

## أساليب تقييم جودة الاختبارات مرجعية المحك

## Methods for assessing the quality of the criterion referenced test,

هاني دانة<sup>1</sup>، رقية بوقندورة<sup>2</sup><sup>1</sup>جامعة محمد خيضر، بسكرة، الجزائر [hani.dana@univ-biskra.dz](mailto:hani.dana@univ-biskra.dz)<sup>2</sup>جامعة محمد لمين دباغين، سطيف2، الجزائر [bougandourarokaya@gmail.com](mailto:bougandourarokaya@gmail.com)

تاريخ الاستلام: 2023/08/26 تاريخ القبول: 2024/04/27 تاريخ النشر: 2024/05/31

## Abstract

## ملخص

Achievement tests are one of the most used evaluation tools in the educational process, as they provide an actual picture of the level of learner performance during the educational stage, and the process of the learning process, through which the learner is built and his behavior is developed according to the expected goal. The achievement tests have many types, including the one based on how to interpret the results, through which the tests are classified into: Norm through which the level of individual performance is classified based on the reference group and highlighting individual differences, and the criterion which aims to verify the level of the learner's ability to achieve the educational goals set.

Therefore, the criterion referenced test and the norm referenced test differ from each other greatly, both in terms of the stages of construction and interpretation or even in terms of how to verify the quality of the test from the indicators of sensitivity to education and indicators of mastery and discrimination, and how to estimate the psychometric properties, so we will try in this article to review the most important methods used in evaluating the quality of the the criterion referenced test

Keywords: the criterion referenced test, difficulty index, sensitivity index for education, proficiency index, psychometric characteristics.

تعد الاختبارات التحصيلية من أكثر الأدوات التقييمية استخداما في العملية التعليمية، فهي تقدم صورة فعلية حول مستوى أداء المتعلم خلال المرحلة التعليمية، وسيرورة عملية التعلم و التي من خلالها يتم بناء المتعلم وتطوير سلوكه تبعا للهدف المنتظر. وللاختبارات التحصيلية أنواع عديدة منها القائم على كيفية تفسير النتائج والتي تصنف من خلالها الاختبارات إلى: المعيارية والتي من خلالها يتم تصنيف مستوى أداء الفرد استنادا للجماعة المرجعية وإبراز الفروق الفردية، والمحكية والتي تستهدف التحقق من مستوى قدرة المتعلم على تحقيق الأهداف التعليمية المسطرة.

لذا يختلف الاختبار المحكي والاختبار المعياري عن بعضها اختلافا كبيرا سواء من حيث مراحل البناء والتفسير أو حتى من حيث كيفية التحقق من جودة الاختبار من مؤشرات الحساسية للتعليم ومؤشرات الاتقان والتمييز، وكيفية تقدير الخصائص السيكومترية، لذا سنحاول في هذه المقال استعراض أهم الأساليب المستخدمة في تقييم جودة الاختبارات المرجعية المحك **كلمات مفتاحية:** الاختبار المحكي، مؤشر الصعوبة، مؤشر الحساسية للتعليم، مؤشر الاتقان، الخصائص السيكومترية.

## 1. مقدمة

تعد الاختبارات التحصيلية من الأدوات التقييمية الواسعة الاستخدام في العملية التعليمية، فهي تمثل مجموعة من الوضعيات التي تهدف إلى قياس مدى تمكن المتعلم من اكتساب المعارف والمهارات خلال الفترة التعليمية، وتوضح للمعلم مدى سيره في تطبيق المنهاج الدراسي وفعالية الطرق التعليمية والتدريسية التي استخدمها وجودتها من خلال تفاعل المتعلم مع الاختبار، وتقدير درجة التغير الحاصل في سلوك المتعلم، فمن خلال الاختبارات التحصيلية يتم اتخاذ القرارات التعليمية المناسبة.

غير أن اتخاذ القرار التعليمي المناسب يرتبط بطبيعة الاختبار التحصيلي في حد ذاته، سواء من حيث مراحل اعداده، شكل أسئلته، وكيفية تفسير النتائج المتحصل عليها من خلال تطبيق هذا الاختبار، لذا يتبين لنا أن للاختبار عدة تقسيمات، ولعل من أهم هذه التقسيمات التقسيم القائم على كيفية تفسير النتائج، ذلك أن هذا التقسيم من الاختبارات يبين لنا نوع الاختبار المناسب المستخدم في اصدار الأحكام على أداء المتعلمين من حيث النجاح - الرسوب، واتخاذ القرارات سواء تسكينا- توجيهها- او تشخيصا، وعلى هذا الأساس تنقسم الاختبارات إلى: الاختبارات معيارية المرجع- اختبارات محكية المرجع.

فالاختبارات المعيارية المرجع تعد من أدوات القياس التقليدي الجماعي، فهي تستند في تحليل نتائجها إلى توزيع الخاصية المعرفية المراد قياسها لدى التلاميذ والتميز في مدى توافرها معا، فهي بالتالي تركز بالأساس على تبيان الفروق الفردية، وهذا من خلال تقدير درجات انحراف الأفراد عن المتوسط الحسابي للجماعة المرجعية. (يوسف خنيش، 2005، ص 63)

فالاختبارات المعيارية حسب ( صلاح الدين محمود علام، 2000، ص 335) تعتبر اختبارات مقننة كونها أنها تعتمد أساسا على إجراءات تقنين منظمة من خلال اعداد جداول معايير يرجع إليها لمقارنة أداء الفرد بالجماعة المرجعية وتحديد توقعه بدقة.

غير أن في الكثير من الأحيان نكون القرارات التربوية المتخذة لا تهدف أساسا إلى توجيه وانتقاء الأفراد، بل تستهدف مدى قدرتهم على تحقيق الأهداف التعليمية المنشودة بعيدا عن الأداء المرجعي، وأيضا تشخيص مستويات الأفراد خلال العملية التعليمية وتحديد المشكلات التي يواجهونها، وتظهر خلال عملية التقويم التكويني، لذا فإن الاختبارات المعيارية في هذه الحالة لاتصلح لذا وجب ايجاد نمط آخر من الاختبارات لتقييم الفرد مع الأهداف التعليمية. مما ساعد على بروز وظهور نمط الاختبارات المحكية.

فقد ظهر هذا النمط من الاختبارات على يد **GLASER** سنة 1963 الذي قام بنشر مقال قصير سنة(1963) بعنوان " تكنولوجيا التعليم وقياس مخرجات التعلم"، حيث استخدم في هذا المقال مصطلحين جديدين هما "القياس المعياري المرجع" والذي أشار فيه إلى الاختبارات التقليدية التي هدفها الأساسي

المقارنة بين أداء الأفراد لتحديد موقعهم النسبي. والمصطلح الثاني "القياس المحكي المرجع" الذي هو البديل ويستهدف تقدير مستوى التمكن لدى الأفراد وتحقيقهم للأهداف السلوكية.

( امحمد بوزيان تيغرة، 1998، ص 116).

وتعتمد الاختبارات المحكية المرجع في تفسير أداء الأفراد وقدرتهم على الوصول إلى الأهداف التعليمية على ما يعرف بدرجة القطع، فهذه الدرجة تساعد على تقدير مستوى الأداء الفعلي الذي وصل إليه الفرد، وتمثل الحد الفاصل بين الاتقان وعدم الاتقان، والتي من خلالها يتم اتخاذ القرارات التربوية المناسبة. ( هاني دانة، وشفيفة كحول، 2022، ص 188).

ويتضح لنا أنّ الاختلاف واضح بين الاختبارات المعيارية والمحكية المرجع في كيفية تفسير النتائج واتخاذ القرارات التربوية المناسبة، غير أنّ هذا الاختلاف لا يعد الوحيد، بل يختلفان أيضا في خطوات بناء الاختبار، والاختلاف الأهم يمكن في كيفية تحليل جودة الاختبار. ففي الاختبارات المعيارية يتم تحليل جودة الاختبار وفقراته انطلاقا من التباينات والمتوسطات الحسابية، غير أنّ الاختبارات المحكية تنطلق في تحليل جودة الاختبار من حيث بنائه عن طريق خصائصه السيكومترية، أو من حيث جودة فقراته عبر درجة القطع، والذي يعد الجوهر الرئيس في الاختبارات المحكية.

لذا سنحاول في هذا المقال استعراض أهم الأساليب المستخدمة في تقييم جودة الاختبارات المحكية المرجع من خلال تناول العناصر التالية:

• تعريف الاختبارات المحكية.

• أوجه التشابه والاختلاف بين الاختبارات المعيارية والمحكية.

• - أساليب تقييم جودة الاختبارات المحكية

**1. - الاختبارات محكية المرجع:** عرفت كل من مريم قارة وصباح ساعد (2018) الاختبار المحكي المرجع بأنه مجموعة من الفقرات الاختبارية التي تقيس أهدافا سلوكية محددة يتم من خلالها تحديد مستوى اتقان الفرد لتلك الأهداف انطلاقا من درجة القطع المحددة سلفا..

ولقد عرف (صباح حمزة العجيلي، 2005) الاختبار المحكي بأنه اختبار يتم فيه تحديد مستوى أداء الفرد بالنسبة لمحك ثابت دون الرجوع إلى أداء الآخرين، ويرتبط عادة بالأهداف السلوكية للمادة الدراسية.

فهذه الاختبارات تكون تشخيصية، إذ أنّها تركز على تقدير أداء الفرد تبعا للمجال السلوكي المحدد، والتي تمثلها مجموعة من الأهداف السلوكية. فالاختبار المحكي يعمل على تشخيص سيرورة التعلم لدى الأفراد خلال الموقف التعليمي ومدى اتقانهم وتحقيقهم للأهداف بدقة. ومن خلال نتائج هذا الاختبار يتم وضع خطط تدريبية لتغطية الاحتياجات التعليمية لكل فرد على حدى. ( هاني دانة، و شفيفة كحول، 2022،

ص 188)

2. - أوجه التشابه والاختلاف بين الاختبارات المعيارية والمحكية: للاختبارات المعيارية والمحكية المرجع مجموعة من أوجه التشابه والاختلاف النظري يمكن ذكرها في الجدول التالي:

جدول رقم 1: أوجه التشابه والاختلاف بين الاختبارات المعيارية والمحكية

الاختبارات المحكية	الاختبارات المعيارية	
انتقان الأهداف التعليمية	إبراز الفروق الفردية	الهدف
التشخيص	التوجيه والانتقاء	الاستخدام
تقويم تشخيصي و تكويني	تقويم ختامي	طبيعة التقويم
وفق درجة القطع	وفق جدول معايير محدد	كيفية تفسير مستوى أداء الفرد
المعلم	الجهات الرسمية	المسؤول عن الاختبار

وفيمايلي شرح تفصيلي للجدول:

2-1- الهدف من الاختبار: فالاختبار المعياري يستهدف أساسا الكشف عن الفروق الفردية بين الأفراد في درجة تملك تلك السمة أو اكتساب المعارف، بينما الاختبار المحكي فإنه يركز على الفرد ومدى قدرته على الوصول إلى الأهداف التعليمية المحددة.

2-2- مجال الاستخدام: تستخدم الاختبارات المعيارية في عملية الانتقاء والتوجيه الأكاديمي والمهني، أما الاختبارات المحكية فهي تشخيصية بالأساس وتساعد على تحديد احتياجات الفرد والمشكلات التي يواجهها في عملية التعلم انطلاقا من عدم التمكن من بعض الأهداف.

2-3- طبيعة التقويم: فالاختبارات المعيارية تستخدم من أجل اتخاذ قرارات الانتقاء والتوجيه، فهذه القرارات تكون بعد عملية التقويم الختامي، ومنه فالاختبارات المعيارية تستخدم مع التقويم الختامي، أما الاختبارات المحكية فهي تعمل على تشخيص سيورة التعلم وعملية التكوين وتحديد الاحتياجات التعليمية المناسبة، وهذا يظهر خلال عملية التقويم التشخيصي والتكويني.

2-4- كيفية تفسير أداء الفرد: تعتمد الاختبارات المعيارية على المعايير (Standard) في تفسير الأداء والتي تتضمن طرق احصائية مختلفة من مئينيات ودرجات معيارية ودرجات تائية. أما الاختبارات المحكية فهي تعتمد على تحديد درجة القطع، والتي من خلالها يتم تحديد الحد الفاصل بين الاتقان وعدم الاتقان للأهداف التعليمية.

2-5- المسؤول عن الاختبار: الاختبارات المعيارية تشرف عليها عدة جهات سواء المعلم داخل الغرفة الصفية أثناء الاختبارات الفصلية، أو الديوان والمركز الوطني للتقويم في الاختبارات الرسمية مثل

شهادة البكالوريا، فالاختبارات المعيارية تكون مقننة. أما الاختبارات المحكية فهي خاصة ترتبط بالمعلم الذي يستخدم هذا النوع من الاختبارات للتشخيص وتحديد مسار التعلم وكيفية علاج المشكلات التعليمية.

### 3- أساليب تقييم جودة الاختبارات المحكية:

لقد ذكرنا أنفاً أن الاختبارات التحصيلية المعيارية والمحكوية المرجع يختلفان في العديد من النقاط بدءاً من الهدف من الاستخدام وكيفية تحديد وتفسير مستوى الأداء وأيضاً في مراحل بناء الاختبار. ومن خلال هذه الاختلافات يمكن جلياً أن تكون هناك اختلافات عملية في تحليل جودة الاختبارات المعيارية والمحكوية. فعملية تحليل الاختبارات بصفة عامة يمكن تقسيمها إلى جزئين: تحليل عام للاختبار وتحليل خاص تفصيلي وهذا مرتبط بفقرات الاختبار من حيث كيفية بنائها، وعليه فإن الاختبارات مرجعية المحك تعتمد على مجموعة من الأساليب للتحقق من جودة الاختبار والتي من بينها.

### 3-1- مؤشر الحساسية للتعليم

اقترح هذا المؤشر كل من "Vargas & Cox" يستخدم هذا المؤشر لتحديد أي من البنود لها فعالية في التعليم. ويحسب من خلال إيجاد الفرق بين مؤشر الصعوبة قبل وبعد التعليم وفق العلاقة التالية:

$$S = P_{\text{post}} - P_{\text{pre}}$$

حيث أن:

S: مؤشر الحساسية للتعليم.

$P_{\text{post}}$ : مؤشر الصعوبة بعد التعليم.

$P_{\text{pre}}$ : مؤشر الصعوبة قبل التعليم.

وكلما ازداد الفرق بين مؤشري الصعوبة، دل ذلك على أن البند مميز، أما في حال وجود قيمة S (مؤشر حساسية التعليم) سالبة أو تساوي 0 فإنه يمكن تفسيرها وفق حالتين.

- أن الفقرة غير متوافقة ولا تدخل في نطاق التعليم.
- التعليم ليس له أثر في نجاح التلميذ.

غير أن أهم صعوبة تواجه استخدام هذا المؤشر هو طول أعمال الوحدة التعليمية حتى يتم تطبيق

الاختبار، وبالتالي فإن الفاصل الزمني مؤثر على أداء الأفراد. (Dany Leveaut & Jacques

Gregory, 2014, p. 224)

أ. مؤشر التمييز عبر المجموعات

يعد أحد الطرق المستخدمة في إيجاد قيمة مؤشر التمييز للفقرات في الاختبارات المحكية، تقوم هذه

الطريقة على انتقاء مجموعتين من الأفراد؛ المجموعة الأولى تكون متلقية للتعليم والأخرى لم تتلق للتعليم،

ويطبق عليهم الاختبار في الوقت نفسه. ويتم استخراج مؤشر التمييز عن طريق حساب الفرق بين مؤشر الصعوبة للفقرة عند المجموعتين. (طارق بن ابراهيم جابر الجهني، 2005، ص 37)

#### ب. مؤشر التمييز في عتبة الاتقان

يشير هذا المؤشر إلى أنّ تحديد قدرة الفقرة على التمييز تتم عن طريق درجة القطع، أي مستوى التمكن من تحقيق الأهداف التعليمية، وتختلف هذه الطريقة عن طريقة المقارنة الطرفية المستخدمة في حساب مؤشر التمييز في الاختبارات المعيارية في تقسيم الأفراد، فالمقارنة الطرفية تعتمد على تصنيف الأفراد استنادا إلى الدرجات في الاختبار بغض النظر عن مدى اتقانهم للأهداف التعليمية، أما في الاختبار المحكي فيتم تصنيف الأفراد تبعا للأهداف المحققة.

ويتم حساب مؤشر التمييز في عتبة الاتقان وفق العلاقة التي وضعها "Brenann" والتالي كانت

كالتالي:

$$B = P_m - P_{nm}$$

حيث أن:

**B:** مؤشر التمييز عند عتبة الاتقان.

**P<sub>n</sub>:** مؤشر صعوبة البند للأفراد الذين بلغوا عتبة الاتقان للاختبار ككل.

**P<sub>nm</sub>:** مؤشر صعوبة البند للأفراد الذين لم يبلغوا عتبة الاتقان للاختبار ككل.

وتتراوح قيمة مؤشر التمييز بين (-1) و(+1)، وتشير القيمة السالبة إلى أن البند لا يميز بين مجموعة

المتقنين وغير المتقنين. (Dany Leveaut, & Jacques, Gregory, 2014, p. 225)

#### ج. مؤشر الاتفاق

يستخدم هذا المؤشر لقياس درجة التوافق بين بندين يقيسان نفس الهدف السلوكي، إذ يجب أن يكون البندان

منتميين إلى نفس المحتوى الدراسي وتكون هذه الحالة في وضعية اختيار بطريقة عشوائية عبر مجموعة

من الفقرات اختبرت ميدانيا.

ويتم استخراج مؤشر التمييز عبر دليل الاتفاق من خلال معامل كاف تربيع وفق العلاقة التالية:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

( Linda Crocker & James Algina, 2008, p. 332)

د. طريقة معامل التوافق المرجعي

يعد من أبسط الطرق المستخدمة في تقدير مؤشر التمييز لبند الاختبارات المحكية، الهدف منه تقدير احتمالية التوافق بين نواتج سؤال معين ونواتج الاختبار، يعتمد هذا المؤشر على تطبيق الاختبار مرة واحدة، ومن ثم يتم تصنيفهم إلى مجموعتين (متقنة-غير متقنة) بناء على درجة القطع .  
ويحسب مؤشر التوافق المرجعي وفق الطريقة التالية:

$$P = \frac{a+d}{n}$$

حيث أن:

a: عدد الأفراد المتقنين الذين أجابوا اجابة صحيحة على البند

d: عدد الأفراد فير المتقنين الذين أجابوا اجابة خاطئة على البند.

n: عدد الأفراد الكلي.

وتتخصر قيمة المعامل بين (0) و(1)، ويمكن حساب الحد الأدنى لمعامل التوافق المرجعي في حال عدم وجود علاقة بين مستوى الاتقان والاستجابة على السؤال وفق الصيغة التالية:

$$P = \frac{(a+b)(a+c)(c+d)(b+d)}{n}$$

حيث أن:

a: عدد الأفراد المتقنين الذين أجابوا اجابة صحيحة على البند.

b: عدد الأفراد غير المتقنين الذين أجابوا اجابة صحيحة على البند.

c: عدد الأفراد المتقنين الذين أجابوا اجابة خاطئة عن البند.

d: عدد الأفراد غير المتقنين الذين أجابوا اجابة خاطئة عن البند.

n: عدد الأفراد الكلي. (طارق بن ابراهيم جابر الجهني، 2005، ص-ص 37-39)

هـ. معامل ماكنمار

يستخدم هذا الأسلوب في حساب مؤشر الصعوبة للمفردات في الاختبار المحكي، والهدف من الاعتماد عليه هو معرفة مدى تساوي مستوى صعوبة المفردتين وأنّ الفروق في مستوى الصعوبة تعزى إلى أخطاء المعاينة.

وقد اقترح كل من "Birlman& Harris" المعادلة التالية لتقدير مؤشر الصعوبة:

$$X_{MC} = \frac{([b-c]-1)^2}{b+c}$$

(Linda Crocker & James Algina, 2008, p. 334)

وتتم مقارنة النتيجة المتحصل عليها مع القيمة الجدولية بالدرجة الحرية عند مستوى دلالة، فإذا كانت القيمة المحسوبة أكبر من الجدولة دل ذلك على وجود اختلاف في مستوى صعوبة الفقرة. ( عماد غضاب عابنة، 2009، ص 102 )

### 3-2- التحقق من صدق وثبات الاختبار

تعد عملية التحقق من الخصائص السيكومترية للاختبارات جد مهمة، فهي تقدم صورة مهمة حول فعالية الاختبار على تقديم معلومات حول ملمح وأداء المفحوص الحقيقي، وكذا مدى انتماء المفردات للمجال السلوكي للسمة محل القياس، عبر تقدير صلاحية وموثوقية الاختبار.

فقد أكد ( فاروق طباع ، و يوسف خنيش، 2021 ) أنّ كلا من الصدق والثبات عنصرين أساسيين وجوهريين في تقييم جودة الاختبارات، إذ اعتبر أن الصدق يمثل حكم تقييمي لمدى قدرة الأدلة الامبريقية والأسس المنطقية النظرية على تدعيم كفاية وملاءمة التفسيرات والعمليات المعتمدة على درجات الاختبار. بينما الثبات وجب التحقق منه عبر أداء الأفراد من خلال استجاباتهم على الاختبار واتساقها عبر مختلف أبعاد المواقف.

ويتضح من خلال هذا القول أنّ هناك العديد من الأساليب الاحصائية المستخدمة في تقدير مؤشرات الصدق والثبات منها التقليدية لشائعة الاستخدام، ومنها ما هو مرتبط بشكل معين من الاختبارات. والاختبارات المرجعية المحك هو نمط خاص من الاختبارات التحصيلية، يتم تقدير فيه مؤشرات الصدق والثبات بطرق مختلفة إضافة للطرق التقليدية.

### 3-2-1- التحقق من صدق الاختبار مرجعي المحك

تعد عملية التحقق من صلاحية الاختبار من أهم العناصر التي وجب متابعتها، إذ أنّها تعطي فكرة حول جودة المفردات وتمثيلها للسمة المقاسة عبر مستوياتها المختلفة، فهي أكثر أهمية من الثبات، لذا فهي تحتاج إلى التدقيق فيها. ويتم تقدير مؤشرات صلاحية الاختبار المحكي عبر تقدير معامل الصدق وفق مجموعة من الطرق

### 3-2-1-1- الصدق الوصفي

يعد من أبسط الطرق المستخدمة في تقدير صلاحية الاختبار مرجعي المحك، ويتصف بأنه أكثر عمومية من صدق المحتوى كون أنّ الاختبار مرجعي المحك يصف السمة وصفا دقيقا. ويمكن تقدير الصدق الوصفي مثلما أشار إليه علام الوارد في (أميرة عابد البقاعين، 2014) بطرق مماثلة للتي تستخدم في تقدير صدق المحتوى، عبر الاستناد إلى المحكمين للتحقق من أنّ مواصفات الاختبار محددة بدقة،



وكذا مدى تمثيل مفردات الاختبار للنطاق السلوكي. ويتم تقدير مؤشر صدق الوصفي حسب (سالم عبد الله سالم الثبيتي، 2014) من خلال المعادلة التالية:

$$L_{ik} = \frac{(N-1) \sum_{j=1}^n a_{ijk} - \sum_k^n N \sum_{i=1}^n a_{ijk} + \sum_{j=1}^n a_{ijk}}{2(n-1)}$$

حيث أن:

$L_{ik}$ : معامل تجانس المفردة (i) مع الهدف (k)

N: عدد الأهداف

n: عدد المحكمين

ترمز إلى أخذ التقديرات (+1، 0، -1) للمفردة (i) كمقياس بواسطة المحكم (j).

3-2-1-2-الصدق الوظيفي

يهتم هذا الصدق بطبيعة الوظيفة التي صمم من أجلها الاختبار، يساهم هذا النوع من الصدق في التنبؤ بالأداء عبر الوصف الدقيق للسمة. (محمد ابراهيم محاسنة، 2013، ص 158)

3-2-1-3-صدق الانتقاء السلوكي

يعتبر هذا الشكل أدق الأساليب المستخدمة في تقدير صلاحية الاختبار مرجعي المحك حسب (James popham, 1978)، حيث يعد مؤشرا لإيضاح مستوى المفحوص بالنسبة إلى النطاق العام. ويمكن استخدام طريقتين للتوصل إلى أدلة عن صدق الانتقاء السلوكي:

- القيام بتدريس المعارف التي يتضمنها كل نطاق من النطاقات السلوكية المطلوبة لمجموعات صغيرة من الطلبة حتى يصلوا إلى مرحلة التمكن، ومن ثم اختبار امكانية تعميم نطاق منها على نطاقات أخرى.

- الموازنة بين أداء مجموعتين في عدد من الاختبارات تقيس المعارف والمهارات التي يتضمنها كل نطاق من النطاقات السلوكية المطلوبة، ويجب أن تكون أحد المجموعتين تلقت التعليم بهذه النطاقات والأخرى لم تتلق، ثم اختيار النطاق السلوكي المميز بين هاتين المجموعتين باستخدام مؤشر الحساسية للتعليم. (محمد بن علي بن عمر الزيلعي، 2014، ص 29)

ويرى (صلاح الدين محمود علام، 1986) أنه يجب عند التحقق من هذا الصدق من خلال معرفة مدى صلاحية النطاق السلوكي المختار في تمثيل البعد أو الأبعاد المراد قياسه، ويشير إلى أن صدق النطاق السلوكي مرتبط بمفهوم امكانية التعميم التي نادى بها كرونباخ.

3-2-2-التحقق من ثبات الاختبار مرجعي المحك

تتنوع وتختلف طرق تقدير ثبات الاختبارات مرجعية المحك تبعاً لعدد مرات تطبيق الاختبار واستخداماتها، ومن بين هذه الطرائق:

1-2-3 الطرق المعتمدة على تطبيق الاختبار مرة واحدة:

• معامل Livingeston

يعتمد هذا المؤشر في تقدير الثبات على أساليب إحصائية بسيطة من خلال إيجاد انحراف درجات الأفراد عن درجة القطع، كما أنه يركز على مفهوم متوسط مجموع مربعات انحرافات الدرجة الملاحظة وبذلك بحساب قيمة انحراف درجات كل طالب عن درجة القطع في الاختبار، ويتم استخراج قيمة مؤشر الثبات وفق العلاقة التالية:

$$K^2(X.T) = \frac{6^2 T + (M_T - N - C)^2}{6^2 X + (M_X - N - C)^2}$$

حيث أن:

$K^2(X.T)$ : مؤشر Livingeston

$6^2 T$ : متوسط مجموع انحرافات الدرجات حول درجة القطع (C)

$6^2 X$ : متوسط مجموع انحرافات الدرجات الحقيقية عن درجة القطع (C)

M: متوسط درجات الطلبة في النطاق السلوكي الذي يقيسه الاختبار.

N: عدد الأسئلة C: درجة القطع. (سوسن شاكر مجيد، 2014، ص 199)

• معامل الاعتمادية Brennan&kean

يعتمد هذا المعامل على مبادئ نظرية امكانية التعميم الذي يقوم على العلاقة بين تباين الدرجة الشاملة والتباين المتوقع للدرجة الملاحظة كما أوضحها علام.

( طارق الجهني، 2005، ص 45 )

• معامل الاتفاق Subkoviatic:

تقوم هذه الطريقة في تقدير مؤشر الثبات من خلال إيجاد معامل الاتفاق باستخدام أسلوب الاحتمالات، عبر توظيف احتمال تصنيف كل فرد مفحوص إلى صنفين (متمكن-غير متمكن) حسب درجة القطع. ويتم تقدير احتمال تصنيف كل مفحوص تصنيفاً صائباً، بعدها يتم جمع هذه الاحتمالات وإيجاد قيمة المتوسط الحسابي لها، والذي يمثل معامل اتفاق خاص بالمفحوصين الذي يطبق عليهم الاختبار. (محمد محاسنة، 2013، ص 140)

ويمكن حساب مؤشر الاتفاق لـ "Subkoviatic" كالتالي:

$$P_0 = \frac{\sum P_X}{N}$$

حيث أن:

N: عدد أفراد المجموعة

P<sub>X</sub>: معامل الاتفاق التقديري لكل فرد من أفراد المجموعة، والذي يتم حسابه وفق العلاقة التالية

$$P_X = KR20\left\{\frac{x}{n}\right\} + (1 - KR20\left\{\frac{M}{N}\right\})$$

حيث:

: معامل التجانس للاختبار KR20

: درجة الفرد في الاختبار x

: عدد الفقرات n

• معامل "Kappa ل Hayna"

تقوم هذه الطريقة على حساب مؤشر الثبات من خلال تطبيق الاختبار مرة واحدة، ويستند هذا

المؤشر إلى مجموعة من الافتراضات منها:

- أن يشتمل توزيع درجات الطلبة في الاختبار في شكل توزيع  $\beta$  (مشابه لتوزيع ذو الحدين).
  - أن يقيس الاختبار سمة واحدة وأن يكون متضمنا لأهداف محتواها متجانس.
  - انتقاء مفردات الاختبار يكون عشوائيا من النطاق الشامل للمفردات المحددة تحديدا دقيقا.
- ( محمد الزيلعي، 2014، ص 33 )

ويتم تقدير مؤشر الثبات حسب "Hayna" كالتالي:

$$K = \frac{P_{ZZ}^1 - P_{mZ}^2}{P_{1Z}^1 - P_{2AZ}^2}$$

حيث أن:

$P_{ZZ}^1$ : القيمة الاحتمالية من جدول التوزيع  $\beta$  الثنائي وتكون أقل من قيمة z.

$P_{1Z}^1$ : القيمة الاحتمالية المجدولة في توزيع  $\beta$  وتكون قيمة  $P_{1Z}^1$  أقل من قيمة z.

( سوسن شاكر مجيد، 2014، ص-ص 201-202 )

• -معامل Harris

أشار ( سالم الثبيتي، 2014) إلى أنّ هذا المعامل يعتمد على فكرة الارتباط تعتمد على معامل

الارتباط الثنائي، والتي تمثل ارتباط درجة الطالب في الاختبار مع متغير تصنيف الطالب حسب درجة

القطع، ويتم حسابه وفق معامل الارتباط الثنائي الحقيقي (Point Biserial Correlation) كالتالي:

$$MC = \frac{x_1 - x_0}{\sigma} \sqrt{y_1 - y_0}$$

حيث أن:

$x_1$ : متوسط درجات الأفراد المتقنين

$x_0$ : متوسط درجات الأفراد غير المتقنين

$\sigma$ : الانحراف المعياري

$y_1$ : نسبة الأفراد المتقنين

$y_0$ : نسبة الأفراد غير المتقنين.

ولقد أشارت ( سوسن شاكر مجيد، 2014) إلى طريقة أخرى تقوم على تقدير مؤشر الثبات وفق طريقة هاريس تكون عبر حساب مجموع المربعات بين وداخل المجموعات، وذلك عبر العلاقة التالية:

$$Mc = \frac{SSb}{SSb + SSw}$$

حيث:

معامل ثبات هاريس  $MC =$

= مجموع المربعات داخل المجموعات  $SSw$

= مجموع المربعات بين المجموعات  $SSb$

3-2-2-2- الطرق المعتمدة على تطبيق الاختبار مرتين

#### • مؤشر Carver

يعد من الطرق الأولى المقترحة في تقدير مؤشر ثبات الاختبارات مرجعية المحك والمتعلقة باتساق قرارات التصنيف، حيث أشارت (إيمان ذيب، 2013) أنّ هذه الطريقة تقوم على تطبيق اختبارين متوازيين على مجموعة واحدة من الأفراد، بعدها يتم مقارنة نسبة الأفراد المتقنين في كلا الاختبارين، فإذا تقاربت النسبتين دل ذلك على ثبات الاختبارين.

ويمكن إيجاد قيمة معامل Carver وفق مايلي:

$$C_r = \frac{A + D}{N}$$

حيث أن:

A: عدد الأفراد المتقنين

D: عدد الأفراد غير المتقنين

N: المجموع الكلي لعدد الأفراد

• معامل "Kappa" لـ "Swaminthan, Hambelton&Algina"

تعد هذه الطريقة امتدادا لمعامل "Carver" فهي تعتمد على حساب فرق النسبة بين الطلبة المتمكنين وغير المتمكنين في صورتى الاختبارين المتوازيين. ويمكن تقدير مؤشر الثبات حسب "Hambelton&Algina" وفق مايلي:

$$K = \frac{\sum P_{KK} - \sum P_K * P_K}{1 - \sum P_K P_K}$$

حيث أن:

$P_{KK}$ : نسبة الأفراد المتقنين في صورتى الاختبار + نسبة الأفراد غير المتقنين في صورتى الاختبار.  
 $P_K P_K$ : نسبة الأفراد المتقنين في صورتى الاختبار \* نسبة الأفراد المتقنين في الصورة الأولى وغير متقنين في الصورة الثانية \* نسبة الأفراد غير المتقنين في صورتى الاختبار \* نسبة الأفراد غير المتقنين في الصورة الأولى والمتقنين في الصورة الثانية. ( سالم الثبيتي، 2014، ص 35)

4. خلاصة

يتضح لنا أنّ الاختبارات المعيارية والمحكى يعدان امتدادا للاختبارات التحصيلية، والاختلافات فيما بينهما لا تظهر فقط في كيفية تفسير النتائج فقط، بل في خطوات البناء وتحليل الفقرات، فالاختبارات المحكّية المرجع تركز على تحليل جودة الفقرات من خلال تقدير مؤشر حساسية التعليم وعتبة الاتقان، أما الثبات، فهو يتفق مع الاختبارات المحكّية حول شكل تطبيق الاختبار مرتين أو مرة واحدة، غير أنّ الأساليب تختلف، وتعد درجة القطع في الاختبارات المحكّية هي العنصر الرئيسي، فمن خلالها يتم تحديد وتقدير الثبات، والفصل بين الأفراد في عتبة الاتقان.

لذا وجب على الباحثين المهتمين ببناء الاختبارات التحصيلية عند اختيار النمط المناسب في تفسير نتائج الاختبار مراعاة خصوصية كل نمط، وأساليب التحقق من جودة كل من الاختبار المعيارى والمحكى.

5. قائمة المراجع

المراجع باللغة العربية:

1. أميرة عابد البقاعين، بناء اختبار محكي هدفي المرجع في الرياضيات للصف العاشر أساسى. رسالة ماجستير، كلية الدراسات العليا، قسم علم النفس، جامعة مؤتة الأردن. (2014)
2. إيمان عبد الكريم ذيب.. أثر طول الاختبار على تحديد درجة القطع لاختبار تحصيلي محكي المرجع في مادة علم النفس الشخصية. مجلة مداد الآداب، المجلد (1)، العدد(6)، (2013).

3. احمد بوزيان تيغرة.. الاتجاه الايديومتري المحكي في إنشاء الاختبارات التقييمية وتأويلها: منطقه وعلاقته بالاتجاه السيكومتري. تأليف نقادي وآخرون،، قراءات في التقييم التربوي (الإصدار 2). باتنة: جمعية الإصلاح الاجتماعي والتربوي. (1998)
4. سالم عبد الله سالم الثبتي، بناء اختبار تشخيصي مرجعي المحك لقياس مهارات المعلمين في بناء الاختبارات التحصيلية لمقرر الحاسب الآلي. رسالة ماجستير، كلية التربية، قسم علم النفس، جامعة أم القرى، السعودية. (2014).
5. سوسن شاكور مجيد، أسس بناء الاختبارات والمقاييس النفسية والتربوية (الإصدار 3)، عمان، الأردن: مركز ديونو لتعليم التفكير. (2014).
6. صلاح الدين محمود علام، تطورات معاصرة في القياس النفسي والتربوي، (الإصدار 1)، الكويت، الكويت: منشورات كلية الآداب، (1986).
7. صلاح الدين محمود. علام.. القياس والتقييم التربوي والنفسي. أساسياته وتطبيقاته وتوجهاته المعاصرة، القاهرة، مصر: دار الفكر العربي. (2000).
8. صباح حسين حمزة. العجيلي.. القياس والتقييم التربوي (الإصدار 3)، صنعاء- اليمن: مركز التربية للنشر والتوزيع. (2005).
9. طارق بن ابراهيم جابر الجهني، أثر اختلاف طرق تحليل بنود الاختبار المحكي المرجع على اختيار الفقرات والثبات، رسالة ماجستير، كلية التربية، جامعة أم القرى، السعودية. (2005).
10. عماد غصاب عابنة. الاختبارات محكية المرجع. فلسفتها وأسس تطويرها، (الإصدار 1)، عمان، الأردن: دار المسيرة للنشر والتوزيع. (2009)
11. فاروق طباع، ويوسف خنيش.. بعض الأخطاء الناجمة عن جمع أدلة الصدق والثبات في البحوث النفسية والتربوية. مجلة العلوم الانسانية والاجتماعية، المجلد(22)، العدد(1)، (2021).
12. محمد إبراهيم. محاسنة، القياس النفسي في ظل النظرية التقليدية والنظرية الحديثة، عمان- الأردن: دار جرير للنشر والتوزيع. (2013).

13. محمد بن علي بن عمر الزيلعي،. بناء اختبار تشخيصي محكي المرجع لقياس مهارات البحث العلمي لدى طلاب الدراسات العليا بكلية التربية. رسالة ماجستير، كلية التربية، جامعة أم القرى، السعودية. (2014)

14. مريم قارة، وصباح ساعد.. بناء الاختبارات التحصيلية مرجعية المحك: خطوات إجرائية لبناء اختبار تحصيلي مرجعي المحك في مادة الرياضيات للسنة الخامسة ابتدائي. مجلة علوم الإنسان والمجتمع، المجلد (2)، العدد (27)، (2017).

15. هاني، دانة ، وكحول شفيقة، درجة القطع في الاختبارات المحكية المرجع وطرق تحديدها، مجلة الروائز، المجلد(6)، العدد(1)، (2022).

16. يوسف خنيش، صعوبات التقويم في التعليم المتوسط واستراتيجيات الأساتذة للتغلب عليها ، رسالة ماجستير، كلية العلوم الانسانية والاجتماعية، قسم علم النفس وعلوم التربية، جامعة الحاج لخضر، باتنة - الجزائر، (2005)

**المراجع باللغة الأجنبية:**

17. Dnay Leveaut & Jacques Gregory, Introduction aux Theories des Tests en Psychologie et en des Science de L'education, (éd. 3), bruxelle, Belgique: De Book. (2014).

18. James Popham, Criterion Referenced Measurement. Eglewood, U.S.A, (1978)

19. Linda Crocker., & James Algina,. Introduction to Classical and Modern Theory test (1 ed.). U.S.A: Wadswork ,(2008)..