

# جامعة سعد دحلب البلدية

كلية الآداب والعلوم الاجتماعية

قسم علم النفس وعلوم التربية والأرطوفونيا

## مذكرة ماجستير

تخصص: القياس في علم النفس و التربية

تحديد درجة القطع على اختبار تحصيلي محكي المرجع في الرياضيات

لتلاميذ السنة الرابعة متوسط .

(دراسة مقارنة بين طريقتي أنجوف و نيدلسكاي)

من طرف

مزايبي جيلالي

أمام اللجنة المشكلة من :

رئيسا  
مشرفا و مقورا  
عضوا مناقشا

أستاذ محاضر، جامعة البلدية  
أستاذ محاضر، جامعة البلدية  
أستاذ محاضر، جامعة البلدية

محي الدين عبد العزيز  
بوسالم عبد العزيز  
العبروزي ربيع

البلدية، نوفمبر 2012

## ملخص

تهدف هذه الدراسة إلى مقارنة درجة القطع الناتجة عن تطبيق إجراءات أنجوف و نيدلسكاي بالإضافة إلى معرفة مدى توفر كل طريقة على أدلة للصدق.

تمثلت أداة الدراسة في اختبار محكي المرجع في الرياضيات لمستوى السنة الرابعة متوسط يتكون من عشرين (20) فقرة ذات الاختيار من متعدد.

اشتملت الدراسة على نوعين من العينات. أما النوع الأول فهو عينة المحكمين، و اشتمل على اثنا عشر محكما يمثلون أساتذة السنة الرابعة متوسط. ويملك جميعهم الخبرة في تدريس هذا المستوى. و تمثلت مهمتهم في تقدير أداء المفحوص ذو الحد الأدنى المقبول من الكفاية على فقرات الاختبار و ذلك لغرض التوصل إلى درجة القطع بالطريقتين.

أما النوع الثاني فيمثل عينة تلاميذ السنة الرابعة متوسط الذين يجيبون على أسئلة الاختبار. و قد بلغ عددهم 173 تلميذ.

و قد توصلت الدراسة إلى أن قيمة درجة القطع الناتجة عن استخدام طريقة أنجوف كانت أكبر بصفة دالة إحصائيا عن درجة القطع الناتجة عن تطبيق إجراءات نيدلسكاي.

و بينت النتائج كذلك أن الارتباط بين معاملات الصعوبة الفعلية لفقرات الاختبار و تقديرات المحكمين عند استخدام الطريقتين (نيدلسكاي و أنجوف) كان قويا و موجبا . كما أظهرت النتائج أنه لا يوجد فرق في الاتساق الداخلي للطريقتين ، إلا أن الاتساق الداخلي في طريقة أنجوف كان أكبر.

كما بينت النتائج أن قيمة معامل كابا للتنبؤ بمستوى تحصيل التلاميذ في اختبار الرياضيات

في شهادة التعليم المتوسط كانت موجبة و مقبولة.

## **Abstract**

### **Setting a Cut Score on a Criterion Referenced Achievement Test in Mathematics for the fourth –Grade middle school pupils.**

**( A Comparative Study between Angoff's and Nedelsky's Methods)**

This study aims at comparing Angoff's and Nedelsky's methods to determine the cut score .It also aims at detecting the validity evidence in favour of each method. Therefore, we developed a criterion referenced test in Mathematics for the fourth- grade middle school pupils. The test is composed of (20) twenty multiple - choice items.

The study consisted of two kinds of samples.The first sample consisted of twelve (12) subject matter experts (SEM)(experienced fourth-grade middle school teachers) whose task was to estimate, item by item, the performance of the barely acceptable examinees on the test for which a cut score is desired. The second sample consisted of 173 fourth- grade middle school pupils who were randomly selected and whose task was to answer the test questions.

The final results demonstrated that the group of judges arrived at a significantly higher cutting score value using the Angoff method than when using The Nedelsky procedures.

Furthermore, judges estimates using both methods correlate strongly and positively with the actual difficulty of test items. The results also indicated that the difference in the internal consistency between the two methods was not statistically significant, though, the Angoff method demonstrated somewhat a higher internal consistency .

Finally the study resulted in a positive kappa coefficient for the prediction of the examinees level classification in the middle education certificate (BEM).

## شكر

عن النبي صلى الله عليه و سلم قال " لا يَشْكُرُ اللهُ مَنْ لا يَشْكُرُ النَّاسَ. "  
رَوَاهُ أَحْمَدُ.

نحمد الله الذي سخر لنا عباده المخلصين فأمدونا بالتوجيه والمساعدة راجيا  
من الله العلي أن ينقلها في ميزان الحسنات، وأخص بالذكر الأستاذ  
الأستاذ المشرف الدكتور "بوسالم عبد العزيز" الذي لم يبخل علي بالإرشادات  
والتوجيهات التي كانت لي عون في تقديم هذا البحث.  
كما لا يفوتني تقديم مثل هذا الشكر إلى جميع أعضاء لجنة مناقشة هذا العمل.  
و ما من سبيل في آخر هذه الكلمة سوى أن أتوجه بالشكر الجزيل إلى كل  
من ساعد في إنجاز هذا العمل من قريب أو من بعيد، في السر و العلن، و لو  
بكلمة...

## قائمة الجداول

الرقم	الصفحة
01	عناصر تقييم عملية تحديد درجة القطع. 66
02	إجراءات تصنيف الأسئلة و تقديرات المحكمين في طريقة ايبيل. 68
03	حسابات طريقة نيدلسكي المطبقة على الاختبار . 69
04	المحاور ،الوحدات و الحجم الساعي في مادة الرياضيات. 78
05	نسبة اتفاق المحكمين على وضوح الصياغة الإجرائية للأهداف السلوكية. 80
06	التعديلات التي تم إجراؤها على الأهداف السلوكية. 81
07	مدى انتماء الأهداف السلوكية للكفايات التابعة لها 81
08	نسبة اتفاق المحكمين على ملائمة الفعل السلوكي للمستوى المعرفي. 82
09	عدد ونسب كل مستوى من المستويات المعرفية حسب الأهداف. 83
10	نسبة اتفاق المحكمين حول مدى وضوح صياغة الفقرات. 85
11	مدى تطابق الفقرة مع الهدف السلوكي. 86
12	معاملات التمييز وفاعلية المموهات. 88
13	قيم معاملات الثبات الناتجة عن تطبيق الطرق الثلاث. 91
14	درجات القطع المحددة بواسطة المحكمين حسب كل طريقة 94
15	نتيجة اختبار ولكوكسن. 95
16	16 بعض الدراسات التي توصلت إلى ارتفاع في درجة القطع الناتجة عن طريقة أنجوف مقارنة بطريقة نيدلسكاي 96
17	متوسط تقديرات المحكمين للفقرات (أنجوف، نيدلسكاي) ومعاملات الصعوبة الفعلية لها. 97
18	معاملات الصعوبة الفعلية للمفحوصين ذوي الحد الأدنى من الكفاية قيم معامل صدق القرار وكابا لتصنيف التلاميذ حسب درجات القطع الناتجة 98
19	عن تقديرات المحكمين للطريقتين ونتائج الرياضيات في شهادة التعليم المتوسط. 102

## قائمة الأشكال

الصفحة	الرقم
57	01
59	02
73	03
75	04

## الفهرس

ملخص	
شكر	
قائمة الجداول و الأشكال	
الفهرس	
12.....	مقدمة
	الفصل التمهيدي : مدخل نظري عام للبحث
15.....	1.1. الإشكالية.....
17.....	2. 1. الفرضيات.....
17.....	3.1. أهداف البحث.....
18.....	4.1. أهمية البحث.....
18.....	5.1. حدود البحث.....
19.....	6.1. تحديد مفاهيم البحث.....
22.....	7.1. الدراسات السابقة.....
22.....	7.1. 1. الدراسات العربية.....
27.....	7.1. 2. الدراسات الأجنبية.....
33.....	8.1. تعقيب على الدراسات السابقة.....
34.....	9.1. علاقة البحث الحالي بالدراسات السابقة.....
	الفصل الثاني : بناء الاختبارات التحصيلية محكية المرجع
35.....	تمهيد.....
35.....	1.2. مفهوم الاختبار التحصيلي.....
36.....	2.2. تصنيفات الاختبارات التحصيلية.....
36.....	2.2. 1. الاختبارات معيارية المرجع.....
38.....	2.2. 2. الاختبارات محكية المرجع.....

39.....	3.2. الفرق بين الاختبارات محكية المرجع و الاختبارات معيارية المرجع.
41.....	4.2. خطوات بناء الاختبارات محكية المرجع.
41.....	4.2. 1. تحديد النطاق السلوكي المراد قياسه.
42.....	4.2. 2. تحديد الكفايات التي سيقاسها الاختبار.
42.....	4.2. 3. تحليل الكفايات إلى أهداف إجرائية (سلوكية).
42.....	4.2. 4. تحديد مواصفات الاختبار.
43.....	4.2. 5. كتابة مفردات الاختبار.
44.....	4.2. 6. تجميع مفردات الاختبار.
44.....	4.7.2. تجريب الاختبار.
44.....	4.2. 8. تحليل فقرات الاختبار.
44.....	4.2. 8.1. صعوبة البند.
45.....	4.2. 8.2. معامل القياس القبلي _ البعدي.
46.....	4.2. 8.3. طريقة معامل التمييز لمجموعة الدارسين و غير الدارسين.
47.....	4.2. 8.1. طريقة معامل التوافق المرجعي.
48.....	4.2. 8.5. طريقة معامل فاي.
49.....	4.2. 9. تقدير صدق الاختبارات محكية المرجع.
50.....	4.9. 1. الصدق الوصفي.
50.....	4.9. 2. الصدق الوظيفي.
51.....	4.9. 3. صدق انتقاء النطاق السلوكي.
51.....	4.10. ثبات الاختبارات محكية المرجع.
52.....	4.10. 1. طرق تعتمد على تطبيق الاختبار مرة واحدة.
54.....	4.10. 2. طرق تعتمد على تطبيق الاختبار مرتين.
55.....	ملخص الفصل.



## الفصل الثالث : تحديد درجة القطع

تمهيد

- 1.3 . مفهوم درجة القطع..... 56
- 2.3. العلاقة بين مستويات المحتوى و مستويات الأداء ..... 58.
- 3.3 . لمحة تاريخية حول مراحل تطور منهجية تحديد درجات القطع ..... 58
1. 3.3 . مرحلة السذاجة..... 59.
2. 3.3 . مرحلة النهضة ..... 59
3. 3.3 . مرحلة الخيبة ..... 60.
- 4.3.3 . مرحلة التقبل الحقيقي..... 61.
- 4.3. التطورات الحالية لعملية تحديد درجة القطع..... 61.
- 5.3 . لخطوات المشتركة بين مختلف طرق تحديد درجة القطع..... 62
- 1.5.3 . تحديد الهدف..... 62.
- 2.5.3 . اختيار نموذج لتحديد درجة القطع..... 63.
- 3.5.3 . كتابة مواصفات مستويات الأداء..... 63.
- 4.5.3 . المفاهيم الأساسية..... 63.
- 5.5.3 . اختيار و تدريب المشاركين في تحديد درجة القطع ..... 64.
- 6.5.3 . تزويد المشاركين بالتغذية الراجعة. .... 65.
- 7.5.3 . تقييم العملية..... 65.
- 6.3 . طرق تحديد درجة القطع ..... 67.
- 1.6.3 . طرق تركز على بنود الاختبار..... 67.
1. 1.6.3 . طريقة ابييل..... 67.
2. 1.6.3 . طريقة نيدلسكاي..... 68.
3. 1.6.3 . طريقة أنجوف ..... 70.
- 2.6.3 . طرق تركز على المفحوصين..... 72.
- 1.2.6.3 . طريقة المجموعات المحكية ..... 72.
- 2.2.6.3 . طريقة المجموعات المتضادة..... 73.
- 3.2.6.3 . طريقة المجموعات الحدية ..... 74.

76.....	ملخص الفصل
	الفصل الرابع: الإجراءات المنهجية للبحث.
77.....	1.4 . مجتمع و عينة البحث
77.....	1. 1.4 . مجتمع التلاميذ
77.....	1. 1. 1.4 . عينة التلاميذ
77.....	2 . 1.4 . مجتمع المحكمين
77.....	1. 2 . 1.4 . عينة المحكمين
78.....	2.4 . منهج البحث
78 .....	3.4 . بناء الاختبار
78.....	1. 3.4 . تحديد النطاق السلوكي
79.....	2. 3.4 . تحديد الكفايات
79.....	3. 3.4 . صياغة الأهداف السلوكية
83.....	4. 3.4 . مواصفات الاختبار
83.....	5. 3.4 . صياغة الفقرات
87.....	6. 3.4 . تطبيق الاختبار
87.....	6.1. 3.4 . التطبيق على عينة استطلاعية
88.....	6.2. 3.4 . التطبيق النهائي للاختبار
89.....	7. 3.4 . تقدير صدق وثبات الاختبار
91 .....	4.4 . تطبيق إجراءات الطريقتين (أنجوف، نيدلسكاي)
91.....	1. 4.4 . طريقة أنجوف
92.....	2. 4.4 . طريقة نيدلسكاي
	الفصل الخامس: عرض و تفسير النتائج
93.....	1.5. التحقق من الفرضية الأولى
96.....	2.5 . التحقق من الفرضية الثانية
98.....	3.5 . التحقق من الفرضية الثالثة
101.....	4.5 . التحقق من الفرضية الرابعة
104.....	خلاصة النتائج

107.....	خاتمة
109.....	الملاحق
126.....	المراجع

## مقدمة

تحظى عملية تصنيف المتعلمين، حسب مستوى التمكن من الأهداف التدريسية باهتمام كبير في مجال القياس و التقييم التربوي و النفسي. حيث أنها تساعد المربين على اتخاذ قرارات تعليمية مهمة، تتعلق بحاضر المتعلمين و مستقبلهم ، ويعتمد هذا التصنيف على مدى ما حققه المتعلم من معارف و مهارات تتعلق بمحتوى دراسي معين في ضوء محك أداء مطلق .

ونظرا لأهمية تصنيف المتعلمين في الميدان التربوي لأغراض مختلفة من جهة، وأهمية القرارات الناتجة و أبعادها المستقبلية ، ظهر ما يسمى بطرق تحديد درجة القطع. و قد ارتبط ظهورها وتطورها بتطور القياس محكي المرجع. و يختلف هذا الأخير عن القياس معياري المرجع في مجموعة من النقاط. و من بينها كيفية تفسير أداء الفرد. حيث يفسر أداءه في القياس معياري المرجع عن طريق مقارنته مع متوسط الجماعة المعيارية. أما في القياس محكي المرجع فيفسر أداء عن طريق مقارنته بمستوى محكي معد مسبقا. ويمثل هذا المستوى الحد الأدنى المقبول من الكفاية و يترجم عن طريق إجراءات تحكيمية إلى نقطة أو مجموعة من النقاط على متصل الأداء تسمى بدرجة القطع .

وقد تعرضت عملية تحديد درجة القطع بصفة عامة إلى حملة واسعة من الانتقادات أدت إلى التشكيك في جميع الطرق المستخدمة في تحديد درجة القطع، و يرجع ذلك لكون المقارنات التي تمت بين مختلف الطرق توصلت إلى درجات قطع مختلفة. و لم ينحصر الاختلاف بين الطريقتين فحسب، بل ظهر الاختلاف بين تقديرات الحكمين الذين ينتمون لمجموعة واحدة و يطبقون طريقة واحدة. وقد وصل الأمر ببعض المختصين في مجال القياس إلى الدعوة للتخلي نهائيا عن الطرق التحكيمية في تحديد مستويات الأداء و الاعتماد بالمقابل على بيانات الإختبار. و قد كانت هذه الانتقادات بمثابة حجر الزاوية للانطلاق في مراجعة مختلف الطرق وذلك بالبحث في النقائص العالقة بكل طريقة ومحاولة ضبط شروط تطبيقها حتى تكون أكثر موضوعية وتكون بذلك درجة القطع أكثر صدقا.

لكن الواضح من الواقع التربوي في بلادنا أن الثورة العلمية التي حصلت في مجال القياس التربوي وفي مجال تحديد مستويات الأداء لم تؤثر في النظام التربوي المحلي، فما زال هذا الأخير يعتمد اعتمادا كلياً على الطرق التقليدية لتحديد درجة القطع في التصنيف الثنائي للتلاميذ

(ناجح/راسب). و غالبا ما تكون هذه الدرجة موحدة و ثابتة بين مختلف المواد و مختلف المستويات حيث تمثل نسبة 50 % من الدرجة الكلية للاختبار، وتتخذ على أساسها قرارات تربوية مهمة تمس حاضر التلميذ و مستقبله. فهذه الوضعية تجعلنا نتساءل عن مدى صحة القرارات المتخذة في ضوء هذه الدرجة. ذلك أن عدم استناد درجة القطع على أساليب منهجية منظمة يجعل قرارات التصنيف مشوبة بقدر كبير من الخطأ. [1](ص:261).

وقد سعينا من خلال هذا البحث على تطبيق طريقتين لتحديد درجة القطع على اختبار محكي المرجع في الرياضيات لتلاميذ السنة الرابعة متوسط. وهما طريقتا نيدلسكاي و أنجوف. و تعتبر الطريقتان من الطرق الأكثر شيوعا في مجال القياس. حيث تصنفان ضمن الطرق التي تركز على فقرات الاختبار (test-centred) إذ تعتمدان على تقديرات المحكم لأداء المفحوص ذو الحد الأدنى من الكفاية على الفقرة لتحديد درجة القطع و التي تتم غالبا قبل تطبيق الاختبار و الحصول على المعلومات المرتبطة ببيانات الصعوبة الفعلية لأداء المفحوص على فقرات الاختبار. وكان الغرض من تطبيق الطريقتين هو المقارنة بينها لمعرفة الطريقة التي تتوفر على أدلة كافية على صدق درجة القطع. و لتحقيق هذا الغرض تم تقسيم البحث الحالي إلى مايلي:

تم في الفصل تمهيدي ، و هو عبارة عن مدخل نظري عام للدراسة ، عرض الإشكالية حيث أردنا معرفة وجود أو عدم وجود اختلاف في قيمة درجة القطع الناتجة عن تطبيق الطريقتين، كما حاولنا التعرف على مدى وجود ارتباط بين تقديرات المحكمين الناتجة عن تطبيق الطريقتين و معاملات الصعوبة الفعلية لفقرات الاختبار، إلى جانب التعرف على الطريقة التي تتمتع باتساق داخلي أكبر و كذا الطريقة التي لها قدرة على التنبؤ بتصنيف تحصيل التلاميذ في شهادة التعليم المتوسط. كما تم عرض في نفس الفصل الدراسات السابقة التي لها علاقة بالدراسة الحالية ، وقد تمت مناقشة نتائجها و أخذ بعض التوصيات التي خرجت بها لاستغلالها لأغراض الدراسة الحالية.

و قد قمنا في الفصل الثاني بتقديم مفهوم الاختبارات التحصيلية و أنواعها إلى جانب تبيان أوجه الاختلاف بين الاختبارات محكية المرجع و الاختبارات معيارية المرجع، بالإضافة إلى تحديد الخطوات المنهجية المعتمدة في بناء الاختبارات محكية المرجع.

أما في الفصل الثالث فقد تطرقنا إلى منهجية تحديد درجة القطع و مراحل تطورها بالإضافة إلى شرح مختلف الخطوات المشتركة بين مختلف الطرق لتحديد درجة القطع

والخطوات الخاصة ببعض الطرق التي تركز على فقرات الاختبار و الطرق التي تركز على أداء المفحوصين و من بينها الطريقتان المعتمدتان في الدراسة الحالية. أما الفصل الرابع فتم تخصيصه لشرح و توضيح الإجراءات المنهجية المتبعة لتحقيق أهداف البحث، ومن أهمها الخطوات التطبيقية التي اتبعها الباحث لبناء الاختبار التحصيلي محكي المرجع في وحدة المعالم، إلى جانب الإجراءات المنهجية المتبعة لتطبيق طريقتي أنجوف و نيدلسكاي لتحديد درجة القطع. و تم في الفصل الأخير تحليل النتائج المتوصل إليها و تفسيرها و مناقشتها.

## الفصل التمهيدي 1

### مدخل نظري عام للبحث

#### 1.1. الإشكالية:

تشكل عملية تحديد درجات القطع الخطوة الأكثر صعوبة و الأكثر جدلا في خطوات بناء الاختبارات محكية المرجع، حيث يشير هانش (Hansche,1998) أن المشكل الأكثر تحديا اليوم في القياس التربوي يتعلق بتحديد درجات القطع على سلم درجة الاختبار في تصنيف المتعلمين إلى فئتين أو أكثر حسب مستوى الأداء. [2](ص:87). ويرجع سبب هذا التحدي إلى عدم وجود درجة قطع صحيحة (True Cut Score) تحتاج أن تكتشف من خلال البحوث و الدراسات، ولكن يحتاج تحديد هذه الدرجة إلى إتباع طرق تحكيمية وفق معايير معينة. ولهذا الغرض تم اقتراح العديد من الطرق حيث بلغ عدد ها حسب بارك (Berk ,1986) . [3] ثمان وثلاثين طريقة، كما تضاعف عددها منذ ذلك الوقت، لتبلغ حسب وانغ (Wang et al,2003) . [4] أكثر من خمسين طريقة سنة 2003. و يرى بارك (Berk ,1986) أن مستويات الأداء أخذت الاهتمام الأكبر من بين مواضيع القياس التي تم فحصها في أدبيات القياس محكي المرجع خلال الثلاثين سنة الماضية. [3].

ونظرا لاعتماد أغلب طرق تحديد مستويات الأداء على تقديرات المحكمين، والتي تمثل القاعدة التي تبنى عليها درجة القطع، حيث يرتبط صدق هذه الدرجة بمدى دقة و صدق التقديرات وانعكاس ذلك على القرارات المتخذة في المجالات المختلفة، كان من الضروري و المهم القيام بالدراسات و البحوث لإيجاد الإطار الذي يساعد المحكمين في تقديم تقديرات ذات معنى، يمكن الدفاع عنها و من ثم الوصول إلى درجات قطع صادقة و قرارات صحيحة. وقد تمحورت أغلب هذه الدراسات حول البحث في كفاية المحكمين، عددهم ومستوى تدريبهم، خصائصهم و كم و نوع المعلومات التي يحتاجونها وعدد الجولات المناسب للوصول إلى درجة القطع النهائية. بالإضافة إلى مدى قدرة المحكمين على تكوين تصور مشترك وصحيح لمفهوم الفرد ذو الحد الأدنى من الكفاية و قدرتهم على تقديم تقديرات دقيقة لأداء هذا الفرد الافتراضي على البند .

و من النتائج المترتبة عن عدم دقة التقديرات عند تطبيق مختلف الطرق هو ظهور أخطاء التصنيف. وهما خطأ التصنيف الأول ( تصنيف الطالب المتقن ضمن فئة غير المتقنين) و يشار إليه بالخطأ السلبي و الثاني ( تصنيف الطالب غير المتقن ضمن فئة المتقنين) و يشار إليه بالخطأ الايجابي. وفي هذا الصدد أشارت نادية عبد السلام (2006) بأن خطورة خطأ القياس المرتبط بدرجات القطع تكمن في القرارات التي يجب أن يتخذها المربون بناء على هذه الدرجات ، حيث أنه في حالة الخطأ ألفا ( $\alpha$ ) ينتقل الطلاب غير المتقنين إلى مرحلة تعليمية تالية دون تمكنهم من المهارات السابقة ، فلا يمكنهم استيعاب مهام هذه المرحلة الجديدة. وفي حالة الخطأ بيتا ( $\beta$ ) يكلف الطلاب المتقنون بإعادة تعلم مهام قد أتقنوها وهم في غنى عن تعلمها من جديد ، فضلا عن الأثر النفسي السلبي الذي يلحق بهم . [5](ص:20).

و يعتبر الاتساق الداخلي لتقديرات المحكمين من أهم المحكات الدالة على الصدق الداخلي وكفاية درجة القطع. إذ يمثل واحدا من الاحتمالات القليلة "الموضوعة الذاتية" Objectifying the (subjectivism) من خلال ربط تقديرات المحكمين بالحقيقة.[6]. ويرتبط تفضيل طريقة على أخرى بمدى إفراز الطريقة لتقديرات حول صعوبة الفقرة تكون أكثر اتساقا مع الأداء الفعلي للمفحوصين المستهدفين. [7]. كما يمثل ضعف الاتساق الداخلي لتقديرات المحكمين أحد أهم المعوقات الأساسية في معظم الطرق التي تركز على فقرات الاختبار. إذ تشير أغلب المعطيات التجريبية المتراكمة عبر السنوات إلى صعوبة التنبؤ بمستوى كفاية المفحوصين. [8].

وإضافة إلى ما سبق فإنه بالرغم من العدد الكبير من الدراسات التي اهتمت بطرق تحديد درجة القطع إلا أن النقاش لا يزال مفتوحا حول أهمية كل طريقة، فلا توجد لحد اليوم إرشادات واضحة في اختيار الطريقة المناسبة. و تشير سوميه شكري إلى أن هناك ندرة في الدراسات التي حاولت دراسة صدق درجات القطع ، بل ركزت معظم الدراسات على مقارنة الدرجات الناتجة عن استخدام طرق مختلفة في حساب درجات القطع. [9] (ص:7)

و يرى كاين [10] (Kane,2001) أن هناك ثلاثة أنواع من الأدلة على الصدق (إجرائية داخلية و خارجية). يهتم النوع الأول بمدى إتباع الخطوات المنهجية للطريقة. أما النوع الثاني فيرتبط بالبيانات الداخلية الناتجة عن إجراءات تحديد درجة القطع مع التركيز على اتساق النتائج. أما النوع الثالث فيرتكز على مقارنة النتائج مع مصادر خارجية.



ونحاول من خلال هذا البحث المقارنة بين طريقتين تحكيميتين من طرق تحديد درجة القطع (نيدلسكاي و أنجوف) وبحث مدى توفر الأدلة المرتبطة بصدق الطريقة، و ذلك من خلال الإجابة على الأسئلة التالية:

- أ- هل تختلف قيمة درجة القطع المحددة على الاختبار التحصيلي محكي المرجع في الرياضيات تبعاً لاختلاف الطريقة المستخدمة (أنجوف، نيدلسكاي) ؟
- ب- ما مدى ارتباط تقديرات المحكمين الناتجة عن استخدام طريقتي أنجوف و نيدلسكاي بمعاملات الصعوبة الفعلية لفقرات الاختبار؟
- ج- هل يوجد فرق في الاتساق الداخلي بين طريقتي أنجوف نيدلسكاي ؟
- د- ما مدى قدرة طريقتي أنجوف و نيدلسكاي على التنبؤ بتصنيف مستوى التلاميذ في مادة الرياضيات في شهادة التعليم المتوسط ؟

### 1.2. الفرصيات :

1.2.1. تختلف قيمة درجة القطع المحددة على الاختبار التحصيلي محكي المرجع في الرياضيات تبعاً لاختلاف الطريقة المستخدمة.(أنجوف، نيدلسكاي).

2.2.1. يوجد ارتباط موجب بين تقديرات المحكمين باستخدام طريقتي أنجوف و نيدلسكاي و معاملات الصعوبة الفعلية لفقرات الاختبار.

3.2.1. لا يوجد فرق في الاتساق الداخلي بين طريقتي أنجوف و نيدلسكاي .

3.2.1. تعتبر طريقتي أنجوف و نيدلسكاي ذات أهمية في التنبؤ بتصنيف مستوى التلاميذ في مادة الرياضيات في شهادة التعليم المتوسط .

### 3.1. أهداف البحث :

يهدف البحث الحالي إلى معرفة :

- أ- مدى اختلاف درجة القطع باختلاف الطريقة المستخدمة (نيدلسكاي و أنجوف ) .
- ب- أي الطريقتين أفضل في تحديد درجة القطع لاختبار محكي المرجع في الرياضيات.
- ج- مدى ارتباط تقديرات المحكمين الناتجة عن استخدام طريقتي نيدلسكاي و أنجوف بمعاملات الصعوبة الفعلية .
- د- الطريقة التي تتميز بأكبر اتساق داخلي لتقديرات المحكمين.

هـ - مدى أهمية استخدام كل من طريقتي نيدلسكاي و أنجوف في التتبؤ بتصنيف التلاميذ في مادة الرياضيات .

#### 4.1. أهمية البحث:

ترتبط أهمية البحث الحالي بأهمية القرارات الناتجة عن تطبيق طريقتي نيدلسكاي وأنجوف لتحديد درجة القطع التي تفسر على أساسها درجات الاختبار. حيث كلما كانت هذه الطرق أكثر فعالية و أكثر صدقا وتوفرت على المعايير المطلوبة كلما قلت أخطاء القياس ليصبح تفسير النتائج صادقا و القرارات الناتجة عن ذلك صحيحة. إضافة إلى أن مرحلة السنة الرابعة متوسط تعتبر مهمة وحساسة ، حيث تعتبر منعرجا للتوجيه نحو تخصص معين 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 إيجاز أهمية هذا البحث في النقاط التالية:

- أ- إعداد اختبار محكي المرجع في الرياضيات، وحدة المعالم، حتى يكون مرجعا يسهم في مساعدة المربين في التقويم و بناء الاختبارات بإتباع نفس الخطوات .
- ب- المقارنة بين طريقتي أنجوف و نيدلسكاي ، حيث تنعكس أهمية هذه المقارنة في معرفة الطريقة الأكثر فعالية و التي تتوفر على أدلة للصدق، أي الطريقة التي تعطي درجة قطع يمكن الثقة فيها و الدفاع عنها و الدفاع عن القرارات التربوية الناتجة عنها.
- ج- يفتح هذا البحث المجال أمام البحوث الأخرى التي تتناول الطرق المختلفة لتحديد درجة القطع.

#### 5.1. حدود البحث:

سوف يقتصر البحث الحالي على ما يلي :

- أ- بناء اختبار محكي المرجع ذات الإختيار من متعدد في وحدة المعالم في الرياضيات لمستوى السنة الرابعة متوسط.
- ب- عينة من مجتمع تلاميذ السنة الرابعة متوسط في الفصل الثالث من السنة الدراسية 2011/2010.

## 6.1. تحديد مفاهيم البحث:

### 1.6.1. الاختبار محكي المرجع:

عرف بابام، (Popham, 1978) الاختبار محكي المرجع بأنه " ذلك الاختبار الذي يستخدم لتقدير أداء الفرد في نطاق سلوكي محدد تحديداً دقيقاً. " [11]

و يعرف كل من جليزر و نيتكو، (Nitko & Glaser, 1978) الاختبار المحكي المرجع بأنه: "اختبار يبنى عن قصد لإعطاء درجات قابلة للتفسير المباشر في ضوء مستويات أداء محددة ". [12] (ص:325)

أما التعريف الإجرائي للاختبار محكي المرجع فهو: اختبار الكفايات الرياضية في موضوع المعالم لتلاميذ السنة الرابعة متوسط من إعداد الباحث الذي اتبع في بناءه الأساليب و الطرق المتعارف عليها في بناء مثل هذا النوع من الاختبارات.

### 2.6.1. المحك:

يعرف كل من جليزر و نيتكو المحك بأنه : " نطاق شامل من المعارف و المهارات المحددة تحديداً جيداً بحيث يمكن موازنة أداء الفرد في الاختبار بهذا النطاق و أن نعرف ما يستطيع أن يؤديه هذا الفرد و ما لا يستطيع ". [1] (ص:26)

و يقصد بالمحك في الاختبارات محكية المرجع ما يمتلكه الفرد من معارف و مهارات و كفايات في وقت معين . [1] (ص:27).

و يصاغ أداء المحك على صورة كفايات محددة أو نواتج متوقعة أو أهداف سلوكية مرتبة بحيث تصف مختلف مستويات الأداء. [5] (ص:26)

أما في التعريف الإجرائي فيعبر عن المحك بالحد الأدنى من الأهداف السلوكية المرتبطة بوحدة المعالم و التي تم تحديدها مسبقاً ليوافق بها أداء أفراد العينة قصد تصنيفهم إلى متمكنين أو غير متمكنين من الوحدة.

### 3.6.1. الفرد ذو الحد الأدنى من الكفاية:

عرف ريكايس ( Reckase, 2000 ) الفرد ذو الحد الأدنى من الكفاية بأنه: "الفرد الذي يحصل على الحد الأدنى المقبول لدخوله ضمن فئة المتمكنين و هو شخصية افتراضية يحددها المحكم". [5].

و يمكن تعريفه في البحث الحالي بأنه تلميذ السنة الرابعة متوسط الذي يمتلك حد أدنى من الكفايات و المعارف اللازمة في وحدة المعالم ليصنف ضمن فئة المتمكنين.

### 4.6.1. الاتساق الداخلي:

يقصد بالاتساق الداخلي مدى الاتفاق بين تقديرات المحكمين و البيانات التجريبية. وأول من أدخل هذا المصطلح إلى ميدان القياس هو فندر لندن (Van der Linden) عام 1982. غير أن له معاني أخرى مختلفة . حيث يمكن أن يستخدم للتعبير عن درجة الاتساق بين تقديرات المحكم الواحد في الجولات المختلفة لتحديد درجة القطع [8] .

و قد استخدم مصطلح الاتساق الداخلي في هذه الدراسة للتعبير عن مدى الاتفاق بين تقديرات المحكمين لفقرات الاختبار و معاملات الصعوبة الفعلية للفرد ذو الحد الأدنى المقبول من الكفاية.

### 4.6.1. درجة القطع:

يعرف همبلتون [13] (ص: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 (1991) 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". [5] (ص:13).

أما التعريف الإجرائي لدرجة القطع في هذه الدراسة فهي : الدرجة الناتجة عن تطبيق الطريقتين التحكيميتين (نيدلسكاي و أنجوف) والتي يتم على أساسها تصنيف أفراد عينة الدراسة إلى متمكنين و غير متمكنين.

### 5.6.1. طريقة نيدلسكاي : Nedelsky Method

اقترحت هذه الطريقة من طرف ليو نيدلسكاي (Leo Nedelsky) سنة 1954، لتستعمل مع الاختبارات ذات الاختيار من متعدد فقط، بحكم أنها تتطلب حكماً على الإجابة الخاطئة المحتملة. و تتمثل مهمة المحكم في تحديد الإجابات الخاطئة التي يمكن أن يتعرف عليها الفرد ذو الحد الأدنى من الكفاية على أنها خاطئة و يقوم بإلغائها .

وتقوم طريقة نيدلسكاي على فكرة بأن الفرد ذو الحد الأدنى للكفاية يبدأ أولاً في الإجابة على السؤال متعدد الاختيارات بإلغاء الإجابات التي يتعرف عليها بأنها خاطئة، و بعد ذلك يقوم بالتخمين بطريقة عشوائية مع الإجابات المتبقية. و للوصول إلى درجة الفرد ذو الحد الأدنى من الكفاية يتم تطبيق القواعد التالية :

أ- تحديد البدائل في كل مفردة من مفردات الاختبار و التي يحتمل أن يستبدها الشخص ذو الحد الأدنى من الكفاية من تفكيره عند اختيار البديل الصحيح.

ب- لإيجاد الدرجة المتوقعة للمفحوص في الاختبار ككل يتم جمع درجات المفحوص المتوقعة لكل سؤال. فمثلاً إذا ألغى الفرد ذو الحد الأدنى من الكفاية بديلاً واحداً من أربع بدائل، له حظ واحد من ثلاثة لاختيار الإجابة الصحيحة. و بالتالي تكون الدرجة المتوقعة للسؤال  $3/1$  أو 0,33.

ج- بحساب متوسط هذه الدرجات نحصل على درجة القطع للاختبار.

### 6.6.1. طريقة أنجوف: Angoff Method

هي طريقة اقترحها أنجوف عام 1971، وتعد من أشهر الطرق التي تستخدم في تحديد درجة القطع و هي مشابهة لطريقة نيدلسكاي ، حيث تحسب الدرجة المتوقعة على كل مفردة من مفردات الاختبار و الفرق هنا أن طريقة أنجوف لا تتطلب من المحكم تحديد الإجابة الخاطئة التي يتعرف عليها الفرد ذو الحد الأدنى من الكفاية بأنها خاطئة ، بل 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 النسب المئوية لكل المفردات لكي نحصل على درجة القطع من وجهة نظر المحكم . بعد ذلك يتم أخذ المتوسط لدرجات القطع لكل المحكمين لكي نحصل على درجة القطع للاختبار.

## 7.1. الدراسات السابقة:

من الإشكالات العالقة اليوم في مجال بناء الاختبارات هو عدم وجود الأدلة الكافية للدفاع عن طريقة معينة من طرق تحديد درجة القطع ، لهذا الغرض تعددت الدراسات من قبل الباحثين لاختبار مختلف المتغيرات التي تؤثر في صدق وثبات درجة القطع الناتجة. ويذكر كاين (Kane,1994) [14] أنه "لا توجد درجة قطع من ذهب و لا حتى من فضة " ، ولكن دراسة مصممة بصفة جيدة وتجري بطريقة متأنية كفيلة بأن تعطي دليلا جيدا على درجة القطع المناسبة.

ونظرا لتعدد الدراسات التي تطرقت إلى موضوع تحديد درجة القطع، و التي ركزت أغلبها على المقارنة بين مختلف الطرق ارتأينا تقسيمها إلى دراسات عربية و دراسات أجنبية.

### 1.7.1. الدراسات العربية :

#### 1.1.7.1. دراسة صلاح الدين محمود علام (1985):

تمثل موضوع الدراسة في استخدام نموذج ذي الحدين في تقدير درجة القطع لاختبار محكي المرجع وهدفت إلى بناء توزيع نظري احتمالي على أساس النموذج ذي الحدين لدرجات الاختبار المستخدمة في الدراسة لبعض مستويات التمكن بعد تصحيح هذه الدرجات من أثر التخمين. واستخدم هذا التوزيع الاحتمالي لبحث إمكانية تحديد درجة القطع للاختبار على أساس البيانات الواقعية المستخدمة من عينة الدراسة. وقد أعد الباحث اختبارا يتكون من صورتين متكافئتين (أ، ب) يشتمل كل منهما على أربعة اختبارات فرعية تقيس المهارة في صياغة الأهداف السلوكية.

تمثلت عينة الدراسة في مجموعة من طلاب كلية التربية بجامعة الأزهر ، و بلغ عدد أفرادها (154) طالبا. كون الباحث توزيعا نظريا احتماليا للنسب المئوية لأعداد الطلاب الذين يمكن اعتبارهم متمكنين وفقا لدرجات قطع و مستويات تمكن مختلفة ، مستخدما في ذلك النموذج ذي الحدين، ثم اختبر حسن مطابقة التوزيع الواقعي للنسب المئوية لأعداد طلاب العينة اعتمادا على درجاتهم الفعلية في الصورة (ب) ، وكان التوزيع النظري الاحتمالي عند مستوى تمكن ( 70 . بين المتغيرين الثنائيين.) و الذي اختاره الباحث للمهارة المطلوبة.

توصل الباحث إلى عدم وجود فروق ذات دلالة إحصائية بين كل من التوزيعين، مما أدى إلى اعتبار مستوى التمكن (70) مناسباً لمجموعة الطلاب المشاركين في البرنامج. [15].

#### 2.1.7.1. دراسة صلاح الدين علام (1991):

تمثل موضوع الدراسة في مقارنة بعض طرق تحديد مستويات الأداء في اختبار مرجعي المحك. وهدفت هذه الدراسة إلى الموازنة بين أربع طرق لتحديد مستويات أداء الاختبار الذي أعده الباحث اثنتان منها تعدان من الطرق المطلقة، وهما طريقتا نيدلسكاي و أنجوف وطريقتين من الطرق النسبية وهما طريقتا المجموعة الحدية و المجموعات المتضادة.

تكونت عينة المحكمين من (26) محكماً مقسمين على ثلاث مجموعات متباينة في مستوى تمكنها في مجال القياس و التقويم النفسي و التربوي، حيث اشتملت المجموعة الأولى على أربعة عشر (14) معلماً من معلمي مدارس الثانوية العامة الحاصلين على دبلوم خاص في التربية أو على دورة تدريبية في مجال القياس و التقويم واشتملت المجموعة الثانية على ستة (06) مدرسين مساعدين حاصلين على درجة الدكتوراه في علم النفس التعليمي و لديهم خبرة لا تقل عن ثلاثة أعوام في تدريس مادة الاختبارات و المقاييس. كما اشتملت المجموعة الثالثة على ستة (06) من أساتذة علم النفس التعليمي. أما بالنسبة لعينة الطلاب الذين اجري تحكيم أدائهم في الاختبار فكانت (372) طالباً من طلاب كلية التربية ممن يدرسون مقرراً سيكولوجياً في الفروق الفردية، و يقوم بتدريسهم عينة المحكمين المكونة من الأساتذة و المدرسين المساعدين التي اختيرت في هذه الدراسة .

أدى تطبيق الطرق الأربعة المختلفة المستخدمة في الدراسة إلى الحصول على درجات قطع مختلفة نسبياً فقد كانت متوسطاتها (40، 44، 43، 41) لكل من طريقة ندلسكاي و المجموعة الحدية و أنجوف و المجموعات المتناقضة على التوالي . [16].

#### 3.1.7.1. دراسة عصام الدسوقي (1996):

تمثل هدف الدراسة في البحث عن مدى فاعلية طريقة أنجوف في تحديد درجة القطع. و ذلك من خلال مقارنة نتائجها مع نتائج طريقة هوفستي. و استخدم الباحث اختباراً من إعدادة في وحدة معادلات الدرجة الثانية من مقرر الجبر للصف الأول الثانوي يتضمن كل منهما (20) مفردة من نوع الاختيار من متعدد ذات الأربع استجابات . و قد استعان الباحث بستة محكمين

من موجهي الرياضيات بالإضافة إلى أربعة مدرسين أوائل ، فضلا عن ذلك تكونت عينة الدراسة من (65) طالبا . و قد استخدم الباحث رسم المنحنى في طريقة هوفستي.

توصلت الدراسة إلى فاعلية طريقة أنجوف في تحديد درجة القطع ، حيث لم تختلف نتائجها عن نتائج طريقة هوفستي، فكانت درجة القطع بطريقة أنجوف 74 بالمائة بينما بلغت درجة الفصل باستخدام طريقة هوفستي 86 بالمائة أي أن الاختلاف بينهما كان في مدى قليل . [17]

#### 4.1.7.1. دراسة أحمد الشريم ويوسف سوالمه،(2006):

هدفت هذه الدراسة إلى المقارنة بين نموذجي أنجوف ونيديلسكاي لتقدير درجة القطع لاختبار محكي المرجع في الرياضيات. و ذلك من خلال وجود مؤشرات عن صعوبة الفقرات أو عدم وجودها. وقد تكون الاختبار من ثلاثين فقرة من نوع الاختيار من متعدد لكل منها أربعة بدائل. وتكونت عينة الدراسة من ثمانين محكما و محكمة. و قد تم تقسيم المحكمين بطريقة المزاجية العشوائية إلى أربع مجموعات متساوية، حددت كل منها درجة قطع للاختبار مرتين وفق الأسلوب المعين لها. وأشارت نتائج الدراسة أن درجة القطع للاختبار تراوحت بين 0.62 و 0.68 باستخدام نموذج أنجوف وبين 0.49 و 0.57 باستخدام نموذج نيديلسكاي. وأنه لا يوجد فرق دال إحصائيا عند مستوى الدلالة 0.05 بين معاملي ثبات النموذجين. كما بينت النتائج أنه لا يوجد فرق دال إحصائيا بين معاملي الارتباط لصعوبات الفقرات وتقديرات المحكمين لها في النموذجين عند عدم تزويد هم بصعوبة فقرات. بينما يوجد فرق دال إحصائيا بين معاملي الارتباط للنموذجين عند اطلاع المحكمين على صعوبة الفقرات. [18].

#### 5.1.7.1. دراسة غانم حجاج (2007):

يتمثل موضوع الدراسة في قرار التصنيف الناتج عن طريقتي أنجوف و نيديلسكاي في تحديد درجة القطع لاختبار محكي المرجع . وقد هدفت هذه الدراسة إلى المقارنة بين درجة القطع المحددة بطريقة نيديلسكاي و درجة القطع المحددة بطريقة أنجوف. كما هدفت أيضا إلى مقارنة مؤشرات تصنيف الطلاب بين الطريقتين. وقد استخدمت الدراسة نوعين من العينات. تمثل النوع الأول في عينة المحكمين و البالغ عددهم (12) قاموا باستخدام طريقتي أنجوف و نيديلسكاي لتحديد درجة القطع على اختبار محكي المرجع في التدريبات العملية لعلم النفس. أما النوع الثاني فتمثل في عينة المفحوصين، وقد تكونت من مجموعتين، المجموعة الأولى و هي المجموعة المتعلمة



و تكونت من (93) طالبا من طلاب الفرقة الثالثة شعبة تربية عامة، أما المجموعة الثانية وهي المجموعة غير المتعلمة فتكونت من (85) طالبا من طلاب الفرقة الثالثة شعبة التربية الرياضية .  
و توصلت النتائج إلى ارتفاع درجة القطع المحددة بطريقة أنجوف عن درجة القطع المحددة بطريقة نيدلسكاي، وهذه النتيجة أدت إلى ارتفاع المؤشرات الايجابية لقرار التصنيف الخاصة بطريقة نيدلسكاي مقارنة بمؤشرات قرار التصنيف الناتج عن طريقة أنجوف. [5].

#### 6.1.7.1. دراسة زيد بن عطا و يوسف سوالمة ( 2007 ):

تمثلت الدراسة في استخدام نموذج المحاولات ذات الحدين في فحص تقديرات المحكمين لدرجة القطع لاختبار مرجعي المحك في الرياضيات لمعرفة مدى اختلاف تقديرات المحكمين لدرجة القطع باختلاف جولة التحكيم. و كذا جودة تقديرات المحكمين لدرجة القطع قبل فحصها باستخدام نموذج المحاولات ذات الحدين وبعده. تكونت عينة الدراسة من ( 110 ) طالبا وطالبة. وقد تم بناء اختبار تحصيلي محكي المرجع في مبحث الرياضيات لطلبة الصف العاشر الأساسي تقيس المهارات و المفاهيم و المصطلحات و التعليمات المتعلقة بوحدة الهندسة التحليلية . و قد احتوى الاختبار على 30 فقرة لكل منها أربعة بدائل، وتمثلت عينة المحكمين في ( 15 ) معلم و معلمة تم اختيارهم بطريقة عشوائية من بين معلمي الرياضيات للصف العاشر الذين لديهم مؤهل تربوي بالإضافة لبيكالوريوس في الرياضيات. و تم تحديد درجات القطع باستخدام طريقة أنجوف عبر ثلاث جولات. تم في الجولة الأولى استخدام نموذج أنجوف لتحديد درجة القطع و تم في الجولة الثانية توزيع نموذج خاص لمستوى الأداء وفق أسلوب أنجوف يشتمل على التعليمات و الفقرات و الوسط الحسابي لمعامل الصعوبة المقدر من طرف المحكمين لكل فقرة من فقرات الاختبار الذي تم إيجاده في الجولة الأولى. و تم تحديد مستوى الأداء بنفس أسلوب الجولة الأولى. و في الجولة الثانية تم تزويد المحكمين بمعاملات الصعوبة لفقرات الاختبار وتم تحديد مستوى الأداء و درجة القطع بالأسلوب السابق نفسه .

و توصلت الدراسة إلى وجود فروق ذات دلالة في درجات القطع تعزى إلى جولة التحكيم. و بعد استبعاد تقديرات المحكمين الذين لم تتطابق تقديراتهم مع نموذج المحاولات ذات الحدين قلت الفروق بين درجات القطع جميعها، و ارتفعت قيمة معامل دقة القرار لتصنيف الطلبة. [19] .

### 7.1.7.1. دراسة أحمد عودة و أحمد الشريم (2010):

تمثل موضوع الدراسة في تطوير أسلوب قائم على الأهداف السلوكية لتحديد درجة القطع ومقارنته مع أسلوب أنجوف كما هدفت الدراسة إلى معرفة مدى اختلاف ثبات تقديرات المحكمين لدرجة القطع باستخدام النموذج القائم على الأهداف عن معامل الثبات لنموذج أنجوف. كما هدفت أيضا إلى معرفة قيمة معامل ثبات التصنيف للمفحوصين إلى متمكنين و غير متمكنين باستخدام كل من النموذج القائم على الأهداف، و درجة القطع الناتجة عن نموذج أنجوف. أما عينة الدراسة فتكونت من ثلاثين محكما تم تقسيمهم إلى ثلاث مجموعات على النحو التالي :

أ- مجموعة التأليف: مكونة من عشرة محكمين ممن شاركوا في تأليف منهاج الرياضيات في وزارة التربية و التعليم.

ب- مجموعة الخبراء: مكونة من عشرة محكمين ممن يحملون مؤهل دكتوراه أو ماجستير في أساليب تدريس الرياضيات و لديهم خبرة في تدريس الصف التاسع الأساسي .

ج- مجموعة المعلمين: وهم المعلمون الذين يحملون درجة البكالوريوس في الرياضيات و يدرسون الصف التاسع لمدة لا تقل عن سنتين. أما عينة الطلبة فقد تكونت من (171) طالبا .

وتمثلت أدوات الدراسة فيما يلي:

أ- تكونت الأداة الأولى من قائمة الأهداف بحيث تكتب الهدف بشكل مميز و يتبعه مباشرة مجموعة من الفقرات المحتملة التي تقيس هذا الهدف بالتحديد .

ب- الاختبار ، وتطلبت إجراءات الدراسة وجود صورتين لاختبار محكي المرجع . تتكون الصورة الأولى من (30) فقرة من نوع الاختبار من متعدد و لكل فقرة أربعة بدائل .أما الصورة الثانية فتتكون كذلك من (30) فقرة من نوع الاختبار من متعدد.

و توصلت الدراسة إلى النتائج التالية :

بلغت المتوسطات الحسابية لتقديرات المحكمين لدرجة القطع لقائمة الأهداف في الجولة الأولى 56 % و في الجولة الثانية 56 % و بالمقابل بلغ المتوسط الحسابي لتقديرات المحكمين لدرجة القطع حسب إجراءات نموذج أنجوف 67 % في الجولة الأولى و 66 % في الجولة الثانية. و للصورة الثانية 63 % في الجولتين. و قد كانت جميع معاملات الارتباط بين تقديرات المحكمين في جولتي التحكيم مرتفعة و دالة إحصائيا. وبينت النتائج أنه لا يوجد فرق دال إحصائيا بين معاملي الثبات للنموذجين. [20].

### 2.7.1. الدراسات الأجنبية :

#### 1.2.7.1. دراسة لوك وود و برينان ( Brennan & Lockwood 1980 ) :

تناولت هذه الدراسة المقارنة بين درجات القطع الناتجة عن طريقتي أنجوف و نيدلسكاي. حيث استخدمت نظرية القابلية للتعميم لمعرفة خطأ التباين في كل طريقة، مقارنة البيانات الناتجة عن تطبيق الطريقتين بالإضافة إلى التحقق من مدى تأثير اختلاف المحكمين على ثبات الاختبار. تكون الاختبار من متعدد من (126) مفردة في مجال الصحة العامة. وتكونت العينة من (5) محكمين. حيث قام نفس المحكمين بتطبيق الطريقتين لتحديد درجة القطع. توصلت الدراسة إلى وجود اختلاف في درجة القطع الناتجة عن كل من طريقتي نيدلسكاي و أنجوف. فكانت (0.84) باستخدام طريقة أنجوف بانحراف معياري قدره (4.7) و (0.70) باستخدام طريقة نيدلسكاي بانحراف معياري قدره (3.9).

كما توصلت هذه الدراسة إلى أن الخطأ المعياري في تقدير درجة القطع كان أقل في طريقة أنجوف مقارنة بطريقة نيدلسكاي. ويعزى هذا التباين إلى الاختلاف في منهجية تقدير المفردات (الاحتمالات) الخاصة بكل طريقة أو إلى كيفية تصور الفرد ذو الحد الأدنى من الكفاية. بالإضافة إلى كون إجراءات نيدلسكاي تحصر المحكمين في عدد قليل من الاحتمالات المنفصلة و غير المتساوية.

كما توصلت الدراسة إلى نتيجة ، قد لا ترتبط مباشرة بالتباين في متوسط درجة القطع. حيث بينت المعطيات بأنه رغم اتفاق المحكمين في طريقة نيدلسكاي على عدد البدائل التي يمكن أن يقوم الفرد ذو الحد الأدنى من الكفاية بالغائها، إلا أن نفس المحكمين يختلفون في تحديد ماهية البدائل التي يمكن أن تلغى. كما توصلت الدراسة إلى أن المحكمين في طريقة نيدلسكاي قاموا بإلغاء عدد كبير من البدائل الصحيحة عند استخدام إجراءات نيدلسكاي. [21].

#### 2.2.7.1. دراسة سكاكين و كلينج ( Skakun, N & Kling,S, 1980 ) :

كانت الدراسة عبارة عن مقارنة لتحديد مستويات الأداء. وهدفت إلى معرفة اختلاف المستويات الناتجة باختلاف الطريقة المستخدمة، و معرفة مدى ثبات متوسط تقديرات المحكمين لدرجات القطع لكل طريقة. و تمثلت أداة الدراسة في اختبار الجراحة العامة يتكون من (149)

مفردة. استخدم الباحثان طريقتي ايبيل و نيدلسكاي. وشملت عينة الدراسة (168) طالبا من كلية الطب تخصص جراحة عامة. و قسم الباحثان العينة إلى قسمين: الأول يحتوي على (48) طالبا يشكلون المجموعة المرجعية و القسم الثاني يحوي (120) طالبا يمثلون المجموعة غير المرجعية. و بلغ عدد عينة المحكمين ثمانية . شارك سبعة منهم في تحديد درجة القطع وفق طريقة نيدلسكاي وشارك ستة منهم في تحديد درجة القطع وفق طريقتي ايبيل المعدلتين، و شارك خمسة منهم في تحديد درجة القطع وفق الطرق الثلاثة مجتمعة.

و توصلت الدراسة إلى أن استخدام الطرق المختلفة لتحديد درجة القطع يؤدي إلى درجات قطع مختلفة حيث كانت الدرجات كما يلي :

طريقة نيدلسكاي 66.7 بالمائة و طريقة ايبيل (1) 69.7 بالمائة و طريقة ايبيل (2) 71.7 بالمائة. أرجعت الدراسة هذه الاختلافات لعدة أسباب منها :

- أن طريقة نيدلسكاي تحصر المحكمين في عدد محدود من الاحتمالات في الاختبار متعدد الاختيارات.

- اختلاف مفهوم الطالب المتقن ، بالإضافة إلى أن اختيار الإجابات الخاطئة أسهل من اختيار الإجابات الصحيحة .

كما نتج عن هذه الطرق درجة ثبات عالية بالنسبة لمتوسط تقديرات المفردات ومتوسط درجات القطع. [22].

### 3.2.7.1. دراسة روك و آخرون (Rock,D.A et al,1980):

تمثل موضوع الدراسة في مقارنة تجريبية بين المقاربات التحكيمية لإجراءات تحديد درجة القطع. وهدفت هذه الدراسة إلى معرفة إذا ما كانت إحدى الطريقتين (أنجوف و نيدلسكاي) تعطي درجات قطع أعلى من الأخرى. كما هدفت إلى معرفة ما إذا كانت إحدى المجموعات تعطي درجات قطع أعلى من المجموعات الأخرى. وهل تختلف درجة القطع الخاصة بالمفحوص ذو الحد الأدنى من الكفاية و درجة القطع الخاصة بالمفحوص ذو الكفاية المتوسطة. إلى جانب الطريقة التي تعطي أكثر اتفاق داخلي للمحكمين.

تكونت عينة الدراسة من ستة عشر محكما موزعين على أربع مجموعات. تكونت المجموعة الأولى من أربعة أفراد يمثلون الأقلية العرقية. أما المجموعة الثانية فتكونت من ثلاثة وسطاء تجاريين. كما تكونت المجموعة الثالثة من أربعة قانونيين و تكونت المجموعة الرابعة من خمسة مستشارين.

تمثلت أداة الدراسة في اختبار الإجازة في العقار. و تكون من 64 فقرة. وتقوم مجموعات المحكمين بتقدير درجات القطع على فقرات الاختبار باستخدام طريقتي نيدلسكاي و أنجوف مع تطبيق التعليمية الأولى و التي تتمثل في تحديد درجة القطع وفق مفهوم المفحوص ذو الحد الأدنى من الكفاية. ثم استخدام التعليمية الثانية و التي تتعلق بتحديد درجة القطع وفق مفهوم المفحوص ذو الكفاية المتوسطة.

توصلت الدراسة إلى أن ثلاث مجموعات من الأربعة أعطت درجات قطع أعلى في طريقة أنجوف مقارنة بطريقة نيدلسكاي. كما أعطت كلتا الطريقتين درجات قطع مميزة بين المفحوص ذو الحد الأدنى من الكفاية و المفحوص ذو الكفاية المتوسطة. غير أن طريقة أنجوف كانت أكثر تمييزا من طريقة نيدلسكاي. حيث بلغ متوسط درجات القطع 39.18 بالمائة عند استخدام التعليمية الأولى مع تطبيق طريقة أنجوف و 31 بالمائة باستخدام طريقة نيدلسكاي ، كما بلغ متوسط درجات القطع باستخدام التعليمية الثانية 47.10 بالمائة عند تطبيق طريقة أنجوف و 34.56 بالمائة عند تطبيق طريقة نيدلسكاي. أما فيما يخص الاتفاق الداخلي فقد أعطت طريقة أنجوف أكثر اتفاق داخلي من طريقة نيدلسكاي. [23].

#### 4.2.7.1. دراسة بيهونياك (Behuniak et al ,1982) :

هدفت هذه الدراسة إلى المقارنة بين طريقتي نيدلسكاي و أنجوف لمعرفة مدى اختلاف درجة القطع الناتجة باختلاف مجموعة المحكمين بالإضافة إلى معرفة مدى تأثير درجة القطع بالخصائص الديمغرافية المختلفة للمحكمين.

و تمثلت أداة الدراسة في اختبارين محكيا المرجع في مادتي القراءة و الرياضيات. يحتوي اختبار مادة الرياضيات على ثلاثة نماذج، يحتوي كل نموذج على (90) سؤالا . أما اختبار القراءة فيحتوي كذلك على ثلاث نماذج، و يحتوي كل نموذج على 80 سؤالا من الأسئلة ذات الاختيار من متعدد . و اشتملت عينة الدراسة على (460) طالبا بالإضافة إلى (14) عضوا من أعضاء هيئة

التدريس المتخصصين في مادة القراءة و (13) عضوا من أعضاء هيئة التدريس المتخصصين في مادة الرياضيات، وتوصلت الدراسة إلى نتائج عديدة منها :

- عدم وجود علاقة بين خصائص المحكمين الديمغرافية و بين درجات القطع الناتجة .
- كما توصلت الدراسة إلى تشابه في تقديرات المحكمين في المجموعة الواحدة وعند استخدام طريقة موحدة. ولكن عند تقسيم مجموعة المحكمين إلى عينتين أدى ذلك إلى درجات قطع مختلفة بصفة دالة إحصائيا . [24].

#### 5.2.7.1. دراسة ليفنجستون و زاكي (Livingston & Zieky, 1983) :

هدفت هذه الدراسة إلى المقارنة بين طريقتي نيدلسكاي و أنجوف، وقد صمم لهذا الغرض اختباران. أحدهما في الرياضيات و تكون من (65) مفردة . والآخر في القراءة وتكون من (70) مفردة من نوع الاختيار من متعدد (رباعية البدائل). أما عينة المحكمين في الاختبارين فتكونت من عدد يتراوح بين (12 - 20) محكما في المجال الذي يقيسه الاختبار، تم اختيارهم من أربعة مدارس مختلفة.

توصلت الدراسة إلى نتيجتين متناقضتين. ففيما توصلت الدراسة إلى أن قيمة درجة القطع المحددة بطريقة أنجوف و التي بلغت 36 بالمائة تعتبر اعلى من قيمة درجة القطع المحددة بطريقة نيدلسكاي و التي بلغت 31 بالمائة بالنسبة لاختبار الرياضيات، نجد أنه في اختبار القراءة و بصورة مجملة توصلت إلى العكس ، حيث بلغت قيمة درجة القطع عند استخدام طريقة أنجوف 34 بالمائة. بينما بلغت 42 بالمائة في طريقة ندلسكاي. [25].

#### 6.2.7.1. دراسة ديمورو و باورز ( Demauro & Powers, 1990 ):

هدفت هذه الدراسة إلى معرفة الاتساق المنطقي لطريقة أنجوف، و تكونت عينة المحكمين من (19) محكما يمثلون أعضاء المؤسسة الدولية لعلم النفس المدرسي في ذلك الوقت. أما أداة الدراسة فتمثلت في اختبار محكي المرجع في علم النفس المدرسي و تكونت من (130) مفردة. أما عينة الدراسة فبلغ عددها (2373) طالبا.

توصلت الدراسة إلى عدة نتائج منها أن طريقة أنجوف تحظى بدرجة عالية نسبيا من الثبات المنطقي حيث يوجد اتساق بين تقديرات المحكمين و الأداء الفعلي للطلاب بلغ (0,71) ، كما يوجد ارتباط بين تقديرات المحكمين و الصعوبات الحقيقية للمفردات تتراوح بين (0,25) و (0,56) بوسيط

بوسيط قيمته (0,42). كما أوصى الباحث باستخدام طريقة أنجوف لتحديد درجة القطع ، و لكن أشار إلى أن نتائجه مرتبطة بالاختبار المستخدم و العينة المستخدمة. [26].

#### 7.2.7.1. دراسة شانج لاي (Chang Lei 1996) :

تمثلت الدراسة في المقارنة بين طريقتي أنجوف و نيدلسكاي ، حيث تركزت المقارنة على الاتساق الداخلي بين المحكمين في كل طريقة و قيمة درجة القطع الناتجة. و تكونت عينة المحكمين من 22 محكما و تكون الاختبار من تسعة 09 مفردات من نوع الاختيار من متعدد. تكونت عينة المفحوصين من (274) طالبا بالدراسات العليا.

توصلت الدراسة إلى عدة نتائج منها: اختلاف درجة القطع باختلاف الطريقة المطبقة. فكانت 71 بالمائة باستخدام طريقة أنجوف و 57 بالمائة باستخدام طريقة نيدلسكاي. و توصلت الدراسة كذلك إلى أن طريقة نيدلسكاي تتمتع باتساق داخلي أكبر مقارنة بطريقة أنجوف . [27].

#### 8.2.7.1. دراسة دونو و أماتو (Donnoe & Amato, 1997) :

هدفت هذه الدراسة إلى المقارنة بين ثلاث طرق لتحديد درجة القطع وهي: أنجوف، نيدلسكاي و اييل. وتوصلت إلى اختلاف درجة القطع باختلاف الطريقة المطبقة ، فكانت 50 بالمائة باستخدام طريقة أنجوف و 33 بالمائة باستخدام طريقة نيدلسكاي و 60 بالمائة باستخدام طريقة اييل. كما توصلت الدراسة إلى وجود فروق ذات دلالة إحصائية بين الطرق الثلاثة، حيث كانت قيمة (ف) تساوي 0,27 عند درجة حرية (2,25)، كما لا توجد ارتباطات بين طريقة نيدلسكاي و طريقتي أنجوف و اييل. بينما وجد ارتباط بين طريقة أنجوف وطريقة اييل. [28]

#### 9.2.7.1. دراسة ناستروم و نايستروم (Nasstrom & Nystrom,2008) :

هدفت الدراسة إلى مقارنة درجات القطع الناتجة عن تطبيق طريقتي أنجوف المعدلة و المجموعة الحدية و التحقق من أدلة الصدق الداخلية الخارجية و الإجرائية.

تمثلت أداة الدراسة في الاختبار الوطني السويدي للرياضيات، الذي يشتمل على 22 فقرة . كما بلغ عدد المفحوصين 948 طالبا مقسمين على 46 فوج. و تكونت عينة المحكمين لطريقة المجموعة الحدية من 44 أستاذا من أساتذة الرياضيات. ستة و ثلاثون أستاذا لهم خبرة تفوق ستة

سنوات .أستاذان لهم خبرة بين ثلاثة و خمسة سنوات و ستة أساتذة تصل خبرتهم إلى عامين .  
و قد اشتملت عينة المحكمين لطريقة المجموعة الحدية على الجنسين (ذكور - إناث) بالتساوي.

أما عينة المحكمين بطريقة أنجوف المعدلة فاشتملت على إحدى عشر محكما يمثلون أساتذة الرياضيات (4 إناث و 7 ذكور) و لديهم خبرة تفوق خمسة سنوات.ويقوم مجموع المحكمين في كل طريقة بتحديد مستويين للأداء على فقرات الاختبار. وهما ناجح (P) و ناجح بتميز (PD).

توصلت نتائج الدراسة إلى اختلاف في درجات القطع، و كان الاختلاف أكثر في مستوى الأداء (PD). فقد بلغت درجة القطع في طريقة أنجوف لتحديد مستوى الأداء (P) 10 و 22 بالنسبة لمستوى (PD). كما بلغت درجة القطع في طريقة المجموعة الحدية 9 و 19 بالنسبة للمستويين (P) و (PD) بالترتيب. [29].

#### 10.2.7.1. دراسة بوكندال، فردوس وجيراو: (Buckendahl,Fedous & Gerrow,2010)

هدفت هذه الدراسة إلى استخدام طريقتي أنجوف المعدلة لتحديد درجة القطع على اختبارين في طب الأسنان لدراسة إمكانية تطبيق درجة القطع الناتجة عن الاختبار الجزئي (Subset) على الاختبار الكلي. بالإضافة إلى استخدام طريقة كتاب العلامات (Bookmark) كمحك خارجي.

تمثلت أداة الدراسة في اختبارين لنيل الشهادة في طب الأسنان. تكون الاختبار الأول من 300 فقرة. أما الاختبار الثاني فتكون من 150 فقرة تم اختيارها بصفة عشوائية من الاختبار الأول . كما تكونت عينة المحكمين من 35 محكما ينتمون إلى سلك طب الأسنان. و قسمت هذه العينة إلى مجموعتين :

المجموعة (أ) و التي اشتملت على 18 محكما، تقوم بتقدير درجات القطع على الاختبار ككل (300 فقرة) . أما المجموعة (ب) و التي اشتملت على 17 محكما فتقوم بتقدير درجة القطع على الاختبار الجزئي (150 فقرة) . كما تقوم كل مجموعة بتحديد درجة القطع على الاختبار ككل باستخدام طريقة علامة الكتاب (Bookmark) .



و قد توصلت الدراسة إلى اختلاف في درجات القطع الناتجة، حيث بلغت تقديرات المجموعة (أ) لدرجة القطع باستخدام طريقة أنجوف 214,1 و 204,4 بالنسبة للمجموعة (ب). كما بلغت درجة القطع بالنسبة للمجموعتين باستخدام طريقة علامة الكتاب 216,3 و 218,4 على التوالي. [30].

### 8.1. تعقيب على الدراسات السابقة :

مع تزايد الإهتمام بالقياس محكي المرجع ، زاد الإهتمام أكثر بطرق تحديد درجة القطع . فقد أشار بارك ( Berk,1996 ) بأن طريقة أو طريقتين تقترح كل عام أو عامين، فأصبح بالتالي اختيار الطريقة المناسبة يتطلب مقارنتها مع بعضها البعض في ضوء بعض المعايير، و دراسة إمكانية تطويرها لنتاسب مع أهمية القرارات التربوية المتعلقة بقضية التصنيف و استخداماتها الواسعة في مختلف مجالات الحياة. [31].

و قد تمحورت أغلب المواضيع المعالجة من قبل الدراسات السابقة حول المقارنة بين مختلف الطرق و ذلك لغرض التوصل إلى الطريقة التي تتميز بفعالية أكبر و ذلك من خلال اختبار و تفحص مختلف المتغيرات التي تؤثر في درجة القطع النهائية. ونجد من بين هذه المتغيرات اختلاف كفايات المحكمين و اختلاف خصائصهم الديمغرافية إلى جانب عدد جولات التحكيم . كما حاولت بعض الدراسات البحث في مدى ثبات تقديرات المحكمين في مختلف الطرق . و ذهبت دراسات أخرى إلى مقارنة مؤشرات التصنيف .

و قد خلصت أغلب الدراسات إلى أن نموذج نيدلسكاي يعطي درجة قطع أقل من درجة القطع الناتجة عن نموذج أنجوف مما يعني نسبة النجاح عند استخدام نموذج نيدلسكاي تكون أكبر. فقد أثبتت دراسات كل من ( Chang,1996.Donnoe & Amato,1997, Behuniak,1982 ) مثلا أن تقديرات المحكمين لدرجة القطع وفق طريقة أنجوف حققت أعلى نسبة درجة قطع من الطرق الأخرى. في حين توصلت دراسة (Rock et al ،1980) إلى نتيجة متناقضة حيث بينت ثلاثة أفواج ارتفاع درجة القطع المحددة بطريقة أنجوف عن تلك المحددة بطريقة نيدلسكاي. غير أن المجموعة الرابعة وهي مجموعة القانونيين توصلت إلى عكس نتيجة المجموعات الثلاثة.

كما توصلت دراسة ليفنجستون و زاكي ( Livingston & Zieky,1983 ) كذلك إلى نتائج متعارضة.حيث كانت درجة القطع باستخدام طريقة أنجوف أعلى من درجة القطع الناتجة عن طريقة نيدلسكاي في اختبار الرياضيات. بينما كان العكس بالنسبة لاختبار القراءة، حيث كانت

درجة القطع الناتجة عن نموذج نيدلسكاي أعلى. غير أن هذه الدراسة تبقى محدودة (inclusive) ذلك أن كل نتائج الطرق المستخدمة كانت متباينة بصفة كبيرة من مدرسة إلى أخرى، رغم أن الاختبار المطبق على كل مدرسة كان موحدًا. كما أن اتساق درجات القطع، بغض النظر عن الطريقة المطبقة، كان ضعيفًا، لذا أوصى ليفنجستون و زاكي بفحص التباين في درجات القطع المحددة بالطرق المختلفة بالإضافة إلى الاتساق بين المحكمين عوض مقارنة درجات القطع الناتجة. و من خلال النتائج التي المتوصل لاسيما دراسة بيهونيياك ( Behuniak,1982 )، إذ يؤكد هذا الأخير أن استخدام مجموعتين من المحكمين لتقدير درجة القطع باستخدام طريقة موحدة يؤدي إلى نتائج مختلفة. و يدعم هذه النتيجة الاقتراح الذي جاء به ليفنجستون (Livingston) أن هذه الدراسات يجب أن تستخدم نفس المحكمين للمقارنة بين طريقتين مختلفتين من طرق تحديد درجة القطع وإلا فلن نتمكن من فصل تأثيرات المحكمين عن تأثيرات اختلاف الطرق نفسها.[32] (ص:6).

### 9.1.علاقة الدراسة الحالية بالدراسات السابقة:

يتضح من الدراسات السابقة التي تم عرضها الاستخدام الشائع لطريقتي نيدلسكاي و أنجوف لتحديد درجة القطع في مختلف الميادين (الصحة، التربية، التعليم العالي) مما يشير إلى أهمية هذه الطرق و فعاليتها. فقد أشار ديمورو و باورز ( Demauro & Powers, 1990 ) إلى أن نموذج أنجوف يعد ملائماً لتقدير درجة القطع للاختبار كونه يتمتع بمؤشرات ثبات و صدق مقبولة. هذا ويشير جيجر ( Jaeger,1989 ) أن الكثير من الدراسات ركزت على الطرق التحكيمية دون غيرها وذلك نظراً لأهميتها و سهولة استخدامها و سهولة تحليل نتائجها. [33].

و بالإضافة إلى ما سبق فإن معظم أدوات الدراسات السابقة هي اختبارات ذات الاختيار من متعدد رباعي البدائل، وهذا ما اتبعه الباحث في إعداد للأداة المستخدمة في البحث. مع أخذنا بعين الاعتبار لما أوصت به بعض الدراسات حول استخدام مجموعة واحدة من المحكمين لتقدير أداء المفحوص ذو الحد الأدنى من الكفاية على فقرات الاختبار في الطريقتين (نيدلسكاي و أنجوف)، أي أن استخدام تصميم القياسات المنكررة لغرض الدراسة الحالية يساعد على تفادي تأثير خصائص المحكمين على درجة القطع النهائية، بالإضافة إلى زيادة صدق نتائج المقارنة. كما لا نكتف في هذه الدراسة بالبحث في مدى اختلاف درجات القطع الناتجة عن تطبيق الطريقتين، بل سنبحث كذلك في مدى توفر كل طريقة على أدلة كافية للصدق مع التركيز على الصدق الداخلي.

## الفصل 2

### بناء الاختبارات التحصيلية محكية المرجع

تمهيد :

إن الهدف الأساسي من عملية التعليم هو مساعدة المتعلم على تحقيق الأهداف التعليمية المسطرة التي يجب أن تتضمن التغييرات المنظرة في سلوكياته. وعندما ينظر إلى عملية التعليم من هذه الناحية، فإن التقويم بصفة عامة و الاختبارات التحصيلية بصفة خاصة تصبح جزء مهما في العملية التعليمية. كما تظهر أهمية الاختبارات التحصيلية المصممة بصفة جيدة في حصول المربي على المعطيات التي تساعد في التعرف على مستوى التحصيل الدراسي الذي وصل إليه المتعلم وتزوده بالتغذية الراجعة حول مقدار ما يحدث لدى البعض من تحسن، و تشخيص صعوبات التعلم لدى بعض الأخر، كما تساعد تلك المعطيات على تسطير الأهداف التعليمية الموالية .

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 إصدار العديد من القرارات الفنية ( مثل : عملية التوجيه ، و انتقاء ذوي الحاجات الخاصة، إعداد و تقويم البرامج الدراسية). ولقد أشار جرونلند ( Gronlund ) أن المعطيات التي نحصل عليها من الاختبارات المصممة بصفة جيدة ومن مختلف أدوات القياس تساعدنا في الحكم على مدى قابلية الأهداف التعليمية للتحقق ومدى فائدة الوسائل التعليمية و فعاليتها. [34] (ص:34).

#### 1.2. مفهوم الاختبار التحصيلي :

لقد أوجدت الكثير من التعريفات للاختبار التحصيلي، فقد عرفه جرونلند ( Gronlund ) بأنه " إجراء منظم لتحديد الكم الذي تعلمه التلميذ " . [35]. 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." [36] (ص: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". [37] (ص:387).

يتبين من خلال عرضنا لمختلف التعريفات بأن 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 مثلا 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.

## 2.2. تصنيفات 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. [38].

### 1.2.2. الاختبارات معيارية المرجع : ( Norm- referenced test )

يرجع تاريخ القياس معياري المرجع إلى الأيام الأولى لظهور اختبار الذكاء، أي عندما صمم بينيه ( Binet ) أدواته الأولى لقياس ما يسمى ب "العمر العقلي". [39]. وتم الانتقال من قياس الذكاء إلى القياس التربوي في بداية أربعينيات القرن العشرين عندما قامت مصلحة الاختبارات التربوية ببناء اختبار الاستعدادات الدراسية Scholastic Aptitude Test ( 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 السيكومترية". 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 يوفام و هيو سك (Popham & Husek، 1969) الاختبارات معيارية المرجع بأنها " الاختبارات التي تستخدم لتحديد أداء الفرد بالنسبة لأداء أقرانه ". [40] و أجمعت 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. [41] (ص: 22).

و لما كان الغرض من الاختبارات معيارية المرجع هو المقارنة بين الأفراد، فقد أصبح يعتمد عموماً 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 المتينة التي يركز عليها ، حيث أكد همبلتون أن أغلب الأساليب الإحصائية في كل من النظرية الكلاسيكية و الحديثة في القياس بنيت على خلفية الاختبارات معيارية المرجع [42]، إلا أنه تعرض إلى كثير من الانتقادات التي أفقدته شيئاً من قوته. فقد أشار كل من أنور الشرفاوي و نادية 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 كيفية براعة هذا الطالب أو تفوقه في أعمال الموضوع المتضمنة في الاختبار [43] (ص: 90). حيث أن تحديد موقع الفرد بالنسبة لأقرانه ليس دليلاً على تفوقه وليس من الضروري أن يكون متوسط المجموعة المعيارية معياراً كافياً يعكس كفاية 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 تقييد في تطوير هذه البرامج و زيادة فعاليتها. [44] (ص: 329). وقد دفعت الانتقادات المتواصلة الكثير من المختصين في القياس التربوي إلى البحث عن منهجية بديلة تكشف عما حققه المتعلم فعلاً و ما لم يحققه من أهداف تربوية، وهذا ما عجل بظهور ما يسمى بالقياس محكي المرجع. وقد أشار في هذا الصدد همبلتون أن "الاختبارات محكية المرجع صممت لغرض تلبية الاحتياجات الخاصة بالنماذج التعليمية الحديثة". [45] (ص: 15).

## 2.2.2. الاختبارات محكية المرجع: Criterion- Referenced Test

يعود الفضل في ظهور مصطلح القياس محكي المرجع إلى جليزر (Glaser) 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 . [38].

و قد تعددت التعريفات الخاصة بالقياس محكي المرجع و تنوعت ولكن اتفقت في جوهرها. فقد عرف جليزر و نيتكو (Glaser & Nitko,1971) الاختبار محكي المرجع بأنه: "اختبار صمم عن قصد إعطاء قياسات تفسر مباشرة بالنسبة إلى مستويات أداء محددة." [46] (ص: 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". [11] 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 أنور الشرقاوي و نادية محمد عبد السلام بأن "المحك الذي سيقارن بناء عليه أداء الطالب هو السلوك الذي يحدد كل نقطة على امتداد متصل التحصيل". [46] (ص: 160).

ولقد تعددت الأسماء الخاصة بالاختبارات محكية المرجع. و نجد من بينها "اختبارات مرجعية المجال" (Domain-Referenced Tests)، "اختبارات مرجعية الأهداف" (Objectives- Referenced Tests) "اختبارات الإتقان" (Mastery Tests). ويرتبط اختيار الإسم المناسب بالسياق الذي يستخدم فيه الاختبار. [47] (ص: 378).

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. [45] (ص: 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 كلا النوعين.

### 3.2. الفرق بين الاختبارات محكية المرجع و الاختبارات معيارية المرجع :

إذا كانت كل من الاختبارات محكية المرجع و الاختبارات معيارية المرجع تخدم أغراضا مختلفة فمن المنطقي أن مقاربات بناء و تفسير و تقييم الاختبارات تختلف اختلافا كبيرا. و في هذا الصدد أشار المختصون في القياس النفسي و التربوي إلى وجود مجموعة من نقاط الاختلاف بين الاختبارات محكية المرجع و الاختبارات معيارية المرجع، نوجزها فيما يلي:

#### أ- الهدف الرئيسي للاختبار:

يعتبر الفرق في الهدف بين الاختبارات محكية المرجع و الاختبارات معيارية المرجع "بالمميز الأساسي" [40]. فالاختبار معياري المرجع يصمم لوضع قاعدة للمقارنة بين المفحوصين، أي مقارنة أداء الطالب بأداء مجموعته المعيارية. أما الاختبار محكي المرجع فيصمم لتسهيل تفسير أداء المفحوصين حسب المحتوى الذي يغطيه الاختبار كما يصمم كذلك قصد الحصول على درجات تستعمل (1) لوصف مستوى أداء المفحوصين (2) وضع تصنيفات لمستويات الأداء (3) تقييم فعالية البرنامج التعليمي .

#### ب- بناء الاختبار:

يتطلب بناء الاختبار، في كلا النوعين، إعداد المواصفات التي تحدد المحتوى المراد قياسه إلى جانب الخصائص والغرض من الاختبار. وبالرغم من أنه يمكن للقياس معياري المرجع استخدام الأهداف لوصف و تحديد النطاق السلوكي، إلا أنه غالبا ما تكون توقعات وضوح النطاق أكبر بالنسبة للقياس محكي المرجع، و يعزى ذلك لطريقة استخدام درجات الاختبار. [47] (ص: 379). كما يتم تصميم الاختبارات معيارية المرجع بصفة مقصودة لغرض ضمان تباين الدرجات للزيادة في ثبات ترتيب المفحوصين عن طريق درجات الاختبار. ويتم ضمان زيادة تباين الدرجات حسب همبلتون عن طريق انتقاء فقرات ذات صعوبة معتدلة (صعوبة البنود يتراوح بين 30. و 70). و قدرة تمييزية عالية (الارتباط بين البند و درجات الاختبار الكلي تفوق 30). [47] (ص: 379). ورغم أن العمل على زيادة تباين درجات الاختبار قد يزيد في صدقها و ثباتها إلا أنه قد ينتج

عنه فشل الاختبار في احتواءه على فقرات تمس المفاهيم الأساسية (Central Concepts) التي ترتبط بمجال تحصيلي معين. [48]. أما في الاختبارات محكية المرجع فيعتمد في بناء الفقرات على مدى تحقيق المفردة للهدف السلوكي . فالفقرات التي لا ترتبط مع مواصفات النطاق السلوكي المتعلق بالاختبار تلغى من البنك.

أما إحصائيات الفقرة فتعتبر أقل أهمية من اعتبارات المحتوى في اختيار الفقرة بالنسبة للاختبارات محكية المرجع. فكل الطرق الإحصائية الكلاسيكية المعروفة تعتبر ذات أهمية بالنسبة للاختبارات معيارية المرجع (إعادة الاختبار، الصور المتكافئة، التجزئة النصفية... الخ). أما فيما يتعلق بالاختبارات محكية المرجع فينتقل التركيز إلى طرق إحصائية جديدة مثل: التوافق بين الفقرة و الهدف أو صدق الفقرة (هل تقيس الفقرة السمة المراد قياسها). و دقة التصنيفات الناتجة عن درجات الاختبار

#### ج- تفسير الأداء :

يتم تفسير درجة الطالب في الاختبار معياري المرجع بناء على درجات معيارية تبين موقع الطالب بين مجموعته الصفية، بينما لا تعتمد تفسيرات الاختبارات محكية المرجع على المقارنة بين درجات المفحوصين، و إنما تفسر مباشرة بإرجاع الدرجات إلى النطاق السلوكي المناسب و مستويات الأداء المرتبطة به . [49].

كما أن التفسيرات معيارية المرجع يمكن أن تتم على أساس الاختبارات محكية المرجع، غير أن المقارنة بين المفحوصين تعتبر إشكالية عند ما تكون درجات الاختبار متجانسة نسبياً. و هذا هو الحال غالباً مع الاختبارات محكية المرجع، حيث أنها لم تصمم لضمان تباين درجات الاختبار مثلما تهدف إليه الاختبارات معيارية المرجع.

#### د- قابلية تعميم الأداء :

إن تعميم الأداء في الاختبارات معيارية المرجع يعتبر غير مبرر. و يرجع هذا إلى طبيعة الطريقة المتبعة في انتقاء بنود الاختبار، و غالباً ما تنتهي المهمة عندما تقارن درجات الاختبار مع جدول المعايير. ونجد بالمقابل أن تعميم الأداء في الاختبارات محكية المرجع يعتبر ذات قيمة. حيث يتم تعميم إتقان الطلاب لنطاق سلوكي معين إلى نطاقات أخرى. فإذا كانت الأهداف التي يقيسها الاختبار واضحة و كانت الفقرات ممثلة للنطاق السلوكي ، أمكن تعميم أداء المفحوص



على فقرات الاختبار إلى أداءه في الاختبار في النطاق السلوكي الأوسع. و في هذا الصدد يرى علام أن "السلوك الذي تتضمنه الكفايات التي يتم اختيارها يجب أن يكون قابلاً للتعميم على نطاق المحتوى الذي تمثله هذه الكفايات". [1] (ص: 379).

فالتلميذ الذي ينجح في التعرف على الفكرة الرئيسية عند قراءة نص في مستوى معين يمكن أن تعمم قدرته للتعرف على الفكرة الأساسية في المواضيع الأخرى التي تكون في نفس المستوى. وفي ضوء ما سبق يمكن القول أنه رغم أن بعض المختصين في القياس التربوي ضيقوا كثيراً من نطاق الاختبارات معيارية المرجع، حيث اعتبروا بأن هذه الأخيرة صممت لغرض المقارنة بين درجة الفرد وجماعته المعيارية دون أن تقدم لنا أي معلومات حول ما يستطيع الفرد أداءه. كأن هذا النوع من الاختبارات لا يرتبط بأهداف ممثلة للمحتوى المراد قياسه. فقرات الاختبار الجيد سواء كان معيارياً أو محكياً يجب أن يرتبط بمجموعة الأهداف الممثلة للنطاق السلوكي المراد قياسه لضمان صدق المحتوى. كما أن أهمية كل نوع ترتبط بالغرض الذي يستخدم لأجله. فقد أشار مهران و ايبل ( Mehren & Ebel ) أنه يوجد مكان في القياس التربوي لكل من التفسيرات محكية الرجوع والتفسيرات معيارية المرجع للاختبار، غير أن السؤال لا يكمن في أيهما نستخدم لكن متى نستخدم أيهما. [50] (ص: 7).

#### 4.2. خطوات بناء الاختبارات محكية المرجع:

لقد تعرفنا من خلال عرضنا لتعاريف الاختبارات محكية المرجع بأنها تستخدم في تحديد مستوى أداء كل فرد بالنسبة إلى نطاق سلوكي من المعارف و المهارات الأساسية المعرفة تعريفاً دقيقاً دون موازنة هذا الأداء بأداء غيره من الأفراد، فالنطاق السلوكي يكون بمثابة المحك الذي ينسب إليه أداء الفرد وهذا بالطبع يجعل تحديد النطاق السلوكي الذي يقيسه الاختبار تحديداً دقيقاً هو المشكلة الأساسية التي تواجه من يقوم ببناء اختبار محكي المرجع. كما يتطلب بناء الاختبار محكي المرجع مجموعة من الخطوات المنظمة نوضحها فيما يلي:

#### 1.4.2. تحديد النطاق السلوكي المراد قياسه:

يقصد بالنطاق السلوكي مجموع المعارف و المهارات المحددة تحديداً دقيقاً بما يمكننا من معرفة ما يستطيع الفرد أداءه وما لا يستطيع. ويرتبط تحديد النطاق السلوكي بتحديد محتوى البرنامج المراد قياسه وتحليله إلى عناصره الأولية. فمن بين أهم الأسئلة التي تتطلب إجابة

في المراحل الأولى لبناء الاختبار هي: ما هو المحتوى المراد قياسه؟ ويشير داونين (Downing) أنه لا توجد خطوة أكثر أهمية في المراحل الأولى لبناء الاختبار مثل تحليل نطاق المحتوى حتى تكون مفردات الاختبار عينة ممثلة فعلا لهذا المحتوى. فإذا لم يتم تحديد نطاق المحتوى بصفة جيدة فلن يتسن تدارك هذا الخلل في الخطوات الموالية . [51] (ص: 7).

ويرتبط تحليل المحتوى المراد قياسه بطبيعة و حدود هذا المحتوى. فإذا كان هذا الأخير محددا مثل وحدة دراسية معينة أو موضوع دراسي واضح المعالم فإنه يمكن الاكتفاء بمعرفة مكونات هذه الوحدة. أما إذا كان النطاق المراد دراسته واسعا و كبيرا فإنه من المستحسن تقسيمه إلى موضوعات فرعية مرتبطة ببعضها البعض، بحيث يمكن قياسها كوحدة منفصلة ، و يتطلب ذلك بناء عدة اختبارات كل منها يرتبط بأحد هذه الموضوعات الفرعية. و لا توجد قاعدة ثابتة لتقسيم المحتوى المراد قياسه ولكن يجب أن يسمح هذا التقسيم ببناء اختبار أو اختبارات تكون مفرداتها بمثابة عينة ممثلة تمثيلا كافيا للموضوعات المختلفة التي يشتمل عليها المحتوى. كما يجب مراعاة طبيعة الطلاب المختبرين و الفترة الزمنية اللازمة لتعليم هذا المحتوى .

#### 2.4.2. تحديد الكفايات التي سيقاسها الاختبار:

يتضمن كل محتوى أو نطاق سلوكي على مجموعة من الكفايات، فيستخدم الاختبار لقياس مدى تحققها لدى المتعلم. و تعبر الكفاية عن نواتج متوقعة من عملية التعليم. غير أنها تحتاج إلى إعادة صياغة حتى تعبر عن أنشطة سلوكية تكون دليلا على أن المتعلم قد اكتسب سلوكا معينا .

#### 3.4.2. تحليل الكفايات إلى أهداف إجرائية (سلوكية):

و نعني بذلك إعادة صياغة الكفايات عن طريق وصف عينة من الأهداف السلوكية التي يمكن أن تتخذ دليلا على تحقق كل كفاية من الكفايات. ويهدف هذا التحليل إلى تحديد المتطلبات اللازم توفرها لدى المتعلم لكي يحقق الكفاية التي يقيسها الاختبار .

#### 4.4.2. تحديد مواصفات الاختبار:

يختلف الاختبار محكي المرجع عن غيره من الاختبارات، في تحديد مواصفاته تحديدا تفصيليا دقيقا. [1]. ويرى بوبام (Popham) أن تحديد الأهداف التي يقيسها الاختبار و صياغتها

بطريقة إجرائية (سلوكية) لا يعد كافيا لبناء الاختبار محكي المرجع ، إذ يجب أن تعد مواصفات أكثر تفصيلا للنطاق السلوكي الذي يتضمن هذه الأهداف. [11]. كما أن عملية بناء مواصفات الاختبار يعمل على توجيه النشاطات المفصلة لبناء الاختبار. وتعتبر مواصفات الاختبار تعريفا إجرائيا كافيا لخصائص الاختبار. ويجب أن تشتمل مواصفات الاختبار على مايلي:

أ - الوصف العام للسلوك: ويمكن اعتبار هذا الوصف هو صياغة للأهداف السلوكية.

ب- خصائص المثير: ويقصد بذلك تحديد ووصف المثيرات التي تشتمل عليها مفردة الاختبار. فمفردة الاختبار تعد بمثابة مثير يقدم للطالب لكي يستجيب له بطريقة معينة. لذلك يجب تحديد ووصف مكونات هذا المثير وتحليله إلى عناصره الأولية.

ج- خصائص الاستجابة: ويقصد بذلك تحديد ووصف طريقة استجابة الفرد لخصائص المثيرات التي تشتمل عليها مفردة الاختبار. فإذا كانت المفردة من نوع الاختيار من متعدد فإنه يجب تحديد طبيعة الإجابة الصحيحة، وكذلك قواعد تكوين المشتتات، بحيث يمكن أن يسترشد بها الفرد الذي يقوم بكتابة المفردات.

#### 5.4.2. كتابة مفردات الاختبار:

إن كتابة فقرات اختبار ذات فاعلية تعتبر فنا أكثر منها علما [51] (ص: 10). وذلك بالرغم من وجود قواعد علمية صلبة لكثير من المبادئ في بناء فقرات الاختبار. كما أن بناء أسئلة اختبار فعالة مصممة لغرض قياس محتوى مهم حسب مستوى معرفي معين، يعتبر من أكبر التحديات بالنسبة للمختصين في بناء الاختبارات، ذلك أن معرفة مبادئ كتابة فقرات الاختبار لا تضمن بالضرورة قدرة المختص على إيجاد أسئلة اختبار مناسبة. لهذا فقد أشار عبدي (Abedi) أن من أهم قضايا الصدق (validity issue) المرتبطة ببناء الاختبار تخص انتقاء و تدريب المشاركين في بناء فقرات الاختبار. [52] .

لذا فإن كتابة فقرات الاختبار تحتاج إلى عناية كبيرة. فمفردات الاختبار محكي المرجع تبنى على أساس مواصفات النطاق السلوكي الذي سبق إعداده. لذلك يجب أن تقيس هذا النطاق بدرجة كبيرة من الدقة، و أن يكون مستوى صعوبة كل مفردة مناسبة لمستوى صعوبة الهدف الذي تقيسه و لمستواه المعرفي . و أن تكون عينة المفردات ممثلة للنطاق السلوكي للأهداف.

#### 6.4.2. تجميع مفردات الاختبار :

إن عملية تجميع الفقرات للوصول إلى الشكل النهائي يعتبر خطوة مهمة في بناء الاختبار .  
فصدق درجات الاختبار ترتبط بمدى دقة هذه الخطوة. فقد أشار داونين أن الأخطاء في عملية  
تجميع الفقرات تقلل من صدق درجات الاختبار. [51] (ص: 12).

أشار صلاح الدين محمود علام [44] (ص: 322). إلى أنه بعد الانتهاء من كتابة المفردات  
ينبغي تنسيقها في ضوء الأهداف التي تقيسها لتكوين الاختبار. وهذا يتطلب مراعاة التسلسل  
المنطقي للمفردات. كما ينبغي أن تجمع المفردات المتشابهة من حيث النوع معا في ضوء مكونات  
المهارة التي تم تحليلها و مستوياتها المعرفية. و ينبغي أن يسبق كل نوع منها تعليمات توجيهية  
دقيقة توضح للتلميذ ، مثلا ، كيف يجب وأين يكتب إجابته ، و تعد هذه التعليمات جزءا  
من الاختبار، كما ينبغي تحديد الزمن المخصص للاختبار .

#### 7.4.2. تجريب الاختبار :

بعد الإنتهاء من كتابة المفردات وتجميعها ينبغي تجريب الاختبار ميدانيا على عينة مناسبة  
من الأفراد لمعرفة مدى ملاءمة الفقرات للمختبرين و درجة تمييزها بين الأفراد الذين حققوا هدفا  
معينا و الذين لم يتمكنوا من تحقيقه ، وذلك لتقييم صدق الفقرات في قياس النطاق المتعلق بالأهداف  
المحددة.

#### 8.4.2. تحليل فقرات الاختبار:

يتمثل الهدف العام من القياس محكي المرجع في تقييم الأداء على مجموعة المهام الممثلة  
لنطاق سلوكي معين. لهذا الغرض فان المختصين في بناء الاختبارات يستخدمون تقنيات مختلفة  
لضمان صدق محتوى فقرات الاختبار و أنها تعكس حقيقة المحتوى المراد قياسه و ذلك من خلال  
عرضها على الخبراء في المادة. فإذا وجدت هناك فقرات غير مناسبة فيجب أن تحذف. وفيما يلي  
عرض لأهم التقنيات المستخدمة لغرض تحليل فقرات الاختبار:

#### 1.8.4.2. صعوبة البند:

يعرف مستوى صعوبة البند بصفة عامة في الاختبارات محكية المرجع على أنه نسبة المفحوصين الذين أجابوا على البند إجابة صحيحة ( p ) . و أشار كروكر و أجاينا ( Croker & Algina ) على أنه يمكن تحديد متوسط أو وسيط مستوى الصعوبة لمجموعة البنود التي تقيس نفس الهدف، حيث يمكن لهذه القيمة أن تكون مفيدة لمعرفة مدى فعالية تلقين هذا الهدف [53]. فإذا كان البند سهلا جدا في القياس القبلي مثلا، يجعلنا نتساءل فيما إذا كان من الضروري أو من غير الضروري تدريس محتوى معين. كما أن البنود التي تبدو أكثر صعوبة بعد عملية التعليم قد تدل على أن عملية التعلم لم تكن فعالة أو أن مواصفات البند تشتمل على محتوى لم يتم تغطيته في الأهداف التعليمية. فقد ترجع سهولة البنود أو صعوبتها إلى غموض في صياغة السؤال أو إحياء غير مقصود للإجابة... الخ. كما أن التباين في صعوبة أو سهولة البنود التي تقيس نفس الهدف تقدم لنا معلومات مفيدة.

#### 2.8.4.2. معامل القياس القبلي - البعدي:

ويسمى أيضا بمعامل حساسية البنود لعملية التعلم (كوكس و فرجاس ) ويكمن الهدف من قياس الحساسية لعملية التعلم في معرفة مدى جودة البند في التمييز بين المفحوصين الذين تلقوا التعليم و المفحوصين الذين لم يتلقوا التعليم. كما يعتبر القياس القبلي - البعدي جزءا مكملًا للتعلم من أجل الإلتقان و للتدريس التشخيصي لسنوات طويلة. [54] (ص: 53).

كما يفيد القياس القبلي - البعدي أيضا في متابعة التقدم الدراسي للمتعلمين أثناء انتقالهم من وحدة تعليمية إلى وحدة موالية. [54] (ص: 54).

و قد اقترح كل من ( فرجاس Vargas و كوكس Cox ، 1966 ) إجراء يتم بواسطته حساب معامل التمييز القبلي - البعدي و ذلك عن طريق تطبيق الاختبار مرتين على مجموعة واحدة ، قبل و بعد عملية التعلم، حيث يتم تكوين مصفوفة تدون في خلاياها الدرجة التي حصل عليها كل فرد في كل مفردة من مفردات الاختبار و تعطى الدرجة (1) على الإجابة الصحيحة على المفردة و الدرجة (0) على الإجابة الخاطئة أو المتروكة، و يفضل إعطاء وقت كاف للأفراد للإجابة على جميع المفردات. [1] (ص: 188).

و تتمثل الطريقة الإحصائية لحساب معامل التمييز كما يلي : [53].

$$D = P \text{ pos} - P \text{ pre}$$

حيث :

P post: تمثل نسبة الذين أجابوا إجابة صحيحة على البند في الاختبار البعدي.

P pre : تمثل نسبة الذين أجابوا إجابة صحيحة على البند في الاختبار القبلي.

وتتراوح قيمة D بين ( 1.00 و -1.00 ). [53] .

ولكن رغم أهمية طريقة القياس القبلي- البعدي، إلا أنها لاقت مجموعة من الانتقادات ، حيث يعاب على هذه الطريقة صعوبة تطبيق القياس البعدي، إذ ينبغي الانتظار إلى غاية الانتهاء من تدريس الوحدة التعليمية . وهذا قد يؤدي إلى تأجيل تطبيق الاختبار إلى وقت أطول إذا كان الزمن المخصص للوحدة طويل ، أما إذا كان الزمن المخصص للوحدة قصير فقد يؤدي كذلك إلى احتمال تأثر أداء الأفراد في القياس البعدي بأدائهم في القياس القبلي.

ويمكن التغلب على مشكلة إعادة تطبيق الاختبار من خلال بناء مجموعتين متكافئتين من الفقرات أو نلجأ إلى إطالة المدة الزمنية الفاصلة بين التطبيقين على مجموعة واحدة، بحيث لا يسمح هذا الفاصل الزمني بتأثير عوامل النضج على الدرجات. [1] (ص: 198).

#### 3.8.4.2. طريقة معامل التمييز لمجموعة الدارسين و غير الدارسين :

تعتبر هذه الطريقة إحدى الأشكال التي استخدمت لسنوات عديدة في التعرف على صدق اختبارات الشخصية و التي طبقت حديثاً على الاختبارات محكية المرجع. [54] (ص: 54). كما تعتبر من الأساليب التي تعتمد في تحليل مفردات الاختبارات المحكية المرجع على اختيار مجموعتين مختلفتين من الأفراد في وقت واحد. إحداهما لم تتلق التعليم و الأخرى تلقت التعليم و يطبق عليهما الاختبار في الوقت نفسه .

يمكن اختيار المجموعة الأولى من بين الطلاب الذي تلقوا تعليماً فاعلاً في أحد الفصول المدرسية و يعرف عنهم معلموهم أنهم حققوا أهداف الوحدة التعليمية التي يقيسها الاختبار. أما المجموعة الثانية فيمكن اختيارها من بين الطلاب الذين لم يتلقوا تعليماً في هذه الوحدة. [1] (ص: 199). و يجب أن تكون المجموعتان (الدارسين ، غير الدارسين ) متشابهتين قدر الإمكان

في مستوى القدرة و جميع الخصائص الأخرى ذات العلاقة. و على الرغم من أن هاتين المجموعتين غالباً ليستا متساويتين في الحجم فإن التوزيع النسبي للخصائص يجب أن يكون متساوياً. فعلى سبيل المثال إذا تكونت مجموعة من 60 بالمائة أولاد و 40 بالمائة بنات فإن هذا التوزيع يجب أن يتضمن في المجموعة الأخرى، و أن الفرق الحقيقي الوحيد الموجود بين المجموعتين هو في تعرضهم للتعليم. [54] (ص: 56).

و من مميزات الأساليب التي تعتمد على قياس مجموعتين من الأفراد أن يكون استخدامها في تحليل مفردات الاختبار في وقت واحد، و ربما يكون ذلك قبل بدء عملية التعليم إذا وجدت مجموعة من الأفراد تلقت التعليم بالفعل، و هذا يبسر الاستخدام الفوري للاختبار أثناء عملية التعليم في تصنيف الأفراد إلى متمكنين و غير متمكنين من النطاق السلوكي للأهداف التعليمية. و لكن يعاب على هذه الأساليب صعوبة تعريف المحك الذي يجري في ضوءه اختيار كل من المجموعتين تعريفاً إجرائياً. كما يعاب عليها صعوبة التأكد من تكافؤ هاتين المجموعتين. [1] (ص: 200).

#### 4.8.4.2. طريقة معامل التوافق المرجعي:

يعتبر هذا المعامل من أحدث معاملات التمييز الخاصة بتحليل فقرات اختبارات الإلتقان و يهدف معامل التوافق المرجعي إلى معرفة احتمالية التوافق بين نواتج فقرة معينة و نواتج الاختبار وهو مشابه في قاعدته النظرية لطريقة معامل استجابة البند [54]. و تعتمد هذه الطريقة على تطبيق الاختبار مرة واحدة على مجموعة واحدة من الأفراد و من ثم يتم تصنيف أفراد هذه المجموعة إلى متقنين و غير متقنين بناء على مدى تحقيقهم للمستوى المطلوب للإلتقان. و قد اقترح كل من هاريس (Harris) و سبكوفياك (Subkoviak) المعادلة التالية لحساب معامل التوافق المرجعي. [55] (ص: 112)

$$\text{معامل التوافق المرجعي} = \frac{a + d}{n}$$

حيث :

ا = عدد الأشخاص المتقنين الذين أجابوا عن السؤال إجابة صحيحة .

د = عدد الأشخاص من غير المتقنين الذين أجابوا عن السؤال إجابة خاطئة .

ن = العدد الكلي للأشخاص .

و ينحصر مدى المعامل بين (صفر) و (1+). و يمكن حساب الحد الأدنى لمعامل التوافق حينما لا تكون هناك علاقة بين مستوى الإلتقان و الاستجابة على السؤال.

و يحسب الحد الأدنى لمعامل التوافق المرجعي من خلال الجدول الثنائي على النحو التالي :

الجدول الثنائي (2 x 2)

الأداء على الاختبار			
غير متقن	متقن		
ب	أ	صح	الإجابة على الفقرة
د	ج	خطأ	

و يتم حسابه باستخدام العلاقة التالية:

$$\frac{(أ+ب)(ج+د)+(ب+د)(أ+ج)}{ن^2} = \text{الحد الأدنى لمعامل التوافق المرجعي}$$

- حيث: أ = عدد الأشخاص المتقنين الذين أجابوا عن السؤال إجابة صحيحة.  
 ب = عدد الأشخاص غير المتقنين الذين أجابوا عن السؤال إجابة صحيحة.  
 ج = عدد الأشخاص المتقنين الذين أجابوا عن السؤال إجابة خاطئة .  
 د = عدد الأشخاص من غير المتقنين الذين أجابوا عن السؤال إجابة خاطئة .  
 ن = العدد الكلي للأشخاص .

يمكن اعتبار الفقرة جيدة وفق معامل (ت.م) إذا كان الفرق بين الحد الأدنى لمعامل التوافق المرجعي و معامل التوافق المرجعي أكبر من أو يساوي 0,05 .

#### 5.8.4.2. طريقة معامل فاي :

يبين هذا المعامل درجة التوافق في التصنيف بين الفقرة والاختبار للمفحوصين، و هو من الطرق التي يتم فيها تطبيق الاختبار مرة واحدة على مجموعة واحدة من الأفراد. و يتم اختيار درجة فاصلة تمثل مستوى الإتقان. وتتحدد فاعلية الفقرة بقدرتها على التمييز بين المفحوصين عند درجة فاصلة محددة على العلامة الكلية على الاختبار .



و يتم إيجاد معامل فاي عن طريق الجدول الثنائي (2 في 2) حيث يبين هذا الجدول عدد الإجابات الصحيحة و الخاطئة للمتقنين و عدد الإجابات الصحيحة و الخاطئة لغير المتقنين .  
الجدول الثنائي (2 x 2)

الأداء على الاختبار			
راسب(غير متقن)	ناجح (متقن)		
ب	أ	صح	الأداء على الفقرة
د	ج	خطأ	

و يحسب معامل فاي للجدول الثنائي على النحو التالي : [56].

$$\text{معامل فاي} = \frac{\text{أد - ب ج}}{\sqrt{(\text{أ} + \text{ب})(\text{ج} + \text{د})(\text{أ} + \text{ج})(\text{ب} + \text{د})}}$$

حيث:

- أ = عدد المتقنين الذين أجابوا عن الفقرة إجابة صحيحة.
  - ب = عدد غير المتقنين الذين أجابوا عن الفقرة إجابة صحيحة.
  - ج = عدد المتقنين الذين أجابوا عن الفقرة إجابة خاطئة .
  - د = عدد غير المتقنين الذين أجابوا عن الفقرة إجابة خاطئة .
- و تعتبر الفقرة جيدة وفق معامل فاي إذا كانت قيمته أكبر من أو تساوي 0,30 .

#### 9.4.2. تقدير صدق الاختبارات محكية المرجع :

يرتبط معنى الصدق في الاختبارات محكية المرجع بمدى دقة استخدام درجات الاختبار في تحقيق ما ينتظر تحقيقه من الاختبار، أي أن يدل على مدى ملاءمة القرارات المتخذة و التي تتم على أساس نتائج الاختبار. فصدق الاختبار لا يرجع إلى الاختبار في حد ذاته بل إلى استخدام أو تفسير درجات الاختبار. وفيما يلي عرض لبعض نماذج تقدير الصدق في الاختبارات محكية المرجع :

#### 1.9.4.2. الصدق الوصفي:

ذكر بابام [11] بأن الصدق الوصفي عبارة عن محاولة لتأكيد مدى ما يقيسه فعلا الاختبار محكي المرجع. أي أن الصدق الوصفي لا يقتصر على التحقق من صدق الفقرات فحسب بل يرمي إلى التحقق من مدى تمثيلها للأهداف التي تعكس المحتوى المراد قياسه. ومفهوم الصدق الوصفي يناظره صدق المحتوى في الاختبارات معيارية المرجع ، إلا أن الصدق الوصفي يعد أعم وأشمل . و يعتبر تحديد الصدق الوصفي شرطا أساسيا لأنواع الصدق الأخرى، و ذلك لأنه أفضل وسيلة لتفسير أداء الطالب في الاختبار. ويمكن تحديد الصدق الوصفي باستخدام أساليب مختلفة، و من بينها الاستعانة بمجموعة من المتخصصين في المجال الذي يقيسه الاختبار لتحديد مدى مطابقة مفردات الاختبار مع المجال السلوكي المراد قياسه .

#### 2.9.4.2. الصدق الوظيفي :

أشار علام أنه ما يقصد بالصدق الوظيفي هو دقة الاختبار محكي المرجع في تحقيق الغرض الذي صمم لأجله الاختبار ، أي أنه لا يقتصر على وصف أداء الطالب في الاختبار فحسب ، بل يستخدم عندما يكون الهدف من تطبيق الاختبار هو وصف و تحديد الوظيفة المراد تحقيقها من الاختبار. [1] (ص: 290). لذا فإن الصدق الوصفي يعتبر شرطا لازما لحساب الصدق الوظيفي . و هذا النوع من الصدق يناظره الصدق التجريبي في الاختبارات معيارية المرجع غير أن مفهوم الصدق الوظيفي يعتبر أعم من الصدق التجريبي.

و يمكن تحديد صدق القرار الذي يعتبر نوعا من أنواع الصدق التجريبي ، بمقارنة أداء مجموعتين محكيتين أو أكثر بدرجة قطع محددة . كما يمكن تحديد صدق القرارات التعليمية المبنية على درجات الاختبارات محكية المرجع بمقارنة أداء المجموعتين (الأولى تلقت التعليم و الثانية لم تتلق التعليم ) بدرجة قطع الاختبار. و يمكن تقييم صدق القرار لكل هدف تعليمي بجمع النسب المئوية للطلاب المتقنين الذين اجتازوا درجة القطع و النسب المئوية للطلاب غير المتقنين الذين لم يجتازوا درجة القطع، و ذلك بهدف تحديد الطلاب المتقنين فعلا و الطلاب غير المتقنين فعلا. و قد أشار علام أنه يمكن حساب قيمة معامل صدق القرار أيضا بإيجاد معامل الارتباط بين متغيرين ثنائيين ، الأول متغير الانتماء إلى إحدى المجموعتين و الثاني متغير قرار الإلتقان. [1] (ص: 291).

### 3.9.4.2. صدق انتقاء النطاق السلوكي :

يرتبط صدق انتقاء النطاق بمدى دقة باني الاختبار في اختيار نطاق سلوكي محدد يمكن أن يستخدم كمؤشر لإيضاح مستوى الطالب بالنسبة إلى نطاق عام يعد أكثر عمومية. وأشار بابام (Popham) أنه يمكن تحديد صدق انتقاء النطاق السلوكي باستخدام أحد الأسلوبين التاليين: [11].  
أ- تدريس مجموعات صغيرة من الأفراد المهارات التي يشتمل عليها كل نطاق من النطاقات السلوكية حتى يصل إلى درجة الإتقان ، يلي ذلك قياس إمكانية تعميم إتقان الطلاب لنطاق واحد تجاه النطاقات الأخرى .

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

من خلال استعراضنا لمختلف أنواع النماذج لتقدير الصدق في الاختبارات محكية المرجع، فقد اختار الباحث استخدام الصدق الوصفي الذي يعتمد على آراء المختصين في مجال الاختبار لتحديد مدى اتفاق كل فقرة من فقرات الاختبار مع مواصفات النطاق السلوكي و ذلك يرجع إلى كونه أكثر شيوعا و أكثر استخداما من قبل المختصين في ميدان القياس التربوي.

### 10.4.2. ثبات الاختبارات محكية المرجع:

تستخدم الاختبارات محكية المرجع في تصنيف الأفراد إلى مجموعتين أو أكثر بحسب درجة تمكنهم من العناصر الأساسية التي تقيسها الاختبارات ، لذلك يعتبر علماء القياس التربوي أن ثبات درجات الاختبارات محكية المرجع تعني اتساق قرارات التصنيف بناء على هذه الدرجات. و المعروف أن مدى التباين بين الأفراد يصبح أقل ما يمكن عندما يتقن جميع الأفراد العناصر الأساسية التي يقيسها الاختبار. لذلك فإن المفاهيم الإحصائية التي تعتمد أساسا على حساب مدى التباين بين الأفراد لا تصلح أن تستخدم في تقدير ثبات الاختبارات محكية المرجع ولكن تستخدم بدل ذلك الطرق الإحصائية التي تعتمد على قرارات تصنيف الأفراد إلى متمكنين و غير متمكنين. [1] (ص: 291).

و قد توصل العلماء في هذا المجال إلى عدد من الطرق لحساب ثبات الاختبارات المحكية المرجع حيث يمكن تصنيفها في مجموعتين، إحداهما تتطلب تطبيق الاختبار مرة واحدة و الأخرى تتطلب تطبيقه مرتين. و فيما يلي سنتطرق إلى أهم الطرق في كل مجموعة :

#### 1.10.4.2. طرق تعتمد على تطبيق الاختبار مرة واحدة:

##### أ- معامل ليفنجستون: Livingston Index

ويعتمد على مفاهيم النظرية الكلاسيكية في تقدير ثبات الاختبارات معيارية المرجع والتي تركز على معرفة انحراف درجة الفرد عن متوسط العينة. إلا أنها تستخدم درجة القطع بدلا من المتوسط. فقد ذكر كل من كروكر و الجاينا بأن هذا المعامل يهتم بإيجاد انحراف درجات الأفراد عن درجة القطع و يهتم هذا المعامل أيضا بمفهوم متوسط مجموع مربعات انحرافات الدرجات الملاحظة المتوقعة و ذلك بحساب قيمة انحراف درجات كل طالب عن درجة القطع في الاختبار. و يحسب وفق العلاقة التالية: [53] (ص: 203).

$$K(X, T) = \frac{\sigma^2(T) + (M - C)^2}{\sigma^2(T) + (M - C)^2}$$

حيث:  $\sigma^2(T)$ : ترمز إلى التباين الملاحظ للاختبار.

$\sigma^2(X)$ : ترمز إلى التباين الحقيقي للاختبار.

$C$ : ترمز إلى درجة القطع.

$M$ : ترمز إلى متوسط الدرجات.

يتميز معامل ليفنجستون بسهولة تطبيقه، لكن يعاب عليه تأثره بدرجة القطع في الاختبار. لذلك ينبغي أن تؤخذ درجة القطع بعين الاعتبار عند تفسير قيمة هذا المعامل. [1] (ص: 299).

##### ب - طريقة هاريس: Harris Method

و تعتمد هذه الطريقة على فكرة الارتباط بين متغير متصل يمثل درجة التلميذ في الاختبار و متغير ثنائي يمثل تصنيف التلميذ إلى متمكن أو غير متمكن في ضوء درجة القطع. [11] (ص: 286-287). و يطلق على معامل الثبات الناتج عن تطبيق هذه الطريقة بمعامل هاريس.

و يرمز له بالرمز (Mc). و يتم حسابه وفق معامل ارتباط (بوينت بايسيريال) التالية: [57].

$$Mc = \frac{\bar{x}_1 - \bar{x}_0}{\sigma} \sqrt{Y_1 \cdot Y_0}$$

حيث:

$\bar{x}_1$ : ترمز إلى متوسط درجات التلاميذ المتمكنين.

$\bar{x}_0$ : ترمز إلى متوسط درجات التلاميذ غير المتمكنين.

$\sigma$ : ترمز إلى الانحراف المعياري.

$Y_1$ : ترمز إلى نسبة التلاميذ المتمكنين.

$Y_0$ : ترمز إلى نسبة التلاميذ غير المتمكنين.

### ج- معامل الاتفاق لسبكوفياك (Pc): Subkoviak Coefficient of Agreement

اقترح سبكوفياك Subkoviak طريقة تعتمد على مدى اتساق تصنيف الأفراد في مجموعتين إحداهما متقنة والأخرى غير متقنة باستخدام نتائج تطبيق الاختبار مرة واحدة فقط. و تتلخص الطريقة في تقدير مقدار احتمال تصنيف كل فرد تصنيفا صحيحا (ف)، و جمع هذه المقادير الاحتمالية ، و إيجاد متوسطها ، وبذلك نحصل على معامل إتقان خاص بالمجموعة التي يطبق عليها الاختبار. وقد رمز سبكوفياك لهذا المعامل بالرمز (Po). و يتم حسابه وفق العلاقة الرياضية التالية: [57].

$$Po = \frac{\sum P_x}{N}$$

حيث:

$N$ : ترمز إلى عدد أفراد المجموعة.

$P_x$ : معامل الاتفاق التقديري لكل فرد من أفراد المجموعة.

و يحسب  $P_x$  من العلاقة الرياضية التالية:

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

حيث:

$KR20$ : ترمز إلى معامل التناسق الداخلي للاختبار.

$x$ : ترمز إلى درجة التلميذ في الاختبار.

$n$ : ترمز إلى عدد المفردات.

$M$ : ترمز إلى متوسط درجات التلاميذ.

و يتميز معامل سابكوفياك بعدم تأثره بحجم عينة الدراسة ويمكن استخدامه في حالة وجود أكثر من درجة قطع واحدة.

#### 2.10.4.2. طرق تعتمد على تطبيق الاختبار مرتين:

حيث يتطلب تطبيق الاختبار نفسه مرتين أو تطبيق صورتين متوازيتين لنفس الاختبار و من هذه المعاملات: معامل كارفر: Carver Index.

أشارت سابكوفياك إلى أن معامل كارفر يعد من الطرق الأولى المقترحة لتقدير ثبات الاختبارات محكية المرجع التي تتعلق باتساق قرار التصنيف، و تقوم هذه الطريقة على تطبيق اختبارين متوازيين على مجموعة واحدة من الأفراد ، يليه مقارنة نسبة عدد الطلاب الذين تم تصنيفهم ضمن المتقنين في كلا الاختبارين فإذا كانت النسبتين متماثلتين أو قريبتين من بعضهما فبذلك يعتبر الاختباران ثابتين، ويعاب على هذه الطريقة بأنها غير حساسة لاتساق قرارات كل طالب في الاختبارين. و يمكن إيجاد قيمة معامل كارفر وفق الجدول التالي : [1] (ص: 305).

#### الاختبار (ب)

متمكن

غير متمكن

أ	ب
د	ج

متمكن

الاختبار (أ)

غير متمكن

ويحسب معامل كارفر باستخدام الصيغة التالية : [1] (ص: 305).

$$\frac{أ + ج}{ن}$$

ن

حيث:

أ = مجموع الطلاب المتمكنين بالفعل.

ج = مجموع الطلاب غير المتمكنين بالفعل .

ن = أ + ب + ج + د

### ملخص الفصل:

يعتبر التقويم التربوي جزءا لا يتجزأ من العملية التربوية، كما تعتبر الاختبارات التحصيلية أداة فعالة لقياس مخرجات التعليم. وقد قسمها جايغر إلى قسمين بناء على طريقة تفسير درجات الاختبار. يتمثل القسم الأول في الاختبارات معيارية المرجع، حيث يقارن أداء المفحوص بمتوسط الجماعة المعيارية. أما القسم الثاني فيتمثل في الاختبارات محكية المرجع، حيث يقارن أداء المفحوص مباشرة مع مستوى محكي معد مسبقا. و يختلف القسم الأول عن القسم الثاني في مجموعة من الأوجه من بينها: الهدف، طريقة بناء الفقرات و طريقة تفسير الأداء، لذلك فإن أغلب الطرق الإحصائية المستخدمة في القياس معياري لا تتناسب مع متطلبات القياس محكي المرجع. غير أن الاختلاف لا يعني التعارض بقدر ما يعني أن لكل وظيفته في القياس التربوي.

كما قمنا ، من خلال هذا الفصل، بتقديم عرض لمختلف الخطوات المنهجية المطلوبة لبناء اختبار محكي المرجع.

### الفصل 3

#### تحديد درجة القطع

تمهيد :

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 واضحا أن عملية تحديد درجات القطع تعتبر من أهم القواعد في بناء الاختبارات حيث أن الفهم الواضح لتصنيف الأفراد إلى فئات حسب الأداء يؤثر على جميع خطوات بناء الاختبار. غير أن إجراءات هذه العملية تعتبر مهمة معقدة و تؤدي إلى صعوبة التوصل إلى درجة القطع التي تحدد مختلف المستويات و التصنيفات. و في هذا الصدد يشير لارنر (Lerner) بأن مشكل تحديد درجة القطع يشبه تماما ما يتعامل معه المحامون والقضاة في كل وقت وهو: متى و كيف ننهي القضية ؟ [6] (ص: 8-9).

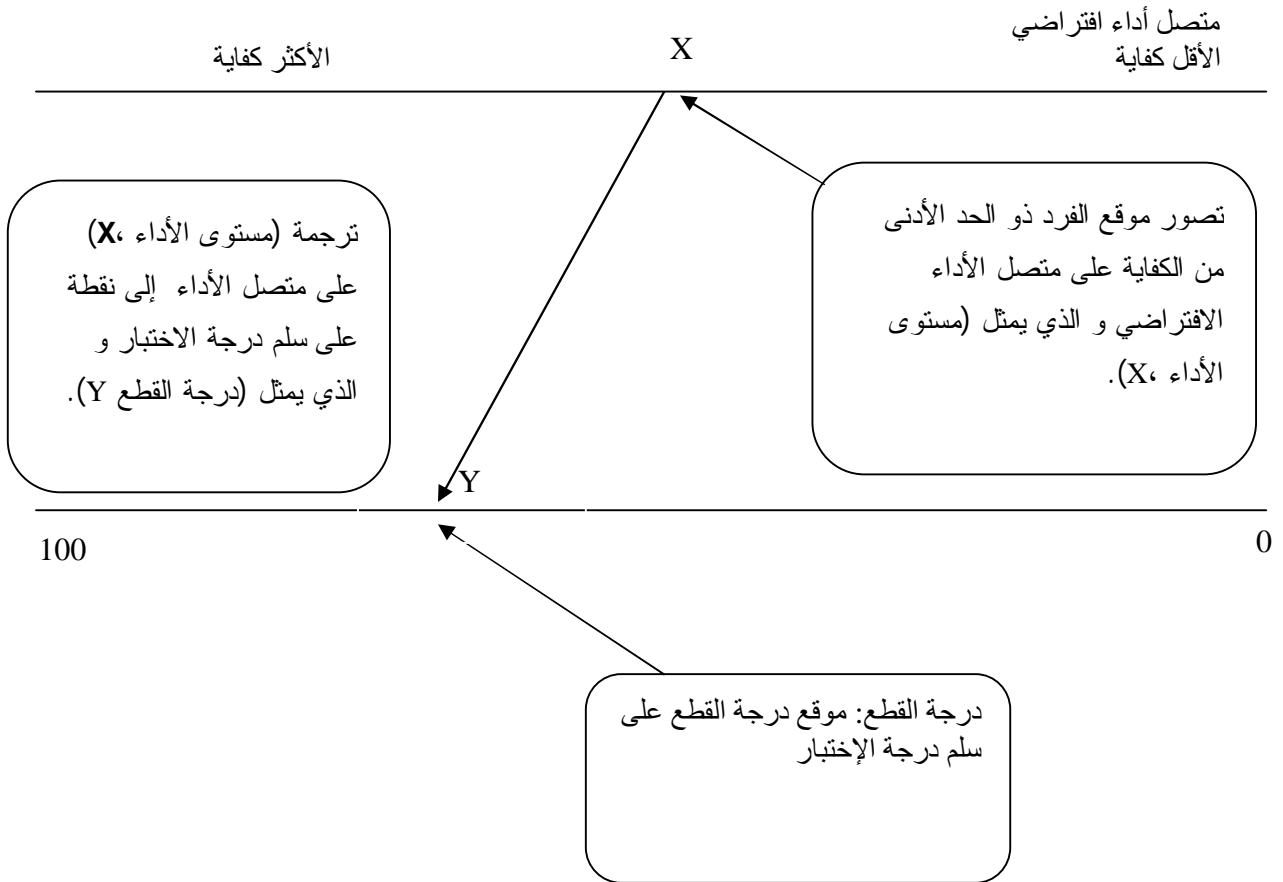
#### 1.3. مفهوم درجة القطع:

أشار زاكي و ليفنجستون (Zieky & Livingston,1983) أن تحديد درجة القطع يتمثل في الإجابة على السؤال التالي: "كم يعتبر كافيا ؟ " . (How much is enough?) [25]. 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." [6] (ص: 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" [14]. 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 مثلما هو مبين في الشكل رقم(1).

شكل رقم (1) يبين الفرق بين مستوى الأداء و درجة القطع. المصدر سيزاك وبانش [6].



يمثل الخط الأفقي الأعلى في الشكل رقم(1) متصل افتراضي للأداء. وتمثل نهايته اليمنى أقل كم ممكن من السمة المقاسة. كما تمثل نهايته اليسرى أكبر كم ممكن من هذه السمة. و يمثل (X) الحد الفاصل بين المقبولين و غير المقبولين، أي مستوى الأداء. أما الخط الأفقي الأدنى في الشكل فيمثل متصل درجة الاختبار أو سلم نسبة الإجابات الصحيحة. والذي يمتد من (0) إلى (100). يقوم المحكمون وفق الطرق التحكيمية المختلفة بترجمة الموقع المجرد (X)

على متصل الأداء الافتراضي إلى نقطة محسوسة على سلم درجة الاختبار، ويشار إلى هذه النقطة بـ (Y) وتمثل درجة قطع.

ولكن حسب سيزاك و بانش، فبالرغم من محاولة ( كايين ) التمييز بين مستويات الأداء و درجة القطع إلا أن هذا التمييز لا يزال غامضا. [6]. ولهذا الغرض فإننا سنستخدم مصطلح مستويات الأداء في هذا البحث كمرادف لمصطلح درجة القطع.

ولقد تعددت مسميات درجات القطع، فأحيانا يطلق عليها مستويات الأداء Performance standards أو مستويات التمكن Mastery Levels ، أو درجات النجاح أو الاجتياز Passing scores، أو الحد الأدنى لمستوى الكفاية Minimum Competency ، أو مستويات المحك Criterion Levels . [1] (ص: 261).

### 2.3. العلاقة بين مستويات المحتوى و مستويات الأداء:

وقد أشار سيزاك (Cizek,2006) أن مفهوم مستويات المحتوى (Content Standards) قد تم إدراجه في الأونة الأخيرة ضمن مجال القياس. حيث تصف مجموعة مخرجات، كفايات، أو أهداف تعليمية خاصة. [58] (ص: 227). فهي ترتبط بالمقرر الدراسي و ما يتوقع من المتعلم أن يعرفه و أن يستطيع أداءه. فيمكن أن يتوقع من المتعلم مثلا، أن يحل مسألة حسابية بسيطة، قراءة نص للفهم أو إجراء تجربة علمية. بينما تعبر مستويات الأداء على ما يجب على المتعلم إظهاره في الاختبار المصمم لغرض قياس مستويات المحتوى. كأن نطلب مثلا من المتعلم أن يحل عشرة مسائل حسابية حتى يعتبر مقبولا أو أن يجيب على خمسين بالمائة من أسئلة القراءة حتى يصنف في فئة المقبولين.

### 3.3. لمحة تاريخية حول مراحل تطور منهجية تحديد درجات القطع :

لقد ميز زاكي (Zieky,1994) بين عدة مراحل لتطور منهجيات تحديد مستويات

الأداء عرضها فيما يلي: [59]

### 1.3.3. مرحلة السذاجة: The Age of Innocence

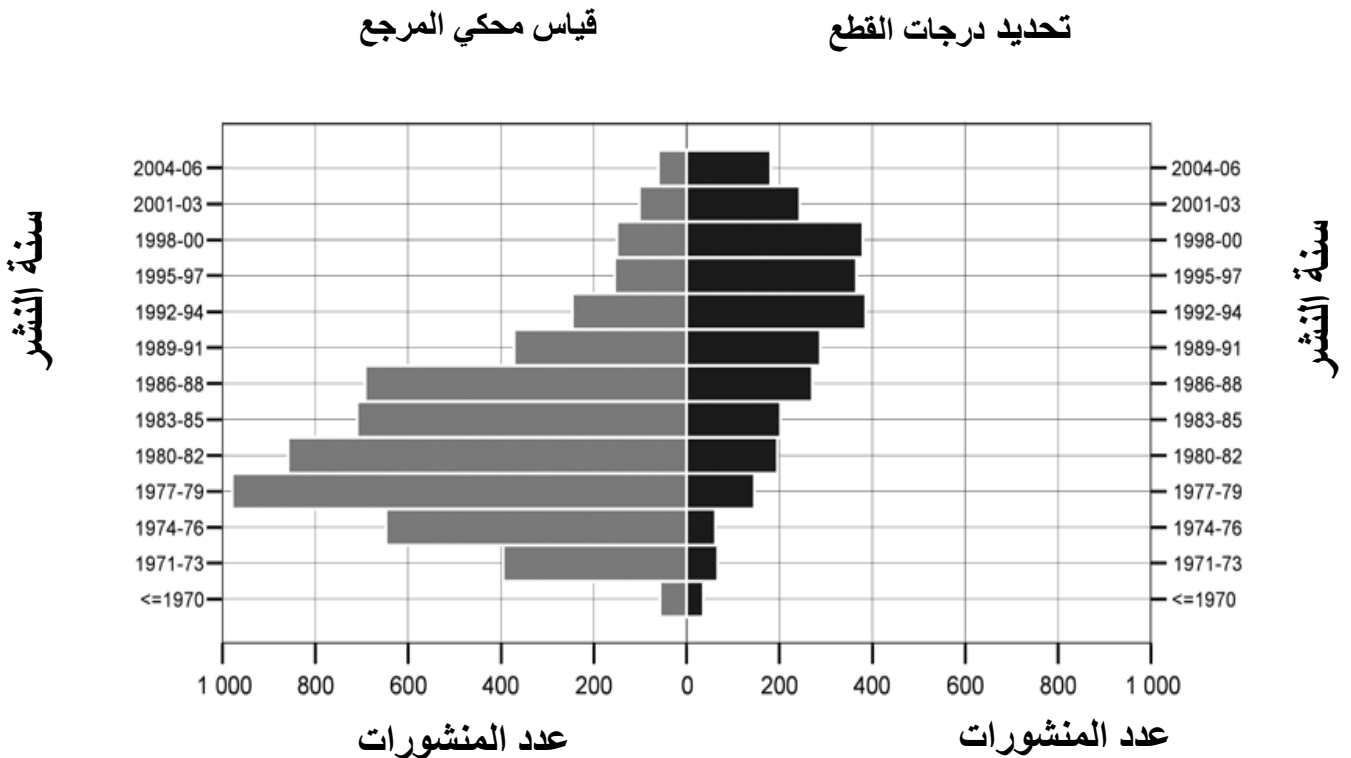
خلال هذه المرحلة التي امتدت لفترة طويلة والتي انتهت مع نهاية ستينيات القرن الماضي، لم يكن يعطى أي اهتمام لكيفية تحديد مستويات الأداء، و لكن هذا لا يعني أنها لم تكن موجودة، لكن لم تكن تخضع للمنهجية العلمية. فحسب زاكي [59] (ص: 4). فإن الطرق السائدة حينذاك كانت طرق تقليدية، كاتخاذ نسبة معينة و ثابتة دون مبرر منطقي كاعتبار (50% درجة قطع مناسبة لأنها دائما كذلك). أو طريقة الموازنة (The Goldilocks method) التي توازن بين نسبة الناجحين و الراسبين. ولكن طول عمر هذه المرحلة انتهى مع ظهور ما يسمى بالقياس محكي المرجع.

### 2.3.3. مرحلة النهضة: The Age of Awakening

ارتبطت هذه المرحلة بظهور القياس محكي المرجع، ما أدى إلى ظهور العديد من طرق تحديد درجة القطع. وقد شهدت هذه المرحلة العديد من البحوث في هذا المجال. الشكل رقم (2) يوضح عدد المنشورات العلمية (باللغة الانجليزية) الخاصة بالقياس محكي المرجع (في الجانب الأيمن) و تحديد درجة القطع (الجانب الأيسر).

شكل رقم (2) المنشورات العلمية المتعلقة بتحديد درجات القطع و القياس محكي المرجع.

[8] (Kaftandjieva,2010)



ما يلاحظ في الشكل رقم (2) أعلاه هو أن عدد المنشورات العلمية المرتبطة بالقياس محكي المرجع قد بلغ ذروته مع نهاية السبعينيات ثم بدأ بعد ذلك في التناقص، غير أن البحوث الخاصة بتحديد درجة القطع استمرت في التزايد، حيث أن أغلب البحوث تظهر في التسعينيات.

ولقد كان للمقال الذي نشره جلاس (Glass,1978) بالغ الأثر لظهور الفترة الحرجة في تطور عملية تحديد مستويات الأداء، حيث وجه فيه انتقادات لاذعة لعملية تحديد درجة القطع بصفة شاملة. إذ وصفها بأنها اعتبارية بشكل صارخ "Blatantly Arbitrary" [59] (ص: 16) وقد دعا جلاس إلى الحد من الطرق العشوائية في اتخاذ القرارات، أو الكف عن البحث في المستويات المحكية لصالح الطرق التي تستخدم بيانات الاختبار، إذ تعتبر هذه الأخيرة أقل عشوائية وبالتالي أكثر أمنا. ولكن بالرغم من أن البعض من المختصين في هذا المجال كان ينظر إلى جلاس نظرة خيبة نظرا للانتقادات التي قدمها، إلا أنه كان لمقاله الأثر العميق في ظهور المرحلة الموالية.

### 3.3.3. مرحلة الخيبة: The Age of Disillusionment

قد أشار زايكي أن مرحلتي النهضة و الخيبة كانتا متداخلتين "بعض المختصين في القياس كانوا يعيشون في مرحلة الخيبة في حين كان يعيش البعض الآخر في مرحلة النهضة [59]. وقد كانت النقطة الحاسمة في هذه المرحلة هو تحقق المختصين أنه لم تتمكن أي طريقة من طرق تحديد درجة القطع من تغيير الواقع الذي أثبت أن هذه الطرق ذاتية بطبيعتها. وتشير Kaftandjieva أنه " في الواقع، الطبيعة الذاتية لتحديد مستوى الأداء هي نقطة الخلاف في ميدان قياس التحصيل. وهذا الأمر يعتبر غريبا، ذلك أنه في نظرية اتخاذ القرار (the theory of decision-making) كانت هذه الخاصية معروفة من قبل إلا أنه لم يثر كل هذا النقاش المحتدم" [8] (ص: 19) . و يمكن أن نرجع دوافع اتساع رقعة النقاش في هذه المرحلة إلى الأهمية المرتبطة بتحديد درجة القطع التي لا يمكن تجاهلها أو الاستغناء عنها. فهي لا تؤثر على مستقبل المفحوصين فحسب بل تعتبر قاعدة للعديد من القرارات الإدارية المرتبطة بتسيير النظام التربوي. و في هذا الصدد يشير سيزاك (2001) بأن "تحديد درجات القطع يعتبر الفرع الوحيد في القياس النفسي الذي يمزج في مكوناته بين الجوانب الفنية السياسية و الثقافية للوصول إلى النتيجة النهائية". [60] (ص: 5).

وقد تولدت عن هذه النقاشات نتائج مهمة دعت إلى مزيد من البحث في مختلف الخطوات الموضوعية الخاصة بإجراءات تطبيق طرق تحديد درجة القطع، ما ساهم في ظهور المرحلة الموالية.

### 4.3.3. مرحلة التقبل الحقيقي: The Age of Realistic Acceptance

شهدت بداية الثمانينيات من القرن الماضي اهتماما خاصا بموضوع تحديد درجات القطع حيث بدأ التعريف المفصل لعدد من طرق تحديد درجات القطع، طريقة نيدلسكاي (نيدلسكاي، 1954) ( طريقة أنجوف (أنجوف، 1971)، طريقة ايبيل (ايبيل ، 1972 )، طريقة المجموعة الحديدية، طريقة المجموعات المتضادة. فقد أصدر كل من زاكي و ليفنجستون عام 1982 دليلا تحت عنوان "درجة القطع: دليل لتحديد مستويات الأداء على الاختبارات التربوية و المهنية". و قد شرحا فيه مختلف الخطوات المتبعة للوصول إلى تحديد درجة القطع الخاصة ببعض الطرق. [61].

فمع بداية 1983، أضحى موضوع تحديد درجات القطع معروفا إلى درجة أن المعلومات الخاصة بمختلف الطرق أصبحت متضمنة في ما وصفه نيتكو ( Nitko, 1983 ) "النصوص القاعدية الخاصة بمدخل القياس للمربين في الابتدائي و الثانوي". فلم يعد موضوع تحديد درجات القطع يهم فئة خاصة ( esoteric topic ) تقتصر على المختصين في القياس النفسي أو الإحصاء فقط، وإنما أصبحنا نجده في جميع النصوص التربوية ". [62] (ص: 26)

### 4.3. التطورات الحالية لعملية تحديد درجات القطع:

واجه الكثير من المختصين في تحديد درجات القطع العديد من التحديات مع بداية الثمانينيات حيث أصبحت مهمتهم أكثر تعقيدا بسبب التطورات التي حصلت في هذا الميدان. فقد انتقل التصنيف الثنائي لمستوى الأداء (ناجح/راسب) إلى تصنيف متعدد المستويات (مبتدئ، متوسط متقدم...). كما انتقل التركيز على الاختبارات البسيطة إلى اختبارات أكثر تعقيدا. و مع ارتفاع نسبة فشل التلاميذ ازداد الضغط على المختصين في تحديد درجة القطع لفرض نوع من التوازن بين الصرامة التي تتطلبها عملية تحديد درجة القطع وبين الواقع الذي يتطلب التخفيف من نسبة الفشل. كما تطورت عملية تحديد درجة القطع لمواكبة التغيرات التي شهدتها منهجية القياس ومنها مايلي:

- تغير في شكل بنود الاختبار من الاستجابات المقيدة أو المنتقاة إلى الاستعمال المتزايد لفقرات الإجابة المفتوحة أو المبنية ( Constructed Response ).

- تغيير في تفسير نتائج الاختبارات، التي انتقلت من التصنيف الثنائي ( dichotomous ) مثل (ناجح/راسب) إلى التصنيف المتعدد ( polytomous ) والذي يمثل تدرجا أكثر تفصيلا في التصنيف.

- تغييرات مرتبطة بالاستخدام المكثف لنظرية استجابة البند (IRT) في بناء و تحليل نتائج الاختبارات والتي أصبحت تقوم بتعويض النظرية الكلاسيكية. بالإضافة إلى تطبيق درجات القطع على الاختبارات المكيفة.

- تغيير في تطبيق الاختبار بإدخال التكنولوجيا الحديثة (الإعلام الآلي، الانترنت). حيث ساهم التطور التكنولوجي في تخفيض التكاليف التي كان يتطلبها تطبيق الطرق المختلفة، مما ساعد على مضاعفة عدد المشاركين، وعدد دورات التدريب، وسيولة المعلومات. كما أصبح من الممكن إدراك بعض الأخطاء بصفة آلية ليتم تصحيحها مباشرة من قبل المشاركين.

و لكن بالرغم من التطورات الحاصلة في ميدان تحديد درجات القطع، إلا أن بعض الخصائص الأساسية ما تزال ثابتة. فقد أكدت الدراسات أنه لا توجد درجة قطع صحيحة (True). بحكم أن درجة القطع تبنى ولا تكتشف (cutscores are constructed, not found). و قد أشار جيجر أنه لا توجد الإجابة صحيحة، ربما إلا في أذهان المحكمين [33] (ص: 492). وبالتالي فإيجاد درجة القطع التي تعتبر مناسبة تعتمد على قيم ذاتية وترتبط بمدى الضرر الذي تسببه أخطاء التصنيف. حيث أن تعديل درجة القطع لتكون مرتفعة أو منخفضة، للتقليل من أحد أنواع الخطأ يعتمد على اعتبارات عملية عديدة. قد تكون أهمها النتائج المترتبة عن أداء المفحوص في سوق العمل. فقد يفضل المحكمون في مواقف معينة الرفع من درجة القطع لتقليل النوع الأول من الخطأ لتفادي الضرر الناتج عنه مثل: (قبول مرشحين ليكونوا طيارين في الملاحة الجوية) بينما يفضل في مواقف أخرى تخفيض درجة القطع للتقليل من الخطأ الثاني، مثل: (الانتقال إلى وحدة دراسية موالية).

### 5.3. الخطوات المشتركة بين مختلف طرق تحديد درجة القطع:

#### 1.5.3. تحديد الهدف :

تنطلق كل طرق تحديد درجة القطع من الاتفاق على الهدف من تحديد مستويات الأداء وعدد درجات القطع الواجب تحديدها و كيفية استخدامها. و في هذا الصدد لخص كاين (Kane, 1994) [14] أهمية تحديد الهدف فيما يلي : " قبل اختيار أي نموذج لتحديد درجة القطع، يتعين

علينا التحقق فيما إذا كان من الضروري أو المفيد استخدام درجة القطع ، فإذا اعتبرنا بأن استخدامها ضروري أو مفيد، فيجب أن يكون ما نريد تحقيقه من خلال استخدامها لقرارات مثل (ناجح/راسب) واضحا. حتى تكون أهدافنا موجهة لاختياراتنا عبر مختلف مراحل تحديد درجة القطع."

### 2.5.3. اختيار نموذج لتحديد درجة القطع:

بالرغم من البحوث المستمرة في هذا المجال لغرلة وتقنين الشروط الملازمة لتطبيق طرق تحديد درجة القطع قصد الوصول إلى نتائج صادقة ، إلا أن تحديد كيفية تفضيل طريقة على أخرى لغرض معين مازال غامضا، إذ أنه لا توجد هرمية واضحة لترتيب الطرق الموجودة حسب معيار معين.

### 3.5.3. كتابة مواصفات مستويات الأداء:

كانت أغلب طرق تحديد درجة القطع تعتمد على التصنيف الثنائي لمستويات الأداء (ناجح/راسب). ولكن ظهرت في بعض البرامج الحديثة تصنيفات متعددة مثل (مبتدئ ، متوسط متقدم ) و في هذه الحالة يجب تقديم وصف دقيق و مفصل لما يجب أن يعرفه و يتقنه الفرد في كل مستوى، إلى جانب التمييز الواضح بين المستويين المتتابعين و المتصلين من الكفاية. فقد أشار سيزاك و بانث ( Cizek & Bunch ) أن التخطيط الأولي لأي اختبار يستخدم لتصنيف الأفراد إلى فئات حسب الأداء يجب أن يشمل دراسة مفصلة حول طبيعة هذه الفئات. [6]. لذا فتحديد درجة القطع لا يجب أن يسبق الاتفاق على مواصفات المستوى. كما يتوجب على كل المشاركين أن يكونوا في نفس مستوى الفهم للحد الأدنى المقبول لمستوى أداء معين.

### 4.5.3. المفاهيم الأساسية:

تعتبر هذه الخطوة مشتركة بين مختلف طرق تحديد درجة القطع، و تتمثل في حاجة المشاركين في تكوين تصور مشترك حول مفاهيم معينة يعتمد عليها في تقديم التقديرات، بالإضافة إلى اعتبارها مرجعا يعود إليه المشاركون خلال مختلف مراحل تحديد درجة القطع حتى يكون للدرجة الناتجة معنى. و الأمثلة على هذه المفاهيم كثيرة و متنوعة ، فعند استخدام طريقة أنجوف مثلا، يطلب من المشاركين تفحص كل بند من بنود الاختبار لتقدير احتمال إجابة الفرد ذو الحد الأدنى من الكفاية إجابة صحيحة. فقدره المشاركون على التصور الصحيح لهذا الفرد الافتراضي

يعد مهما لنجاح الطريقة ومن ثم صدق درجة القطع الناتجة. و الشيء نفسه بالنسبة لطريقة المجموعات المتضادة، و طريقة المجموعة الحدية [61]. ففي الطريقة الأولى يطلب من المشاركين تكوين تصور صحيح حول مفهوم الفرد المتمكن و الفرد غير المتمكن. كما يطلب من المشاركين في الطريقة الثانية تكوين تصور مشترك و صحيح حول الفرد الذي يقع في الحد الفاصل بين التمكن و عدم التمكن.

### 5.5.3. اختيار و تدريب المشاركين في تحديد درجة القطع:

تعتبر طريقة انتقاء المشاركين في تحديد درجة القطع إلى جانب فعالية التدريبات، عاملان مهمان لهما تأثير كبير على نجاح العملية، كما يعتبران مصدرا للتباين في درجة القطع الناتجة. و قد أشار سيزاك و باناش أن هذين العاملين لهما تأثير كبير على درجة القطع الناتجة أكثر من الطريقة ذاتها. [6]. كما أن العوامل الأكثر تأثيرا في درجة الاتساق كثيرة و مختلفة. و أهمها الخبرة السابقة للمحكم في ميدان الاختبارات بالإضافة إلى مدة و طبيعة النشاطات التدريبية قبل دورة التقديرات الحقيقية. [8]. غير أن السؤال الذي يتكرر كثيرا عند القيام باختيار المشاركين لتحديد درجة القطع هو: "ما هو العدد الأدنى من المحكمين الذين يشاركون في تحديد درجات القطع على اختبار معين؟". و للإجابة على هذا السؤال، قام بعض المختصين ببعض الدراسات لتحديد العدد المناسب، حيث اقترح ليفنجستون و زاكي أن لا يقل عدد المحكمين عن خمسة. [61] (ص: 16). و باعتمادهما على التحليل باستخدام نظرية القابلية للتعميم اقترح كل من هاتز و هارتز (Hutz & Hertz, 1999) [63] من عشرة إلى خمسة عشرة محكما. و أشارت كافتجيفا أنه كلما زاد عدد المحكمين كلما كان أحسن حيث يؤدي زيادة المحكمين إلى التقليل من الخطأ المعياري. [8]. لهذا الغرض يجب أن تكون عينة المشاركين كبيرة بما فيه الكفاية حتى تكون ممثلة لمجتمع الأفراد المؤهلين للمشاركة. و بعد الانتهاء من هذه الخطوة، يجب إعداد برنامج تدريبي يتضمن:

أ- تدريب المشاركين على الطريقة المستخدمة: ويتم في هذه الخطوة تزويد المشاركين بوثائق ترتبط بإجراءات تحديد درجة القطع. حيث تتضمن معلومات حول خصوصيات الطريقة المستخدمة، الهدف من تحديد درجة القطع و المحتوى الذي تطبق عليه درجة القطع. بالإضافة إلى القيام بدورات قصد تسهيل عملية الاتصال بين المشاركين و فتح المجال لكشف الغموض عن جميع جوانب عملية تحديد درجة القطع.



ب- تدريب المشاركين على المحتوى: يقوم المشاركون في هذه المرحلة بالإجابة على أسئلة الاختبار إلى جانب القيام بالتصحيح الذاتي قصد التعود على محتوى الاختبار، وتتم هذه العملية قبل القيام بتحديد درجة القطع. ثم تقدم لهم بعد ذلك الإجابة النموذجية.

### 6.5.3. تزويد المشاركين بالتغذية الراجعة:

تعتبر هذه الخطوة مشتركة بين مختلف طرق تحديد درجة القطع، حيث يتم تزويد المشاركين بالتغذية الراجعة وذلك بعد أن يقدموا تقديراتهم حول البنود، المهام، المفحوصين... الخ حسب طبيعة كل طريقة. وتتلخص التغذية الراجعة عموماً في (التوزيعات التكرارية، المتوسط... الخ). حيث تبلغ للمشارك بصفة فردية. كما تتضمن الكثير من طرق تحديد درجة القطع على جولات حيث يتم في كل جولة تزويد المشاركين بمجموعة من المعلومات تتضمن ملخصاً حول تقديرات المحكمين، التناسق الداخلي، مقارنة بين تقديراتهم و تقديرات مشاركين آخرين، التباين بين التقديرات و احتمال تأثير تقديرات الفرد أو المجموعة على مجتمع المفحوصين. و يمكن تقسيم نوع المعلومات المقدمة للمشاركين إلى ثلاثة أقسام :

أ- معلومات وصفية Normative Information: و تتمثل في تزويد المشاركين بمعطيات حول تقديراتهم بالمقارنة مع تقديرات الآخرين.

ب- معلومات واقعية Reality Information: و تتمثل في تزويد المشاركين بمعطيات حول الأداء الحقيقي للمفحوصين على البنود مثل معاملات الصعوبة .

ج- معلومات التأثير Impact Information : و التي تتمثل في تزويد المشاركين بمعطيات حول تأثير تقديراتهم على مجموعة المفحوصين ( نسبة المصنفين في فئة معينة في حالة تطبيق درجة القطع الناتجة ).

### 7.5.3. تقييم العملية:

يخدم تقييم عملية تحديد درجة القطع هدفين أساسيين وهما: (1) تقديم التغذية الراجعة حول التدريبات ومستوى فهم المشاركين و(2) تحديد مستوى الرضا عن العملية و درجة القطع الناتجة والتي تعتبرها عنصراً هاماً في ضمان صدق مستويات الأداء. [64]. كما أن تكرار عملية التقييم يتم عبر مختلف مراحل عملية تحديد درجة القطع.

و يجب أن تتم عملية التقييم، على الأقل، بعد التدريب و عند نهاية دورة درجة القطع. و ينصح بالقيام بالتقييم بعد الجولة الأولى من اقتراح درجة القطع، كما يفضل في بعض الأحيان تقييم فهم

المشاركين بعد كل جولة. وعند نهاية العملية يجب القيام بالتقييم النهائي (جدول رقم 1) لتوثيق مستوى رضا المشاركين عن النتيجة. [6].

### جدول رقم (1) عناصر تقييم عملية تحديد درجة القطع. ( سيزاك و بانش، 2007) [6].

عناصر التقييم	الوصف
1- إجرائية	
الوضوح	مدى وضوح عملية أو أهداف تحديد درجة القطع.
القابلية للتطبيق	إمكانية تطبيق الإجراءات و تحليل المعطيات.
التطبيق	إلى أي مدى كانت الإجراءات التالية ( اختيار و تدريب المشاركين ، تعريف مستويات الأداء و جمع المعطيات ) معقولة ، منتظمة و تمت بصفة صحيحة.
التغذية الراجعة	مدى ثقة المشاركين في العملية و في درجة القطع الناتجة.
التوثيق	مدى توثيق مميزات و خصائص العملية.
2- داخلية	
الاتساق داخل الطريقة	دقة التقديرات الخاصة بدرجة القطع.
اتساق تقديرات المشارك	إلى أي مدى استطاع المشارك تقديم تقديرات متسقة مع صعوبات البنود، و إلى أي مدى تتغير التقديرات مع كل جولة.
الاتساق بين المشاركين	مدى اتساق تقديرات البند ودرجة القطع بين المشاركين.
اتساق القرار	مدى اتساق تصنيف المفحوصين عند إعادة تطبيق درجة القطع المقترحة.
قياسات أخرى	اتساق درجة القطع مع أشكال البند، المحتوى ، والعمليات المعرفية.
3- خارجية	
المقارنة مع محكات خارجية	علاقة القرارات الناتجة عن استخدام الاختبار مع محكات أخرى ذات صلة، مثل (الدرجات، الأداء على اختبارات تقيس نفس السمة.. الخ )
المقارنة مع الطرق الأخرى	مدى توافق درجات القطع الناتجة عن الطريقة المستخدمة مع درجات القطع الناتجة عن الطرق الأخرى .
واقعية درجة القطع	إلى أي مدى تعتبر درجة القطع الناتجة واقعية و قابلة للتطبيق.

### 6.3. طرق تحديد درجة القطع:

توجد العديد من الطرق في ميدان القياس تتعلق بتحديد درجة القطع على الاختبارات محكية المرجع. وقد حاول المختصون في القياس تقسيم هذه الطرق إلى مجموعات منطقية، إلا أن هذه التقسيمات تبقى إلى يومنا هذا غير دقيقة [8] وبالتالي يبقى التصنيف الأكثر شيوعاً والمستخدم بصفة واسعة هو ما جاء به جيغر، حيث صنف طرق تحديد درجة القطع إلى فئتين: طرق ترتكز على بنود الاختبار. (Test- Centred) و طرق ترتكز على المفحوصين. (Examinee- Centred). [33] (ص: 493).

#### 1.6.3. طرق ترتكز على بنود الاختبار :

إن أغلب طرق تحديد درجة القطع المستخدمة في الوقت الراهن، تتطلب من المشاركين في تحديد درجة القطع التركيز على كل فقرة أو مهمة في الاختبار، أو مجموعة من المهام وال فقرات بطريقة أو بأخرى ، و تقديم أحكام حول أداء المفحوص الافتراضي، مثل (الفرد ذو الحد الأدنى من الكفاية) على هذه الفقرات أو المهام واضعين في اعتبارهم المتغيرات التي تؤثر في تحديد هذه الكفاية قصد الحصول على درجة القطع المناسبة. ومن بين أهم الطرق التي ترتكز على فقرات الاختبار مايلي:

#### 1.1.6.3. طريقة ايبيل :

اقترح ايبيل (Ebel) عام 1972 طريقة لتحديد درجة القطع تعتمد على تعليقات المحكمين على مفردات الاختبار، حيث أنه على خلاف الطرق الأخرى فإنه يطلب من كل محكم تقديم تعليقين و ذلك باعتبار أن مواصفات المفردات لها بعدين هما: بعد الصلة و بعد الصعوبة . و يشتمل بعد الصعوبة على ثلاثة مستويات (سهلة، متوسطة، صعبة) . و يشتمل بعد الصلة على أربعة مستويات (أساسية ، مهمة، مقبولة، قابلة للطرح). (جدول رقم 2). و بذلك يستخدم المحكم أعداداً مختلفة من تصنيفات الصعوبة والصلة، ثم يطلب من كل محكم ما يلي :

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

ب- تحديد النسبة المئوية لعدد الطلاب ممن لديهم الحد الأدنى من الكفاية . الذين يجب أن يجيبوا عن كل مفردة بصورة صحيحة و ذلك لكل خلية من الخلايا. ثم يضرب عدد مفردات كل خلية

في النسبة المئوية، ثم يقسم مجموع نواتج جميع الخلايا على عدد المفردات، و بالتالي نحصل على درجة القطع . [33].

الجدول ( 2 ) يوضح إجراءات تصنيف الأسئلة و تقديرات المحكمين في طريقة ايبيل

الصلة	الصعوبة		
	سهل	متوسط	صعب
أساسي	الأسئلة: 1، 4، 7، 8، 13. التقديرات: 95 % صحيحة	الأسئلة: 11، 15، 22. التقديرات: 85 % صحيحة	الأسئلة: 21 التقديرات: 80 % صحيحة
مهم	الأسئلة: 2، 6، 9. التقديرات: 90 % صحيحة	1، 14، 20 التقديرات: 75 % صحيحة	الأسئلة: 16، 25 التقديرات: 60 % صحيحة
مقبول	الأسئلة: 5 التقديرات: 80 % صحيحة	الأسئلة: 12، 18 التقديرات: 55 % صحيحة	الأسئلة: 19، 23 التقديرات: 35 % صحيحة
قابل للطرح	الأسئلة: 5 التقديرات: 50 % صحيحة	الأسئلة: لا يوجد لا تحتاج إلى تقدير	الأسئلة: 17، 24 التقديرات: 20 % صحيحة

المصدر: درجات الإجتياز ( ليفنجستون و زاكي، 1982). [61]

### 1.6.3. 2. طريقة نيدلسكاي :

اقترحت هذه الطريقة من طرف ليو نيدلسكاي (Lio Nedelsky) عام 1954 لتستخدم فقط مع الاختبارات ذات الاختيار من متعدد، و من مميزاتا أنها تطبق مع وجود أو عدم وجود معطيات أداء المفحوصين على الاختبار. و بصيغة أخرى يمكن تحديد درجة القطع قبل أو بعد التطبيق الفعلي للاختبار. كما تعتمد هذه الطريقة على الاسترشاد بالمتخصصين في المجال الذي يقيسه الاختبار لكي يقوموا بثلاث خطوات (جدول رقم3) وهي :

- تحديد البدائل في كل مفردة من مفردات الاختبار و التي يستبعدها الفرد ذو الحد الأدنى من الكفاية من تفكيره عند اختيار البديل الصحيح.
- تسجيل مقلوب عدد البدائل المتبقية.
- تجميع مقلوبات مفردات الاختبار لكي ينتج درجة القطع من وجهة نظر المحكم ، و بأخذ متوسط هذه الدرجات على كل المحكمين نحصل على درجة القطع. [1].

و تفترض هذه الطريقة أن المفحوص ذو الحد الأدنى المقبول من الكفاية يمتلك مستويات مقبولة من المعرفة و المهارات و يستطيع استبعاد البدائل التي يعتبرها غير صحيحة أثناء التفكير في اختيار البديل الصحيح ، و مهمة المحكمين هي تحديد هذه البدائل. [26].

جدول رقم(3) يمثل طريقة نيدلسكي المطبقة على الاختبار دون التصحيح من أثر التخمين. [61]

السؤال	الإجابات	الإجابات غير المستبعدة	الدرجة المتوقعة
1	ا (ب) <del>ب</del> <del>ج</del> <del>د</del> <del>هـ</del>	2	.50 = 2/1
2	<del>ا</del> <del>ب</del> <del>ج</del> <del>د</del> (هـ)	1	1 = 1/1
3	<del>ا</del> <del>ب</del> (ج) <del>د</del> <del>هـ</del>	2	.50 = 2/1
4	ا (ب) <del>ج</del> <del>د</del> هـ	3	.33 = 3/1
5	<del>ا</del> <del>ب</del> <del>ج</del> <del>د</del> (هـ)	1	1 = 1/1
6	ا ب (ج) د هـ	5	.20 = 5/1
7	ا ب <del>ج</del> (د) هـ	4	.25 = 4/1
8	(ا) ب <del>ج</del> د هـ	4	.25 = 4/1
9	ا (ب) ج د هـ	5	.20 = 5/1
10	ا (ب) ج د هـ	5	.20 = 5/1

المجموع = 4.43

مجموع الدرجة المتوقعة 4.43

- تمثل الدائرة الإجابات الصحيحة .

- تدل علامة ( x ) على الإجابة التي يستبدها الفرد ذو الحد الأدنى من الكفاية.

و من مزايا هذه الطريقة أنها سهلة الاستعمال و يمكن تطبيقها قبل تطبيق الاختبار. و من عيوبها أنها محدودة في الاختبارات ذات الاختيار المتعدد.

وأشار ندلسكاي إلى أنه لتلافي مشكلة خطأ القياس يمكن الحصول على الانحراف المعياري لدرجات القطع الناتجة، ثم يتم ضرب قيمة الانحراف المعياري في مقدار ثابت يتفق عليه المحكمون وهذا المقدار يعتمد على التوزيع الطبيعي لدرجات القطع للاختبار، ثم تضاف القيمة الأخيرة إلى درجة القطع التي تم تحديدها لنصل إلى درجة القطع بالطريقة المعدلة. وقيم الثابت التي يمكن أن يختار منها المحكم هي ( 2 أو ) وهي تقابل النسبتين المئويتين (84،98) على الترتيب، وهذه النسبة تعني نسبة فشل المجموعة الحدية (ذوي الحد الأدنى من الكفاية) في الوصول لدرجة القطع. [1] (ص: 266).

غير أن الطريقة المعدلة تعترها الصعوبة من ناحية فهم المحكمين لها بالإضافة إلى أن معظم الدراسات تستخدم الطريقة الأصلية لسهولة فهمها من جانب المحكمين . [5].

### 3.1.6.3. طريقة أنجوف :

أقترحت هذه الطريقة من طرف أنجوف (Angoff, 1971) وهي طريقة مماثلة لطريقة نيدلسكاي . غير أن طريقة أنجوف لا تتطلب من المحكم تحديد الاختيار غير الصحيح على كل مفردة من مفردات الاختبار، بل يطلب من المحكم في هذه الطريقة فحص كل مفردة من مفردات الاختبار، ثم يطلب منه تصور مجموعة الأفراد الذين حققوا الحد الأدنى من الكفاية التي يقيسها الاختبار، ثم تقدر نسبة عدد الأفراد الذين يحتمل أن يجيبوا إجابة صحيحة على كل مفردة من مفرداته ويمثل متوسط هذه النسب الحد الأدنى لمتوسط الاجتياز في الاختبار.

وقد أجرى أنجوف تعديلات على طريقته بحيث تيسر تقدير القيم الاحتمالية لمفردات الاختبار من متعدد لذلك يمكن تصميم استمارة تشتمل على التعليمات، وأرقام مفردات الاختبار، وتوضع أمام كل منها الحرف المناظر للإجابة الصحيحة للمفردة داخل مربع. وكذلك ميزان يشتمل على عشر فترات متساوية تقريبا (1=صفر-10)، (2=0.11-0.20)، (3=0.21-0.30)، (4=0.31-0.40)، (5=0.41-0.50)، (6=0.51-0.60)، (7=0.61-0.70)، (8=0.71-0.80)، (9=0.81-0.90)، (10=0.91-1.0). ثم يطلب من كل محكم تقدير احتمال أن يعرف الفرد ذو الحد الأدنى من الكفاية المطلوبة الإجابة عن المفردة دون أن يلجأ إلى التخمين، و أن يضع دائرة حول الرقم الذي يحتوي على هذه القيمة على الميزان المعطى. أي أن هذه الطريقة تتطلب من كل محكم التنبؤ بمستوى صعوبة كل مفردة في الاختبار لمثل هذا الفرد. [1] (ص: 271-272).

و لتقدير درجة القطع بطريقة أنجوف يقوم الباحث بتجميع تقديرات المحكمين لجميع مفردات الاختبار و التي تمثل درجة القطع التي حددها كل محكم للاختبار ككل، و من ثم يتم إيجاد متوسط درجات القطع لجميع المحكمين و يكون هذا المتوسط هو درجة القطع النهائية للاختبار التحصيلي.

وهناك صورة أخرى لطريقة أنجوف ظهرت عام 1978 وهي أيضا تعتمد على اختيار المحكمين لنسبة مئوية ، و لكن يكون الاختيار من احتمالين ، حيث يتم سؤال المحكمين عما إذا كان الفرد ذو الحد الأدنى من الكفاية قادرا على الإجابة على المفردة بصفة صحيحة ، و بالتالي يختار المحكم (نعم أو لا). ويتم تحديد درجة القطع في ضوء نسبة الموافقة على (نعم) ثم يتم تحرير الدرجة من أخطاء القياس ، ولكن يعاب على هذا التعديل أنه يقيد المحكمين بنسبتين فقط (100 % "نعم" ) أو (صفر % "لا" ) مما يزيد أخطاء القياس إلى أقصى حد. [3] (ص: 148).

وحدث تعديل آخر لطريقة أنجوف عام 1984 حيث يسترشد الخبراء في عملية التحكيم ببعض البيانات المستمدة من التجريب الميداني لمفردات الاختبار، وفي هذه الطريقة يطلب من كل محكم تقدير احتمال إجابة الفرد من ذوي الحد الأدنى من الكفاية المطلوبة عن مفردات الاختبار ككل إجابة صحيحة أي أن تقديرات المحكمين تتعلق بالاختبار ككل و ليس بكل مفردة كما سبق وبعد الحصول على هذه التقديرات يقدم لكل منهم بيانات تجريبية عن مفردات الاختبار تتضمن متوسط صعوبة كل مفردة، ثم يطلب منهم تعديل تقديراتهم الاحتمالية في ضوء ذلك ، وتكون درجة القطع هي متوسط تقديراتهم المعدلة. [1] (ص: 275).

و لقد أوضح رايد (Reid) عام 1984 أن هناك تعديل آخر لطريقة أنجوف وهو مرتبط أيضا بطريقة نيدلسكاي ، و يعتمد على أخذ المتوسط لدرجتي القطع الناتجتين عن طريقة أنجوف و نيدلسكاي إذ تتميز هذه الطريقة بسهولة تنفيذها و فهمها. [3] (ص: 152).

و جدير بالذكر أن طريقة أنجوف و صورها المعدلة بالرغم من النقد الموجه لها، إلا أنها حظيت باهتمام و تفضيل علماء القياس النفسي و التربوي لسهولة فهم المحكمين للإجراءات التي تتطلبها هذه الطريقة. [65] (ص: 252).

ومن مزايا طريقة أنجوف أنها سهلة الاستعمال. [66] و يمكن استخدامها قبل تطبيق الاختبار. [67]. ومن عيوبها صعوبة تقدير الأداء على الفقرات بالنسبة لمجموعة الأفراد ذوي الحد الأدنى من الكفاية، و الميل إلى المبالغة في تقدير الفقرات السهلة. [66] .

### 2.6.3. طرق ترتكز على المفحوصين:

وصف جايجر (Jaeger, 1989) المجموعة الثانية من طرق تحديد درجة القطع على أنها ترتكز على المفحوصين (Examinee Centred) [33]. و أشار كل من (زاكي و ليفنجستون) بأن أي طريقة تحكيمية في تحديد درجة القطع تتطلب بالضرورة من كل محكم التزود بمعلومات حول محتوى الاختبار وخصائص المجتمع الذي تطبق عليه مستويات الأداء. فلا توجد طريقة تستطيع أن تركز على بنود الاختبار فقط، أو على المفحوصين فقط، ولكن تختلف الطرق مبدئياً في التركيز أكثر على بنود الاختبار أو على المفحوصين. ويقوم المحكمون في هذه الطرق بتقسيم المفحوصين إلى عدة فئات حسب مستوى الأداء المطلوب (متمكنين، غير متمكنين و الحد بين التمكن و عدم التمكن) معتمدين في ذلك على محك خارجي [68]. ثم يتم تطبيق الاختبار على هذه الفئات حيث يرتكز تحديد درجة القطع على نتائج الاختبار [58]. و من بين أهم الطرق التي ترتكز في تطبيقها على المفحوصين طريقة المجموعات المحكية، طريقة المجموعات المتضادة وطريقة المجموعة الحدية.

### 2.6.3.1. طريقة المجموعات المحكية :

اقترح هذا النموذج بارك عام 1976، ويتطلب هذا النموذج وجود مجموعتين من الأفراد. إحدهما متقنة للكفاية المطلوبة، وتحقق درجات عالية، و الأخرى غير متقنة للكفاية المطلوبة، وتحقق درجات منخفضة. و باستخدام الاختبارات التشخيصية يمكن اعتبار الطلاب الذين تلقوا تعليماً فاعلاً في الفصل الدراسي متقنين، و الطلاب الذين لم يتلقوا تعليماً في المجال المطلوب غير متقنين. و يطبق الاختبار محكي المرجع على المجموعتين، و يتم عمل التوزيع التكراري لدرجات كل منهما، ويعتبر هذا التوزيع في المجموعتين المتقنة و غير المتقنة المحدد الأولي لمدى دقة تصنيف الاختبار للطلاب كمتقنين و غير متقنين فعلاً للهدف المراد قياسه . و تعد الدرجة ذات التمييز الأفضل بين المجموعتين بمثابة درجة القطع. و يمكن الحصول على هذه الدرجة عن طريق حساب القيم الاحتمالية للتصنيف للوصول إلى أعلى احتمال للتصنيف الصحيح و أقل احتمال للتصنيف الخاطئ. ويمكن استخدام أسلوب تحليل المنفعة و الذي يشتمل على ثلاث مقاربات منفصلة وهي :

(أ) تصنيف الفرد على أساس تحقيق الإتقان دون أن يتلقى التعليم.

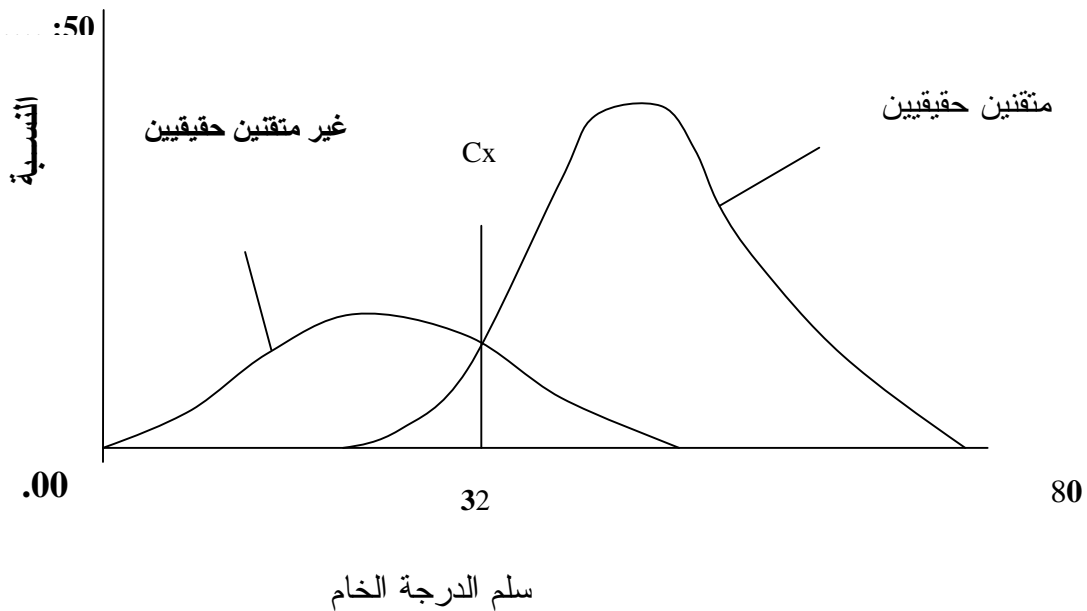
(ب) تصنيف الفرد على أساس عدم تحقيق الإتقان مع أنه تلقى التعليم.



ج) تصنيف الفرد على أساس تحقيق الإتقان مع تلقي التعليم. وكل مؤشر من هذه المؤشرات الثلاثة يقوم أساسا على تقويم محدد للتصنيفات المختلفة لحالة الخطأ و القرارات الصحيحة. و يترك المجال للباحث لتحديد مؤشر واحد مسبقا ليكون النقطة التي يركز عليها التفسير. [5].

### 2.2.6.3. طريقة المجموعات المتضادة :

ذكر كل من ( بارك و جايفر) بأن هذه الطريقة مماثلة لحد ما لطريقة المجموعات المحكية و طريقة المجموعة الحدية ، غير أنها تختلف في كيفية تحديد درجة القطع في الاختبار. ففي هذه الطريقة يقوم المحكمون المتخصصون بتحديد مجموعتين من الطلاب بحيث تكون الأولى متقنة و الأخرى غير متقنة بناء على درجات الطلاب و أدائهم داخل الفصل. ثم يطبق الاختبار على كلتا المجموعتين و يرسم منحنيان يمثلان التوزيعين التكراريين لدرجات الطلاب في الاختبار. و تمثل نقطة تقاطع المنحنين درجة قطع الاختبار، مثل ما هو مبين في الشكل رقم(3)، و يفترض هذا النموذج تداخل التوزيعين التكراريين و عدم تطابقهما. كما يمكن تعديل هذه الدرجة مع الأخذ بعين الاعتبار القيم الاحتمالية التقديرية الناجمة عن أخطاء التصنيف. [1] (ص: 277).



شكل رقم(3) يبين التوزيعات تكرارية لطريقة المجموعات المتضادة: (سيزاك ، 006)[58].

يبين الشكل رقم (3) منحنيان يمثلان التوزيعين التكراريين لدرجات الطلاب في الاختبار، و تمثل نقطة تقاطع المنحنيين ( $Cx=32$ ) درجة قطع الاختبار.

### 3.2.6.3. طريقة المجموعات الحدية :

بالرغم من أهمية طريقة المجموعات المتضادة ، إلا أنه ينظر إليها على أنها تتطلب من المشاركين في العملية مهاما صعبة إلى حد معين ، فهي تتطلب من كل مشارك تصنيف الأفراد إلى فئتين ( متقنين أو غير متقنين ) ، (ناجحين أو راسبين) ، حيث لا يمكن تصنيف كل الأفراد في إحدى الفئتين، إذ يوجد حتما أفراد يقعون خارج هذا التصنيف. فقد يفكر المحكم أن فردا معينا لا يملك المعرفة أو المهارة الكافية التي تؤهله ليصنف ضمن فئة المتقنين ، كما أنه في نفس الوقت لا يفتقر إلى المعرفة أو المهارة إلى درجة تجعله يصنف ضمن فئة غير المتقنين ولكن قد يقع مستوى هذا الفرد في الحد الفاصل بين الإتيقان و عدم الإتيقان. و كبديل لهذه الطريقة أو جدت طريقة المجموعة الحدية (زاكي ، ليفنجستون ، 1977 ) [69]، فعوض إرغام المحكمين على تطبيق التصنيف الثنائي الذي لا يشمل حتما على جميع مستويات المفحوصين، تسمح طريقة المجموعة الحدية بتصنيف المفحوصين ضمن ثلاث فئات و هي:

أ- أفراد متقنون بصفة قطعية.

ب- أفراد غير متقنين بصفة قطعية.

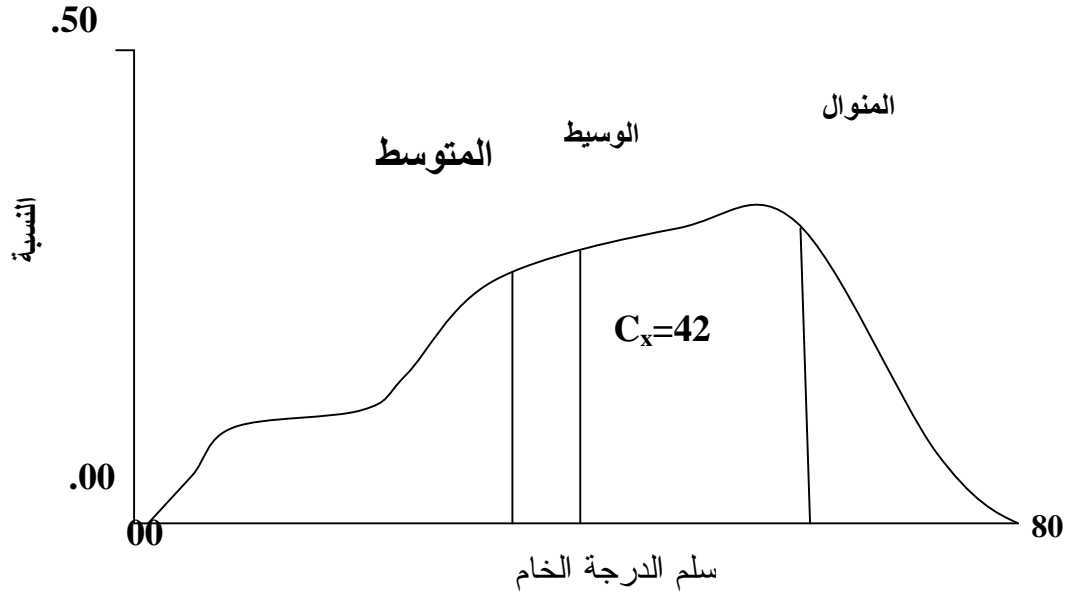
ج- أفراد يقعون في الحد الفاصل.

وبنفس طريقة المجموعات المتضادة ، يجب على المحكمين الذين تم انتقاءهم للمشاركة في إجراء المجموعة الحدية أن يكونوا على دراية بمحتوى الاختبار و لديهم اطلاع بمعارف و مهارات المفحوصين. و لتطبيق طريقة المجموعة الحدية، يجب أن يصل المشاركون عبر النقاش إلى وصف موحد لخصائص المفحوص الذي يقع في الحد الفاصل. وبهذا التصور في الذهن و بدون الإطلاع على أداء المفحوصين على الاختبار، يستخدم المشاركون طريقة المجموعة الحدية لتحديد المفحوصين الذين يقعون في الحد الفاصل بين الإتيقان و عدم الإتيقان.

ثم تستخدم درجات الاختبار الفعلية، ويوضع توزيع تكراري لأداء المجموعة الحدية، ويمثل الوسيط في التوزيع درجة القطع. مثل ما هو موضح في الشكل رقم(4).

و الغرض من استخدام الوسيط عوض المتوسط، هو كون الوسيط أقل تأثراً بالقيم المتطرفة .  
و تعتبر هذه الصفة للوسيط مهمة بالنسبة لطريقة المجموعة الحديدية.

#### توزيع المجموعة الحديدية



شكل رقم (4) يبين التوزيعات الافتراضية بطريقة المجموعة الحديدية. (سيزاك ، 2006) [58].

يبين الشكل رقم (4) التوزيع التكراري لأداء المجموعة الحديدية، ويمثل الوسيط ( $C_x=32$ ) في التوزيع درجة القطع.

و من مزايا هذه الطريقة أنها سهلة الاستعمال و سهلة الشرح، غير أن العائق الأساسي هو كون المجموعة الحديدية تشكل نسبة قليلة جدا ، بالإضافة إلى أن تحديد الفرد الذي يقع عند الحد الفاصل بين التمكن و عدم التمكن أصعب كثيرا من تحديد الفرد المتمكن و غير المتمكن. كما أن هناك متغيرات كثيرة تتدخل في عملية انتقاء أفراد هذه المجموعة، لذلك ينصح البعض بعدم استخدام هذه الطريقة و الاعتماد على الطرق الأخرى . [1] (ص: 277).

## ملخص الفصل:

تم في هذا الفصل التطرق إلى التعريف بدرجة القطع، بالإضافة إلى محاولة التمييز بينها و بين مستوى الأداء. كما تم تحديد الفرق بين مستوى الأداء و مستوى المحتوى. ومحاولة منا إثراء موضوع البحث قدمنا لمحة تاريخية حول مراحل تطور منهجيات تحديد درجة القطع و خصوصيات كل مرحلة بالإضافة إلى توضيح التطورات التي حصلت في عمليات تحديد درجات القطع في الفترة الراهنة . كما حاولنا من خلال هذا الفصل كذلك شرح و توضيح مختلف الخطوات المشتركة بين طرق تحديد درجة القطع. وقد صنفنا هذه الطرق إلى قسمين: طرق تركز على الاختبار ( test centred ) و طرق تركز على أداء المفحوصين ( examinee centred ). و قمنا بعد ذلك بشرح خطوات تطبيق إجراءات بعض الطرق و ركزنا على الطريقتين المستعملتين في البحث الحالي.

## الفصل 4

### الإجراءات المنهجية للبحث:

1.4. مجتمع و عينة البحث: جرت الدراسة الحالية على مجتمعين وهما:

#### 1.1.4. مجتمع التلاميذ:

تكون مجتمع البحث من جميع تلاميذ السنة الرابعة متوسط المنتظمين في الدراسة بالمتوسطات التابعة لبلدية عين الدفلى في الفصل الثالث من العام الدراسي 2010/2011، وقد بلغ عددهم 1290 تلميذ موزعين على ثمانى متوسطات.

#### 1.1.1.4. عينة التلاميذ:

قام الباحث باختيار عينة البحث من مجتمع التلاميذ بطريقة عشوائية على النحو التالي:

- تحديد أسماء جميع المتوسطات ببلدية عين الدفلى التابعة لمديرية التربية لولاية عين الدفلى.
- تم ترتيب الأفواج التربوية التابعة لمتوسطات مجتمع العينة و التي يبلغ عددها 32 فوج تربوي.
- تم عشوائيا، عن طريق القرعة، اختيار 06 أفواج تربوية من المجموع الكلي للأفواج، أي ما نسبته 19 بالمائة. وقد بلغ عدد أفراد العينة 173 .

#### 2.1.4. مجتمع المحكمين:

تكون مجتمع المحكمين من أساتذة الرياضيات الذين لديهم خبرة في تدريس مستوى السنة الرابعة متوسط، ينتمون للمتوسطات التابعة لإقليم بلدية عين الدفلى، ويقومون بتدريس تلاميذ السنة الرابعة متوسط في السنة الدراسية 2010-2011. و قد بلغ عددهم اثنا عشر محكم.

#### 1.2.1.4. عينة المحكمين:

تم اختيار جميع أفراد مجتمع البحث ليمثلوا عينة المحكمين . وقد بلغ عددهم اثنا عشر (12) محكما. وهم يمثلون أساتذة الرياضيات الذين يقومون بتدريس تلاميذ السنة الرابعة متوسط في السنة الدراسية 2010-2011، و الذين لهم الخبرة في تدريس تلاميذ هذا المستوى، لكونهم أكثر معرفة بمحتوى المادة وبمستوى التلاميذ. و سيقوم نفس المحكمين الإثنا عشر بتقدير فقرات الاختبار موضوع البحث باستخدام طريقتي أنجوف و نيدلسكاي للحصول على درجة القطع الناتجة عن كل طريقة وهذا لتفادي تأثير خصائص المحكمين على الإخلاف في النتائج.

## 2.4. منهج البحث:

لغرض تحقيق أهداف البحث استخدم الباحث المنهج الوصفي المقارن وذلك لغرض وصف طريقتي أنجوف ونيديلسكاي من حيث خصائصهما وإجراءاتهما وشروط وظروف تطبيقهما ومقارنة درجات القطع الناتجة عن تطبيق الطريقتين.

## 3.4. بناء الاختبار:

لغرض بناء الاختبار تم إتباع الخطوات المذكورة في الجانب النظري و التي تتمثل في الخطوات التالية:

### 1.3.4. تحديد النطاق السلوكي:

وتم ذلك عن طريق تحديد محتوى المقرر الدراسي المراد قياسه و تحليله إلى عناصره الأساسية. وقد اقتصر اختيار الباحث على وحدة المعالم. فبعد الاطلاع على المقرر الدراسي لتلاميذ ودليل الأستاذ لمستوى السنة الرابعة متوسط التابع لوزارة التربية الوطنية [70]، بين بأن 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 مثل ما هو مبين في الجدول رقم (4).  
جدول رقم (4) يبين المحاور، الوحدات و الحجم الساعي في مادة الرياضيات.

المحور	الوحدات	عدد الساعات
الأنشطة العددية	الحساب الحرفي	10
	الحساب على الجذور	05
	الأعداد الطبيعية و الأعداد الناطقة	10
الدوال و تنظيم المعطيات	جملة معادلتين من الدرجة الأولى بمجهولين.	05
	الدالة الخطية-الدالة التالفية.	20
	المترجمات من الدرجة الأولى بمجهول واحد.	05
	الإحصاء.	10
الأنشطة الهندسية	خاصية طاليس	10
	حساب المثلثات على المثلث القائم.	05
	الأشعة و الانسحاب	10
	المعالم	10
	الهندسة في الفضاء.	15

يتبين من الجدول أعلاه أن وحدة المعالم تدخل ضمن محور الأنشطة الهندسية وقد خصص لها حجم ساعي يقدر بعشر ساعات، وهذا يمثل نسبة 20% من الحجم الساعي المخصص للأنشطة الهندسية.

#### 2.3.4. تحديد الكفايات:

لغرض تحقيق هذه الخطوة قام الباحث بالاتصال بأساتذة مادة الرياضيات للطور المتوسط . و قد بلغ عددهم ثمانية أساتذة فاقت خبرتهم التعليمية خمسة وعشرون سنة. حيث تم الاعتماد على خبرتهم الطويلة بالإضافة إلى دليل الأستاذ. و قد تم التوصل إلى ثمان كفايات نهائية ينبغي أن يتمكن منها كل متعلم و تمثلت فيما يلي:

- 1- يعرف إحداثيتي شعاع.
- 2- يقرأ إحداثيتي شعاع في معلم.
- 3- يمثل شعاع بمعرفة إحداثيتيه.
- 4- يحسب المسافة بين نقطتين في معلم متعامد و متجانس.
- 5- يعرف خصائص الشعاعين المتساويين.
- 6- يعرف أنواع المعالم.
- 7- يحسب إحداثيتي منتصف قطعة.
- 8- 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.

#### 3.3.4. 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 ، و قد تم التوصل إلى صياغة عشرين هدفا سلوكيا يعكس الكفايات ، مثل ما هو مبين في الملحق رقم (1).

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 ملاءمة 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. و قد طلب منهم الإجابة على الأسئلة التالية:

أ- هل الصياغة الإجرائية للأهداف السلوكية ملاءمة ؟

ب- ما مدى انتماء الأهداف السلوكية للكفايات التابعة لها؟

ج - مامدى ملاءمة الفعل السلوكي للمستوى المعرفي؟

و يقوم المحكمون بالإجابة على السؤالين ب(نعم) أو(لا ) مع تقديم الاقتراحات في الخانة المخصصة لذلك. وبعد الحصول على استجابات المحكمين تم تلخيصها في الجدول رقم(5):

جدول رقم(5) يبين نسبة اتفاق المحكمين على وضوح ودقة الصياغة الإجرائية للأهداف السلوكية.

رقم الهدف	عدد المحكمين	عدد مرات الاتفاق(نعم)	نسبة الاتفاق	رقم الهدف	عدد المحكمين	عدد مرات الاتفاق(نعم)	نسبة الاتفاق
01	09	07	%78	11	09	08	%89
02	09	07	%78	12	09	08	%89
03	09	08	%89	13	09	08	%89
04	09	08	%89	14	09	08	%89
05	09	08	%89	15	09	07	%78
06	09	07	%78	16	09	08	%89
07	09	07	%78	17	09	08	%89
08	09	08	%89	18	09	08	%89
09	09	08	%89	19	09	08	%89
10	09	08	%89	20	09	08	%89

يتبين من الجدول رقم (5) أن نسبة اتفاق المحكمين على وضوح ودقة الصياغة الإجرائية للأهداف السلوكية تراوحت 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 مثلما هو مبين في الجدول رقم (6).

جدول رقم (6) يبين التعديلات التي تم إجراؤها على الأهداف السلوكية.

الهدف السلوكي قبل التعديل	الهدف السلوكي بعد التعديل
أن يعرف المتعلم العلاقة بين إحدائتي نقطة وإحدائتي شعاع مبدؤه مبدأ المعلم.	أن يعرف المتعلم العلاقة بين إحدائتي نقطة وإحدائتي شعاع مبدؤه مبدأ المعلم ونهايته تلك النقطة.
أن يحدد المتعلم شعاعا أعطيت إحدائتا مبدئه وإحدائتا نهايته.	أن يحدد المتعلم شعاعا أعطيت إحدائتا مبدئه وإحدائتا نهاية ممثله.
أن يذكر المتعلم إحدائتي شعاع من بيان.	أن يحدد المتعلم إحدائتي شعاع من بيان.

أما نتائج السؤال الخاص بمدى انتماء الأهداف السلوكية للكفايات التابعة فيوضحها الجدول رقم (7).

جدول رقم (7) يبين مدى انتماء الأهداف السلوكية للكفايات التابعة لها.

رقم الهدف	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق	رقم الهدف	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق
01	09	09	100%	11	09	09	100%
02	09	09	100%	12	09	09	100%
03	09	09	100%	13	09	09	100%
04	09	09	100%	14	09	09	100%
05	09	08	89%	15	09	09	100%
06	09	09	100%	16	09	09	100%
07	09	09	100%	17	09	09	100%
08	09	08	89%	18	09	09	100%
09	09	09	100%	19	09	09	100%
10	09	09	100%	20	09	09	100%

يوضح الجدول رقم (7) أن نسبة اتفاق المحكمين حول ملاءمة الأهداف السلوكية للكفايات التابعة لها تراوحت بين 89% و 100%. و تعتبر هذه النسب مرتفعة و تعبر عن تمثيل الأهداف السلوكية للكفايات التي تنتمي إليها.

أما فيما يتعلق بالسؤال المتعلق بمدى ملاءمة 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 تم الخروج بالنتائج المبينة في الجدول رقم (8)، حيث تراوحت نسبة اتفاق المحكمين على ملاءمة الفعل السلوكي للمستوى المعرفي بين 78% و 100%. و هي كذلك نسب اتفاق مرتفعة ومناسبة.

جدول رقم(8) يبين نسبة اتفاق المحكمين على ملاءمة الفعل السلوكي للمستوى المعرفي.

رقم الهدف	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق	رقم الهدف	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق
01	09	09	100%	11	09	09	100%
02	09	09	100%	12	09	09	100%
03	09	09	100%	13	09	09	100%
04	09	09	100%	14	09	09	100%
05	09	07	78%	15	09	09	100%
06	09	08	89%	16	09	09	100%
07	09	09	100%	17	09	09	100%
08	09	09	100%	18	09	08	89%
09	09	09	100%	19	09	09	100%
10	09	09	100%	20	09	09	100%

و قد تم بعد ذلك إحصاء عدد و نسب المستويات المعرفية مثلما هو مبين في الجدول رقم (09).

جدول رقم (09) يبين توزيع عدد ونسب كل مستوى من المستويات المعرفية حسب الأهداف.

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

#### 4.3.4. 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 مثلما هو مبين في الملحق رقم (1) على العناصر التالية:

1.4.3.4. الهدف السلوكي: لتكوين صورة واضحة لما تقيسه الفقرة.

2.4.3.4. خصائص المثير: و قد قسم الباحث هذه الخطوة إلى قسمين وهما :

أ- مدخل المثير : و هو وصف موجز لمقدمة السؤال.

ب- وصف المثير: و هو وصف موجز لما هو مطلوب من السؤال.

3.4.3.4. خصائص الاستجابة: و هي عبارة عن وصف للبدائل.

و في هذا الصدد قام الباحث باختيار الشكل المناسب لأسئلة الاختبار و هو الاختيار

من متعدد. مثلما هو مبين في مخطط مواصفات الاختبار بالملحق رقم (2).

#### 5.3.4. 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 جرونلاند [35] (ص: 39-53). ونعرض منها مايلي:

- تصميم كل سؤال ليقيس ناتج تعليمي مهم.
- يحوي متن السؤال مشكلة واحدة واضحة يستطيع التلميذ فهمها دون اللجوء إلى قراءة البدائل.
- البعد عن عبارات النفي، وإذا كان لابد من وجود النفي فيجب تمييزه.
- التأكد من أن البدائل تحوي إجابة صحيحة واحدة فقط.
- تكون البدائل متفقة نحويا مع متن السؤال و متوازية في الشكل.
- عدم وجود تلميحات تساعد على الإجابة.
- تساوي طول الإجابة الصحيحة مع طول المشتتات.
- تجنب استخدام عبارات "كل مما سبق" أو "لاشيء مما سبق".

بعد الانتهاء من صياغة فقرات الاختبار والتحقق من توفر المعايير العلمية. حيث تحصلنا على فقرتين لكل هدف سلوكي كصورة أولية، للحصول على أكثر مما هو مطلوب. لأن بعضها قد يستبعد عند مراجعته من طرف المحكمين بعدها 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. حيث يجيب المحكمون على مايلي:

أ- هل تم صياغة الفقرات بأسلوب واضح؟

ب- ما مدى انطباق كل فقرة اختباريه على الهدف السلوكي التابعة له ؟

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 مثلما هو موضح في الجدول رقم(10) . 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.

جدول رقم (10) يبين نسبة اتفاق المحكمين حول مدى وضوح صياغة الفقرات.

رقم الفقرة	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق	رقم الفقرة	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق
01	09	09	%100	21	09	09	%100
02	09	07	%78	22	09	09	%100
03	09	09	%100	23	09	09	%100
04	09	09	%100	24	07	07	%78
05	09	07	%78	25	09	09	%100
06	09	08	%89	26	09	09	%100
07	09	09	%100	27	09	09	%100
08	09	09	%100	28	08	08	%89
09	09	09	%100	29	07	07	%78
10	09	08	%89	30	07	07	%78
11	09	09	%100	31	09	09	%100
12	09	08	%89	32	07	07	%78
13	09	08	%89	33	09	09	%100
14	09	07	%78	34	09	09	%100
15	09	09	%100	35	07	07	%78
16	09	08	%89	36	08	08	%89
17	09	08	%89	37	09	09	%100
18	09	08	%89	38	09	09	%100
19	09	09	%100	39	09	09	%100
20	09	09	%100	40	08	08	%89

أما فيما يخص نسبة اتفاق المحكمين حول مدى تطابق الفقرات الاختبارية مع الأهداف السلوكية التابعة لها فقد تراوحت نسبة الاتفاق بين 33% و 100%. مثلما هو مبين في الجدول رقم (11).

جدول رقم (11) يبين مدى تطابق الفقرة مع الهدف السلوكي.

رقم الفقرة	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق	رقم الفقرة	عدد المحكمين	عدد مرات الاتفاق (نعم)	نسبة الاتفاق
01	09	08	%89	21	09	08	%89
02	09	09	%100	22	09	08	%89
03	09	09	%100	23	09	09	%100
04	09	09	%100	24	09	08	%89
05	09	09	%100	25	09	09	%100
06	09	09	%100	26	09	09	%100
07	09	09	%100	27	09	09	%100
08	09	08	%89	28	09	08	%89
09	09	09	%100	29	09	09	%100
10	09	09	%100	30	09	09	%100
11	09	04	%44	31	09	08	%89
12	09	09	%100	32	09	09	%100
13	09	09	%100	33	09	03	%33
14	09	09	%100	34	09	09	%100
15	09	08	%89	35	09	09	%100
16	09	09	%100	36	09	09	%100
17	09	03	%33	37	09	09	%100
18	09	09	100	38	09	08	%89
19	09	09	100	39	09	08	%89
20	09	09	100	40	09	09	%100

و قد تم إلغاء الفقرات رقم: 37، 31، 13. لضعف نسبة اتفاق المحكمين حول ملاءمتها للأهداف السلوكية. أما الفقرات المتبقية فقد تراوحت ما بين 89% و 100% ، مما يدل على أنها تعكس صدق الفقرات في تمثيلها للنطاق الذي يقيسه الاختبار.

وبعد إتمام هذه الخطوة ، اختيرت فقرة واحدة لكل هدف سلوكي بصفة عشوائية.، حيث يصبح كل هدف سلوكي ممثلاً بفقرة واحدة.

### 6.3.4. تطبيق الاختبار:

#### 1.6.3.4. التطبيق على عينة استطلاعية:

أ- تم تطبيق الاختبار بصورته الأولية (20) فقرة على فوج تربوي من خارج عينة الدراسة بلغ عدده 40 تلميذا من تلاميذ السنة الرابعة متوسط. وقد 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 مايلي:

- كانت تعليمات الاختبار وفقراته واضحة لدى مجموع أفراد العينة.

- استغرق متوسط زمن الاختبار 50 دقيقة.

(ب)- وقصد دراسة حساسية الفقرات لعملية التعليم و قدرتها التمييزية وفعالية المشتتات فقد تم تطبيق الاختبار على فوج تربوي تكون من أربعين (40) تلميذا من تلاميذ السنة الرابعة متوسط . تم اختياره بصفة عشوائية، عن طريق القرعة، من خارج عينة الدراسة وذلك قبل تدريسهم وحدة المعالم وبعد تدريسهم لها، وقد تم تطبيق معامل كوكس وفرجاس (Cox & Vergas) لحساب معامل حساسية الفقرات لعملية التدريس. ويعتمد هذا المعامل على تطبيق الاختبار مرتين على مجموعة واحدة من الأفراد ، إحداهما قبل عملية التعليم والأخرى عقب الانتهاء منه. [1] (ص: 189). وقد تراوحت قيم مؤشر حساسية الفقرات لعملية التدريس بين 0.2 للفقرة رقم 14 و0.58 للفقرة رقم 18، وجميعها قيم موجبة تدل على حساسية الفقرات لعملية التدريس. والذي يعتبر مؤشرا من مؤشرات صدق الفقرة في الاختبار مرجعي المحك . [19] (ص: 41). ولغرض التحقق من وجود فرق بين المتوسطين (أداء التلاميذ) قبل تدريسهم وحدة المعالم وبعد تدريسهم لها فقد تم استخدام اختبار t للبيانات المرتبطة، حيث بلغ الفرق بين المتوسطين (7.57) وبلغت قيمة الاختبار t (12.76) وهي دالة إحصائيا عند مستوى الدلالة 0.01 .

- أما فيما يتعلق بفاعلية المموهات فقد تم حساب معاملات التمييز لجميع المموهات، حيث تم حساب الفرق بين نسبة التلاميذ الذين اختاروا بدائل خاطئة في القياس القبلي ، و نسبة التلاميذ الذين اختاروا بدائل خاطئة في القياس البعدي . [71] (ص: 205). وكانت معاملات التمييز للمموهات كلها سالبة و قد تراوحت بين (-0.08 و -0.25) (جدول رقم 12).

الجدول (12) يبين معاملات التمييز وفاعلية الموهبات لكل فقرة اختباريه حيث رتبت الخيارات في الجدول ترتيبيا أبجديا، بعد استبعاد الإجابة الصحيحة.

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

#### 2.6.3.4. 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 بأساتذة 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. (ملحق رقم 3).

#### 7.3.4. تقدير صدق وثبات الاختبار:

##### 1.7.3.4. تقدير الصدق:

لقد قام الباحث بتقدير الصدق الوصفي وهذا ما تحقق سابقا في مرحلة بناء الاختبار، حيث تم إعداد مواصفات الاختبار وتم التحقق من اتفاق المحكين حول ارتباط الأهداف السلوكية بالكفايات التي تنطوي تحتها وتطابق الفقرات مع الأهداف السلوكية، بالإضافة إلى تحليل البيانات التجريبية المتعلقة بالفقرات ( معاملات التمييز، وفاعلية المشتتات) فهذه الخطوات كلها تضمن الصدق الوصفي للاختبار.

##### 2.7.3.4. تقدير الثبات:

نظرا لاختلاف طرق تقدير الثبات في الاختبارات محكية المرجع عن نظيراتها في الاختبارات معيارية المرجع فقد استخدم الباحث الطرق الثلاثة التالية:

أ- طريقة  $\alpha$  لكرونباك.

ب- طريقة البنجستون

ج- طريقة هاريس.

و لقد تم استخدام الطرق الثلاثة للمبررات التالية:

- تم استخدام أكثر من طريقة لغرض التأكد من النتائج و الثقة فيها.
- الطرق الثلاثة تقتضي تطبيق الاختبار مرة واحدة فقط، وتعتبر مناسبة للاختبارات محكية المرجع.

وفيما يلي عرض للطرق المستخدمة في تقدير ثبات الاختبار:

### أ-طريقة $\alpha$ لكرونباك.

وقد تم حساب معامل  $\alpha$  لكرونباك باستخدام العلاقة التالية:

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum S^2}{\sigma^2} \right)$$

و فيما يلي بيان تطبيق هذه الطريقة:

$$\alpha = \frac{20}{20-1} \left( 1 - \frac{3.3}{11.05} \right) = 1.05 \cdot (1.03) = 0.74$$

وما يلاحظ عن هذه الطريقة أنها مناسبة للاختبارات محكية المرجع إلا أنها تعطي تقديرات منخفضة [18]. الأمر الذي يستدعي الثقة بثبات الاختبار لغرض الدراسة.

### ب- طريقة ليفنجستون:

تعتمد هذه الطريقة على فكرة نسبة التباين الحقيقي و التباين الملاحظ حول درجة القطع. وقد تم افتراض 0.65 كدرجة قطع لتقدير الثبات بهذه الطريقة، وهي تمثل الدرجة 13 إذا ضربت في عدد فقرات الاختبار. وقد تم حساب معامل ليفنجستون وفق العلاقة التالية:

$$K(X.T) = \frac{\sigma^2(t) + (M - C)^2}{\sigma^2(x) + (M - C)^2}$$

و فيما يلي بيان تطبيق هذه الطريقة:

$$K(X.T) = \frac{8.18 + (14.7 - 13)^2}{11.05 + (14.7 - 13)^2} = \frac{11.07}{13.96} = 0.79$$

و بالرغم من سهولة تطبيق معامل ليفنجستون إلا أنه يعاب عليه تأثره بدرجة القطع.

ج-معامل هاريس:

تم حساب معامل هاريس و الذي يرمز له بالرمز (Mc) باستخدام العلاقة التالية :

$$M_c = \frac{\bar{X}_1 - \bar{X}_0}{\sigma} \cdot \sqrt{Y_1 \cdot Y_0}$$

وفيما يلي بيان تطبيق هذه الطريقة :

$$M_c = \frac{16.75 - 10.95}{3.32} \cdot \sqrt{0.65 \cdot 0.35} = 0.83$$

ويخلص الجدول رقم (13) قيم معاملات الثبات الناتجة عن تطبيق الطرق الثلاث.

جدول رقم (13) يبين قيم معاملات الثبات الناتجة عن تطبيق الطرق الثلاث.

المعامل	$\alpha$	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.

#### 4.4. تطبيق إجراءات الطريقتين (أنجوف، نيدلسكاي):

##### 1.4.4. طريقة أنجوف:

تم الالتقاء بعينة المحكمين ، حيث عرض عليهم الاختبار محكي المرجع في وحدة المعالم بعد التحقق من صدقه و ثباته و طلب منهم تحديد درجة القطع للاختبار بعد تقديم معلومات تفصيلية لهم عن هدف الدراسة، وتوضيح إجراءات استخدام طريقة أنجوف مزودة بأمثلة تطبيقية للوصول

إلى درجة القطع. كما تم تقديم شرح تفصيلي لمفهوم التلميذ ذو الحد الأدنى من الكفاية (المعارف المكتسبة في وحدة المعالم) والتي تؤهله للانتقال إلى الوحدة الموالية. و ذلك قصد الوصول إلى تصور صحيح و مشترك لهذا الفرد الافتراضي، حيث أن الكثير من الدراسات السابقة أثبتت بأن أغلب المحكمين يفكرون في الفرد المتوسط عوض الفرد ذو الحد الأدنى من الكفاية. و لتسهيل تكوين نموذج ذهني حول الفرد ذو الحد الأدنى من الكفاية ، طلب من مجموع المحكمين تصور أداء تلميذ من تلاميذ القسم يمثل هذا المستوى.

بعد إتمام هذه الخطوة ، 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؟

بعد الانتهاء من هذه الخطوة، تم جمع تقديرات كل المحكمين في جدول خاص ، وتم بعد ذلك عرض جدول التقديرات على المحكمين حيث سمح لهم بمناقشتها مع بعضهم كما سمح لهم بإجراء التعديلات اللازمة عليها .

كما تم بعد إتمام هذه الخطوة استرداد النماذج الخاصة بتحديد درجات القطع باستخدام طريقة أنجوف وذلك لغرض حساب درجة القطع لكل فقرة وحساب درجة القطع للاختبار ككل. وتكون درجة القطع النهائية هي متوسط درجات القطع لمجموع المحكمين.

#### 2.4.4. طريقة نيدلسكاي:

بعد أسبوعين من تطبيق طريقة أنجوف لتحديد درجة القطع على الاختبار تم الاتصال بنفس المحكمين الاثنا عشر لتحديد درجة القطع باستخدام طريقة نيدلسكاي، حيث طلب من كل محكم أن يحدد عدد المشتتات ( البدائل الخاطئة) التي يمكن أن يستبدها التلميذ ذو الحد الأدنى المقبول من الكفاية في المجال الذي يقيسه الاختبار عندما يفكر في اختيار بديل من البدائل الأربعة. تم بعد ذلك اطلاع المحكمين على تقديراتهم وسمح لهم بتعديلها إذا تطلب الأمر ذلك.

وقد تم استرداد النماذج النهائية الخاصة بتقديرات المحكمين لحساب درجة القطع الناتجة عن طريقة نيدلسكاي، حيث تصبح درجة القطع المحددة بواسطة كل محكم هي مجموع مقلوبات عدد البدائل المتبقية وتكون درجة القطع النهائية، بمثابة الوسط الحسابي لدرجات القطع المقدره من طرف جميع المحكمين.

## الفصل الخامس

### عرض و تفسير النتائج

#### 1.5. التحقق من الفرضية الأولى:

ولغرض التحقق من الفرضية الأولى التي تنص على ما يلي " تختلف قيمة درجة القطع المحددة على الاختبار التحصيلي محكي المرجع في الرياضيات تبعا لاختلاف الطريقة المستخدمة". (أنجوف، نيدلسكاي)، قام الباحث بحساب متوسط تقديرات المحكمين حسب ما هو متبع في طريقتي أنجوف و نيدلسكاي، حيث تم التوصل إلى درجة القطع الخاصة بكل محكم، ثم تم حساب متوسط درجات القطع لمجموع المحكمين وهذا بالنسبة لكلتا الطريقتين. وقد تراوحت الأوساط الحسابية لتقديرات المحكمين لدرجات القطع للاختبار ككل، باستخدام طريقة أنجوف بين (0,66) أو (66%) و (0,74) أو (74%) بمتوسط حسابي مقداره (0,70) أو (70%) وانحراف معياري قيمته (2,57). وإذا ضربنا درجة القطع النهائية (0,70) في عدد فقرات الاختبار العشرين (20) نجدها تساوي (14)، أي أنه يمكن اعتبار التلميذ متمكنا إذا أجاب على أربعة عشرة فقرة إجابة صحيحة كحد أدنى من المجموع الكلي لفقرات الاختبار. ويعتبر غير متمكن إذا أجاب على أقل من أربعة عشرة فقرة إجابة صحيحة من المجموع الكلي لفقرات الاختبار. كما تراوحت الأوساط الحسابية لتقديرات المحكمين لدرجات القطع للاختبار، باستخدام طريقة نيدلسكاي بين (0,50) أو (50%) و (0,61) أو (61%) بمتوسط حسابي قيمته (0,55) أو (55%) وانحراف معياري قيمته (3,86). وإذا ضربنا درجة القطع النهائية (0,55) في عدد فقرات الاختبار (20) نجدها تساوي (11)، أي أنه يمكن اعتبار التلميذ متمكنا إذا أجاب على إحدى عشرة فقرة إجابة صحيحة كحد أدنى من المجموع الكلي لفقرات الاختبار. ويعتبر غير متمكن إذا أجاب على أقل من إحدى عشر إجابة صحيحة من المجموع الكلي لفقرات الاختبار. وقد تم عرض البيانات في الجدول رقم (14):

جدول رقم (14) يبين درجات القطع المحددة بواسطة المحكمين حسب كل طريقة.

درجة القطع		المحكم
نيدلسكاي	أنجوف	
60	68	01
59	71	02
54	69	03
60	71	04
54	74	05
54	71	06
52	70	07
50	74	08
51	68	09
52	66	10
61	67	11
55	72	12
<b>55</b>	<b>70</b>	المتوسط
<b>3.86</b>	<b>2.57</b>	الانحراف المعياري
<b>14.88</b>	<b>6.63</b>	التباين

ولغرض الكشف عن وجود فرق ذات دلالة بين درجة القطع الناتجة عن تطبيق طريقة أنجوف ودرجة القطع الناتجة عن تطبيق طريقة نيدلسكاي، تم استخدام اختبار ولكوكسن (Wilcoxon test). ويعتبر هذا الاختبار كأسلوب إحصائي لابرامتري يستخدم لمعرفة الفرق بين عينتين مرتبطتين للبيانات الفترية [72] (ص: 547)، ويصبح بديلا للنسبة التائية لعينتين مرتبطتين عندما يكون حجم العينة صغيرا (n) تتراوح بين (6 - 25). [73] (ص: 173). و نظرا لأن العينتين مرتبطتين و حجم العينة صغير (ن=12) فقد تم اللجوء إلى استخدام اختبار ولكوكسن وتم حسابه عن طريق استخدام الحزمة الإحصائية للعلوم الاجتماعية ( spss ). وتم التوصل إلى النتائج المبينة في الجدول رقم (15).

جدول رقم (15) يبين نتيجة اختبار ولكوكسن.

مجموع الرتب	متوسط الرتب	حجم العينة	النسبة الحرجة	الدلالة عند مستوى الطرفين
78	6.50	12	3.06	0.01

ويتبين من الجدول رقم (15) بأن النسبة الحرجة 3.06 دالة إحصائياً عند مستوى الدلالة 0.01. وهذه النتيجة تدل على أن درجة القطع الناتجة عن تطبيق طريقة أنجوف تختلف عن درجة القطع الناتجة عن تطبيق طريقة نيدلسكاي، وذلك لصالح طريقة أنجوف، مما يعني أن تطبيق طريقة أنجوف يعطي درجات قطع أعلى من درجات القطع الناتجة عن تطبيق طريقة نيدلسكاي. وبالتالي يتم قبول الفرضية الأولى.

وما يبرر انخفاض درجة القطع الناتجة عن طريقة نيدلسكاي مقارنة بطريقة أنجوف، هو ما أشار إليه Shepard سنة 1984 إلى أن التقديرات الخاصة بالفقرات في طريقة نيدلسكاي على اختبار رباعي البدائل هي عبارة عن أربع قيم ثابتة (0.33، 0.25، 0.5 و 1)، حيث لا تسمح بإعطاء تقديرات تتحصر بين (0.5) و (1)، وتبعاً لذلك لا يميل معظم المحكمين إلى تعيين القيمة الاحتمالية (1) أو مستوى الإتيان 100% ، وبذلك يضطر المحكمون لتعيين القيمة (0.5)، لمعظم الفقرات والذي يؤدي بدوره إلى انخفاض درجة القطع الناتجة عن طريقة نيدلسكاي. [5] (ص: 72).

و يرتبط كذلك الاختلاف في قيمة درجة القطع بين الطريقتين بمدى معرفة المحكم لمحتوى الاختبار. أي أن درجة القطع تتأثر بمعرفة المحكم و تمكنه من محتوى الاختبار، فمن المنطقي أن يضع المحكم درجة قطع مرتفعة للفقرات التي يقدر بأن الإجابة عليها تكون سهلة ، بينما يضع درجات قطع منخفضة للفقرات التي يقدر بأن الإجابة عليها تكون صعبة، وهذا التأثير يكون أكبر عند استخدام طريقة نيدلسكاي، والتي تتطلب من المحكم التفحص الكامل لمختلف البدائل وملاحظة المحكم لمدى التشابه بينها ومنطقية كل بديل كإجابة محتملة مما يجعله أقل ثقة بصعوبة الفقرة لتشابه المشتتات وبالتالي يضع درجة قطع منخفضة للفقرة. كما أن تعرض المحكم لفقرة سهلة يبقي احتمال تقديم تقديرات منخفضة عليها قائماً ، لأن التشابه في المشتتات يزيد من تعقيدها. وفي هذا الصدد أشار كل من برتون و كروص (Burton,1978 Cross,1982) ، أن تقدير صعوبة الفقرة بطريقة نيدلسكاي لا يرجع إلى صعوبة المفهوم (concept) المراد قياسه فحسب بل كذلك إلى تشابه المشتتات. [27] (ص: 17-18). أما فيما يتعلق بطريقة أنجوف، حيث لا يضطر

المحكم إلى تفحص البدائل ويقوم وضع تقدير للمفردة بمجرد قراءته للسؤال، مما قد يؤدي به إلى تقديم تقدير مبالغ فيه إذا وجد بأن السؤال سهل مما ينتج عنه ارتفاع في درجة القطع.

وتتفق نتائج الكثير من الدراسات مع نتائج الدراسة الحالية، أي ارتفاع درجة القطع المحددة بواسطة أنجوف مقارنة بطريقة نيدلسكاي ، وقد تم عرض البعض من هذه الدراسات في الجدول رقم (16).

جدول (16) يبين بعض الدراسات التي توصلت إلى ارتفاع في درجة القطع الناتجة عن طريقة أنجوف مقارنة بطريقة نيدلسكاي.

الدراسة	عدد المفردات	درجة القطع %	أنجوف	نيدلسكاي
		درجة القطع %	درجة القطع %	درجة القطع %
Brennan & Lockwood(1980)	126	67%	84	56%
Livingston & zieky( 1989)	70	51%	36	44%
Chang lei (1996)	09	71%	06	57%
صلاح الدين علام ( 1991 )	60	71.7%	43	66.77%
حجاج غانم أحمد علي(2007)	87	79%	69	54%
الدراسة الحالية	20	70%	14	55%

يتبين من الجدول رقم(16) أن درجات القطع الناتجة عن استخدام طريقة أنجوف في الدراسات الستة كلها مرتفعة مقارنة بطريقة نيدلسكاي . وقد بلغت أعلى قيمة لدرجة القطع الناتجة عن طريقة أنجوف 79 % في دراسة حجاج غانم مقابل 54 % بطريقة نيدلسكاي. أما أدنى قيمة لدرجة القطع في طريقة أنجوف فقد بلغت 51 % في دراسة ليفنجستون و زاكي مقابل 44 % باستخدام طريقة نيدلسكاي. و قد بلغ متوسط درجات القطع الناتجة عن تطبيق طريقة أنجوف في الدراسات الستة 68 % ، كما بلغ متوسط درجات القطع 55 % في طريقة نيدلسكاي.

## 2.5. التحقق من الفرضية الثانية:

ولغرض التحقق من الفرضية الثانية التي تنص على وجود ارتباط موجب بين تقديرات المحكمين للفقرات باستخدام كل من طريقتي أنجوف ونيدلسكاي ومعاملات الصعوبة الفعلية لها، فقد تم إعداد البيانات لهذا الغرض مثل ما هو مبين في الجدول رقم(17)، بعد ذلك تم استخدام معامل



بيرسون لحساب الارتباط. وقد تم استخدام الحزمة الإحصائية للعلوم الاجتماعية (spss) لهذا الغرض.

جدول رقم (17) يبين متوسط تقديرات المحكمين للفقرات ومعاملات الصعوبة الفعلية لها.

رقم الفقرة	معاملات الصعوبة	تقديرات المحكمين		رقم الفقرة	معاملات الصعوبة	تقديرات المحكمين	
		أنجوف	نيدلسكاي			أنجوف	نيدلسكاي
01	0.98	0.88	0.71	11	0.75	0.57	0.43
02	0.72	0.62	0.47	12	0.59	0.69	0.58
03	0.93	0.85	0.71	13	0.39	0.40	0.40
04	0.57	0.65	0.50	14	0.60	0.58	0.58
05	0.77	0.81	0.54	15	0.58	0.65	0.51
06	0.86	0.84	0.71	16	0.44	0.69	0.42
07	0.88	0.78	0.67	17	0.81	0.72	0.67
08	0.94	0.76	0.67	18	0.72	0.74	0.46
09	0.88	0.78	0.54	19	0.66	0.71	0.44
10	0.80	0.68	0.54	20	0.82	0.60	0.43

ولقد بلغت قيمة معامل الارتباط بين تقديرات المحكمين بطريقة أنجوف ومعاملات الصعوبة الفعلية لفقرات الاختبار 0.72 وهي تدل على وجود ارتباط قوي وموجب ، أما قيمة معامل الارتباط بين تقديرات المحكمين بطريقة نيدلسكاي ومعاملات الصعوبة الفعلية للفقرات فقد بلغت 0.69 وهي تدل كذلك على وجود ارتباط قوي وموجب. و تدعم نتائج الارتباط صدق الطريقتين.

ولغرض اختبار دلالة معامل ارتباط بيرسون بين تقديرات المحكمين لكل من طريقة نيدلسكاي وأنجوف ومعاملات الصعوبة الفعلية، تم استخدام معادلة كاندال. وتستخدم هذه المعادلة لاختبار دلالة معامل الارتباط بين متغيرين عندما يكون حجم العينة أقل من خمسين (50) فردا ويستعمل الباحث معامل بيرسون أو معامل سبيرمان . [73] (ص: 82).

وقد بلغت قيمة t المحسوبة 5.89 في طريقة أنجوف كما بلغت قيمة t المحسوبة 5.59 في طريقة نيدلسكاي وهذه القيم دالة إحصائياً عند مستوى الدلالة 0.05 لاختبار الطرفين.

و من الدراسات التي تدعم النتيجة الحالية هو ما توصلت إليه دراسة [18] التي تشير 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. أما معامل الارتباط بين تقديرات المحكمين بطريقة نيدلسكاي و معاملات الصعوبة فقد بلغ 0.77 في حالة معرفة المحكمين لمعاملات الصعوبة و0.52 في حالة عدم تزويدهم بها. و هذه المعاملات جميعها دالة إحصائيا عند مستوى الدلالة 0.05 .

وبناء على ما سبق، و من خلال هذه النتائج يمكننا القول بتحقق الفرضية الثانية للدراسة التي تنص على وجود ارتباط موجب بين تقديرات المحكمين بطريقة أنجوف و نيدلسكاي ومعاملات الصعوبة الفعلية لفقرات الاختبار. وهذا يدل على أهمية الطريقتين في تقديم تقديرات تتناسب طرديا مع معاملات الصعوبة الفعلية مما يجعلنا أكثر ثقة بدرجات القطع الناتجة كما يزيد في القدرة التنبؤية للطريقتين.

### 3.5. التحقق من الفرضية الثالثة:

ولغرض التحقق من الفرضية الثالثة والتي تنص على عدم وجود فرق ذات دلالة في الاتساق الداخلي لتقديرات المحكمين بين طريقتي أنجوف و نيدلسكاي، تم حساب معامل الصعوبة الفعلية لفقرات الاختبار للمفحوصين الذين يمتلكون الحد الأدنى من الكفاية بناء على درجتي القطع الناتجة عن تطبيق طريقتي نيدلسكاي و أنجوف مثل ما هو موضح في الجدول رقم (18). و يشير شانج أن أحسن وسيلة للتحقق من الاتساق الداخلي لأية طريقة هو مقارنة تقديرات المحكمين للمستوى الأدنى المقبول من الأداء مع معاملات الصعوبة الفعلية الخاصة بالمفحوصين الذين يمثلون الحد الأدنى من الكفاية. [27] (ص: 4).  
جدول رقم (18) يبين معاملات الصعوبة الفعلية للمفحوصين ذوي الحد الأدنى من الكفاية.

رقم الفقرة	معاملات الصعوبة (م.ح.أ.ك)		رقم الفقرة	معاملات الصعوبة (م.ح.أ.ك)	
	أنجوف	نيدلسكاي		أنجوف	نيدلسكاي
01	1,00	1,00	11	0,80	0,73
02	0,75	0,55	12	0,55	0,27
03	1,00	0,82	13	0,30	0,27
04	0,50	0,18	14	0,45	0,36
05	0,75	0,36	15	0,65	0,55
06	0,80	0,64	16	0,50	0,10
07	0,80	0,55	17	0,75	0,45
08	0,95	0,91	18	0,65	0,45
09	0,85	1,00	19	0,60	0,36
10	0,85	0,82	20	0,85	0,73

و قد تم في هذا البحث اتباع نفس الخطوات التي اتبعها شانج لاي في حساب الاتساق الداخلي حيث تم حساب التباين بين متوسط تقديرات المحكمين لكل الفقرات الناتجة عن طريقتي أنجوف ونيدلسكاي ومعاملات الصعوبة الفعلية للمفحوصين الذين يمثلون الحد الأدنى من الكفاية (م.ح.أ.ك). حيث تم استخدام العلاقة التالية: [27]

$$\sigma_i^2 = \frac{\sum (x_i - y_i)^2}{n_i}$$

حيث:

$x_i$  : تمثل تقديرات المحكمين الاثني عشر للفقرة  $i$  باستخدام طريقة نيدلسكاي أو أنجوف.

$y_i$  يمثل معامل الصعوبة الفعلية للفقرة  $i$ .

$n_i$  : يمثل عدد الفقرات المقدره بعشرين فقرة.

و يتبين من خلال حساب التباين المستخدم لتقدير الاتساق الداخلي لتقديرات المحكمين لكل طريقة مع معاملات الصعوبة الفعلية للاختبار، بأن قيمة التباين لطريقة أنجوف بلغت 0.028 كما بلغت 0.048 بالنسبة لطريقة نيدلسكاي، وعند حساب الجذر التربيعي لهذه التباينات، نجد بأنه قد بلغ 0.17 بالنسبة لطريقة أنجوف و 0.22 بالنسبة لطريقة نيدلسكاي، مما يعني أن انحراف درجات القطع عن قيم الصعوبة الفعلية بلغ حسب طريقة أنجوف 3.4 أو 17% من مجموع الفقرات و 4.4 أو 22 % من مجموع الفقرات حسب طريقة نيدلسكاي. حيث كلما كانت قيمة الانحراف المعياري لتقديرات المحكمين منخفضة كلما دل ذلك على ارتفاع الاتساق الداخلي. و يلاحظ من خلال النتائج أن قيمة الانحراف المعياري لتقديرات المحكمين عند استخدام طريقة أنجوف كانت منخفضة مقارنة بطريقة نيدلسكاي مما يدل على انخفاض الاتساق الداخلي لطريقة نيدلسكاي مقارنة بطريقة أنجوف.

وما يبرر هذه النتيجة هو أن تقديرات المحكمين بطريقة أنجوف تعطي تقديرات مفتوحة من (0) إلى (1) أي أن المحكم يستطيع أن يعطي أية نسبة تتناسب مع درجة سهولة أو صعوبة الفقرة. أما التقديرات بطريقة نيدلسكاي على اختبار رباعي البدائل فهي عبارة عن قيم احتمالية منفصلة (0.33، 0.25، 0.5، 1) وبالتالي لا يميل أغلب المحكمين إلى اختيار القيمة الاحتمالية (1) ويختارون القيمة الاحتمالية (0.5). فإذا أراد المحكم أن يقدم تقديرا بين (0.5 و 1) فلا

يستطيع، مما ينتج عنه مستوى أدنى من الكفاية أقل أو أكبر مما يريده المحكم. وهذا يدل على أن طبيعة القيم الاحتمالية المنفصلة الخاصة بطريقة نيدلسكاي تجبر المحكم على عدم الاتساق إلى حد معين.

و يلاحظ من خلال الجدول رقم (18) أن أعلى قيمة لتقديرات المحكمين بطريقة نيدلسكاي قد بلغت 0.71 في حين أن نسبة قيم معاملات الصعوبة الفعلية للمفحوص ذو الحد الأدنى من الكفاية حسب طريقة نيدلسكاي و التي تفوق 0.71 فقد بلغت 35 بالمائة. و إضافة إلى ذلك فإن أدنى قيمة للتقدير حسب متطلبات طريقة نيدلسكاي هي 0.25 ، بمعنى أنه لا يسمح للمحكم أن يعطي تقديراً أقل من هذه القيمة في حين نرى من خلال الجدول رقم (18) أن قيمة معامل الصعوبة الفعلي للفقرة رقم (04) هو 0.18 مما يزيد في قيمة التباين الحاصل بين تقديرات المحكمين و معاملات الصعوبة العلية للمفحوص ذو الحد الأدنى من الكفاية.

أما في طريقة أنجوف فيلاحظ أن نسبة تقديرات المحكمين التي تفوق 0.50 قد بلغت 95 بالمائة ، كما بلغت نسبة معاملات الصعوبة الفعلية للمفحوص ذو الحد الأدنى من الكفاية حسب طريقة أنجوف 90 بالمائة مثل ما هو موضح في الجدول رقم(18).

أما المبرر الثاني لانخفاض الاتساق الداخلي لطريقة نيدلسكاي مقارنة بطريقة أنجوف هو خبرة المحكم المرتبطة بميدان الاختبارات . فقد أشارت الكثير من الدراسات أن الخبرة السابقة للمحكم وطبيعة ومدة التدريب الذي يتلقاه حول عملية تحديد درجة القطع تعتبر من العوامل الهامة التي تؤثر على قيمة الاتساق الداخلي . [8]. ومن خلال هذا نستنتج بأن خبرة مجموعة المحكمين في ميدان الاختبارات تسمح لهم بالتنبؤ بأداء المفحوص ذو الحد الأدنى من الكفاية بصفة أسهل عند استخدام طريقة أنجوف، ذلك أن المحكم ( الأستاذ ) في الدراسة الحالية، يستطيع تقدير مستوى صعوبة الفقرة عند قراءة متن السؤال بصورة أفضل من قراءة البدائل، وهذا لتعوده على استخدام هذا السلوك مع مختلف الاختبارات خلال مساره المهني. أما تقدير المحكم لمستوى صعوبة الفقرة من خلال تفحصه لمختلف البدائل فيمثل له نوع من الصعوبة مما يؤدي إلى ضعف الدقة في التقدير.

وما يزيد من ثقتنا في نتيجة الدراسة الحالية فيما يتعلق بالاتساق الداخلي هو ما توصلت إليه دراسة ( Demauro & Powers, 1990 ) [26] التي تشير نتائجها بأن طريقة أنجوف تحظى بدرجة عالية نسبياً من الثبات المنطقي، حيث يوجد اتساق بين تقديرات المحكمين والأداء الفعلي

للطلاب قدره 0.71. ولكن يشير الباحثان إلى أن نتائجه محكومة بالاختبار المستخدم والعينة المستخدمة. بالإضافة إلى نتائج دراسة فندر لندن التي أظهرت انخفاض الاتساق الداخلي لإجراءات نيدلسكاي مقارنة بطريقة أنجوف.

و لاختبار الفرضية الصفرية التي تنص على عدم وجود فرق في الاتساق الداخلي لتقديرات المحكمين في طريقتي أنجوف ونيدلسكاي تمت المقارنة بين قيمتي التباين للطريقتين عن طريق قسمة التباين الأكبر على التباين الأصغر للحصول على النسبة الفائية، وقد بلغت قيمة F المحسوبة (1.71)

وهي أصغر من القيمة المجدولة (2.15) عند مستوى الدلالة 0.05. و عليه نقبل الفرضية الصفرية مما يعني أنه لا يوجد فرق في الاتساق الداخلي للتقديرات المحكمين بين الطريقتين . و بالرغم من أن النتيجة التي توصل إليها شانج لاي [27] هي عدم وجود فرق في الاتساق الداخلي لطريقتي أنجوف و نيدلسكاي إلا أنه توصل إلى أن طريقة نيدلسكاي تتميز باتساق داخلي أكبر. غير أن شانج استخدم المفحوص ذو المستوى المتوسط من الأداء عوض المستوى الأدنى.

#### 4.5. التحقق من الفرضية الرابعة:

ولغرض التحقق من قدرة الطريقتين (أنجوف ونيدلسكاي) على التنبؤ بتصنيف مستوى تحصيل التلاميذ في مادة الرياضيات في شهادة التعليم المتوسط، تم الحصول على نتائج أفراد العينة من المؤسسات التي ينتمون إليها لحساب معامل كبا (Kappa Coefficient) ومعامل صدق القرار لكل من درجتي القطع الناتجتين عن تطبيق طريقتي أنجوف و نيدلسكاي، وذلك بالاعتماد على ثلاث مستويات للتمكن (55،60،70). و يعبر معامل صدق القرار عن نسبة عدد الأفراد الذين يتم تصنيفهم في المجموعة نفسها في مرتي التطبيق، كما يعتبر معامل كبا كأسلوب إحصائي يأخذ أخطاء التصنيف بعين الاعتبار ويعمل على تحديد نسبة اتساق التصنيف وفقا لدرجة قطع معينة ويصح هذه النسب من أخطاء التصنيف التي تعزى للصدفة. وتجدر الإشارة إلى أن قيم معامل كبا تتراوح بين (-1 و +1) و كلما اقتربت القيمة من (+1) كلما دل ذلك على أن النسب الهامشية تقترب من بعضها البعض، وكلما اقتربت القيمة من (-1) كلما دل ذلك على الاختلاف الكبير في تصنيف الأفراد في مرتي التطبيق. [20] (ص: 199). وبيين الجدول رقم (19) قيم معامل كبا للطريقتين مع نتائج الرياضيات لشهادة التعليم المتوسط.

جدول (19): يبين قيم معامل صدق القرار وكابا لتصنيف التلاميذ حسب درجات القطع الناتجة عن تقديرات المحكمين للطريقتين ونتائج الرياضيات في شهادة التعليم المتوسط.

الطريقة	درجة القطع	مستويات التمكن	معامل دقة القرار	معامل كابا
أنجوف	0.70	0.70	0.83	0.65
		0.60	0.88	0.73
		0.55	0.86	0.67
نيدلسكاي	0.55	0.70	0.65	0.28
		0.60	0.83	0.49
		0.55	0.90	0.67

يتبين من الجدول رقم (19) أن قيم معامل كابا لنموذج أنجوف تراوحت بين 0.65 و0.73 وهي مؤشرات موجبة للتنبؤ بمستوى التحصيل المدرسي للتلاميذ في مادة الرياضيات، غير أنه يلاحظ بأن أعلى قيمة لمعامل كابا كانت عند مستوى التمكن 0.60 حيث بلغ معامل الدقة 0.88 وهي أقل من درجة القطع الناتجة عن طريقة أنجوف والتي بلغت 0.70. و بما أن طريقة أنجوف تعطي درجات قطع مرتفعة ما يزيد في احتمال الوقوع في الخطأ السالب.

أما فيما يتعلق بطريقة نيدلسكاي فقد تراوحت قيم معامل كابا بين 0.28 و0.67 وكانت أعلى قيمة لمعامل كابا 0.67 ومعامل صدق القرار 0.90 عند مستوى التمكن 0.55 وهذا مؤشر ايجابي على قدرة النموذج على التنبؤ بتصنيف مستوى التحصيل الدراسي في مادة الرياضيات. وما يلاحظ في نتائج طريقة نيدلسكاي هو ارتفاع قيمة معامل كابا كلما اقتربنا من مستوى التمكن المطابق لدرجة القطع.

و من خلال عرضنا لنتائج كابا عند استخدام الطريقتين، فإننا نقول بتحقق الفرضية الرابعة التي تنص على قدرة الطريقتين في التنبؤ بمستوى تصنيف التلاميذ في مادة الرياضيات في شهادة التعليم المتوسط.

وإذا ما قارنا قيم معاملات كابا لتصنيف التلاميذ عند مستويات التمكن المقابلة لدرجاتي القطع الناتجة عن الطريقتين في هذه الدراسة مع القيم التي توصلت إليها دراسة الشريم و سواملة (2006) [18] نجد بأن هذه الأخيرة قد توصلت إلى قيم أعلى عند مستويات التمكن المقابلة لدرجات القطع الناتجة عن الطريقتين ، حيث بلغت قيمة معامل كابا 0.86 في طريقة أنجوف

و 0.87 في طريقة نيدلسكاي . ويرد تدني قيم معاملات كبا للدراسة الحالية مقارنة مع الدراسة المذكورة إلى مجموعة من العوامل نذكر منها مايلي :

أ- اختلاف طبيعة الاختبارين(اختبار الدراسة الحالية و اختبار شهادة التعليم المتوسط ) فالأول هو اختبار ذات الاختيار المتعدد و يقيس المستويات المعرفية الثلاثة الأولى (التذكر ، الفهم ، التطبيق) أما الثاني فقد يقيس المستويات المعرفية التي لا تقيسها أداة الدراسة الحالية.

ب- أداة الدراسة الحالية صممت لغرض قياس مدى تمكن التلاميذ من وحدة معينة من برنامج تلاميذ السنة الرابعة متوسط وهي وحدة المعالم، بينما الاختبار المصمم لشهادة التعليم المتوسط فيشمل وحدات مختلفة. وعليه يمكن القول بأنه كلما كان الاختبار أشمل لنطاق المحتوى كلما كان زاد الصدق التنبؤي لطريقتي أنجوف و نيدلسكاي.

## خلاصة النتائج

هدفت هذه الدراسة إلى المقارنة بين طريقتي أنجوف ونيدلسكاي لتقدير درجة القطع على اختبار محكي المرجع في الرياضيات لمستوى السنة الرابعة متوسط و معرفة مدى توفر كل طريقة على المعطيات الدالة على صدقها.

ولغرض تحقيق هدف الدراسة تم بناء اختبار محكي المرجع في وحدة المعالم، حيث تكون الاختبار من عشرين فقرة من الاختيار من متعدد لكل منها أربعة بدائل . وتضمنت الدراسة نوعين من العينات. أما الأولى فعينة المحكمين، وقد تكونت من اثنا عشرة محكما قاموا بتقدير فقرات الاختبار موضوع البحث بطريقتي أنجوف ونيدلسكاي للوصول إلى تحديد درجة قطع على الاختبار، وقد تم اختيار مجموعة المحكمين من بين الأساتذة ذوي الخبرة والذين يقومون بتدريس مستوى السنة الرابعة متوسط خلال السنة الدراسية (2010-2011). أما النوع الثاني فهو عينة التلاميذ، حيث تكونت من 173 تلميذ من تلاميذ السنة الرابعة متوسط وقد طبق عليهم الاختبار خلال الفصل الثالث من السنة الدراسية (2010-2011).

وقد نتج عن استخدام الطريقتين درجتى قطع مختلفتين حيث بلغت درجة القطع الناتجة عن تطبيق إجراءات أنجوف (0,70) أو (70%) وكانت أعلى من طريقة القطع الناتجة عن تطبيق إجراءات نيدلسكاي والتي بلغت (0,55) أو (50%).

وقد أشارت نتائج التحليل اللابارامتري (wilcoxon test) أنه يوجد فرق دال إحصائيا عند مستوى الدلالة 0,01 بين درجتى القطع الناتجة عن الطريقتين، وجاءت نتيجة هذه الدراسة لتضاف إلى نتائج العديد من الدراسات التي توصلت إلى أن طريقة أنجوف تعطي درجة قطع أعلى من طريقة نيدلسكاي و من بين هذه الدراسات (Chang Lei,1996) [27] (الشريم وسوالمه، 2006) [18] (حجاج غانم، 2007). [5].



أما نتيجة التساؤل الثاني حول وجود ارتباط بين تقديرات المحكمين الناتجة عن كل طريقة ومعاملات الصعوبة الفعلية لفقرات الاختبار، فقد توصلت الدراسة إلى وجود ارتباط قوي وموجب بين تقديرات المحكمين بطريقة أنجوف ومعاملات الصعوبة الفعلية، حيث بلغ معامل الارتباط 0,72 وهو دال إحصائياً، كما بلغ معامل الارتباط بين تقديرات المحكمين بطريقة نيدلسكاي ومعاملات الصعوبة الفعلية 0,69 وهو كذلك دال إحصائياً، و تعتبر هذه النتيجة دليلاً داخلياً للصدق.

وللتأكد من وجود فرق في الاتساق الداخلي بين طريقتي أنجوف ونيدلسكاي. فقد دلت النتائج على أنه لا يوجد فرق في الاتساق الداخلي بين الطريقتين، غير أن طريقة أنجوف كانت تتميز باتساق داخلي أكبر من طريقة نيدلسكاي، و يرتبط ارتفاع الاتساق الداخلي بانخفاض الانحراف المعياري مما يؤدي إلى زيادة الثقة في درجة القطع الناتجة. [66] و قد كانت قيمة الانحراف المعياري أقل في طريقة أنجوف مقارنة بطريقة نيدلسكاي مما يعني تفوق طريقة أنجوف من حيث الاتساق الداخلي. ويرجع هذا التفوق إلى عدة عوامل من بينها مايلي:

أ- طبيعة و متطلبات كل طريقة: فطريقة نيدلسكاي تتطلب تقديم تقديرات حول الفقرات بناء على قيم احتمالية منفصلة (0.25، 0.33، 0.5، 1) و يحاول في أغلب الأحيان المحكمون تجنب اختيار النسبة الكاملة 100%، و يميلون بالتالي إلى اختيار نسبة 50 %، مما يزيد في التباين بين التقديرات و بالتالي انخفاض الاتساق الداخلي. بالإضافة إلى أن طبيعة الطريقة (نيدلسكاي) تتطلب من المحكم فحص البدائل ، حيث يزيد تشابهها في تعقيد السؤال مما يؤدي بالمحكم إلى تقديم تقديرات منخفضة على الفقرة رغم سهولتها.

أما طريقة أنجوف فترتكز على قيم احتمالية متصلة من الصفر إلى الواحد، و هذا يساعد المحكم في تقديم التقديرات التي يراها مناسبة. كما أن هذه التقديرات تتم مباشرة بعد قراءة السؤال دون الخوض في البدائل مما يساعد المحكم إيجاباً بوضع تقديرات بناء على خبرته في ميدان الاختبارات و معرفته لمستوى التلاميذ دون تردد لعدم تأثر تفكيره بمدى التشابه بين المشتتات.

ب- خبرة المحكم في ميدان الاختبارات: تتماشى خبرة المحكم في ميدان الاختبارات مع طريقة أنجوف أكثر من طريقة نيدلسكاي. فممارسات الأستاذ لعمليات التقويم في ميدان التربية أكسبته خبرة جعلته يستطيع التنبؤ بأداء الفرد ذو الحد الأدنى من الكفاية على الفقرة عند استخدام طريقة أنجوف و ذلك من خلال قراءته للسؤال و تقديم التقدير المناسب مباشرة. على عكس طريقة نيدلسكاي التي تجعل المحكم يركز على صعوبة أو سهولة البدائل أكثر من صعوبة أو سهولة السؤال.

و لكن رغم تميز طريقة أنجوف باتساق داخلي أكبر مقارنة بطريقة نيدلسكاي، إلا النتائج تبين بأن الطريقتان تتمتعان باتساق داخلي مقبول مما يعتبر دليلاً يدعم صدق الطريقتين. غير أن هذه النتيجة تبقى مرتبطة بطبيعة الاختبار وطبيعة المفحوصين .

أما فيما يخص أهمية الطريقتين في التنبؤ بتصنيف مستويات التلاميذ في اختبار مادة الرياضيات في شهادة التعليم المتوسط فقد بينت النتائج أن أعلى قيمة لمعامل كابا عند استخدام المستوى 55 كأدنى مستوى للتمكن فبلغت 0.67 في الطريقتين غير أن معامل صدق القرار كان أعلى في طريقة نيدلسكاي حيث بلغ 0.90، أما قيمة معامل كابا عند مستوى 60 كأدنى مستوى للتمكن فقد بلغ 0.73 حسب طريقة أنجوف. و تعتبر النتائج المتوصل إليها مؤشرات موجبة لقدرة النموذجين على التنبؤ بتصنيف مستوى التلاميذ في شهادة التعليم المتوسط .

و في الأخير يمكن القول بأنه بالرغم من وجود فرق دال إحصائياً بين درجتي القطع الناتجتين عن تطبيق إجراءات الطريقتين، إلا أن نتائج الدراسة الحالية بينت أنه يوجد ارتباط قوي و موجب لتقديرات المحكمين عند استخدام الطريقتين و معاملات الصعوبة الفعلية لفقرات الاختبار. كما بينت النتائج أن الطريقتان تتمتعان باتساق داخلي كبير مما يعتبر دليلاً كذلك على الصدق الداخلي للطريقتين. هذا بالإضافة إلى أهمية الطريقتين في التنبؤ بتصنيف مستوى تحصيل التلاميذ. كما تبرر هذه النتائج الاستعمال الواسع لهاتين الطريقتين في المجالات المختلفة رغم تاريخهما الطويل. غير أنه لا يمكن من خلال النتائج المتوصل إليها من خلال مقارنة الطريقتين، تفضيل طريقة على أخرى ، بالرغم من اختلاف درجتي قطع الناتجة ، و ذلك لسبب حصولهما على نفس المزايا فيما يتعلق بالأدلة الداخلية للصدق (validity evidence). بالإضافة إلى غياب محك خارجي يؤكد فعالية طريقة على أخرى. و هذا ما أشار إليه كاين (Kane,2001) أن الدراسات التي تناولت خصائص مختلف طرق تحديد درجات القطع توصلت إلى نقاط القوة و الضعف لمختلف الطرق، غير أنها لم تستطع تفضيل طريقة على أخرى بصفة قطعية ، و ذلك يرجع لغياب محك خارجي يمكن من معرفة كيف يجب أن تكون درجة القطع. لكن رغم عدم وجود طريقة مفضلة بصفة واضحة ، إلا أن يوجد اتفاق بأن تحديد درجة القطع يتم بطريقة متأنية و منتظمة. [10].

و يشير داونين أن لا طريقة من الطرق المستعملة في تحديد درجة القطع يمكن اعتبارها أصح من الطرق الأخرى، فمهمة المحكمين من ذوي الخبرة ليس اكتشاف درجة القطع الصحيحة (True) ولكن تحديد المستوى الأدنى المقبول من الأداء و الذي يجب أن يظهره المفحوص لاجتياز الاختبار. [51].

## خاتمة

لقد أصبح موضوع تحديد درجة القطع أكثر من ضروري في مختلف الميادين لاسيما في الميدان التربوي، حيث أصبحت الكثير من الدول الأجنبية تعتمد اعتمادا كليا على طرق تحديد درجة القطع في اتخاذ قراراتها ، غير أن التطبيق الميداني للدراسة الحالية كشف لنا بأن أغلب المربين ليس لهم اطلاع على هذه الطرق ، ذلك لكون درجة القطع المعتمدة ثابتة مع كل الاختيارات ولجميع المواد. فعلى الرغم من الانتقادات التي تعرضت لها طرق تحديد درجة القطع لتميزها بالذاتية و التناقض إلا أنها ليست أسوأ من درجة القطع الثابتة ( fixed cut score ) حيث أن التناقض أو الاختلاف بين درجات القطع الناتجة عن مختلف الطرق لا يعني وجود درجة قطع صحيحة وأخرى خاطئة. كما أن غياب محك خارجي لا يسمح بصفة قطعية تقرير أفضلية طريقة على أخرى، و لكن يبقى احتمال اختيار طريقة معينة يعتمد على ايجابيات و حدود الطريقة المختارة. هذا بالإضافة إلى توفير الشروط التي يتطلبها تطبيق مختلف الطرق و التي تم التطرق إليها في الجانب النظري.

و قد أفرز البحث الحالي مجموعة من الاقتراحات نلخصها فيما يلي:

- دعم طريقة أنجوف بإطار مرجعي لتقليل نسبة الخطأ وذلك بتزويد المشاركين في تحديد درجة القطع بمعطيات حول معاملات الصعوبة الفعلية لل فقرات. فتقديرات المحكمين في طريقة نيدلسكاي تعتمد على إطار مرجعي مشترك يتمثل في كون كل محكم يصدر عدة قرارات حسب عدد المشتتات وبالتالي فإن نسبة الخطأ في قرار واحد لا يكون له نفس التأثير مقارنة بطريقة أنجوف التي يطلب فيها من الحكم إصدار قرار واحد على كل فقرة. و في هذا الصدد يشير ( Abott,2003 ) أن درجة القطع المحددة بواسطة طريقة أنجوف صعب اجتيازها لارتفاعها مما يجعلها تحتاج إلى بيانات معيارية و امبريقية يسترشد بها المحكمون في تقديراتهم لفقرات الاختبار مما قد يساعد في تحديد درجة قطع أكثر موضوعية. [5].

- تدريب المفتشين والمربين على تصميم اختبارات محكية المرجع خاصة بكل وحدة دراسية. بالإضافة إلى اختيار محكمين أكفاء ، من أهل الاختصاص، وتدريبهم التدريب الجيد حول تحديد درجة القطع لكل اختبار باستخدام طريقة أنجوف أو نيدلسكاي مع تقديم المعلومات الكافية حول

إجراءات كل طريقة وضمان فهمهم للفرد ذو الحد الأدنى المقبول من الكفاية ، فأغلب أساتذة التعليم المتوسط لا يفرقون بينه وبين الفرد المتوسط، وهذا ما لاحظناه خلال تطبيق إجراءات الدراسة. كما نتمنى أن تخصص الكثير من البحوث حول صحة إدراك المعلمين و الأساتذة لمفهوم الفرد ذو الحد الأدنى من الكفاية.

و نتمنى في الأخير أن يتم إدراج تدريس مختلف طرق بتحديد درجة القطع ضمن البرامج التربوية حتى ينتفع بها المربي الجزائري، ويعتمدها في عمليات التقويم المستمرة. وحتى لا تبقى حكرا على المختصين في القياس النفسي و التربوي . كما نتمنى أن ننقل من مرحلة السداجة ، حيث تستخدم الطرق التقليدية في عمليات التصنيف و القبول في مختلف المجالات و خاصة المجال التربوي. والتي انتهت مع نهاية الستينيات من القرن الماضي في الغرب، إلى مرحلة التقبل الحقيقي للطرق و تطبيقها و اعتمادها في اتخاذ القرارات حتى يجد أصحاب القرار الأدلة الكافية للدفاع عن القرارات المفروزة.

## ملحق رقم (1) الكفايات و الأهداف السلوكية المتفرعة عنها و المستوى المعرفي لكل هدف سلوكي

المستوى المعرفي	الأهداف السلوكية	الكفايات
- تذكر. - تطبيق.	- أن يتعرف المتعلم على العلاقة التي تسمح بحساب المسافة بين نقطتين في معلم متعامد و متجانس. - أن يطبق المتعلم قاعدة حساب المسافة بين نقطتين. إحداثياتهما معلومة.	يحسب المسافة بين نقطتين في معلم متعامد و متجانس.
- تطبيق.	- أن يحدد المتعلم إحداثيتي نقطتين علمت المسافة بينهما.	يعرف خصائص الشعاعين المتساويين.
- تذكر. - تذكر.	- أن يتعرف المتعلم على المعلم المتعامد و المتجانس. - أن يميز المتعلم بين مختلف المعالم.	يعرف أنواع المعالم.
- فهم.	- أن يعرف المتعلم القاعدة التي تعطي إحداثيتي منتصف قطعة.	يحسب إحداثيتي منتصف قطعة.
- فهم. - فهم. - فهم. - فهم.	- أن يعرف المتعلم تعليم نقطة في مستو منسوب إلى معلم. - أن يقرأ المتعلم إحداثيتي نقطة من مستو مزود بمعلم. - أن يحدد المتعلم التمثيل الذي يمثل نقطة علمت إحداثياتها. - أن يتعرف المتعلم على إحداثيتي نقطة في معلم متعامد و متجانس.	يعرف إحداثيتي نقطة.

## ملحق رقم (1) الكفايات و الأهداف السلوكية المتفرعة عنها و المستوى المعرفي لكل هدف سلوكي

المستوى المعرفي	الأهداف السلوكية	الكفايات
- تذكر. - تطبيق.	- أن يتعرف المتعلم على العلاقة التي تسمح بحساب المسافة بين نقطتين في معلم متعامد و متجانس. - أن يطبق المتعلم قاعدة حساب المسافة بين نقطتين. إحدائيهما معلومة.	يحسب المسافة بين نقطتين في معلم متعامد و متجانس.
- تطبيق.	- أن يحدد المتعلم إحدائتي نقطتين علمت المسافة بينهما.	يعرف خصائص الشعاعين المتساويين.
- تذكر. - تذكر.	- أن يتعرف المتعلم على المعلم المتعامد و المتجانس. - أن يميز المتعلم بين مختلف المعالم.	يعرف أنواع المعالم.
- فهم.	- أن يعرف المتعلم القاعدة التي تعطي إحدائتي منتصف قطعة.	يحسب إحدائتي منتصف قطعة.
- فهم. - فهم. - فهم. - فهم.	- أن يعرف المتعلم تعليم نقطة في مستو منسوب إلى معلم. - أن يقرأ المتعلم إحدائتي نقطة من مستو مزود بمعلم. - أن يحدد المتعلم التمثيل الذي يمثل نقطة علمت إحدائياتها. - أن يتعرف المتعلم على إحدائتي نقطة في معلم متعامد و متجانس.	يعرف إحدائتا نقطة.

## مواصفات الاختبار

ملحق رقم (2)

الرقم التسلسلي	01	02	03
الهدف السلوكي	أن يتعرف المتعلم على المعلم المتعامد و المتجانس.	أن يميز المتعلم بين مختلف المعالم.	أن يعرف المتعلم تعليم نقطة في مستو منسوب إلى معلم.
خصائص المثير	مدخل المثير	من بين المعالم التالية:	إليك أربع أجوبة لتلاميذ طرح عليهم السؤال التالي: علم النقطة؟.
	وصف المثير	تعطى للتلميذ أربعة أشكال لمعلم مختلفة من بينها شكل واحد يمثل المعلم المتعامد و المتجانس و يطلب من التلميذ تحديد المعلم المتعامد و المتجانس.	تعطى للتلميذ أربع أشكال لمعلم مختلفة، علمت عليهم نقطة معطاة. و يطلب منه اختيار الشكل الذي علمت عليه النقطة المعطاة بصفة صحيحة.
خصائص الاستجابة	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، واحد منها يمثل الإجابة الصحيحة (المعلم المتعامد و المتجانس) و البدائل الأخرى تمثل ثلاث أنواع أخرى من المعالم و هي: (المتجانس، الكيفي و المتعامد).	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا واحد منها يمثل البديل الصحيح و البدائل الأخرى عبارة عن أسماء المعالم المختلفة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا واحد منها يمثل الإجابة الصحيحة. (شكل علم بصفة صحيحة) و البدائل الأخرى عبارة عن نقطة معلمة في معلم في وضعيات مختلفة غير صحيحة .

06	05	04	الرقم التسلسلي	
أن يتعرف المتعلم على إحداثيتي نقطة في معلم متعامد و متجانس.	أن يقرأ المتعلم إحداثيتي نقطة من مستو مزود بمعلم.	أن يحدد المتعلم التمثيل الذي يمثل نقطا علمت إحداثياتها.	الهدف السلوكي	
إحداثيات النقطة A في المعلم المتعامد و المتجانس هما:	ما هما إحداثيات النقطة E .	$(O, \vec{i}, \vec{j})$ المستوي مزود بمعلم من بين التمثيلات الآتية:	مدخل المثير	خصائص المثير
يعطى للتلميذ شكل لمعلم متعامد و متجانس علمت عليه النقطة A و يطلب منه إيجاد إحداثياتها.	يعطى للتلميذ بيان عبارة عن معلم علمت فيه نقطة و يطلب منه التعرف على إحداثياتها.	تعطى للتلميذ أربع تمثيلات لأربع نقط و يطلب منه تحديد التمثيل الصحيح الذي يمثل النقط الأربعة.	وصف المثير	
أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، كل بديل يمثل إحداثية مختلفة و واحد منها يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، واحد منها يمثل الإجابة الصحيحة و البدائل الأخرى عبارة عن إحداثيات مختلفة لنفس النقطة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، واحد منها يمثل الإجابة الصحيحة و البدائل الأخرى عبارة عن تمثيلات تحوي أربع نقط معلمة في وضعيات مختلفة.	خصائص الاستجابة	



09	08	07	الرقم التسلسلي	
أن يحدد المتعلم النقطة التي تمثل نهاية شعاع، مبدؤه مبدأ المعلم و إحدائياته معلومة.	أن يعين المتعلم شعاعا علمت إحدائتيه.	أن يعرف المتعلم العلاقة بين إحدائتي نقطة و إحدائتي شعاع مبدؤه مبدأ المعلم و نهايته تلك النقطة.	الهدف السلوكي	
$(O, \vec{I}, \vec{J})$ المستوى منسوب إلى معلم	عين الشعاع الذي إحدائياته هما:	نقطة من المستوى المزود بمعلم $F$ بحيث $(O, \vec{I}, \vec{J}) . F (2.3)$	مدخل المثير	خصائص المثير
يعطى للتلميذ معلم له أربعة أشعة و تعطى له إحدائيات نقطة و يطلب منه تحديد النقطة التي تمثل نهاية النقطة.	يعطى للتلميذ أربعة أشعة في معلم متعامد و متجانس و يطلب منه التعرف على الشعاع الذي إحدائياته هما إحدائيتا تلك النقطة.	يعطى للتلميذ نقطة علمت إحدائياتها، و يطلب منه إيجاد إحدائيات شعاع مبدؤه مبدأ المعلم و نهايته نقطة علمت إحدائياتها.	وصف المثير	
أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، تمثل أربع نقاط مختلفة، واحد منها يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، و لها نفس المبدأ و نهايات مختلفة، واحد منها يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، كل بديل يمثل شعاع بإحدائية مختلفة و بديل واحد يمثل الإجابة الصحيحة.	خصائص الاستجابة	

18	15	20	الرقم التسلسلي	
أن يتعرف المتعلم على العلاقة التي تسمح بحساب المسافة بين نقطتين في معلم متعامد و متجانس.	أن يحدد المتعلم إحداثيتي شعاع في مستو مزود بمعلم متعامد و متجانس.	أن يحدد المتعلم إحداثيتي نقطتين علمت المسافة بينهما.	الهدف السلوكي	
نقطتان من المستوى $W, V$	إحداثيتا الشعاع $\vec{T}$ في المعلم الآتي هما:	المسافة بين نقطتين هي:	مدخل المثير	خصائص المثير
يعطى للتلميذ نقطتان من المستوى. ويطلب منه معرفة العلاقة التي تسمح بحساب المسافة $V, W$ .	يعطى للتلميذ إحداثيتا أي شعاع في معلم متعامد و متجانس و يطلب منه تحديد إحداثيتا الشعاع في المعلم.	يعطى للتلميذ المسافة بين نقطتين و يطلب منه تحديد إحداثيتي النقطتين اللتين تكون المسافة بينهما هي المسافة المعطاة.	وصف المثير	
أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، كل بديل يمثل علاقة معينة وواحد منها يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، يمثل كل واحد منها إحداثيتين و بديل واحد فقط يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، و البدائل هنا تحتوي أربع ثنائيات نقطية إحداثيات كل نقطة منها معلومة و احد منها فقط يمثل الإجابة الصحيحة.	خصائص الاستجابة	

14	12	19	الرقم التسلسلي	
أن يحدد المتعلم إحداثيتي شعاع من بيان.	أن يقرأ المتعلم إحداثيتي شعاع معرف بمبدأ ونهاية ممثله.	أن يطبق المتعلم قاعدة حساب المسافة بين نقطتين إحداثياتهما معلومة.	الهدف السلوكي	
ما هما إحداثيتا الشعاع (يعطى أي شعاع) في البيان التالي:	هل الشعاع (يعطى أي شعاع) المرسوم في البيان التالي هو:	لتكن النقطتين (أي نقطتين) في المستوي المنسوب إلى معلم متعامد و متجانسو المعرفتين كما يلي:	مدخل المثير	خصائص المثير
يعطى للتلميذ شعاع مرسوم في بيان ويطلب منه تحديد إحداثيتا هذا الشعاع.	يعطى للتلميذ شعاع مرسوم في بيان ويطلب منه تحديد هذا الشعاع.	يعطى للتلميذ طرفي قطعة مستقيمة معرفتين بإحداثيتيهما و يطلب منه تحديد طول القطعة المستقيمة (يقوم التلميذ بتطبيق العلاقة لحساب المسافة)	وصف المثير	
أربع بدائل مرتبة عشوائيا ومرقمة هجائيا، و البدائل هنا تحتوي على أربع إحداثيات مختلفة. بديل واحد فقط يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا كل بديل منها يمثل شعاعا بإحداثيتين . بديل واحد يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، و البدائل هنا تمثل أربع أطوال مختلفة . بديل واحد يمثل الإجابة الصحيحة.	خصائص الاستجابة	

10	17	الرقم التسلسلي	
أن يقرأ المتعلم إحدائتي شعاع في مستو مزود بمعلم متعامد ومتجانس.	أن يعرف المتعلم القاعدة التي تعطي إحدائتي منتصف قطعة.	الهدف السلوكي	
من البيان، هل هل إحدائتا الشعاع (يعطى أي شعاع) هما:	ليكن في المستوي المنسوب لمعلم النقطتين (تعطى أي نقطتين)	مدخل المثير	خصائص المثير
يعطى للتلميذ شعاع في بيان و يطلب منه إيجاد إحدائتا هذا الشعاع.	تعطى للتلميذ نقطتان مرفقتان بإحدائتيهما و يطلب منه التعرف على العلاقة التي تسمح بحساب إحدائتا منتصف قطعة.	وصف المثير	
أربع بدائل مرتبة عشوائيا و مرقمة هجائيا. كل بديل يمثل إحدائية مختلفة لنفس الشعاع. بديل واحد فقط يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا. كل بديل يمثل علاقة معينة لحساب منتصف قطعة. بديل واحد يمثل الإجابة الصحيحة.	خصائص الاستجابة	

13	16	11	الرقم التسلسلي	
أن يقرأ المتعلم إحدائيتي شعاع مزود بمعلم.	أن يحسب المتعلم إحدائيتي شعاع بمعرفة إحدائيتي مبدأ و نهاية ممثله.	أن يحدد المتعلم شعاعا أعطيت إحدائيتا مبدئه و إحدائيتا نهاية ممثله.	الهدف السلوكي	
مثلث متقايس الأضلاع: A B C	ليكن في مستوى منسوب إلى معلم متعامد و متجانس النقطتين:	حدد الشعاع المعين بنقطتين .	مدخل المثير	خصائص المثير
يعطى للتلميذ مثلث متقايس الأضلاع و يطلب منه تحديد إحدائيتا الشعاع (ABC) في المعلم المتجانس. $\overline{AB}$	يعطى للتلميذ إحدائيتي مبدأ و نهاية ممثل شعاع و يطلب منه حساب إحدائيتي هذا الشعاع.	يعطى للتلميذ إحدائيتي نقطتين و شكل معلم متعامد و متجانس به أربعة أشعة و يطلب منه تحديد الشعاع المعين بهاتين النقطتين.	وصف المثير	
أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، و البدائل هنا عبارة عن أربع احداثيات لنفس الشعاع و بديل فقط يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، كل بديل يمثل زوج مختلف من الإحداثيات و بديل واحد فقط يمثل الإجابة الصحيحة.	أربع بدائل مرتبة عشوائيا و مرقمة هجائيا، كل بديل يمثل شعاع معين، و بديل واحد فقط يمثل الإجابة الصحيحة.	خصائص الاستجابة	

## الاختبار

ملحق رقم (3)

### تعليمات الاختبار

يهدف هذا الاختبار لقياس مدى تمكنك من الكفايات الأساسية في وحدة المعالم. ويشتمل الاختبار على عشرين (20) سؤالاً من نوع الاختيار من متعدد. يوجد لكل سؤال أربعة بدائل واحد منها فقط صحيح.

يطلب منك أن تقرأ كل سؤال بعناية، و أن تختار الإجابة الصحيحة من بين الإجابات المعطاة، ثم تضع العلامة (X) تحت الحرف المناظر لهذه الإجابة في الخانة المعدة لذلك على ورقة الإجابة.

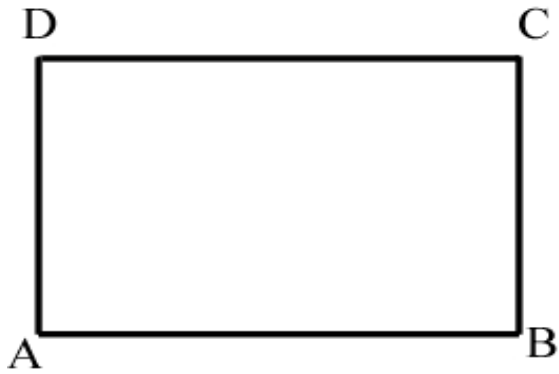
قبل تسليم ورقة الإجابة، حاول أن تتأكد من وضعك للعلامة التي تشير إلى إجابتك عن كل سؤال في الخانة المعدة لذلك على ورقة الإجابة المرفقة.

## الاختبار

1) من بين المعالم التالية، حدد المعلم المتعامد والمتجانس .



2)  $ABCD$  مستطيل. المعلم  $(A, B, D)$  هو:



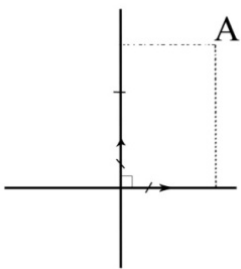
(أ) معلم متعامد ومتجانس

(ب) متجانس

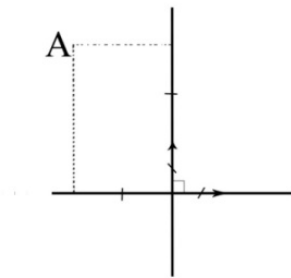
(ج) متعامد

(د) كافي

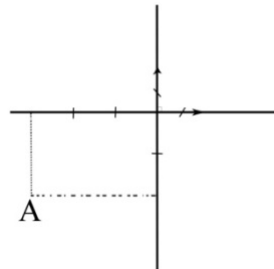
3) إليك أربعة أجوبة لتلاميذ طرح عليهم السؤال التالي : علم النقطة  $A (-2, 3)$  ما هي الإجابة الصحيحة؟



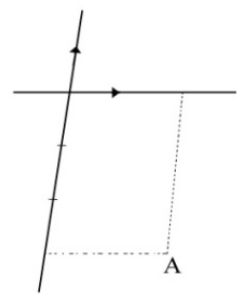
(د) إجابة مريم



(ج) إجابة مراد



(ب) إجابة زهرة



(أ) إجابة علي

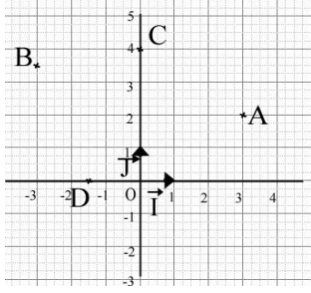
4) المستوي مزود بمعلم  $(O, \vec{i}, \vec{j})$ ، من بين التمثيلات أدناه أي التمثيل الذي يمثل النقط الآتية:

$$D\left(-\frac{3}{2}, 0\right) \text{ (د)}$$

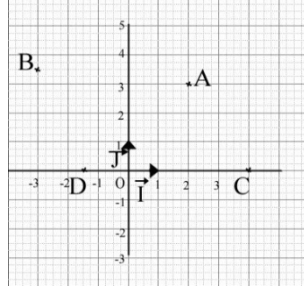
$$C(0,4) \text{ (ج)}$$

$$B\left(-3, \frac{5}{2}\right) \text{ (ب)}$$

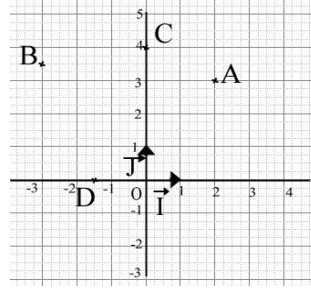
$$A(2,3) \text{ (أ)}$$



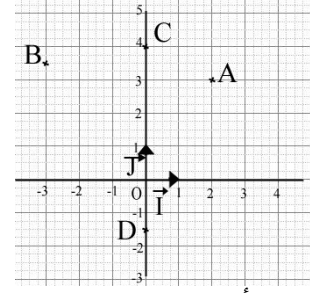
(د)



(ج)

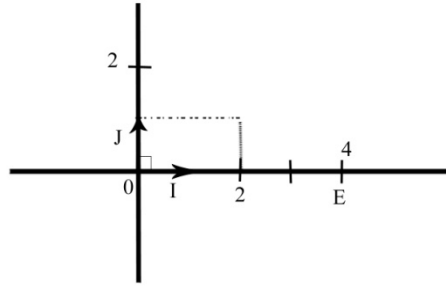


(ب)



(أ)

5) ما هما إحداثيتا النقطة  $E$  في البيان التالي؟



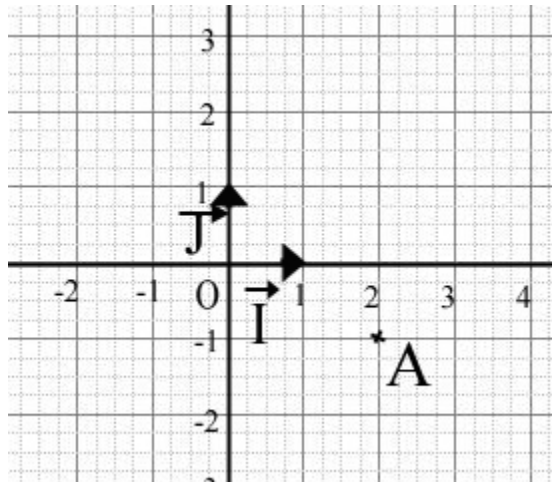
$$E(4,4) \text{ (د)}$$

$$E(1,4) \text{ (ج)}$$

$$E(4,0) \text{ (ب)}$$

$$E(0,4) \text{ (أ)}$$

6) إحداثيتا النقطة  $A$  في المعلم المتعامد والمتجانس الآتي هما: أ  $(2,1)$  ب  $(-2,-1)$  ج  $(2,-1)$  د  $(-1,2)$

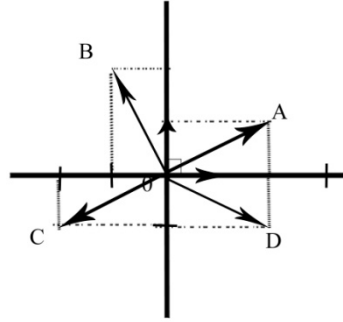




7 ( F نقطة من المستوي المزود بمعلم  $(O, \vec{i}, \vec{j})$  بحيث  $F(2,3)$  هل إحداثيتي الشعاع  $\vec{OF}$  هي :

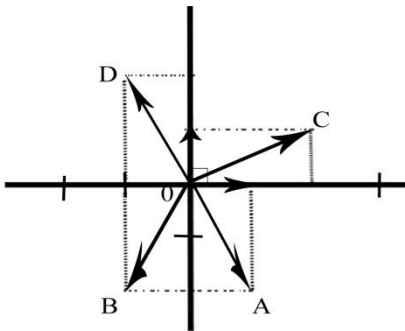
- أ)  $\vec{OF}(3,2)$       ب)  $\vec{OF}(3,3)$       ج)  $\vec{OF}(2,2)$       د)  $\vec{OF}(2,3)$

8 ( عين الشعاع الذي إحداثياته هما  $(2,1)$  من بين الأشعة التالية :



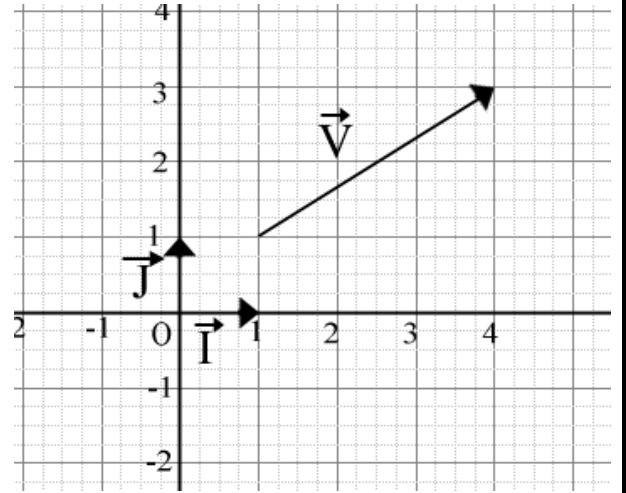
- أ)  $\vec{OA}$       ب)  $\vec{OB}$       ج)  $\vec{OC}$       د)  $\vec{OD}$

9 ( المستوي منسوب إلى معلم  $(O, \vec{i}, \vec{j})$  , وليكن الشعاع الذي مبدؤه النقطة O وإحداثيته هما  $(-1,-2)$  حدد النقطة التي هي نهاية هذا الشعاع .

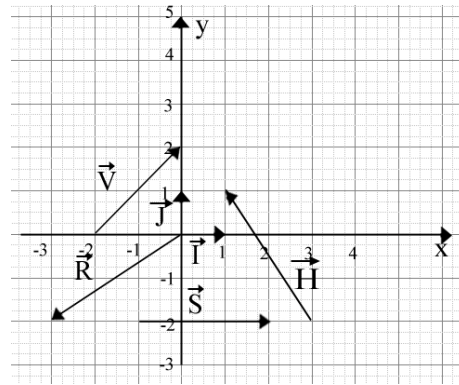


- أ) A      ب) B      ج) C      د) D

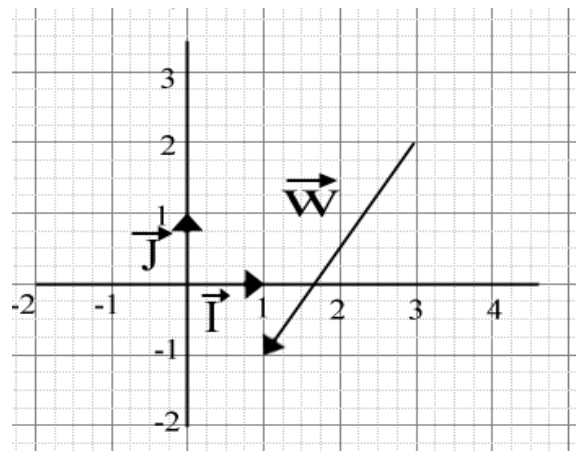
10) من البيان أدناه ، هل إحداثيتا الشعاع  $\vec{V}$  هي: أ)  $\vec{V} (2,3)$  ب)  $\vec{V} (3,2)$  ج)  $\vec{V} (-2,3)$  د)  $\vec{V} (2,-3)$

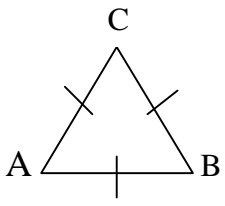


11) هل الشعاع المعين بالنقطتين  $(-1,-2)$  و  $(2,-2)$  هو: أ)  $\vec{V}$  ب)  $\vec{R}$  ج)  $\vec{S}$  د)  $\vec{H}$



12)- هل الشعاع  $\vec{W}$  المرسوم في البيان التالي هو: أ)  $\vec{W} (3,2)$  ب)  $\vec{W} (2,3)$  ج)  $\vec{W} (-3,2)$  د)  $\vec{W} (-2,-3)$





13 ( )  $ABC$  مثلث متقايس الأضلاع إحداثيتا الشعاع  $\vec{AB}$  في المعلم المتجانس  $(A, B, C)$  هما :

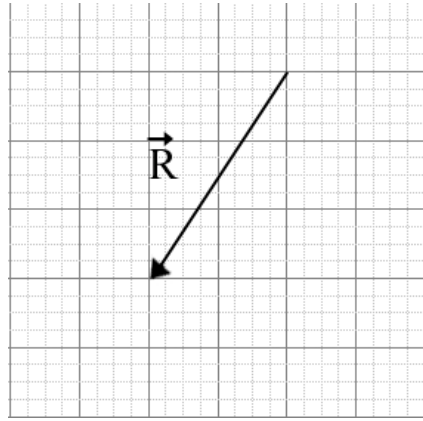
(أ)  $\vec{AB}(0,1)$

(ب)  $\vec{AB}(1,0)$

(ج)  $\vec{AB}(-1,0)$

(د)  $\vec{AB}(0,-1)$

14 ( ) ما هما إحداثيتا الشعاع  $\vec{R}$  في البيان التالي؟:



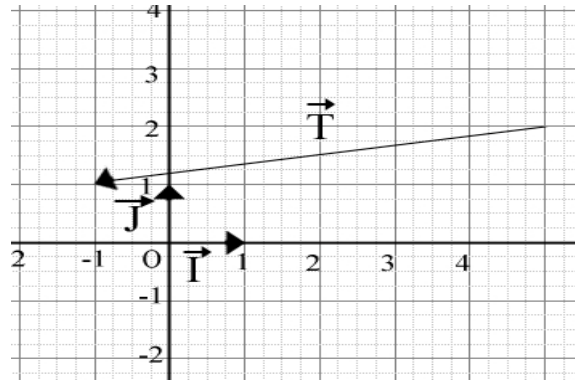
(أ)  $(-2,3)$

(ب)  $(-2,-3)$

(ج)  $(2,3)$

(د)  $(2,-3)$

15 ( ) إحداثيتا الشعاع  $\vec{T}$  في المعلم الآتي هما : (أ)  $(-6, 1)$  (ب)  $(6,1)$  (ج)  $(-6,-1)$  (د)  $(6,-1)$



16) ليكن في مستو منسوب إلى معلم متعامد ومتجانس النقطتين  $A(0,-2)$  ،  $B(3,0)$  . ماهي الإجابة الصحيحة من بين الأجوبة التالية ؟ :

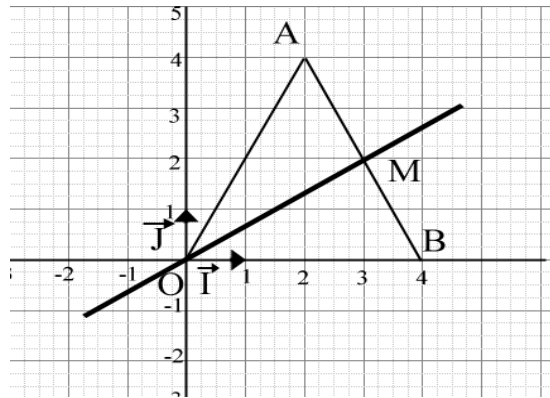
(د)  $\vec{AB}(3,2)$

(ج)  $\vec{AB}(-3,-2)$

(ب)  $\vec{AB}(3,-2)$

(أ)  $\vec{AB}(-3,2)$

17) ليكن في المستوي المنسوب للمعلم النقطتين  $A(2,4)$   $B(4,0)$  . هل العلاقة التي تسمح بحساب إحداثيات النقطة  $M$  منتصف القطعة  $[AB]$  هي:



$YM = YA + YB$

$X_M = X_A + X_B$  (أ)

$YM = \frac{(YA + YB)}{2}$

$X_M = \frac{(X_A + X_B)}{2}$  (ب)

$YM = YA - YB$

$X_M = X_A - X_B$  (ج)

$YM = \frac{(YA - YB)}{2}$

$X_M = \frac{(X_A - X_B)}{2}$  (د)

18 (  $V, W$  نقطتان من المستوي حيث  $V(x_v, y_v)$  ،  $W(x_w, y_w)$  . من بين العلاقات التالية عين العلاقة التي تسمح بحساب المسافة  $VW$  .

$$VW^2 = (x_v - x_w)^2 + (y_v - y_w)^2 \quad (\text{أ})$$

$$VW = \sqrt{(x_v - x_w)^2 + (y_v - y_w)^2} \quad (\text{ب})$$

$$VW^2 = (x_w - x_v)^2 + (y_w - y_v)^2 \quad (\text{ج})$$

$$VW = \sqrt{(x_w - x_v)^2 + (y_w - y_v)^2} \quad (\text{د})$$

19) لتكن النقطتين  $L, K$  في المستوي المنسوب إلى معلم متعامد ومتجانس والمعرفتين كما يلي:  $L(3,0)$  و  $K(0,4)$

ما هو طول القطعة المستقيمة  $KL$  ؟

(د) 12

(ج) 5

(ب) 3

(أ) 4

20 ( المسافة بين نقطتين هي 3. حدد النقطتين  $B, A$  التي تكون المسافة بينهما 3 من بين الثنائيات النقطية التالية:

(د)  $A(2,1)$  ،  $B(4,2)$

(ج)  $A(2,2)$  ،  $B(1,4)$

(ب)  $A(2,2)$  ،  $B(2,4)$

(أ)  $A(1,2)$  ،  $B(4,2)$

## قائمة المراجع

- 1-علام، صلاح الدين محمود "الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية و التدريبية " ، القاهرة ، دار الفكر العربي.( 2007 ) .
- 2.Hansche, L" Handbook for the Development of Performance Standards: Meeting the Requirements of Title I". Washington, US Department of Education and the Council of Chief State School Officers. .(1998)
3. Berk, R. A. "A Consumer's Guide to Setting Performance Standards on Criterion Referenced Tests". *Review of Educational Research*, 56, 137–172. (1986).
- 4.Wang, L., Pan, W., & Austin, J. T." Standards-setting procedures in accountability research: Impacts of conceptual frameworks and mapping procedures on passing rates". Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL(2003, April) .
- 5.حجاج غانم "بحوث معاصرة في القياس النفسي و علم النفس التربوي"، ط1 القاهرة عالم الكتب. (2007).
- 6 .Cizek, G J. Bunch. B , M "Standard setting: A guide to establishing and evaluating performance standards on tests." Sage Publications, Inc.London. (2007) .
- 7.Chang, L. " Judgemental item analysis of the Nedelsky and Angoff standard-setting methods". *Applied Measurement in Education*, 12, 151-165. (1999).
- 8.Kaftandjieva ,F " Methods for Setting Cut Scores in Criterion-referenced Achievement Tests". EALTA.Cito, Arnhem. .(2010) .
9. سوميه شكري "فاعلية بعض طرق تقدير درجة الفصل في التنبؤ بالتحصيل اللاحق في الهندسة للمرحلة الإعدادية " رسالة ماجستير غير منشورة ، كلية التربية ، جامعة المنيا . (2006) .
- 10.Kane, M. T" So Much Remains the Same: Conception and Status of Validation in Setting Standards". In G. J.Cizek (Ed.), *Setting performance standards: Concepts, methods,and perspectives* (pp. 53-88). Mahwah, NJ: Erlbaum. . (2001).
- 11.Popham, W. J. " Criterion Referenced Measurement". Englewood cliffs NJ. prentic hall. (1978) .
12. علي محمد سليمان، أمين "القياس و التقويم في العلوم الإنسانية ( أسسه و أدواته و تطبيقاته)" ط1 ، القاهرة، دار الكتاب الحديث. (2009) .
- 13.Hambleton,R,K.Eignor,D,R" A Practioner Guide to Criterion Referenced Test Development,Validation and Test Score Usage".(2<sup>nd</sup> ed).Laboratory of Psychometric and Evaluative Research.Washington,D.C. .(1979) .
14. Kane, M. ". Validating the performance standards associated with passing scores". *Review of Educational Research*,64(3), 425–461. (1994)
15. علام، صلاح الدين محمود " استخدام النموذج ذي الحدين في تقدير درجة القطع لاختبار محكي المرجع (دراسة احصائية و تجريبية)، المجلة العربية للعلوم الإنسانية،جامعة الكويت ،العدد 19 المجلد الخامس ،ص 26 - 43 ( 1985) .

16. علام، صلاح الدين محمود" دراسة مقارنة لبعض طرق تحديد مستويات الأداء في اختبار مرجعي المحك " ، المجلة المصرية للدراسات النفسية ، تصدرها الجمعية المصرية للدراسات النفسية ، ع1 ، سبتمبر ، ص ص 77 - 96 . ( 1991 ) .

17. الجبة، عصام الدسوقي إسماعيل "مدى فاعلية نموذج "أنجوف " في تحديد المستوى لاختبار محكي المرجع". مجلة كلية التربية جامعة المنصورة، العدد 36 ، ص 43-73.(1998).

18. 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 1 - 10 . ( 2006 )

19. بن عطا، زيد . سوالمة، يوسف." استخدام نموذج المحاولات ذات الحدين في فحص تقديرات المحكمين لدرجة القطع لاختبار مرجعي المحك في الرياضيات". المجلة الأردنية في العلوم التربوية. مجلد 4 ، عدد 1، ص ص 37 - 48 . (2008).

20. عودة، أحمد. الشريم، أحمد." تطوير أسلوب قائم على الأهداف السلوكية لتحديد درجة القطع .(دراسة مقارنة مع أسلوب أنجوف)". المجلة الأردنية في العلوم التربوية. مجلد 6 ، عدد 3 ، ص ص 191 - 202 . (2010).

21. Brennan, R. & Lockwood, R " A comparison of Nedelsky and Angoff Cutting Score Procedures using Generability Theory". *Applied Psychological Measurement*. (Vol.4.N°1, pp 219-240). (1980).

22. Skakun, E. N & Kling, S " Comparability Of Methods For Setting Standards". *Journal Of Educational Measurement*. Vol 17 , PP 229 – 235). (1980).

23 . Rock, D, a. et al "An Empirical Comparisons of Judgemental Approaches to Standard Setting Procedures". Educational Testing Service, Princeton, N.J. (1980) .

24. Behuniak, P., Archambault, F. X., and Gable, R. K" Angoff and Nedelsky Standard Setting Procedures: Implications For The Validity of proficiency Test Score Interpretation. *Educational and Psychological Measurement*. Vol 42: PP 247-255). (1982).

25. Livingston, S. A., & Zieky, M. J. " A Comparative Study of Standard-Setting Methods". Educational Testing Service ,Princeton, N.J. (1983).

26. Demauro, G & Powers, D" Logical consistency of The Angoff Method of Standard Setting", Paper Presented at The Annual Meeting of The National Council on Measurement in Education. Boston, MA, 17-19 April. (1990) .

27. Chang, L" A Comparison Between The Nedelsky and Angoff Standard Setting Methods "? Paper presented At The Annual Meeting of The National Council On Measurement In Education, New York. (1996) .

28. Donnoe, W. & Amato, R" Supportive Data & Guidelines For Using The Angoff , Ebel and Nedelsky Cut Off Score Methods". Paper Presented At The International Personnel Management Assessment Council Conference ;New Port Beach. (1997).

29. Nastrom.G,Nystrom.p" A Comparison of Two Different Methods For Setting Performance Standards for a Test With Constructed- Response Items". Practical Assessment ,Research & Evaluation.(Vol.13 N°9,pp 1-12) (2008) .
30. Buckendahl.C.W,Ferdous.A.A,Gerrow.J"Recommending Cut scores with Subset of Items:Empirical Illustration".Practical Assessment,Research & Evaluation.(Vol. 15 N°6,pp1-10). (2010).
31. Shepard , L. " Setting Performance Standards, N R.A.Berk. (Ed) :AGuide To Criterion – Referenced Test Construction " , pp . 169 – 198 The Johns Hopkins University press , London , (1984).
32. Livingston, S.A"Assumptions of Standard Setting Methods". Paper Presented at the Annual Meeting of NCME, March, New York. (1982).
33. Jaeger, R. M. " Certification of student competence". In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 485–514). Washington, DC: American Council on Education. (1989).
34. Gronlund, N, E.Miller,D ,L.Linn,R,L"Measurement and Assessment in Teaching". (10th Ed).Pearson Education,Inc,Upper Saddle River,New Jersey. .(2009) .
35. Gronlund, N, E, " Constructing achievement tests".(2<sup>nd</sup> Ed). Prentice-Hill, INC., Englewood Clifls, N.J. (1977) .
36. عودة، احمد" القياس و التقويم في العملية التدريسية".الأردن ، أريد. دار الأمل للنشر. ( 2005 ) .
37. أبو حطب، فؤاد".التقويم النفسي". القاهرة، الأنجلو المصرية . ( 1987 ) .
38. Glaser, R" Instructional technology and the measurement of learning outcomes. Some questions", *American Psychologist*, 18, 519-523. (1963).
39. Gregory, R. J" Psychological testing: history, principles and applications" (4th edition). Needham Hights: Allyn & Bacon. (2004).
40. Popham, W.J., & Husek, T.R. " Implications of Criterion-Referenced Measurement." *Journal of Educational Measurement*, 6(1), 1-9.Cliffs NJ. (1969).
41. أنور محمد الشرقاوي ونادية محمد عبد السلام " 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. (1996).
42. Hambleton, R.K. (1994). The rise and fall of criterion-referenced measurement"?*Educational Measurement: Issues and Practice*, 13(4), 21-27. (1994).
43. أنور محمد الشرقاوي ونادية محمد عبد السلام " بناء اختبار هدفي المرجع وجماعي 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. (1996).



44. علام، صلاح الدين محمود "القياس و التقويم التربوي (أساسياته و تطبيقاته و توجهاته المعاصرة)" ، ط 1 ، القاهرة ، دار الفكر العربي. ( 2000 ) .
45. Hambleton ,R.Olszewski,F"Woodworking Objective and Test Item Bank".Center for Occupational Education.New York. (1972) .
46. أنور محمد الشرقاوي ونادية محمد عبد السلام " القياس محكي المرجع و تطويع التعليم (دراسة تجريبية)" ، أنور محمد الشرقاوي، سليمان الخصري 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. (1996).
47. Hambleton,R.Zenisky,A"Advances in Criterion –Referenced Testing Methods and Practices".in Reynolds,C,R & Kamphaus,R,W (ed).Handbook of Psychological & Educational Assessment of Children.(2<sup>nd</sup> ed pp 377-404). the Guilford Press.New york,NY. (2003) .
48. Popham, W. J. " Specifying the domain of content or behaviors". In R. A. Berk (Ed.), A guide to criterion-referenced test construction (pp. 29–48). The Johns Hopkins University press , London (1984).
49. Hambleton, R. K. " Criterion-referenced testing, technical advances, and evaluation guidelines".In C. Reynolds & T. Gutkin (Eds.), Handbook of school psychology (3rd ed., pp. 409 – 434). New York: Wiley. (1998).
50. Mehrens,W,A.Ebel,L,R " Some Comments On Criterion-Referenced and Norm-Referenced tests".Measurement In Education .Vol,10,n1.pp 1-9. (1979)
51. Downing ,S, M. " 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. (2006) .
52. Abedi, J" 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. (2006) .
53. Linda Crocker, L. James Algina J" Introduction to Classical and Modern Test Theory". Cengage Learning.USA. (2006) .
54. Berk,R.A " Item Analysis". In R.A.Berk.Criterion-Referenced Measurement : The State of The Art.The Gohns Hopkins University Press. (1980).
55. القاطعي، عبد الله علي " دراسة مقارنة لبعض طرق تحليل بنود الاختبارات مرجعية المحك وفعاليتها في الاختبار، دراسات تربوية"، المجلد الثامن، الجزء 50، ص ص 106-131. (1993).
56. الجهني طارق " أثر اختلاف بعض طرق تحليل بنود الاختبار محكي المرجع على اختيار الفقرات و الثبات"، رسالة ماجستير غير منشورة،جامعة الملك سعود،كلية التربية،المملكة العربية السعودية. (2004).
57. بشناق عدنان" بناء اختبار محكي المرجع لقياس الكفايات الرياضية في العمليات الحسابية على الأعداد بالمرحلة الابتدائية"، رسالة ماجستير غير منشورة،جامعة أم القرى،كلية التربية، المملكة العربية السعودية. (1991).

58. Cizek, G. J. "Standard Setting". In S. M. Downing & T. M. Haladyna (Ed.), *Handbook of Test Development*. (pp. 225–258). Lawrence Erlbaum Associates, Inc. Mahwah, New Jersey. (2006) .
59. Zieky, M. "A Historical Perspective on Setting Standards. Joint Conference on Standard Setting for Large-Scale Assessments. Proceedings": Vol.2, Ed. By L. Crocker & M. Zieky, Washington, U.S. Government Printing Office, 1-38( 1994) :
60. Cizek, G. J. "Conjectures on the Rise and Fall of Standard Setting: An Introduction to Context and Practice". In G. Cizek (ed) *Standard-setting: Concepts, methods, and perspectives*. Hillsdale, Erlbaum, 2001, 3-18. (2001) .
61. Livingston, S. & M. Zieky. " Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests". Princeton, NJ: ETS(1982) .
62. Zieky, M. J. " So Much Has Changed: How the Setting of Cutscores Has Evolved Since the 1980s". In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 19–52). Mahwah, NJ: Lawrence Erlbaum. (2001).
63. Hertz, G. & N. Hertz " How Many Raters Should be Used for Establishing Cutoff Scores with the Angoff Method? A Generalizability Theory Study". *Educational and Psychological Measurement*, 59 (6), 885-897. (1999) .
64. Hambleton, R. K. " \_Setting performance standards on educational Assessments and criteria for evaluating the process". In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, methods, and perspectives* (pp. 89–116). Mahwah, NJ: Lawrence Erlbaum. (2001).
65. Goodwin, L. "Focus on Quantitative Methods Determining Cut – Off Scores". *Research in Nursing & Health*. 19(1). 857-872. (1996).
66. Hambleton, R. K., & Pitoniak, M. J. "Setting performance standards". In R. L. Brennan (Ed.) *Educational Measurement* (4th ed.). Westport: American Council on Education & Praeger Publishers. (2006).
67. Kane, M. " Choosing between examinee-centred and test-centred standard-setting methods". *Educational Assessment*, 5(3), 129-145. (1998).
68. Giraud, G., Impara, J. C., & Buckendahl, C. " Making the cut in school districts: alternative methods for setting cutscores". *Educational Assessment*, 6(4), 291-304. (1999/2000).
69. Zieky, M., & Livingston, S. " Manual for setting standards on the Basic Skills Assessment Tests. Princeton, NJ: Educational Testing Service. . (1977) .
70. وزارة التربية الوطنية، الوثيقة المرافقة لمناهج السنة الرابعة من التعليم المتوسط. جويلية (2005).
71. Kubisyn, T. Borich, G. " Educational Testing and Measurement : Classroom Application and Practice". (7th ed). John Wiley & Sons. Inc. USA. (2003) .
72. Dunn, D. S. "Statistics and Data Analysis for the Behavioral Sciences". Mc Graw-Hill. New York. NY. (2001) .
73. بوعلام محمد " الموجه في الإحصاء الوصفي و الاستدلالي في العلوم النفسية و التربوية و الاجتماعية. دار الأمل. (2009) .