

جــامعــة صنعــــــاء
نيابة الدراسات العليا والبحث العلمي

كلية التربية

قسم مناهج العلوم وطرائق تدريسها

تقويم اختبارات العلوم لنهاية مرحلة التعليم الأساسي وفق معايير ومواصفات الاختبار الجيد

بحث مقدم لاستكمال متطلبات الحصول على درجة الماجستير في التربية

(مناهج العلوم وطرائق تدريسها)



**إعداد**

**محمد علي ثابت سعيد**

**إشــــراف**

|  |  |
| --- | --- |
| **الدكتور/ عبدالولي حسين دهمش** | **الدكتورة/ أمة الكريم طه أبو زيد** |
|  |  |
| أستاذ مناهج العلوم وطرائق تدريسها المشارككلية التربية – جامعة صنعاء | أستاذ مناهج العلوم وطرائق تدريسها المشارككلية التربية – جامعة صنعاء |

**1430 هـ / 2009م**

**ملخص البحث**

**تقويم اختبارات العلوم لنهاية مرحلة التعليم الأساسي**

**وفق معايير ومواصفات الاختبار الجيد**

**إعداد:**

محمد علي ثابت

2010م

**إشراف:**

|  |  |
| --- | --- |
| الدكتور/ عبدالولي حسين دهمش | المشرف الرئيـس |
| الدكتورة/ أمة الكريم طه أبو زيد | المشرف المشارك |

يهدف هذا البحث إلى التعرف على واقع اختبارات العلوم لنهاية مرحلة التعليم الأساسي وفق معايير ومواصفات الاختبار الجيد في الجمهورية اليمنية ,والتي تم التوصل إليها من خلال هذا البحث أيضاَ.

وكانت أسئلة البحث على النحو التالي:

1. **ما أهم معايير ومواصفات الاختبار التحصيلي الجيد لمادة العلوم لنهاية مرحلة التعليم الأساسي؟**
2. **ما مدى توافر هذه المعايير والمواصفات في كل نموذج من النماذج الاختبارية المشمولة في هذا البحث؟**
3. **ما مدى تحقق التوازن بين نماذج الاختبارات المتعددة في السنة الدراسية الواحدة؟**

استخدم الباحث العينة العشوائية البسيطة, والتي تحددت بـ (15) نموذجاً اختبارياَ ممثلاً لخمسة أعوام دراسية متتالية، وهي من العام 2003/2004م إلى العام 2007/2008م, أي ما يعادل نسبة 34% تقريباً من حجم المجتمع، وقد بلغت الأسئلة التي تم تحليلها (90) سؤالاَ, والتي اشتملت على (194) فرعاَ وكانت متضمنة عدد (641) فقرة اختبارية.

وتم إعداد ثلاث أدوات تحقق الإجابة عن أسئلة البحث, حيث كانت الأداة الأولى: استبانة شبه مفتوحة، بغرض تحديد معايير ومواصفات اختبار العلوم الجيد, وكانت الأداة الثانية: جدول مواصفات إعداد الاختبار, والأداة الثالثة: بطاقة تحليل نماذج الاختبارات حيث تم إعدادهما في ضوء قائمة المعايير التي تم التوصل إليها, وقد تم التأكد من صدق وثبات الأدوات،وقد تبين صلاحيتهما للبحث بدرجة كبيرة.

استخدم الباحث المنهج الوصفي، وتم اتباع أسلوب تحليل المحتوى للنماذج الاختبارية المشمولة في هذا البحث, لتسهيل التفسير والتقويم والمقارنة بينها .

وكانت فئة التحليل المناسبة للبحث هي المعيار والتي تم تطبيقها على المضمون لوضع مقياس كمي, وتم اختيار الفقرة الاختبارية كوحدة للتحليل.

**الوسائل الإحصائية المستخدمة وهي:**

التكرارات والنسب المئوية والمتوسطات الحسابية ومعادلة هولستي لحساب الثبات. وتوصل البحث إلى ما يلي :

**أولاَ:أهم معايير ومواصفات الاختبار الجيد هي :**

1. التمثيل النسبي لمجالات المحتوى ( أحياء , الأرض والفضاء, فيزياء, كيمياء ) .
2. التمثيل النسبي لمستويات المعرفة ( تذكر, فهم , تطبيق وما فوق ) .
3. شمولية الاختبار لأهداف تدريس العلوم.
4. تنوع الأسئلة.
5. مناسبة زمن الاختبار.

 **ثانياَ: مدى توافر هذه المعايير والمواصفات في كل نموذج من النماذج الاختبارية المشمولة في هذا البحث وكانت على النحو التالي:**

1. **التمثيل النسبي لمجالات المحتوى :** ظهر التمثيل النسبي لمجال الأحياء متدرجا بين (22.2 % - 42.9%) وكانت تسعة نماذج قريبة من نسبة التمثيل النسبي المطلوبة والمحددة بنسبة 29% , وظهر التمثيل النسبي لمجال الأرض والفضاء متدرجا بين (صفر% - 11.1 %) وكانت عدد ستة نماذج قريبة من نسبة التمثيل المطلوبة والمحددة بنسبة 5%, أما التمثيل النسبي لمجال الفيزياء فقد ظهر متدرجا بين( 19% - 34.2 % ) ولم يحقق أي من النماذج النسبة المطلوبة والمحددة بنسبة 40%, وظهر التمثيل النسبي لمجال الكيمياء متدرجا بين ( 26.3% - 46.8% ) متجاوزا النسبة المطلوبة والمحددة بنسبة 26% .
2. **التمثيل النسبي لمستويات المعرفة:** ظهر التمثيل النسبي لمستوى التذكر متدرجا بين (37.1% - 76.9% ), ومتجاوزاَ للنسبة المحددة له في جدول المواصفات وهي 30%, وظهر التمثيل النسبي لمستوى الفهم متدرجا بين ( 22.7% - 62.9%) , وكانت ثلاثة نماذج من العينة فيها التمثيل النسبي لمستوى الفهم قريب من النسبة المطلوبة والمحددة بنسبة 30% , وظهر مستوى التطبيق وما فوق بنسب ضئيلة أو مفقودة متدرجا بين (صفر % - 8.3% ) بينما النسبة المطلوبة محددة بنسبة 40 % .
3. **شمولية الاختبار لأهداف تدريس العلوم :** غير محقق في النماذج الاختبارية, حيث ظهرت نسبة ارتباط فقرات الاختبارات بالأهداف متدرجة بين (23% - 40% ) وهي نسبة متدنية جداَ عن النسبة المطلوبة (درجة القطع في هذا البحث) والمحددة بنسبة لا تقل عن 70% من الأهداف الواردة في دليل معلم العلوم للصف التاسع الأساسي.
4. **التنوع في الأسئلة:** تحقق معيار التنوع في الأسئلة بدرجة عالية،حيث ظهرت نسبة النوع المقالي متدرجة بين (36.6% - 63.2%) وعلى العكس منه في النوع الموضوعي, وكان التقارب واضحا في معظم النماذج المشمولة في هذا البحث .
5. **مناسبة زمن الاختبار:** غير محقق في جميع النماذج الاختبارية حيث ظهر تقديره بين (55-66 دقيقة) من قبل المحكمين لمعيار الزمن, بينما الزمن المحدد للاختبار وقدره (180 دقيقة ) من قبل واضعيه كان طويلا جداَ قياساَ إلى ما تم التوصل إليه في هذا البحث.

**ثالثاَ: مدى تحقق التوازن بين نماذج الاختبارات المتعددة في السنة الدراسية الواحدة وكانت على النحو التالي:**

النماذج الاختبارية المتعددة في السنة الدراسية الواحدة غير متوازنة (أي غير مرتبطة بأهداف مشتركة من أهداف المقرر الواردة في دليل المعلم) حيث ظهر التوازن التام متدرجاً بين (9.6 %- 23.4%)، وهي قيم منخفضة جداً, أي أن ظهور أهداف مشتركة بين النماذج الثلاثة- العينة الممثلة للسنة الدراسية الواحدة - كانت قليلة جداَ، وظهر التوازن الجزئي متدرجاً بين (23.1%- 37.5%) وهي أيضاً قيم منخفضة، أي أن ظهور أهداف مشتركة بين نموذجين من النماذج الثلاثة - العينة الممثلة للسنة الدراسية الواحدة - كان قليلاً جداَ, بينما انعدام التوازن ظهر متدرجاً بين (45.8%- 67.3%)، أي أن ظهور عدد كبير من الأهداف بشكل منفرد في كل نموذج من النماذج الثلاثة - العينة الممثلة للسنة الدراسية الواحدة- وهذا يعني أن انعدام التوازن هو الصفة السائدة لهذه الاختبارات.

وبناء على هذه النتائج التي توصل إليها البحث, خرج الباحث بعدد من التوصيات ومقترحات لبحوث مستقبلية.

**قائمة المحتويات**

| **الموضــــــــوع** | **الصفحة** |
| --- | --- |
| قرار لجنة المناقشة والحكم | ج |
| الإهداء | هـ |
| شكر وتقدير | و |
| ملخص البحث باللغة العربية | ي |
| قائمة المحتويات | ط |
| قائمة الجداول | ك |
| قائمة الأشكال | ل |
| قائمة الملاحق | م |
| **الفصل الأول: مشكلة البحث وأهميته** | 1 |
| المقدمة | 2 |
| مشكلة البحث | 4 |
| أهداف البحث | 5 |
| أهمية البحث | 5 |
| حدود البحث | 5 |
| مصطلحات البحث | 6 |
| **الفصل الثاني: الإطار النظري** | 8 |
| أولاً: الخلفية النظرية  | 9 |
| ثانياً: الدراسات السابقة | 30 |
| **الفصل الثالث: منهجية البحث وإجراءاته** | 36 |
| أولاً: منهج البحث | 37 |
| ثانياً: مجتمع البحث | 37 |
| ثالثاً: عينة البحث | 38 |
| رابعاً: أدوات البحث  | 39 |
| خامساً: عملية التحليل | 42 |
| سادساً: بناء جدول المواصفات | 44 |
| سابعاً: معيار الزمن | 45 |
| ثامناً: الوسائل الإحصائية المستخدمة | 45 |
| **الفصل الرابع: عرض نتائج البحث ومناقشتها والتوصيات والمقترحات** | 46 |
| أولاً: النتائج والمناقشات | 47 |
| ثانياً: التوصيات | 62 |
| ثالثاً: المقترحات | 62 |
| المراجع | 63 |
| الملاحق | 69 |
| نماذج من عينة أسئلة الاختبارات | 86 |
| الملخص باللغة الإنجليزية | 96 |

**Abstract**

***Evaluation of Science Tests for Basic Education's Final Grade According to Good Test's Standards and Criteria***

**By Mohammad Ali Thabit**

This research aims to identify the reality of science tests for the end of the stage of basic education in accordance with the standards and specifications of the test well in the Republic of Yemen , which has been reached by this research as well.

 **The research questions are as follows:**

1. What is the most important standards and specifications of the achievement test for good science to end-stage basic education?
2. What is the availability of these standards and specifications in all of the models covered by the test in this search?
3. What is the balance between the multiple models tests per school year?

The researcher used simple random sampling , which Thddtb (15) model of a test representative of a five-year study in a row , one of the year 2003/2004 AD to the year 2007/2008 m , the equivalent ratio of approximately 30% of the size of the community , has reached the questions that have been analyzed ( 90 ) question , which included 194 branches and a number were included ( 641 ) paragraph test .

Was developed three tools verify answer research questions , where she was the first tool : identification of semi-open , in order to determine the standards and specifications of the test of science good , and the second tool : Specification Table test preparation , and the third tool : card analysis models tests have been prepared in light of the list of criteria that have been reached , has been confirmed validity and reliability of the tools , has been found disqualified for search significantly .

The researcher used the descriptive approach , was to follow the method of content analysis of the models covered in this test

Find , to facilitate the interpretation and evaluation and comparison .

The analysis of the appropriate category to search is the standard that has been applied to content to develop a quantitative measure , was chosen as a test paragraph for analysis.

Statistical methods used are:

Frequencies, percentages and averages and Holsti equation to calculate stability

The research found the following:

First, the most important Maiaromuasfat good test are:

1. Proportional representation of the content areas ( neighborhoods , the land of science , physics , chemistry).
2. Proportional representation of the levels of knowledge (remember , understand , apply and above(.
3. Comprehensiveness of the test objectives for the teaching of science .
4. The diversity of questions.
5. Manasphzmn test.

 Second, the availability of these standards and specifications in all of the models covered by the test in this research were as follows:

1. Proportional representation of the content areas : the back of proportional representation for the field of biology graded between (22.2% - 42.9% ) and was nine models close to the ratio of proportional representation required and specified by 29 % , and the back of proportional representation for the field of the Earth graded between (zero % - 11.1% ) and was the number of six models close to the ratio of representation required and specified by 5 %, while proportional representation for the field of physics has appeared graded between (19% - 34.2 % ) did not achieve any of the forms required percentage identified by 40 % , and the back of proportional representation for the field of chemistry graded between ( 26.3 % - 46.8 % ) , surpassing the required percentage and 26% specific .
2. Proportional representation of the levels of knowledge : the back of proportional representation to the level of remembering graded between ( 37.1 % - 76.9 % ) , and surpassing the percentage set for it in the table of specifications , a 30% , and the back of proportional representation to the level of understanding graded between ( 22.7 % - 62.9 % ) , and the three models of sample the proportional representation of the level of understanding of the relative ratio required and specified by 30% , and appeared at the application level and above in small amounts or missing graded between (zero % - 8.3% ), while the specific ratio required by 40%.
3. Comprehensiveness of the test to the goals of teaching science : is an investigator in the models test , which showed the proportion of link paragraphs testing objectives gradient between (23 % - 40% ), a very low rate for the required percentage ( cut-off grade in this research ) and defined by at least 70% of the objectives contained in the guide science teacher for ninth-grade primary .
4. Diversity in the questions : Check the standard questions of diversity in a high degree , where she developed the gender ratio gradient between pans (36.6 % - 63.2% ) and on the contrary it in the kind of substantive , and convergence was evident in most of the models covered in this research .
5. Manasphzmn test: is an investigator in all forms of test where the back of appreciation between ( 55-66 minutes) by the arbitrators to standard time, while the allotted time for the test of ( 180 minutes ) by the authors was a very long compared to what has been reached in this research .

Third: the extent to which the balance between the multiple models tests per school year were as follows:

      Multiple experimental models per school year is balanced (ie not linked to the objectives of the common objectives of the course in the Teacher's Guide ) where it appeared the perfect balance between gradual ( 9.6% - 23.4 % ) , which is very low values ​​, that is, the emergence of shared goals between the three models - representative sample for the year per school - too few , and the back of the partial equilibrium graded between ( 23.1 % - 37.5 %), which is also low values ​​, that is, the emergence of shared goals between the two models of the three models - the representative sample for the year per school - was Kulailajda, while the lack of balance between the back of a gradual ( 45.8 % - 67.3 % ) , ie, that the emergence of a large number of targets individually in each model of the three models - the representative sample of one academic year - this means that the imbalance is the dominant characteristic of these tests.

Based on these findings of the research, the researcher came out a number of recommendations and suggestions for future research .