

بناء اختبار تحصيلي محكي المرجع في الرياضيات لتلاميذ السنة الرابعة متوسط

Development and validation of Mathematics criterion referenced achievement test for the fourth grade middle school pupils

مزائني جيلالي¹

جامعة الجيلالي بونعامة خميس مليانة (الجزائر)

mezainid@gmail.com

تاريخ الوصول 2021/11/16 القبول 2023/05/03 النشر على الخط 2023/06/05

Received 16/11/2021 Accepted 03/05/2023 Published online 05/06/2023

ملخص

هدفت هذه الدراسة إلى تطوير اختبار محكي المرجع في الرياضيات لمستوى السنة الرابعة متوسط. و قد طبقت الدراسة على عينة حجمها 173 تلميذ موزعين على ستة أفواج تم اختيارهم بطريقة عشوائية من الأفواج التابعة للمتوسطات ولاية عين الدفلى. و قد تكون الاختبار في صيغته النهائية على عشرين فقرة رباعية البدائل. و قد تم الاعتماد على دلالات صدق المحتوى للتحقق من صدق الاختبار. كما تم استخدام ثلاث طرق مختلفة للتحقق من ثبات الاختبار و هي: ألفا لكرونباك، ليفنجستون، و هاريس. وقد تم اختيار اثنا عشر محكما لتحديد درجة القطع للاختبار باستخدام إجراءات أنجوف. وقد بينت النتائج أن الاختبار الذي تم تصميمه يتوفر على الخصائص السيكومترية المطلوبة لقياس التحصيل.

الكلمات المفتاحية: اختبار محكي المرجع، السنة الرابعة متوسط، الرياضيات، أنجوف.

Abstract :

This study aimed at developing an achievement criterion referenced test in mathematics for the fourth level meduim school pupils. The study sample consisted of 173 pupils randomly chosen from the pupils of the forth level meduim school in Ain Defla . The developed test cosisted of 20 items with four alternatives for each. The content validity was used to assess the test validity inferences.As for reliability, three common types of reliability were used :Alpha, Livingston, and Harris, Besides,twelve subject matter experts (SME) were chosen to set the cut score on the test using the Angoff procedure. The findings of the study revealed that the test developed was valid and reliable for measuring achievement in mathematics.

Keywords: Criterion referenced forth level ,mathematics,Angoff

مقدمة:

إن الهدف الأساسي من عملية التعليم هو مساعدة المتعلم على تحقيق الأهداف التعليمية المسطرة التي يجب أن تتضمن التغييرات المنتظرة في سلوكياته. وعندما ينظر إلى عملية التعليم من هذه الزاوية، فإن التقويم بصفة عامة و الاختبارات التحصيلية بصفة خاصة تصبح جزءا مهما في العملية التعليمية. و تظهر أهمية الاختبارات التحصيلية المصممة بصفة جيدة في حصول المربي على المعطيات التي تساعد في التعرف على المستوى الفعلي للتحصيل الدراسي الذي وصل إليه المتعلم وتزوده بالتغذية الراجعة حول مقدار ما يحدث من تطور لديه. كما تساعد نتائج الاختبارات التحصيلية في إصدار العديد من القرارات الإدارية (مثل: الانتقال من مستوى إلى مستوى أعلى أو الانتقال من مرحلة دراسية إلى مرحلة أخرى) ، إلى جانب المساعدة في إصدار العديد من القرارات المهمة (مثل: عملية التوجيه، و انتقاء ذوي الحاجات الخاصة، إعداد و تقويم البرامج الدراسية). ولقد أشار جرونلند (Gronlund,2009,34) أن المعطيات التي نحصل عليها من الاختبارات المصممة بصفة جيدة ومن مختلف أدوات القياس تساعدنا في الحكم على مدى قابلية الأهداف التعليمية من التحقق ومدى فائدة الوسائل التعليمية و فعاليتها.

الإشكالية:

إن تعرفنا على مدى تحقيقنا للأهداف التربوية المسطرة في المناهج التعليمية يعتمد أساسا على التقويم الموضوعي الذي يجب أن يعكس بصفة شاملة مجال المحتوى المراد قياسه، و الذي يسمح بدوره بالتشخيص الدقيق لنقاط القوة و الضعف و اتخاذ القرارات المناسبة. و تعتبر الاختبارات التحصيلية محكية المرجع من بين الأدوات المناسبة و الضرورية لتحقيق ذلك. فهي ترمي إلى تسهيل تفسير أداء المفحوصين حسب المحتوى الذي يغطيه الاختبار. كما يهدف تطبيق هذه الاختبارات إلى الحصول على درجات تستعمل (1) لوصف مستوى أداء المفحوصين (2) وضع تصنيفات لمستويات الأداء (3) تقييم فعالية البرنامج التعليمي .

وتعتبر برامج مادة الرياضيات من ضمن البرامج التعليمية الأساسية و المهمة في المسار الدراسي و المستقبل المهني للتلاميذ. و يتطلب تقويم الكفايات المكتسبة من خلال دراستهم للبرنامج الدراسي الكثير من الدقة و الموضوعية لاسيما عندما يتعلق الأمر بمستوى الرابعة متوسط. فهذه المرحلة تعتبر حساسة، إذ يتوقع أن ينتقل فيها التلميذ من نمط التعليم العام إلى شعب و تخصصات مختلفة حسب ملامحه الدراسي. و تعتبر الدرجة التي يحصل عليها التلميذ في مادة الرياضيات متغيرا أساسيا يعتمد عليه لتحديد ملامحه و اتخاذ على أساسه قرارات التوجيه. و حتى تقترب هذه القرارات من الموضوعية يجب أن تعتمد على أدوات تقويمية مصممة وفق خطوات علمية تسمح نتائجها بالتنبؤ بالمسار الدراسي و المهني المناسب للتلميذ. كما يجب تشجيع المختصين و الخبراء على تطوير اختبارات تحصيلية تتوفر على الخصائص السيكمترية المطلوبة. و لما كانت طبيعة الاختبارات محكية المرجع تتطلب تفصيلا للمحتوى المراد قياسه قصد تغطية جميع الأهداف، ما يجعل بناء اختبار يسمح بتقويم مدى تحقق جميع الأهداف المسطرة في البرنامج التعليمي غير ممكن ، فيفضل بناء اختبارات تقيس كل وحدة بصفة مستقلة. و لقد أردنا من خلال هذه الدراسة بناء اختبار محكي المرجع في وحدة المعالم و التحقق من خصائصه السيكمترية و قمنا بطرح التساؤلات التالية:

1- ما هو مؤشر صدق المحتوى للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

2- ما مستوى معاملات التمييز وفعالية للمشتتات للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

- 3- ماهي معاملات ثبات الاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟
- 4- ما هي درجة القطع الناتجة عن استخدام طريقة أنجوف للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟ -
- 5 ما مدى ارتباط تقديرات المحكمين الناتجة عن استخدام طريقة أنجوف بمعاملات الصعوبة الفعلية للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

أهمية الدراسة:

ترتبط أهمية الدراسة الحالية بأهمية الاختبار التحصيلي محكي المرجع في الرياضيات في تقييم كفاية تلاميذ السنة الرابعة متوسط في وحدة المعالم، بالإضافة إلى أهمية القرارات الناتجة عن تطبيق طريقة أنجوف لتحديد درجة القطع التي تفسر على أساسها درجات الاختبار. حيث كلما كانت هذه الطرق أكثر فعالية و أكثر صدقا وتوفرت على المعايير المطلوبة قلت أخطاء القياس ليصبح تفسير النتائج صادقا و القرارات الناتجة عن ذلك صحيحة. إضافة إلى أن مرحلة السنة الرابعة متوسط تعتبر مهمة وحساسة. فهي تمثل منعرجا للتوجيه نحو تخصصات مختلفة في المرحلة الثانوية، كما قد ترتبط عند البعض مع نهاية السن الإلزامي للتمدرس. و بالتالي تعتبر القرارات التربوية في هذه المرحلة ذات تأثير كبير على مصير التلميذ. كما يمكن أن تسهم هذه الدراسة في تمكين المربين من التعرف على الخطوات العلمية في بناء الاختبارات التحصيلية محكية المرجع و على إجراءات تحديد درجة القطع باستخدام طريقة أنجوف.

أهداف الدراسة:

تهدف الدراسة الحالية إلى :

- بناء اختبار تحصيلي محكي المرجع في الرياضيات في وحدة المعالم لمستوى السنة الرابعة متوسط.
- توفير أدوات تقويم موضوعية للمعلمين.
- تحديد درجة القطع المناسبة للاختبار وفق الطرق العلمية.

تحديد المفاهيم:

تعرف المفاهيم التي تناولتها الدراسة إجرائيا كالتالي:

الاختبار محكي المرجع

عرف بابام (Popham, 1978) الاختبار محكي المرجع بأنه " ذلك الاختبار الذي يستخدم لتقدير أداء الفرد في نطاق سلوكي محدد تحديدا دقيقا." و يعرفه كل من نيتكو و جليزر (Nitko & Glaser) بأنه: "اختبار يبنى عن قصد لإعطاء درجات قابلة للتفسير المباشر في ضوء مستويات أداء محددة." (سليمان، 2009، 325).

المحك

يعرف كل من جليزر و نيتكو المحك بأنه: " نطاق شامل من المعارف و المهارات المحددة تحديدا جيدا بحيث يمكن موازنة أداء الفرد في الاختبار بهذا النطاق و أن نعرف ما يستطيع أن يؤديه هذا الفرد و ما لا يستطيع ". (علام، 2007، 26). ويقصد بالمحك في الاختبارات محكية المرجع ما يمتلكه الفرد من معارف و مهارات و كفايات في وقت معين (علام، 2007، 27). ويصاغ أداء المحك على صورة كفايات محددة أو نواتج متوقعة أو أهداف سلوكية مرتبة بحيث تصف مختلف مستويات

الأداء (حجاج، 2007، 26). ويعرف إجرائيا بالحد الأدنى من الأهداف السلوكية المرتبطة بوحدة المعالم و التي تم تحديدها مسبقا ليوازن بها أداء أفراد العينة قصد تصنيفهم إلى متمكنين أو غير متمكنين من الوحدة.

درجة القطع

يعرف همبلتون (Hambleton, 1979, 324) درجة القطع بأنها: " نقطة على متصل درجات الاختبار تستخدم لتصنيف الطلاب إلى فئات تعكس مستويات الأداء المختلفة بالنسبة لمجموعة الأهداف المراد قياسها في الاختبار". وعرفها صلاح الدين علام أنها: " الدرجة أو الدرجات التي تفصل بين المتمكنين و غير المتمكنين أو بين مستويات مختلفة من التمكن على متصل درجات الاختبار، و يطلق عليها عدة مصطلحات مثل مستويات التمكن، درجات النجاح أو الاجتياز الحد الأدنى للكفاية، مستويات المحك." (حجاج، 2007، 26). وأما التعريف الإجرائي لدرجة القطع في هذه الدراسة فهي : الدرجة الناتجة عن تطبيق الإجراءات التحكيمية (أنجوف) والتي يتم على أساسها تصنيف أفراد عينة الدراسة إلى متمكنين و غير متمكنين.

طريقة أنجوف Angoff Method

هي طريقة اقترحها أنجوف عام 1971، وتعد من أشهر الطرق التي تستخدم في تحديد درجة القطع من خلال حساب الدرجة المتوقعة على كل مفردة من مفردات الاختبار. و تتطلب من المحكم تقدير احتمال (النسبة المئوية) لإجابة الفرد ذو الحد الأدنى من الكفاية على كل مفردة من مفردات الاختبار إجابة صحيحة، و يتم جمع هذه النسب لكي نحصل على درجة القطع من وجهة نظر كل المحكم . ويتم بعد ذلك حساب متوسط درجات القطع لكل المحكمين لنحصل على درجة القطع النهائية للاختبار.

الاختبارات التحصيلية

مفهوم الاختبار التحصيلي

لقد أوجدت الكثير من التعريفات للاختبار التحصيلي، فقد عرفه جرونلاند (Gronlund, 1977) بأنه " إجراء منظم لتحديد الكم الذي تعلمه التلميذ ". كما عرفه عودة بأنه: "طريقة منظمة لتحديد مستوى تحصيل الطالب لمعلومات و مهارات في مادة دراسية كان قد تعلمها مسبقا بصفة رسمية من خلال إجاباته على عينة من الأسئلة التي تمثل محتوى المادة." (عودة، 2005، 77). و يعرفه أبو حطب بأنه " الأداة التي تستخدم في قياس المعرفة و الفهم و المهارة في مادة دراسية أو تدريبية معينة أو مجموعة من المواد " (أبو حطب، 987، 387).

يتبين من خلال عرضنا لهذه التعريفات أن الاختبار التحصيلي هو أداة منظمة تصمم لقياس مكتسبات المتعلم في مجال دراسي معين، و تستخدم نتائج هذا القياس في اتخاذ القرارات التربوية. وترتبط صحة هذه القرارات بمدى صدق تفسير النتائج. كما أن تطبيق الاختبارات التحصيلية لا ينحصر في نهاية الوحدة الدراسية فحسب ، بل يجب أن يكون خلال جميع مراحل العملية التعليمية وذلك أن المربي مثلاً يحتاج إلى معلومات موضوعية عند بداية، خلال وعند نهاية الوحدة الدراسية. ففي بداية الوحدة قد يحتاج إلى معلومات حول مدى توفر المتعلم على القدرات الضرورية التي تتطلبها الوحدة. كما يحتاج إلى معرفة الصعوبات التي يلاقيها المتعلم خلال العملية التعليمية. أما عند نهاية الوحدة فيحتاج المربي إلى المعلومات التي تخص مدى اكتساب المتعلم للأهداف التعليمية المسطرة.

تصنيف الاختبارات التحصيلية

لقد تعددت التصنيفات المتعلقة بالاختبارات التحصيلية، و ذلك حسب نوع الاستجابة ، طريقة التطبيق صياغة الأسئلة و تفسير الدرجات الناتجة. و لعل أهم تصنيف للاختبارات التحصيلية هو التصنيف المرتبط بطريقة تفسير الدرجات الناتجة عن تطبيق الاختبار والذي ينسب إلى جليزر (Glaser,1963) حيث قسم هذا الأخير الاختبارات التحصيلية إلى قسمين وهما: الاختبارات معيارية المرجع و الاختبارات محكية المرجع.

أ - الاختبارات معيارية المرجع (Norm- referenced test)

رجع تاريخ القياس معياري المرجع إلى الأيام الأولى لظهور اختبار الذكاء، أي عندما صمم بينيه (Binet) أداته الأولى لقياس ما يسمى ب "العمر العقلي" (Gregory,2004). وتم الانتقال من قياس الذكاء إلى القياس التربوي في بداية أربعينيات القرن العشرين عندما قامت مصلحة الاختبارات التربوية ببناء اختبار الاستعدادات الدراسية (SAT) التي أصبحت تستخدم لأغراض الانتقاء و القبول في الولايات المتحدة الأمريكية. و يطلق عادة على الاختبارات معيارية المرجع أسماء مختلفة من بينها "الاختبارات المرجعة للجماعة" أو "الاختبارات السيكمومترية". وهذا النوع من الاختبارات يمدنا بمعلومات عن الوضع النسبي للفرد بالنسبة لأقرانه الآخرين، و ذلك بمقارنة أداء الفرد (درجته) بمعيار مستمد من جماعة معيارية مشابهة. و تتحدد أفضلية الفرد في ضوء ترتيبه النسبي بين أقرانه بغض النظر عن كفاءته. وقد عرف كل من بوفام و هيويسك (Popham & 1969) (Husek, الاختبارات معيارية المرجع بأنها " الاختبارات التي تستخدم لتحديد أداء الفرد بالنسبة لأداء أقرانه ".

تتفق التعريفات المختلفة للاختبارات مرجعية المعيار على أن هذه الاختبارات تعتمد في تفسير الدرجات على مقارنة أداء الفرد بأداء أقرانه. أي أن الدرجات التي نحصل عليها من هذا الاختبار نرجعها إلى أحد الجداول المعيارية. ونستطيع أن نحصل من هذه الجداول على مستوى الأداء النسبي للفرد في ضوء الدرجة التي ينتمي إليها هذا الفرد أو بالنسبة إلى جماعته المعيارية (الشرقاوي و عبد السلام ، 1996 ، 22).

و لما كان الغرض من الاختبارات معيارية المرجع هو المقارنة بين الأفراد، فقد أصبح يعتمد عموما في بناءها على مهام أو فقرات تختلف من حيث الصعوبة و التمييز.

ولكن بالرغم من التاريخ الطويل للقياس معياري المرجع الذي أكسبه وضعية قوية في ميدان القياس بالإضافة إلى الأسس النظرية المثبتة التي يركز عليها ، حيث أكد هملتون أن أغلب الأساليب الإحصائية في كل من النظرية الكلاسيكية و الحديثة في القياس بنيت على خلفية الاختبارات معيارية المرجع (Hambleton,1994)، إلا أنه تعرض إلى كثير من الانتقادات التي أفقدته شيئا من قوته. فقد أشار كل من الشرقاوي و عبد السلام (1996) إلى أنه بالرغم من أن القياس معياري المرجع يزودنا بمعلومات خاصة بقدرة الطالب، و ذلك في ضوء مقارنة أدائه بأداء زملائه، إلا أن هذا النوع من القياس لا يعطينا معلومات عن كفاءة براءة هذا الطالب أو تفوقه في أعمال الموضوع المتضمنة في الاختبار (الشرقاوي و عبد السلام ، 1996 ، 90). حيث أن تحديد موقع الفرد بالنسبة لأقرانه ليس دليلا على تفوقه وليس من الضروري أن يكون متوسط المجموعة المعيارية معيارا كافيا يعكس كفاءة أداء الفرد. كما أشار علام (2000) بأن القياس محكي المرجع قد أثبت قصوره في التقويم التشخيصي وكذلك قصوره في تقويم فعالية البرامج التعليمية من حيث تحقيقها للأهداف المرجوة وبذلك لا يساعد في اتخاذ قرارات تفيد في تطوير هذه البرامج و زيادة فعاليتها(علام ، 2000 ، 329). وقد دفعت الانتقادات المتواصلة الكثير من المختصين في القياس التربوي إلى البحث عن منهجية بديلة تكشف عما حققه المتعلم فعلا و ما لم يحققه من أهداف تربوية، وهذا ما عجل

بظهور ما يسمى بالقياس محكي المرجع. وقد أشار في هذا الصدد همبلتون و أولزاسكي أن "الاختبارات محكية المرجع صممت لغرض تلبية الاحتياجات الخاصة بالنماذج التعليمية الحديثة" (Hambleton & Olzsewski,1972,15)

ب- الاختبارات محكية المرجع

يعود الفضل في ظهور مصطلح القياس محكي المرجع إلى جليزر وذلك منذ أكثر من أربعين سنة، حيث نشر سنة 1963 بحثاً بعنوان "التكنولوجيا التعليمية و قياس نواتج التعليم. بعض الأسئلة". (Instructional Technology and the learning Outcomes.Some questions.) والذي يمثل نقطة الانطلاق الفعلي لعهد القياس محكي المرجع (Glaser,1963). و قد تعددت التعريفات الخاصة بالقياس محكي المرجع و تنوعت ولكن اتفقت في جوهرها. فقد عرف جليزر و نيتكو الاختبار محكي المرجع بأنه: "اختبار صمم عن قصد إعطاء قياسات تفسر مباشرة بالنسبة إلى مستويات أداء محددة" (Glaser & Nitko,1971,160). وتحدد مستويات الأداء عموماً حسب نيتكو و جليزر، بتحديد مجال المهام التي سيؤديها الفرد. وتنظم العينات الممثلة لهذه المهام من هذا المجال في الاختبار و تؤخذ القياسات التي تستخدم لشرح أداء كل فرد لهذا المجال. كما عرف بابام (Popham , 1978) الاختبار محكي المرجع بأنه "الاختبار الذي يستخدم لتقدير أداء الفرد بالنسبة لمحك أو مستوى أداء مطلق دون الحاجة إلى مقارنة أداء الفرد بأداء أقرانه الآخرين". أي أن درجة الفرد تفسر في ضوء محك سابق التحديد و ليس متوسط أداء الجماعة. و هذا المحك هو الذي يحدد مستوى الأداء المطلوب تحقيقه. و قد أشار كل من الشرقاوي و عبد السلام بأن "المحك الذي سيقارن بناء عليه أداء الطالب هو السلوك الذي يحدد كل نقطة على امتداد متصل التحصيل". (الشرقاوي و عبد السلام ، 1996، 160). وتخدم الاختبارات محكية المرجع، حسب همبلتون ، وظيفتين أساسيتين. تتمثل الوظيفة الأولى في أن هذا النوع من الاختبارات يزودنا بمعلومات خاصة حول مستويات أداء الأفراد على الأهداف التعليمية، أما الوظيفة الثانية فتتمثل في تقويم فعالية العملية التعليمية (Hambleton & Olzsewski,1972,15). وبالتالي فإن هذا النوع من الاختبارات يناسب المواقف التعليمية التي تتطلب تحديد درجة إتقان الطالب لمجموعة من المهارات الأساسية المتعلقة بمادة دراسية أو محتوى دراسي معين. ولغرض ضبط حدود ومجالات استخدام الاختبارات معيارية المرجع و الاختبارات محكية المرجع، ذهب الكثير من المختصين في القياس التربوي إلى تحديد نقاط الاختلاف بين النوعين.

تحديد درجة القطع

تعتبر عملية اتخاذ القرارات التربوية في تصنيف المتعلمين إلى متمكنين أو غير متمكنين أو تصنيفهم إلى فئات مختلفة حسب مستوى الأداء ، عملية لا مفر منها. ولكن تحتاج هذه التصنيفات إلى منهجية علمية صحيحة للوصول إلى قرارات تربوية سليمة. و لتحقيق هذا الغرض ظهرت عدة منهجيات لتحديد نقطة فاصلة أو عدة نقاط على متصل الأداء يتم من خلالها تصنيف مستويات التحصيل سميت بطرق تحديد درجات القطع.

كما يبدو واضحاً أن عملية تحديد درجات القطع تعتبر من أهم القواعد في بناء الاختبارات. فهي تؤثر على جميع خطوات بناء الاختبار. فإجراءات هذه العملية تعتبر مهمة معقدة. و في هذا الصدد يشير لارنر بأن مشكل تحديد درجة القطع يشبه تماماً ما يتعامل معه المحامون والقضاة في كل وقت وهو: متى و كيف ننهي القضية ؟ (Cizek & Bunch,2007,8-9).

مفهوم درجة القطع:

أشار زاياكي و ليفنجستون (Livingston & Zieky, 1983) أن تحديد درجة القطع يتمثل في الإجابة على السؤال التالي: "كم يعتبر كافيا ؟" (How much is enough ?). وعليه فإن درجة القطع تدل على مستوى المعرفة أو المهارة التي تعتبر كافية لغرض معين. عرف سيزاك (Cizek) عملية تحديد درجة القطع بأنها: "إتباع صحيح لنظام منطقي من القواعد أو الإجراءات المؤدية إلى تحديد عدد أو رقم يميز بين حالتين أو أكثر أو درجات للأداء." (Cizek & Bunch, 2007, 226). كما أشار كاين (Kane, 1994) إلى أنه من المفيد التمييز بين درجة القطع التي هي عبارة عن نقطة على سلم الدرجة، ومستوى الأداء المعرف بأنه مستوى الحد الأدنى المقبول من الكفاية لغرض معين. فمستوى الأداء هو بمثابة الصيغة النظرية "Conceptual Version" للمستوى المطلوب من الكفاية. أما درجة القطع فتتمثل الصيغة العملية "Operational Version". ففي تحديد مستوى الأداء يتصور المختصون نقطة على طول متصل الأداء، تفصل بين الأداء المقبول و الأداء غير المقبول حسب الهدف. لتكون عملية تحديد درجات القطع عبارة عن ترجمة لهذه النقطة التي تعتبر بمثابة مفهوم مجرد، إلى موقع محسوس على سلم درجة الاختبار بواسطة طرق تحكيمية منتظمة. و قد أطلقت عدة تسميات على درجة القطع و هي: مستويات التمكن Mastery Levels، أو درجات النجاح أو الاجتياز Passing scores، أو الحد الأدنى لمستوى الكفاية Minimum Competency، أو مستويات المحك Criterion Levels (علام، 2007، 261).

الإجراءات المنهجية

منهج البحث

لغرض تحقيق أهداف البحث استخدم الباحث طريقة الاختبار Instrumentation method و التي تعرف بأنها علم بناء الاختبارات (Kpolovie, 2010, 467).

عينة البحث

لغرض جمع بيانات الدراسة تم الإعتماد على عيتين مختلفتين:

عينة التلاميذ: تكونت عينة التلاميذ من 173 تلميذا يزاولون دراستهم في مستوى الرابعة متوسط و موزعين على 06 أفواج تربوية تم اختيارهم بطريقة عشوائية، عن طريق القرعة، من بين 32 فوجا تربويا ينتمون إلى المتوسطات التابعة لمديرية التربية لولاية عين الدفلى.

عينة المحكمين: و قد تكونت عينة المحكمين من اثنا عشر محكما تم اختيارهم بصفة عشوائية من بين أساتذة التعليم المتوسط الذين لديهم خبرة في تدريس الرياضيات لمستوى السنة الرابعة متوسط.

أدوات البحث

لغرض بناء أداة الدراسة المتمثلة في اختبار تحصيلي محكي المرجع في الرياضيات "وحدة المعالم" تم إتباع الخطوات التالية المتفق عليها:

تحديد النطاق السلوكي:

تم تحديد النطاق السلوكي للاختبار عن طريق تحديد محتوى المقرر الدراسي المراد قياسه و تحليله إلى عناصره الأولية. حيث اقتصر اختيار الباحث على وحدة المعالم. وقد اتضح بعد الاطلاع على المقرر الدراسي ودليل الأستاذ لمستوى السنة الرابعة متوسط

التابع لوزارة التربية الوطنية أن المحتوى المعرفي لمقرر مادة الرياضيات لمستوى السنة الرابعة متوسط مقسم إلى ثلاث محاور كبرى، و كل محور يشتمل على مجموعة من الوحدات لكل منها حجمه الساعي. و قد صنفنا وحدة المعالم ضمن محور الأنشطة الهندسية. حيث خصص لها حجم ساعي يقدر بعشر ساعات. وهو يمثل نسبة 20 % من الحجم الساعي المخصص للأنشطة الهندسية.

تحديد الكفايات:

لغرض تحقيق هذه الخطوة قام الباحث بالاتصال بأساتذة مادة الرياضيات للطور المتوسط . و قد بلغ عددهم ثمانية أساتذة فاقت خبرتهم التعليمية خمسة وعشرين سنة. حيث تم الاعتماد على خبرتهم الطويلة بالإضافة إلى دليل الأستاذ في تحديد الكفايات. و قد تم التوصل إلى ثماني كفايات نهائية تمثلت فيما يلي:

- يعرف إحداثيتي شعاع.
 - يقرأ إحداثيتي شعاع في معلم.
 - يمثل شعاع بمعرفة إحداثيتيه.
 - يحسب المسافة بين نقطتين في معلم متعامد و متجانس.
 - يعرف خصائص الشعاعين المتساويين.
 - يعرف أنواع المعالم.
 - يحسب إحداثيتي منتصف قطعة.
 - يعرف إحداثيتا نقطة.
- و قد عرض الباحث الكفايات الثمانية على ثلاث مفتشين للتعليم المتوسط لمادة الرياضيات بالإضافة إلى أستاذين من ذوي الخبرة في تدريس الرياضيات لمستوى السنة الرابعة متوسط لمعرفة مدى أهمية ووضوح هذه الكفايات. وقد اتفقوا بالإجماع على أهمية هذه الكفايات ووضوحها دون حذف أو تعديل.

صياغة الأهداف السلوكية

قصد صياغة الأهداف السلوكية التي تعكس الكفايات، و تحديد المستوى المعرفي الخاص بكل هدف سلوكي ، تمت الاستعانة بنفس أساتذة مادة الرياضيات الثمانية بالإضافة إلى مفتش مادة الرياضيات للطور المتوسط . كما تم الرجوع إلى دليل الأستاذ المصمم من طرف وزارة التربية الوطنية ، وقد تم التوصل إلى صياغة عشرين هدفا سلوكيا يعكس الكفايات .

بعد إتمام خطوة صياغة الأهداف السلوكية تم تصميم استبيان لغرض التحقق من مدى وضوح الصياغة الإجرائية للأهداف السلوكية ومدى انتماء الأهداف السلوكية للكفايات التابعة لها ومدى ملائمة الفعل السلوكي للمستوى المعرفي. و قد تم عرض الاستبيان على تسعة محكمين ستة منهم يمثلون أساتذة الرياضيات لمستوى السنة الرابعة متوسط بالإضافة إلى ثلاثة مفتشين في التعليم المتوسط.

وقد تراوحت نسبة اتفاق المحكمين على وضوح ودقة الصياغة الإجرائية للأهداف السلوكية بين 78% و 89%. وتعتبر هذه النسب مقبولة مما يدل على وضوح الأهداف السلوكية. وقد تم إجراء بعض التعديلات الخاصة بصياغة الأهداف السلوكية بناء على اقتراحات المحكمين.

أما فيما يتعلق بمدى ملائمة الفعل السلوكي للمستوى المعرفي، فبالرغم من أن بعض المحكمين وجدوا صعوبة في الإجابة عليه، و ذلك لصعوبة التمييز بين مختلف المستويات المعرفية و خاصة المستويين الأول و الثاني (التذكر و الفهم)، إلا أنه بعد المناقشات تم التغلب على هذه الصعوبات، و تراوحت نسبة اتفاق المحكمين على ملائمة الفعل السلوكي للمستوى المعرفي بين 78% و 100%. و هي كذلك نسب اتفاق مرتفعة ومناسبة. وقد تم بعد ذلك إحصاء عدد و نسب المستويات المعرفية مثلما هو مبين في الجدول (01).

جدول 01: بين توزيع عدد ونسب كل مستوى من المستويات المعرفية حسب الأهداف.

المستويات المعرفية	تذكر	فهم	تطبيق	المجموع
الأهداف السلوكية	03	12	05	20
النسبة المئوية	15%	06%	25 %	100%

مواصفات الاختبار

قام الباحث بإعداد مواصفات كل فقرة من فقرات الاختبار بصفة واضحة حتى يتسنى بناء فقرات واضحة تحقق هدف الاختبار. وقد احتوى المخطط التفصيلي لمواصفات الاختبار على العناصر التالية: الهدف السلوكي، خصائص المثير، و خصائص الاستجابة.

صياغة الفقرات

قام الباحث في هذه الخطوة بصياغة الفقرات الاختبارية التي تمثل انعكاسا للأهداف السلوكية وذلك بالاسترشاد بجدول مواصفات الاختبار. و قد حرص الباحث على أن يمثل كل هدف من الأهداف السلوكية العشرين التي تم وضعها، بفقرة على الأقل. وقد تمت مراعاة الأسس العلمية في صياغة الفقرات ذات الاختيار من متعدد و التي أشار إليها (Gronlund, 1977, 39-53).

بعد الانتهاء من صياغة فقرات الاختبار والتحقق من توفر المعايير العلمية، قام الباحث بعرض الصورة الأولية على تسعة محكمين متخصصين في ميدان تدريس الرياضيات لمستوى السنة الرابعة متوسط للحكم على مدى وضوح أسلوب صياغة الفقرات ومدى تغطيتها للأهداف السلوكية. ولتحقيق هذا الغرض تم تقديم جدول يضم الأهداف السلوكية و الفقرات الاختيارية، وقد وضعت ثلاث خانات لكل سؤال تضم الأولى التقدير (+1) في حالة الإجابة بنعم. وتضم الثانية التقدير (0) في حالة عدم التأكد من الإجابة. و تضم الخانة الثالثة التقدير (-1) في حالة الإجابة بلا. بالإضافة إلى خانة رابعة تضم الملاحظات لإجراء أي تعديل أو حذف أو إضافة للفقرات الاختيارية بما يتوافق مع الأهداف السلوكية. وبعد جمع المعطيات من المحكمين تم حساب نسبة اتفاقهم حول مدى وضوح الفقرات الاختيارية. وقد تراوحت نسبة الاتفاق بين المحكمين على وضوح الفقرات بين 78 % و 100%. وهي نسب مرتفعة تدل على وضوح صياغة الفقرات. كما أدخلت بعض التعديلات النحوية الطفيفة حسب اقتراحات المحكمين.

تطبيق الاختبار

أ- التطبيق على عينة استطلاعية:

تم تطبيق الاختبار بصورته الأولى (20) فقرة على فوج تربوي من خارج عينة الدراسة بلغ عدده 40 تلميذا من تلاميذ السنة الرابعة متوسط. و قد اختير بصفة عشوائية بغرض تحديد الزمن الفعلي الذي يستغرقه التلميذ في الإجابة على فقرات الاختبار، وكذا التحقق من مدى وضوح فقراته وتعليماته. وبعد تنفيذ إجراءات تطبيق الاختبار على العينة الاستطلاعية توصل الباحث إلى أن تعليمات الاختبار وفقراته كانت واضحة وقد استغرق الاختبار 50 دقيقة.

ب - التطبيق النهائي للاختبار:

تم تطبيق الاختبار على أفراد العينة المكونة من 173 تلميذ من تلاميذ السنة الرابعة متوسط في الأسبوع الأول من الفصل الثالث للسنة الدراسية 2011/2010. و قد تم تطبيق الاختبار على أفراد العينة في مدارسهم وبشكل جماعي و بالاستعانة بأستاذة الرياضيات التابعين للمؤسسة التعليمية التي طبق فيها الاختبار. و قد تكون الاختبار في صورته النهائية من عشرين فقرة رباعية البدائل. يحصل التلميذ على (1) في حالة الإجابة الصحيحة و (0) في حالة الإجابة الخاطئة.

تحديد درجة القطع:

بعد تقديم معلومات تفصيلية عن هدف الدراسة للأستاذة المحكمين، وتوضيح إجراءات استخدام طريقة أنجوف مزودة بأمثلة تطبيقية للوصول إلى درجة القطع و تقديم شرح تفصيلي حول مفهوم التلميذ ذو الحد الأدنى المقبول من الكفاية، قصد الوصول إلى تصور صحيح و نموذج ذهني موحد لهذا الفرد الافتراضي، طلب من كل محكم تقديم تقديرات لكل فقرة من فقرات الاختبار وذلك من خلال الإجابة على السؤال التالي:

من بين مائة تلميذ يمثلون الحد الأدنى المقبول من الكفاية (حسب نظرك) في محتوى المعالم في مادة الرياضيات لمستوى السنة الرابعة متوسط، ما هو عدد التلاميذ الذين يستطيعون الإجابة على الفقرة إجابة صحيحة؟

عرض نتائج البحث ومناقشتها

عرض و مناقشة نتائج التساؤل الأول:

- ما هو مؤشر صدق المحتوى للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

لقد قام الباحث بتقدير صدق المحتوى وهذا ما تحقق سابقا في مرحلة بناء الاختبار، حيث تم إعداد مواصفات الاختبار وتم التحقق من اتفاق المحكمين حول ارتباط الأهداف السلوكية بالكفايات التي تنطوي تحتها وتطابق الفقرات مع الأهداف السلوكية. فقد تراوحت نسبة اتفاق المحكمين حول مدى تطابق الفقرات الاختبارية مع الأهداف السلوكية التابعة لها ما بين 89% و 100%، مما يدل على أنها تعكس صدق الفقرات في تمثيلها للنطاق السلوكي الذي يقيسه الاختبار. كما تم التحقق من مدى انتماء الأهداف السلوكية للكفايات التابعة لها. فقد تراوحت نسبة اتفاق المحكمين بين 89% و 100%. و تعتبر هذه النسب مرتفعة و تعبر عن تمثيل الأهداف السلوكية للكفايات التي تنتمي إليها.

عرض و مناقشة نتائج التساؤل الثاني:

- ما مستوى معاملات التمييز وفعالية للمشتتات للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

قصد التحقق من مستوى معاملات التمييز و فعالية المشتتات للاختبار قمنا بتحليل البيانات التجريبية المتعلقة بالفقرات. وقد تراوحت قيم مؤشر حساسية الفقرات لعملية التدريس بين 0.2 للفقرة رقم 14 و 0.58 للفقرة رقم 18، وجميعها قيم موجبة تدل على حساسية الفقرات لعملية التدريس. والذي يعتبر مؤشرا من مؤشرات صدق الفقرة في الاختبار محكي المرجع (بن عطا و السوالمه، 2008، 41). ولغرض التحقق من وجود فرق بين المتوسطين (أداء التلاميذ) قبل تدريسهم وحدة المعالم وبعد تدريسهم لها تم استخدام اختبار (t) للبيانات المرتبطة، حيث بلغ الفرق بين المتوسطين (7.57) وبلغت قيمة الاختبار (t) (12.76) وهي دالة إحصائية عند مستوى الدلالة أقل من 0.05.

أما فيما يتعلق بفاعلية المشتتات، فقد تم حساب معاملات التمييز لجميع المشتتات، حيث تم حساب الفرق بين نسبة التلاميذ الذين اختاروا بدائل خاطئة في القياس القبلي، و نسبة التلاميذ الذين اختاروا بدائل خاطئة في القياس البعدي (Kubisyn, 2003, 205). وبينت النتائج أن معاملات التمييز للموهبات كانت كلها سالبة و قد تراوحت بين (-0.08 و -0.25). مثلما هو موضح في الجدول (2).

الجدول 2: يبين معاملات التمييز وفاعلية المشتتات لكل فقرة اختبارية.

الفقرة	التمييز	فاعلية المشتتات			الفقرة	التمييز	فاعلية المشتتات		
		الأول	الثاني	الثالث			الأول	الثاني	الثالث
01	0,53	-0,13	-0,1	-0,25	11	0,38	-0,1	-0,1	-0,13
02	0,43	-0,18	-0,2	-0,15	12	0,38	-0,18	-0,15	-0,1
03	0,23	-0,1	-0,1	-0,1	13	0,3	-0,1	-0,1	-0,1
04	0,45	-0,23	-0,15	-0,15	14	0,20	-0,08	-0,08	-0,13
05	0,38	-0,1	-0,1	-0,1	15	0,38	-0,1	-0,15	-0,2
06	0,43	-0,15	-0,15	-0,18	16	0,35	-0,08	-0,12	-0,1
07	0,35	-0,15	-0,15	-0,18	17	0,38	-0,18	-0,12	-0,1
08	0,4	-0,15	-0,15	-0,18	18	0,58	-0,18	-0,18	-0,20
09	0,38	-0,15	-0,15	-0,15	19	0,43	-0,13	-0,15	-0,13
10	0,45	-0,18	-0,13	-0,13	20	0,43	-0,15	-0,05	-0,1

عرض و مناقشة نتائج التساؤل الثالث:

- ماهي معاملات ثبات الاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

نظرا لاختلاف طرق تقدير الثبات في الاختبارات محكية المرجع عن نظيراتها في الاختبارات معيارية المرجع، فقد استخدم الباحث الطرق الثلاثة التالية: أ- طريقة α لكرونباك، ب- طريقة البفنجستون، ج- طريقة هاريس لتقدير ثبات الاختبار الذي تم بناءه. ويلخص الجدول (3) قيم معاملات الثبات الناتجة عن تطبيق الطرق الثلاثة.

جدول 3 : يبين قيم معاملات الثبات الناتجة عن تطبيق الطرق الثلاث.

المعامل	α	$K(XT)$	Mc
الاختبار	0.74	0.79	0.83

يتبين من خلال الجدول أن قيم معاملات الثبات جميعها مقبولة للاستدلال بها على ثبات الاختبار. و يتبين كذلك أن قيمة معامل ليفنجستون أعلى من قيمة ألفا لكرونباك و ذلك أن قيمة معامل ليفنجستون تزداد كلما زاد الفرق بين المتوسط و درجة القطع. و قد تم الحصول على أعلى درجة ثبات عند تطبيق طريقة هاريس.

عرض و مناقشة نتائج التساؤل الرابع:

- ما هي درجة القطع الناتجة عن استخدام طريقة أنجوف للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

قام الباحث بحساب متوسط تقديرات المحكمين حسب ما هو متبع في طريقة أنجوف حيث تم التوصل إلى درجة القطع الخاصة بكل محكم، ثم تم حساب متوسط درجات القطع لمجموع المحكمين. وقد تراوحت المتوسطات الحسابية لتقديرات المحكمين لدرجات القطع للاختبار ككل بين (0,66) و (0,74) بمتوسط حسابي مقداره (0,70) أو ما نسبته وانحراف معياري قيمته (2,57) مثلما يوضحه الجدول (4). وإذا ضربنا درجة القطع النهائية (0,70) في عدد فقرات الاختبار العشرين (20) نجدها تساوي أربعة عشر (14)، أي أنه يمكن اعتبار التلميذ متمكناً من وحدة المعالم إذا أجاب على أربعة عشرة فقرة على الأقل إجابة صحيحة. و يعتبر غير متمكن إذا أجاب على أقل من أربعة عشرة فقرة إجابة صحيحة من المجموع الكلي لفقرات الاختبار.

جدول 4 : يبين درجات القطع المحددة بواسطة المحكمين حسب كل طريقة.

المحكم	درجة القطع	المحكم	درجة القطع
01	68	07	70
02	71	08	74
03	69	09	68
04	71	10	66
05	74	11	67
06	71	12	72
المتوسط	70		
الانحراف المعياري	2.57		
التباين	6.63		

عرض و مناقشة نتائج التساؤل الخامس:

- ما مدى ارتباط تقديرات المحكمين الناتجة عن استخدام طريقتي أنجوف بمعاملات الصعوبة الفعلية للاختبار التحصيلي محكي المرجع في الرياضيات الذي تم بناؤه؟

لغرض التحقق من وجود ارتباط بين تقديرات المحكمين لل فقرات باستخدام طريقة أنجوف ومعاملات الصعوبة الفعلية لها، تم إعداد البيانات لهذا الغرض مثل ما هو مبين في الجدول (5)، بعد ذلك تم استخدام معامل بيرسون لحساب الارتباط. وقد تم استخدام الحزمة الإحصائية للعلوم الاجتماعية (spss) لهذا الغرض.

جدول 5 :يبين متوسط تقديرات المحكمين لل فقرات ومعاملات الصعوبة الفعلية لها.

رقم الفقرة	معاملات الصعوبة	تقديرات المحكمين	رقم الفقرة	معاملات الصعوبة	تقديرات المحكمين
01	0.98	0.88	11	0.75	0.57
02	0.72	0.62	12	0.59	0.69
03	0.93	0.85	13	0.39	0.40
04	0.57	0.65	14	0.60	0.58
05	0.77	0.81	15	0.58	0.65
06	0.86	0.84	16	0.44	0.69
07	0.88	0.78	17	0.81	0.72
08	0.94	0.76	18	0.72	0.74
09	0.88	0.78	19	0.66	0.71
10	0.80	0.68	20	0.82	0.60

يتبين من خلال الجدول (5) أن قيمة معامل الارتباط بين تقديرات المحكمين بطريقة أنجوف ومعاملات الصعوبة الفعلية لفقرات الاختبار بلغت 0.72 . و تدل هذه القيمة على وجود ارتباط قوي وموجب بين معاملات الصعوبة للاختبار الذي تم تطويره و تقديرات المحكمين لدرجات القطع المتعلقة بفقرات الاختبار. وتؤكد هذه النتيجة على أهمية طريقة أنجوف في تقديم تقديرات تتناسب طرديا مع معاملات الصعوبة الفعلية للاختبار مما يجعلنا أكثر ثقة بدرجة القطع المتعلقة بالاختبار.

ومن الدراسات التي تدعم النتيجة الحالية هو ما توصلت إليه دراسة الشريم و سؤالة (2006) التي توصلت إلى أن معامل ارتباط بيرسون بين تقديرات المحكمين بطريقة أنجوف و معاملات الصعوبة الفعلية لفقرات الاختبار بلغ 0.57 في حالة عدم معرفة المحكمين لمعاملات الصعوبة . كما بلغ معامل الارتباط 0.98 بعد تزويدهم بها.

خاتمة

هدفت الدراسة الحالية إلى بناء اختبار تحصيلي محكي المرجع في وحدة المعالم.. وتضمنت الدراسة نوعين من العينات. أما الأولى فعينة المحكمين، وقد تكونت من إثنا عشرة محكما قاموا بتقدير فقرات الاختبار موضوع البحث بطريقة أنجوف للوصول إلى تحديد درجة قطع على الاختبار، وقد تم اختيار مجموعة المحكمين من بين الأساتذة ذوي الخبرة والذين يقومون بتدريس مستوى السنة الرابعة متوسط. أما النوع الثاني فهو عينة التلاميذ التي تكونت من 173 تلميذا من تلاميذ السنة الرابعة متوسط التابعة لمؤسسات التعليم المتوسط لمديرية التربية بعين الدفلى .

وقد تكون الاختبار في صيغته النهائية من عشرين فقرة من الاختبار من متعدد لكل منها أربعة بدائل. وقد تم التحقق من صدق المحتوى للاختبار. كما تم التحقق من مدى فعالية المشتتات و معاملات التمييز للفقرات الاختبارية. وقد بلغت قيم معاملات الثبات باستخدام الطرق الثلاثة التالية : ألفا، ليفنجستون، و هاريس على التوالي: 0.74 ، 0.79 ، 0.83. أما فيما يتعلق بدرجة القطع الناتجة عن تطبيق إجراءات أنجوف فقد بلغت (0,70) أو (70%). و تدل هذه النتيجة على أن الدرجة التي يجب أن يحصل عليها التلميذ في اختبار الرياضيات (وحدة المعالم) حتى يصنف ضمن فئة المتمكنين يجب أن تساوي أو تفوق 12. وقد بينت النتائج وجود ارتباط قوي بين معاملات الصعوبة للاختبار و تقديرات المحكمين لدرجة القطع مما يزيد في ثقتنا في درجة القطع الناتجة عن استخدام إجراءات أنجوف.

قائمة المراجع

المراجع باللغة العربية

- 1- أبو حطب، فؤاد (1987) " التقويم النفسي". القاهرة، الأنجلو المصرية .
- 2- أنور محمد الشرقاوي ونادية محمد عبد السلام (1996) "القياس محكي المرجع و تطويع التعليم (دراسة تجريبية)" ، أنور محمد الشرقاوي، سليمان الخضري الشيخ، أمينة محمد كاظم، نادية محمد عبد السلام (محررون): اتجاهات معاصرة في القياس و التقويم النفسي و التربوي، القاهرة، مكتبة الأنجلو المصرية، ص ص 153-195.
- 3- أنور محمد الشرقاوي ونادية محمد عبد السلام (1996) " الاختبارات المرجعة إلى محك : وسائل جديدة في القياس النفسي و التربوي"، في نادية محمد عبد السلام، سليمان الخضري الشيخ، أمينة محمد كاظم، في نادية محمد عبد السلام (محررون): اتجاهات معاصرة في القياس و التقويم النفسي و التربوي، القاهرة، مكتبة الأنجلو المصرية، ص ص 17-39.
- 4- أنور محمد الشرقاوي ونادية محمد عبد السلام (1996) " بناء اختبار هدي المرجع و جماعي المرجع.(دراسة مقارنة بين النموذجين)(1)"، أنور محمد الشرقاوي، سليمان الخضري الشيخ، أمينة محمد كاظم، نادية محمد عبد السلام (محررون): اتجاهات معاصرة في القياس و التقويم النفسي و التربوي، القاهرة، مكتبة الأنجلو المصرية، ص ص 86-120.
- 5- بن عطاء، زيد .سوالمة، يوسف (2008). " استخدام نموذج المحاولات ذات الحدين في فحص تقديرات المحكمين لدرجة القطع لاختبار مرجعي المحك في الرياضيات". المجلة الأردنية في العلوم التربوية. مجلد4، عدد 1، ص ص 37 - 48 ..
- 6- حجاج غانم (2007) "بحوث معاصرة في القياس النفسي و علم النفس التربوي"، ط1 القاهرة عالم الكتب.
- 7- الشريم، أحمد . و سوالمة، يوسف " تحديد درجة القطع لاختبار محكي المرجع في الرياضيات باستخدام نموذجي " أنجوف " و "نيدلسكي". دراسة مقارنة بمعرفة صعوبة الفقرات و عدم معرفتها". المجلة الأردنية في العلوم التربوية. المجلد 2. العدد الأول. ص ص 10 - 11 (2006)،

- 8- علام، صلاح الدين محمود (2000) "القياس و التقويم التربوي(أساسياته و تطبيقاته و توجهاته المعاصرة)" ، ط 1 ، القاهرة ، دار الفكر العربي.
- 9- علام، صلاح الدين محمود (2007) "الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية و التدريسية " ، القاهرة ، دار الفكر العربي.
- 10- علي محمد سليمان، أمين (2009) "القياس و التقويم في العلوم الإنسانية (أسسه و أدواته و تطبيقاته)" ط 1 ، القاهرة، دار الكتاب الحديث.
- 11- عودة، أحمد (2005) "القياس و التقويم في العملية التدريسية".الأردن ، أريد. دار الأمل للنشر.

المراجع الأجنبية

- 12- Abedi, J(2006) " Language Issues in Item Development". In S. M. Downing & T. M. Haladyna (Ed.), *Handbook of Test Development* (pp. 377–398). Lawrence Erlbaum Associates, Inc. Mahwah, New Jersey.
- 13- Cizek, G J. Bunch. B , M (2007) " Standard setting: A guide to establishing and evaluating performance standards on tests." Sage Publications, Inc.London.
- 14- Downing ,S, M. (2006) " Twelve Steps for Effective Test Development". In S. M. Downing & T. M. Haladyna (Ed.), *Handbook of Test Development*. (pp. 3–25). Lawrence Erlbaum Associates, Inc. Mahwah, New Jersey.
- 15- Glaser, R" Instructional technology and the measurement of learning outcomes. Some questions", *American Psychologist*, 18, 519-523. (1963).
- 16- Gregory, R. J" *Psychological testing: history, principles and applications* " (4th edition). Needham Hights: Allyn & Bacon. (2004).
- 17- Gronlund, N, E, (1977)" *Constructing achievement tests*".(2nd Ed). Prentice-Hill, INC., Englewood Cliffs, N.J.
- 18- Gronlund, N, E.Miller,D ,L.Linn,R,L(2009) "*Measurement and Assessment in Teaching* ". (10th Ed).Pearson Education,Inc,Upper Saddle River,New Jersey.
- 19- Hambleton ,R.Olszewski,F"(1972)*Woodworking Objective and Test Item Bank*".Center for Occupational Education.New York.
- 20- Hambleton,R,K.Eignor,D,R"(1979) *A Practioner Guide to Criterion Referenced Test Development,Validation and Test Score Usage*".(2nd ed).Laboratory of Psychometric and Evaluative Research.Washington,D.C.
- 21- Hambleton, R.K. (1994). The rise and fall of criterion-referenced measurement"?*Educational Measurement: Issues and Practice*, 13(4), 21-27. (1994).
- 22- Kane, M. ". (1994) Validating the performance standards associated with_passing scores". *Review of Educational Research*,64(3), 425–461.
- 23- Kubisyn,T.Borich,G(2003)" *Educational Testing and Measurement : Classroom Application and Practice*".(7th ed).JohnWiley & Sons.Inc.USA.
- 24- Linda Crocker, L. James Algina J(2006) "*Introduction to Classical and Modern Test Theory*". Cengage Learning.USA.
- 25- Livingston, S. A., & Zieky, M. J. (1983)" *A Comparative Study of Standard-Setting Methods*". Educational Testing Service ,Princeton,N.J.
- 26- Popham, W.J., & Husek, T.R. " Implications of Criterion-Referenced Measurement," *Journal of Educational Measurement*, 6(1), 1-9.Cliffs NJ. (1969).
- 27- Popham, W. J. (1978) "*Criterion Referenced Measurement* ". Englewood cliffs NJ. prentic hall.