CHAPTER - 3

REVIEW OF THE RELATED LITERATURE

3.1 Introduction

3.2 Source

3.3 Hambleton and Eignor’s Review

3.4 Review of CRTs given MMY

3.4.1 Buros’s Review of CRT in MMY

3.4.2 Bhogayata’s Review of CRT in MMY

3.5 Review of CRTs: Developed in India

3.5.1 Verma’s Test

3.5.2 Raitha’s Test

3.5.3 Singh et.al.’s Test

3.6 Review of CRTs: Developed in Abroad

3.6.1 Smith et.al.’s Test

3.6.2 Ganopola’s Test

3.6.3 Royce’s Test

3.6.4 D.Regina’s Test

3.7 Review of Studies related to CRT in Abroad

3.7.1 Rovinelli, R.J.’s Study

3.7.2 Thrash, S.K.’s Study

3.7.3 Wright, E.B.’s Study

3.7.4 Berk, R.A.’s Study

3.7.5 Eigner’s Study

3.7.6 Peng’s Study
3.7.7 Cristensen’s Study

3.7.8 Peter’s Study

3.7.9 J.Jen’s Study

3.7.10 Aiken’s Study

3.8 Review of Articles of Journals published in India

3.8.1 Bhogayata’s Article

3.8.2 Raitha’s Article

3.9 Review of M.Ed. Dissertations published in Gujarat

3.10 Review of literature reviewed

3.11 Significance and Special Characteristics of the Present Study

Footnotes
CHAPTER - 3

REVIEW OF THE RELATED LITERATURE

3.1 Introduction

Man is the only animal that can take advantage of the knowledge which has been preserved or accumulated through the centuries. Human knowledge has the three phases: preservation, transmission and advancement. This fact is of particular importance in research which operates as a continuous function of ever-close approximation to the truth.

According to John W. Best:

"Practically all human knowledge can be found in books and libraries and money in the Banks. Unlike other animals that must start a new with each generation, man builds upon the accumulated and recorded knowledge of the past." ¹

It is known to all that the study of the past is useful to the present generation in several ways. This is true in each walk of life. In the various fields of research, the posterior researchers can learn useful lessons from the sincere efforts made by the pioneering researchers.

According to W. R Borrg:

"The literature in any field forms the foundation upon which all future work will be built. If we fail to build the foundation of knowledge provided by the review of literature our work is likely to be shallow and native and will often duplicate work that has already been done better by some one else." ²

The researchers of criterion-referenced tests (CRTS) also are not exception to it Therefore it is necessary to take into consideration the studies done upto
the present day in the field of CRT before the consideration of construction and standardization of CRT. The work done in the field of CRT in the past serves as a guideline to the present researchers and helps a researcher to know the merits and demerits of the study made in the past. Consequently, a researcher tries to overcome the limitations and to avoid the pitfalls. Such a study helps a researcher in giving a picture regarding funds, energy and time required for an attempt in the direction.

Looking to the past, it may be noted that the conceptual fore-runners are evident in the work of Rev. George Fisher (1864), Thorndike (1913), Flanagan (1951) and Eble (1962) but the term criterion-referenced measure (CRM) was first used by Glaser and Klaus (1962) in their description of the assessment of the performance of man machines systems. (Gray3 (1978); Nitko4 (1980))

A year later the name ‘CRT’ was mentioned for the first time by Glaser in his articles on ‘Instructional Technology and the Measurement of learning Outcome’ published in 1963 in the well-know periodical ‘American Psychologist.’ Since then, over 700 papers on the topic have been published and the scope and direction of educational testing have been changed dramatically [Hambleton (1982)]

There has been an appreciable amount of work in the field of evaluation focussed on evaluation of effective learning during the last ten years. The eminent educationists like Popham, Hambleton, Nitko, Millman and Berk have presented their views relating to the role of criterion-referenced testing and construction of domain-referenced test. Though there are divergent views held by different scholars regarding CRT , yet the fact remains that each one of them has the same goal in view and is keen to relate testing to expected level of performance or criteria of the intended learning outcomes and thus emphasizing criterion-referenced testing.
The origin of CRTs lies in minimum competency testing movement in California. It was found from the trial of this movement that the educationists still required to adopt better measurable standard of proficiency in the basic skill areas. Moreover, the origin of CRT is largely from the emphasis on behavioural objectives, individualization of instruction, sequancing of learning and the concept of mastery learning. Added to this, the prevailing belief that norm-referenced tests promote unhealthy competition and leaves behind harmful psychological impressions on low scoring students, is another factor which contributes to the origin of CRTs.

3.2 Sources

As the subject of CRT is a recent innovation particularly in developing countries like India, the books exclusively devoted to CRT are comparatively much limited in number and even they are rarely available in India. The present chapter deals with the review of CRT literature upto the present day. Therefore, the following sources were consulted to fulfill the purpose:

(i) Articles published in foreign leading journals;
(ii) The Eighth Mental Measurement Yearbook (MMY) edited by Buros (1978);
(iii) Dissertation Abstracts International (DAI) (1974-1987);
(iv) Three surveys compiled by Bucy (1973;1979;1983);
(v) The International Encyclopedia of Education (IEE) (1985);
(vi) Criterion-Referenced Assessment-An Annotated Bibliography &

3.3 Hambleton and Eignor’s Review

Hambleton’s (1982) article on ‘Advances in Criterional-Referenced Testing Technology’ described the review of CRTs done by him and Eignor. Eleven of
the more popular commercially available CRTs were evaluated by them. They used evaluating system in their work. It was their considered opinion that there was substantial room for improvement in the development of CRTs and in the reporting of content and technical information. They found the tests under their review wanting in one or more than one characteristics. None of the eleven CRTs possessed all the characteristics required for a CRT. On the basis of the above mentioned review and from the available literature on CRT, it could be seen that an ideal CRT should have the following psychometric Characteristics.

(1) Behavioural or content domains should be well-defined.
(2) Items should be reviewed empirically for their instructional sensitivity.
(3) Items should be reviewed logically for item-objective congruence.
(4) Items should be randomly selected for each domain or each stratum of a domain.
(5) A cut off score for each domain should be set purposefully and not as an only inevitable psychometric property of a CRT.
(6) A cut off score for each domain should be set by employing one of the different technical procedures for it.
(7) A domain score for each domain should be estimated.
(8) Some special reliability procedures for a CRT should be adopted.
(9) Special procedures for establishing validity of a CRT should be employed.

These psychometric characteristics have been utilised by the investigator for the review of CRTs given in MMY.

3.4 Review of CRTs in MMY

3.4.1 Buros's Review of CRT in MMY

The forty five CRTs, described and primarily reviewed in the Eighth Mental Measurement Yearbook, were review by the investigator in the light of the above mentioned psychometric characteristics.
Of the forty five CRTs, twenty two related to reading, fourteen to mathematics, three to language arts, two each to social studies and reading mathematics as well as one each to home economics and Spanish. The construction of these CRTs covers a period from 1966 to 1977 and standards from preprimary to XII. Most of the tests were developed by the individual developers and very few through institutions. On analysing the forty five CRTs, it was found that,

(i) seven out of the forty five CRTs were developed from well-defined behavioural domain. Out of these seven, four were Instructional Objective Exchange (IOX) and they were developed by employing 'amplified objective' technique;

(ii) only one of them was developed after selecting the items randomly from behavioral domains;

(iii) in twenty one of forty five CRTs, cut off scores were determined but the technical procedure was not adopted for determining the cut off score;

(iv) traditional item indices were reported only in four of forty five CRTs;

(v) in two tests items were logically reviewed for item objective congruence;

(vi) only two tests contained reliability data report;

(vii) validity evidences were reported in three of the tests.

(viii) the norms were established in four of forty five CRTs but these norms were not utilized even in one test for standard setting.

Over and above this, in none of these forty five CRTs,

(i) domain score was estimated;

(ii) test items were empirically reviewed for instructional sensitivity;

(iii) standard error of measurement was reported for estimating domain score;

(iv) decision consistency index was reported and

(v) validity evidence for standard setting classification was reported.
It could be seen on the evidence of the analytical study that all the forty five CRTs were not CRTs but they were only ‘slogan’ CRTs. Popham (cited by Nitko, 1980) called them ‘cloud-referenced tests’. Tindal et.al. (1985) commented that among published CRTs, there was scant conceptual and empirical support for technical adequacy. The tests based on poorly articulated behavioural objectives might be called pseudo referenced tests and represented a misapplication of the idea of criterion-referencing (Nitko (1980)).

3.4.2 Bhogayata’s Review of CRTs in MMY (1987)

Bhogayata has reviewed forty five CRTs described in the Eighth Mental Measurement Year Book. From them, fourteen CRTs were of Mathematics subject. These fourteen CRTs were reviewed according in the compound review method of Jekson (1980). These fourteen CRTs were reviewed according to following questions:

1. Were the behavioural or content domains well-defined?
2. Were the items selected randomly from the item-universe?
3. Were the portable cut-off scores estimated?
4. Were the cut-off scores set up?
5. Was the method appropriate for setting the cut-off scores?
6. Were the norms of items given traditionally?
7. Were the items logically reviewed for item-objective congruence index (IOCI)?
8. Were the items empirically reviewed for their instructional sensitivity?
9. Were the reliability index found?
10. Was the standard error estimated for the domain score?
11. Was the Kappa Index calculated?
12. Were the validity index found?

62
In the following Table 3.1 a List of fourteen CRTS of Mathematics reviewed according to above questions is given. In the Table 3.2 the rating according to above fourteen CRTs as per above given questions.

**Table-3.1**

**Review of Fourteen CRTs of Mathematics**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Serial No. of Test</th>
<th>Test Description</th>
<th>Developer or the Institute</th>
<th>Year</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>251</td>
<td>Analysis of skills Maths</td>
<td>Ander Helter</td>
<td>74-76</td>
<td>1-8</td>
</tr>
<tr>
<td>2</td>
<td>253</td>
<td>Basic Education Skills Inventory : Mathematics.</td>
<td>Adamson et.al.</td>
<td>72-73</td>
<td>Kgn-6</td>
</tr>
<tr>
<td>3</td>
<td>263</td>
<td>Diagnosis : An Instructional Aid Mathematics.</td>
<td>Se.Research Association</td>
<td>72-74</td>
<td>1-6</td>
</tr>
<tr>
<td>4</td>
<td>264</td>
<td>Diagnostc Mathematics Inventory.</td>
<td>Gessel</td>
<td>71-75</td>
<td>1-5</td>
</tr>
<tr>
<td>5</td>
<td>270</td>
<td>Fountain Valley Teacher Support System in Maths</td>
<td>Eweig Association</td>
<td>72-74</td>
<td>Kgn-8</td>
</tr>
<tr>
<td>6</td>
<td>274</td>
<td>Individual Pupil Monitoring System Mathematics.</td>
<td>Houghton Miffin Co.</td>
<td>73</td>
<td>1-8</td>
</tr>
<tr>
<td>9</td>
<td>279</td>
<td>Mathematics : IOX objectives based tests.</td>
<td>Moscow</td>
<td>73-76</td>
<td>Kgn-9</td>
</tr>
<tr>
<td>10</td>
<td>282</td>
<td>Mathematics Topic Tests: Elementary Level.</td>
<td>Marrison</td>
<td>74</td>
<td>4-9</td>
</tr>
<tr>
<td>11</td>
<td>287</td>
<td>Objectives Referenced Bank of Items and Tests in Maths</td>
<td>CBT / McGraw Hills</td>
<td>75</td>
<td>Kgn-12</td>
</tr>
<tr>
<td>12</td>
<td>293</td>
<td>Test of Achievement in Basic Skills :</td>
<td>Young and Knapp</td>
<td>70-76</td>
<td>Kgn-12</td>
</tr>
<tr>
<td>13</td>
<td>303</td>
<td>Basic Arithmetics skills Evaluation.</td>
<td>May and Hood</td>
<td>73-74</td>
<td>1-9</td>
</tr>
</tbody>
</table>

* Serial number of the test is as per Eighth Mental Measurement Year Book.
### Table 3.2
Rating of Fourteen CRTs of Mathematics According to Twelve Questions of Jekson's Method (Characteristics of CRTs of Mathematics)*

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Serial No. of Test</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
<th>Q12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>251</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>253</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>263</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>264</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>270</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>274</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>275-6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>278</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>279</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>282</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>287</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>293</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>303</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>312</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*1 is for 'Yes' and 0 is for 'No'*

**Analysis and Interpretation of CRTs given in The Eighth Mental Measurement Year Book:**

1. Out of 14 only 2 CRTs were constructed on well-defined behaviour. Hence, only 14% CRTs were constructed on well-defined behaviour. Remaining 86% CRTs were not constructed on well-defined behaviour. So they are not called CRT in true sense.

2. Items were not randomly from any item-universe, so it was difficult to estimate domain score.

3. Domain score estimation was not done in any CRT.

4. Cut off score was estimated in seven CRTs out of fourteen CRTs.
5 Cut off score was not estimated systematically in any CRT.
6 Statistical values of items were given in only one CRT.
7 Logical review was not applied to find IOCI in any CRT because domains were not well-defined. This was a serious mistake.
8 Empirical review was not applied to find instructional sensitivity in any CRT.
9 Standard error of measurement of domain score, reliability index and Kappa index were not estimated in any CRT.

3.5 Review of CRTs: Developed in India

3.5.1 Verma’s Test

Verma\(^9\) developed and validated a CRT as a part of his Ph.D. degree. Fifty Hindi Suffixes were selected for that CRT. The empirical review of the test items was conducted for difficulty value and instructional sensitivity. The logical review of items was conducted by a panel of three judges. For establishing content validity, the pre-to-post test gain (PPG) technique was adopted. Mastery-non mastery classification decision was adopted for establishing the reliability. The standard was set at 0.85 percent achievement. The investigator found the following limitations in Verma’s test:

(i) Instead of generating the test items from well defined behavioural domains, they were generated on the basis of objectives; 
(ii) Test items were not randomly selected from the domains; 
(iii) The results of the logical review by the panel of the judges were not reported; 
(iv) No report was made regarding procedure adopted for setting standard; 
(v) The domain score was not estimated and 
(vi) The test lacked in the report regarding any evidence for establishing construct and criterion-related validity.
3.5.2 Raitha's Test

Raitha B.C.\textsuperscript{10} of Bhavnagar University, Bhavnagar (India), developed CRT for her Ph.D. degree in 1989. The title of her study was: 'Development and validation of Criterion-Referenced Test on Vowel Coalition (Svar Sandhi) in Sanskrit Language.

Sixteen domains (on Svar Sandhi) were developed adopting Hively's item form technique. The logical review was conducted to assess the congruence between items and objectives. The empirical review was conducted to examine the instructional sensitivity of items. Then the items were selected randomly from the pool of items.

Content validity, construct validity and criterion related validity were established.

Two types of reliability—the domain score estimation and mastery classification decision—were established. The cut-off score was determined for all the behavioural domains of both the parallel forms.

3.5.3 Singh's Test

Pritam Singh\textsuperscript{11} developed and validated a CRT on the subject of environment for standard III but that test was found lacking in all the psychometric characteristics reported earlier in Section 2.3. Hence, it was not a CRT in the true sense.

3.6 Review of CRTs : Developed in India and Abroad

Very few CRTs have been developed and validated in India because they are at the infancy stage.

The summary of the CRTs was available from the foreign articles and dissertation abstracts. Such CRTs having both the aspects developmental and validation were available in a much limited number. Moreover the available
number of the studies discussing most of the aspects of a CRT was much less than those discussing only one or other aspect of CRT.

3.6.1 Smith et al.'s Test

Smith et al. developed and validated CRTs entitled, 'Standard Achievement Recording Systems (STARS)' for the measurement of reading and writing achievement of first to six graders. Those tests were developed from well-defined behavioural domains. For selecting the items from the domains R-S model was employed. The decision theoretic approach was implemented. Internal consistancy was found in the test. The tests were Criterion-related and valid measures. The following limitations of STARS come to the notice of the investigator:

(i) The test items were not logically reviewed for item-objective congruence; (ii) The test items were not empirically reviewed for instructional sensitivity; (iii) The standard setting of 100 percent was not procedural and (iv) There was no report about an evidence for establishing construct validity.

3.6.2 Ganopola's Test

Ganopola made a study for this Ed.D. degree of California University under the supervision of Popham. The title of his study was, 'The construction and validation of a Criterion-Referenced Test of Fundamental Reading Competencies.' His study progressed through three principal stages:

(i) The selection of reading competencies to be assessed,

(ii) The generation of test items to measure those competencies and

(iii) The establishment of the validity and reliability of the instrument.

An extensive pool of tentative competencies was gathered by drawing upon such sources like National Assessment of Educational Programme. He
established a viable strategy for selecting competencies from the universe of plausible reading competencies for assessment. Test specification represented the set of rules to generate items for a CRT. Fundamental Reading Competencies Test (FRCT) consisted of two types of validity-descriptive validity and domain-selection validity. Three procedures were used to assess the reliability of FRCT. A test-retest reliability procedure, an equivalent forms reliability procedure and a combination of these procedures. In assessing the reliability of the test, reliability was considered and calculated from two rather distinct approaches.

(i) The traditional, co-relational approach typically used in computing estimates of norm-referenced measures and

(ii) A decision-consistency approach suitable for use with CRMs.

Reliability estimates were calculated for each of the subtests as well as the test as a whole. The main limitations of FRCT as noticed by the investigator, were as follows:

(1) The test items were not logically and empirically reviewed;

(2) The evidence of the determination of cut off score was not reported;

(3) Construct as well as criterion-related validity were not established and

(4) The domain score was not estimated.

3.6.3 Royce's Test\textsuperscript{14}

Royce under the guidance of McCurly, developed and validated a diagnostic criterion-referenced test of science process for students enrolled in high school Biology course for the Omaha Public Schools, Omaha, Nebraska. That project consisted of three phases viz.
(i) the development of Biology Test of Science Process (BTSP)

(ii) the development of the support for the BTSP included a computer programme for reporting results and a test manual for interpreting results and

(iii) validation of BTSP.

In BTSP development, the selection of objectives, domain specification and the generation of an item-pool were carried out as well as two techniques (phi co-efficient and the Cox-Vargas pretest post test difference index) of item analysis were utilized.

A computer programme was developed to fulfil the purpose of the diagnostic nature of criterion-referenced test. As a part of the diagnostic functions of the BTSP, an analysis of biology texts and related media was included in the test manual to identify activities which involved students in science.

The validation of BTSP involved an analysis of item quantity, content and concurrent validity as well as Subkoviak reliability index.

The main limitation of the test under review was that the evidence of construct validity was not reported.

3.6.4 D.Regina’s Test

D.Regina made a study entitled, ‘A Criterion-Referenced Instrument of Functional Mathematics Skills in Moderately and Severely Retarded Adults.’

In this CRT, the domain was mapped out, objective were generated as well as categorized into 11 subtests and item specifications were written as well as submitted to a panel of judges for logical review. The test was piloted and then administered to 34 sheltered workshop clients who had passed a screening pre-test.
Item analysis for each subtests was completed by computing point biserial co-relation between the item score and the respective subtest score. Content validity was established to enhance the quality of items. Concurrent validity was established by co-relating subtest scores with teacher ratings. Factor analysis on selected objectives and on the set of subtest scores supported the construct validity of the test. The reliability measure used was Subkoviak’s coefficient of agreement for CRTs. Item scores were used to predict objective scores.

The test under review can be considered a satisfactory CRT barring that the method of item selection was not reported.

3.7 Review Studies related to CRT in Abroad

3.7.1 Rovinelli R.J.’s Study


This study was designed to achieve three goals: (i) to provide a synthesis and organization of the various item validity methods that have appeared in the literature, (2) to conduct an empirical investigation of available item validity methods to determine which ones provide the most useful informations and in which situations, (3) to produce an item validation methodology that reflects the results of work in the first two areas above.

Regarding the use of item generation rules, the study attempted to organize and literature and discuss strength and weakness of the approach. The results of the empirical studies on two sets of test data clearly suggested that empirical analysis could not be used to establish item validity. Two models for the construction of C.R.Tests were presented. The basis of the models was the review of existing research and the empirical studies carried out by the author.
The procedures were outlined for item validity. They represented important theoretical contribution to the field and provided practical means for enabling practicenor to construct valid C.R.Test items.

3.7.2 Thrash, S.K.'s Study 17


The purpose of the study were:

(i) to propose a theoretical conception of C.R.Testing and to explain two basic item analysis techniques (Cox and Vargas-C.V.) and Roudabush-R theoretically with respect to this general model.

(ii) to determine the adequacy of the C.V. and R procedures using the theoretical model,

(iii) to compare three item analysis techniques, the ‘C-V, ‘R’ and the ‘Brennas and Stolurow’ (8-6) using real data.

A theoretical model for C.R.testing was proposed. It included 12 parameters that completely described the pretest-post test situation. The R and C-V indices can be explained in terms of this general model by making certain assumptions. There are two parts to the study : (1) It attempted to determine if the C-V and R indices adequately estimated the true values or if one was better than the other, and (2) If compared the three item analysis procedures.

The results of several analysis suggested that the best item analysis procedure to use with C.R.testing or pretest posttest situations is the C-V technique. It provided a reasonably accurate and stable estimate of its true value and gave very similar results when compared to the R index and B-S procedure.
3.7.3 Wright, E.B.'s Study\textsuperscript{18}


The focus of this study centered around this question: 'With a well-defined instructional system, is it effectively possible to direct student learning activity on a contingency basis so as to ensure M.S.?'

The study suggested that scores might drop if instruction on the same topic continued after mastery was reached. The implications were that a study should be done to determine an optimum time to begin differential treatment so that instructional plans could be as adaptive as possible.

3.7.4 Berk, R.A.'s Study\textsuperscript{19}


Sixteen item statistics which are recommended for use in the development of CRTs were evaluated. There were two major criteria: (1) practicability in terms of the ease of computation and interpretation, and (2) meaningfulness in the context of the development process.

Most of the statistics were based on a comparison of performance changes (pretest, posttest) or differences (uninstructed-instructed) between criterion groups. Descriptions and critiques of the difficulty, discrimination and homogeneity indices are presented in the form of a 'Consumer's Guide.'

3.7.5 Eignor's Study\textsuperscript{20}

Eignor (1979), made a study under the direction of Hambleton. The title of his study was 'Psychometric and Methodological Contributions of Criterion-
Referenced Testing Technology’. He discussed three problem areas and offered their solutions.

He pointed out the present lack of CRTs and test manuals. A considerable amount of CRTs and manuals came to be published in the last few years. A set of 39 guidelines is offered with a rationals and procedure for applying them to eleven of the most popular commercially prepared CRTs in the field.

He pointed out that the relationship of test length to reliability and validity was unexamined in the psychometric area of CRTesting research. This relationship was investigated via simulation techniques.

He pointed out the placing of emphasis on minimum competency testing in the Nation's Schools. He again pointed out that while there existed a variety of well-known methods for setting cut-off scores, there did not exist a suitable set of guidelines to help the individual concerned. He recommended the application of Modified Angoff Technique to minimum competency tests in the Insurance field.

3.7.6 Peng’s Study

Peng (1979), under the supervision of Subkoviak, carried out an investigation of Huynh's normal approximation procedure for estimating criterion-referenced reliability. In the context of mastery tests, various procedures existed which were intended to assess the reliability of a test. Among these procedures, Huynh’s single administration approach received more attention. Inspite of the comparative popularity of that procedure, the practical disadvantages arised in the actual computation, usually performed on a computer. Hence, Peng sought to explore the possibility of simplifying that procedure and his simple normal approximation yielded better estimates than the normal approximation suggested by Huynh.
3.7.7 Cristensen's Study

Cristensen (1981) in his study on 'Validation of A Criterion-Referenced Reading Skills Mastery Test for the Fifth Level of GEMS' reported the validation procedure. This test was designed to measure fifth grade student's mastery of specified reading goals and objectives at the fifth level in GEMS (Goal-Based Educational Management System). This reading skills mastery test (RSMT-S) consisted of two sub-tests (96 items). Evaluation of the test items by four reading specialists indicated higher content validity. The criterion-related validity of both subtests of RSMT was established. The RSMT had corelation coefficients in the 0.65 to 0.85 range with the ITBS vocabulary subtests. The RSMT had reliability coefficients of above 0.85 which validated the internal consistency of the test.

3.7.8 Peter's Study

Peter (1981) made a study to investigate several aspects of the problems related to standard setting on proficiency test at the University of Connecticut. He presented the problems in the following two ways:

(1) What methods should be used for standard settings.

(2) How those methods should be implemented.

The intent of his study was to provide some directions to the problems.

This study discussed the following questions regarding the above mentioned two problems: (1) Do different methods of setting standards yield different cut off scores? (2) Do different groups of judges employing the same method of standard setting yield different cut off scores? (3) Is the variability of cut off scores generated by individual judges different for the different standard setting methods? (4) Is there a relationship between prior measures of student achievement and the proficient-non proficient categorization of students based on each standard set? (5) Does the combination of empirical and a judgmental
method of standard setting alter the relationship between prior achievement and the proficiency test and improve classification results?

Two judgmental methods (Angoff, 1971, Nedelky, 1954) and an empirical method (Kriewall, 1972) were used to set standards. The judgmental standards were set by 27 content experts. The empirical standards were set by reanalysing the judgmental standards using a technique based upon the binomial distribution. The investigation of the remaining research questions was made possible by also collecting content experts' demographic data and student achievement data.

The following of the above queries were as follows:

(1) The use of different standard setting methods will produce different cut-off scores. (2) The use of more than one group of judges employing the same standard setting method will produce different cut-off scores. (3) The Angoff and Nedelsky methods produced the same cut-off scores. (4) Group dynamics and content related differences appear to be relevant factors in the setting standards judgementally. (5) Proficiency test classifications are highly related to prior student achievement regardless of which standard setting method is used.

3.7.9 J.Jen’s Study (1983)

J.Jen (1983) made a monte carlo comparative study at University of Minnesota to investigate the accuracy of four decision consistency approaches in estimating reliability of criterion-referenced tests viz. (1) Swaminathan method, (2) Huynh method; (3) Subkoviak method and (4) Marshall method. After the comparative study of these four methods, he found that Swaminathan’s estimation was closer to ‘true’ agreement with compared to the other methods. The variability of estimator for Swaminathan method was always larger than the variability of estimator for other methods.
3.7.10 Aiken's Study (1988)

Aiken (1988) constructed and evaluated reliability coefficients for CRTs. Berk (1980) had discussed many aspects for deciding reliability for CRT. From them decision varying index (Po) and Kappa Index is most useful. Therefore, to calculate these three BASIC, PASCAL and FORTRAN programmes are prepared for small and large samples. In this article, a guideline for calculating methods are given for small and large samples. With this limitations of the programme is also shown.

3.8 Review of Articles in Journals published in India

The researcher studied the articles published in the leading (prominent) educational magazines and list of those is presented here:

3.8.1 Bhogayata’s Article (1987)

He has mentioned CRT as a tool for education improvement according to the New Education Policy. It has been presented as a handy tool for eradication of the present malpractices of the examination system. Moreover he has expressed hope that this would guide both the teacher and the student in fulfilling the concept of continuous evaluation.

3.8.2 Raithatha B.C.’s Article

(1) In February 1990’s article CRT has been considered as the best, complete and result oriented teaching evaluation technique and information has been given regarding it. It holds information regarding the origin of CRT, its development meaning of terminology and a brief report about the investigators working in the field of CRT. The author believes that CRT works appreciably in the field of evaluation keeping effective learning in view plus where as NRT does not succeed in establishing norms, CRT in successful in showing the flows of the pupils and therefore ahead of NRT. This article being available in Hindi would prove as prominent guide.
In August 1990’s article teaching oriented and evaluation oriented situation has been explained. Weightage has been given to CRT by showing the inefficiency of NRT for assessing the individual’s achievement. By indicating the difference between NRT and CRT, the use of CRT as a meaningful tool for evaluation in America has been established from the first article of Glaser in 1963 to the 700 and more articles published so far various aspects of CRT have been studied and the vast area of education and the dramatic change in this direction, has been the central idea for discussion. In the end hope has been expressed that it would prove helpful for the teachers looking to its increasing importance day by day.

In January 1991’s article the reforms in the field of learning, teaching and testing has reached the classroom in evaluating the pupil. There is mention of the various techniques of item construction and item in a test for evaluation. Light has been thrown on the technique of Timon-Markley for item construction. Various techniques of item construction are appropriate according to the personal fit of the content. How to construct items for evaluating concept; that has been clarified by giving the various steps of item construction. Clarification has been done by giving an illustration of a topic of grammar - ‘Samas’. Finally by explaining the importance of this technique, item constructions have been requested to make use of this technique.

In 1991 March’s article assessment of the pupil is last but necessary for academic growth. CRT decides whether the pupil has reached the finalised (determined) measure or not criterion.

3.9 Review of M.Ed. Dissertation published in Gujarat

Review of M.Ed. Dissertation in CRTs published in Bhavnagar University, Bhavnagar in Mathematics subject are as below:
Table 3.3
Review of M.Ed. Dissertations in CRTs Published in Bhavnagar University, Bhavnagar

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Test</th>
<th>Researcher</th>
<th>Year</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Development and Validation of a Criterion-Referenced Test in Maths for addition-subtraction of fractions for Std. V.</td>
<td>Mehta Akshay Sagar, Bhavnagar Uni.</td>
<td>1988</td>
<td>V</td>
</tr>
<tr>
<td>2</td>
<td>Development and Validation of a Criterion-Referenced Test in Maths for additions</td>
<td>Shah Pallaviben, Bhavnagar Uni.</td>
<td>1989</td>
<td>III</td>
</tr>
<tr>
<td>3</td>
<td>Development and Validation of a Criterion-Referenced Test in Maths for additions</td>
<td>Shamji J. Beradia, Bhavnagar Uni.</td>
<td>1989</td>
<td>IV</td>
</tr>
<tr>
<td>4</td>
<td>Development and Validation of a Criterion-Referenced Test in Maths for addition-subtraction of integers.</td>
<td>Raval Kirikumar, Bhavnagar Uni.</td>
<td>1991</td>
<td>VI</td>
</tr>
<tr>
<td>5</td>
<td>Development and Validation of a Criterion-Referenced Test in Maths for addition-subtraction of fractions.</td>
<td>Jani Rajeshkumar, Bhavnagar Uni.</td>
<td>1991</td>
<td>V</td>
</tr>
</tbody>
</table>

The characteristics seen in above CRTs developed in Mathematics at M.Ed. level, in Bhavnagar University's Education Department as per twelve questions given in 3.4.2 is given in the following table No.3.4

Table 3.4
(Characteristics of CRTs of Mathematics at Bhavnagar Uni.)

<table>
<thead>
<tr>
<th>Sr. No. of test</th>
<th>Serial of questions for Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  Q11  Q12</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>2</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>3</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>4</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>5</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>Total</td>
<td>5 5 5 5 5 5 5 5 5 5 5 5</td>
</tr>
</tbody>
</table>

1 is for 'Yes' 0 is for 'No'

The answers of all twelve questions were obtained in 'Yes' for CRTs of Mathematics developed at Bhavnagar University only. The analysis shows that the CRTs developed in Bhavnagar University, were as per standard methods for development of CRTs.

(2) Review of M.Ed./M.Phil. Theses of Gujarat Vidyapith, Ahmedabad-14.
Table 3.5
Review of M.Ed./M.Phil. Dissertations on CRTs Published in Gujarat Vidyapith, Ahmedabad 380 014.

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Title / Researcher's Name</th>
<th>Test Domain</th>
<th>Un. Year</th>
<th>Domain Items</th>
<th>Writing Sample</th>
<th>Kappa Index</th>
<th>PPDIIOCI</th>
<th>Item Writing</th>
<th>Cutoff Score</th>
<th>Content Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Construction and Validation of a Criterion-Validity of Mathematics for Standard VI (Jayprakash G. Pandya)</td>
<td>Math</td>
<td>1990</td>
<td>200</td>
<td>Yes</td>
<td>0.11 to 0.56</td>
<td>8</td>
<td>0.11 to 0.00</td>
<td>0.16 to 0.70</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>To Construct and Standardize a Criterion-Validity in Mathematics for Pupils of Standard VII of Primary Schools of Ahmedabad City (Bhikhabhai G. Patel)</td>
<td>Math</td>
<td>1991</td>
<td>200</td>
<td>Yes</td>
<td>0.11 to 0.56</td>
<td>8</td>
<td>0.11 to 0.00</td>
<td>0.16 to 0.70</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Construction and Validation of a Criterion-Validity in Mathematics for Standard VII (Ramanbhai B. Patel)</td>
<td>Math</td>
<td>1991</td>
<td>300</td>
<td>Yes</td>
<td>0.11 to 0.56</td>
<td>8</td>
<td>0.11 to 0.00</td>
<td>0.16 to 0.70</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Construction and Validation of a Criterion-Validity in Mathematics for Standard V (Ashokbhai A. Patel)</td>
<td>Math</td>
<td>1992</td>
<td>300</td>
<td>Yes</td>
<td>0.11 to 0.56</td>
<td>8</td>
<td>0.11 to 0.00</td>
<td>0.16 to 0.70</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Construction and Validation of a Criterion-Validity in Mathematics for Standard IV (Akshay N. Mehta)</td>
<td>Math</td>
<td>1993</td>
<td>200</td>
<td>Yes</td>
<td>0.11 to 0.56</td>
<td>8</td>
<td>0.11 to 0.00</td>
<td>0.16 to 0.70</td>
<td>Yes</td>
</tr>
</tbody>
</table>
3.10 Review of the Literature Reviewed

On reviewing the available literature on the subject, the researcher found that the development and validation of a fine CRT is still a challenging task.

The present test is developed and standardized after taking into consideration the lessons learnt from the work done till date in the field. The foundation of the present study was based on CRT literature especially articles and dissertation abstracts mentioned in the bibliography appended to the present study. These articles and abstracts mainly dealt with the following aspects of CRT:

1. Item Writing techniques
2. Item Review procedures
3. Test Length
4. Cut-off scores
5. Validity
6. Reliability

Over and above, these aspects, the last word about the other final procedure or method regarding several aspects is yet to be said. It will not be considered out of place to quote Hambleton (1985) in this connection:

"At present, research is still under way on (1) method for setting standards, (2) formats for reporting scores to maximize test score usefulness, and (3) approaches for describing objectives."41
3.11 Significance and Special Characteristics of the present Study

The present study tries to remove deficiencies mentioned earlier in the chapter and to include following special characteristics:

1. Behaviour domains were well defined before the development of the test.
2. The test items were generated by adopting the special item writing techniques.
3. The test items were logically reviewed.
4. The test items were empirically reviewed.
5. Final form for the test was prepared after pre-pilot test.
6. The items were prepared according to domain specification.
7. The cut-off score was determined for each behavioural domain by employing a particular method.
8. The content validity, construct validity and criterion-related validity were established.
9. The domain score was estimated for each behavioural domain and standard error of measurement was calculated for the reliability of the estimated domain score.
10. The reliability of mastery-non mastery classification decision was established for each behavioural domain.
11. The present study in Mathematics for standard X is in the very important unit: Trigonometry.
12. The present test is standardized as per guidelines of Hambleton's twelve steps.
13. A manual for the administration of the test and for the technical data was prepared.

The succeeding chapter deals with the construction and try-out of the CRT.

2. Ibid., p.65.


7. Ibid., p.465.


22 Byron James Christensen: Validation of Criterion-Referenced Reading Skills Mastery Test for the Fifth level of GEMs. *DAI*, 1982 (March), 41(9), p.3963 A.


