضافة إلى احتمالية زيادة مصادر الخطأ المختلفة وذلك عندما نحسب متوسط المعيارين.

٥ ـ طريقة تقدير أهمية الصعوبة Difficulty importance estimate في هذه الطريقة يقوم المحكمون بفحص كل هدف، وذلك لتقدير الحد الأدنى للنجاح ويجب أن تتفق عينة المحكمين على الحد الأدنى وعلى كل هدف أيضاً، وذلك لأن للحد الأدنى وزن خاص طبقاً لأهمية صلته بالهدف المنشود، ثم تضرب النسبة المئوية لدرجة النجاح في نسبة أهمية الدرجة لكل هدف، والناتج يجمع من خلال جميع الأهداف، ويضرب في عدد بنود الاختبار وذلك للوصول للمعيار، ويجب أن يضع الحكام في الاعتبار الخطأ المعياري وذلك عند اتخاذهم القرارات الخاصة بدرجة القطع.

المميزات: أ ـ سهلة الفهم، ب ـ سهلة الاستخدام وتعمل على ربط درجة القطع بقائمة خصائص النطاق السلوكي.

العيوب: استخدمت تقديرات المحكمين للأهداف لتقدير معيار بنود الاختبار، بينما نجد أنه من المفروض أن تكون العلاقة بين الأسئلة والأهداف مباشرة، وليس بالضرورة افتراض أن تقديرات الأسئلة تساوي تقديرات الأهداف، ومن المحتمل أن يكون نوع الخطأ المعياري المستخدم غير ملائم لتفسير درجة القطع.

٦ ــ طريقة مواصفات المفردة (ميلز و بار ١٩٨٣م): في هذه الطريقة يقوم المحكمون بقراءة مواصفات كل مفردة والتي تحتوي على المهارة المراد قياسها، ووصف محتوى البند المقبول، وخواص المثير، والقوانين التي تحكم إحداث خيارات الاستجابة وعينة المفردات، ومن ثم فإن على المحكمين إدراك وتصور الإطار الكلي للمفردات والمتمثل في المواصفات والتنبؤ بنسبة البنود في الإطار الكلي، والتي يمكن أن يجيب عليها الممتحن من الحد الأدنى من الكفاية إجابة صحيحة.

المميزات: سهلة الفهم، وسهلة الاستخدام، وتقوم بربط درجة القطع بمواصفات النطاق السلوكي، وربطهما بكافة البنود التي تقيس النطاق السلوكي، وتستغرق زمناً أقل مما لو عمدنا إلى تقدير كل مفردة بشكل منفصل، وتكون درجة القطع فيها مرتبطة بالنطاق السلوكي وليس بصيغة الاختيار المحدد.

العيوب: تستخدم تقديرات المحكمين لمواصفات البنود لتقدير درجة القطع لمفردات الاختبار، بينما العلاقة بين المفردات ومواصفات البنود مباشر ة، لذا ليس من الضروري أن تكون تقديرات درجة القطع مماثلة لتقديرات مواصفات البنود.

ثانياً: الطرق التي تعتمد جزئياً على التحكيم وتسترشد ببيانات تجريبية:

أشار علام ١٩٩٥م، بأن الباحثين أدركوا قصورا ملحوظا في الطرق التحكيمية ومن ذلك القصور اعتمادها على تحكيم مفردات الاختبار دون النظر إلى الأداء الفعلي للمختبرين مما يؤدي ذلك إلى الحصول على معايير غير متميزة بدرجة عالية من الصدق.

وأشار كل من سكون وجاليون وفيرار ا بأن درجة القطع المحددة باستخدام بعض من الطرق التحكيمية تؤدي إلى رسوب أكثر من نصف المتقدمين للحصول على ترخيص مهنة الطب في أمريكا، أما في حالة استخدام الطرق التجريبية فإنه يتبين خطأ تلك الدرجات حيث يجتاز معظم الراسبين الحد الأدنى لمستوى النجاح. لذا اقترح علماء القياس العديد من الطرق المختلفة لتحديد مستويات الأداء التي تتضمن طرقا تجريبية محاولين في ذلك تجاوز عيوب الطرق التحكيمية والحصول على مستويات أداء دقيقة يمكن الاعتماد عليها ومن هذه الطرق مايلي:

١ـ الطريقة التحكيمية المعززة بالمعلومات Informed Judgement method (popham , ١٩٨١)يتضح من مسمي هذه الطريقة أنها تعتمد على بيانات مسبقة عن أداء الطلاب حيث تفيد هذه البيانات في إعطاء دلائل للمحكمين عن المستوى العام للطلاب، وبالتالي توضع هذه البيانات في اعتبار المحكمين لكي يسترشدوا بها أثناء تحديدهم للمعايير وفي هذا الشأن ذكر بيرك (١٩٨٦) بأن هذه الطريقة تعتمد على بيانات تتعلق بالأداء الفعلي للأفراد، وتتضمن هذه البيانات إحصاءات خاصة بالمفردات مثل معاملات الصعوبة والتمييز ومتوسط درجات الاختبار بعد تطبيقه على العينة المناسبة من الأفراد، وبعض تعليقات وأحكام الخبراء حول مفردات الاختبار، واقتراحات المختصين مثل المدرسين والإداريين وأساتذة الكليات وأعضاء مجالس المدارس، وتقدم جميع هذه المعلومات للمحكمين لفحصها وتبادل

المقترحات حولها ومناقشتها وإعادة النظر فيها وذلك لكي يتم التوصل إلى اتفاق حول درجة القطع المناسبة.

المميزات: سهلة التطبيق و سهلة الفهم و يسهل حسابها، كما أنها تجمع وتوحد بيانات الأداء والبيانات المتميزة من مدى واسع بواسطة المجموعات المختصة بتلك البيانات.

العيوب: من عيوب هذه الطريقة أنها تتطلب جمع كثير من البيانات التجريبية والمعلومات المتعلقة بالاختبار ومحتواه من مصادر متعددة، وقد يستغرق ذلك وقتاً وجهداً كبيرين، ويبدو أن تحليلات المحكمين للبيانات غير منظمة، كما تبدو الحاجة لوجود خطوط تفصيلية تسهل استخدام أنواع البيانات المختلفة والمناقشة الجماعية للأداء الفردي، والتي لا تبدو مفيدة في هذه الطريقة بسبب تأثيراتها المعيارية.

٢- طريقة أنجوف التجريبية المعدلة Modified Angoff empirical ذكر بيرك أن هذه الطريقة عبارة عن تعديل لطريقة أنجوف التحكيمية، وتتطلب هذه الطريقة المعدلة تزويد المحكمين ببعض البيانات المستمدة من التجريب الميداني لمفردات الاختبار، وذلك لكي يسترشد بها المحكمون في عملية التحكيم، ويطلب من كل محكم تقدير احتمال إجابة الأفراد ذوي الحد الأدنى من الكفاية المطلوبة لمفردات الاختبار ككل وليس من كل مفردة على حده كما في الطريقة السابقة (التحكيمية)، وبعد الحصول على هذه التقديرات يتم تزويد المحكمين ببيانات تجريبية عن مفردات الاختبار تشتمل على متوسط صعوبة كل مفردة، ثم يطلب منهم تعديل تقديراتهم الاحتمالية في ضوء تلك البيانات التجريبية، ويكون الحد الأدنى لمستوى الاجتياز في الاختبار هو متوسط مجموع تقديراتهم المعدلة.

المميزات: سهلة التطبيق، سهلة الفهم ويسهل حسابها وتعمل على ربط المعيار بالكفايات المقاسة، وتدمج معامل صعوبة البند الحقيقي خلال عمليات التحكيم.

العيوب: إن عملية استخدام تقديرات المحكمين للكفايات يجب أن تكون على درجة عالية من التوافق والانسجام، وذلك بين المفردات والكفاية ولا يمكن افتراض أن تكون التقديرات الواقعية متطابقة. وأضاف جايقر Jaeger بأنه أجريت تعديلات في عام ١٩٧٨م على طريقة أنجوف الأساسية وذلك بهدف تبسيط مهمة

المحكمين بحيث يطلب من كل محكم تقييم كل مفردة من مفردات الاختبار بالإجابة (بنعم) أو (لا)، ومن ثم يتم تحديد الحد الأدنى للكفاية وبعد الحصول على هذه التقديرات يستعرض عدد مختلف من المحكمين ثلاثة أنواع من البيانات التجريبية، وذلك خلال ثلاثة لقاءات منفصلة وتشتمل هذه البيانات التجريبية على التوزيع التكراري لدرجات الطلاب في الاختبار والأساليب الإحصائية الوصفية ودرجات القطع المحددة من قبل المحكمين، ثم يطلب من عينات المحكمين المختلفة تعديل تقديراتهم بناءً على تلك البيانات التجريبية، وتصبح درجة القطع عبارة عن متوسط تلك التقديرات التي حددتها عينات المحكمين في اللقاءات الثلاثة المنفصلة.

المميزات: تتيح للمحكمين فرصة لتحسين تقديراتهم بناء على ثلاثة أنواع من البيانات المختلفة، ويحدد درجة القطع عينات مختلفة من المحكمين وتميل هذه الطريقة إلى زيادة ثبات درجة القطع.

العيوب: تحديد الإجابة بنعم أو لا، بالإضافة إلى أن هذه الطريقة تفرض على المحكمين أن يلتقوا ثلاثة لقاءات منفصلة، كما أنها تستغرق كثيراً من الوقت، وتعتبر أكثر تعقيداً من الطرق السابقة، ويعتبر تفسير المعيار النهائي صعب جداً بالنسبة للمحكمين. لتلافي بعض من هذه العيوب قام كل من سنديرس ومابوس بإجراء تعديلات على الطريقة فأصبحت تحتوى على عينة واحدة من المحكمين، وأصبح تحديد المعيار النهائي يتم خلال لقاء واحد يجمع بين المحكمين.

كما قام شيبرد أيضاً بإجراء تعديلات على طريقة أنجوف الأساسية بحيث أصبحت هذه الطريقة تتطلب أربعة أنواع من المعطيات التي يجب أن توضع بعين الاعتبار وهي:

أ- أحكام مطلقة مبنية على فحص المفردات (مثل طريقة أنجوف).

ب- بيانات تجريبية على عينة محكمة صنفت المتقنين وغير المتقنين (مثل طريقة المجموعات المتضادة).

ج- تقديرات صانعي القرار (المحكمين) لمستوى النجاح المقبول.

د- التوفيق بين تقديرات المحكمين المختلفة.

فإذا كانت درجات القطع الممكنة والمختلفة قد نشأت من هذه المعطيات فيجب على محدد المعايير أن يوفق بين هذه الاختلافات.

المميزات: من مميزاتها أنها تضم أربعة أنواع من البيانات التجريبية والتحكيمية وتضع في حسابها معظم المصادر الرئيسة للقيام بعملية تحديد المعايير، وتحتفظ بمميزات طريقتي أنجوف والمجموعات المتضادة.

العيوب: لها نفس عيوب طريقتي أنجوف والمجموعات المتضادة فضلا عن صعوبة التوفيق بين الاختلافات بين المعايير وتعديل المعايير المبنية على معدلات النجاح المقبولة وعدم الثبات، بينما نجد أن (التثليث) لها مميزات معينة وخطوط إرشادية محددة، والتي يمكن أن تكون مقبولة بشكل كبير حتى تصل بشكل منظم إلى المعيار النهائي.

٣ـ طريقة توفق بين الطرق المطلقة والطرق النسبية Absolute Relative (I)(II) Compromise أشار بيرك إلى أن هذه الطريقة تعمل على التوفيق بين أحكام المحكمين المتعلقة بأداء الأفراد في الامتحان والأداء الفعلي المستمد من البيانات التجريبية المبنية من تطبيق الاختبار على العينة، وفي هذه الطريقة يطلب من المحكمين تحديد قيمتين هما:

أ ـ الحد الأدنى للنسبة المئوية لعدد المفردات التي يجب أن يجيب عليها الفرد إجابة صحيحة لكي يعتبر ناجحاً في الاختبار، ويرمز لهذه النسبة بالرمز (K).

ب ـ النسبة المئوية لعدد الأفراد المتوقع اجتيازهم للاختبار ويرمز لهذه النسبة بالرمز (V)، ويمكن التوصل إلى معادلة تربط بين النسبتين K، V، والتي يمكن استخدامها لتعديل تلك القيم التي لا تتفق مع التوزيع الفعلي لدرجات الاختبار، إذ يمكن رسم منحنى للمتوسطات والانحرافات المعيارية لقيم K، V التي حددها المحكمون، وكذلك النسبة المئوية لعدد الأفراد المتوقع نجاحهم ويرمز لها بالرمز Y كدالة تناقصية لدرجات الاختبار ويرمز لها بالرمز X، ثم يرسم منحنى لتحديد النقطة ويرمز لها بالرمز (I) والتي تمثل الحد الأدنى لمستوى النجاح، بحيث توفق بشكل أمثل بين المستويات المطلقة (التحكيم) والمستويات النسبية (الأداء الفعلي للأفراد في الاختبار).

المميزات: من مميزات هذه الطريقة أنها توفق بين الأحكام المتعلقة بمفردات الاختبار وبين الأداء الفعلي للمختبرين أي بين المستويات النسبية.

العيوب: من عيوبها أنها تتطلب من المحكمين أن يوفروا تقديرين وهذا يجعل تنفيذ هذه الطريقة أكثر صعوبة من غيرها سواء في عملية التحكيم أو من الناحية الإحصائية، وكذلك تتطلب جهداً كبيراً من جانب المحكمين، بالإضافة إلى أن تفسير المعيار من قبل المتخصصين قد يكون صعباً جداً. لذلك أجرى هوفستي تعديلات لهذه الطريقة حيث يتطلب من المحكمين تحديد أربع قيم هي:

أ ـ أعلى نسبة مئوية لعدد المفردات التي يمكن أن يجيب عليها الفرد إجابة صحيحة ويرمز لها بالرمز (Kmax) وهي عبارة عن درجة القطع المرضية حتى إذا حصل عليها جميع الطلاب.

ب ـ أدنى نسبة مئوية من عدد المفردات التي يمكن أن يجيب عليها الفرد إجابة صحيحة ويرمز لها بالرمز (Kmin) وهي عبارة عن درجة القطع الدنيا التي لا يمكن لأي طالب أن يحرز أقل منها.

ج ـ أعلى نسبة مئوية مقبولة من الراسبين ويرمز لها بالرمز.(Fmax)

د ـ أدنى نسبة مئوية مقبولة من الراسبين ويرمز لها بالرمز(Fmin).

وباستخدام توزيع درجات الاختبار المتجمع (التراكمي) فإننا نستطيع تحديد العلاقة التجريبية بين (K و F)، ويكون المعيار هو نقطة التقاطع بين النموذج (المطلق) والمنحنى (النسبي).

المميزات:

من مميزات هذه الطريقة أنها توفق بين الأحكام المتعلقة بمفردات الاختبار وبين الأداء الفعلي للمختبرين، أي بين المستويات المطلقة والمستويات النسبية كما تضع في اعتبارها الحد الأعلى للنسبة المئوية للمفردات التي من الممكن أن يجيب عليها الفرد إجابة صحيحة (Kmax)، كما تقوم بتبديل النسبة المئوية من الأفراد المتوقع اجتيازهم للاختبار) (V (بأعلى نسبة مئوية مقبولة من الراسبين Fmax وبأدنى نسبة مئوية مقبولة من الراسبين (Fmin) .

العيوب: من عيوب هذه الطريقة أنها تتطلب جهداً كبيراً من المحكمين، وتتطلب توفير أربعة تقديرات وهذا يجعل تنفيذ هذه الطريقة أكثر صعوبة من غيرها سواء، في عملية التحكيم أو الإحصاء.

ثالثاً: طرق تعتمد على البيانات التجريبية وتسترشد بالتحكيم: Empirical Judgmental:

١ـ طريقة المجموعات المحكية:CRITERION GROUPS أشار كل من بيرك وهامبلتون وآخرون (Hambleton et l,١٩٧٨ ;BERK.١٩٨٦) إلى أن هذه الطريقة تتطلب انتقاء مجموعتين من الطلاب الأولى (متقنة) تلقت التعليم (والثانية غير متقنة) لم تتلق التعليم، ويكون مدى توزيع درجات الطلاب من الدرجة (صفر) إلى الدرجة (ن) حيث (ن) تمثل عدد الأسئلة التي تقيس كل هدف على حده، وتتطلب الطريقة تقسيم الطلاب بناء اً على درجة القطع (س) إلى قسمين (متقنين – غير متقنين)، وبافتراض أن الطلاب في المجموعة المتقنة (التي تلقت التعليم) عبارة عن طلاب متقنين فإنه يمكن تقسيمهم بناء.على درجة القطع إلى فئتين:

الأولى: متقنين بالفعل TM **والثانية:** غير متقنين غير حقيقين (FN).

وبالمثل يتم تصنيف طلاب المجموعة الثانية غير المتقنة (التي لم تتلق التعليم) إلى فئتين الأولى غير متقنين فعلاً (TN) والثانية متقنين غير حقيقين.(FM)

وتعتبر توزيعات درجات الطلاب في المجموعتين المتقنة وغير المتقنة المحددات الأولية(Primary Determinants) لمدى دقة تصنيف الاختبار للطلاب كمتقنين وغير متقنين فعلاً للهدف المراد قياسه. وتحدد قيمة درجة القطع (س) من خلال فحص خلية الاحتمالات لكل درجة قطع محتملة، وتكون درجة القطع المثلى تلك التي تزيد نسبة التصنيفات الصحيحة (TN ـ TM) وفي المقابل تحجم نسبة التصنيفات الخاطئة.(FM-FN) ويتم تحديد درجة الإتقان بشكل أولي من خلال توزيعات تكرارات درجات الطلاب في المجموعتين المتقنة (التي تلقت التعليم) وغير المتقنة (التي لم تتلق التعليم) حيث يصنف الطلاب بناء على نقطة تقاطع التوزيعين إلى فئتين متقنين بالفعل وغير متقنين بالفعل للهدف المقاس. ثم استخدم بيرك (Berk

١٩٧٦) طريقة بديلة تهدف لإيجاد معامل الصدق لكل درجة قطع محتملة (س)، لذا فإن متغيرات التصنيف تعطي قيم ثنائية هي:

أ ـ متغير التصنيف المنبئ The Predictor Classification حيث يقدر للطالب القيمة (١) عندما تكون درجته أعلى من درجة القطع، ويقدر للطالب القيمة (صفر) عندما تكون درجته أدنى من درجة القطع.

ب ـ متغير التصنيف المحك(Criterion Classification) يقدر لكل طالب ضمن مجموعة المتقنين القيمة (١) بينما يقدر لكل طالب ضمن مجموعة غير المتقنين (صفر)، ويحسب معامل الصدق من قيمة معامل (فاي) بين المتغيرين الثنائيين.

يلي ذلك حساب المنفعة والضرر النسبيين لأخطاء التصنيف، ويمكن تمثيل قيمة الضرر للخطأ (FN) بالقيمة (ـ١)وللخطأ (FM) بالقيمة (ـ٢)، في المقابل فإنه يتم حساب الفائدة النسبية لتحديد الإتقان بالفعل (TM) بالقيمة (٢) ولتحديد عدم الإتقان بالفعل (TN) بالقيمة (١).

مميزاتها:

أنها نموذج للصدق التكويني وتعتبر هذه الطريقة ذات حساسية خاصة للخصائص الإحصائية للمفردة وكذلك الصدق البنائي.

عيوبها:

من عيوبها أنها لا تحدد المعيار الحقيقي وفي أفضل الحالات تقوم بتحديد موقع الأداء على متصل الدرجات حيث يتعذر التمييز بين المجموعتين المتقنة وغير المتقنة.

٢ـ طريقة المجموعات المتضادة(LIVINGSTON&ZIEKY,١٩٨٢) ONTRASTING GROUPS ذكر كل من بيرك وجايقر بأن هذه الطريقة مماثلة لحد ما لطريقة المجموعات المحكية وطريقة المجموعة الحدية، ولكنها تختلف في كيفية تحديد درجة القطع في الاختبار، ففي هذه الطريقة يقوم المحكمون المتخصصون بتحديد مجموعتين من الطلاب بحيث تكون الأولى متقنة والأخرى غير متقنة بناءاً على مايلي:

المعلومات المتوفرة حول مستويات مهارات الطلاب.

استناداً على درجات الطلاب وأداءهم داخل الفصل حيث يتم تصنيف الطلاب إلى فئتين الأولى الحاصلون على تقدير (جيد جدا فما فوق) والثانية الطلاب الحاصلون على تقدير (مقبول فمادون).

ثم يطبق الاختبار على كلتا المجموعتين ويرسم منحنيان يمثلان التوزيعين التكراريين لدرجات الطلاب في الاختبار، وتمثل نقطة تقاطع المنحنين درجة قطع الاختبار، هذا الاقتراح يفترض تداخل التوزيعين التكراريين وعدم تطابقهما، ويمكن تعديل هذه الدرجة آخذةً في اعتبارها خطأي التصنيفين، أي يمكن اختيار الدرجة التي تقلل من خطأ التصنيف الموجب FALSE POSITIVE أو خطأ التصنيف السالب.(FALSE NEGATIVE)

عيوبها: لها نفس عيوب طريقة المجموعات المحكية، بالإضافة إلى أنه من الصعب تجاوز الخطأ الناجم من المحك

٣ـ طريقة المجموعة الحدية BORDERLINE GROUP (LIVINGSTON&ZIEKY,١٩٨٢) ذكر كل من بيرك وجايقر (BERK, ١٩٨٦; JAEGER, ١٩٨٩) بأن هذه الطريقة تتطلب من المحكمين انتقاء عينة من الطلاب الذين يرون بأن كفايتهم المراد قياسها تقع عند الحد الفاصل بين التمكن وعدم التمكن، ومن المحتمل أن تكون نسبة الطلاب قليلة بالنسبة للمجموعة الكلية، ثم يطبق الاختبار على أفراد العينة وبعد ذلك يتم إيجاد وسيط درجات هذه المجموعة ليكون بمثابة مستوى الأداء المطلوب.

مميزاتها:

تمتاز هذه الطريقة بأنها سهلة الفهم والتطبيق والإحصاء غير أنها تحتاج إلى عدد كاف من الطلاب وذلك لكي نحصل على تقديرات مستقرة نسبياً.

عيوبها: من عيوبها أنه يصعب على المحكمين أو المدرسين انتقاء المجموعة الحدية، إذ أن تعريف أو تحديد الطالب الذي يقع عند الحد الفاصل بين الإتقان وعدم الإتقان أصعب بكثير من تحديد الفرد المتقن أو غير المتقن، كما تتدخل بعض المتغيرات في عملية انتقاء أفراد هذه المجموعة وقد يصعب تحديد الحجم المناسب للعينة.

٤ـ طريقة المحك معياري المرجع: NORM REFENCEED CRITERION ذكر بيرك (BERK)١٩٨٦بأنه في هذه الطريقة يتم اختيار المعيار بناءاًعلى العلاقة بين درجات القطع في الاختبار محكي المرجع وبين درجات الأداء في الاختبار معياري المرجع(NRT) ويتم اختيار درجة القطع أولاً في الاختبار معياري المرجع بطريقة اختيارية على سبيل المثال: مناظرة الدرجة الخام بالدرجة المئينية ٥٠أو٧٥ ثم تختار درجة القطع في الاختبار محكي المرجع على أساس أنها تزيد مايلى:

أ ـ نسبة ثبات القرارات الناتجة عن تطبيق الاختبار معياري المرجع ويرمز لها بالرمز (P)

ب ـ نسبة ثبات القرارات المصححة من عامل الصدفة ويرمز لها بالرمز(K)

ج ـ تعيين دالة الفقد (Loss Function) ويرمز له بالرمز (L)

المميزات: من مميزاتها أنها سهلة التطبيق والفهم والإحصاء ويمكن توفير بيانات محك الاختبار معياري المرجع بسهولة ويسر، بالإضافة إلى أن هذه الطريقة تضع في اعتبارها دالة الفقد(LOSS FUNCTION) والعلاقة المباشرة بين الاختبارين.

العيوب:إن استخدام درجات القطع الاختيارية في الاختبار معياري المرجع كمحك يعد خطأً فنياً وغير منطقي.

٥ـ طريقة النتائج التربوية EDUCATIONAL CONSEQUENCES (BLOCK,١٩٧٢)

ذكر بيرك (١٩٨٦) BERK بأن هذه الطريقة تعتمد أساساً على اختيار المعيار من خلال العلاقة بين درجات الاختبار ومحك التعلم المستقبلي نتائج انفعالية أو معرفية حيث يتوقع أن يزداد المنحنى طردياً ليربط بين المتغيرين، وتحدد درجة القطع وفقاً على أثرها في زيادة أداء المحك.

المميزات: سهلة التطبيق والفهم والحساب.

العيوب: إن معيار (التعلم المستقبلي) يحتاج إلى تحديد واضح ولا تبدو الافتراضات في هذا النموذج متطابقة مع البيانات الواقعية، بالإضافة إلى أن العلاقة

بين المتغيرات المقترحة لا توفر دالة درجيةً (STEP FUNCTION) أو خاصية فاعلة في المحك لتحديد درجة القطع.

٦ـ طريقة إيمريك لتقييم اختبار التمكن Emrick's mastery Testing Evaluation ذكر إيمريك ١٩٧١ EMRICK بأن هذه الطريقة تناسب الاختبارات الفرعية القصيرة التي تتكون من خمس فقرات أو أقل، وتعتمد هذه الطريقة على نظرية القرار، إذ يتم من خلالها تحليل ثلاثة من أخطاء القرار وهي الخطأ الأول(α) والخطأ الثاني (β) ونسبة الفاقد لأخطاء القرارين الأول والثاني (RR) لذلك يجب تحديد الخطأ الأول (α) والخطأ الثاني (β) ونسبة الفاقد وطول الاختبار، ويتم حساب نسبة الفاقد من خلال تقييم الأنواع المختلفة لأخطاء القرار، ومن ثم تسجيل مجموع المخاطرات بتحديد درجة القطع باستخدام المعادلة التالية:

$$K = \text{Log}(\beta / 1 - \alpha) + 1/n(\text{Log RR}) \div \text{Log}(\alpha \beta) / (1- \alpha)(1- \beta)$$

حيث أن: k: ترمز إلى درجة القطع ، α: ترمز إلى الخطأ الأول ، β: ترمز إلى الخطأ الثاني، RR: ترمز إلى نسبة الفاقد للخطأين الأول والثاني.، N: ترمز إلى عدد الأسئلة.

٧ـ طريقة نوفيك ومعاونيه: THE WORK OF NOVICK& COLLABORATORS ذكر(١٩٧٦ MESKAUSKAS) بأن هذه الطريقة عبارة عن مزيج من الطرق المختلفة، وتهدف إلى تحديد كل من درجة القطع والتوزيع الاحتمالي للأ داء و الفواقد الناتجة عن الخطأ الأول (α) والخطأ الثاني (β) ويفيد تحديد هذه المتغيرات في زيادة دقة القرارات المتخذة، حيث، يتم وضعها في شكل توزيع احتمالي، ويضاف إلى هذا التوزيع درجات الاختبار بحيث ينتج عنه توزيع احتمالي أكثر حساسية، ويتم تحديد القيمة المقدرة للمستوى الحقيقي للطالب المتقن عن طريق تحديد عينات من المفردات، ثم يضرب الاحتمال الذي تزيد فيه الدرجة الملاحظة على التحصيل المعطى، ثم تقارن النتيجة بالدالة المشابهة التي تشير إلى الاحتمال الذي لا تزيد فيه الدرجة الملاحظة.

٨ ـ طريقة رودابش للدرجة الحقيقية الثنائية ROUDABUSH'S DICOTOMOUS TRUE SCORE ذكر (١٩٧٦) MESKAUSKAS بأن طريقة رودابش تستخدم مع

الاختبارات ذات الفقرة الواحدة، وتشتمل هذه الطريقة على نموذجين من النماذج القطعية وكلا النموذجين يشتملان على مقياس ثنائي للدرجة الحقيقية، وتقوم هذه الطريقة على تحديد أربعة من الأخطاء المحتملة وهي:

(α2 α ١) احتمال أن يكون الفرد المتقن ضمن غير المتقن في المحك و(β2 β1) احتمال ان يكون الفرد المتقن ضمن غير المتقن في المحك. ويتم تحديد الاخطاء الاربعة وفق العلاقات التالية:

$$a_1 = P(x \geq Xc / T = o) \qquad \alpha_1 = P(x \geq Xc / T = \alpha_2 = o)$$

$$\beta_1 = P(x \geq Xc / T = o) \qquad \beta_2 = P(x \geq Xc / T = o)$$

حيث إن: T: الدرجة الحقيقية X: الدرجة الناتجة Xc: درجة القطع

وتستخدم العلاقات التالية لتحديد درجة القطع وهي كالتالي:

$$F_{oo} = N_0 (1 - \alpha_1)(1 - \alpha_2) + N_1 \beta_1 \beta_2$$

$$F_{o1} = N_0 (1 - \alpha_1) \alpha_2 + N_1 \beta_1 (1 - \beta_2)$$

$$F_{10} = N_0 \alpha_1 (1 - \alpha_2) + N_1 \beta_2 (1 - \beta_1)$$

$$F_{11} = N_0 \alpha_1 \alpha_2 + N_1 (1 - \beta_2)(1 - \beta_1)$$

٩ـ طريقة ملمان ذات القرار المبني على الحدين MILLMAN'S BINOMIAL DECISION BASED ذكر كل من هامبلتون، وآخرون (Hambleton et al) ١٩٧٨ وميسكاوسكاس بأن ملمان طور مجموعة من الجداول التي تعتمد على التوزيع ذي الحدين، والذي يشير إلى كل من تقديرات الخطأ المتوقعة لمجموعة متعددة من الدرجات الحقيقية وطول الاختبار ودرجة القطع، وهذه الطريقة تقوم على افتراض أن الاختبار يتكون من مجموعة عشوائية من الدرجات تتراوح مابين (١ـ صفر)، ويستخدم التوزيع ذو الحدين لإيجاد البيانات حول قيم محددة مفترضة للدرجة الحقيقية في كل من أطوال الاختبار والاختبارات الكبيرة والتوزيع الهندسي الفوقي أو ذي الحدين، فيتم التنبؤ بالتكرارات المتصلة، ثم يرسم التكرار المتصل للنسب المئوية المختلفة من الأسئلة التي تمت الإجابة عليها بشكل صحيح.

١٠ـ طريقة ديفز ودايموند THE DAVIS DIAMOND BAYESIAN Method أشار ميسكاوسكس ١٩٧٦ MESKAUSKAS إلى أن هذه الطريقة تعتمد على نموذج بييز، وقد حدد كل من ديفز ودايموند جدولاً يضع تقديرات تعتمد على الدرجة الملاحظة واحتمال أن يكون المستوى الحقيقي لكفاءة الطالب بالقرب أو أعلى من المستويات المختارة.

بالنسبة للاختبار (ذو خمس فقرات) فإن احتمال أن يكون الطالب قد أحرز خمس درجات وله مستوى كفاءة حقيقي يساوي (٠,٩٩) أو أعلى من ذلك يساوي (٠,١١٣٥) وهو رقم أدنى، ويعرض الجدول درجات القطع الضرورية في مستويين من التصنيفات الحقيقية هما (٠,٨٥) ودرجة القطع للاختبارات حتى عشرون فقرة. فإذا أراد صانع القرار أن يقتطع نسبة ٨٥ ٪ من التصنيفات الحقيقية للطلاب مع وجود مستويات كفاية عند نسبة ٩٠٪ أو أعلى فيجب أن تكون هنالك مجموعة متساوية لعدد المفردات في الاختبار، ولتحقيق هذا الهدف يجب استخدام اختبار تتراوح فقراته مابين ١٢ إلى ٢٠ فقرة.

أسس تقويم طرق تحديد درجة القطع: ذكر بيرك (Berk, ١٩٨٦) بأنه يوجد معيارين لتقويم طرق تحديد درجة القطع وهما كما يلي:

١ ـ الكفاية الفنية: Technical - Adequacy

أ ــ أن تحدث الطريقة تصنيفاً مناسبا للمعلوماتً بمعنى أن تحدث الطريقة تصنيفات ثنائية مثل الإتقان / عدم الإتقان، الفشل / النجاح، ويعتبر التصنيف أساسياً لمعظم القرارات المتخذة .

ب ـ أن تكون حساسة لأداء المختبر يجب أن تكون الطريقة حساسة لمستويات صعوبة الأسئلة المختلفة ولأي خصائص أخرى مثل معامل التمييز.

ج ـ أن تكون الطريقة حساسة لكل من المتعلم أو لعملية التدريب

يجب أن تراعي الطريقة عملية التعلم والتدريب الفعلي الذي تلقاه الطالب بحيث تقاس المهارات والكفايات التي تلقاها الطالب بالفعل.

د ـ أن تتضمن الطريقة طرقاً إحصائية مناسبة يجب استخدام الطرق الإحصائية المناسبة، ويجب أن تفسر بشكل صحيح.

هـ ـ أن تحدد الطريقة المعيار الصحيح.بمعنى أنه يجب أن تراعى الطريقة أخطاء القياس.

و ـ أن تثبت الطريقة صدق القرار

بمعنى أنه يجب الحصول من الطريقة تقدير احتمالات قرارات التصنيفات الصحيحة والخاطئة.

٢ـ الجانب التطبيقي (العملي): Practicability

أ ـ أن تكون الطريقة سهلة التطبيق: Easy to implement يجب أن تكون الخطوات المستخدمة في تطبيق الطريقة واضحة وسهلة الفهم للمشاركين فيها، وأن تتم في قدر معقول من الوقت.

ب ـ أن تحسب الطريقة بسهولة: Easy to Compute يجب أن تكون الطرق الإحصائية المستخدمة للحصول على المعيار النهائي سهلة الاستخدام سواء باستخدام الآلة الحاسبة أو البرامج الإحصائية المتوفرة في الحاسب الآلي الشخصي أو المركزي.

ج ـ أن تكون قابلة للتفسير بسهولة للعامة Easy to interpet to lay people يجب أن تكون الطريقة المستخدمة في تحديد درجة القطع قابلة للتفسير وقابلة للفهم، ويجب أن تكون تفسيرات الطريقة واضحة وسهلة الإدراك والفهم للعامة والمهتمين على حد سواء.

د- أن تكون الطريقة جديرة بثقة العامة Credible to lay people يجب أن تكون الطريقة مقنعة وجديرة بثقة العامة.كما أضاف بيرك (Berk, ١٩٨٢) بأن هناك عدة عوامل تساعد على اختيار الطريقة المناسبة وهي:

أ ـ مدى أهمية القرارات المبنية على تحديد درجة القطع.

ب ـ الكمية المتاحة من الوقت لتحديد درجة القطع.

ح ـ مدى توفر المصادر (المادية - الطاقة البشرية) لإنجاز مهمة تحديد درجة القطع.

د ـ كفاية المحكمين) فبعض الطرق تتطلب دراية أكثر بالمحتوى المراد قياسه وبمستوى الطلاب التحصيلي.

هـ ـ مدى مناسبة طريقة تحديد درجة القطع لنوعية الاختبار المراد قياسه.

(الاختبارات التشخيصية)

ومن أنواع الاختبارات المحكية الاختبارات التشخيصية التي تهدف للتحقق من اكتساب المتعلم كفايات أو مهارات أساسية تعبر عن نواتج تعليمية محدودة ومحددة، وتشخيص الصعوبات التي تصادفه أثناء تعلمه أو تدربه، والتعرف على مصادر الاخطاء سواء كانت ناجمة عن سو الفهم أو عن عدم التمكن من

الاجراءات أو العمليات التي تنطوي عليها هذه الكفايات أو المهارات. وبذلك تساعد المعلم في تصميم أساليب تعليمية علاجية مناسبة تيسر على المتعلم تصحيح هذه الأخطاء ومتابعة التعلم من أجل تحقيق الكفايات أو المهارات المرجوة

خطوات بناء الاختبارات التشخيصية:

الخطوة الأولى: تحديد الكفايات أو المهارات الرئيسة المرجوة ومحتواها:

فكل برنامج تعليمي أو تدريبي يهدف لتحقيق مجموعة من الكفايات أو المهارات الرئيسة لدى المتعلمين. ونقصد بالكفاية مجموعة متكاملة من المعارف والمهارات الوظيفية المحددة تحديداً دقيقاً والمتعلقة بمجال تعليمي أو تدريبي معين بحيث يمكن تحقيقها وقياسها من خلال البرنامج. ويمكن إجراء ذلك عن طريق:

١ ــ الاستعانة بمجموعة من خبراء المادة الدراسية الذين يقومون بتحديد الكفايات أو المهارات الرئيسة التي يرون أهمية تحققها لدى المتعلمين.

٢ ـ التحليل المتعمق لمحتوى المنهج الدراسي أو المجال التدريبي.

٣ ـ إجراء دراسة لتقدير احتياجات الفئة المستهدفة.

٤ ــ في المجال التدريبي يمكن التوصل إلى المهارات والكفايات اللازم إتقانها في مهنة معينة عن طريق تحليل العمل.وتتطلب تحديد الكفايات اعتبارات يمكن إيجازها:

١ ـ مدى اتساع الكفاية.٢ ـ إمكانية تعليم الكفاية.٣ ـ قابلية انتقال الأثر.

٤ ـ تمثيل الكفايات للسلوك الختامي المستهدف.

إن تحديد الكفايات أو المهارات الرئيسة في الخطوة الأولى ليس كافيا لبناء الاختبار التشخيصي. فالمهارات الرئيسة تعد بمثابة نواتج مركبة وتتضمن معارف وعمليات عقلية ونفسية حركية. لذلك فهي تتطلب تحليلا يرتب مكوناتها ترتيبا منطقيا يكشف عن العلاقات القائمة بينها.

أما الخطوة الثانية: تحليل الكفايات أو المهارات الرئيسة إلى مكوناتها:

وهناك طرق عدة في تحليل الكفايات من أهمها:

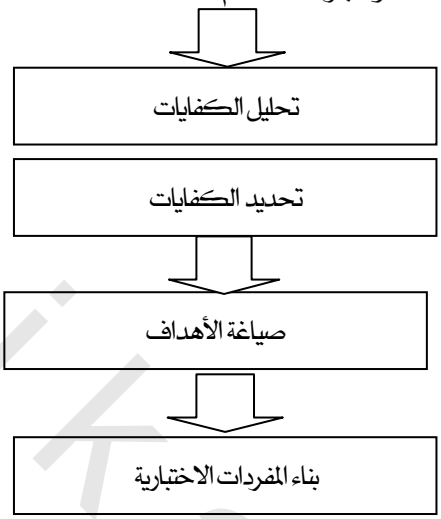
طريقة التحليل الهرمي للكفاية: تهتم هذه الطريقة بتحديد المعارف والمهارات المساعدة التي ينبغي تعلمها بترتيب وتتابع بنائي هرمي لكي تتحقق الكفاية أو

المهارة الرئيسة المطلوبة. ولإجراء التحليل البنائى الهرمي لكفاية أو مهارة رئيسة معينة نبدأ بسؤال أنفسنا: ما المتطلبات أو المكونات السلوكية اللازم توافرها لدى المتعلم لكي يحقق الكفاية أو المهارة الرئيسة ؟ وبعد تحديد هذه المتطلبات نكرر السؤال نفسه لكل متطلب منها، وبالتالي نحدد متطلبات كل من هذه المكونات السلوكية، وهكذا حتى نصل إلى السلوك المدخلي للمتعلم أي السلوك الذي يكون قد سبق تعلمه، واكتسبه المتعلم بالفعل. ويتم ترتيب هذه المتطلبات أي المعارف والمهارات المساعدة ترتيبا هرميا بنائيا بحسب أولويات إسهاماتها في تكوين المهارة الرئيسة، حيث يمثل السلوك المدخلي قاعدة الهرم، ثم تتدرج مستويات السلوك من الأبسط إلى الأكثر صعوبة حتى نصل إلى قمة الهرم الذي تمثله الكفاية أو المهارة الرئيسة.غير أنه من الضروري قبل البدء في إجراء هذا التحليل أن يكون الباحث أو المعلم على دراية بأنماط التعلم الرئيسة ومستوياته المتدرجة ونواتجه لكي يكون التحليل متكاملاً (أنظر الشكل رقم ١)

الخطوة الثالثة: صياغة الأهداف السلوكية: بعد الانتهاء من عملية تحليل الكفايات أو المهارات الرئيسة للبرنامج التعليمي أو التدريبي، فإن الخطوة الثالثة هي صياغة الأهداف السلوكية المتعلقة بنواتج التحليل. فقد اتضح لنا أن هذه النواتج تتضمن المعارف والمهارات المساعدة التي تسهم في تحققها أو اكتسابها إذا استخدمت طريقة التحليل البنائي الهرمي، بحيث يتم إعادة صياغة هذه النواتج صياغة إجرائية أو سلوكية يمكن ملاحظتها ملاحظة مباشرة وقياسها بمفردات اختباريه. فالاختبار التشخيصي مرجعي الهدف يعتمد في بنائه على الأهداف السلوكية التي تحدد نواتج التحليل التعليمي تحديداً واضحاً، وتعبر عن هذا الأداء تعبيراً دقيقاً لا يحتمل الجدل أو التأويل.لذلك يجب العناية بصياغة الأهداف السلوكية المتعلقة بإجراءات أو مكونات الكفايات أو المهارات الرئيسة التي تم تحليلها. وتشمل عادة عبارات الأهداف السلوكية على أربع عناصر أساسية هي: (١) وصف السلوك المتوقع، (٢) المحتوى المرجعي ، (٣) شروط الأداء ، (٤) مستوى الأداء

الخطوة الرابعة: بناء المفردات الاختبارية :وتتطلب هذه الخطوة دراية تامة من جانب الباحث أو المعلم بكيفية انتقاء أنسب أنواع المفردات التي تقيس الأهداف

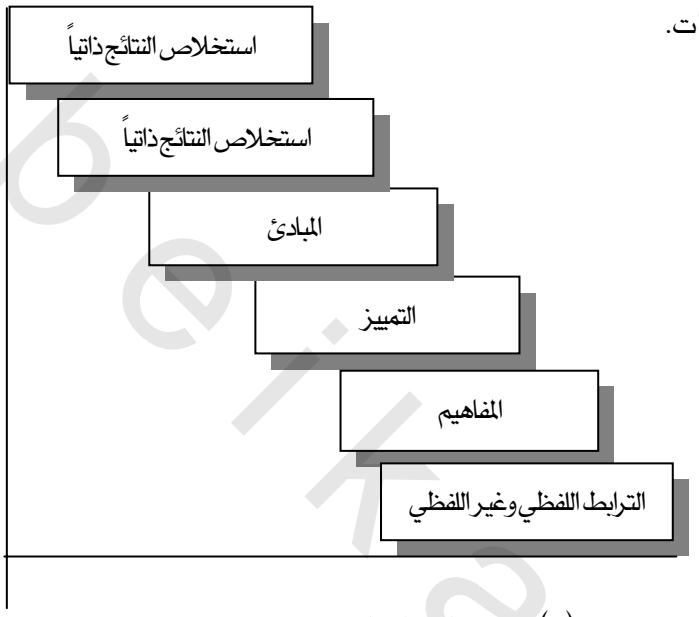
السلوكية المحددة قياساً مباشراً، كما تتطلب التمكن من محتوى البرنامج التعليمي أو التدريبي المعين وفهم خصائص المتعلمين. وذلك لأن هذه المفردات تستخدم في التمييز بين الذين استطاعوا تحقيق الأهداف المحددة والذين واجهتهم صعوبات، وتشخيص أخطاء وفجوات التعلم.



شكل رقم (١) خطوات الاختبارات التحصيلية

أنماط التعلم ومستوياته: يصنف جانييه الإمكانيات العقلية في ثلاثة أنماط رئيسة يعتمد كل منها على الآخر وهي: المعلومات، المهارات العقلية، الاستراتيجيات المعرفية. فالمتعلم يتعلم المعلومات ويختزنها في ذاكرته، ويمكن أن تتعلق هذه المعلومات بمادة دراسية أو مجال تعليمي أو تدريبي معين، أو تتعلق بمحتوى يتميز باستمرارية توظيفه في الحياة مثل معرفة الحروف والأعداد والحقائق المتعلقة بالإنسان والبيئة. ويعتمد المتعلم على هذه المعلومات في اكتساب المهارات العقلية والاستراتيجيات المعرفية. فالمهارات العقلية تتعلق بكيفية أداء نشاط عقلي معين كالتمييز بين مجموعة من الأعداد أو الرموز أو المفاهيم وتطبيق القواعد وحل المشكلات، أما الاستراتيجيات المعرفية فتعد نوعا من المهارات العقلية المتعلقة بسلوك المتعلم بغض النظر عن محتوى مادة التعلم. فهي " إمكانيات " ذات تنظيم داخلي خاص بالمتعلم ويستخدمها دون عون من الآخرين في توجيه عمليات استقبال المثيرات والتذكر والتفكير والإبداع.وقد حدد جانييه عدة مستويات متدرجة فيما

يتعلق بهذه الأنماط أو الإمكانيات الرئيسة الثلاث للتعلم، وكل من هذه المستويات يعتمد على ما يسبقه في إطار البنية الهرمية.ولكي يتمكن الباحث والمعلم من توظيف هذا التصنيف في إجراء التحليل البنائى الهرمي للكفايات والمهارات الرئيسة سوف نلقى الضوء على كل من هذه المستويات.



شكل (٢) مستويات التعلم

١- الترابط اللفظي وغير اللفظي

وهذا النمط الذي يمثل المستوى الأول للتعلم يقصد به العلاقة الترابطية بين مجموعة من المثيرات والاستجابات في تسلسل معين. وقد يكون هذا النمط لفظياً أو غير لفظي. ويتمثل الترابط اللفظي في تعلم الطفل التتابع اللفظي للأعداد حيث يشتمل هذا التتابع على روابط بين المثيرات والاستجابات، كما يتمثل في تعلم اللغات حيث يقوم المتعلم بالربط التسلسلي للكلمات بهدف تكوين جملة أو مجموعة من الجمل، وكذلك سرد المتعلم تواريخ بعض الأحداث في ترتيبها الزمني أو سرد الشعر.

أما الترابط الحركي فيتمثل في المشي الذي يتطلب حركات متسلسلة ومتآزرة بين الرجل والذراع والخصر، أو في فتح باب الغرفة باستخدام المفتاح، أو في تشغيل جهاز معين.

٢ـ التمييز:

وهذا المستوى الثاني للتعلم يعد من المهارات الأساسية وبخاصة لدى الأطفال حيث تكون هناك استجابات متباينة لمثيرات تختلف عن بعضها في صفة أو أكثر، مثل التمييز بين الأشخاص أو أنواع النباتات أو الألوان أو الكلمات الجديدة وغير ذلك.

٣ـ المفاهيم:

ويتعلق هذا المستوى بتعلم المفاهيم واستخدامها، أى تعلم تصنيف المثيرات تبعا لخصائص مجردة كاللون أو الشكل أو الحجم أو العدد. وعندئذ يستطيع المتعلم أن يستجيب بأسلوب واحد لمجموعة من الأشياء التي تكون فئة أو مجموعة معينة. لذلك يعتمد تعلم المفاهيم على تعلم التمييز، وهذا بدوره يعتمد على تعلم الترابط اللفظي وغير اللفظي بين المثيرات والاستجابات.

٤ـ المبادئ:

ويختلف هذا المستوى من التعلم عن المستوى السابق المتعلق بتعلم المفاهيم في أنه يتعلق بتكوين ترابط تسلسلي بين مفهومين أو أكثر في شكل نمط سلوكي استجابة لمجموعة من المثيرات بينما تعلم المفاهيم يتعلق بالتمييز بين سمات أو خصائص مجموعات من الأشياء أو الأحداث أو الأماكن. لذلك فإن تعلم المبادئ التي تربط بين عدد كبير نسبيا من المفاهيم يتطلب تعلم المبادئ الأبسط أي التي تربط بين مفهومين مثلا. ويعتمد هذا أيضاً على مستويات التعلم السابقة.

٥ـ حل المشكلات:

يعد هذا المستوى من مستويات التعلم امتدادا طبيعيا لتعلم المبادىء. فتعلم حل المشكلات يتضمن اختيار مبادىء معينة من مجموعة كبيرة من المبادئ التي سبق تعلمها من أجل حل مشكلة محددة أو تحقيق هدف معين. وفى مواقف حل المشكلات لا يخبر المتعلم بالمبدأ أو المبادىء التي يستخدمها، بل تقدم له المشكلة وعليه أن يبحث بمفرده عن حل لها. غير أنه يمكن تعلم حل المشكلات من خلال

عملية تعليم كيفية تحديد المشكلة وتحليلها، وكيفية اختيار المبادئ المناسبة التي سبق تعلمها وتطبيقها للتوصل إلى الحل الصحيح للمشكلة.

٦ـ استخلاص النتائج ذاتياً:

ويتطلب هذا المستوى من التعلم اكتشاف العلاقات القائمة بين المفاهيم والمبادئ ذاتياً، أي دون تقديم أي عون. ويعد هذا المستوى أعلى مستويات التعلم حيث يصبح المتعلم معتمداً على نفسه ويفكر تفكيراً مستقلاً. وعلى الرغم من أن الإبداع يمكن أن يتضح في جميع مستويات التعلم، إلا أنة يتمثل بدرجة أكبر في مستوى استخلاص النتائج ذاتياً.

obeikan.com