Item Analysis of History Achievement Test on Difference Index (DI) in the Criterion Referenced Measurement

Gowdhaman K. and Nachimuthu K.
Department of Education, Periyar University, Periyar Palkalai Nagar, Salem, Tamil Nadu, INDIA

Abstract

This piece of work was aimed to analyze test items of a achievement in the subject of history (Indian) standard of 11th. Objective tests are widely used to measure intelligence, aptitude, and achievement (or attainment). Item analysis allows selecting or omitting items from the test, but more important, item analysis is a tool to help the item writer improve an item. Item statistics used in the item analysis of such criterion-referenced test: the difference index (DI). The difference index shows the gain, or difference in performance, on each item between the pretest and posttest. Purposive sampling was adopted for this study and 15 students were taken from municipal girl's higher secondary school, Gugai, Salem district in Tamil Nadu state, country of India. A test of 190 items was used for data collection. Photographs (358) based teaching history of instruction, before and after the tests were conducted. Findings of the study were total 190 items of achievement test for the difference index like 80 items are much better related to the curriculum, 25 items are higher values and 85 items are not fitting because they reflect only small gains. So these items were rejected. Finally 105 items were selected form of history achievement test.

Keywords: Achievement test, Objective test, Item analysis, Difference index, Criterion-referenced test.

Introduction

Educational evaluation is the evaluation process of characterizing and appraising some aspects of an educational process. There are two common purposes in educational evaluation which are, at times, in conflict with one another. Educational institutions usually require evaluation data to demonstrate effectiveness to funders and other stakeholders, and to provide a measure of performance for marketing purposes. Educational evaluation is also a professional activity that individual educators need to undertake if they intend to continuously review and enhance the learning they are endeavoring to facilitate[1].

Objective tests

An objective test is one for which the rules for scoring it are so specific and comprehensive that anyone who marks a test script in accordance with these rules will arrive at the same test score. Most objective tests used in education are composed of a sequence of individual "objective" test items in which the testees must choose their answers from a specified list of alternatives rather than by creating them for themselves. It is important to remember, however, that the definition relates to the method of scoring the test and not to the format of its constituent items as such. Not all objective tests require the student to select from a presented list. Items which require the student to write down a phase, a word, or a number, and in the scoring of which there are clear and unequivocal rules for deciding whether the response is right or wrong, also qualify as "objective".

Objective tests are widely used to measure intelligence, aptitude, and achievement. The appropriateness of objective tests for the measurement of achievement is much more controversial. Essay tests are still preferred to objective tests for most educational purposes in many countries. Where the teacher scores the tests for a single class, objectivity as such is less important, and the advantages of getting the students to express themselves fully and openly tend to outweigh the demands for a reliable score[2].

Objective test item formats: Objective test items include items with the following formats. i. Completion or Short Answer Item, ii. True-False Item, iii. Multiple Choice Item, iv. Matching Item.

Completion or short answer item: Completion items are relatively easy to write. Perhaps the first tests classroom teachers construct and students take is completion test. Like items of all other formats, though there are good and poor completion items[3]. Unlike the other types in which the student is selecting from a list of a alternative responses presented to him or her, the supply type of item requires a student to construct a response. However, the question is so structured as to limit the score of the student's response so that there will be one, and only one, acceptable answers. Demanding that the student construct rather than recognize the response avoids some of the common criticisms of objective tests.

Specification of the acceptable answers is an essential part of the item construction process, but there is always the danger that certain students will invent unforeseen answers that could arguably be accepted as correct[2].

True/false item: True/false items are popular probably because they are quick and easy to write, or at least they seem to be. Actually true/false items do take less time to write than good objective items of any other format, but good true/false items are not that easy to write[3]. The true/false item presents a declarative statement and requires the examinee to indicate whether he or she judges it to be true or false. Although such items are easy to construct, this is not a format to be generally recommended. Tests composed of true/false items tend to be rather unreliable, and are particularly susceptible to contamination by guessing. Ebel (1970) argues that this is not serious and that true/false tests may be very efficient, but few other writers support this view. True/false items can be quite effective for assessing factual knowledge, but are usually inappropriate for testing more complex skills[2].

Multiple choice items: Another popular item format is the multiple-choice question. These items are unique among objective test items because, contrary to popular belief, they enable to measure at the higher levels of the taxonomy of educational objectives. Multiple-choice items will be in two parts. The first part will consider the mechanics of multiple-choice item construction applied to lower level questions. The second part will deal with the construction of higher level multiple-choice items[3]. The multiple-choice item is by far the most frequently used in educational achievement testing. The number of alternative answers offered varies but is usually four or five. As a rule only one of the alternatives is correct, the others (the distractors) being constructed so as to provide plausible examples of common errors. If the distractors are carefully written a wrong-answer analysis can yield valuable diagnostic information about the types of error being made by students. The chief difficulty in constructing good multiple-choice questions is to find appropriate distractors. To be effective they must be plausible at least to a substantial minority of students, yet they must be clear and unambiguously wrong in the judgment of experts. Distracters should not give inadvertent clues which permit test wise students to eliminate them irrespective of their ability to solve correctly the question. Although they too may effectively be used to assess specific knowledge, multiple-choice items are readily adaptable to measure more complex skills involving reasoning and analysis. It has been found to be relatively straightforward to construct an achievement test, all of whose items are in the multiple-choice format which assesses student performance on a wide range of objectives involving different skill levels[2].

Matching item: The fourth widely used item format to be considered is the matching exercise. The matching items represent a popular and convenient testing format; good matching items are not as easy to write as might think[3]. The most obvious advantage of the matching item format is one of economy. More responses are obtained from the student for the same amount of reading. The format can be effectively employed to test knowledge of specific facts, but is generally unsuitable for more complex objectives. However, it has been suggested variation of the format, the classification task, can be used to appraise comprehension and application-type objectives[2].

Item analysis: Item analysis is a general term that refers to the specific methods used in education to evaluate test items, typically for the purpose, of test construction and revision[4]. Item analysis is the set of qualitative and quantitative techniques and procedures used to evaluate the characteristics of items of the test before and after the test development and construction. An item is a basic building block of a test, and its analysis provides information about its performance. Item analysis allows selecting or omitting items from the test, but more important, item analysis is a tool to help the item writer improve an item. Item analysis can be used both for dichotomously scored (correct or incorrect) items and polytomously scored items.

Purpose of item analysis: The main purpose of item analysis is to improve internal consistency or internal structure validity, focused on confirming a single factor or one-trait test. If the trait is not one factor, then the use of item analysis might tend to lower validity. If a test has two factors, the calculation of item statistics for each item should be focused on the subtotal for the relevant set of items rather than on the total test score. Item analysis in this case is used to improve the internal consistency of each subset of items with no intention to change the dimensionality of the entire set. In these cases, an overall reliability index would be stratified alpha or battery alpha rather than the regular co efficient alpha as used for a one-dimensional test[2].

Criterion-referenced test: Criterion-referenced tests aim to show whether students have achieved a given learning objective, with performance on a test item treated as a behavior that demonstrates learning. Ideally, 100 percent of students pass each item, thereby showing that they have the required knowledge; but it is equally possible that everyone could fail. In mastery tests, the pass mark is usually set at rightly answering 80 percent of the questions. Students can be ranked by their scores, so criterion-referenced tests can also be used for selection, although that not their prime educational purpose. Implicitly, most classroom tests are criterion-referenced, although they might not be written formally against learning objectives. For valid research on student achievement, formal learning objectives are required, they need to be taught and the tests must measure student performance systematically against the learning objectives[5]. Item statistics such as item discrimination index and item difficulty level of non-referenced test are not used in the

same way for criterion-referenced interpretations. Item analysis must focus on group differences and might be more helpful in identifying problems with instruction and learning rather than guiding the item selection process. Two item statics are often used in the item analysis of such criterion-referenced tests: the difference index (DI) and the *B*-index.

The difference index (DI): The difference index is defined as the item facility (IF) on the particular item for the posttest minus the item facility for that same item on the pretest. In other words, the difference index shows the gain, or difference in performance, on each item between (DI= IF posttest – IF pretest) the pretest and posttest[6].

Objective of the study: i. To find the item facility index calculated from pre test and post test. ii. To find the difference index of history achievement test items in the criterion referenced test. iii. To find the performance of achievement test.

Population: In the study all history students, studying in the Tamil medium school of 11th class in Salem district of Tamil Nadu constitute the population of the study.

Sample: Purposive sampling was adopted for this study and 15 students were taken from municipal girls higher secondary school, Gugai, Salem district in Tamil Nadu, India.

Design of the study: The pre test and post test design used in this study. By comparing students' test scores before and after instruction, teachers can at least partly determine what students have learned in class. A test administered as a diagnostic test is also going to be used as an achievement test, they may only study the parts of the class content that are on the test (figure 1).
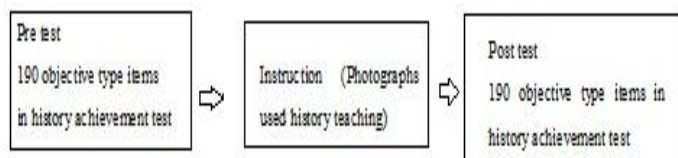
Figure-1
A single form pretest/posttest design

Instrument: A test of 190 items was used for data collection. This test was developed from the text book (2011) of Indian history for class 11th (2012-2013) by the researcher himself. The research has 10 years' experience of teaching of history in government and private sector at secondary level with involvement of five other relevant teachers. Bloom taxonomy was used as framework for test construction. A test was administered in Tamil language (state mother tongue) in this school.

Data collection: A test was administered by the researcher himself for data collection. The researcher enjoyed full support from the working teachers in this school. First researcher conducted a pretest. Photographs (358) based teaching history of instruction and after that post test was conducted (figure 1).

Data organization and analysis: Total scores of the students were entered in excel sheet. Table 1 shows that the item facility (IF) was calculated. The formula for difference index was discussed (table-2).

Results

The Criterion referenced index, difference index – item analysis
Calculating the item facility goes like this: if 5 out of 15 students answered item 1 correctly on the pretest for a course, the pretest item facility (IF pretest) would be 5/15=0.33; if 12 out of the same 15 students answered that same item correctly on the posttest, the post test item facility (IF post test) would be 12/15= 0.80 (table-1).

Calculating the difference index (DI) given that IF posttest = 0.80 and IF pretest =0.33, the DI would be 0.47 (DI=IF posttest‒ IF pretest=0.80‒ 0.33=0.47 (table-2).

The level of difference index calculated.
>.60 = items highly values
>0.30 <0.60 = items better
<0.30 = items rejected

Table 2 shows that items 1, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 22, 24, 27, 28, 29, 30, 31, 32, 33, 34, 36, 38, 39, 41, 42, 45, 46, 47, 50, 52, 56, 57, 63, 64, 65, 68, 70, 71, 72, 74, 75, 77, 78, 80, 82, 83, 86, 88, 95, 96, 99, 106, 107, 108, 110, 113, 116, 124, 137, 139, 151, 152, 154, 155, 156, 157, 169, 171, 172, 173, 176, 182, 183, 185, 186, 187 and 188. These items (80) are much better related to the curriculum.

Items are 67, 79, 81, 87, 97, 101, 103, 104, 105, 112, 114, 120, 130, 131, 132, 134, 140, 148, 160, 161, 166, 168, 177, 181, and 184. These items (25) are higher values (figure -2).

Table 2 items shows that 2, 3, 4, 5, 13, 17, 19, 20, 21, 23, 25, 26, 35, 37, 40, 43, 44, 48, 49, 50, 53, 54, 55, 58, 59, 60, 62, 66, 69, 70, 73, 76, 84, 85, 89, 90, 91, 92, 93, 94, 99, 100, 102, 109, 111, 115, 117, 118, 119, 121, 122, 123, 125, 126, 127, 128, 129, 133, 135, 136, 138, 141, 142, 143, 144, 145, 146, 147, 149, 153, 158, 159, 162, 163, 164, 165, 167, 170, 174, 175, 178, 179, 180, 189 and 190. These items (85) are not fitting because they reflect only small gains. So these items were rejected.

Table-1
The item facility (IF) index calculated from pretest and posttest

Ite	post	Di	stu	IF post	pr	Di	stu	IF pre	Ite	post	Di	stu	IF post	pr	Di	stu	IF pre
1	12	/	15	0.80	5	/	15	0.33	51	14	/	15	0.47	6	/	15	0.40
2	4	/	15	0.27	6	/	15	0.40	52	12	/	15	0.80	4	/	15	0.27
3	8	/	15	0.53	6	/	15	0.40	53	7	/	15	0.47	6	/	15	0.40
4	7	/	15	0.47	5	/	15	0.33	54	6	/	15	0.40	7	/	15	0.47
5	7	/	15	0.47	9	/	15	0.60	55	2	/	15	0.13	4	/	15	0.27
6	13	/	15	0.87	6	/	15	0.40	56	4	/	15	0.27	7	/	15	0.47
7	14	/	15	0.93	7	/	15	0.47	57	10	/	15	0.67	2	/	15	0.13
8	12	/	15	0.80	6	/	15	0.40	58	8	/	15	0.53	7	/	15	0.47
9	14	/	15	0.93	6	/	15	0.40	59	6	/	15	0.40	5	/	15	0.33
10	14	/	15	0.93	5	/	15	0.33	60	7	/	15	0.47	6	/	15	0.40
11	11	/	15	0.73	5	/	15	0.33	61	9	/	15	0.60	8	/	15	0.53
12	12	/	15	0.80	8	/	15	0.53	62	13	/	15	0.87	12	/	15	0.80
13	8	/	15	0.53	7	/	15	0.47	63	12	/	15	0.8	5	/	15	0.33
14	5	/	15	0.33	11	/	15	0.73	64	7	/	15	0.47	11	/	15	0.73
15	9	/	15	0.60	2	/	15	0.13	65	10	/	15	0.67	4	/	15	0.27
16	8	/	15	0.53	3	/	15	0.20	66	13	/	15	0.87	11	/	15	0.73
17	9	/	15	0.60	8	/	15	0.53	67	13	/	15	0.87	2	/	15	0.13
18	12	/	15	0.80	6	/	15	0.40	68	3	/	15	0.20	11	/	15	0.73
19	8	/	15	0.53	6	/	15	0.40	69	13	/	15	0.87	12	/	15	0.80
20	6	/	15	0.40	5	/	15	0.33	70	1	/	15	0.07	2	/	15	0.13
21	11	/	15	0.73	9	/	15	0.60	71	14	/	15	0.93	6	/	15	0.40
22	13	/	15	0.87	6	/	15	0.40	72	12	/	15	0.80	6	/	15	0.40
23	6	/	15	0.40	7	/	15	0.47	73	7	/	15	0.47	6	/	15	0.40
24	11	/	15	0.73	6	/	15	0.40	74	14	/	15	0.93	7	/	15	0.47
25	8	/	15	0.53	6	/	15	0.40	75	10	/	15	0.67	3	/	15	0.20
26	10	/	15	0.67	8	/	15	0.53	76	14	/	15	0.93	13	/	15	0.87
27	8	/	15	0.53	3	/	15	0.20	77	12	/	15	0.80	5	/	15	0.33
28	5	/	15	0.33	10	/	15	0.67	78	3	/	15	0.20	10	/	15	0.67
29	10	/	15	0.67	6	/	15	0.40	79	14	/	15	0.93	5	/	15	0.33
30	7	/	15	0.47	3	/	15	0.13	80	10	/	15	0.67	4	/	15	0.27
31	7	/	15	0.47	3	/	15	0.13	81	14	/	15	0.87	2	/	15	0.13
32	12	/	15	0.80	6	/	15	0.40	82	11	/	15	0.73	3	/	15	0.20
33	10	/	15	0.67	5	/	15	0.33	83	8	/	15	0.53	2	/	15	0.13
34	11	/	15	0.73	6	/	15	0.40	84	6	/	15	0.40	7	/	15	0.47
35	4	/	15	0.27	6	/	15	0.40	85	8	/	15	0.53	9	/	15	0.60
36	8	/	15	0.53	3	/	15	0.20	86	12	/	15	0.80	4	/	15	0.27
37	6	/	15	0.40	5	/	15	0.33	87	13	/	15	0.87	4	/	15	0.27
38	11	/	15	0.73	6	/	15	0.40	88	13	/	15	0.87	5	/	15	0.33
39	11	/	15	0.73	6	/	15	0.40	89	8	/	15	0.53	7	/	15	0.47
40	7	/	15	0.47	8	/	15	0.53	90	7	/	15	0.47	5	/	15	0.33
41	11	/	15	0.73	5	/	15	0.47	91	5	/	15	0.33	6	/	15	0.40
42	12	/	15	0.80	6	/	15	0.40	92	7	/	15	0.47	9	/	15	0.60
43	7	/	15	0.47	5	/	15	0.33	93	10	/	15	0.67	8	/	15	0.53
44	7	/	15	0.47	6	/	15	0.40	94	3	/	15	0.20	4	/	15	0.27
45	11	/	15	0.73	3	/	15	0.20	95	11	/	15	0.73	3	/	15	0.20
46	10	/	15	0.67	2	/	15	0.13	96	2	/	15	0.13	9	/	15	0.60
47	5	/	15	0.47	10	/	15	0.67	97	14	/	15	0.93	4	/	15	0.27
48	6	/	15	0.40	5	/	15	0.33	98	11	/	15	0.73	7	/	15	0.20
49	8	/	15	0.53	6	/	15	0.40	99	5	/	15	0.33	4	/	15	0.27
50	7	/	15	0.93	8	/	15	0.53	100	8	/	15	0.53	9	/	15	0.60

(Continue)

Ite	post	Di	stu	IF post	pr	Di	stu	IF pre	Ite	post	Di	stu	IF post	pr	Di	stu	IF pre
101	14	/	15	0.93	4	/	15	0.66	146	12	/	15	0.80	10	/	15	0.13
102	10	/	15	0.67	9	/	15	0.07	147	14	/	15	0.93	13	/	15	0.06
103	12	/	15	0.80	3	/	15	0.60	148	14	/	15	0.93	5	/	15	0.60
104	14	/	15	0.93	5	/	15	0.60	149	15	/	15	1.00	13	/	15	0.13
105	13	/	15	0.87	3	/	15	0.67	150	11	/	15	0.33	3	/	15	0.13
106	3	/	15	0.20	10	/	15	-0.47	151	14	/	15	0.93	6	/	15	0.53
107	12	/	15	0.80	4	/	15	0.53	152	13	/	15	0.87	7	/	15	0.40
108	13	/	15	0.87	5	/	15	0.54	153	13	/	15	0.87	12	/	15	0.07
109	14	/	15	0.93	12	/	15	0.13	154	12	/	15	0.80	6	/	15	0.40
110	12	/	15	0.80	6	/	15	0.40	155	10	/	15	0.67	4	/	15	0.40
111	11	/	15	0.73	10	/	15	0.06	156	11	/	15	0.73	6	/	15	0.33
112	13	/	15	0.87	4	/	15	0.60	157	14	/	15	0.93	6	/	15	0.53
113	5	/	15	0.33	12	/	15	-0.47	158	12	/	15	0.80	11	/	15	0.07
114	12	/	15	0.80	3	/	15	0.60	159	8	/	15	0.53	7	/	15	0.06
115	13	/	15	0.87	12	/	15	0.07	160	13	/	15	0.87	4	/	15	0.60
116	10	/	15	0.67	2	/	15	0.54	161	11	/	15	0.73	2	/	15	0.60
117	12	/	15	0.80	10	/	15	0.13	162	14	/	15	0.93	13	/	15	0.06
118	14	/	15	0.93	13	/	15	0.06	163	14	/	15	0.93	13	/	15	0.06
119	6	/	15	0.40	5	/	15	0.07	164	13	/	15	0.87	12	/	15	0.07
120	12	/	15	0.80	1	/	15	0.73	165	13	/	15	0.87	14	/	15	-0.06
121	8	/	15	0.53	7	/	15	0.06	166	12	/	15	0.80	3	/	15	0.60
122	6	/	15	0.40	5	/	15	0.07	167	13	/	15	0.87	14	/	15	-0.06
123	6	/	15	0.40	5	/	15	0.07	168	13	/	15	0.87	4	/	15	0.60
124	14	/	15	0.93	7	/	15	0.46	169	13	/	15	0.87	5	/	15	0.54
125	14	/	15	0.93	13	/	15	0.06	170	14	/	15	0.93	13	/	15	0.06
126	14	/	15	0.93	13	/	15	0.06	171	13	/	15	0.87	6	/	15	0.47
127	12	/	15	0.80	10	/	15	0.13	172	9	/	15	0.60	1	/	15	0.53
128	14	/	15	0.93	13	/	15	0.06	173	14	/	15	0.93	7	/	15	0.46
129	4	/	15	0.27	5	/	15	-0.06	174	11	/	15	0.73	8	/	15	0.20
130	14	/	15	0.93	4	/	15	0.66	175	6	/	15	0.40	5	/	15	0.07
131	12	/	15	0.80	3	/	15	0.60	176	11	/	15	0.73	5	/	15	0.40
132	13	/	15	0.87	2	/	15	0.74	177	13	/	15	0.87	4	/	15	0.60
133	7	/	15	0.47	5	/	15	0.14	178	6	/	15	0.40	5	/	15	0.07
134	13	/	15	0.87	2	/	15	0.74	179	6	/	15	0.40	4	/	15	0.13
135	15	/	15	1.00	13	/	15	0.13	180	4	/	15	0.27	5	/	15	-0.06
136	6	/	15	0.40	5	/	15	0.07	181	13	/	15	0.87	3	/	15	0.67
137	10	/	15	0.67	5	/	15	0.34	182	14	/	15	0.93	6	/	15	0.53
138	7	/	15	0.47	5	/	15	0.14	183	13	/	15	0.87	5	/	15	0.54
139	12	/	15	0.80	6	/	15	0.40	184	13	/	15	0.87	4	/	15	0.60
140	14	/	15	0.93	5	/	15	0.60	185	15	/	15	1.00	7	/	15	0.53
141	8	/	15	0.53	6	/	15	0.13	186	8	/	15	0.53	14	/	15	-0.40
142	13	/	15	0.87	12	/	15	0.07	187	11	/	15	0.73	5	/	15	0.40
143	14	/	15	0.93	13	/	15	0.06	188	12	/	15	0.80	7	/	15	0.33
144	14	/	15	0.93	12	/	15	0.13	189	6	/	15	0.40	4	/	15	0.13
145	15	/	15	1.00	13	/	15	0.13	190	7	/	15	0.47	5	/	15	0.14

Ite=Item, Pr=pretest, Post=posttest, Di=Divide, Stu= Students, IF= Item Facility, IF post=Item Facility posttest, IF pre=Item Facility pretest

Table-2
The difference index (DI) was calculated from item facility in pretest and posttest

Ite	post test	pre test	IF Post	Min	IF Pre	Equ	DI	Ite	Post test	Pre test	IF Post	Min	IF Pre	Equ	DI
1	12	5	0.80	-	0.33	=	0.47	51	14	6	0.47	-	0.40	=	0.07
2	4	6	0.27	-	0.40	=	-0.13	52	12	4	0.80	-	0.27	=	0.53
3	8	6	0.53	-	0.40	=	0.13	53	7	6	0.47	-	0.40	=	0.07
4	7	5	0.47	-	0.33	=	0.14	54	6	7	0.40	-	0.47	=	-0.07
5	7	9	0.47	-	0.60	=	-0.13	55	2	4	0.13	-	0.27	=	-0.14
6	13	6	0.87	-	0.40	=	0.47	56	4	7	0.27	-	0.47	=	-0.20
7	14	7	0.93	-	0.47	=	0.46	57	10	2	0.67	-	0.13	=	0.54
8	12	6	0.80	-	0.40	=	0.40	58	8	7	0.53	-	0.47	=	0.06
9	14	6	0.93	-	0.40	=	0.53	59	6	5	0.40	-	0.33	=	0.07
10	14	5	0.93	-	0.33	=	0.60	60	7	6	0.47	-	0.40	=	0.07
11	11	5	0.73	-	0.33	=	0.40	61	9	8	0.60	-	0.53	=	0.07
12	12	8	0.80	-	0.53	=	0.27	62	13	12	0.87	-	0.80	=	0.07
13	8	7	0.53	-	0.47	=	0.06	63	12	5	0.8	-	0.33	=	0.47
14	5	11	0.33	-	0.73	=	-0.40	64	7	11	0.47	-	0.73	=	-0.26
15	9	2	0.60	-	0.13	=	0.47	65	10	4	0.67	-	0.27	=	0.40
16	8	3	0.53	-	0.20	=	0.33	66	13	11	0.87	-	0.73	=	0.14
17	9	8	0.60	-	0.53	=	0.07	67	13	2	0.87	-	0.13	=	0.74
18	12	6	0.80	-	0.40	=	0.40	68	3	11	0.20	-	0.73	=	-0.53
19	8	6	0.53	-	0.40	=	0.13	69	13	12	0.87	-	0.80	=	0.07
20	6	5	0.40	-	0.33	=	0.07	70	1	4	0.07	-	0.13	=	-0.06
21	11	9	0.73	-	0.60	=	0.13	71	14	6	0.93	-	0.40	=	0.53
22	13	6	0.87	-	0.40	=	0.47	72	12	6	0.80	-	0.40	=	0.40
23	6	7	0.40	-	0.47	=	-0.07	73	7	6	0.47	-	0.40	=	0.07
24	11	6	0.73	-	0.40	=	0.33	74	14	7	0.93	-	0.47	=	0.46
25	8	6	0.53	-	0.40	=	0.13	75	10	3	0.67	-	0.20	=	0.47
26	10	8	0.67	-	0.53	=	0.14	76	14	13	0.93	-	0.87	=	0.06
27	8	3	0.53	-	0.20	=	0.33	77	12	5	0.80	-	0.33	=	0.47
28	5	10	0.33	-	0.67	=	-0.34	78	3	10	0.20	-	0.67	=	-0.47
29	10	6	0.67	-	0.40	=	0.27	79	14	5	0.93	-	0.33	=	0.60
30	7	3	0.47	-	0.13	=	0.34	80	10	4	0.67	-	0.27	=	0.40
31	7	3	0.47	-	0.13	=	0.34	81	14	2	0.87	-	0.13	=	0.74
32	12	6	0.80	-	0.40	=	0.40	82	11	3	0.73	-	0.20	=	0.53
33	10	5	0.67	-	0.33	=	0.34	83	8	2	0.53	-	0.13	=	0.40
34	11	6	0.73	-	0.40	=	0.33	84	6	7	0.40	-	0.47	=	-0.07
35	4	6	0.27	-	0.40	=	-0.13	85	8	9	0.53	-	0.60	=	-0.07
36	8	3	0.53	-	0.20	=	0.33	86	12	4	0.80	-	0.27	=	0.53
37	6	5	0.40	-	0.33	=	0.07	87	13	4	0.87	-	0.27	=	0.60
38	11	6	0.73	-	0.40	=	0.33	88	13	5	0.87	-	0.33	=	0.54
39	11	6	0.73	-	0.40	=	0.33	89	8	7	0.53	-	0.47	=	0.06
40	7	8	0.47	-	0.53	=	-0.06	90	7	5	0.47	-	0.33	=	0.14
41	11	5	0.73	-	0.47	=	0.26	91	5	6	0.33	-	0.40	=	-0.07
42	12	6	0.80	-	0.40	=	0.40	92	7	9	0.47	-	0.60	=	-0.13
43	7	5	0.47	-	0.33	=	0.14	93	10	8	0.67	-	0.53	=	0.14
44	7	6	0.47	-	0.40	=	0.07	94	3	4	0.20	-	0.27	=	-0.07
45	11	3	0.73	-	0.20	=	0.53	95	11	3	0.73	-	0.20	=	0.53
46	10	2	0.67	-	0.13	=	0.54	96	2	9	0.13	-	0.60	=	-0.47
47	7	10	0.47	-	0.67	=	-0.20	97	14	4	0.93	-	0.27	=	0.66
48	6	5	0.40	-	0.33	=	0.07	98	11	7	0.73	-	0.20	=	0.53
49	8	6	0.53	-	0.40	=	0.13	99	5	4	0.33	-	0.27	=	0.06
50	7	8	0.93	-	0.53	=	0.40	100	8	9	0.53	-	0.60	=	-0.07

Ite	post test	pre test	IF Post	Min	IF Pre	Equ	DI	Ite	post test	pre test	IF Post	Min	IF Pre	Equ	DI
101	14	4	0.93	-	0.27	=	0.66	146	12	10	0.80	-	0.67	=	0.13
102	10	9	0.67	-	0.60	=	0.07	147	14	13	0.93	-	0.87	=	0.06
103	12	3	0.80	-	0.20	=	0.60	148	14	5	0.93	-	0.33	=	0.60
104	14	5	0.93	-	0.33	=	0.60	149	15	13	1.00	-	0.87	=	0.13
105	13	3	0.87	-	0.20	=	0.67	150	11	3	0.33	-	0.20	=	0.13
106	3	10	0.20	-	0.67	=	-0.47	151	14	6	0.93	-	0.40	=	0.53
107	12	4	0.80	-	0.27	=	0.53	152	13	7	0.87	-	0.47	=	0.40
108	13	5	0.87	-	0.33	=	0.54	153	13	12	0.87	-	0.80	=	0.07
109	14	12	0.93	-	0.80	=	0.13	154	12	6	0.80	-	0.40	=	0.40
110	12	6	0.80	-	0.40	=	0.40	155	10	4	0.67	-	0.27	=	0.40
111	11	10	0.73	-	0.67	=	0.06	156	11	6	0.73	-	0.40	=	0.33
112	13	4	0.87	-	0.27	=	0.60	157	14	6	0.93	-	0.40	=	0.53
113	5	12	0.33	-	0.80	=	-0.47	158	12	11	0.80	-	0.73	=	0.07
114	12	3	0.80	-	0.20	=	0.60	159	8	7	0.53	-	0.47	=	0.06
115	13	12	0.87	-	0.80	=	0.07	160	13	4	0.87	-	0.27	=	0.60
116	10	2	0.67	-	0.13	=	0.54	161	11	2	0.73	-	0.13	=	0.60
117	12	10	0.80	-	0.67	=	0.13	162	14	13	0.93	-	0.87	=	0.06
118	14	13	0.93	-	0.87	=	0.06	163	14	13	0.93	-	0.87	=	0.06
119	6	5	0.40	-	0.33	=	0.07	164	13	12	0.87	-	0.80	=	0.07
120	12	1	0.80	-	0.07	=	0.73	165	13	14	0.87	-	0.93	=	-0.06
121	8	7	0.53	-	0.47	=	0.06	166	12	3	0.80	-	0.20	=	0.60
122	6	5	0.40	-	0.33	=	0.07	167	13	14	0.87	-	0.93	=	-0.06
123	6	5	0.40	-	0.33	=	0.07	168	13	4	0.87	-	0.27	=	0.60
124	14	7	0.93	-	0.47	=	0.46	169	13	5	0.87	-	0.33	=	0.54
125	14	13	0.93	-	0.87	=	0.06	170	14	13	0.93	-	0.87	=	0.06
126	14	13	0.93	-	0.87	=	0.06	171	13	6	0.87	-	0.40	=	0.47
127	12	10	0.80	-	0.67	=	0.13	172	9	1	0.60	-	0.07	=	0.53
128	14	13	0.93	-	0.87	=	0.06	173	14	7	0.93	-	0.47	=	0.46
129	4	5	0.27	-	0.33	=	-0.06	174	11	8	0.73	-	0.53	=	0.20
130	14	4	0.93	-	0.27	=	0.66	175	6	5	0.40	-	0.33	=	0.07
131	12	3	0.80	-	0.20	=	0.60	176	11	5	0.73	-	0.33	=	0.40
132	13	2	0.87	-	0.13	=	0.74	177	13	4	0.87	-	0.27	=	0.60
133	7	5	0.47	-	0.33	=	0.14	178	6	5	0.40	-	0.33	=	0.07
134	13	2	0.87	-	0.13	=	0.74	179	6	4	0.40	-	0.27	=	0.13
135	15	13	1.00	-	0.87	=	0.13	180	4	5	0.27	-	0.33	=	-0.06
136	6	5	0.40	-	0.33	=	0.07	181	13	3	0.87	-	0.20	=	0.67
137	10	5	0.67	-	0.33	=	0.34	182	14	6	0.93	-	0.40	=	0.53
138	7	5	0.47	-	0.33	=	0.14	183	13	5	0.87	-	0.33	=	0.54
139	12	6	0.80	-	0.40	=	0.40	184	13	4	0.87	-	0.27	=	0.60
140	14	5	0.93	-	0.33	=	0.60	185	15	7	1.00	-	0.47	=	0.53
141	8	6	0.53	-	0.40	=	0.13	186	8	14	0.53	-	0.93	=	-0.40
142	13	12	0.87	-	0.80	=	0.07	187	11	5	0.73	-	0.33	=	0.40
143	14	13	0.93	-	0.87	=	0.06	188	12	7	0.80	-	0.47	=	0.33
144	14	12	0.93	-	0.80	=	0.13	189	6	4	0.40	-	0.27	=	0.13
145	15	13	1.00	-	0.87	=	0.13	190	7	5	0.47	-	0.33	=	0.14

Ite=Item, IF=Item Facility, MIN=Minus, Equ=Equals, DI=Difference Index

Conclusion

Totally 190 items were used for item analysis of criterion referenced test. The difference index (DI) was calculated. 85 items are not fitting because they reflect only small gains. So these items were rejected. Finally 105 items (80+25) were accepted for the history achievement test. This item analysis used to construct the history achievement test.

Acknowledgement

References

1. http://en.wikipedia.org/wiki/Educational_evaluation dated 01.06.2013

2. Keeves P. John., Educational Research Methodology and Measurement: An International Handbook (Second Edition), Hambleton R.K., (Eds.), Criterion Referenced Measurement *(719-725)*, Stocking M.L., (Eds.), Item Response Theory (836-839) Choppin B. H., (Eds.), Objective Tests (771-775) UK, USA, Japan, Elsevier Science Ltd **(1997)**

3. Tom Kubiszyn and Gary Borich., Educational Testing and Measurement Classroom Application and Practice, (7th Edition) Singapore, John. Wiley and Sons (Asia) Pvt. Ltd **(2003)**

4. http://www.education.com/reference/article/item-analysis/ dated 06.06.2013

5. Gerard Guthrie., Basic Research Methods an entry to Social Science Research, Los Angeles, London, New Delhi, Singapore, Washington DC, SAGE Publications **(2010)**

6. James Dean Brown., Criterion-referenced item analysis (the difference index and *B*-index), *SHIKEN: The JALT testing and Evaluation SIG Newsletter*, **7(3)**, 18-24 **(2003)**