الفصل الثاني

الإطار النظري

- الاختبارات معيارية المرجع .
- الاختبارات محكية المرجع .
- استخدامات الاختبارات محكية المرجع .
- مقارنة بين الاختبارات محكية المرجع ومعيارية المرجع
- طرق بناء الاختبارات محكية المرجع .
- صدق الاختبارات محكية المرجع .
- ثبات الاختبارات محكية المرجع .
- معامل الحساسية للتعليم .
  - ☜- التحليل المنطقي.
  - ☜- التحليل الإحصائي.
  - ☜- معامل الحساسية للتعليم ومعامل التمييز.
  - ☜- تعريف معامل الحساسية للتعليم.
- طرق حساب معامل الحساسية للتعليم.
  - أولاً- معامل كوكس – فارجاس .
  - ثانياً- معامل برينان–ستوليورو .
  - ثالثاً- معامل برينان.
  - رابعاً- معامل بابام.
  - خامساً- معامل رودابوش.
  - سادساً- معامل كوسكوف- كلاين.
  - سابعاً- معامل هالادينا.
  - ثامناً- معاملات بايزن.
  - تاسعاً- المعامل المقترح.

obeikandl.com

الفصل الثاني

الإطار النظري

❋ مقدمة :

يقـــدم الباحث في هذا الفصل فكرة عن الاختبارات معيارية المرجع من حيث تعريفاتها ، وأوجه القصور التي تعاني منها هذه الاختبارات .

ومـــن ثـــم يعرض للاختبارات محكية المرجع من خلال نشأة هذه الاختبارات وتعريفاتها وأنواعهـــا ، ومن ثم استخداماتها والمقارنة بينها وبين الاختبارات معيارية المرجع ، ثم يبين بعض الطرق المتبعة في بناء الاختبارات محكية المرجع .

ويركـــز الباحث على الصدق وأنواعه ، وعلى الثبات وطرق حسابه من خلال عرضه لبناء الاختبارات محكية المرجع .

وبعد ذلك يقدم لمفهوم معامل الحساسية للتعليم ، ثم يعرفه ويشرح بعد ذلك بعض الطرق المتبعة في حساب معامل الحساسية للتعليم .

❋ الاختبارات معيارية المرجع Norm - Referenced Tests :

تشابهت تعريفات الاختبار معياري المرجع رغم تعدد هذه التعريفات :

- حيث **يعرفه بابام وهيسك** (Popham & Husek , 1969, p2) بأنه اختبار يستخدم لتحديد أداء الفرد بالنسبة لأداء أفراد مجموعته . ويتفق معهما **فارجاس** (Vargas ,1969, p 23910) .

- **تعريف صلاح الدين علام** (١٩٨١، ٦٠) :الاختبار معياري المرجع هو ذلك الاختبار الذي يستخدم لتقدير أداء الفرد بالنسبة للأفراد الآخرين في القدرة التي يقيسها ذلك الاختبار .

- **تعـــريف رجاء أبو علام** (رجاء أبو علام ، قاسم الطرف ،١٩٨٥ أ ، ٤٧) رجاء أبو عـــلام ، قاســـم الطرف ،١٩٨٥ ب ، ص١٤) التقدير المعياري بأنه التقدير الذي يقوم على أســـاس مقارنة درجة الطالب بدرجة زملائه في شعبة ما ، بحيث يتم ترتيب الدرجات ترتيباً تنازلـــياً أو تصاعدياً وبحسب مركز درجة الطالب بين درجات زملائه ، ويكون هذا التحديد قائمـــاً علـــى أساس حساب المتوسط والانحراف المعياري ، أو على أساس الوسيط أو على

obeikandl.com

الميئنيات . ويتفق مع هذه التعريف كل من إبراهيم مبارك (٢٠٠٠،١٢١) كيوبيسزن (Kubiszyn , 2000 , p33) .

- **تعريف كرونلند ولن** (Gronlund & Linn , 1990 , p 14)عرف الاختبار معياري المرجع بأنه اختبار مصمم لقياس الأداء الذي يفسر في ضوء الفروق الفردية لبعض المجموعات المعروفة .

ويلاحظ على جميع التعاريف أنها اتفقت على ضرورة أن يتم تحديد أداء الطالب وتحديد أداء مجموعته ثم مقارنة أداء الطالب بأداء المجموعة، وتفسير درجة الطالب في ضوء درجات مجموعته من خلال المقارنة أو ترتيب الطلاب وذلك بهدف معرفة موقعه .

ويمكن للباحث أن يستخلص التعريف التالي :**الاختبار معياري المرجع هو ذلك الاختبار الذي يستخدم لتقدير أداء الفرد مقارنة بأداء مجموعة مماثلة له ، وذلك بهدف تحديد موقع الفرد أو ترتيبه بالنسبة لأفراد هذه المجموعة .**

أما استخدام هذه الاختبارات :بما أن الاختبار معياري المرجع يركز على ترتيب الأفراد وتوضيح الفروق بينهم فإن هذه الخاصية لها استخدامات هامة .

حيث تتطلب بعض القرارات تشتتاً أكبر قدر الإمكان بين درجات الأفراد لتسهيل التمايز والترتيب فيما بينهم ، ففي قرارات القبول لبعض البرامج الدراسية أو الاختيار لبعض الوظائف من المهم أن تتوفر أداة تمكن من ترتيب المرشحين للمساعدة في اختيارهم (إبراهيم مبارك ، ٢٠٠٠ ، ١٢٢-١٢٤) ويمكن استخدامها في شتى الأغراض التعليمية ، ويساعد وجود جماعة معيارية يمكن المقارنة بأدائها على معرفة مستويات التلاميذ الحالية ، ورصد التغير والاختلاف في الأداء ودراسة أسباب التغير والاختلاف من أجل القيام بالتدخل المناسب للتطوير .

وكذلك يمكن استخدام الاختبارات معيارية المرجع في مجال القرارات الانتقائية والإرشادية ، ورغم هذه الاستخدامات وغيرها أيضاً إلا أن هذه الاختبارات تعاني من بعض القصور في النواحي التالية :

١- يكون التأكيد فيها على الدرجة الكلية ، وبذلك يقل الاهتمام فيها على دقائق المحتوى بينما يزيد التركيز على التجريدات الأكثر عمومية . ولهذا فإن الاختبار يحدد الدرجة

الكلية في العمليات الحسابية مثلاً ، ولكنه لا يوضح شيئاً عن نوع العملية الحسابية التي يستطيع الطالب أو لا يستطيع إجراءها ، فالدرجة الكلية يصعب استخدامها في تحديد مواطن الضعف والأخطاء الشائعة أو فجوات التحصيل.

٢- تبنى على فلسفة تؤكد إبراز الفروق الفردية بين الطلاب ، وجعل توزيع درجاتهم في الاختبار يتخذ شكل التوزيع الاعتدالي ، الذي تستند إليه معايير هذه الاختبارات (صلاح الدين علام ، ٢٠٠٠ ، ٧٢٨ ، ٣٤٢) .

٣- أنها تعطي انطباعاً بأن النجاح في الامتحان أو التحصيل الدراسي المرتفع لا يعني أي شيء لأنه ينبغي أن يكون هناك مجموعة مرجعية Reference Group من غير المشاركين في أداء الاختبار. ونظام التعزيز في هذا النوع من تقدير الدرجات هو تقسيم التلاميذ إلى فائزين وخاسرين أو ناجحين وراسبين في سباق التحصيل الدراسي ، وهذا يزيد من التنافس المرضي بين المتعلمين (محمود منسي ، ٢٠٠٣، ١٩١) .

٤- أدى استخدام المقياس جماعي المرجع إلى اختلاف بين نظام الإثابة (أي بمعنى التقديرات Grades) وبين نظام الأداء الفعلي ، بمعنى الإثابة على أساس تحصيلهم بالنسبة لزملائهم وليس على أساس تحصيلهم بالنسبة إلى محك مطلق للأداء (نادية، ١٩٨١ ، ٤٢) .

٥- طرق اختيار المفردة يشوبها الكثير من العيوب فهي تعتمد على عينة الطلاب الذين أجري عليهم الاختبار ، وبذلك تختلف الخصائص للمفردة باختلاف عينة الطلاب . ولا تقدم لنا هذه الطرق أساساً سيكولوجياً لتوضيح ما يفعله الطالب عندما يحاول إجابة مفردة اختبار ، فهذه المعلومات تعد ضرورية إذا أراد مصمم الاختبار أن يتنبأ بخصائص درجة الاختبار في مجتمع معين أو مجتمعات مختلفة من الطلاب (صلاح الدين علام ، ١٩٨٢ أ ، ٣٩) .

٦- مقارنة الفرد بزملائه قد يكون لها انعكاساتها الضارة على تحصيله الدراسي وتقدمه سواءً أكان متفوقاً أو ضعيفاً ، فالطالب المتفوق صاحب القدرة المرتفعة قد يؤدي به احتلاله مركزاً متقدماً في الصف بشكل مستمر إلى الكسل والتقليل من الجهد الذي يبذله في الدراسة ، وكذلك إذا كان الطالب ذا قدرة متوسطة أو منخفضة فإن احتلاله مكانة متدنية بين زملائه بشكل دائم على الرغم مما يبذله من جهد قد يؤدي به إلى

اليأس ، والكف عن بذل مزيد من الجهد (جابر عبد الحميد وآخرون، ١٩٩٧ ،٤٢٥ – ٤٢٦) .

❋ الاختبارات محكية المرجع Criterion-Referenced Tests :

أ- نشأة الاختبارات محكية المرجع والاهتمام بها :

وجدنا أن هناك عدة أسباب جعلت الاختبارات التقليدية معيارية المرجع غير مفيدة بالدرجة الكافية في تحقيق أغراض التقويم التربوي ، ولحسن الحظ فقد ظهرت طريقة جديدة مختلفة للقياس وهي أكثر اتساقاً مع متطلبات التقويم التربوي ، وهذه الطريقة الجديدة يطلق عليها القياس محكي المرجع (جابر عبد الحميد ، ١٩٩٦ ، ٢٢٨) .

وترجع نشأتها إلى رغبة الاختصاصيين في الاختبارات ، في معرفة وسائل جديدة تمكنهم من الوصول إلى معلومات أكثر عن مستوى أداء الأفراد موضع القياس ، تفيد في اتخاذ القرارات . ومن هنا أخذت الفكرة تنتشر بين المشتغلين والمهتمين بالاختبارات والقياس النفسي والتربوي ، واتسع نطاق الاهتمام من مجرد فكرة إلى دعوة لإنشاء أسلوب جديد في القياس التربوي والنفسي ، بما يشمله من دراسة المناهج وما يتعلق بها من جوانب تربوية هامة . (أنور الشرقاوي ، ١٩٩٩٦ ، ١٩ – ٢٠) .

وعلى الرغم من أن ثورنديك (Therondike , 1913) وفلاناجان (Flanagan , 1951) وندلسكي (Nedelesky , 1954) وايبل (Ebel , 1962) ميزوا بين فكرة الأداء مقابل المعيار النسبي في القياس التربوي إلا أن هذا التمييز لم يسهم إسهاماً ملحوظاً في ظهور مصطلح الاختبار محكي المرجع . فكان أول من أوجد هذا المصطلح - Criterion Referenced هو كليزر وكلوس (Claser & Klaus 1963) ، (Berk , 1982 , p3) (Hambleton , 1978 , p277) ، (نادية عبد السلام ، ١٩٩٩٠ ، ٤٥) .

وميز كليزر في مقالته بين الاختبار محكي المرجع (CRT) وبين الاختبار معياري المرجع (NRT) على أساس أداء الفرد ، ففي الاختبار الثاني يفسر بالنسبة إلى أداء الأفراد الآخرين الذين ينتمون إلى جماعة معينة ، أما في الاختبار محكي المرجع فإن تفسير الأداء على الاختبار هو عبارة عن تقرير سلوكي Behavioral Statement أو مجموعة من القرارات السلوكية التي تحدد بدون الرجوع إلى الآخرين ، وبالتالي فإننا في الاختبار محكي

المرجع (CRT) تنسب أداء الفرد على الاختبار إلى محتوى الاختبار ذاته ، وليس إلى أداء الأفراد الآخرين (أنور الشرقاوي ،١٩٩٦ ، ٢٥) .

وأعقب مقالة كليزر (Claser 1963) مقالة لبابام وهاسك Popham & Husek في دورية القياس التربوي عام ١٩٦٣ ، التي حفزت المشتغلين بالقياس على الاهتمام بتطوير أساليب التقويم باستخدام الاختبارات محكية المرجع ، وترتب على ذلك حركة بحث كبيرة في هذا المجال . (صلاح الدين علام ، ٢٠٠٠ ، ٣٠٥) .

ب- تعريف الاختبارات محكية المرجع :

يوجد للاختبارات محكية المرجع تعاريف كثيرة ومتنوعة وهذه التعريفات متشابهة ومتداخلة مع بعضها ، وسوف يقدم الباحث بعض هذه التعاريف :

- **يرى فارجاس** (Vargas , 1969 , p 2391) **فينلي** (Fenley, 1973, p3377) (أن الاختبار محكي المرجع هو الاختبار الذي يقيس أداء الفرد أو المجموعة بالنسبة لمحتوى الأداء المطلق . ولا يكون الاهتمام منصباً على مقارنة الفرد بغيره من الأفراد .
- **ويعرف بابام** (Popham , 1978 , p94) **وصلاح الدين علام** (١٩٨١ ، ٦١ ، ١٩٨٥، ٢٨- ٢٩) الاختبار محكي المرجع بأنه ذلك الاختبار الذي يستخدم لتقدير أداء الفرد بالنسبة إلى محك Criterion أو مستوى أداء مطلق Absolute Standar دون الحاجة إلى موازنة أدائه بأداء الأفراد الآخرين .
- **ويعرف كليزر ونيتكو** Claser & Nitko (Nitko , 1980 , p463) **ونادية عبد السلام** (١٩٩٠، ٤٧) الاختبار محكي المرجع هو الاختبار الذي صمم عن قصد لإعطاء قياسات تفسر مباشرة بالنسبة إلى معايير محددة ، وتحديد هذه المستويات عن طريق تعريف فئة أو نطاق من المهام ينبغي أن يؤديها الفرد . ويتفق معهم جود Good , 1973 , p357) (ويضيف إلى ذلك صلاح أبو ناهية (١٩٩٤ ، ٤٩) أن تفسير الدرجات يكون في ضوء أهداف سلوكية معينة تصف الأداء المتوقع من الطالب في نهاية دراسته لوحدة دراسية معينة.
- **ويعرف رجاء أبو علام وقاسم الصراف** (١٩٨٥(أ) ، ٤٧ ،١٩٨٥(ب) ، ١٤) التقدير محكي المرجع هو التقدير القائم على أساس إعطاء درجة مطلقة تعبر عن مستوى الطالب في الامتحان ، فالطالب الذي يحصل على ٩٠ % من الدرجة النهائية يعني أنه أتقن

٩٠ % مـــن المقـــرر المدرسي . وعلى هذا يجب أن يكون تصميم الاختبار ملتزماً بأهداف تفصيلية محددة تعكس عناصر المقرر الدراسي .

• **تعـــريف أنور الشرقاوي** (أنور الشرقاوي وآخرون ، ١٩٩٦ ، ٢٧-٢٨) الاختبار محكـــي المـــرجع هو ذلك الاختبار الذي يذكر لنا شيئاً عن الفرد بدون الرجوع إلى أداء فرد آخـــر . ففـــي المجـــال المعرفي مثلاً يذكر لنا الاختبار ماذا يعرف الفرد أو ماذا يستطيع أن يفعـــل، وفـــي المجـــال غير المعرفي يكشف لنا عن اتجاهات الفرد وقيمة ومستوى أدائه في النواحي الأخرى .

• **تعـــريف جابـــر عبد الحميد** (جابر عبد الحميد ، ١٩٩٦ ، ٢٣٣ ، ٢٣٧) يستخدم الاختـــبار محكي المرجع لتحديد مكانة الفرد بالنسبة لمجال سلوكي حسن تعريفه ويضيف أن القياس المحكي المرجع هو :

١- أن نحدد مجالاً سلوكياً تحديداً واضحاً .

٢- أن نحدد العلاقة بين أداء الفرد وهذا المجال السلوكي .

أي أننا نصل أداء الفرد بمحك وهو فئة من الأنماط السلوكية حسن تحديدها .

• **تعـــريف كيوبيســـزن** (Kubiszyn , 2000 , p33) الاختبار محكي المرجع هو الاختـــبار الذي يمدنا بمعلومات تبين لنا مستوى الطالب في إتقان مهارة ما ، أو مجموعة من المهـــارات ، ويـــتم ذلك بمقارنة أداء الطالب بمستوى إتقان يدعى محك Criterion . وهذه المعلومات لا تقدم شيئاً عن موقع الفرد أو ترتيبه بالنسبة للآخرين .

تعليق :

من خلال عرض التعاريف السابقة للاختبار محكي المرجع يلاحظ الباحث ما يلي :

١- اتفقـــت هـــذه الـــتعاريف على أن الاختبار محكي المرجع يعطي معلومات عن أداء الفرد، ويربطها مباشرة مع المحك . دون أية معالجات (كما في الاختبارات المعيارية) .

٢- تســـعى جميع هذه التعاريف إلى تحديد موقع الفرد (أداء الفرد) على متصل الإتقان الذي يبدأ من عدم الإتقان تماماً إلى الإتقان التام للنطاق .

٣- يقـــوم بعـــض هذه التعريفات بمقارنة أداء الفرد على الاختبار بنطاق سلوكي معرّف جـــيداً ، أي أنهـــا تنظر إلى وظيفة الاختبار من خلال هذه المقارنة وتحدد موقع الفرد بالنسبة للـــنطاق السلوكي ، ومعرفة ما حققه الفرد في هذا النطاق مع إهمال وضع الفرد بالنسبة إلى أقرانه ، مثل تعريف فارجاس ، بابام ، صلاح الدين علام.

obeikandl.com

٤- و نظر البعض إلى الاختبار محكي المرجع من خلال مكوناته بأنه عينة مفردات مأخوذة من نطاق سلوكي معرف جيداً ، وتحديد ما يمكن أن يقوم به الفرد وما لا يمكنه القيام به في هذا النطاق ، وذلك من خلال إجابته على هذه المفردات . كما في تعريف جود وتعريف أنور الشرقاوي .

٥- و نظر البعض إليه من خلال تفسير الدرجة المطلقة ، وأن هذه الدرجة تعادل النسبة المئوية من أهداف المجال . كما في تعريف رجاء أبو علام وصلاح أبو ناهية وكيوبيسزن .

والمتتبع لأدبيات القياس التربوي والنفسي يلاحظ تعدد وتنوع تعريفات الاختبارات محكية المرجع . وهذه التعريفات تتشابه في بعض عناصرها وتختلف في البعض الآخر . وربما يرجع ذلك إلى حداثة مفهوم الاختبارات محكية المرجع وتعدد أنواع هذه الاختبارات ، كما يرجع إلى عدم الاتفاق حول مفهوم " المحك Criterion " الذي تنسب إليه درجات الأفراد في هذه الاختبارات ، إذ يرى بعض علماء القياس من أمثال جليز Claser ونيتكو Nitko أن مفهوم "المحك" يعني نطاقاً شاملاً من المعارف والمهارات المحددة تحديداً جيداً بحيث يمكن، نتيجة لموازنة أداء الفرد في الاختبار بهذا النطاق ، أن نعرف ما يستطيع أن يؤديه هذا الفرد وما لا يستطيع أن يؤديه. بينما يرى البعض الآخر أن المحك يشير إلى أداء " Standard " أو درجة قطع " Cut-Score " (صلاح الدين علام ،٢٣، ٢٠٠١- ٢٤). وأما ايكن Aiken 1982 فيرى أن المحك هو أساس للحكم على المتعلمين من خارج الظاهرة موضع القياس ويتم إعداده قبل تطبيق الاختبار(محمود منسي،١٩٢، ٢٠٠٣-١٩٣).

ويشير بابام Pobham 1975 إلى أن كلمة محك Criterion في الاختبار محكي المرجع تأخذ المعاني الآتية : الهدف التعليمي Instructional Objective ، أو تستخدم أيضاً بمعنى مستوى الكفاءة المرغوب تحقيقه لدى المتعلم A desired Level of Learner Proficeincy أو بمعنى الناتج التعليمي المرغوب تحقيقه A desired Postinstruction Outcome (آمال محروس ، ١٩٨٨ ، ٢٢).

وعلى ضوء ذلك نجد أننا نواجه اتجاهين الأول تستخدم فيه كلمة محك Criterion بمعنى " مستوى Standard " والاتجاه الآخر تستخدم فيه كلمة محك Criterion بمعنى بعض مظاهر السلوك في الحياة الواقعية والذي نرد إليه السلوك أو الأداء الناشئ عن تطبيق الاختبار (أنور الشرقاوي ، ١٩٩٦ ، ٢٤) .

وربمـــا يـــرجع اخـــتلاف تعـــريفات الاختبار محكي المرجع إلى اختلاف طبيعة النطاق الســـلوكي الـــذي تنســـب إليه الدرجة وتفسر في إطاره ، ومحتوى هذا النطاق ودرجة تعقده (صـــلاح عـــلام ، ٢٠٠١ ، ٢٤) . وهناك ثلاث تسميات (أنواع) لهذه الاختبارات اعتماداً على درجة تحديد النطاق السلوكي الذي يستند إليه بناء كل منهما وهي :

١-الاختبارات هدفية المرجع : Objective - Referenced Tests

- **تعـــريف ناديـــة عـــبد الســـلام** (١٩٨٦ ، ٩٦) الاختبار هدفي المرجع هو الاختبار المكون بحيث يقدر مستويات أداء المختبرين بالنسبة إلى مجموعة أهداف محددة جيداً .

- **وقـــد عـــرف كل** من ميلمان Millman 1974 وبابام Popham 1978 وهامبلتون Hambleton 1978 وصـــلاح الدين علام ١٩٨٠ ، ١٩٨٢ الاختبار هدفي المرجع بأنه ذلك الاختـــبار الـــذي يقيس مجموعة من الأهداف التعليمية السلوكية ، وهو يشتمل عادة على عدد قلـــيل نســـبياً مـــن المفـــردات أو الأسئلة التي يفترض أنها تمثل جميع المفردات التي تقيس مجموعة الأهداف (إسماعيل حسن فهيم ، ١٩٩٦ ، ١٦) .

٢-الاختبارات نطاقية المرجع : Domain - Referenced Tests

- **يعـــرفه مـــيامان** Millman 1974 بأنـــه عيـــنة مفردات مأخوذة من مفردات نطاق Domain معـــرف جـــيداً تقيس هدف تعليمي . والاختبار النموذجي يتضمن مفردات اختبار تقيس أهداف تعليمية متعددة (Hambleton , & Cook , 1977 , p 93) .
- **تعريف صلاح الدين علام** (١٩٨١ ، ٦١) يعرف الاختبار نطاقي المرجع بأنه ذلك الاختـــبار الذي يتكون من عينة عشوائية أو طبقية من المفردات ، تسحب من عينة شاملة من المفـــردات التـــي تقـــيس متطلبات أو مهارات معينة تسمى النطاق Domain . وهذا النطاق يكـــون معـــرفاً تعريفاً جيداً عن طريق التحديد الواضح ، أو عن طريق قواعد دقيقة لاختيار وبناء المفردات.

٣- اختبارات الإتقان : Mastery Tests (MTS)

- **تعريف شرينر** (Schreiner , 1988 , p 21) اختبارات الإتقان تستخدم في التأكد من أن الطلاب قد اكتسبوا مستوى الأداء المحدد ، وذلك من خلال وحدة أو منهج دراسي ما. فـــإذا أمكن تحديد الأهداف الهامة في المجال السلوكي ، يكون الاعتبار الهام هو التأكد ما إذا

كان الطالب قد حقق هذه الأهداف أكثر من أن نعرف موقع الطالب بالنسبة للطلاب الآخرين، والطالب إما أن يتقن الهدف بصورة مرضية أو أن لا يتقن .

• و **يقـول صلاح الدين علام** (٢٠٠١ ، ٢٥) تصمم هذه الاختبارات من أجل تحديد ما إذا كان فرد معين اكتسب السلوك الذي يهدف برنامج تعليمي أو تدريبي إلى تنميته لديه . فاختبار الإتقان أو التمكن يساعد في اتخاذ قرارات تتعلق بإتقان الفرد لهدف تعليمي أو مهارة معينة أو نطاق من المهارات المحددة ، ولذلك يمكن أن يكون هذا الاختبار مرجعي الهدف أو مرجعي النطاق .

ممـا تقـدم يخلص الباحث إلى التعريف التالي :**الاختبار محكي المرجع هو اختبار يبنى مـن مفردات تغطي تماماً نطاق سلوكي حسن تعريفه ، ويستخدم لتقدير أداء الفرد بالنسبة إلى هذا النطاق السلوكي دون الحاجة إلى موازنة أداء الفرد بأداء أفراد آخرين .**

❋ استخدامات الاختبارات محكية المرجع :

١- تسـتخدم درجـات الاختـبار محكـي المرجع في تفسير النتائج تفسيراً مباشراً لأنها تتضـمن حقائق وصفية دقيقة ومحددة ، عما يفعله الطلاب ، وفي اتخاذ القرارات التعليمية وتقويم البرامج وتأثيرها (Hambleton , 1978 , p 280) .

٢- إذا لـم يكن هناك قيود على عدد الأفراد الذين يمكن أن يحققوا الأداء المطلوب ، فإن الاختـبارات محكـية المرجع تكون مناسبة في الكشف عن هؤلاء الأفراد . (صلاح الدين علام ، ٢٠٠٠ ، ٣١٦) .

٣- أوضـح كل من هامبلتون Hambleton 1978 وإيجنور Hambleton & Eignor 1978 ومـيلمان Millman 1974 وآخرون ، أن الاختبارات محكية المرجع تستخدم في تشـخيص نقـص تعلـيم الطـلاب ومراقبة تقدمهم وتقيم البرامج التعليمية وتعليم الأفراد ومصداقية أمتلاكهم الحد الأدنى من الكفايات Competency ، p (Hambleton , 1978 , 277)

٤- يمكـن أن " تفصـل Tailored " هـذه الاختبارات على الأهداف المحددة التي يقوم علـيها الـبرنامج ، وبذلـك يمكـن الحصول على معلومات محددة عن الأهداف التي يتم تحقيقها، والأخرى التي لم يتم تحقيقها .

٥- يمكـــن الاستفادة منها في حل بعض المشكلات التي ترتبط بقياس الأهداف المركبة ، وكذلك يمكن قياس المهارات الدقيقة .

٦- يمكـــن بواسطتها قياس بعض الأهداف التربوية المتخصصة التي ترتبط بمجموعات معينة من الطلاب والتي تفسر في ضوء مستويات محددة .

٧- تستطيع أن تحقق مرونة البرنامج التربوي من حيث المضمون والزمن المخصص له.

٨- القـــرارات التـــي تتخذ داخل هذا النموذج أما أن تكون قرار وضع ، أو قرار إتقان . وذلـــك عـــن طريق استخدام نموذج قبلي لاختبار محكي المرجع . ونموذج بعدي لاختبار محكـــي المـــرجع ، ويحـــدد اختبار الوضع موقع الفرد على هذا التتابع التعليمي ، فالهدف الأول لاختـــبار الوضع إذن هو تحديد مستوى تقدم التلميذ في كل موضوع من موضوعات الاختـــبار . وهـــذه المعلومة محكية المرجع تخبرنا بما يستطيع التلميذ عمله وما لا يستطيع عملـــه بالنســـبة للمهـــارات المتضمنة في كل موضوع ، وبالتالي فهي تساعدنا على اتخاذ قرارات تعليمية متعلقة بما يجب على التلميذ دراسته (نادية عبد السلام ، ١٩٩٠ ، ٥٨) .

٩- يقـــول رونـــالد Ronald أن الاختـــبارات محكية المرجع تستخدم في تقديم معلومات مفـــيدة فـــي عـــلاج صـــعوبات التعلـــيم ، وهـــذا العلاج مصاحب لفكرة التقويم التكويني (Hambleton , 1978 , p 180) .

١٠- ويشـــير صـــبكوفياك (Subkoviak , 1976 , 265) إلـــى أن الاختبارات محكية المـــرجع تستخدم غالباً في توجيه البرامج التعليمية ، التي تحاول أن يكون العدد الأكبر فيها هو من الطلاب الذين أتقنوا المراحل .

١١- وتقول إنستاذي (Anastasi , 1976 , p 97) بأن استخدام الاختبارات محكية المرجع يفيد في تحسين وتطوير طريقة إعداد المدرسين للاختبارات ، التي يستخدمونها داخل غرفة الصف .

١٢- يفضـــل الاختبار محكي المرجع ، عندما يحتاج المرشد Counselor إلى معرفة هل المفحـــوص قـــد حصل المستويات الدنيا في موضوع أو مجال محدد (نادية عبد السلام ، ١٩٩٦ ، ٦٩) .

١٣- يهـــدف تقويم البرنامج التربوي لتعرف مدى تحقيقه لأهدافه المرجوة ، والاختبارات محكـــية المرجع تفيد ليس فقط في التقويم الختامي للبرنامج لتعرف مدى فاعليته في تحقيق

تلك الأهداف ، وإنما تفيد بدرجة أكبر في التقويم البنائي للبرنامج وذلك بالكشف عن جوانب القوة والضعف في أهدافه وتصميمه وإجراءاته وأساليب تنفيذه ، بحيث يمكن تعديل مسار البرنامج بناءً على نتائج هذه الاختبارات .(صلاح الدين علام، ٢٠٠٠ ،٣٣) .

❋مقارنة بين الاختبارات محكية المرجع ومعيارية المرجع :

من خلال التعاريف السابقة لكل من الاختبارات محكية المرجع ومعيارية المرجع واستخداماتها ، يتبين للباحث (وكما هو معروف) أن كل نوع من أنواع هذه الاختبارات، نشأ مرتبطاً بفلسفة تربوية معينة . فعندما كانت الفلسفة التربوية السائدة هي مقارنة أداء الأفراد فيما بينهم وإبراز الفروق الفردية ، ظهرت الاختبارات معيارية المرجع تلبية لهذا الغرض . ولكن وبسبب العيوب التي ظهرت في هذا النوع من الاختبارات ، وحيث أنها لا تهتم بما حققه الفرد من كسب في المهارات والمعلومات ، وتحوّل الاهتمام من المقارنة بين الأفراد إلى مقارنة أداء الفرد بمحك للأداء محدد مسبقاً . وذلك لمعرفة ما حققه الفرد وما يمكنه القيام به من أعمال ، ظهرت الاختبارات محكية المرجع لتلبي هذا الطلب .

ومهما يكن فإن كل من هذين النظامين يقدم معلومات مناسبة تفيد في تقويم الأفراد ، واختيار أحد الأنواع من هذه الاختبارات ينبغي أن يعتمد على الغرض من عملية التقويم ، واستخدامات هذه الاختبارات .

ولكن بسبب الاختلاف في الخلفية التي يعتمد عليها كل اختبار ، وجدت فروق بين هذين النوعين .

- نحصل في الاختبار معياري المرجع على الفروق بين أداء الفرد بالنسبة إلى أداء الجماعة التي ينسب إليها .أما في الاختبار محكي المرجع فإننا يمكن أن نحصل على الفروق في أداء الفرد بما حققه وما لم يحققه في الاختبار.(أنور الشرقاوي ، ١٩٩٦،٢٨).

- ويميز ليفنجستون Livingston المقاييس محكية المرجع عن المقاييس معيارية المرجع، فعندما نستخدم المقاييس معيارية المرجع نريد أن نعرف مدى انحراف درجة الطالب من متوسط المجموعة . وعندما نستخدم المقاييس محكية المرجع فإننا نريد أن نعرف مدى انحراف درجة من معيار محدد " المحك " (Livingston , 1972 , 14).

- المقاييس معيارية المرجع تستخدم حيث يتطلب الموقف نوعاً من الانتقاء Selectivit. بينما المواقف التي يرغب فيها الفرد معرفة إذا ما كان هذا الشخص لديه مهارة خاصة ولا يهمه كم من الأفراد لديهم هذه المهارة ، فإن المقاييس محكية المرجع تكون مناسبة .

• عندما تستخدم الاتجاهات محكية المرجع للاختبار ، فإننا لا نحتاج لضم الدرجات على الاختبارات المختلفة . فكل درجة في حد ذاتها تمثل نموذجاً لبعض الأداء المحدد . ولا يهمنا هنا أن تضم درجة الطالب في اختبار للفهم مع درجته في اختبار للمعرفة . فنحن نريد أن نعرف فقط في كل حالة ، إذا ما كان الطالب حقق الغرض أم لا . لكن عندما تستخدم برامج الاختبار معيارية المرجع ، فإن الاختبارات العديدة التي اختارها المدرس لتمثل أداء التلميذ ، ربما تحتاج إلى أن تضم معاً في ملخص مركب للحصول على الدرجة الكلية .

• يعتبر التباين أحد الفروق الجوهرية بين الاختبارات معيارية المرجع ومحكية المرجع. وتعتبر زيادة التباين أمراً مفضلاً في الاختبار معياري المرجع، ذلك لأننا نريد أن نفرق بين الطالب س وبين الطالب ص ، ع .أما بالنسبة للاختبار محكي المرجع ، فإن التباين غير ذي أهمية . فمعنى الدرجة لا يعتمد على المقارنة مع الدرجات الأخرى . إنما ينشأ مباشرةً من العلاقة بين الوحدات والمحك . (نادية عبد السلام ، ١٩٨١ ، ٤٦– ٥٣) .

• ويشير ايكن Aiken إلى أن الاختبارات محكية المرجع تؤكد على الفروق داخل الفرد، وعادةً ما نصنف الطلاب إلى مجموعتين ، إحداهما أتقنت المادة والأخرى لم تتقن ، وعندما نحدد مستوى الأداء على المقياس في الاختبارات محكية المرجع لا بد من تحديد المحك مسبقاً ، وليس درجات الأفراد الآخرين . (Aiken 1985 , p 116) .

• ويشير شرينر Schreiner إلى أن هناك إجماعاً بأن المعلومات التي تقدمها الاختبارات معيارية المرجع تؤدي إلى معلومات قليلة حول ما يستطيع الطالب أن يفعله وعلى سبيل المثال ، ماذا تعني الدرجة الخام ١٢١ أو الدرجة الميئنية ٦٣ بالنسبة لمهارة القراءة التي يستطيع أن يؤديها الطالب ؟ إن مثل تلك الدرجة تكون مفيدة في القرارات التي لها طبيعة المقارنة ، أكثر من كونها مفيدة في كشف وتحديد ما يستطيع الطالب أن يؤديه وما لا يستطيع أن يؤديه (Schreiner , 1988 , p 21) .

• أفضل البنود من وجهة النظر معيارية المرجع هي الأقدر على التمييز ، وهي تلك التي يساوي فيها كل من معاملي سهولة البند وصعوبته المقدار (٠,٥) . أما أسوؤها فتلك التي لا تستطيع أن تميز بين الأفراد ، كأن يخفق في الإجابة عليها جميع الأفراد ، أو أن ينجح في الإجابة عليها جميع الأفراد . أما من وجهة النظر محكية المرجع فأفضل البنود تلك الأقدر على قياس النمو أو التحصيل ، وهو البند الذي يكون مستوى سهولته قبل البدء في البرنامج التربوي صفراً ، أي لا يستطيع أي من الأفراد الإجابة على السؤال قبل دراستهم للبرنامج ،

ثم يصبح معامل سهولة هذا البند واحداً صحيحاً ، بعد تعلم البرنامج حيث يستطيع جميع الأفراد الإجابة على هذا السؤال . (أمينة محمد كاظم ، ١٩٨٨ ، ١٨- ١٩) .

• ويبين هامبلتون Hambleton 1988 أن الاختبارات معيارية المرجع تعتمد على مخطط الاختبار ، وتستخدم الأهداف والسلوكيات المتضمنة في المخطط لكتابة الأسئلة ، وإن الاختبارات محكية المرجع تعتمد على التوصيف التفصيلي الدقيق للمحتوى . أي أن الأهداف وحدها لا تكفي لكتابة العناصر . والأساس في تفسير الدرجة في الاختبارات المعيارية هو أداء المجموعة التي قنن عليها الاختبار ، لذا فإن تعميم درجات الطالب بالنسبة للمحتوى أو المجال الذي يقيس الاختبار يكون محدداً . أما في الاختبارات مرجعية المحك فإن التعميم ضروري من حيث الدرجة التي اكتسبها الطالب أو مجموعة من الطلاب على المجال الذي يقيس الاختبار . (إبراهيم مبارك ، ٢٠٠٠ ، ١٣٥ – ١٣٦) .

• الاختبارات معيارية المرجع تكون واسعة ، وتغطي عدة موضوعات . أما الاختبارات محكية المرجع فهي ضيقة وتغطي مستوى ضيق . والاختبارات معيارية المرجع تكون ضحلة ، وعادة تحتوي على مفردة أو اثنتين لكل هدف . بينما الاختبارات محكية المرجع تكون شاملة ، وعادة يكون هناك ثلاث مفردات أو أكثر لكل هدف . وعند تفسير الاختبارات معيارية المرجع تستخدم الرتب الميئنية ، والدرجات المعيارية . وفي الاختبارات محكية المرجع يستخدم عدد الناجحين أو الفاشلين أو مدى الأداء المقبول. Kubiszyn & Borich) (2000 , p 39 , . ويتفق جرونلند ولن Gronlad & Linn مع كيوبيسزن Kubiszyn في ذلك ، ويضيف عليه أن الاختبارات معيارية المرجع تستخدم كاختبارات مسحية ولكن الاختبارات محكية المرجع تستخدم كاختبارات إتقان(Gronland & Linn , 1990, p 16).

• صدق الاختبار محكي المرجع يختلف عن صدق الاختبار معياري المرجع ، فمثلاً استخدام معاملات الارتباط في معظم طرق حساب صدق الاختبار معياري المرجع لا تناسب معاملات ثبات وصدق الاختبار محكي المرجع . وتستخدم في هذا المجال فنيات تعطي نتائج ذات معنى وقابلة للتفسير ، بغض النظر عن مقدار التباين في توزيع درجات الاختبار . (محمود منسي ، ٢٠٠٣ ، ٢٠٠) .

بناءً على ما تقدم يستطيع الباحث أن يقدم المقارنة التالية بين الاختبارات محكية المرجع ومعيارية المرجع في الجدول (١) التالي :

الجدول (١)

المقارنة بين الاختبارات محكية المرجع ومعيارية المرجع

صفة المقارنة	الاختبار معياري المرجع	الاختبار محكي المرجع
الغرض من الاختبار	قياس أداء الفرد بالنسبة لأداء الأفراد الآخرين في المجموعة التي ينتمي إليها من أجل ترتيب الأفراد .	قياس أداء الفرد بالنسبة لمحك سلوكي معين وتحديد مدى الإتقان وما توصل إليه الفرد وما يمكنه القيام به .
بناء الاختبار	– مجال الاختبار عام وشامل ويغطي عدة موضوعات. – متوسط الصعوبة وهو ٠,٥٠ للمفردة ليكون معامل التميز مرتفع. – تستبعد المفردات التي يجيب عليها جميع الطلاب إجابة صحيحة وكذلك المفردات التي يجيب عليها جميع الطلاب إجابة خاطئة لأن هذه المفردات لا تميز بين الطلاب. – يعطي كل هدف مفردة أو اثنتين فقط بسبب وسع الموضوعات.	– مجال الاختبار يكون ضيقاً ويهتم بالجوانب التفصيلية للمجال . – لا يوجد أهمية للصعوبة ولكن المفردة قبل التعليم تكون صعبة جداً وبعد التعليم تصبح سهلة جداً (إذا كان التعليم فعالاً). – لا تستبعد أية مفردة لأن كل مفردة تغطي جزءاً من النطاق السلوكي الذي نرغب تحديد وضع الفرد بالنسبة له (شرط المفردة سليمة الصياغة). – يعطي كل هدف ثلاث مفردات أو أكثر.
درجات الاختبار	– انتشار الدرجات مرغوب (التباين) وهو مفضل لأنه يساعد على إظهار الفروق الفردية . – التباين كبير بين الدرجات . – نقطة ارتكاز الدرجات هو المتوسط الأداء.	– انتشار الدرجات غير مهم . – يتوقع أن يكون التباين قليل وفي حالات الإتقان التام أو عدم الإتقان تماماً يكون التباين معدوماً . – نقط الارتكاز تكون عند الطرفين والدرجة في أعلى المقياس تدل على الإتقان التام وفي أدنى المقياس تدل على عدم الإتقان تماماً.

	– الاهـتمام يكون بمدى انحراف درجة الفرد عن متوسط درجة المجموعة. – تـتوزع الدرجـات وفـق المنحنـي الاعتدالي. – تـأخذ الدرجة معناها من الترتيب مع درجات الأفراد الآخرين . – تفسـر الدرجـات فـي ضوء الرتب المئينية والدرجات المعيارية .	– الاهـتمام يكون بمدى انحراف درجة الفرد عن المحك المحدد. – لـيس لهـا تـوزع منتظم ولكن غالباً تـتوزع وفق منحني ملتوي نحو اليسار إذا كان التعليم فعالاً أو ملتوياً نحو اليمين إذا كان التعليم غير فعال . – تـأخذ الدرجـة معـناها من ارتباطها بالمحك . – تفسر الدرجة في ضوء عدد الناجحين أو الفاشلين .
الاستخدام	– مقارنة أداء الفرد مع أداء زملائه . – تستخدم في المواقف التي تتطلب انتقاء الأفراد . – تعطي معلومات ضعيفة في معرفة ما حققه الفرد وتكون مفيدة في حال المقارنة بين الأفراد. – تسـتخدم كاختـبارات مسـحية لأنها عامة.	– مقارنة أداء الفرد بمستوى أداء محدد مسبقاً . – تسـتخدم لمعـرفة ما يمتلك الفرد من معارف ومهارات وقدرات وبما يمكنه أن يعمله وما لا يمكنه . – تعطـي معلومـات تفصيلية عما حققه الفرد وقد تكون المعلومات غير مفيدة في حال المقارنة بين الأفراد . – تسـتخدم كاختبارات تشخيصية وتقويم مدى نجاح البرامج .

❋طرق بناء الاختبارات محكية المرجع :

تخـتلف خطـوات بـناء الاختـبار محكي المرجع عن خطوات بناء الاختبار معياري المرجع، وذلك بسبب اختلاف هدف كل نوع منهما ، والقرارات التي تتمخض عن نتائجهما .

فالاختبارات معيارية المرجع كما تبين سابقاً تهدف إلى مقارنة أداء الفرد بأداء المجموعة المعـيارية ، مـن خـلال مقارنة درجته بمتوسط أداء المجموعة لتحديد موقعه بالنسبة لهذه

obeikandl.com

المجموعـــة . والدرجـــة في هذا النوع تتسم بالعمومية والشمول للمجال الذي تقيسه ، فهي لا تركز على جزئيات المجال المقاس .

أمـــا الاختبارات محكية المرجع فتهدف إلى مقارنة أداء الفرد بمحك أداء مسبق ، لمعرفة مـــا حققـــه الفرد وما فشل في تحقيقه في البرنامج التعليمي ، وتقدم وصفاً تفصيلياً لحالة الفرد بالنســـبة للمحـــك أو المجـــال الســـلوكي الذي يقيسه . فهي تساهم في تشخيص جوانب القوة والضعف عند المتعلم وفي البرامج التعليمية .

ولعـــدم وجـــود اتفـــاق تام حول معنى المحك Criterion الذي يقاس الأداء في ضوئه وينســـب ويـــرد إليه ، تعددت تعريفات الاختبارات محكية المرجع . ولذلك يعتبر تحديد هذا المحـــك أو الـــنطاق الســـلوكي أو الكفايات التي يقيسها الاختبار ، والتي ينسب إليها ، حجر الأساس في بناء هذه الاختبارات .

وســـوف يعرض الباحث بعض الطرق المختلفة التي تساهم في تحديد هذا النطاق أو في بناء الاختبارات محكية المرجع .

١- صيغة المفردات Item Form :

يـــرجع اســـتخدام هذه الطريقة إلى كل من هافلي Hively وباترسون Paterson وبيج Page عام ١٩٦٨ . ويعرف بابام بأنها مجموعة قواعد شديدة التفصيل لبناء مفردات اختبار، وهـــذه المفـــردات ذات طبيعة متجانسة وهي تناسب المواد ذات الطابع التركيبي ، مثل تقدير المهارات في الرياضيات والعلوم . (Popham , 1982 , p 20) .

٢- قواعد التحويلات اللغوية : Linguistic Transformations

وهي طريقة لوصف المجال وكتابة مفردات الاختبار المرتبطة ارتباطاً وثيقاً بهذا المجال. حيـــث يتألف المجال من مفردات نص تعليمي يتم تحويل العبارات في هذه الفقرات إلى أسئلة وذلـــك باستخدام التحويلات اللغوية . وبذلك نضمن كتابة جميع المفردات المماثلة ، ومن هذه المفردات يمكن توليد اختبارات متكافئة معتمدة على هذه المفردات المتماثلة .

(Millman , 1982 , p 38-39).

٣- الأهداف المكبرة : Amplified Objectives

نظراً لأن الأهداف السلوكية التي يستخدمها المعلمون في تحديد النواتج التعليمية المرجو تحقـــيقها في وحدة دراسية أو مجال تدريبي معين ، تصاغ عادةً في عبارة موجزة ، فإن هذه

الأهداف تتطلب مزيداً من التفصيل بحيث تمكن القائمين ببناء الاختبار من كتابة مفردات اختبارية متجانسة تقيس الهدف المحدد قياساً دقيقاً وصادقاً .

لذلك اقترح بابام Popham أن يصاغ الهدف السلوكي مبدئياً بالطريقة المعتادة ، أي يشتمل بقدر الإمكان على الفعل السلوكي ، والمحتوى المرجعي وشروط الأداء ومستوى الأداء .

ويلي ذلك توضيحاً تفصيلياً لكيفية بناء المفردات الاختبارية التي تقيس ذلك الهدف . وعلى الرغم من هذه التفاصيل التي يضيفها الهدف المكبر ، إلا أنه يظل هناك بعض الغموض فيما يتعلق بالفئة الشاملة للسلوك المرجو قياسه . (صلاح الدين ، علام ، ٢٠٠١، ١٠٣- ١٠٥) .

ومن خلال الدراسات السابقة التي وجدها الباحث ، تبين له أن طريقة مواصفات الاختبار والتي تنسب إلى بابام هي الأكثر انتشاراً واستخداماً في بناء الاختبار محكي المرجع. وبالإضافة إلى ذلك فهي تحدد بدقة المجال السلوكي الذي يقيسه الاختبار ، وتتبع خطوات واضحة ودقيقة لبناء المفردات ، وذلك من خلال مواصفات الاختبار التي تحتوي على مراحل دقيقة لوصف كل هدف تعليمي من أهداف محتوى الاختبار .(صلاح الدين علام، ٢٠٠٠ ،٣١٧-٣٢٥) (سعاد حسانين ،٢٠٠٠) (اعتدال عبد الله ، ٢٠٠٠)

وبما أن أداة البحث الحالي هي اختبار محكي المرجع فإن الباحث سوف يتبع هذه الطريقة في بناء اختباره ،و هذه الطريقة تمر بالمراحل التالية:

١- تحديد النطاق السلوكي المراد قياسه وتحليله إلى مكوناته .
٢- تحديد الأهداف العامة وصياغتها في أهداف سلوكية .
٣- إعداد مواصفات الاختبار .
٤- كتابة مفردات الاختبار .
٥- تحديد صدق الاختبار وثبات درجاته .

وسوف يتناول الباحث هذه الطريقة بشيء من التفصيل .

المرحلة الأولى:تحديد النطاق السلوكي المراد قياسه وتحليله إلى مكوناته.

التعريف الدقيق للنطاق السلوكي Behavioral Domain الذي يقيسه الاختبار محكي المرجع ، يعد متطلباً أساسياً من متطلبات هذا النوع من الاختبارات . ويقصد بالنطاق السلوكي المعارف والمهارات المحددة تحديداً دقيقاً مما يمكننا من معرفة ما يستطيع الفرد

أداءه ومـــا لا يستطيع أداءه ، ويذكر بابام Popham أن التحديد الدقيق للنطاق السلوكي يعد الركيزة الأساسية لمفهوم المحك .

وتخـــتلف درجـــة تعقد النطاق السلوكي اختلافا ملحوظاً في درجة تعقدها ونوعية الأداء الـــذي تتضـــمنه متطلبات النمو أو التعلم . وكلما زاد تعقد النطاق السلوكي تطلب تحليله إلى مكونـــات يمكـــن قياســـها جهـــداً ملحوظـــاً ، ويفضـــل الاستعانة بخبراء المادة الدراسية أو اختصاصـــيي تصميم البرامج التدريبية في تحديد النطاق السلوكي المراد قياسه ، وتحليله إلى مكوناته وما تشتمل عليه من معارف ومهارات وسلوكيات .

ويهدف هذا التحليل لتحديد المتطلبات أو المكونات السلوكية اللازم توافرها لدى المتعلم، لكـــي يحقـــق الكفايـــة أو المهـــارة الرئيسية التي يقيسها الاختبار ، ونكرر عملية تحليل هذه المتطلبات حتى نصل إلى السلوك المدخلي للمتعلم . (صلاح الدين علام ، ٢٠٠٠ ، ٣١٩).

ومـــنه فـــإن تحديـــد النطاق السلوكي المراد قياسه ، ثم تحليله إلى مكوناته من المعارف والمهـــارات والكفايات الأساسية ، بالاستعانة بخبراء في هذا المجال ، يساعد مصمم الاختبار في تحديد الخطوات التالية من بناء الاختبار . وإن إهمال أي عنصر أو جزء من هذا المجال السلوكي سواءً عند تحديده أو عند التحليل ، سوف يؤثر بالتأكيد على بناء الاختبار ، وسوف لـــن يقـــيس الاختـــبار هذا العنصر المهمل . لذلك يعتبر تحديد النطاق السلوكي وتحليله إلى مكوناته الأساس الذي ترتكز عليه باقي مراحل بناء الاختبار .

المرحلة الثانية : تحديد الأهداف العامة وصياغتها في أهداف سلوكية .

بعد الانتهاء من عملية تحليل الكفايات أو المهارات الرئيسية للمحتوى التعليمي إلى نواتج مـــن معـــارف ومهارات فرعية ، ينبغي تحديد الأهداف العامة لذلك المحتوى ، فلكل محتوى تعليمـــي أو تدريبي أهداف مرتبطة به يسعى المعلم لقياس مدى تحققها لدى المتعلمين . وتلك الأهـــداف تعـــبر عن نواتج متوقعة لعملية التعلم ، ولكنها في نفس الوقت تحتاج إلى صياغة ســـلوكية إجرائية ، يمكن أن تتخذ دليلاً علماً أن المتعلم قد حقق كل هدف من الأهداف العامة السابق تحديدها . (صلاح شريف ، ٢٠٠٠ ، ٢٤- ٢٥) .

لذلـــك يجب الاهتمام بصياغة الأهداف السلوكية التي تقيس مكونات الكفايات والمهارات الرئيســـية التي تم تحليلها ، والتي يتكون منها النطاق السلوكي . وسوف يعرض الباحث فيما يلـــي تعـــريف الهـــدف الســـلوكي وعناصر الهدف السلوكي ومستويات الأهداف السلوكية ، وتصنيف الأهداف .

١- تعريف الهدف السلوكي :

الهدف السلوكي هو العبارة التي تصف السلوك المعرفي أو المهاري المطلوب تحقيقه من الطالب بعد دراسته لمفهوم أو حقيقة أو مهارة . (مدحت السيد ، ١٩٩٥ ، ٩٦) .والهدف السلوكي الجيد هو ذلك الهدف الذي يصف السلوك المطلوب أداؤه بعبارات إجرائية يمكن ملاحظتها وقياسها ، ويحدد شروط وظروف الأداء التي تعتمد على ضبطه والتحكم في المتغيرات التي تؤثر فيه . كما يضع أدنى أداء مقبول يشير إلى مدى تحقق الغاية التعليمية . (على ماهر ، ٢٠٠١ ، ٤٦) .

٢- عناصر الهدف السلوكي :

كل هدف سلوكي يتكون من أربعة عناصر أساسية وهي :

١- تحديد السلوك أو الأداء الظاهري للمتعلم :

أي تحديد وتميز نوع الأداء أو التغيير الذي طرأ على سلوك المتعلم بعد الانتهاء من تعلم وحدة دراسية معينة ، بحيث يكون هذا التحديد واضحاً وصريحاً ودقيقاً . ويتطلب هذا استخدام كلمات أو أفعال إجرائية تتضمن ملاحظة وقياساً مباشراً للأداء المتوقع حدوثه من جانب المتعلم . ويقترح بلوم ١٩٥٦ وميجر ١٩٦٢ استخدام أفعال إجرائية (مثل : يفرق ، يقارن ، يستخدم ، يحسب ، .....) . (علي ماهر ، ٢٠٠١ ، ٤٤) .

وهذه الأفعال تسمى في بعض الأحيان أفعال سلوكية .

٢- المحتوى المرجعي :

يمكن أن يحدد هذا المستوى تحديداً عرضياً ، أي مجموعة متجانسة من المحتويات الفرعية مثلاً الأشكال الهندسية ، أو أن يكون أكثر تحديداً ، بحيث يشتمل على عنصر محدد من عناصر هذا المحتوى ويفضل استخدام التحديد الضيق للأهداف . (صلاح الدين علام ، ٢٠٠١ ، ٦١) .

٣- تحديد شروط الأداء :

وهي الظروف والشروط أو القيود التي ستعرض على المتعلم ، والتي يجب أن تتوافر لدى القيام به .

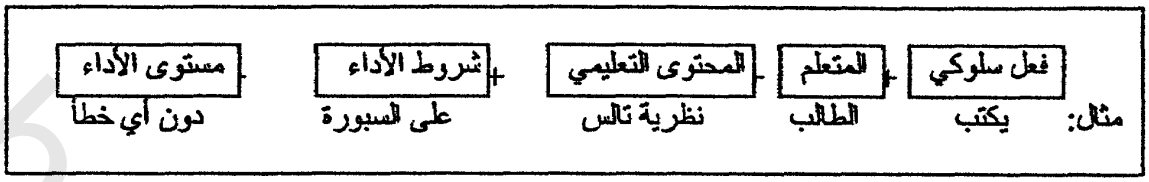
٤- تحديد مستوى الأداء :

وهو تحديد معيار أو مستوى الأداء المقبول والذي يستخدم لتعليم الأداء أو الناتج السلوكي المتوقع تحقيقه لدى المتعلم ، ويتم ذلك بإضافة كلمات أو عبارات تصف أدنى مستوى يمكن قبوله . (علي ماهر ، ٢٠٠١ ، ٤٥- ٤٦) .

obeikandl.com

ويجد الباحث أن عبارة الهدف لها الهيكل أو الشكل التالي :

شكل رقم (١) شكل عبارة الهدف السلوكي



٣- مستويات الأهداف العامة :

تقسم الأهداف السلوكية من حيث عموميتها إلى ثلاثة مستويات ، وهي مدرجة من العام إلى الخاص كما يلي : (صلاح أبو ناهية ، ١٩٩٦ ، ١٣٥- ١٣٧) .

أ- مستوى الأهداف العامة :

في هذا المستوى توصف الأهداف بأنها أهداف عامة أو أهداف تربوية عامة ، وتتميز بدرجة عالية من التجريد ، كما أنها تمثل الأهداف النهائية أو الغايات القصوى للعملية التربوية وتعبر في نفس الوقت عن الفلسفة التربوية للمجتمع .

ب-الأهداف التعليمية الخاصة :

تتميز الأهداف في هذا المستوى بأنها أقل عمومية وتجريداً من الأهداف العامة ، بحيث تقع في مستوى متوسط من حيث العمومية ، ويطلق عليها الأهداف التعليمية الضمنية . وتظهر هذه الأهداف في أهداف المقرر الذي يضمه الكتاب المدرسي ، أو أهداف وحدة تعليمية من المقرر الدراسي ، أو مجموعة من المقررات . ويقوم بوضع هذه الأهداف وصياغتها مؤلفو المناهج والكتب المدرسية .

ج- الأهداف التعليمية النوعية :

وهي أهداف خاصة ومحددة ويتوقع أن تتحقق لدى المتعلم خلال فترة زمنية قصيرة نسبياً ، وبعد الانتهاء من تدريس موضوع دراسي معين ، أو حصة دراسية أو محاضرة . وتعرف هذه الأهداف بالأهداف التعليمية الظاهرية أو الأهداف السلوكية Behavioral Objectives

٤- تصنيف الأهداف التربوية :

تصنف الأهداف التربوية إلى ثلاثة مجالات كبرى تتضمن : المجال المعرفي ، والمجال الوجداني، والمجال النفسي الحركي . ويتضمن المجال المعرفي الأهداف التربوية التي تؤكد

على نواتج التعلم المعرفية وتنمية المهارات والقدرات العقلية . (علي ماهر ،٥١،٢٠٠١) .

وسوف يتم التركيز في هذه الدراسة على الجانب المعرفي ، لأن الدراسة الحالية تهتم بهذا الجانب من التصنيف ، وسوف يتم التركيز على تصنيف بلوم لهذا الجانب . وبلوم اعتمد في تصنيفه للأهداف في هذا المجال على بناء هرمي ، قاعدته البسيط والمحسوس والخاص ، ورأسه المركب والمجرد والعام على الترتيب . أما فئات هذا التصنيف فهي :

١- مستوى التذكر (المعرفة) : Knowledge

هذا المستوى هو أبسط مستويات الأهداف المعرفية ، ويتم في هذا المستوى التعرف على المعلومات واستظهارها ، ويطلب في هذا المستوى من التلميذ إعادة نفس المعلومات المتعلمة بحذافيرها .

٢- مستوى الفهم : Comprehension

وهذا المستوى أرقى قليلاً من مستوى التذكر ، حيث يتطلب أن يفهم المتعلم ما تعلمه . ومن أمثلة السلوكيات التي تدل على فهم المتعلم للمعلومات ، قدرته على تلخيص أو إعادة المعلومة بشكل لا يخل بمحتواها ، كأن يقرأ جدولاً أو رسماً بيانياً معبراً عن معناه . كما يستدل على وجود الفهم من خلال قدرة المتعلم على التفسير ، حيث يمكن للمتعلم أن يقدم شرحاً للمعلومة المتعلمة ، يفسرها ويشرح معناها بلغته هو . (حسن ،٥٥،١٩٩٥- ٥٦) .

٣- مستوى التطبيق : Application

وهو استخدام المجردات في مواقف خاصة وملموسة ، ويمكن أن تكون التعميمات على شكل أفكار عامة ، أو قواعد إجرائية أو طرق معممة . ويمكن أيضاً أن تكون التجريدات مبادئ فنية ، أو أفكار ونظريات يجب تذكرها وتطبيقها .

والقدرة على تطبيق المبادئ والتعميمات على مشكلات ومواقف ، هي نوع من هدف تربوي ، يوجد في معظم مقررات التعليم ،والتلميذ لا يفهم حقاً فكرة أو مبدأ ما لم يستطع استخدامها في مواقف مشكلة جديدة .

٤- مستوى التحليل : Analysis

التحليل هو تفتيت (تجزئة) معلومة إلى عناصرها أو أجزائها المكونة ، بحيث أن الترتيب الهرمي للأفكار يكون واضحاً وتكون المعلومات بين الأفكار المعبرة عنها صريحة.

والقصد من هذه التحليلات هو توضيح المعلومات لبيان كيفية تنظيم المعلومات ، والطريقة التي تصل بها تأثيرها ، وأسس ترتيبها . (بلوم ، ١٩٨٣ ، ٢٤١- ٢٦٨) .

٥- مستوى التركيب : Synthesis

ويقصد به التأليف بين العناصر والأجزاء بحيث تكون كلاً واحداً، ويتضمن ذلك ترتيب الوحدات والأجزاء والعناصر ، والربط بينها على نحو يؤلف بوضوح نمطاً لم يكن موجوداً من قبل.

٦- مستوى التقويم : Evaluation

ويقصد بالتقويم تلك العملية العقلية التي يصدر بها المتعلم أحكاماً حول قيمة المحتوى الذي يدرسه ، وهذه الأحكام قد تكون كمية أو كيفية لتحديد مدى ملاءمة المحتوى لمحكات معينة. (مجدي حبيب ،٦٠،١٩٩٦-٦١) .

المرحلة الثالثة : إعداد مواصفات الاختبار .

بعد تحديد النواتج التعليمية التي نتوقع تحقيقها من دراسة محتوى أو موضوع أو مادة دراسية معينة ، فإن المرحلة التالية هي إعداد قوائم بمواصفات النطاق السلوكي أو مواصفات مفردات الاختبار التي ستبنى لقياس هذا النطاق . فهذه المواصفات تحدد النطاق السلوكي الذي سيقيسه الاختبار تحديداً أكثر تفصيلاً من عبارات الأهداف السلوكية الإجرائية، كما أنها توضح ما يستطيع أن يؤديه الطالب وما لا يستطيع أن يؤديه ، كما أنها تساعد في بناء مجموعة متجانسة من مفردات الاختبار ، طبقاً للنطاق السلوكي الذي تحدده هذه المواصفات . (صلاح الدين علام ، ١٩٨٢ب ، ٢٩) .

وتتكون مواصفات الاختبار من أربع مراحل أساسية بالإضافة إلى ملحق المواصفات وهي كما يلي :

١- الوصف العام : General Description

وهو العنصر الأول من مجموعة مواصفات الاختبار ، ويشتمل على جملة أو جملتين عامتين تصفان ما يقيسه الاختبار . والهدف من هذا الوصف العام تقديم وصفاً موجزاً لفئة السلوك الذي ستحدده تماماً الخطوات الأخرى من المواصفات ، وأحياناً يكون هذا العنصر من مواصفات الاختبار محكي المرجع (الوصف العام) على شكل هدف سلوكي ، (Popham , 1978 , p 122) .

٢- عينة المفردة : Sample Item

العنصـــر الثانـــي مـــن مواصفات الاختبار هو عينة المفردة ، وهي تكمل الوصف لمعد الاختـــبار وربما تكون هذه المفردة مستعملة في الاختبار نفسه . والمفردات التوضيحية عادة تكون سهلة ، بينما يتكون الاختبار من مفردات قصيرة نسبياً.

وهناك سببان لعينة المفردة :

الســـبب الأول : بعـــض الـــناس يســـتعملون أجزاء من مواصفات الاختبار ، فهم ربما يعتقدون أنه يكفي أن يأخذوا عبارات الوصف العام للهدف والمثال التوضيحي فقط.

والســـبب الثانـــي : إعطاء تلميحات لهؤلاء الذين يكتبون مفردات الاختبار ، وهذا يجعل المفردة مطابقة للمفردات الأخرى التي يمكن بناؤها في ضوء هذه المواصفات .

وليس من الضروري تحديد الإجابة الصحيحة للمفردة. (Popham ,1978, p 123).

٣- خصائص المثير : Stimulus Attributes

ويعـــرفها Popham بأنها مجموعة من العبارات التي نحاول عن طريقها وصف وتحديد فـــرع ومادة المثير الذي يتم عرضه على الطالب ، وهذا يتطلب صياغة هذه العبارات بحيث يتحدد فيها خصائص المثيرات ، التي تشتمل عليها المفردة الاختبارية .

ومن أهم الشروط الواجب توافرها في خصائص المثير ، هو تحديد المحتوى الذي سوف تختار المفردات على أساسه . (محمد فتح الله ،١٩٩٥، ٣٢- ٣٣).

لذلـــك يعـــد هـــذا العنصـــر من أهم عناصر مواصفات الاختبار ، لأنه يحدد بالتفصيل خصائص المثير الذي سوف يقدم الممتحن الاستجابة عليه .

٤- خصائص الاستجابة Response Attributes

وهـــي العنصـــر الأخـــير مـــن عناصر مواصفات الاختبار ، وهي ترتبط ارتباطاً وثيقاً بخصائص المثير ، ونظراً لوجود نمطين من أنماط الاستجابة على المفردات وهما:الاستجابة المقيدة الاستجابة الحرة أو المفتوحة .

فإذا كان الاختبار يستلزم استجابة مقيدة ، فإن القواعد في تحديد طبيعة الإجابة الصحيحة لا تكفـــي بل يجب تحديد طبيعة الإجابة الخاطئة . فمثلاً مفردة الاختيار من متعدد تتكون من الإجابـــة الصحيحة والمشتتات Distractors ، والإجابة الصحيحة يجب أن تحدد بدقة وتبين

مواصـــفاتها ،وبعد ذلك المشتتات فلا يكفي أن تكون المشتتات إجابات خاطئة ولكن يجب أن تحدد بشكل واضح هذه الأخطاء ، بحيث لا تختلف المشتتات باختلاف كاتب المفردة.

أما تحديد مواصفات الاستجابات الحرة أو ذات الإجابة القصيرة ، فإنه يعد أكثر صعوبة وضـــخامة ، ويجـــب أن يشـــرح المحك الذي يجيز ثبات الكفايات للممتحنين على المفردة ، وتحديـــد محكـــات الإجابـــة الصـــحيحة وعدم استخدام عبارات غامضة أو غير محددة . (Popham , 1982 , p 23 - 24).

٥- ملحق المواصفات Specification Supplement

أحـــياناً تكـــون بعض المراحل من مواصفات الاختبار تتطلب فئات من المحتوى ، مثلاً سلاســـل مـــن القوانيـــن أو قائمة من الأشكال التاريخية الهامة ، فإذا وضعت مع خصائص المثير أو الاستجابة فإنها سوف تكون ضخمة ، لذلك يفضل سرد الخصائص الرئيسة فقط في تلـــك العناصـــر ، وتوضـــع التفاصيل التي تيسر كتابة المفردات الاختبارية في ملحق خاص يسمى ملحق المواصفات . (Popham , 1982 , p 24).

المرحلة الرابعة : كتابة مفردات الاختبار .

إن مرحلة بناء مفردات الاختبار ترتبط بشكل وثيق بالمراحل السابقة ، وبالتحديد بمرحلة إعداد المواصفات ، حيث يتم كتابة هذه المفردات طبقاً لمواصفات الاختبار التي تم إعدادها ، وهذه المرحلة تتطلب من الشخص الذي يكتب المفردات ، أن يكون على دراية تامة بمحتوى المـــادة الدراســـية ، أو النطاق السلوكي الذي تبنى المفردات لقياسه ، وكذلك يجب أن يكون على معرفة بأنواع المفردات التي تناسب أهداف هذا النطاق .

وعـــند كتابة مفردات الاختبار محكي المرجع يجب أن يراعي ابتكار المفردات التي تميز بيـــن التلامـــيذ الذيـــن حققوا الأهداف والنواتج التعليمية ، والذين لم يحققوا هذه الأهداف . (Popham , 1978 , p 147).

وبمـــا إن هـــناك نوعيـــن من المفردات تستخدم في قياس الجانب المعرفي والمهاري ، أحدهمـــا يتطلـــب إجابة مقيدة ، وآخر يتطلب إجابة حرة . ومن أمثلة النوع الأول ، مفردات الاختـــيار مـــن متعدد والصواب والخطأ والمزاوجة ، ومن أمثلة النوع الثاني ، أسئلة المقال والإجابة القصيرة والتكميلية . ويتطلب كتابة كل نوع من هذه الأنواع مراعاة بعض الشروط الفنية ، وبخاصة المفردات التي تتطلب إجابة مقيدة لقياس المهارات المعرفية العليا . وينبغي

تدقـيق المفردات عقب الانتهاء من كتابتها ، بواسطة مجموعة من خبراء المجال الذي يقيسه الاختبار ، وخبراء القياس والتقويم . (صلاح الدين علام ، ٢٠٠٠ ، ٣٢٢) .

وسـوف يقـوم الباحث بعرض نوع واحد فقط من المفردات ذات الإجابة المقيدة ، وهو مفردات الاختيار من متعدد ، وذلك لأن الاختبار في الدراسة الحالية تكون من هذا النوع من المفردات .

مفردات الاختيار من متعدد :

إن مفـردات الاختـيار من متعدد هي المفردات التي تحتوي على استجابتين أو أكثر ، واحـدة مـنها فقـط صحيحة ، أو محددة بدقة أكثر من الاستجابات الباقية ، والأخرى غير صـحيحة . أو قـد تكـون الاسـتجابات كلهـا صحيحة ، وواحدة فقط غير صحيحة يطلب اختيارها. (Stanley , 1978 , p 232) .

وتعتـبر أسئلة الاختيار من متعدد أكثر الأسئلة الموضوعية شيوعاً ،ويتكون السؤال من مشـكلة (قـد تصـاغ علـى صورة سؤال مباشر أو عبارة ناقصة) تسمى الجذر Stem ، وقائمة من الحلول المقترحة (قد تشتمل على كلمات أو أعداد أو رموز أو عبارات) وتسمى الاختيارات Alternatives .

ويطلـب مـن المفحوص قراءة جذر السؤال وقائمة البدائل ، وانتقاء البديل الصحيح أو الأفضـل . والـبديل الصحيح في كل سؤال يسمى ببساطة الإجابة Answer ، بينما تسمى الـبدائل الاختـيارية المحولات (المشتتات) Distractors . (فؤاد أبو حطب وآخرون ، ١٩٩٩ ، ٣٩٧ – ٣٩٨) .

وأهم الأسس التي تتبع في كتابة مفردات الاختيار من متعدد:

- يجـب أن تكتب المفردة بحيث تطرح مشكلة واضحة ومحددة تماماً . وهذا يعني أن الجـذر يجب أن يصاغ بحيث يفهم المتعلم منه السؤال المطروح ، دون قراءة بدائل الإجابة الموجودة في المفردة .
- يجـب أن تحتوي المقدمة على الجزء الأكبر من السؤال ، وأن نجعل البدائل قصيرة ما أمكن .
- يجب تجنب استخدام صيغة النفي في مقدمة مفردة الاختيار من متعدد .
- يجـب أن لا يوجد بين البدائل المعطاة في الفقرة إلا بديل واحد ، أو إجابة واحدة فقط هي الصحيحة ، أو هي الأفضل من بين البدائل المعطاة في الفقرة .

- يجب التأكد من خلو المفردة من أي إشارة ، أو تلميح غير مقصود يمكن أن يدل على الإجابة الصحيحة .
- عدم استخدام عبارات من نوع " كل ما ذكر صحيح " أو " كل ما ذكر أعلاه " أو" جميع ما سبق ذكره " ، أو ما شابه ذلك في البدائل . وعند استخدام عبارات مثل " جميع ما ذكر خطأ " أو " لا شيء مما سبق ذكره " ، فيجب أن يكون بحرص شديد وتحت شروط معينة . (صلاح أبو ناهية ، ١٩٩٦ ، ٢٢٤- ٢٣٨) .
- يجب عدم وضع رمز الإجابة الصحيحة بطريقة تمكن المفحوص من التخمين ، أو الحذر بأن باقي الأسئلة تأتي على نفس الخط أو الأسلوب . كأن يضع البديل (ب) هو الإجابة للمفردة (١) ، والبديل (ج) هو الإجابة للمفردة (٢) وهكذا .....
- يفضل توزيع الإجابات الصحيحة بنسب متوازية بين البدائل ، بحيث لا تطغى واحدة على الأخرى .
- على الفاحص أن لا يضع جميع المفحوصين في درجة واحدة ، من حيث تعين الإجابة الصحيحة ، بل تبعاً لمستواهم العقلي والزمني . فمثلاً تلميذ الصف الرابع الابتدائي يضع إشارة على الإجابة الصحيحة ، ولا يكتب على ورقة أخرى . (توما، ١٩٩١ ، ٨٥ - ٨٦) .

المرحلة الخامسة : تحديد صدق الاختبار وثبات درجاته .

بعد الانتهاء من كتابة مفردات الاختبار يجب أن تعمل هذه المفردات كوحدة متسقة ، لتشكل في مجملها الاختبار الذي هو أداة القياس ، وحتى يكون الاختبار صالحاً للغرض الذي بني من أجله وتكون قراراته صائبة يجب أن تتحقق بعض الخصائص الأخرى ، وهي صدق الاختبار وثبات درجاته . وسوف يتناولها الباحث فيما يلي .

❋ صدق الاختبارات المحكية المرجع : Validity

يشير تايلر Tyler ١٩٨٣ إلى أن أهم ما يؤخذ في الاعتبار هو الصدق الذي يتصل بإجابة التساؤل " عما يقيسه الاختبار ؟. وإذا لم تكن لدينا إجابة كافية بدرجة معقولة فإن الاختبار يصبح عديم الفائدة ، بل قد يؤدي تطبيقه إلى أخطاء جسيمة تؤدي بدورها إلى قرارات خاطئة .(محمود إبراهيم ، ١٩٩٠ ، ٣٥) .

لذلك يعتبر الصدق الخاصية الأساسية التي يجب أن تتوافر في وسيلة القياس بصفة عامة والاختبار بصفة خاصة ، بل أن حساب صدق الاختبار من أفضل الأساليب التي تستخدم

لتقيـيم فاعلـية الاختبار بوصفه أداة قياس ، والصدق بتعبير بسيط هو أن يقيس الاختبار ما وضـع لقياسه . أي يقيس فعلاً الوظيفة التي يفترض أن يقيسها . (صلاح أبو ناهية ،١٩٩٤، ٣٣٦)

ويمكـن القول أن جميع طرق تحديد صدق الاختبارات تبحث عن العلاقة التي تربط أداء الأفـراد فـي الاختـبار ، وبعض الحقائق والمعلومات التي تتصل بالعملية النفسية موضوع الاختبار . والتي (أي هذه المعلومات) يمكن ملاحظتها ملاحظة مستقلة . (فؤاد أبو حطب، ١٩٩٩ ، ١٣٣) .وتوجد ثلاث طرق لتحديد صدق الاختبار محكي المرجع وهي :

١- الصدق الوصفي Descriptive Validity :

يعتـبر الاختـبار محكي المرجع صادقاً بدرجة دقته في قياس فئة عناصر سلوك المتعلم التي وردت أو جاءت في تعريف المجال وتحديده . ولما كان هذا النمط من الصدق يعتمد أو يـتوقف علـى الاختبار من حيث قدراته الوصفية . بمعنى هل يمكن تفسير أداء المتعلم بدقة علـى أساس وصف المجال أم لا ؟ فإننا نستطيع أن نطلق على هذا النوع من الصدق صدقاً وصفياً . (جابر عبد الحميد ، ١٩٩٦ ، ٢٦٣) .

ويمكـن تقديـر الصدق الوصفي بالطرق التي تعتمد على أحكام الخبراء للتحقق مما إذا كانـت مواصفات النطاق محددة تحديداً تفصيلاً واضحاً ودقيقاً ، وإذا ما كان هناك اتفاق على أن مجموعـة المفـردات التي يشتمل عليها الاختبار تعد بمثابة عينة كافية ، وممثلة للنطاق السلوكي الذي يقيسه الاختبار . (صلاح الدين علام ، ٢٠٠١ ، ٢٨٢) .

ويرى بابام Popham أنه يمكن التحقق من الصدق الوصفي للاختبار محكي المرجع إذا كانت مواصفات الاختبار تقوم بالوظيفتين التاليتين :

أ- أن تكون واضحة بحيث يمكن استخدامها في تفسير أداء الفرد .

ب-أن تكـون فـي مواصفات الاختبار سهلة الاستخدام في بناء مفردات ملائمة للنطاق المقاس . (Popham , 1978 , p 156) .

وبالتالـي يؤسس الصدق الوصفي بواسطة حكم الأفراد لدرجة تحديد المجال بدقة لطبيعة مجموعـة من مفردات الاختبار ودرجة مطابقتها مع وصف أو تحديد المجال . ولذلك يسمي واضـعو الاختـبارات محكية المرجع تسمية هذا النوع من الصدق ، بصدق المحتوى نظراً للعلاقة بين المفردات والموضوعات المحددة .(نادية عبد السلام ، ١٩٨٧ ، ١٤٤) .

إلا أنه يفضل تسمية هذا الصدق ، بالصدق الوصفي لأن الاختبارات محكية المرجع تختص بأشياء أخرى عن مجرد المحتوى . فمن الممكن أن نبني أيضاً اختباراً محكي المرجع لقياس سلوك نفسي حركي أو وجداني . في مثل هذه الحالات يبدو أن مصطلح صدق المحتوى يكون غير مناسب . أيضاً بسبب الانطباق الكبير بين تحديد المجال والمفردات ، فإن مصطلح الصدق الوصفي سوف يخدم أهدافنا أفضل عندما نحاول الحصول على صدق الاختبارات محكية المرجع . (Popham , 1978 , p 155) .

٢- الصدق الوظيفي Functional Validity :

ويقصد بالصدق الوظيفي أن يؤدي الاختبار محكي المرجع الوظيفة التي صمم من أجلها أو الغرض الذي بني من أجله ، ولا يقتصر فقط على أداء الفرد في النطاق السلوكي الذي يقيسه الاختبار . (صلاح الدين علام ، ٢٠٠٢ ، ٢٨٢) .

فأحياناً يواجه المربون مواقف يرغبون فيها اختبار إنجاز الفرد لوظيفة ما ، وليس فقط شرح أو وصف أدائه بدقة ، في مثل هذه الحالات يمكن أن نراجع الصدق الوظيفي للاختبار. (Popham , 1978 , p 159) .

٣- صدق اختيار النطاق السلوكي Domain-Selection - Validity :

يقوم هذا النوع من أنواع الصدق على دقة اختيار المجال . وهذا يتطلب من واضع الاختبار تحديد مجال معين للاختبار يحقق خدمة دقيقة وصحيحة ويصلح كمؤشر لبعد أكثر عمومية صمم أو وضع ليعكسه . (مجدي حبيب ، ١٩٩٦ ، ٣٨٦) .

ويرى علام (١٩٨٦) أن صدق النطاق السلوكي يرتبط بمفهوم إمكانية التعميم الذي نادى به كرونباك Cronbach (١٩٧٣) في نظريته التي تناولت هذا المفهوم Generalizability Theory . فالنطاق السلوكي الذي نختاره من بين السلوكيات الممكنة يجب أن يسمح لنا بالتعميم عل النطاق الشامل للبعد الذي يقيسه الاختبار .

وقد اقترح بابام Popham طريقتين يمكن باستخدام أي منهما للتوصل إلى أدلة عن صدق انتقاء النطاق السلوكي للاختبار هما :

أولاً : القيام بتدريس المعارف والمهارات التي يتضمنها كل نطاق من النطاقات السلوكية المطلوبة لمجموعات صغيرة من الطلاب حتى يصلوا إلى درجة التمكن ، ثم اختيار إمكانية تعميم إتقان نطاق منها على النطاقات الأخرى .

ثانـــياً : الموازنـــة بين مجموعتين من الطلاب في عدد من الاختبارات ، تقيس المعارف والمهارات التي يتضمنها كل نطاق من النطاقات السلوكية المطلوبة ، ويجب أن تكون إحدى المجموعتيـــن قـــد تلقت التعليم المتعلق بهذه النطاقات ، والأخرى لم تتلق التعليم . ثم اختيار النطاق السلوكي الذي يميز بين هاتين المجموعتين .(صلاح الدين علام ، ٢٠٠١ ، ٢٨٥).

❋ ثبات الاختبارات محكية المرجع : Reliability

يعتـــبر الثـــبات الخاصية الأساسية الثانية التي يجب أن تتصف بها أداة القياس الجيدة . فالثـــبات يعـــنى بالاتساق في نتائج القياس على نفس الأداة أو الاختبار . فإذا كان الصدق هو أن يقـــيس ما وضع الاختبار لقياسه فإن الثبات هو الاتساق في نتائج هذا الاختبار عن تطبيقه من وقت لآخر . (صلاح أبو ناهية ، ١٩٩٤ ، ٥١) .

ويعـــرف فـــؤاد أبـــو حطـــب (١٩٩٩ ، ١٠١) الثبات بأنه دقة الاختبار في القياس أو الملاحظة وعدم تناقضه مع نفسه ، واتساقه واطراده فيما يزودنا به من معلومات عن سلوك القياسات أو الملاحظات التي يحصل عليها .

ومع بداية ظهور فكرة الاختبار محكي المرجع وجد بابام Popham وهاسك Husek أن الاتســـاق الداخلي والاستقرار، ربما تعتبر خواص هامة لدرجات الاختبار التي نحصل عليها مـــن مقـــياس محكـــي المرجع . ولكن المعاملات التي توصف بواسطة النظرية الكلاسيكية للاختـــبار . لتقدير هذه الخواص ربما تكون غير مناسبة ، لأن درجة ثبات الاختبار لمجموعة مـــن المختبريـــن تعتمد على إمكانية تغير درجة الاختبار . وبما أن درجات الاختبار محكي المرجع شائع ملاحظة توزيعات متجانسة فيها . لذلك يتوقع الحصول على تقديرات كلاسيكية منخفضة للثبات . (Popham & Husek , 1969 , p 5) .

ويـــرى هامبلـــتون ونوفـــيك Hambleton & Novick (1973) أن الطرق الكلاسيكية لـــتقدير ثبات وصدق الاختبارات محكية المرجع يجب أن تستبعد أو تفسر بحذر شديد .و هذه المفاهـــيم الســـيكومترية التـــي تستخدم الطريقة الارتباطية غير مناسبة . (Hambleton , Swaminathan , Algina , Coulson , 1978 , p 15)

ومـــع تطور الاختبارات محكية المرجع حدث تحول في منهجية تقدير ثبات درجات هذه الاختـــبارات بما يتفق مع الاستخدامات المناسبة لها ، حيث أن هذه الاختبارات تستخدم عادة فـــي تصـــنيف الأفراد إلى مجموعتين أو أكثر بحسب درجة تمكنهم من المعارف والمهارات التـــي يشتمل عليها نطاق سلوكي معين ، لذلك اعتبر علماء القياس التربوي أن ثبات درجات

هـذا النوع من الاختبارات يعني اتساق قرارات التصنيف بناءً على هذه الدرجات . (صلاح الديـن عـلام ، ٢٠٠١ ، ٢٨٧) ، وتوصل عدد من العلماء إلى طرق مختلفة لتقدير الثبات بهـذا المفهوم الجديد ، منها ما يعتمد على تطبيق الاختبار مرة واحدة ، ومنها ما يعتمد على تطبيق الاختبار مرتين بفاصل زمني مناسب أو يطبق اختبار واختبار موازي له .

وسوف يستعرض الباحث فيما يلي بعض هذه الطرق وهي :

أ- طرق تتطلب تطبيق الاختبار مرتين :

١- معامل كارفر Carver :

اقـترح كارفـر Carver طريقتيـن لتقدير ثبات الاختبارات محكية المرجع . والطريقة الأولـى تتطلـب تطبـيق نفس الاختبار على مجموعتين متقابلتين ، ومقارنة النسب المئوي للممتحنيـن الذيـن صـنفوا كمتقنين . والطريقة الثانية تتطلب اختبارين متوازيين على نفس المجموعـة ، ومقارنـة النسـب المـئوية للإتقـان علـى الاختباريـن . (Hambleton , Swaminathan , Algina , Coulson , 1978 , p 20 ; Subkoviak , 1982 , p131)

ويحسب معامل كارفر من العلاقة التالية :

$$\text{معامل الثبات} = \frac{أ + د}{ن}$$

أ = عدد الطلاب الذين أتقنوا في مرتي التطبيق (أي متقن / متقن) .

د = عدد الطلاب الذين لم يتقنوا في مرتي التطبيق (أي غير متقن / غير متقن) .

ن = العدد الكلي للطلاب .

٢- معامل هامبلتون ونوفيك Hambleton & Novick :

اقـترح كـل من هامبلتون ونوفيك Hambleton & Novick 1973 أن ثبات قرارات تصـنيف الإتقـان سـوف يحـدد بالنسبة إلى اتساق القرارات من تطبيقين لنفس الاختبار أو الصـور المـتوازية له . فعـندما يتم تصنيف المختبرين إلى ك مستوى (أي يتم تقسيم مدى درجـات الاختـبار إلى ك من المستويات) فإن معامل الثبات الذي اقترح هامبلتون ونوفيك يعطى بالعلاقة :

$$ل. = \text{مجـ}_{ي=1}^{ك} \; ل_{تت}$$

حيث أن ل .هي معامل الثبات ، ك هي عدد مستويات التصنيف (أو التمكن) .

لن ت هـي نسـبة عدد الممتحنين المصنفين في مستوى واحد من مستويات الإتقان على التطبيقيـن . (ناديـة عـبد السـلام ١٩٨٧،١٣٠ (Hambleton , Swaminathan , Algina , Coulson ,1978, 20 - 21 ;

٣- معامل كابا لسواميناثان وهامبلتون وألجاينا :

بسـبب عدم أخذ معامل ل. لنسبة الإتقان التي قد توجد بالصدفة والتي تؤدي بدورها إلى إعطـاء انطـباع مـزيف لمستخدمي مدى اتساق تصنيف الإتقان . لذلك اقترح سواميناثان وهامبلتون وألجانيا استخدام معامل كابا كوهين Cohen 1960 كمعامل ثبات وصيغته :

$$\text{معامل كابا (K)} = \frac{ل_{ء} - ل_{م}}{١ - ل_{م}}$$

حيث أن ل٠ نسبة الإتقان الملاحظ في التصنيفات ، ل. = $\text{مجـ}_{ي=1}^{ك}$ ل ت ت

ل م = أي المجموع لحاصل ضرب النسب لها قيمة الأعداد والأفراد الذين يتم تصنيفهم في المجموعة ت نفسها من مستويات التمكن في مرتي التطبيق .

ونلاحـظ أن أكبر قيمة له هو +١ والحد الأدنى هو -١ . (نادية عبد السلام ، ١٩٨٧ ، ١٣١) ; (Hambleton , Swaminathan , Algina , Coulson , p 21 ; Huynh , 1976 , p 253)

٤- طريقة ليندمان وميريندا Lindeman & Merenda , 1979 :

بمـا أن الهدف من بناء الاختبار محكي المرجع هو وضع الفرد في إحدى مجموعتين . الأولـى تضـم الطلاب الذين حققوا مستوى المحك (متقن) للمهارة ، والمجموعة الثانية تضم الطـلاب الذيـن لم يحققوا هذا المحك (غير متقن) وبالتالي لا يصلح تطبيق الارتباط التقليدي في ثبات الاختبارات محكية المرجع . لذلك اقترح كل من ليندمان وميرايندا Lindeman & Merenda , 1979 المعادلة التالية لحساب الثبات :

$$\text{معامل الثبات} = \frac{ن\ ن_{.} - ن_{١}\ ن_{٢}}{ن\ ن_{.} + ل(ن + ن_{.} + ل)}$$

حيث أن : ن. = عدد الممتحنين الذين لم يصلوا إلى مستوى المحك في مرتي التطبيق معاً

ن = عدد الممتحنين الذين بلغوا المحك في المرتين معاً .

ن١ = عدد الممتحنين الذين بلغوا المحك في المرة الأولى فقط . (أي لم يبلغوا المحك في المرة الثانية) .

obeikandl.com

ن٢ = عدد الممتحنين الذين بلغوا المحك في المرة الثانية فقط . (أي لم يبلغوا المحك في المرة الأولى) .

ل هي أصغر القيمتين ن ، ن .

مـــثال : افترض أننا طبقنا اختبار مرتين على ١٠٠ طالب ، ن = ٧٥ عدد الأفراد الذين حققوا المحك في المرتين ، ن. = ٥ عدد الذين لم يحققوا مستوى الاختبار في المرتين ، ن١= ١٥ عـــدد الذين حققوا المحك في المرة الأولى فقط ، ن٢= ٥ عدد الأفراد الذين حققوا المحك في المرة الثانية فقط . ومنه ل هي أصغر العددين (أصغر القيمتين ن١، ن.) ⇐ ل = ٥

معـــامل الثبات = $\frac{5\times15-5\times75}{(5+5+15)\times5+5\times75}=\frac{300}{800}=0.375$ ومنه الثبات هنا منخفض .

(Aiken , 1985 , p92)

ب- طرق تتطلب تطبيق الاختبار مرة واحدة :

أ- طريقة ليفنجستون Livingston Method :

اقـــترح ليفنجســـتون Livingston 1972 معادلة لتقدير ثبات الاختبارات محكية المرجع وهـــذه المعادلة هي في الحقيقة تعميم لتقدير الثبات طبقاً لنظرية الثبات الكلاسيكية (وذلك من أجل حالتي إتقان) ، والصيغة الرياضية لهذه المعادلة :

$$ر_{م\,م}=\frac{ر_{س\,س}\,ع_{ك}^{2}+(م-د)^{2}}{ع_{ك}^{2}+(م-د)^{2}}$$

حيـــث أن : ر م م= معـــامل ثـــبات الاختبار محي المرجع Criterion - Referenced Reliability .

ر س س= معـــامل ثـــبات الاختـــبار مقدراً بأية طريقة أخرى من طرق تقدير الثبات طبقاً لنظرية الثبات الكلاسيكية .

م = متوسط درجات الأفراد على النطاق السلوكي .

د = درجة المحك Criterion Score .

ويتضـــح مـــن هذه المعادلة أنه عندما د = م فإن تقدير قيمة معامل ثبات الاختبار محكي المـــرجع باستخدام هذه المعادلة يساوي تقدير قيمة معامل ثبات الاختبار محكي المرجع مقدراً بأيـــة طـــريقة أخـــرى مـــن طـــرق تقديـــر الثبات طبقاً للنظرية الكلاسيكية .(علي ماهر ، ٢٠٠١،٢٩٥- ٢٩٦ ; Subkoviak , 1978 , p265) .

٢- معامل الاعتمادية لبرينان وكين Brennan & Kane Dependability Coefficient

تم استخدام نظرية إمكانية التعميم لتعريف نوعين من معاملات الاعتمادية في الاختبارات محكية المرجع ، وهذان المعاملان يمكن أن يعطيان تفسيرات مشابهة لنظرية الثبات الكلاسيكية ومعاملات قابلية التعميم . وبما أن معاملات الاعتمادية للاختبارات محكية المرجع مختلفة تماماً عن مثيلاتها من المعاملات في الاختبارات معيارية المرجع فإن معاملات الاعتمادية المتعلقة بالاختبارات محكية المرجع تسمى بمعامل الاعتمادية Indices Dependability أما معاملات معيارية المرجع فتسمى معاملات إمكانية التعميم Genralizability Coefficients . (Brennan , 1982 , p201) .ويعطى معامل الاعتمادية م(س.) بالعلاقة:

$$م(س_{.}) = \frac{ع^{2}(ف) + (س - س_{.})^{2}}{ع^{2}(ف) + (س - س_{.})^{2} + ع^{2}(\Delta)}$$

حيث أن ع^{٢}(ف) تباين الدرجة الشاملة عن درجة القطع س.

و ع^{٢}(Δ) = الخطأ المرتبط بتفسير درجات محكية المرجع .

$\overline{س}$ ترمز إلى المتوسط العام لدرجات الأفراد في النطاق الشامل للمفردات .

وعندما $\overline{س}$ =س نحصل على معامل الاعتمادية بالمفهوم الكلاسيكي للثبات. (صلاح الدين علام ، ٢٠٠١ ، ٢٩٣- ٢٩٥) .

٣- طريقة صبكوفياك Subkoviak :

اقترح صبكوفياك طريقته التي تعتمد على اتساق تصنيف الأفراد إلى مجموعتين إحداهما متقنة والأخرى غير متقنة باستخدام نتائج تطبيق الاختبار مرة واحدة فقط ، وهي تتلخص في تقدير مقدار لاحتمال تصنيف كل فرد تصنيفاً صحيحاً ، وجمع هذه التقديرات الاحتمالية وإيجاد متوسطها وبذلك نحصل على معامل إتقان خاص بالمجموعة التي يطبق عليها الاختبار .(Subkoviak , 1982 , p143) .

❋معامل الحساسية للتعليم

Instructional Sensitivity Index

بعد الانتهاء من كتابة مفردات الاختبار يفضل أن تخضع لنوعين من التحليلات وهما : التحليل المنطقي والتحليل الإحصائي .

التحليل المنطقي Logical Review :

ويهدف لتحديد مدى مطابقة المفردة بالهدف الذي أعدت لقياسه . وهذه المطابقة تتم من قبل المحكمين حيث يسأل أختصاصيي المحتوى لتحديد العلاقة المنطقية بين الأهداف أو مواصفات المجال ومفردات الاختبار ، وإجراء كهذا دليل على صدق الاختبار أيضاً . ولقد قدم روفينلي وهامبلتون Rovinell & Hambleton 1977 طريقة للتحقق من ذلك (وسوف يقدم الباحث هذه الطريقة في الفصل الرابع عند بناء أداة الدراسة) .

التحليل الإحصائي Statistical Review :

بعد التأكد من مطابقة المفردات للأهداف السلوكية استرشاداً بآراء الحكام (المحكمين) يتم تطبيق الاختبار لجمع البيانات ، ويتم تحليلها بالطرق الإحصائية وذلك لتقييم فاعلية المفردات في ضوء ما هو متوقع منها وذلك من خلال تحديد العلاقة بين ما تقيسه المفردة وبين استجابات الأفراد لها ، بهدف تحديد المفردات التي لا تكشف عن التعليم ، أي تحديد المفردات التي تحتاج إلى إعادة نظر فيها ، وليس الهدف من تحليل المفردة التجريبي (الإحصائي) هو اختيار المفردات على أساس إحصائي ولكن لتحسين المفردات (صلاح علام ،١٧٦، ٢٠٠١) (Berk, 1982 , p 51- 52 ; Haladyna &Roid, 1981, p 39).

وبما أن الاختبارات معيارية المرجع تركز على ترتيب الأفراد تبعاً لمستواهم ، وعليه فإن الاختبارات معيارية المرجع تقوم على مبدأ الفروق الفردية بين الأفراد وينعكس ذلك على معايير الجودة للمفردة أو الاختبار بشكل عام . إذ أن جودة الاختبار _ أو المفردة_ تقاس بمدى قدرته على التمييز بين المستويات المختلفة من القدرة لدى الأشخاص . (عبد الله القاضي ، ١٩٩٣ ، ١٠٦) ولمضاعفة قوة الاختبار التمييزية تحذف المفردات السهلة والصعبة (نادية شريف ، ١٩٨٧، ١٢٣) أي أن تحليل المفردة بالمفهوم التقليدي يركز على فكرتي الصعوبة والتمييز للمفردات فالمفردة المعتدلة الصعوبة والتمييز تكون مرغوبة في هذا النوع (Haladyna & Roid , 1981, p 39) .

ولكن في الاختبارات محكية المرجع حذف المفررات السهلة أو الصعبة لن يقلل فقط من صلاحية الاستخدام للاختبارات محكية المرجع ، ولكنه أيضاً سوف يهدم سمة مجالية المرجع، بمعنى أن الاختبار سوف لن يمثل مجموعة فرعية من مجال المفردات التي قصد أصلاً أن يمثلها . (نادية شريف ، ١٩٨٧ ، ١٢٣) . وبما أن الاختبارات محكية المرجع تقوم على مبدأ الإتقان أو المعرفة لمهارة أو مفاهيم محددة ، وهي بهذا لا تركز على مفهوم الفروق بين الأشخاص ، فالكل يفترض أن يصل إلى الحد الأدنى من المعرفة أو المهارة ، وعليه فإن الفروق تتلاشى حينما يجيب الأفراد على جميع مفردات الاختبار أي عندما يصلوا إلى مرحلة الإتقان ، وبتلاشي الفروق بين الأفراد يصبح من الصعب الحكم على صلاحية الاختبار أو بنوده في ضوء معايير تركز على مبدأ الفروق الفردية كما هو الحال في الاختبارات معيارية المرجع (عبد الله القاضي ، ١٩٩٣ ، ١٠٦) .

معامل الحساسية للتعليم ومعامل التمييز:

مما تقدم نجد أن إجراءات تحليل المفردات للاختبارات محكية المرجع تختلف عن إجراءات تحليل مفردات الاختبارات معيارية المرجع ، وذلك لأن الاختبارات محكية المرجع تهدف إلى وصف أداء المتعلمين بالنسبة لأنواع الأداء الذي يمكنهم القيام به ، ولا تهدف إلى ترتيبهم وفقاً لدرجاتهم بين أقرانهم . ولذلك فإن مؤشرات الصعوبة والتمييز ليس لها معنى هنا ، ويمكن القول أنه كلما اتصفت مؤشرات الصعوبة في الاختبارات محكية المرجع بالارتفاع (أي ارتفاع نسبة الناجحين) كان معناه أن التدريس تدريس فعال .(رجاء أبو علام، ١٩٨٧ ، ٢٣٨- ٢٣٩)

أي إذا كان الاختبار حساساً للموضوعات التعليمية وكان التعليم فعالاً فإن غالبية المتعلمين سوف يحصلون على درجات منخفضة قبل التعليم ودرجات مرتفعة بعد التعليم (Kubiszyn, 2000 , p 147)

وبمعنى آخر يجب أن تعكس الاختبارات محكية المرجع التأثيرات المتوقعة للتعليم على تعلم الطالب ، وهذه التأثيرات المتوقعة غالباً ما تكون معرفة عملياً في مصطلح الأهداف التعليمية . وبشكل عام أداء الطلاب غير المتعلمين يكون عند مستوى منخفض قبل التعليم ، وعند مستوى مرتفع بعد التعليم والاختلاف في الدرجات قبل التعليم وبعد التعليم مؤشر للفاعلية التعليمية . وتسمى هذه الفاعلية معامل الحساسية للتعليم Instructional Sensitivity Index وليس معامل صعوبة وتمييز (Haladyna& Roid, 1981, p 39).

ويرى بيرك (Berk , 1982 , p 59) أن معامل التمييز للمفردة في الاختبارات محكية المرجع يقيس تغير الأداء بين الاختبار القبلي – والبعدي Pretest - Posttest أو يبين الاختلاف بين المتعلمين وغير المتعلمين بين المجموعات المحكية ، وهذا الاختلاف يرد إلى أثر التعليم . وقد حث تفسير المعامل في البرامج التعليمية بعض المؤيدين لهذه التقنية أن يسموه معامل الحساسية للتعليم Instructional Sensitivity Index .

إلا أن كوكس وفارجاس أكدا في دراسة لهما أن الارتباط بين معامل الحساسية للتعليم ومعامل التمييز المعروف (المجموعات العليا والدنيا التي تأخذ ٢٧ ٪) كانت بحدود ٠,٣٧ أي أن الارتباط بينهما ضعيف ، وأكدا أيضاً أنه ليس بالضرورة أن يكون هناك ارتباط بين معامل الحساسية للتعليم ومعامل التمييز . (Popham , 1978, 108) .

ويرى رجاء أبو علام ١٩٨٧ أن هناك سؤالاً حيوياً في تقويم الاختبار محكي المرجع وهو : لأي حد تقيس البنود أهداف التدريس ؟ . ورأى أن إجابة هذا السؤال هي مؤشر لفاعلية البند والتي سماها الحساسية لفاعلية التدريس Teaching Sensitivity. (رجاء أبو علام ، ١٩٨٧ ، ٢٤٠)

ولقد أشار إلى وجود فرق جوهري بين معامل الحساسية ومعامل التمييز ، بأن معامل الحساسية للتعليم يعكس فاعلية مقصودة للتدريس من تطبيقين للاختبار . بينما معامل التمييز في المقاييس المعيارية المرجع يشير إلى الدرجة التي تمييز فيها المفردة بين مجموعتين عليا ودنيا من المتعلمين على نفس الاختبار . (رجاء أبو علام ، ١٩٨٧ ، ٢٤٢) .

وقد وجد الباحث من خلال إطلاعه أن مصطلح معامل الحساسية للتعليم استخدم بمعنى معامل التمييز للاختبارات محكية المرجع ، أي أن بعض علماء القياس استخدم معامل التمييز في الاختبارات محكية المرجع والبعض الآخر استخدم معامل الحساسية (أمثال Popham و Haladyna و Perkins وصلاح علام ، ورجاء أبو علام) ويرى الباحث أن معامل الحساسية للتعليم هو الأفضل لأن معامل التمييز أولاً يرتبط بالاختبارات معيارية المرجع ، وثانياً تتم المقارنة (حسابه) فيه بين المجموعات العليا والدنيا ، حيث يتم تقسيم الطلاب إلى مجموعتين حسب درجاتهم الكلية على الاختبار ، المجموعة العليا تضم ٢٧ ٪ من الطلاب الذين حققوا أعلى الدرجات على الاختبار ، والمجموعة الدنيا تضم ٢٧٪ من الطلاب الذين حققوا أدنى الدرجات على الاختبار ، وإذا كان التعليم فعالاً وطبق الاختبار فإن معامل التمييز سوف يكون معدوماً بمفهومه المعلوم ، أي أن هذه المفردة تلغى وهذا غير ممكن في الاختبارات محكية المرجع .

تعريف معامل الحساسية للتعليم:

- **عرف هالدينا** (Haladyna & Roid , 1981 , p 40) معامل الحساسية للتعليم بأنه ميل المفردة لتغير مستوى الصعوبة كدالة للتعليم . فعندما يكون التعليم فعالاً (أي بدون شك) فإن المفردة المقدمة للطلاب غير المتعلمين يجب أن تكون صعبة ، ونفس المفردة عندما تقدم لطلاب متعلمين يجب أن تكون سهلة ، والمفردات التي لا تكشف هذا التغير يجب أن تراجع وتدقق .
- **و يتفق معه بيركينز** (Perkins , 1984) بأن معامل الحساسية للتعليم هو ميل مفردة الاختبار للتغير في الصعوبة كدالة للتعليم .
- **و يعرف رجاء أبو علام** (رجاء أبو علام ، ١٩٨٧ ، ٢٤٢) معامل الحساسية للتعليم بأنه يشير إلى الدرجة التي يعكس بها البند الفاعلية المقصودة للتدريس ، كما يعكس الفرق بين إجرائيين للاختبار قبلي – وبعدي .
- **و يعرف صلاح الدين علام** (صلاح الدين علام ، ٢٠٠١ ، ١٨١– ١٨٤) معامل الحساسية للتعليم بأنه من المقاييس الإحصائية التي يمكن استخدامها في إيجاد معامل تمييز مفردة الاختبار محكي المرجع ، ولكن نظراً لتأثير هذا العامل بفاعلية التعليم يسمى معامل الحساسية للتعليم .
- و يرى الباحث أن : **معامل الحساسية للتعليم : هو أسلوب إحصائي يستخدم لتحليل مفردات الاختبار محكي المرجع بهدف الكشف عن المفردات التي تعكس أثر المعالجة التعليمية وذلك من خلال ما تقيسه المفردة واستجابة الأفراد لها ،** فالمفردة ذات الحساسية المرتفعة للتعليم هي المفردة التي يجيب عليها كل الطلاب إجابة خاطئة قبل التعليم وإجابة صحيحة بعد التعليم والمفردة التي تحقق ذلك تكون مفردة مثالية .

وقبل أن يقدم الباحث بعض الطرق التي تستخدم لقياس معامل الحساسية للتعليم ، فإنه سوف يقدم أنواع المجموعات التي تعتمد عليها هذه المعاملات ، فبعد التأكد من مطابقة المفردة للأهداف السلوكية استرشاداً بآراء الحكام يتم تطبيق الاختبار على مجموعات من الطلاب ، والبيانات التي نحصل عليها من الطلاب يمكن أن تستخدم لتقييم فاعلية المفردات في ضوء ما هو متوقع منها ، واختيار عينات الأفراد يطلق عليه مجموعات المحك

Criterion Groups . ((Berk , 1982 , p 52) .ويوجد نوعان من هذه المجموعات تعتمد عليها وهي :

أ- طريقة القياس القبلي – البعدي للتعليم

: Preinstruction-Postinstruction Measurement Method

وتعتمد هذه الطريقة على اختبار مجموعة واحدة من الطلاب مرتين قبل التعليم Preinstruction وبعد التعليم Postinstruction ، حيث يختبر الطلاب بنفس المجموعة من المفردات وتحت نفس الظروف في مرتي التطبيق أو يتم اختبارهم بمجموعتين متوازيتين أو متكافئتين من المفردات . والقياس للطلاب قبل التعليم وبعده يفيد المعلم في تحديد ما حققه الطلاب من تقدم ، وتعطي معلومات مفصلة لتقييم البرامج التعليمية .

٢- طريقة قياس أداء مجموعتين مستقلتين (متعلمة وغير متعلمة)

: Uninstructed - Instructed Groups Approaches

وهي طريقة مستخدمة منذ فترة طويلة ، وهي تتطلب اختيار مجموعتين منفصلتين من الأفراد في وقت واحد إحداهما لم تتلق التعليم (المجموعة غير المتعلمة The Uninstructed Group) ومجموعة تلقت التعليم (المجموعة المتعلمة The Instructed Group) . وهذه التقنية تقوم بموازنة بين مجموعتين من الأفراد حيث يمتلك كل فرد من المجموعة الأولى قدر مرتفع من المعرفة أو الصفة المعينة مع فرد من المجموعة الثانية يمتلك قدر منخفض من المعرفة أو الصفة المعينة المراد قياسها.(Berk,1982, p 53-56).

طرق حساب معامل الحساسية للتعليم:

أولاً- معامل كوكس- فارجاس Cox & Vargas :

اقترح كوكس وفارجاس عام ١٩٦٦ طريقة الاختبار القبلي والبعدي Pretest - Posttest لاختيار مفردات الاختبار محكي المرجع ، ويرمز لهذه الإحصائية بالرمز PPDI وهي تعطى بالعلاقة التالية :

$$PPDI = \frac{ب - جـ}{ن}$$

حيث أن ب عدد الطلاب الذين أجابوا على المفردة في الاختبار البعدي إجابة صحيحة أو بعد التعليم ، جـ عدد الطلاب الذين أجابوا على المفردة في الاختبار القبلي إجابة صحيحة أو قبل التعليم ، ن العدد الكلي للطلاب .

وكان هذا النموذج هو النموذج الأول لمعامل الحساسية للتعليم وبذلك أصبح هو النموذج الرئيسي لتحليل المفردات ، وهو يحتاج إلى تطبيق الاختبار مرتين وقد أوضح كوكس وفارجاس أن هذا المعامل هو معامل حساسية للتعليم وليس معامل تمييز كما في الإحصائيات معيارية المرجع . (Harris & Subkoviak , 1986 , p 496) .

وأجرى كوكس وفارجاس عام ١٩٦٦ دراسة للمقارنة بين القيم الناتجة عن هذا المعامل والقيم الناتجة عن تطبيق الطرق التقليدية في حساب معاملات التمييز التي تعتمد على إيجاد الفرق بين نسبة الأفراد في المجموعة العليا ٢٧ ٪ من أعلى الدرجات الكلية والمجموعة الدنيا ٢٧ ٪ من أدنى الدرجات الكلية . فوجد أن معامل الارتباط بين رتب المفردات لمجموعتي البيانات الناتجتين عن الطريقتين كانت تصل فقط إلى ٠,٣٧ أو ٠,٤٠ فقط ،وفي دراسات لاحقة أيضاً أكد فيها كوكس وفارجاس أن معامل الحساسية للتعليم لا يرتبط بالضرورة بمعاملات التمييز التقليدية وبخاصة إذا كانت درجات الاختبار البعدي مرتفعة (Popham , 1978 , p 108) . وهذا ما يؤكد رأي الباحث بأن معامل الحساسية أفضل في تحليل مفردات الاختبارات محكية المرجع من معاملات التمييز المستخدمة في الاختبارات معيارية المرجع .

وقيمة هذا المعامل بين (+١) عندما تكون نسبة الإجابة الصحيحة للأشخاص في الاختبار البعدي ١٠٠ ٪ ، و(-١) عندما تكون نسبة الإجابة الصحيحة الأشخاص في الاختبار القبلي ١٠٠ ٪ . (عبد الله القاضي ، ١٩٩٣ ، ١١١) .

ثانياً- معامل برينان وستوليورو : Brennan & Stolurow 1971

اقترحه كل من برينان وستوليورو عام ١٩٧١ وهو يعتمد على معامل كوكس وفارجاس PPDI عام ١٩٦٦ . ويعبر هذا المعامل عن النسبة المئوية لاحتمال الكسب Percent of Possible Gain ويرمز له بالرمز PPG ويعطى بالعلاقة التالية :

$$PPG = \frac{PPDI}{\text{1 – معامل الصعوبة في الاختبار القبلي}}$$

ويبدو أن PPG يعطي قيم أعلى من معامل كوكس وفارجاس PPDI لأنه يأخذ في الحسبان التحسن الكامن في المفردة . (Haladyna & Roid, 1981 , 40) .

مثال : إذا كانت لدينا مفردة أولى صعوبتها في الاختبار القبلي ٠,٤٠ وصعوبتها في الاختبار البعدي ٠,٧٠ ، ومفردة ثانية صعوبتها في الاختبار القبلي ٠,٦٠ وصعوبتها في الاختبار البعدي ٠,٩٠ ، فإن معامل الصعوبة حسب كوكس وفارجاس (أو معامل التطبيق القبلي والبعدي) :

للمفردة الأولى PPDI = ٠,٧٠ – ٠,٤٠ = ٠,٣٠

وللمفردة الثانية سوف يكون لها نفس المعامل .

ولكن النسبة المئوية لاحتمال الكسب PPG للمفردة الأولى

$$PPG = \frac{0.30}{0.40 - 1} = \frac{0.30}{0.60} = 0.50$$

وللمفردة الثانية : $PPG = \frac{0.30}{0.60 - 1} = \frac{0.30}{0.40} = 0.75$

أي أن قيمة PPDI لم تتغير في المفردتين مع وجود تغيير في نسبة الصعوبة للمفردتين في حين أن النسبة المئوية لاحتمال الكسب PPG تغيرت من المفردة الأولى عنها في الحالة الثانية وهذا يعني أن الكسب تحقق في المفردة الثانية أكثر .

تعليق : يرى الباحث أنه عندما يحسب PPG يجب أن يحسب المعامل PPDI ولتسهيل العمل يمكن أن نقوم بتعويض قيمة PPDI بما تساويها في PPG وكذلك تعويض قيمة معامل الصعوبة للمفردة في الاختبار القبلي أيضاً في PPG مع ملاحظة أن عدد الطلاب في الاختبار القبلي والبعدي متساوٍ (لأنها نفس المجموعة يطبق عليها الاختبار مرتين) .

$$PPG = \frac{PPDI}{1 - \text{معامل الصعوبة في الاختبار القبلي}}$$

$$PPG = \frac{\frac{ب - جـ}{ن}}{1 - \frac{جـ}{ن}} = \frac{\frac{ب - جـ}{ن}}{\frac{ن - جـ}{ن}} = \frac{ب - جـ}{ن - جـ}$$

ب عدد الطلاب الذين أجابوا إجابة صحيحة على المفردة بعد التعلم ، جـ عدد الطلاب الذين أجابوا إجابة صحيحة على المفردة قبل التعلم ، ن العدد الكلي للطلاب .

وفي رأي الباحث أن هذه الصفة أفضل وأسهل حسابياً من سابقتها التي ترتبط دوماً مع PPDI .

obeikandl.com

وغالباً ما تكون ب ≤ جـ عندما يكون التعليم فعالاً وعندما يتحقق ذلك فإن :

$$\text{ن} - \text{جـ} \geq \text{ن} \quad \text{إذن :} \quad \frac{\text{ب} - \text{جـ}}{\text{ن} - \text{جـ}} \leq \frac{\text{ب} - \text{جـ}}{\text{ن}}$$

إذن : PPG ≤ PPDI

أي أن قيمة معامل النسبة المئوية لاحتمال الكسب PPG هو دوماً أكبر أو يساوي قيمة معامل كوكس – فارجاس وذلك عندما تكون ب ≤ جـ ، وهذا يعني أيضاً أن PPG هو حساس لعملية التعلم ، ويظهر بشكل واضح التقدم البسيط في الكسب .

ثالثاً– معامل برينان : Brennan 1972

لقد طور برينان معامل الحساسية للتعليم الذي اقترحه كوكس وفارجاس عام ١٩٦٦ وذلك بتقسيم المجموعة إلى متقنين وغير متقنين . وفي البداية أخذ برينان عدد الأفراد في المجموعتين متساوٍ وتوصل إلى معامل الحساسية للتعليم التالي :

$$\mathbf{D} = \frac{\text{ب}}{\text{ن}} - \frac{\text{جـ}}{\text{ن}}$$

حيث أن ب عدد الطلاب المجموعة العليا المتقنة الذين أجابوا على المفردة إجابة صحيحة ، وجـ عدد الطلاب المجموعة الدنيا (غير المتقنة) الذين أجابوا على المفردة إجابة صحيحة ، ون عدد الطلاب في كل مجموعة ويجب أن يكون هذا العدد متساوياً .

وبعد ذلك طور برينان هذا المعامل وذلك للتخلص من العدد المتساوي في المجموعتين المتقنين وغير المتقنين ، وهذا المعامل هو :

$$\mathbf{B} = \frac{\text{ب}}{\text{ن}_1} - \frac{\text{جـ}}{\text{ن}_2}$$

حيث أن ب عدد الطلاب من مجموعة المتقنين الذين أجابوا إجابة صحيحة على المفردة ، جـ عدد الطلاب من مجموعة غير المتقنين الذين أجابوا إجابة صحيحة على المفردة ، ن١ عدد الطلاب لمجموعة المتقنين ، ن٢ عدد الطلاب لمجموعة غير المتقنين .

وهذه الصيغة تسمح لمستخدمها اختيار الطريقة التي يراها مناسبة لتحديد مجموعتي المتقنين وغير المتقنين ، وذلك باستخدام أي مستوى للإتقان أو أي درجة قطع يراها مناسبة. في حين أن الصيغة السابقة كانت تشترط أن يكون العدد متساوياً في المجموعتين العليا والدنيا (Brennan , 1972 , p 290 - 293) .

obeikandl.com

مثال : الجدول رقم (٢) التالي يبين درجات عشرة طلاب على ثماني مفردات حيث أخذ الرقم (صفر) الإجابة الخاطئة أو المتروكة والعدد (١) الإجابة الصحيحة ، فإذا كانت درجة القطع ٨٠ ٪ ، أي أن الطالب المتقن هو الذي يحصل على ٧ درجات فأكثر . ومنه فإن عدد الطلاب المتقنين هو (٦) طلاب وعدد الطلاب غير متقنين (٤) .

الجدول رقم (٢)

درجات عشرة طلاب على ثمانية مفردات

المفردة / المفحوص	١	٢	٣	٤	٥	٦	٧	٨	الدرجة الكلية
أ	١	١	١	١	١	٠	١	١	٧
ب	١	١	٠	١	١	٠	٠	٠	٤
جـ	١	١	١	١	١	١	١	١	٨
د	١	١	١	١	٠	١	١	١	٧
هـ	١	٠	١	١	١	٠	٠	٠	٤
و	١	١	٠	١	١	٠	١	١	٥
ز	١	١	١	١	١	١	١	١	٨
حـ	١	١	٠	٠	١	٠	٠	٠	٣
ط	١	١	١	١	١	١	١	١	٧
ي	١	١	١	١	١	٠	١	١	٨

ولإيجاد معامل الحساسية حسب برينان B للمفردة (٤) مثلاً ، ننظم جدول بيانات خاص بهذه المفردة كما الجدول (٣) حيث نلاحظ من الجدول (٢) أن عدد الطلاب الذين أجابوا على المفردة(٤) إجابة صحيحة من المتقنين هو ٦ وعدد الطلاب الذين أجابوا عليها إجابة صحيحة من غير المتقنين هو٣.

جدول (٣) بيانات خاصة بالمفردة رقم (٤)

		متقن	غير متقن
إجابة المفردة	صحيحة	٦	٣
	خاطئة	٠	١
المجموع		٦	٤

معامل الحساسية للمفردة (٤) حسب بريتان $B = \frac{3}{4} - \frac{6}{6} = \frac{1}{4} = 0.25$ وهكذا بالنسبة لباقي المفردات .

رابعاً- معامل بابام: Popham

اقترح بابام Popham إجراءات بديلة لتلك التي اقترحها كوكس وفارجاس وبريتان لإيجاد معامل الحساسية ، حيث يرى أنه بعد الانتهاء من عملية التعليم يجب أن تظهر المفردات الحساسة للتعليم التغير الذي حدث من خلال نتائج الاختبار القبلي والبعدي وذلك عندما يكون التعليم فعال ،أي التعرف على المفردات الأكثر حساسية للتعليم وذلك من خلال التغيرات التي تحدث في الدرجات ، كما تقاس بالفروق بين نتائج الاختبار القبلي والبعدي (مثلاً أن تعكس بشكل عام ارتفاع في الأداء من الاختبار القبلي إلى الاختبار البعدي) .

فإذا رمزنا لدرجة الطالب الذي يجيب إجابة خاطئة على المفردة بالرمز (٠) والطالب الذي يجيب إجابة صحيحة على المفردة بالرمز (١) ، يكون الرمز (٠، ١) دال على أن الفرد أجاب على المفردة في الاختبار القبلي إجابة خاطئة وفي الاختبار البعدي أجاب على نفس المفردة إجابة صحيحة . وهكذا بالنسبة للقيم التالية : (٠،٠) ، (١، ٠) ، (١، ١) ، وهذه القيم أي (٠، ١) ، (٠،٠) ، (١، ٠) ، (١، ١) هي النتائج الممكنة لأية مفردة في الاختبار القبلي والبعدي . ويمكن أن توضع هذه النتائج الممكنة في الجدول(٤) التالي :

الجدول(٤)النتائج الممكنة لأي مفردة في تطبيقين

		الاختبار البعدي	
		إجابة صحيحة	إجابة خاطئة
الاختبار القبلي	إجابة خاطئة	(٠ ، ١)	(٠ ، ٠)
	إجابة صحيحة	(١ ، ١)	(١ ، ٠)

وللتحقق من مدى اختلاف درجات المفردات بعملية التعليم إذا كانت المفردات تنتمي للنطاق السلوكي نفسه يمكن تكوين مصفوفة لدرجات الأفراد في الاختبارين القبلي والبعدي نضع في خلاياها تكرار الأفراد الذين حصلوا على الدرجات (٠، ١) ، (٠،٠) ، (١، ٠) ، (١، ١) وذلك مقابل كل مفردة كما في الجدول(٥) التالي :

obeikandl.com

جدول (٥) تكرار الاستجابات الممكنة للمفردات

المفردات	نتائج الاختبار القبلي _ البعدي			
	(٠ ، ١)	(٠ ، ٠)	(١ ، ١)	(١، ٠)
١ ٢ ٣	" نضع التكرارات المقابلة لكل مفردة الموافقة للاستجابات "			

وبما أن قيمة الوسيط Medean لا تتأثر بالقيم المتطرفة أي لا تتأثر بالمفردات الشاذة (كما في قيمة المتوسط) فإننا نقوم بإيجاد الوسيط لكل خلية من الخلايا الأربعة الموجودة في الجدول الأول ، ثم نقارن تكرار كل مفردة بالتكرارات الافتراضية التي نحصل عليها من قيمة وسيط تكرارات كل من هذه الخلايا . وبذلك نحصل على مصفوفة تشتمل على صف واحد وأربع أعمدة لكل مفردة ، نوجد قيمة مربع كاي (Chi-Square Test) والمفردات التي يكون فيها اختلاف مفاجئ لقيم مربع كاي تدل على أنها غير حساسة للتعليم وتحتاج إلى مراجعة . (Popham , 1978 , p 108 - 110) .

مثال : لنفرض أن لدينا اختباراً يتألف من عشر مفردات طبق قبل التعليم وبعده على مجموعة مؤلفة من ١٠٠ طالب وكانت النتائج في الجدول (٦) .

نحدد وسيط تكرار الخلايا ، وسيط الخلية (٠، ١) هو ٢٥ ، وسيط الخلية (٠،٠) هو ٥٠,٥ ، وسيط الخلية (١،١) هو ٥٧ ، وسيط الخلية (١، ٠) هو ٢,٥

obeikandl.com

جدول (٦)

تكرار إجابات ١٠٠ طالب على ١٠ مفردات في تطبيقين

المفردات	نتائج الاختبار القبلي _ البعدي			
	د(٠ ، ١)	(٠ ، ٠)	(١ ، ١)	(٠ ، ١)
١	٢٣	١	٧٦	٠
٢	٣٩	١٠	٤٣	٨
٣	٢٧	٣٨	٢٢	١٣
٤	٥٧	١٩	١٦	٨
٥	٣٩	٠	٦١	٠
٦	١٦	٠	٨٢	٢
٧	١٩	٨	٥٢	٢١
٨	٣٦	٨	٥٣	٣
٩	١٧	٠	٨٢	١
١٠	١١	٣	٨٤	٢

نقارن تكرار كل مفردة بالتكرار الافتراضي السابق ونوجد قيمة مربع كاي ، مثلاً المفردة الأولى :

شكل رقم (٢)

مقارنة تكرار المفردة (١) بالتكرار الافتراضي

المفردة الأولى	التكرار الحقيقي – التكرار الافتراضي ك-كَ=٢٣-٢٥=-٢	ك-كَ=١-٥,٥=- ٤,٥	ك-كَ=٧٦-٥٧=١٩	ك-كَ=٠- ٢,٥-=٢,٥

$$كا^{٢} = \sum_{م=0}^{4} \frac{(ك_م - كَ_م)^2}{كَ_م}$$

و منه كا^{٢} للمفردة الأولى :

$$كا^{٢} = \frac{(-2)^2}{25} + \frac{(-4.5)^2}{5.5} + \frac{(19)^2}{57} + \frac{(-2.5)^2}{2.5} = 12.674$$

و بنفس الطريقة كا^{٢} للمفردة الثانية :

obeikandl.com

$$\text{كا}^2 = \frac{^2(25-39)}{25} + \frac{^2(5.5-10)}{5.5} + \frac{^2(57-43)}{57} + \frac{^2(2.5-8)}{2.5} = 27.05$$

وهكذا بالنسبة لباقي المفردات نحصل على قيم كا^2 وهي مبينة في الجدول(٧) التالي:

جدول (٧) قيم كا^2 للمفردات العشرة

المفردات	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
قيم كا^2	١٢,٦٧٤	٢٧,٠٥	٢٥٧,٧٩	١١٥,٦٨	١٦,١	١٩,٨	١٣٩,٩	٦,٤	١٩,٩	٢١,٩

بالنظر إلى هذه القيم نجد أن قيم كا^2 بالنسبة للمفردات ٣ ، ٤ ، ٧ فيها اختلاف عن باقي القيم وبالتالي تعد هذه المفردات الثلاث أقل المفردات العشرة حساسية للتعليم وتحتاج مراجعة.

خامساً- معامل رودابوش : Roudabush

ومن المقاييس الأخرى التي اهتمت بتقديم معلومات تتعلق بالتغير الحادث في أداء مجموعة من المتعلمين مثل طلاب فصل معين نتيجة عملية التعليم مقياس رودابوش Roudabush 1973 . ويتميز هذا المقياس بسهولة إجرائه ، ويرمز له بالرمز DIS_{IG} ويحسب كما يلي :

DIS_{IG} = نسبة عدد الطلاب الذين أجابوا على المفردة إجابة خاطئة في الاختبار القبلي ولكنهم أجابوا عليها إجابة صحيحة في الاختبار البعدي .

وتتراوح قيمة هذا المعامل بين (١) ، صفر إذا لم يكن مصحح من أثر التخمين، وعندما يصحح من أثر التخمين تتراوح بين (-١) ، (+١) ، ويرى بيرك أن التصحيح من أثر التخمين مشكوك فيه . (صلاح علام ،٢٠٠١ ، ١٨٧ ; Berk , 1982 , p 60) .

سادساً- معامل كوسكوف- كلاين : Kosecoff & Klein

يوجد ثلاثة معاملات لكوسكوف وكلاين

أ- قدم كوسكوف وكلاين عام ١٩٧٤ مقياساً للحساسية ، هذا المقياس يأخذ بعين الاعتبار نسبة الطلاب الذين أجابوا على المفردة قبل التعليم وبعد الانتهاء منه إجابة صحيحة، ويعتبر أكثر دقةً وتحفظاً حيث يمثل التغيير الفعلي الحادث نتيجة عملية التعليم فيما تقيسه المفردة ، ويعطى بالعلاقة التالية :

$$\text{معامل الحساسية} = \frac{\text{د} - \text{هـ}}{\text{ن}}$$

حيث أن :

د : عدد الأفراد الذين أجابوا على المفردة إجابة خاطئة قبل التعليم ولكنهم أجابوا عنها إجابة صحيحة بعد التعليم .

هـ : عدد الأفراد الذين أجابوا على المفردة إجابة صحيحة قبل التعليم وبعد التعليم .

ن : العدد الكلي للأفراد الذين طبق عليهم الاختبار .

وقيمة هذا المعامل تتراوح بين (-١) ، (+١) . (صلاح علام ، ٢٠٠١ ، ١٨٨) .

٢- وقدم كوسكوف وكلاين عام ١٩٧٤ معامل آخر لقياس الحساسية ، وهو يعتمد على اعتبار أن جميع الطلاب إجابتهم خاطئة على المفردة في الاختبار القبلي ، وأن معامل الحساسية يرمز له بالرمز DIS_{NG} ويعطى بالعلاقة التالية :

$$DIS_{NG} = \frac{\text{د} - \text{ط}}{\text{ن}}$$

حيث أن : د ، ن لهما نفس المعنى في المعامل السابق .

ط : هي عدد الأفراد الذين أجابوا على المفردة إجابة خاطئة في مرتي التطبيق .

وقيم هذا المعامل تتراوح بين (-١) ، (+١) ، وهذا المعامل يستخدم معلومات أكثر وقيمه الممكنة لها مدى أكبر من معامل رودابوش DIS_{IG} ، وهو أكثر حساسية من معامل كوكس وفارجاس ، ويأخذ في الحسبان الكسب والخسارة الفردية الناتجة عن التعليم وكذلك في المجموعة وهو يعطي تقييم أكثر تحفظاً من المعاملات الأخرى(Berk,1982,p60-61).

٣- وهناك معامل آخر لحساب الحساسية للتعليم قدمه كوسكوف وكلاين عام ١٩٧٦ Kosecoff & Klein ، ويطبق هذا المعامل على مجموعتين مستقلتين إحداهما تلقت التعليم والأخرى لم تتلقَ التعليم ، ويرمز له بالرمز DIS_{UIGD} وهو يعطى بالعلاقة التالية :

$$DIS_{UIGD} = ك_١ - ك_٢$$

حيث أن ك١: نسبة الطلاب الذين أجابوا إجابة صحيحة على المفردة في المجموعة التي تلقت التعليم .

ك٢ : نسبة الطلاب الذين أجابوا إجابة صحيحة على المفردة في المجموعة التي لم تتلقَ التعليم .

وقيم هذا المعامل تتراوح أيضاً بين (-١) ، (+١) . (صلاح علام ، ٢٠٠١ ، ١٩٤; Berk , 1982 , p 60) .

سابعاً- معامل هالادينا : Haladyna 1974

قدم هالادينا Haladyna 1974 طريقة لتحليل مفردات الاختبار محكي المرجع تستخدم معامل التمييز الكلاسيكي ، واستخدم في هذه الطريقة الارتباط الثنائي النقطي للعينات المتجمعة (Combined - Samples Point - Biserial Corelation (COMPBI يبين الأداء على المفردة والاختبار عندما تتضمن العينة مدى كامل للطلاب المتعلمين والطلاب غير المتعلمين ، فإن معامل الارتباط الثنائي النقطي للعينات المتجمعة (COMPBI) يكون أقل تأثراً بالمدى ويميل لأن يعمل كنوع من معامل الحساسية للتعليم .

ويوضح فحص صيغة الارتباط الثنائي النقطي (ر د م) أن متوسط الاختلافات في درجات الاختبار للأفراد الذين أجابوا على المفردة إجابة صحيحة والذين أجابوا عليها إجابة خاطئة هو أداة لتحديد حجم المعامل والذي يعطى بالعلاقة التالية :

$$\text{ر د م} = \frac{(\bar{س}_{\text{ص}} - \bar{س}_{\text{خ}}) \sqrt{\text{ن}_1 \text{ ن}_2}}{\text{أ}}$$

حيث أن : ر د م الارتباط الثنائي النقطي بين درجات الاختبار والمفردة .

$\bar{س}_{\text{ص}}$:متوسط درجة الطلاب (على الاختبار) الذين أجابوا إجابة صحيحة على المفردة .

$\bar{س}_{\text{خ}}$: متوسط درجة الطلاب (على الاختبار) الذين أجابوا إجابة خاطئة على المفردة .

ن_1 : نسبة المفحوصين الذين أجابوا على المفردة إجابة صحيحة .

ن_2 : نسبة المفحوصين الذين أجابوا على المفردة إجابة خاطئة .

أ : الانحراف المعياري لدرجات الاختبار .

وعندما يكون $\bar{س}_{\text{ص}} - \bar{س}_{\text{خ}}$ كبير تكون ر د م كبيرة أيضاً .

وعند الاستخدام مع عينات متجمعة (موحدة) فإن ر د م يميل للتفريق بين الطلاب المتعلمين الذين يميلون للأداء الناجح على المفردات ، والطلاب غير المتعلمين الذين يميلون للأداء غير الناجح على المفردات . وعندما يمتحن الطلاب المتعلمين فقط فإن ر د م لا يمتلك القدرة على التمييز ، ودرجات الاختبار البعدي تميل لأن تكون مرتفعة بانتظام ، وأغلب الطلاب المتعلمين سوف يجيبون بشكل صحيح على المفردات . وهكذا سوف يسمح استخدام كل الدرجات في الاختبار القبلي والبعدي لمعامل التمييز الكلاسيكي أن يصبح معامل حساسية

للتعليم ، لأن ر د م حساس للاختلاف بين الدرجات المتوسطة لمجموعة الطلاب المتعلمين وغير المتعلمين . (Haladyna & Roid , 1981 , p 41 - 42) .

ثامناً- معاملات بايزن : Bayesien

نظرية بايزن هي قاعدة لثلاثة معاملات وجميع هذه المعاملات الثلاث تتطلب معلومات ملائمة لها وتعتبر معلومات الصعوبة للاختبار القبلي هي أكثر المعلومات ملائمة ، ومعاملات بايزن هي :

١- ب١ : هو احتمال أن الطالب يمتلك المعلومة بمعلومية إعطاء إجابة صحيحة على المفردة .

٢- ب٢ : هو احتمال أن الطالب لا يمتلك المعلومة ويعطي إجابة صحيحة على المفردة.

٣- ب٣ : هو احتمال اتخاذ قرار صحيح ، والذي يتطلب أن يصنف إما متقن أو غير متقن .

وإحدى تطبيقات هذا النوع من تحليل المفردة ، هو أن المفردات ذات المعامل ب٣ المرتفع تزيد من احتمال التصنيف الصحيح للطلاب إلى حالات الإتقان أو عدم الإتقان . ومعاملات بايزن تتطلب معرفة صعوبة المفردة في الاختبار القبلي والبعدي ، وهذه المعاملات تعطى بالعلاقات التالية :

$$ب_1 = \frac{(ص_2)\ (ص)}{(ص_2)\ (ص) + (ص_1)\ (1 - ص)}$$

$$ب_2 = \frac{(1 - ص)\ (1 - ص_1)}{(ص)\ (1 - ص_2) + (1 - ص)\ (1 - ص_1)}$$

$$ب_3 = (ص_2 - ص) + (1 - ص_1) + (ص - ص_2)$$

حيث أن كلاً من : ب١ ، ب٢ ، ب٣ معاملات بايزن .

ص١ : صعوبة العينة قبل التعليم .

ص٢: صعوبة العينة بعد التعليم .

ص : هي متوسط ص١ + ص٢ أو هي صعوبة العينة مجتمعة .

إن أساس هذه المفاهيم لكل معامل من هذه المعاملات يستمد من الاحتمالات الشرطية والبناء المنطقي لها هنا يأخذ تعريف معامل الحساسية للتعليم .

ومن العلاقات السابقة نجد أن ب١ تكون أكبر ما يمكن عندما يكون البسط أكبر ما يمكن والمقام أصغر ما يمكن ، ويكون ذلك عندما تكون صعوبة المفردة في الاختبار القبلي والبعدي تساوي (١) ، وكذلك ب٢ تميل للزيادة عندما تقترب صعوبة المفردة في الاختبار القبلي والبعدي من الصفر ، أما ب٣ فإنها تكون أكبر ما يمكن عندما تكون صعوبة المفردة في الاختبار القبلي والبعدي مرتفعتين معاً أو منخفضتين معاً . (Haladyna & Roid , 1981 , p 41 - 42)

مثال : الجدول (٨) يحتوي صعوبات الاختبار القبلي والبعدي ومعاملات بايزن

جدول(٨) يوضح حساب قيم معاملات بايزن

الصعوبة في الاختبار القبلي ص١	الصعوبة في الاختبار البعدي ص٢	ب١	ب٢	ب٣
٠,٧٠	٠,٩٠	٠,٨٤	٠,٤٣	٠,٧٨
٠,٦٠	٠,٨٠	٠,٧٦	٠,٤٦	٠,٦٨
٠,٥٠	٠,٧٠	٠,٦٨	٠,٥٣	٠,٦٢
٠,٤٠	٠,٦٠	٠,٦٠	٠,٦٠	٠,٦٠
٠,٣٠	٠,٥٠	٠,٥٣	٠,٦٨	٠,٦٢
٠,٢٠	٠,٤٠	٠,٤٦	٠,٧٦	٠,٦٨

إذا حسبنا معامل الحساسية وفق كوكس وفارجاس نجد أن معامل الحساسية ثابت ، وهو يساوي ٠,٢٠ وهو منخفض ، أما قيم بايزن فنجد أنها مرتفعة وخصوصاً عندما كانت مستويات الصعوبة مرتفعة في مرتي التطبيق كانت ب١ مرتفعة .

و نلاحظ من الجدول أن : ب١ تتناقص قيمتها مع تناقص الصعوبة في مرتي التطبيق .

و ب 2 تتزايد قيمتها عندما تتناقص قيمة الصعوبة في مرتي التطبيق .

ولكن ب3 تتزايد قيمتها عندما تتزايد قيمتا الصعوبة معاً في مرتي التطبيق أو تتناقصان معاً ، أي أن هذه المعاملات تتأثر كثيراً بمستوى الصعوبة وخصوصاً عند الأطراف لذلك هناك بعض المحاذير من استخدامها .

- بقي أن يشير الباحث إلى أنه هناك معاملات أخرى غير التي ذكرها ، مثل معامل Darlington & Bishop , 1966 ومعامل Saupe , 1966 ومعامل الانحدار المتعدد لملمان Millman , 1974 وغيرها أيضاً ، ولكن هذه المعاملات غير منتشرة وقليلة

الاستعمال إما لأنها تحتاج لجهد كبير في حسابها ، أو أن تفسيرها غير ملائم ولذلك لا تستعمل في أوساط القياس كثيراً . (Berk , 1982 , p 61)

تاسعاً- المعامل المقترح:

وعندما يطبق اختبار على مجموعة من الطلاب قبل التعليم وبعد الانتهاء منه نحصل على الاحتمالات التالي للإجابة على كل مفردة :

ص ص : تعني أن الطالب أجاب إجابة صحيحة في الاختبارين القبلي والبعدي .

ص خ : تعني أن الطالب أجاب إجابة صحيحة في الاختبار القبلي وخاطئة في الاختبار البعدي

خ ص : تعني أن الطالب أجاب إجابة خاطئة في الاختبار القبلي وصحيحة في الاختبار البعدي

خ خ : تعني أن الطالب أجاب إجابة خاطئة في الاختبارين القبلي والبعدي .

فإذا طبقنا اختباراً مكوناً من خمس مفردات على١٠٠ طالب قبل التعليم وبعد الانتهاء منه وكانت النتائج موضحة في الجدول(٩) وأوجدنا قيم هذه المعاملات للمفردات الخمسة وهي مدونة في الجدول(٩).

نلاحظ من هذا المثال أن قيمة معامل كوكس -فارجاس لجميع المفردات متساوٍ وهي ٠,٤٥ وبالتالي هذا المعامل هنا لم يفرق بين هذه المفردات علماً أن استجابات الأفراد لها كانت مختلفة .

ومعامل الحساسية لبرينان – ستوليورو فرق بين المفردتين الأولى والثانية ولكنه لم يفرق بين المفردات الباقية أي أنه يعطي للمفردات ٣ ، ٤ ، ٥ نفس المعاملة من حيث قيم معامل الحساسية .

أما قيمة معامل كوسكوف-كلاين (الصيغة ١) فإننا نلاحظ أنها متساوية من أجل المفردتين الثانية والخامسة ومختلفة بالنسبة للمفردات الباقية ،أي أن هذا المعامل لم يفرق بين المفردتين الثانية والخامسة علماً أن استجابات الأفراد لهما كانت مختلفة.

ولتلافي التساوي في مثل المفردات السابقة يقترح الباحث حساب معامل الحساسية من العلاقة التالية :

$$\text{معامل الحساسية المقترح} = \frac{\text{ب} - \text{جـ}}{1 - \text{جـ}} \quad ; \quad ٢\text{حـ} \geq \text{ب} + ١$$

حيث أن ب نسبة من أجابوا على المفردة في الاختبار القبلي إجابة خاطئة وفي الاختبار البعدي إجابة صحيحة .

أما جـ فهي نسبة من أجابوا على المفردة في الاختبارين القبلي والبعدي إجابة صحيحة.

ومن الجدول (٩) نلاحظ أن قيم هذا المعامل المقترح مختلفة بالنسبة للمفردات الخمسة، فقيمته عند المفردة الأولى ٠,٧٠ ، وعند المفردة الثانية ٠,٤١ ، وعند المفردة الثالثة ٠,٦١ ، وعند المفردة الرابعة ٠,٥٣ وعند المفردة الخامسة ٠,٤٤، وبالتالي أعطى معاملات حساسية تختلف من مفردة إلى أخرى ، أي أن المعامل المقترح يفرق بين المفردات ، في حين أن المعاملات السابقة لم تفرق بين هذه المفردات.

جدول (٩)

تكرار إجابات ١٠٠ طالب على ٥ مفردات في تطبيقين ومعاملات الحساسية لكل مفردة

المفردات	ص ص	ص خ	خ ص	خ خ	كوكس- فارجاس	برينان - ستوليورو	كوسكوف - كلاين	المعامل المقترح
١	٠	٢٥	٧٠	٥	٠,٤٥	٠,٦٠	٠,٧٠	٠,٧٠
٢	١٥	٥	٥٠	٣٠	٠,٤٥	٠,٥٦	٠,٣٥	٠,٤١
٣	١٠	٢٠	٦٥	٥	٠,٤٥	٠,٦٤	٠,٥٥	٠,٦١
٤	١٥	١٥	٦٠	١٠	٠,٤٥	٠,٦٤	٠,٤٥	٠,٥٣
٥	٢٠	١٠	٥٥	١٥	٠,٤٥	٠,٦٤	٠,٣٥	٠,٤٤

مناقشة المعامل المقترح :

$$\text{معامل الحساسية المقترح} = \frac{\text{ب} - \text{جـ}}{1 - \text{جـ}} \quad ; \quad 2\text{حـ} \geq \text{ب} + 1$$

* عندما يجيب جميع الطلاب في الاختبار القبلي إجابة خاطئة وفي الاختبار البعدي إجابة صحيحة يكون : ب = ١ ، جـ = ٠

$$\Leftarrow \text{معامل الحساسية المقترح} = \frac{\text{ب} - \text{جـ}}{1 - \text{جـ}} = \frac{1-0}{1-0} = 1$$ أي يكون معامل الحساسية المقترح مساوياً الواحد عندما يجيب جميع الطلاب في الاختبار القبلي إجابة خاطئة وفي الاختبار البعدي إجابة صحيحة ، وبالتالي يكون عدد الطلاب الذين أجابوا إجابة صحيحة في التطبيقين مساوياً الصفر .

* وعندما يتساوى عدد الطلاب الذين أجابوا إجابة صحيحة في مرتي التطبيق مع عدد الطلاب الذين أجابوا إجابة خاطئة في الاختبار القبلي وإجابة صحيحة في الاختبار البعدي. أي أن $\boxed{\text{ب} = \text{جـ}} \Leftarrow \frac{\text{ب} - \text{جـ}}{1 - \text{جـ}} = 0 \Leftarrow$ معامل الحساسية المقترح = ٠

وبملاحظة أن ب + جـ ≥ ١ ⇐ ب + ب ≥ ١ ⇐ ب ≥ ٠,٥

أي | ب = حـ ≥ ٠,٥ |

ومنه تكون قيمة المعامل المقترح مساوية الصفر عندما تكون نسبة من أجابوا إجابة خاطئة في الاختبار القبلي ثم أجابوا إجابة صحيحة في الاختبار البعدي تساوي نسبة من أجابوا إجابة صحيحة في مرتي التطبيق وكل من هاتين النسبتين أقل من ٠,٥ .

* وإذا كانت نسبة من أجابوا على المفردة إجابة صحيحة في مرتي التطبيق تساوي نصف نسبة من أجابوا على المفردة إجابة خاطئة في الاختبار القبلي وإجابة صحيحة في الاختبار البعدي مضافاً إليها نصف، أي:

جـ = ٠,٥ ب + ٠,٥ ⇐ ب= ٢جـ - ١ ⇐

المعامل المقترح = $\frac{\text{ب} - \text{جـ}}{1 - \text{جـ}} = \frac{2\text{حـ} - 1 - \text{حـ}}{1 - \text{جـ}} = \frac{\text{حـ} - 1}{\text{جـ} - 1} = -1$

وبما أن ب نسبة فإن ٠ ≤ ب ≤ ١ ⇐ ٠ ≤ ٠,٥ ب ≤ ٠,٥ ⇐

٠,٥ ≤ ٠,٥ ب + ٠,٥ ≤ ١ ⇐ ٠,٥ ≤ حـ ≤ ١ ①

ومن جهة ثانية ب+ حـ ≥ ١ ولدينا ب= ٢جـ - ١ ⇐

٢حـ - ١ + حـ ≥ ١ ⇐ جـ ≥ (٣١٢) ②

من ① و ② نجد أن | ٠,٥ ≤ حـ ≤ ٣١٢ |

نعوض حـ بما تساويه فنجد :

٠,٥ ≤ ٠,٥ ب + ٠,٥ ≤ ٣١٢ ⇐ | ٠ ≤ ب ≤ ٣١١ |

أي إذا تحقق أن : معامل الحساسية المقترح =-١ فإن ذلك يعني أن نسبة من أجابوا على المفردة إجابة صحيحة في مرتي التطبيق نصف نسبة من أجابوا على المفردة إجابة خاطئة في الاختبار القبلي وإجابة صحيحة في الاختبار البعدي مضاف إليها نصف مع ملاحظة أن نسبة من أجابوا إجابة صحيحة في مرتي التطبيق محصورة بين ٠,٥ و ٣١٢، ونسبة من أجابوا إجابة خاطئة في الاختبار القبلي وصحيحة في الاختبار البعدي محصورة بين الصفر وبين ٣١١ .

*****