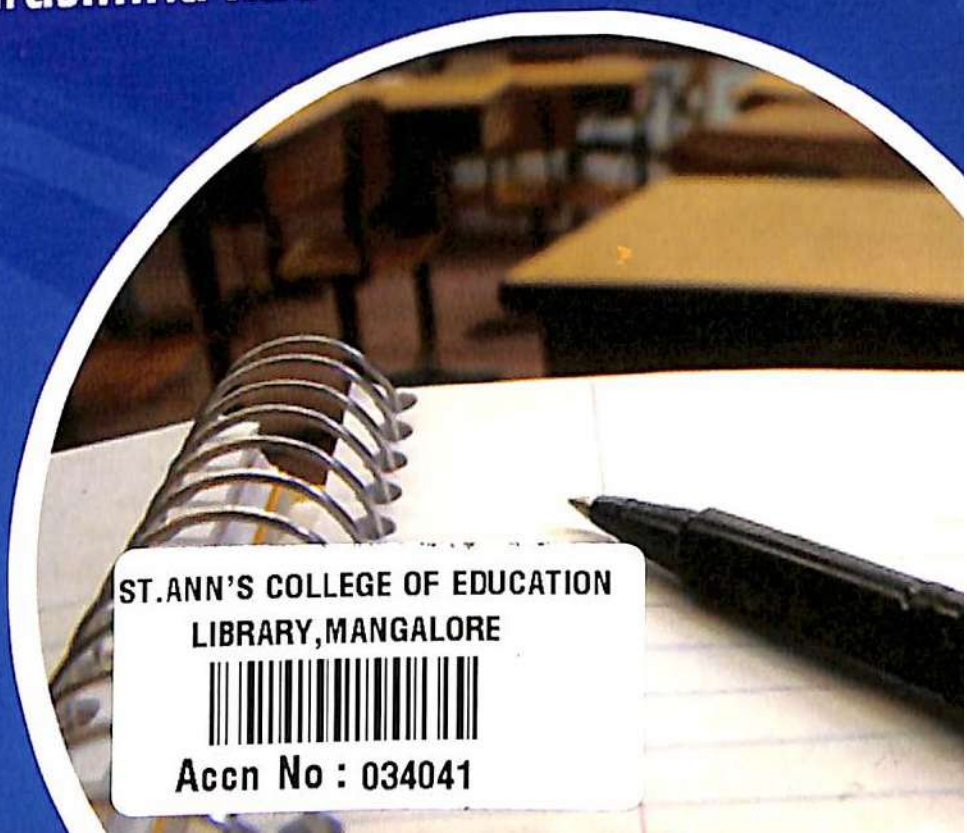
NEELKAMAL

Explorations in Educational Evaluation

Dr. K.V.V.L. Narasimha Rao



Explorations in Educational Evaluation

Explorations in Educational Evaluation

Dr. K.V.V.L. Narasimha Rao
Former Principal
Southern Regional Language Centre (C.I.I.L.)
Govt. of India
Mysore.

NEELKAMAL PUBLICATIONS PVT. LTD.
EDUCATIONAL PUBLISHERS
(E
HYDERABA HI

Explorations in Educational Evaluation

Dr. K.V.V.L. Narasimha Rao

Foreword

P.H. Sethu Madhava Rao

First Edition : 2013

ISBN: 978-81-8316-418-4

Price: ₹250/-

NEELKAMAL PUBLICATIONS PVT. LTD.

Sultan Bazar, Hyderabad - 500 095.

✆ 24757140, 24757197, 24757944, Fax: 040-24757951

Delhi Office:

BG5/9B, Paschim Vihar, New Delhi-110 063,

✆ 011-25285894

website: www.neelkamalpub.com

e-mail: neelkamalbooks@gmail.com; sales_neelkamal@rediffmail.com

---

Published by ***Suresh Chandra Sharma*** for ***Neelkamal Publications Pvt. Ltd.,*** New Delhi, Hyderabad and printed at ***Sri Vinayaka Art Printers,*** Hyderabad, India.

P.H. Sethu Madhava Rao
Former Joint Educational Adviser (Languages)
Ministry of Human Resource Development,
Govt. of India, New Delhi.

Foreword

Education in the Indian context has been an area of great concern. The Union Government, Union Territories and the State Governments have been laying emphasis on Primary Education. In fact, 50-60 percent of the total budget on Education is spent on Primary Education. In a developing country like India, Education is a necessary but not a sufficient condition for development. To attain cent percent literacy has been the main target of the government. However, so far, no distinct break-through in the education system to compete with the international standards has been possible in spite of several efforts. In the context of Indian education system, evaluation has been a matter of serious concern for the educationists, decision makers and the policymakers. The main reason for such a concern has been that, on the one hand, the evaluation and assessment of learners' performance has a number of lacunae and is based primarily on the performance of learners in the term-end examinations. No weightage has been given to the learners' performance during the course of instruction and this has led to an incomplete and biased (onesided) assessment which cannot be relied upon nor can it be construed as valid based on sound principles. At the level of higher education, certain universities and institutions have tried to make up these lacunae by assigning weightages to the in-course performance through "Internal

Assessment". This is not adequate in itself to make a just, realistic and valid assessment of learners in any course of instruction. Therefore, it becomes necessary to devise ways and means to make an objective assessment of the learners' performance through the course of instruction and take an overall view for finally deciding upon the quality of performance.

Several scholars including educationists, language specialists and psychologists have been making efforts to work towards a feasible solution to implement a realistic and valid method of evaluating the learners' performance, taking into consideration their regular and consistent performance all through the course of instruction, besides their performance in the term-end examinations. This issue has been aptly realized by Dr. K.V.V.L. Narasimha Rao, as a result of which he has worked on the problems and issues concerned and come out with reasonable and practical solutions not only in suggesting the strategies to arrive at a valid and reliable performance evaluation, but has also discussed in detail as to how these strategies and methods could be implemented in an educational system. It is a maiden effort, in my opinion, that has been made in this regard and I commend the innovations and explorations presented in this volume entitled *Explorations in Educational Evaluation* by Dr. Rao. I have no hesitation in stating that this valuable work will be a landmark in the area of educational evaluation especially in the Indian context.

Added to the innovative and explorative measures proposed by Dr. Rao in the context of performance evaluation, he has also rightly pointed out the inadequacy and incompleteness of material evaluation. I am gratified to see the viewpoint of Dr. Rao that the Comprehensibility of Textbooks in the context of material evaluation is an integral part. He has discussed in detail as to how the comprehensibility of textbooks could be assessed and related with other aspects of evaluating instructional material.

The present volume is quite comprehensive and self-contained in that the aspects relating to the evaluation of various components of education, like the Programs, Processes and Products and Methods, Materials and Media have been well presented in the present volume besides discussing in detail the aspects relating to the performance evaluation, the general principles of test construction and all related aspects.

I congratulate the author Dr. Narasimha Rao in bringing out the present volume which will be of immense help and use to the educationists in general and language educators in particular.

I have no doubt that the scholars in the fields of education, language and the policy and the decision makers relating to education will welcome this work and duly make use of it in their academic endeavors.

– P.H. Sethu Madhava Rao

Preface

Evaluation and testing are complex and difficult tasks in the field of education. They are very crucial and significant as they become the deciding factors regarding the degree of success of the examinees. Similarly, they form the basis for deciding the quality and efficacy of instructional material, methods and media. In the same way, programs, processes and products can also be evaluated regarding their effectiveness and efficacy. The findings of evaluation and measurement would facilitate in revising, reformulating and improving upon various programs, processes and products mainly in the field of education and so also in the case of the methods, materials and media.

It has been experienced by most people including the educationists, decision makers and policymakers that the evaluation system in education has been improper, inadequate and insufficient. Added to these, experience has shown that the existing system has several lacunae because of which the outcomes of measurement and evaluation have not been realistic, reliable and valid. It is a very common phenomenon that the decisions regarding the learners' success are made solely based upon the evaluation outcomes of examinations. This unilateral and one-sided evaluation method is unjust as it does not really reflect the competence of the learners. Several educational psychologists hold the view that the very concept of examinations and tests develops a fear complex among the learners as a result of which their performance does not reflect their real merits and demerits and their competence. They further hold the view that a consistent and continuous process of assessment and evaluation put together can only be the realistic evaluation. Therefore, a thought emerged

that certain measures should be discussed which facilitate realistic evaluation. The exposure and experience of the author in India and abroad made him realize that factors like the regular classroom observation and other evaluation aspects without using tests and examinations should be explored and suggestions made to incorporate them for arriving at a comprehensive, consolidated and realistic evaluation. From the point of view of performance evaluation, aspects like classroom observation, examination of journals and portfolios, conferences, questionnaires and interviews should be used.

Similarly, evaluation of instructional material is a very sensitive and alarming issue. Mostly, the assessment and evaluation of material is based upon the well established evaluation tools like questionnaires, checklists and rating scales. Little importance is given to the aspect of learners' comprehensibility of instructional material. This cannot be taken care of through the tools and instruments normally made use of. Therefore, an added dimension of assessing learners' comprehensibility of instructional material has to be ensured and integrated with the findings of the questionnaires, checklists, observation schedules, rating scales etc. It is only then that evaluation of instructional material could be realistic.

These two aspects relating to inadequacy and insufficiency made the author think of working on the present work. Detailed discussions and suggestions have been made that lead to a more realistic and dependable evaluation. While the main thrust has been on these two aspects, detailed discussion about almost all aspects of testing and measurement, assessment and evaluation have been made in the present volume. An added feature is that the scope of evaluation is not limited to language alone. The techniques and strategies of non-subject evaluation that might be different from those of language evaluation have also been discussed in this work.

It is with this background that the author has made a humble attempt at highlighting these aspects (Explorations in Educational Evaluation). It is hoped that educationists, evaluators, decision makers and policymakers will give due consideration to the theme and background of the present work and make up the lacunae existing in the evaluation process in the field of education, thereby making the process just and dependable.

The author would greatly appreciate if the readers take the trouble of offering constructive comments and criticism of the work.

– K.V.V.L.Narasimha Rao

Dedicated
To
The Lotus Feet
of
Lord Srimannarayana

K.V.V.L. Narasimha Rao

Acknowledgements

The procedures and techniques existing in the evaluation system in education have become a matter of serious concern for educationists, decision makers and policymakers. The seriousness of drawbacks and lacunae in the evaluation system on the one hand did not do real justice to the learners and on the other hand, several inadequacies and lacunae were taken cognizance of by all those who are actively engaged in the field of education with special reference to measurement and evaluation.

With a view to suggesting ways and means to fill up the gap to make measurement and evaluation as realistic as possible, the author having thought of such steps with special reference to performance evaluation and evaluation of instructional material, has worked on the present book for the possible use and benefits to all those who are concerned with education in general and evaluation and measurement in particular.

Though keenly interested in intensely working in the area of educational evaluation, the author could not have ventured to work on and bring out this book, but for the encouragement and untiring and intense help rendered to him by Smt.L.Anasuya in preparing this volume. It would be the bounden duty of the author to record a word of appreciation and gratitude for Ms. G. Bharathi who has evinced interest and incorporated the corrections in the draft copy of this volume. The author very gratefully acknowledges the tremendous help and efforts of his beloved friends Sri V. Gopal and Sri N.H. Itagi. Several friends and colleagues of the author have contributed directly or indirectly in many ways to the publication of this work. The author records his grateful thanks to all of them.

Special thanks are due to Sri Suresh Chandra Sharmaji Managing Director, Neelkamal Publications Pvt. Ltd., New Delhi-Hyderabad and his team for bringing out this Indian edition.

– *K.V.V.L.Narasimha Rao*

Explorations in Educational Evaluation

Contents

❖ ❖ ❖

CHAPTER

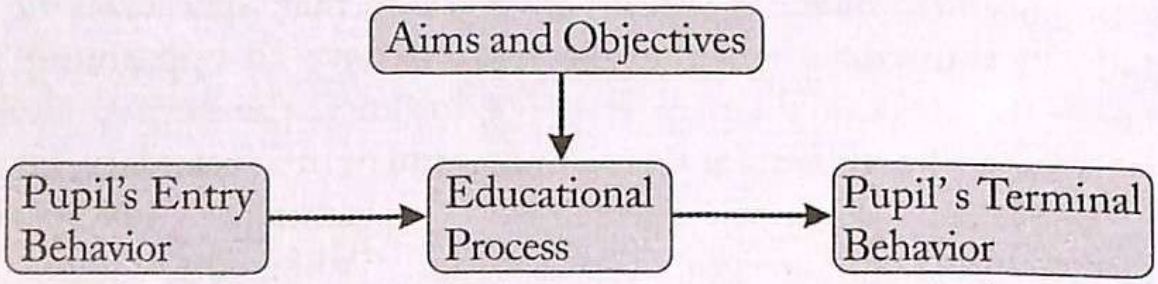
1

Introduction

Education is generally conceived of as a process through which certain defined change in behavior is expected from those who undergo the process of instruction. In other words, the aims and objectives of education are aimed at providing the learners with certain quantum of information, knowledge and exposure, based on which a set of behavioral changes are expected to take place among the learners. Such changes could be with reference to any and all subjects included under the broad realm of education. Looking at the education system, we will notice that until the end of the primary school stage, the scope, length and breadth of the educational framework is general in nature and covers almost all subjects of common interest which include language, physical and natural sciences, social sciences, civics, mathematics etc. It is from the commencement of the secondary school level that "specialization" is gradually introduced whereby the learners have an opportunity to choose the areas of their interest. For instance, the normal choice available for the students at the commencement of a secondary school is choosing between specialized mathematics (composite mathematics) and general mathematics. While the choice of specialized mathematics would lead the learners towards higher education in mathematics, technology and sciences, the choice in favor of general mathematics would normally lead the learners

towards sciences, arts and commerce in their higher education. As the level of instruction/education goes higher and higher, the intensity of specialization multiplies and the area of specialization gets narrowed down. It may thus be noticed that the field of education leads the learners from a broad general framework to a narrow and specialized discipline.

From the above description, one can notice that it becomes essential for any educational program to be very precise and specific about the goals/aims and objectives of education which need to be defined in clear and unambiguous terms and that too in terms of the learners' behavior. Needless to state that these specifications apply to each and every subject at different levels. The general educational objectives clearly indicate the changes/developments among the learners in their behavior through the educational process. The behavioral changes/developments expected among the learners could be with reference to (1) the knowledge of the learners, (2) the skills and abilities of the learners, (3) the interest among the learners and (4) the attitudes among them. The framework of education in terms of aims and objectives will be reflected in terms of what the learners know before undergoing the process of education, what they have and feel and what they do. The takeoff point for any educational program will be the entry behavior of the learners and terminal/landing point will be the end behavior or terminal behavior of the learners. The bridge between these two points is the educational process. It can be diagrammatically represented as:



As stated above, the process of education has to be related to the broad educational objectives and specific classroom objectives. The broad and general educational objectives are the important and

significant aspects of human development in general. Therefore, they form the crux of primary education. On the other hand, the classroom objectives are unique and specific to each course and subject and therefore they have to be planned in such a way that they are in synchrony with the broad and general goals of education. However, it must be taken note of in this context that generally, the educational process at the primary level is focused on the human development and it is only thereafter that different courses and subjects come into picture in the context of education. This is the stage from which the classroom objectives have to be geared towards the subject specifics and course specifics. There is a close knit relationship between the broad and general educational objectives and the specific classroom objectives. The specific classroom objectives relate to and contribute to the total educational process. Broad objectives are indirect while the classroom objectives are direct, specific and functional.

In the context of evaluation, the entire educational framework needs to be taken into consideration for the reason that the scope, breadth and length of both education and evaluation remain the same. The difference between education and evaluation primarily lies in that while the educational process takes care of every minute detail of the course content, the evaluation process gets limited to the selected few elements of the whole educational framework. Another difference between education and evaluation is primarily the quantum of time available for the two processes. Both the processes of education and evaluation take place simultaneously. While the process of education within the formal framework of classrooms is formal and structured, the process of evaluation is informal and unstructured except in the context of administering tests and examinations with a defined and specific structure. The purpose of evaluation is to determine the extent to which the set educational objectives are attained, to determine the effectiveness of learning experience provided during the course of instruction and to determine the degree of efficacy of the accomplished educational goals.

As the processes of education and evaluation have a common base, aspects of evaluation will have to be focused around all the aspects of education. These include:

1. Curriculum and syllabus
2. Educational objectives
3. Instructional material
4. Instructional methods
5. Instructional media
6. Instructor efficacy and expertise
7. Learners' performance

Although, in the real sense of education, all the above aspects have to be included, the common tendency and belief among most people involved in educational evaluation restrict and limit its scope to performance of learners alone. The other factors mentioned above are normally not taken seriously unless specifically insisted upon. Further, even the learners' performance is looked at from the point of view of how the learners exhibit their achievement through the tests and examinations conducted periodically during the course of instruction. A very significant point is lost sight of by almost all involved in the process of education and evaluation in that the decisions and value judgments arrived at regarding the learners' performance are based only on the outcomes of the tests and examinations. Thus, great injustice is done to all the learners as their performance throughout the course of instruction is not considered at all. Even the educational psychologists have argued that the mental framework of any individual, leave alone the learners, keeps changing and fluctuating from situation to situation and time to time. It is therefore possible that the mental framework of the learners might also fluctuate from time to time and for all that we know, most learners in general become tense and nervous in the context of formal tests and examinations. As a result, they could fail to establish their knowledge, skills and intelligence through the formal tests and examinations. This being the case, a realistic, more reliable and

dependable performance can be accomplished if and only if the learners' performance all through the course of instruction is considered simultaneously with the outcome of the formal tests and examinations. It needs to be mentioned here that some educational institutions, particularly of higher learning, assign certain degree of weightage to the internal assessment which is expected to be based on the learners' performance during the process of instruction. However, on the one hand, the weightage so assigned is minimal and internal assessment is generally based on several extraacademic factors. It is in this context that a review of the existing system of evaluating learners' performance has to be considered, the lacunae and drawbacks identified and solutions and remedial processes worked out so as to make the process of evaluating learners' performance is made realistic, fool-proof and dependable.

Similarly, another important point in the context of evaluating the instructional material is that, in most cases, the process is limited to the questionnaires and checklists and may be observation schedules too. Again, this process is not comprehensive and realistic for the reason that it is not evaluated from the point of view of the learners. Therefore, the process of evaluating instructional material also has to be made more realistic and dependable by supplementing the adoption of tools mentioned with the outcome of assessment of comprehensibility of the instructional material from the point of view of learners. Similar is the case with reference to the assessment of teachers' expertise in performance in implementing the educational programs.

The present volume on "Explorations in Educational Evaluation" is specifically focused around the lacunae and drawbacks of the existing evaluation system in the context of education and it attempts to identify these to the extent possible besides making an effort to suggest ways and means so that the processes of education and evaluation become more realistic and dependable.

❖ ❖ ❖

CHAPTER

2

Evolution of Present Day Education System

The origin of education system can be traced back to the Puranic and Vedic period. In the beginning, the process of education was accomplished through the oral tradition in the Gurukula system of education. At that time, the framework of education was informal in that no specific goals, aims and objectives were set up and no formal instructional materials were made use of. The effort of the Gurus was concentrated upon the human development of the sishyas (learners) in exposing them to different facets and aspects of life and building-up a good personality among the learners from the point of view of moral ethics and traditions. With the passage of time, developments and innovations, the educational process has also taken a formal shape in designing and defining the goals, aims and objectives of instruction, developing necessary instructional materials, adopting various methods of instruction etc. For quite a long time, the educational process, having been very informal and unstructured, did not result in the achievement of desired goals and achievements. Therefore, the education system existing those days had to undergo a transformation in such a way that it had to be related intensely with the life and needs of people thereby making it a powerful instrument of social, economic and cultural transformation necessary for the

realization of the national goals. The process of educational reformation was planned to be achieved through

- relating education to productivity
- promoting social and national integration
- giving a value orientation to the entire education system
- adopting modern methods of teaching and evaluation
- making the education system elastic and dynamic

2.1 Relating Education to Productivity

Keeping the above aspects in view, it was visualized that people living in industrialized societies must be given education related to productivity so that the educated individuals make solid and significant contribution to production. Further, for all people in general, education must result in increasing their standards of living. In order to achieve this, the emphasis was laid on (i) science education and research, (ii) work experience and (iii) vocational and professional education. The importance given to science and research is with a view to promoting advances in technology thereby supporting the development of agriculture, industry and in turn, the national wealth. Further, such education would loosen the bonds of dogmatism and would act as a powerful dispeller of fear, superstitions, fatalism or passive resignation and strengthen the commitment of man for fearless search for truth. Further, it was expected that improvement in the quality of science and education would help promote a very deep understanding of basic principles, a concept of free enquiry and experimentation, analytical skills and abilities of problem solving.

Work experience is active participation of the people in general and learners in particular in the productive work in different situations and environments. This is expected to build the gap between the intellectual and manual workers and the kinds of reasons for social stratification. Further, the work experience is expected to contribute to the economic growth by helping the young to develop insights into productive processes and uses of science and by

generating in them respect for manual labor and the fruits of hard and responsible work. Further, this would facilitate establishing a link between the individuals and the community.

The vocational and professional education is based on the manpower needs felt from time to time. It would help basically the illiterates and less educated in developing their vocational skills resulting in their self-dependence.

2.2 Promoting Social and National Integration

Promotion of social and national integration is very essential to develop a strong and united nation in view of the cultural and ethnic diversity that we find in India. Unless unity is established among so many kinds of diversities, it is not possible to build a strong nation. Therefore, to achieve the task of establishing a strong and united nation, education related to social and national integration will have to be built into the normal educational system through formal and informal schooling. Aspects relating to the social and national integration and unity in diversity will have to be indirectly built into the content areas of different subjects in schooling. The education process meant for this purpose must try to minimize, eliminate and eradicate the difference from the point of view of linguistic pluralism and cultural and ethnic multiplicities. Such a process will develop a kind of national awareness and consciousness among people and will help in (a) opening of avenues for all children irrespective of caste, creed, community, religion, economic conditions or social status, (b) provide access to education without reference to wealth, class or talent and (c) will try to maintain adequate standards in all schools.

2.3 Giving Value Orientation to the Entire Educational System

The aspects of value orientation refer to the scientific temper of mind, fearless search after truth, respect for manual labor and capacity, respect for proper pride in the past, faith and confidence in future, national consciousness, sense of social responsibilities and concept of social services. Further, they also include developing a

democratic way of life involving readiness and preparedness to understand and appreciate the intricacies and idiosyncrasies of fellow beings, settlement of disputes and quarrels through mutual discussion, persuasion and compromise, etc. All these aspects become significant in the context of education as they take a lead from the Indian Constitution. The value-oriented education has also to be concerned with the concepts of duty without self-involvement, tolerance, an innate love of peace and reverence for all living beings.

2.4 Adopting Modern Methods of Teaching and Evaluation

With the passage of time, innumerable discoveries and innovations in all fields including education have taken place. With a view to exploit these and make the education process very successful and rapid, it is ideal and desirable to effectively make use of various instructional methods and methods of evaluation in the field of education.

2.5 Making the Educational System Elastic and Dynamic

The potential and interest of people in any society are varied and multiple in nature. Therefore, at certain stage, the interest, potential and abilities of the learners must be kept in view while making available courses of instruction based on their interest and motivation and their personal and societal needs. Thus, it is necessary to ensure elasticity and dynamism in the process of education, thereby taking care of the wide ranging and varied interests and abilities. Further, the aspects of elasticity and dynamism in education will help in developing a multiple approach to problem solving which will, in turn, become a strong base and support for qualitatively better and rapid results.

Keeping the issues and problems faced by our country in view and its multi-linguistic, multi-cultural and multi-ethnic characteristics, the Government of India had come out with a National Policy on Education in 1968 which again underwent a

revision in 1986. In suggesting the reorganization of education at different stages, the Government of India had kept the national goals of bringing about equality among all and enriching one and all with aspects mentioned above besides promoting several moral and other values including cultural heritage, egalitarianism, democracy and secularism, equality, protection of the environment, removal of social barriers, observance of small family norm and inculcation of scientific temper. As certain drawbacks and lacunae were found in the implementation of the education policy defined way back in 1968, the 1986 policy has suggested certain reorganization of steps in the total framework of education in the country. Emphasis was laid on the early childcare and education including therein the child development from the points of view of nutrition, health and social, mental, physical, moral and emotional development. It was also suggested that the 3 R's will be initiated and introduced in the early childcare and education. Further, the pre-primary education was planned in such a way that it could function as a feeder and a fortifying factor for primary education and human resource development.

It is from the stage of secondary level of education that the learners would be exposed to the differentiated roles of science, humanities and social sciences and this is the stage at which emphasis was planned to be laid upon the children with a sense of history of national perspective and provide them with an opportunity to understand their constitutional duties and rights as citizens.

The higher education was aimed at providing people with an opportunity to reflect on the critical, social, economic, cultural, moral and spiritual issues facing humanity. This level of education is expected to contribute towards national development through dissemination of specialized knowledge and goals. In fact, this is the level at which the fields of specialization get narrowed down and become more intense and deep.

A significant feature of the National Policy on Education, 1986 is that the concept of delinking degrees from jobs in selected

areas was conceived of and initiated. However, the areas to which this concept shall not apply included engineering, medicine, law and teaching etc. The concept of delinking degrees is aimed at applying for services for which a formal degree need not be a necessary qualification. Technical and management education have been given priority at the higher levels of education. Special attention was proposed to be given to mathematics teaching, science education, sports and physical education etc. It was felt necessary that the teaching of mathematics should be visualized as a vehicle to train the children to think, reason, analyze and to articulate logically. Besides, its being a subject specific aspect, it is to be treated as concomitant to any subject involving analysis and reason. Similarly, science education was aimed at developing well defined abilities and values like the spirit of enquiry, creativity, objectivity, the courage to question and aesthetic sensibility among the learners.

A very significant step contained in the National Policy on Education, 1986 is with reference to the process of evaluation and reforms in examination system. With a view to introduce qualitative improvement in education, several steps to recast the examination system, making the process of assessment valid and reliable, were initiated. The suggestions in this context were focused on the learners' performance, which include:

i) elimination of excessive element of chance and subjectivity;

ii) de-emphasis on memorization;

iii) continuous and comprehensive evaluation that incorporates both scholastic and non-scholastic aspects of education spread over the total span of instructional time;

iv) effective use of the evaluation process by teachers, students and parents;

v) improvement in the conduct of examinations;

vi) introduction of concomitant changes in instructional materials and methodology;

vii) introduction of the semester system from the secondary stage in a phased manner; and

viii) use of grades in place of numerical scores.

Simultaneously with the re-orientation of the education system and the process of evaluation (limited to the performance evaluation), emphasis was also laid on devising a continuous process of education with reference to teacher training etc. The new programs of teacher education were contemplated to emphasize upon continuing education and need for teachers to meet the thrusts envisaged in the National Policy on Education.

CHAPTER

3

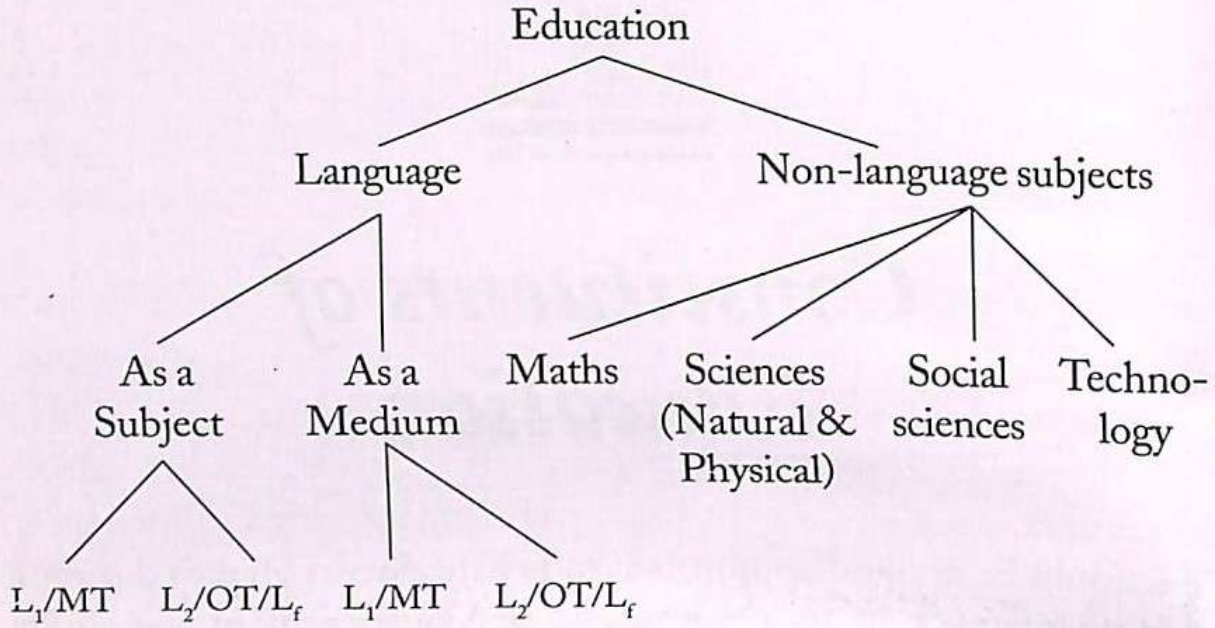
Constituents of Education

It would be apt and appropriate to briefly discuss about education with reference to its constituents. As mentioned in the earlier chapters "Education" is a very broad concept involving all kinds of learning at different levels of instruction.

It has also been pointed out that the process of education starts from a broad and general framework and proceeds towards narrow and specialized disciplines. This implies that the earlier levels of instruction include a number of subjects and as the level of instruction increases, the number of subjects gets reduced and ultimately would narrow down to one specific field of specialization at the highest level. Therefore, in conceptualizing the process of education, it must be kept in mind that several subject areas form the constituents of education. Such constituents include languages and subjects such as mathematics, natural and physical sciences, technology, social sciences, civics, health sciences, physical education, etc.

It is a well established fact that, in any context, language and content go hand in hand and each one supplements the other. To be clearer and more specific, there can be no content without language or expression and there can be no expression/language without content. However, a gulf of difference exists between the processes

involved in teaching languages as opposed to the teaching of subjects. Thus, the concept of education can be visualized diagrammatically as follows:



From the above diagram, it can be seen that the constituents of education could be broadly categorized into language and non-language subjects. Language, again, can be visualized (a) as a subject and (b) as medium of instruction. Again, both language as a subject and language as a medium could apply to the first language (L_1)/ mother tongue (MT) or second language (L_2) or other language (OL)/foreign language (L_f). The non-language subjects include mathematics, sciences – natural and physical, social sciences, technology, history, civics and a host of other subjects.

From the point of view of imparting education and evaluating educational processes, a distinct demarcating line could be drawn between language and non-language subjects. To the extent that the language instruction is concerned, it has a dual role to play – (a) to impart language as a subject and (b) to utilize the language as a medium of instruction with reference to non-language subjects. Language instruction as a subject primarily involves the development of language skills among the learners which enables them to actively and effectively make use of the language in different contexts and situations in day-to-day life. Development of such skills aims at

developing two kinds of abilities among the learners viz., (1) linguistic/grammatical competence and (2) communicative competence. The skills involved in language instruction (as a subject) are wide ranging. At the initial level of language instruction, the skills involved are the 4 primary/basic language skills viz., LSRW (Listening, Speaking, Reading and Writing). As the level of language instruction goes higher, the primary/basic skills have to be supplemented with the development of the higher order language skills viz., RARE (Representation, Argumentation, Refutation and Establishment of one's own point of view). Further, the abilities and skills relating to the development of a running text on a given topic, précis writing/epitomization composition/essay also become necessary depending upon the level, nature and purpose of instruction.

In the context of language instruction as a medium of education, 'language' becomes the vehicle of expression with reference to the subjects involved. Therefore, 'content' plays a very significant role than the language skills. Accordingly, the emphasis would be on the core areas of the subject wherein very active and intense use of specific discipline/subject based vocabulary and the typical sentence patterns figure in. The skills involved in imparting language instruction in the context of language as a medium include 'study skills' which in general are: listening and note-taking, reading and note-taking, listening and reading comprehension, epitomisation/ précis writing, comprehension-cum-composition and free composition, critical and creative writing. Thus, it may be noticed that there is a clear-cut distinction and difference between the skills involved in the context of language as a subject as opposed to the language as a medium of instruction. It is at this stage, that it should be clearly mentioned that the 'study skills' mentioned in the context of language as a 'medium' hold good in the context of subject teaching. It follows that the whole framework of education differs between the two and consequently, the modus operandi/procedures with reference to evaluation also becomes different.

Reverting to the discussion on language instruction as a subject, it becomes necessary to make a distinction between first language/

mother tongue, second language and other language. Innumerable definitions have been proposed by a host of scholars in respect of these three concepts and more with reference to the concept of native language/mother tongue. It is beyond the scope of the present volume to discuss in detail about the merits and demerits of definitions proposed by different scholars and therefore, it is proposed to merely present the generally accepted broad definitions of the three concepts.

Keeping in view different viewpoints of various scholars regarding the concept of mother tongue/native language, a generally acceptable definition can be presented as "Mother Tongue/Native Language" is that language, the environment of which is available to the individual everywhere at home, in the learning environment and elsewhere as well and the degree of such environment is the maximum right from the individual's infancy all through.

Second Language is that language, the quantum of environment of which is next in degree to that of mother tongue/ native language and the environment of which is available in the learning environment and relatively reasonable number of opportunities exist for active use of that language in the outside environment. Besides, the skills relating to the second language are developed through a formal course of instruction.

Other Language/Foreign Language is that language which is learnt through formal course of instruction and the exposure to which is limited to the learning environment and opportunities to make use of that language are limited only to the four walls of the classroom. Opportunities for actively making use of such a language are restricted to the learning environment only and the skills relating to such a language are also picked up through the formal course of instruction.

It may now be relevant to discuss in brief about the contexts in which different kinds of languages are taught as subjects and are made use of as media of instruction. To the extent that the mother tongue/native language is concerned, it is needless to mention that the individuals pick up such a language in a natural and informal

situation and thereby gain control of relevant and reasonable quantum of vocabulary relating to various situations in common day-to-day life and the sentence patterns related to these situations. Further, the competence/ability that individuals gain through such informal and natural environment before entering the formal schooling system relates primarily to the spoken skills viz., understanding and speaking. In most cases, the development of written skills takes place through the formal course of instruction in early stages of schooling. It should be admitted that it is not infrequent that some children get exposed to the written skills also in some degree or the other during the pre-school stage itself. This being the case with mother tongue/native language, certain experiments have shown and established that education/instruction through mother tongue/native language is relatively more effective as compared to the instruction through an alien language – be it second language or other language. Several language scientists have also argued in favor of adopting mother tongue/native language as medium of instruction atleast at the school level and the formal policies of the Central and State Governments have also been in favor of adopting the mother tongue/native language as the medium of instruction. The issue of native language/mother tongue as a medium of instruction remains highly controversial even today for various reasons. Some people hold the view that the regional languages and mother tongues are not self-sufficient for adopting them as media of instruction for teaching of various subjects. They further argue that education gained through the mother tongue/native language as medium of instruction restricts the scope of movement of the individuals for the reason that this prevents them from travelling to other regions where the regional languages are different from their mother tongue/native language. As such, individuals do not possess any knowledge whatsoever pertaining to languages other than their own. In spite of serious controversies raised about the usefulness of instruction through mother tongue/native language, the author feels that the arguments proposed regarding the limitations and constraints presented above are sound and meaningful. One problem could be successfully overcome in making the regional languages/

mother tongues effective media of instruction by developing vocabulary relevant to various disciplines and the styles and registers thereof. As far as the limitations regarding mobility and scope of employment etc., are concerned, there does not appear to be a different view than agreeing upon the arguments that people have about instruction through mother tongue/native language and this prevents people from moving and seeking employment beyond their regions as they will be fish out of water in an alien language atmosphere. Therefore, keeping the pros and cons of the issue in view, it could at best be compromised to the extent that MT/NL could be adopted as medium of instruction up to the end of school level. However, it becomes necessary that the medium of instruction at the higher levels of instruction has to be such a language which can be made use of for effective communication in different regions all over the nation – be it the official language or the associate official language i.e., Hindi or English. Arguments similar to the case of mother tongue/regional languages in the context of using them as media of instruction are also extended to Hindi as well, which in the opinion of the author are true. In spite of the policymakers, the politicians and the decision-makers' views and recommendations being in favor of the official language, keeping in view the wider acceptability and easier mobility not only within the nation but beyond, it is sincerely felt that the medium of instruction should be the associate official language at least from the commencement of the collegiate level of instruction. It is worthy to mention here that most people, whether they are policymakers or politicians or educationists, formally argue in favor of the regional languages/MTs being the media of instruction but when it comes to a question of applying the principles to themselves and their dependents, they always prefer education through English medium and in most cases, education abroad. The author would venture to state that all these reflect on the hypocrisy of the policymakers, the decision makers and the politicians.

The second languages are preferred as media of instruction more in comparison to the MTs/NLs. The reasons for such a view are

the scope for wider acceptability and easier mobility beyond the regions. It needs to be confessed and accepted in this context that developed languages are richer in comparison to the underdeveloped languages, many of which are the regional languages/mother tongues looked at from the point of view of the vocabulary and the styles and registers established for different disciplines. The argument for argument sake can be extended in this context in support of all the regional languages/mother tongues being made the media of instruction in that the reasonable and relevant vocabulary, sentence patterns, styles and registers could be developed in all MTs/NLs which is a remote practicality. Therefore, the author tends to maintain the view that for certain time to come i.e., until the MTs/NLs become self-sufficient to serve as media of instruction, it becomes necessary to adopt English as the medium of instruction keeping in view the advantages and benefits from the point of view of their employment, inter-state and intra-state mobility and international and intra-national mobility besides wider acceptability. Incidentally, it is worthy to mention that language specialists in India and abroad have successfully established that "An Indian variety of English" exists indirectly implying that one variety of English is native to India. Under these circumstances, there should be no reservations to adopt English as the medium of instruction at different levels. Under the circumstances prevailing and the facts, it should be admitted that the English language has come to India to remain for ever, though the English people and rule were in India for sometime and fled away from India after Independence. It does not stand to reason why anybody should have any misgivings or reservations about any language to be used either for communication purposes or for educational purposes. It therefore becomes necessary to further enrich the English language simultaneously with Hindi and other languages so as to provide a wider choice to be left for the individuals to choose any language as medium of instruction keeping their own benefits, advantages, beliefs and attitudes in view. It should also be pointed out in this context that neither there should be any imposition regarding the choice of various subjects nor the languages nor the medium of instruction on anybody. Such imposition could be

harmful and disastrous both from the point of view of the individual's advantage and larger interests.

As far as the other languages/foreign languages are concerned, they should be looked at from the point of view of the definitions presented above. Any language in the context of education should be visualized as a non-native language or a foreign language from the point of view of its geographical existence and not from the point of view of its origin. For instance, though English is a foreign language from the point of view of its origin and regional belongingness, it remains a second language in India for the reason that it is not only used intensively in the education system all over but it is also used very extensively in contexts beyond the scope of education everywhere. Further, the knowledge and use of English has also become a "status symbol". Similarly, languages in regions other than their own could be visualized as foreign languages though from the point of origin they might be Indian in nature. For instance, in a State like Karnataka, a language like Kashmiri or Gujarati or Bengali would practically become a foreign language as the environment of such languages is limited to the teaching-learning situation alone and little opportunities exist beyond the four walls of the school. No opportunity exists for using such languages either at home or outside. Therefore, such languages though Indian in origin, become foreign languages in the context of education.

Some foreign languages like French, German etc., are made use of as media of instruction in certain selected disciplines which are highly technical in nature and no instructional material with reference to such serious technical subjects is available either in any Indian language or English. Therefore, learning of such languages to gain knowledge in specified subjects becomes necessary and need based. It is in this context that the need-based learning of languages as opposed to the general learning of languages has to be made clear. The learning of languages for general purposes involves the development of skills for using such a language in various common day-to-day life situations. The need-based learning of languages refers to the development of control over the specialized discipline-

based vocabulary, sentence patterns, styles and registers and the use of such a language is limited and restricted to the discipline concerned. Such languages are not made use of for general purposes like common day-to-day life situations.

It may now be appropriate to make a distinction between language and subjects (other than language) in the context of education. From the point of view of instruction, a distinction can he made between the two in that that while "skills" play a significant role in the context of language, "content" plays a crucial role in the context of subjects. It again becomes necessary to reiterate and reaffirm that expression and content go hand in hand and they are not watertight compartments. Expression, thus necessarily involves skills in some manner or the other. Vice-versa is also true. The difference between the two could be made clearer looking at from the point of view of preparing the instructional material. In the process of developing language instructional material, selection and gradation refer to the general vocabulary and common sentence patterns whereas such concepts in the case of subjects refer to the content areas of the subject concerned. Further, even from the point of view of skills, the difference between the two could be established in that that while language involves the development of linguistic skills, subjects involve the development of content and "study skills". Another point to be taken note of is that "presentation" is very significant in the context of language instruction whereas it is not as significant in the context of subject instruction. The same arguments hold good with reference to "repetition". Looking at from the point of view of the exercises etc., they should be focused on the linguistic skills in the context of language instruction whereas they should attempt at eliciting knowledge and content of the subject concerned in the context of subject instruction. Linguistic accuracy is much more serious and a matter of concern in the context of language instruction whereas it is relatively flexible in the context of subject instruction. Communicability, however, is equally important in both language instruction and subject instruction. Thus, it may be noticed that there is a gulf of difference between language and non-language

subjects in the context of education. These differences would clearly separate language from subjects from the point of view of instruction and consequently, the process of evaluation also becomes different with reference to language and subjects.

There is a close-knit relation between instruction and evaluation. The scope and basis for both remains the same but the difference lies in that that the process of instruction is in depth and in detail and more time consuming compared to evaluation. Evaluation involves a very careful selection of the skills and content areas whereas such a selection is not possible in the context of instruction.

As mentioned earlier, the constituent elements of education and evaluation could be visualized as the following: (1) Curriculum and syllabus, (2) Educational objectives, (3) Instructional material, (4) Instructional methods, (5) Instructional media, (6) Instructor efficacy and expertise, and (7) Learners' performance.

CHAPTER

4

Evaluation and Measurement

Having identified the constituent elements of evaluation, it may now be apt and relevant to briefly discuss the terms "evaluation" and "measurement" which are very frequently used and then proceed to talk about various steps involved in the process of evaluation. "Evaluation" is defined by Good as " (I) the process of ascertaining or judging the value or amounts of something by use of a standard of appraisal; includes judgements in terms of internal evidence and external criteria, (2) the consideration of evidence in the light of value, standard and in terms of the particular situation and the goals which the group or the individual is striving to attain, (3) judgement of merit sometimes based solely on measurements, such as those provided by test scores, but more frequently involving the synthesis of various measurements, critical incidents subject to impression and other kinds of evidence weighted in the process of carefully appraising the effects of an educational experience".

According to Good, "Measurement is the comparison of a quantity (exhibited by a particular case) within an appropriate scale for the purpose of determining (within the limits of accuracy imposed by the nature of the scale) the numerical value on the scale that corresponds to the quantity to be measured"[2].

Having perused the definitions of evaluation and measurement, it is worthwhile to look at the difference between the two. From the definitions and concepts relating to the two terms, it can be clearly noted that the process of "measurement" is limited to the quantity to descriptions of the phenomena. There is no qualitative description involved in measurement and it does not imply judgments concerning the worth or value of the phenomena measured. Evaluation includes both qualitative and quantitative descriptions of a phenomenon plus value judgments concerning its desirability. Thus, evaluation and measurement are no doubt related terms. While evaluation is a broader concept, measurement is a narrower concept. Evaluation includes measurement and measurement forms the basis for evaluation. Some scholars have argued that evaluation may or may not be based on measurement but it goes beyond quantitative description. The relationship between the two can be presented as follows:

Evaluation = quantitative description of phenomenon (measurement) + value judgment

Or

qualitative description of phenomenon (non-measurement) + value judgment.

It is well established in the earlier discussion that evaluation is a process. Therefore, there are certain well-defined steps in the process. They are:

i) The first step of evaluation is to determine and define what is to be evaluated. This implies a clear definition of the objectives of the instructional program.

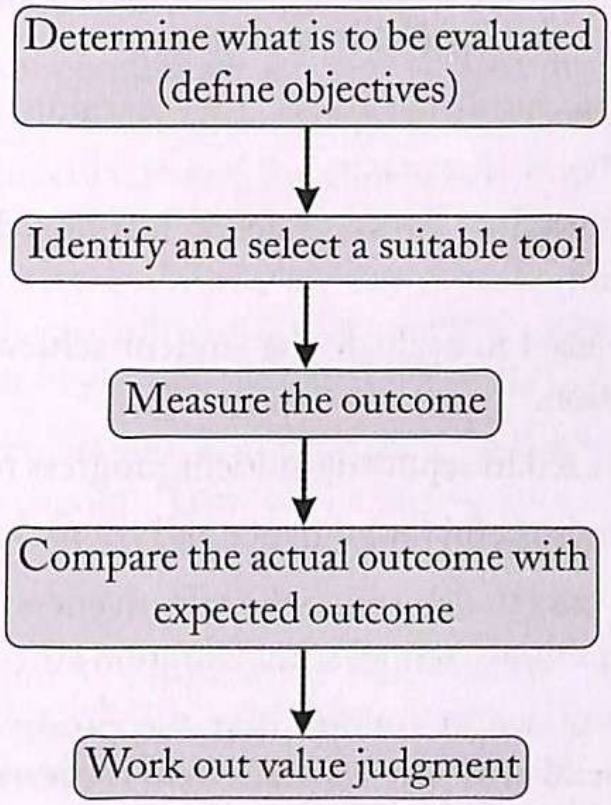
ii) The second step is to select the most suitable tool/instrument for measuring the phenomenon under consideration. For e.g., the item types and tests are the suitable tools for performance evaluation while questionnaires, checklists and observation schedules etc., are the suitable tools/instruments for evaluating constituents like the instructional material, methods, media etc.

iii) The third step is to measure the learning outcomes by using the selected tool/instrument.

iv) The fourth step is to compare the result of measurement with what is expected.

v) The final step is to arrive at a judgment on the basis of comparison of the actual outcome with the expected outcome.

It can be diagrammatically represented as follows:



According to Genesee and Upshur, "purpose of evaluation is to ascertain the effectiveness of a particular instructional activity with a view to re-designing it or re-using it with modification which might rely on observational information ...."[3]. According to them, the process of evaluation consists of the following four steps:

- To determine what information would be useful for making particular kinds of decisions.
- To device ways of collecting relevant information.
- To make sense of the evaluation information.
- To make appropriate decisions.

As mentioned in the earlier pages, most people relate measurement and evaluation only to the learners' performance in an instructional program.

The other aspects are not generally thought of except when specifically warranted. Looking at evaluation and measurement from a broader perspective, its uses in education could be visualized as follows:

- Evaluation may be used to assess the entering behavior of a student in a sequence of instruction.
- It may be used to assess the learning progress during instruction.
- It may be used to assess students' learning difficulties during instruction.
- It may be used to evaluate the student achievement at the end of instruction.
- It may be used in reporting student progress to parents.
- Evaluation is useful for guidance and counseling.
- It may be used to determine the effectiveness of course content, teaching methods, school administration etc.

From the above, it follows that the process of evaluation is applicable to several constituents other than the learners' performance like the instructional objectives, syllabus, instructional material, instructional methods, media etc. The outcomes of such evaluation would facilitate re-designing and re-using and re-applying these constituents in a better and more meaningful manner. Thus, the outcomes of evaluation with reference to these constituents will help in the qualitative improvement and better designing of the constituents.

Having talked about the scope and extent of evaluation in the context of education, it is apt and desirable to discuss about different types of evaluation with reference to any educational program. Broadly speaking, the following types of evaluation are talked about:

1. Ongoing or continuous evaluation
2. Terminal evaluation
3. Brief evaluation
4. Extensive evaluation
5. Formative evaluation
6. Summative evaluation
7. Diagnostic evaluation

Ongoing or continuous evaluation is primarily an informal process of getting feedback from time to time on a regular basis about the impact and effectiveness of the constituent evaluated. This could be at different stages in the educational program viz., planning, preparation, production and application. This facilitates making use of the feedback so obtained for instant improvements and modifications that might be considered necessary.

Terminal evaluation is the process of formal evaluation on completion of a program. This will be an overall and final evaluation to ultimately decide upon the success or failure of a program as a whole.

Brief evaluation involves the opinions and value judgments based on the significant highlights of an educational program without going into in greater depth and in detail. Such judgments are more impressionistic, global and subjective than being objective and realistic. It is possible that the findings of brief evaluation may prove to be contrary to the findings of extensive evaluation. This, however, is a rare phenomenon. The main advantage of brief evaluation is that it is less time consuming and could be useful in making a selection of items of similar nature. Brief evaluation is closer to review. The difference between the two is that brief evaluation is the evaluator's judgment based on systematic measurement of evidence collected by him. Review involves analytical judgment. Brief evaluation is a via-media between review and extensive evaluation.

Extensive evaluation involves a detailed and in-depth analysis of a program with reference to all its main and sub aspects. Each of

these aspects has to be weighed by the evaluator who has to consolidate the total rating based on each and the value judgments arrived at. This is highly objective, reliable and valid compared to the other kinds of evaluation.

Formative evaluation is that process of evaluation which is carried out from time to time in an educational program at different stages. This, in a way, is similar to the ongoing evaluation in that that it provides feedback from time to time whenever undertaken. Formative evaluation again does not provide a total impression of the quality of constituents evaluated. It could be made applicable to the educational program as a whole, methods, materials and media; performance evaluation etc. With reference to the learners' performance, formative evaluation helps in monitoring the learning process and progress of students during the course of instruction. The feedback so obtained provides reinforcement for successful learning and identifies the specific learning difficulties and errors that need correction and remedies. Further, it would also help the instructor to modify his process of instruction and for prescribing group or individual remedial work.

Summative evaluation is designed to find out the extent to which the instructional objectives have been achieved and is used in arriving at the final decisions regarding the quality and effectiveness of the educational constituents evaluated or the degree of success in the case of performance evaluation. This further facilitates placing the learners depending on their final scores arrived at. This is similar to the terminal evaluation.

Diagnostic evaluation is primarily concerned with the students' learning difficulties during the course of instruction which are left unsolved by standard corrective prescriptions based on the outcome of formative evaluation. This is more comprehensive and detailed. This involves the use of specially designed diagnostic tests and techniques of observation. The basic aim of diagnostic evaluation is to determine the causes of learning problems and to formulate a plan of action for devising remedies thereof.

❖ ❖ ❖

CHAPTER

5

Tools and Techniques in Measurement and Evaluation

There are several instruments/tools and techniques adopted in the process of evaluation, and the choice of such tools in different contexts depends upon the kind and nature of the constituents evaluated. Such tools include tests and examinations – both oral and written, observational techniques, socio-metric techniques, inventories, checklists, questionnaires etc. Broadly speaking, they can be classified under three heads of measurement viz., testing techniques, observational techniques and self-reporting techniques. Testing techniques include a wide range of teacher-made and standardized tests wherein the achievement, diagnostic, intelligence and aptitude tests are the integral variants. Rating scales, checklists, anecdotal records and socio-metric techniques can he discussed under observational techniques keeping in view the similarities in nature. Similarly, the self-reporting techniques, interviews and questionnaires are discussed.

Tests could be a task or a set of tasks that elicit observable behavior from the test takers. As already mentioned, there is a wide range of tests and different types of tests are used for eliciting behavior related to different kinds of skills and knowledge. In educational terms, tests possess a subject matter or content. Tests and examinations yield scores that represent attributes or characteristics of

individuals. As these tests have a frame of reference, the scores obtained from the tests form the frame of reference which is called "measurement". Thus, we may notice that there are three main aspects of tests viz., content, method and measure-ment. The content reflects as to what is tested. It could be intelligence; it could be language proficiency or it could be knowledge in the context of subjects and skills. As stated earlier, the tests cover only a selected representative sample of the entire content and skills covered during the course of instruction. The test methods are concerned with the manner in which the testing process is carried out. To elaborate this point, it can be mentioned that the type of response expected from the test takers depends upon the nature and type of test item. For instance, a multiple-choice test item demands that the test taker has to pick and choose the correct answer among the choices available. Incidentally, the choices provided have to be very close to the 'key' (correct answer) and making the correct choice among them could, at times, be extremely difficult for the test takers. Similarly, if the test item is a short answer question, the test taker is expected to provide a very brief answer recalling from his memory. In this case, the question of choice does not arise. If the test item happens to be a true/false item, again it is a question of the test taker's choosing the correct answer between the two choices. Thus, the methodology of testing is related to the item types. While some scholars call them as item types, some others prefer to label them as testing methods. The third aspect is the measurement. Tests describe attributes or quality of things and individuals by assigning numerical scores to them. The measuring types, broadly speaking, are of two kinds. While one type measures the difference in kind, the other type measures the difference in degree. In the context of educational achievement, the task of measurement becomes very difficult as in most cases, the tests are sensitive to differences in degree within the same kind. This requires that the test items measure how much of a particular ability or attribute the test taker possesses and not simply whether the test taker has it. Another problem is that measuring educational achievement requires that all the items in a test measure only attribute or quality as otherwise the test measures different qualities

and the score or measurement is not "pure". It is to be admitted, however, that irrespective of the degree of expertise involved in preparing the test item, it is too difficult, if not impossible, to ensure that the test items are focused around only one attribute. Therefore, the value judgments drawn based on the test scores involve the deriving of inferences by the examiners. Consequently, in spite of the high objectivity of the test, items in the tests some degree of `subjectivity' creeps in the process of drawing inferences.

Having discussed what a test and its constituents are, it may now be appropriate to discuss in brief about various kinds of tests and other tools and techniques used in the process of measurement and evaluation.

5.1 Achievement Tests

Achievement tests are those tests which are commonly and extensively used in a teaching-learning situation. There are several purposes for which they are made use of. They could be used for measuring the knowledge and competence of a learner just at the time of entry to a particular course of instruction or they are used from time to time during the process of instruction to monitor the learners' learning processes and to provide ongoing feedback both to the learners and the teachers. They are also used to clearly identify the learning difficulties of the students – whether such difficulties are persistent or non-recurring. They are also made use of at the termination of a course of instruction whereupon grades are assigned and value judgments about the degree of success or failure of the learners are drawn up.

Achievement tests are broadly classified into two categories viz., (1) Teacher-made tests and (2) Standardized tests. Teacher-made tests are used during the course of instruction and the results indicate the learners' achievement in relation to the instruction provided. Normally, the results of these tests are made use of in improving the teaching-learning process. A significant feature of these tests is that they motivate the students in further learning and lead them towards self-evaluation. The drawbacks of such tests include the insufficiency

of objectivity and non-availability of a standardized procedure for administering them. Added to these, no empirical evidence exists in support of reliability and validity in respect of teacher-made tests.

Standardized tests are those tests which are used commonly by all educational institutions at a given level of instruction. The significant characteristics of standardized tests are:

- They consist of test items of high quality. The items are pre-tested and selected on the basis of difficulty value, discrimination power and relationship to clearly defined objectives in behavioral terms.
- The instructions and directions for administration, the exact time limits and methodology of scoring are precisely defined and spelt out, so that valuation by different examiners remains the same.
- Norms based on representative groups of individuals are provided as the basis for interpreting the test scores. Such norms are based on various aspects like the age, grade, sex, regional area etc.
- Reliability and validity are established before making use of the tests.
- Normally, a detailed test manual is made available which explains the purpose and use of the test briefly describing the manner of its construction and provides clear and specific guidelines and directions for administering, scoring and interpreting results. The manual also contains the table of norms and summarizes available research data on the tests.

Normally, the standardized tests are used for evaluating the general educational development of students in the basic skills and in those learning outcomes common to many courses of study. Tests made use of in several institutions in a wide geographical area and those tests which are made use of for common entrance to various technical and professional courses etc., fall under the category of standardized tests.

The teacher-made tests, if very carefully and cautiously prepared, are very close to the standardized tests. However, the teacher-made tests are not as reliable and valid as the standardized tests are and the test items contained in the teacher-made tests are generally inferior in quality compared to those contained in the standardized tests.

The teacher-made tests and standardized tests are made use of for different purposes. Standardized tests are useful if the purpose of testing is to compare the students' performance in different content areas, to find the status of an individual student, class or school system in a wider population, to compare classes and schools among themselves and to measure growth over a period of time. On the other hand, the teacher-made tests are used to determine whether specific curriculum goal has been achieved and to compare students with their immediate classmates as in assigning grades. TOEFL, GRE, GMAT, etc., are examples of standardized tests used at the national and international levels. (see table in page 36)

5.2 Diagnostic Tests

As the very name indicates, such tests are made use of for purposes of diagnosis. Thus, these tests will help in locating and identifying the problems and difficulties of the learners at different stages of instruction. These tests consist of test items based on a detailed analysis of the specific skills and content. The test difficulty is kept lower so as to provide adequate discrimination among the learners with learning difficulties. Normally, diagnostic tests are designed for students of lower abilities. The drawback of such tests is that they provide only partial information relating to the learner difficulties. These tests are highly questionable from the point of view of reliability and validity. Therefore, the scope and use of such diagnostic tests is limited and restricted and are made use of for the sole purpose of diagnosing the learning difficulties.

Sl.No.	*Point of comparison*	*Teacher-made achievement test*	*Standardized achievement test*
1.	Learning out-comes and content measured	They are generally used to evaluate the outcomes and content of a curriculum.	They are generally used to evaluate outcomes and content common to a number of schools.
2.	Quality of test items	Quality of test items is unknown and is lower than that of items of standardized test.	General quality of items is high. They are pre-tested and selected on the basis of difficulty and discriminating power.
3.	Reliability	Reliability is usually unknown but it can be high if test items are carefully constructed.	Reliability is usually high and the value lies between 0.80 and 0.95.
4.	Administration and scoring	Uniform procedure of administration is possible but it is usually flexible.	The procedure of its administration is standardized and specific instructions for its administration and scoring are provided.
5.	Interpretation of Scores	Scores can be compared and interpreted only in the context of the local school situation.	Scores can be compared to norm-groups. Test manual and other guides aid interpretation and use.

5.3 Intelligence Tests

The intelligence tests which are otherwise called the Proficiency Tests attempt to measure an individual's capacity to undertake difficult tasks, ability to pursue relationships, capacity to undertake activities that are complex and characterized by abstractness, ability to solve problems and apply knowledge in a variety of contexts etc. Intelligence tests are used to measure general scholastic achievement.

The intelligence tests could be either verbal or non-verbal and they could be categorized into two kinds: Group Tests and Individual Tests. The similarities and differences between the two may be seen from the following:

- In group tests, problems are presented in printed booklets. They are read by the examinees (students) and there is minimum contact of the examinees with the examiner. On the other hand, in individual tests, problems are presented orally by the examiner in a face-to-face situation.
- In case of group tests, the test is timed as a unit, or separate time limits are set for each sub-test, while in case of individual tests, problems are presented one at a time, usually without indication of time limits.
- Lastly, in group tests, each individual student usually responds by selecting one of a limited set of response options printed in the test booklet whereas in individual tests, individual student usually responds freely, giving whatever response seems appropriate to him.

There are various ways in which scores of intelligence tests may be used. Some of them are for:

- determining reading readiness,
- determining whether or not any student's achievement approximates his potential,
- determining eligibility for subjects or courses requiring a high degree of intelligence,

- vocational and educational guidance in secondary schools,
- selection of reading materials for given classes,
- assigning students to special classes or curricula, and
- identification of learning disabilities.

5.4 Aptitude Tests

Again, as it is clear from the name of the test, the outcome or findings of such tests will indicate the bent of mind of test takers towards the subject concerned. According to Good, aptitude test is "a group of characters deemed to be symptomatic of an individual's ability to acquire proficiency in a given area; examples might be a particular art, school subject or vocational area"[4]

Aptitude tests tend to be limited to tasks that are equally unfamiliar to all students. They are again of two types – verbal and non-verbal. They are like the intelligence tests administered either individually or in groups. The selection of a particular type of test would depend upon the use for which the test is intended and the grade level at which the test is to be used.

While the above are the kinds of tests made use of for purposes indicated against each, the kinds of test items are manifold. They include the multiple-choice, true/false, matching, re-arrangement, completion, short answer, long answer etc. besides the paragraph and essay type of questions and précis/epitomization. The nature, structure and use of these different types of test items is described and presented later in the present book.

The types of tests discussed above are various kinds of tools/ instruments made use of in the context of performance evaluation i.e., learner evaluation with reference to the process of learning that the learners go through.

It would now be relevant to discuss about various tools/ instruments that are made use of in the process of evaluating other aspects of education viz., curriculum and syllabus, instructional objectives, instructional materials, instructional methods and media and evaluation of instructor's performance. Such tools include the

rating scales, anecdotal records, checklists, questionnaires, interviews etc.

5.5 Rating Scales

Rating scales are made use of for systematically obtaining, consolidating and reporting the judgments of observers on the constituent evaluated. As the process involves close observation, this type of instrument comes under "observational technique". It contains a well planned set of characteristics or qualities to be judged and a predetermined kind of scale to measure the degree to which each attribute is present in the constituent evaluated. The scale prescribed normally is a five-point scale and in some cases, it could be seven-point. Nine-point scale is a rarity. As the process of evaluation has to be realistic and objective in nature, it would become necessary to make the evaluation instrument as effective, precise and efficient as possible to make the observations and conclusions valid, reliable and objective. In the context of making the rating scale meet these requirements, the following points are to be kept in mind:

- Rating scales have to be in consonance with and relevant to the objectives and the learning outcomes. Needless to reinforce that these objectives have to be in behavioral terms so that a thing can he effectively evaluated with the help of the characteristics and the rating scales included in the tool.
- It is necessary that the rating scales should be made applicable only to the behavioral characteristics. Whether the constituent evaluated is the syllabus/curriculum, instructional material, the objectives, methods or media or the teacher performance, the items in the rating scale should be so worked out that they focus on different characteristics of the aspects evaluated. It must be made doubly sure that the characteristics are clearly observable and can be rated with ease with reference to the learners. Rating scales could be applied to the classroom discussion, seminars and the like.
- The characteristics and points in the rating scale should be clearly defined. They should be simple, unambiguous and must

precisely state the characteristics to be measured without any confusion and without their being misleading in nature.

- In order to avoid the possibility of the personal biases of the raters, it is desirable that two or more raters are employed for the same purpose thereby taking into consideration a consolidated viewpoint of the group of raters.
- It is but natural to use the rating scales for rating different dimensions of the constituent evaluated. In order to ensure that the rating of one or some characteristics don't influence the rating of others, it is desirable and necessary that each dimension is looked at as a totally separate entity without any relation whatsoever with the other entities and then rate each of them independently. Another strategy adopted in trying to avoid the influence of one or some ratings on the others is that the items in the rating scale are so arranged that sometimes the right extreme of the scale happens to be high and sometimes low and vice-versa. Incidentally, the effect resulting from the possible influences of one or some characteristics/ratings on the others is called "halo effect". It is defined "as the bias resulting from the tendency of a rater to be influenced in his rating of specific traits by his general feeling or impressions or the ratings already executed". All possible efforts must be made to keep the intensity of the halo effect the least, if it cannot be totally avoided.

Generally speaking, the uses of rating scales are classified into three major areas of evaluation viz., (1) Procedure, (2) Product and (3) Personal and Social Development.

Evaluation of methods, materials and media, programs and processes come under the first category.

Product evaluation refers to the end product of a program or constituent evaluated. This helps us in judging the quality of the end product and its relative assessment with reference to the expected outcomes. Further, the products so evaluated can be arranged in a

rank order or they can be compared to a product scale. A product scale is a series of samples of the product which have been very carefully graded to represent different degrees of quality.

Unlike the evaluation process in the context of procedures, the ratings in the area of personal and social development are obtained at periodic intervals and they represent the kind of consolidating and summing up of the general expressions that the evaluator has about the constituent evaluated. Such evaluation however, is prone to the personal biases of the raters.

5.6 Anecdotal Records

Anecdotal records are the descriptions of significant aspects pertaining to the individuals during the course of instruction – be it the instructor or the learner. Such descriptions pertain to occasional and specific incidents wherein they have opportunities to exhibit their skills and talents with reference to instruction or learning. The quality of performance of the instructors/learners during such occasions will provide an opportunity for the evaluators to assess such an incident which can become the supporting evidence in respect of the overall evaluation. Again, the evaluation of anecdotal records is prone to be subjective and biased. Therefore, such measurement and assessment based on anecdotal records have to be carried out by a group or team of evaluators. This is the only way to minimize the degree of bias and subjectivity. Efforts should also be made to work out various criteria for assessing the anecdotal records to make the assessment as objective as possible. There are several limitations of anecdotal records. Primarily, they consume a lot of time and cause lot of difficulty in maintaining a system of records. The other limitation is that neither these records themselves nor the evaluation thereof can be made highly objective resulting in less reliable value judgments and conclusions. Normally, these anecdotal records are used in examining and assessing the social development of the individuals or even with reference to the performance in any or all constituents of education. The significant characteristics of anecdotal records are the following:

- To a very great extent what exactly is observed is recorded in them. Inferences, causes, assumptions etc., are avoided and omitted unless they are clearly labeled as inferences, causes, and assumptions.
- The observers have to clearly decide for themselves as to what aspects of behavior are related to the dimensions observed.
- Anecdotal records have to be made cumulative and a plan of periodic observation and recording has to be established and stuck on to.
- Words or phrases are made use of which make the item crystal clear instead of complete sentences to ensure brevity.
- The nomenclature made use of in these records can he defined in terms of things. Therefore, the statements included in the anecdotal records consist of concrete statements and not abstract ones. Generalized terminology like *well behaved, highly talented, nervous, industrious* etc. should be avoided as they make little sense looked at from the point of view of objective assessment and evaluation.

5.7 Checklists

Rating scales and checklists are very similar to each other. The basic and primary difference between the two is that while rating scale provides a chance to indicate the degree to which the character is present, the checklist does not provide such an opportunity. Checklists consist of simple *Yes/No* judgment. It is for this reason that checklists are not made use of where a highly objective judgment becomes necessary. Normally, checklists are used in the context of evaluating procedure, products etc. Further, they could be useful in consolidating and summarizing a host of observations made through an extended period of time. Although, the checklists are made use of quite frequently, they are not really efficient tools for objective evaluation.

5.8 Questionnaires

A questionnaire contains a set of written instructions and questions through which information relating to different types and characteristics is elicited. Broadly speaking, there are two kinds of questionnaires, viz., (1) Structured and (2) Unstructured. The former kind of questionnaires is called closed forms of questionnaires and the later open forms of questionnaire. A structured or closed form of questionnaire consists of a host of questions and a set of alternative answers under each. Such answers have to be exhaustive of all possible responses besides their being mutually exclusive. The unstructured or open forms of questionnaires do not include the suggested answers under each of the questions. It may thus be noticed that the administration and scoring of a structured questionnaire is systematic, straightforward, reliable and objective. The unstructured or open forms of questionnaire cannot be objective and reliable as the structured questionnaires are. Further, the analysis of the data obtained from the unstructured or open forms of questionnaires is extremely difficult while it is very easy and dependable in the case of structured questionnaires In order to make the questionnaires more effective, the following aspects are to be kept in view:

- The questionnaire must deal with a significant topic in a particular field in which it is used.
- Each question must be precise, unambiguous and must aim at a specific task and should not provide scope for any overlap with other aspects and skills.
- The questionnaire must contain instructions and guidelines for the respondents in very clear, unambiguous and direct terms as to what precisely is expected from the respondents.

5.9 Interviews

Interviews are very widely used as assessment techniques in several contexts. The most significant and important quality of interviews is their flexibility. However, a word of caution has to be recorded that to a considerable extent, the performance of the

interviewee depends upon the degree of favorable rapport established in the first few minutes of the interview with the interviewers. Again, interviews are of two kinds: Structured and Unstructured. In the case of structured interviews, the questions are predetermined and prearranged and are invariably put to all the interviewees. The advantage of a structured interview is that it is standardized and consequently the analysis of the data becomes easier. The main drawback of a structured interview is that it is inflexible, although, on the other hand, the reliability is increased and the depth is decreased. The unstructured interviews are more informal. The direction in which the unstructured interview proceeds has a strong bearing over how the interviewee and the interviewer lead each other. There is little scope in unstructured interviews for making a relative assessment of different interviewees. One significant advantage of unstructured interview is that the respondents have freedom of expression whereby they have opportunities to establish their individual talents and skills. Added to this, the interviewers have to be highly skilled and they have to possess very high degree of expertise in assessing the interviewees. Following points have to be kept in mind to make the interviews very effective:

- Pertinent information has to be obtained and organized in preparing for the interview,
- The interviewers must create a tension-free, friendly and congenial atmosphere for the respondents to be bold and strong and be able to unhesitatingly express views, feelings, ideas etc.
- The interviewers must be active listeners and passive speakers thereby providing more opportunities for the respondents to speak.
- The interviewers should refrain from expressing their impressions or opinions during the course of interview.
- Adequate opportunities and situations must be created so as to enable the respondents to think critically and creatively as much as possible in responding to the questions,

- Greeting the respondents at the entry time and bidding adieu at the close of the interview must be ensured by the interviewers which help the respondents in reducing their tension.

In addition to the various tools/instruments discussed above in the context of evaluation, there is another category of tools/ instruments called 'socio-metric techniques'. Normally, such techniques are made use of for studying the organization of social groups and for evaluating the social acceptance of individuals in a group. The purposes for which Socio-metric Techniques are made use of include –

1. Classroom organization into groups,
2. Improving the social adjustments of individual students, and
3. Evaluating the influence of school practices on students' social relations.

CHAPTER

6

Program Evaluation in Education

Discussion regarding various kinds of tools/instruments adopted in the process of evaluating, different kinds of constituents in the context of education have been discussed. It has further been clarified that the concept of evaluation is a broader one covering several aspects of education looked at from global point of view including therein curriculum, syllabus, instructional materials, media, methods etc., and is not confined to the process of measuring and evaluating learner performance. Consequently, it would be ideal to take up these constituents one after the other and discuss in detail as to what they are, what their nature is like and suggest the adoption of relevant tools/instruments for evaluating each of them.

6.1 Curriculum Evaluation

'Curriculum' in the educational context primarily refers to the specifications of learning outcomes in behavioral terms. Further, it should be clear from the curriculum as to the level of instruction, the entry behavior and exit behavior of the learners, the nature of materials, methods and media suggested, time limit for the course of instruction and generally focus on the content and skills involved in the program under consideration. It should also indicate whether the takeoff point of the course of instruction has a bearing over the end

point of the previous level of instruction, and similarly whether the end point of the present course of instruction would lead towards the takeoff point of the next level of instruction. In addition, curriculum should also provide in brief the guidelines for the instructors to observe during the course of instruction.

While this is applicable to the curriculum relating to any aspect of education, language curriculum is slightly different in nature from what has been mentioned above. A language curriculum, in addition to ensuring the above aspects, has to relate the skills and content together and integrate them in such a way that the progression of language instruction goes on smoothly and has a cumulative effect as the process continues and it becomes a continuum. The skill and content again have to be related to the instructional materials, methods etc. Looking at curriculum from the point of view of evaluation, there are two main purposes. The first purpose is related to collection and elicitation of information which could be made use of in revising and improving upon the programs, processes and products besides methods, materials and media of instruction. Thus, the curriculum makers get feedback and information about the merits and demerits of the curriculum in question which will facilitate working for betterment of such curricula. The second purpose relates to the decisions regarding the adoption of courses including materials depending upon the appropriateness of aims and objectives. It follows that the organization of curriculum has to be ensured based on certain principles and in a graded manner. In fact, curriculum organization process has been looked at as a technical process involving steps like selection of facts, concepts, ideas etc., with reference to the mental maturity level of the consumers. It further involves the study of relationship among different elements of the content, and their organization into integrated wholes or units on strict academic principles. The arrangement of content etc. has to be executed in a strictly graded manner. The curriculum planner has to take decisions based on highly objective data such as facts and realities, concepts and perceptions, definitions and their relationships, phenomena and

evidences relating to them, interests, needs and problems, capacities and capabilities of the consumers. A highly structured and organized curriculum is an invaluable tool for the classroom instructor. The reason is that it clearly specifies the length and breadth of the content and skills indicating therein the activities and learning experiences that must be transmitted to the consumers. Curriculum also helps the instructor in clearing, understanding and planning teaching strategies and techniques. Normally, the following principles are talked about in the context of curriculum planning:

1. Principle of Utility
2. Principle of Variety
3. Principle of Need
4. Principle of Relevance
5. Principle of Readiness
6. Principle of Flexibility

The principle of utility refers to the utilitarian and useful outcome of instruction which has a bearing over life in general and the personal, social and occupational aspects in particular. From this point of view, the curriculum planning is expected to be organized in such a way that the content, activity and experience included in different stages of the curriculum have a corresponding positive impact on the lives of the learners.

The principle of variety relates to several aspects like the instruction being theoretical, practical, aesthetic, vocational, occupational and professional in nature. Therefore, looking at the overall requirements of an individual's life, all these aspects will have to be included in an ideal curriculum so as to ensure developing a desirable and ideal personality besides taking care of the material advantages of the consumers. The choice, selection and inclusion of these aspects will have to cater to the interests, needs and aspirations of the learners. Further, a wide variety of content and skills makes the teaching-learning process interesting and motivating.

The principle of needs refers to the practical utility of knowledge attained through the course of instruction besides leading the learners to occupations and professions through which they derive their bread and livelihood. Needless to state that the needs of different individuals will be different. Therefore, depending upon the level of instruction, a common base with reference to the needs may have to be drawn up in case of early education and as the level of instruction goes higher and higher, specialized need-based courses may have to be devised catering to the needs of specific groups of learners. Therefore, a good curriculum in the early stages of education has to provide a clear vision to the consumers which will have a focus on their future needs. It has also to inspire and motivate the learners towards disciplines leading to vocations, professions and occupations. Scholars in the field of education have observed that the principle of need in the process of curriculum making has been considerably lacking and recommended that this aspect be invariably taken care of particularly at the school level of instruction.

The principle of relevance refers to the needs and problems of the learners, the society to which they belong and finally the school where they study. These needs and problems and social aspects differ from group to group. This difference necessitates the devising of different curricula for different groups. In the present day context, efforts have been initiated to mould the curricula keeping in view the needs, aspirations and interests of the learners. A word of caution has to be taken note of that, to be idealistic from the point of view of relevance, curriculum making will have to be localized. The implications are that different kinds of courses may have to he designed for which specially designed materials, methods etc., have also to be devised. Although, the principle of relevance is also very important in planning a good educational program, practical constraints do not permit localization of curriculum-making to meet the ideal situation. However, efforts need to be made to the extent possible in this direction atleast from the post-schooling stage of education.

The principle of readiness refers to the psychological bent of minds of the learners to learn and receive instruction. Thus, readiness depends upon the learner and the environment. The learners' factor includes the biological maturity, the individual's abilities, values and goals, the aspirations and previous learning in relation to their knowledge, skills and attitudes. The environmental factor refers to the conditions of classroom, climate and process of teaching. Curriculum, looked at from the point of view of readiness principle, must suggest and indicate the stages at which instruction has to be imparted, the conditions under which it is to be carried out and the manner in which it is to be executed. It has been suggested by scholars in the field of educational psychology to fall back upon the research evidences with reference to biological maturity, psychological preparedness, the requisite knowledge, skills and attitudes. In fact, the principle of readiness in the context of education results in excellent outcomes of instruction.

The principle of flexibility implies autonomy at three different levels – at the school level, at the teachers' level, and the learners' level. At the school level, curriculum prescribes a wide variety of courses like core courses and elective courses, compulsory and operational courses, academic, practical and vocational courses and provides scope in the institution to offer different courses depending upon the interests, needs and problems of the learners. At the teachers' level, a good curriculum gives a free hand for the instructor to choose his methods and strategies and techniques for imparting instruction. At the learners' level, the curriculum must provide an opportunity and good choice for the learners to choose the courses of their liking and such courses which appear to be beneficial for their own development. Therefore, an ideal curriculum at the initial levels must provide a broad vision for the learners so as to enable them to decide for themselves regarding the choice of courses depending upon their own interests, needs etc. Again, as it is true in the case of principle of relevance, it is also true in the case of principle of flexibility to narrow down the scope as the level of instruction goes higher and higher.

Therefore, in the context of evaluating curricula, the aspects discussed above will have to be kept in view and necessary and relevant tools developed for the purpose. Most ideal tools for curriculum evaluation are checklists and rating scales. A suggested checklist/rating scale is presented hereunder.

- Does the curriculum clearly specify the contemplated aims and objectives?
- Are the aims and objectives defined in terms of learner behavior?
- Does each main objective contain the related sub-objectives?
- Does the curriculum indicate the level of instruction?
- Does the curriculum specifically indicate the end point of the preceding level of instruction?
- Does the curriculum indicate the entry point of the next higher level of instruction?
- Does it appropriately integrate the skills and content?
- Does it indicate the clear gradation of content and skills?
- Does the curriculum indicate the type and nature of instructional materials, methods and media?
- Does it provide guidelines and suggestions for the instructor enabling him to choose the right kind of instructional methods, techniques and strategies?
- Does it specify the total time ceiling for the course of instruction and the breakup thereof for different constituents of instruction?
- Does the curriculum emphasize upon the concept formation, critical thinking, creative abilities etc.
- Is the curriculum well organized from the point of view of different principles of utility, variety, need, relevance, readiness and flexibility?

- Is the approach for curriculum making relevant and appropriate?
- Which is the approach adopted in the formation and organization of the curriculum (subject-based approach, activity-based approach, environment-based approach)?
- To what extent does the curriculum lead to working out a suitable syllabus, materials, methods etc.?

The above checklist is only suggestive and depending upon the specific requirements, additions, deletions, modifications and alterations may have to be made. Adopting a 5-point or 7-point scale against each item converts this checklist into a rating scale.

6.2 Measurement and Evaluation of Instructional Objectives

Instructional objectives in any educational program are required to be very specific in nature and have to be formulated in terms of the learners' behavior. Therefore, they have to be related with the entry behavior of the learners and their terminal behavior. These learning objectives have to provide a description of the process which the learners are required to undergo during the course of instruction at different stages. They have a bearing over the content and skills required to be covered during the process of instruction. Normally, these instructional objectives are made the basis for measuring and evaluating the progress of the learners. But, the instructional objectives as such are not generally evaluated. Although, scholars like Dressel[5] have maintained that certain problems and considerations have to be taken into account when formulating objectives for evaluation purposes, there are contradicting views about the evaluation of instructional objectives themselves. It should, however, be accepted in principle that as there is a definitive method and manner of developing instructional objectives, consequently they can be evaluated with reference to the methods and manner involved.

Instructional objectives can be examined from several angles depending upon the characteristics involved. Such objectives could be

wide ranging from very simple to very complex. While the simple objectives could be assessed very objectively, it is not as easy to assess the complex objectives objectively but they could be at best rated on a rating scale. It should be taken note of that excessively simple and specific objectives are prone to break the education into a number of unrelated bits, while the very complex objectives get ignored in the process of instruction. It is, therefore, necessary to assess them from the objectivity point of view preferably on a rating scale than merely deciding upon about their being simple or complex. Certain objectives could be achievable and certain others non-achievable. This characteristic is also similar to the simplicity and complexity of objectives though not identical. The achievable objectives are those which do not specifically provide a description; for e.g., an objective "the learner develops the knowledge ...." does not specify precisely the degree and extent of knowledge and therefore, it can neither be assessed objectively nor will it be an achievable objective. In the case of achievable objectives, objective evaluation is possible whereas in the case of unachievable objectives, evaluation can take place only on a rating scale. Again, the instructional objectives could be explicit or implicit. Just as in the case of achievable versus non-achievable objectives, measurement could be objective in case of the explicit objectives whereas it becomes subjective in the case of implicit objectives. Some scholars have classified the instructional objectives into two categories viz., Intrinsic and Transcendental. Intrinsic objectives are those which are inherent in the nature of the materials or the experience. For instance, the process that the learners are expected to go through in 'drilling' in the case of language instructional programs or "multiplying" in learning mathematics refer to the intrinsic objectives. Thus, the objectives that specify the thought process and value judgments do not come under the intrinsic objectives. They are classified as transcendental objectives. It is but natural that instructional objectives pertaining to any educational program could include both intrinsic and transcendental objectives. Again, the intrinsic objectives could be objectively evaluated whereas transcendental objectives could at best be rated on a rating scale, thereby their evaluation being not highly objective. Certain objectives

could represent the goals of groups and certain others could represent the individual's goals. The individual-oriented objectives refer to the development of an individual whereas group objectives refer to the development of a group of individual as a whole. Thus, while the individual-oriented objectives could be more specific and can be subjected to highly objective evaluation, group objectives can again at best be rated on a rating scale.

Bloom's Taxonomy of Educational Objectives is a monumental work which deals with the nature and kind of educational objectives. The analysis of educational objectives in this monumental work has been based on their cognitive, affective and motor domains. Depending upon the nature and kind of the educational program, the stated objectives have to be categorized as one among these three. Therefore, while evaluating the instructional objectives, they could be examined with reference to these categories.

To be more specific, the instructional objectives provide a direction both for the education program as a whole and the learners. It follows that different constituents of education like the instructional methods, materials, media etc., have to be picked and chosen with reference to the educational objectives. Added to that, the arrangements of these objectives should be such that they provide a continuity, sequence and integration with reference to the skills and content. The question of evaluating instructional objectives as such still remains controversial as scholars differ whether at all the instructional objectives are required to be evaluated while some others hold the view that the instructional objectives could merely be taken as the basis with reference to the evaluation of the progress and outcome of the learners in an educational program. In spite of these contradictions, as already stated, for the mere fact that there is a specific method and manner of developing instructional objectives, there could be no doubt in maintaining that the instructional objectives can be and should be evaluated atleast from the point of view of their relevance, appropriateness and meaningfulness. It is from this angle that a suggested checklist is given hereunder:

- Are the instructional objectives stated in clear, specific and unambiguous terms?
- Are the instructional objectives stated in terms of the learners' behavior?
- Have the instructional objectives been related with the entry and exit behavior with reference to the educational program?
- Are the educational objectives stated in terms of specific achievement of the learners?
- Do the educational objectives consist of both explicit and implicit objectives?
  - (a) If the answer to No.5 is 'Yes', what is the ratio of explicit and implicit objectives?
- Do the instructional objectives include both intrinsic and transcendental objectives?
  - (a) If the answer to No.6 is 'Yes', indicate the ratio of intrinsic and transcendental objectives.
- Do the instructional objectives refer to the individual goals or societal goals?
  - (a) What is the ratio of objectives focused on individual goals and societal goals?
- Do the objectives belong to cognitive, affective or psychomotor domains and do they have bearing over the contemplated educational achievement of the learners?
- Do the objectives clearly indicate the choice of a particular set of instructional materials, a particular instructional method and medium?
- Are the objectives arranged in such a manner that they provide a continuity, sequence and integration?
- Does each main objective consist of related sub-objectives?
- Can the objectives be easily classified into area-wise, subject-wise and unit-wise objectives for the benefit of the instructors and the learners?

The above checklist is only suggestive and has to be reformulated, modified, amended with additions, deletions etc., depending upon the set of instructional objectives chosen for purposes of evaluation. Adopting a 5-point or 7-point scale in the above checklist makes it a rating scale.

6.3 Evaluation of Instructional Material

Instructional material in any educational program can be viewed at from two angles viz., (1) Primary Instructional Material and (2) Supplementary Instructional Material. The purpose of primary instructional material is to make use of it directly in the classroom or any other medium chosen for instruction and this is designed and formulated in accordance with the curriculum and syllabus. The supplementary instructional material is normally made use of for purposes of reinforcement of instruction that goes on in the classroom directly and is expected to facilitate the learners in overcoming their problems/difficulties in the process of learning. Except in the case of distance education, the primary instructional material is made use of for instruction in direct interaction with the learners and supplementary material is made use of during tutorials or specialized individual instruction given to the needy learners from time to time.

As sufficiently clarified in the preceding pages, the designing and development of instructional material has to be in accordance with the aims and objectives of instruction and they have to be moulded in a manner that suits the medium in which instruction is proposed to be imparted. Normally, in the context of evaluating instructional material, an exhaustive questionnaire/checklist is used based on the findings of which conclusions regarding the quality, adequacy and appropriateness of textbooks are arrived at; needless to mention that the factors for consideration in evaluating instructional material depend upon the subject of instruction. For instance, lot of importance is given to the linguistic aspects like the vocabulary, structures, lessons, formats, drills, exercises etc., in the context of evaluating language instructional material. On the other hand, if the

material pertains to a non-language subject, importance is given to factors relating to the appropriate selection of the subject-related themes, and their arrangement in a suitably graded manner thereby ensuring the continuum of the theme and subject. Another aspect with reference to the evaluation of non-language subject instructional material is that the exercises in this case are focused on the content rather than the expression. Similarly, there are some other differences between the language and non-language subject instructional materials which have to be kept in view in the process of material evaluation.

This being the case, it is desirable to attempt discussing about the evaluation process involved in the language instructional material as opposed to that of the non-language subject instructional materials separately.

The language instructional material normally aims at developing the language skills among the learners. The skills involved depend upon the level of instruction. If it is the beginning or primary level of instruction, the basic/primary language skills are emphasized upon which include Listening, Speaking, Reading and Writing. If the level of instruction is higher or advanced, the skills emphasized upon are higher order language skills viz., Representation, Argumentation, Refutation and Establishing one's own point of view. Further, irrespective of the level of the instruction, the skills involved have to be suitably integrated with the content besides ensuring the adherence of the four principles of material production in language. They are Selection, Gradation, Presentation and Repetition. Apart from these aspects, the material evaluation in language has to be with reference to all elements of a language lesson/ language instructional material such as the form of lessons, various constituents of a lesson like the vocabulary, their glosses and meanings, their usage, and exercises, grammatical explanations and cultural notes etc.

It has been the practice that the scope and extent of evaluating language instructional material has been, at best, confined to the examination of the material with reference to the objectives and

syllabus and the aspects of material preparation mentioned above. It has seldom been realized that this is not sufficient for a reliable and valid material evaluation. A more relevant aspect to ensure highly objective, reliable and valid material evaluation is with reference to the comprehensibility of material. Some scholars have viewed the task of assessing comprehensibility as a separate entity in itself without relating it to the other two aspects. However, it is to be made clear that a complete and overall, useful and reliable evaluation of instructional material is possible only when the outcomes of the three aspects mentioned above including the comprehensibility of textbooks are consolidated and comprehensive conclusions drawn about. As far as the general process of material evaluation is concerned, it has two aspects viz., 1 (A) Academic aspects, 1 (B) Other academic aspects and (2) Physical aspects. They could be classified as follows:

1 (A) Academic Aspects

a) Purpose of instruction
b) Learners' age group
c) Learners' previous knowledge
d) Duration of instruction
e) Selection
f) Gradation
g) Presentation
h) Repetition
i) Variety of language

1 (B) Other Academic Aspects

a) Content
b) Planning and presentation
c) Visuals
d) Drills and Exercises
e) Style

2. *Physical Aspects*
   a) Book form
   b) Letter size and shape
   c) Quality of paper used
   d) Quality of printing and printing mistakes
   e) Cost of books
   f) Colors used

The above aspects can be viewed at from two different angles viz., (1) Micro-linguistic factors and (2) Macro-linguistic factors.

The Micro-linguistic factors are:

1) Orthography/Spelling system
2) Morphemic/Lexical system
3) Syntactic/Syntagmatic system
4) Discourse/Textual system

In order to ensure that evaluation from these angles is very intensive, each concept mentioned above has to be sub-categorized and the material evaluated accordingly. Such sub-categorization could be as follows:

1. *Orthographical/Spelling System*
   a) Graphemic Variation
   b) Diacritics
   c) Punctuation
   d) Letter Design
   e) Orthographic/Printing Errors
   f) Frequency and Progression
2. *Morphemic/Lexical System*
   a) Colloquial to Cultivated
   b) Neologisms

c) Incidence of Loan Words

d) Random/Competing Spellings

e) Compound vs. Analytic

f) Difficult Words

g) Spelling Mistakes

h) Use of Wrong Words

i) Wrong Glosses

3. *Syntactic/Syntagmatic System*

a) Grammatically Deviant

b) Wrong Conjunctions

c) Semantic Deviances

d) Ambiguous Phrases/Expressions

e) Sentence Types

4. *Discourse/Textual Norms*

a) Organization of Discourse

b) Paragraph Organization

c) Participant Identification

d) Unity and Coherence

e) Stylistic Appropriateness

f) Ideational (Factual Errors)

The Macro-linguistic factors are:

1) Learner factors

2) Instructional factors

3) Presentational factors

4) Socio-cultural factors

Just as in the case of micro-linguistic factors, the above macro-linguistic factors have also to be sub-categorized and the material

looked at from the point of view of all the sub-categories. The sub-categorization could be as follows:

1. *Learner Factors*

   a) Ability to understand

   b) Aptitude study

   c) Educational achievement

The factors relating to ability to understand are the same as those of comprehensibility. This is basically taken care of by making use of the following tools:

i) Cloze Tests

ii) Comprehension Tests (listening and reading)

iii) Multiple-Choice Tests

The readability and comprehensibility of textbooks is discussed in detail in the further pages to come.

The aptitude study relates to the understanding of learners' readiness to receive instruction. In order to ensure a good aptitude for the learners, the material has to meet the following requirements:

i) Interesting

ii) Exciting

iii) Provoking

iv) Novel

v) Activating

vi) Applicational

vii) Compatible with experience and suitable to the level

The educational achievement is measured by evaluating the kinds of questions and exercises used in the prescribed instructional material. The quality of questions and exercises are judged based on the following criteria:

i) Coverage (the degree of coverage which the questions have in relation to the content of the lesson)

ii) Range (whether the questions represent different item types or not)

iii) Depth (are the questions such that the students have to think critically and creatively before answering)

iv) Mode (do the questions require answers based exclusively on rote learning or do they invite students to think and reason on their own)

The second macro-linguistic factor is the instructional factor which has the following breakup:

a) Self-sufficiency

b) Adequacy

c) Transfer effects

d) Organization of language instruction

e) The language styles and identification of learners

Self-sufficiency

This includes Teachers' Handbooks, aims and objectives, and gradation.

Adequacy

Appropriateness of presentational forms of lessons, exercises and drills *(Spoken and Written Skills)*.

Transfer Effects

Highlighting of significant points and motivating learners to relate these points with their day-to-day life and experience.

Organization of Language Instruction

Simple to complex, known to unknown.

Language Styles

Should be formal, written, and standard. Mix up of spoken and written etc., should be avoided. Ad-hoc use of styles should be avoided.

The third macro-linguistic factor is the presentational factor which can be sub-divided into two parts viz., (a) Extrinsic factors and (b) Intrinsic factors. The extrinsic factor relates to the physical aspects of the material including therein aspects like the size of the book, the cover, the type face, colors and balance of visuals and non-visuals. The intrinsic factors include lesson titles, the text organization, cohesion and recapitulation. These factors include the lesson format, the logical and sequential arrangement of ideas, development of themes, paragraph organization and necessary reinforcement. The socio-cultural factors include location, demography, levels of education, functional distribution of languages and styles in the community. The location of schools refers to whether place of their location is urban or rural. Demography relates to the social and technical background of the learners classifying them into forward or backward communities. Levels of education refer to the educational background of parents/guardians and the functional distribution refers to the specific speech community of the learners, functional use of different languages for different purposes and styles in different contexts.

Suggested evaluation sheets for posting the data may be seen in the Appendix-1.

As already stated above, the comprehensibility aspects in the context of evaluating instructional material is normally not taken care of as it is viewed as an independent and separate entity in itself. However, as it is strongly felt that this aspect is an integral part of the overall process of evaluating instructional material, the nature and processes involved in the assessment of comprehensibility are discussed hereunder.

6.4 Comprehensibility of Textbooks in Indian Languages

Language teaching/learning can be basically construed as consisting of curriculum and syllabus, instructional material and the teaching methodology. All the three aspects are inter-linked in that instructional material or textbooks are expected to be related to the

curriculum and syllabus and consequently, teaching methodology essentially has to have a bearing over the instructional material in terms of content, form of presentation, vocabulary and linguistic structures. Needless to say that teaching techniques, which are idiosyncratic to each individual instructor, are something apart from teaching methods as such.

For any language instructional program to be effective and meaningful, it is necessary that the instructional material based on the syllabus and the curriculum has to be planned keeping in view the requirements of the recipients i.e., students, their educational background, their age group and their proficiency in the language which is the medium of instruction. It is all the more necessary to consider the control of vocabulary and sentence structures with reference to the students while preparing the instructional material. It is also equally essential, if not more, to adopt proper forms of presentation of lessons which should help in motivating the learners properly and maintaining consistency of motivation throughout the course of instruction.

The development of instructional material or the preparation of textbooks etc., in India is a unique phenomenon. It is paradoxical that the syllabus makers, the textbook writers and the teachers are in parallel lines and therefore there is no liaison or coordination among them. As a result, a very powerful adverse effect emerges from the point of view of instruction and learning. One could even proceed to state that the preparation of textbooks and other instructional material has practically become commercial and little importance is given to the relevance of the textbooks with reference to the syllabus and the learners' needs.

While the evaluation of textbooks is a global and in-depth concept which has many facets, 'comprehensibility' is a narrower concept which needs to be looked at from the point of view of the learners. 'Comprehensibility' can be construed as the ability of the children to decipher the printed letters, understand the vocabulary and sentence structures in addition to the ability to conceptualize the theme properly. Looking at this aspect of 'comprehensibility' at the

primary level, certain other facts also need to be kept in view. For instance, the physical aspects of the instructional materials like the size and type of letters made use of in instructional material, spacing in-between lines, illustrations and other visuals used in the textbooks, clarity of print, etc.

If one has to claim that a given instructional material is "comprehensible" for the learners, it has to be ensured that the language used in terms of vocabulary and syntax, ability to conceptualize and the physical aspects mentioned above are properly taken care of and further ensure that they are very much within the reach of the target audience.

In the context of universalization of education, Education for All and in particular DPEP, MLL etc., the aim is to impart education to one and all at the primary level. It becomes essential to assess the 'comprehensibility' of textbooks so as to examine whether the aims and objectives set forth are met with through the use of textbooks.

'Readability' has been defined by some scholars as the learners' ability to read. This does not involve the comprehension aspect on the part of the learners. One's ability merely to read without understanding or comprehending does not make much sense. Therefore, several other scholars have defined the concept of readability as one's ability to decipher the printed letter and comprehend the material. Mechanical reading makes little sense in any context and therefore one might even proceed to argue that comprehension is more crucial than mere deciphering of what is printed. In this context, it may be noted that communication has at least two media viz., the written medium and the oral medium.

Assessment of 'comprehensibility' could both be subjective and objective. Subjective assessment would naturally be highly questionable as it does not depend on any empirical data and it could be highly hypothetical and cannot be relied upon. Therefore, the objective assessment of comprehensibility of textbooks becomes all the more essential so as to determine the suitability, appropriateness

and the applicability of the instructional material to specified target audience.

Objective assessment of comprehensibility is a highly technical task which requires various tools including the tests of various kinds; questionnaires, observation schedules etc. While the tests need to be administered to the target children, the observation schedules are to be adopted by both the teachers and the learners and the questionnaires need to be adopted by instructors or evaluators.

It is also essential and fit to make a content analysis and frequency counts of selected portions of the textbooks from the language point of view, the findings of which may or may not support and substantiate those arrived at through administering tools.

6.5 The Tools for Comprehensibility

A wide variety and range of tools could be thought of for language evaluation including the assessment of comprehensibility; needless to mention that a wide range and variety of tests is available whether it is the testing of language or any other subject for that matter. Each type of test has its own merits and demerits and limitations in achieving something. There have been lots of controversies as to which type of test should be adopted and which could be reliable for an overall assessment of 'comprehensibility'.

Some people have argued in favor of listening and reading comprehension tests to make a global assessment of comprehensibility while some others have argued in favor of vocabulary tests and tests concerned with the sentence structures separately and some others have argued in favor of oral tests. However, there has been a general consensus in favor of both comprehension and cloze tests. The vocabulary tests however, cannot be totally ignored. Considering all the controversies in this regard, it is perhaps appropriate to adopt cloze tests which are aimed at providing an idea of the overall readability and comprehension besides the reading comprehension tests and vocabulary tests. Preparing or developing such tests also depends practically on the subject matter. Although the cloze and the reading comprehension and vocabulary tests could be thought of

with reference to language, environmental sciences; such cloze and comprehension tests cannot be easily constructed with reference to subjects like mathematics. The reasons for such constraints are that the mathematics textbook is primarily teacher-based and consequently a high degree of dependence of the learners on the teacher exists and not much description is warranted in mathematics as more explanation is essential. Further, there cannot be much of a running text material in mathematics and perhaps more of language is found in instructions and exercises than the text materials as such.

6.6 Content Analysis

Content analysis of language textbook or instructional material will have at least two aspects – linguistic content and thematic content. The linguistic content is primarily concerned with the nature and category of vocabulary, complexity of sentences and the discourse pattern. The thematic content is concerned with the complexity of themes/topics, their relevance and applicability to the daily lives of the target audience etc. Not much emphasis can be laid on the thematic content analysis at primary levels as the general assumption is that the themes/topics in these textbooks will be very common in the day-to-day life and the children will be reasonably familiar with the themes dealt with. Even in the linguistic content analysis, not much emphasis is on the structure words and content words become equally important from post-primary levels. At the primary stage, particularly at the second or third standard, it is not an exaggeration to state that the content words form the communication as such and the structure words are not that crucial at this stage. Another reason why adequate weightage is not given to the structure words at the primary level is that the children though familiar with the structure words, perhaps are not reasonably familiar with their usage in expression, At the same time, their range of vocabulary in terms of content words is much less and hence of greater importance.

6.7 Frequency Count

With a view to study the pattern of language use, it would perhaps be enough to pick up 20% of the text material to study the

frequency of occurrence of the content words made use of and the sentence patterns in the instructional material. This is essential to see whether the vocabulary used in the textbooks is within the range of vocabulary control of the learners, the eye-span and whether the complexity of sentence patterns is commensurate to the level of the target learners. Therefore, it is appropriate to select the textual material for this purpose at random.

6.8 The Sample

A high risk is involved in making any generalization or powerful statements particularly with reference to instructional material. This necessitates that a large sample has to be taken to make realistic value judgments.

Within the constraints and limitations, with a view to making comprehensibility studies as reliable as possible, various kinds of tools can be adopted. They are:

- Language (cloze test, reading comprehension test and vocabulary test)
- Natural and Physical Sciences; Social Studies etc. (cloze test, reading comprehension test and vocabulary test)
- Mathematics (vocabulary test)

It may not be out of place to discuss in brief the above kinds of tests and more in particular about the cloze test.

6.8.1 Cloze Test

The coinage of the term "cloze" can be traced back to the year 1953 when it was first used by Taylor (1953) when he tested the first language. The derivation of this expression is apparently from 'closure' of Gestalt psychology. It is based on the nature of human tendency to close or complete a pattern once they are familiar with the structure of that pattern. In course of time, the cloze procedure began to be applied to the second and the other language situations also besides non-language subjects. Basically, the cloze test involves a continuous running text, from which certain words are deleted, which

are expected to be filled – in by the examinees with the appropriate vocabulary items. Such deletion of vocabulary from the running text has been visualized in two ways: (i) Fixed deletion and (ii) Random deletion.

The Fixed deletion method was apparently visualized with a view to assess the ability of the examinees with reference to the specified categories of the vocabulary items like the content words or the structure words or even more specifically, nouns, verbs, adjectives and the like. The Random deletion was visualized with a view to assess the overall expressive abilities of the examinees. Random deletion could be either every third, fifth, seventh or ninth word in the running passage which should be decided upon considering the sentence length and complexity and the level at which the tests are contemplated. Experts in the field of educational testing have reasonably and logically argued that the first and last sentences of the text in question should be retained intact without deletion of words as they, in the normal course, provide a brief and general idea about the nature and the content of the text in question.

There have been differing opinions about the appropriateness of cloze test for assessing comprehensibility. It is a well-known fact that several comprehensibility formulae were discovered in the West particularly from 1948 onwards, for e.g., Dale Chall (1948), Flesch (1948), Fog (Gunning, 1952), Smog (Mc Laughlin,1959), Fry (1977) and so on. Most of the formulae involve different parameters as word frequency, word length, sentence length, etc. They have apparently not considered crucial aspects like the sentence complexity and the idea-density, overall text organization etc. This lacuna apparently makes people doubt about the realistic applicability of the formulae to assess the comprehensibility level of textual material. These controversies still remain. Incidentally almost all these formulae were worked out with reference to a non-Indian language mostly English and in addition, with reference to Western situation. The Indian scene is quite different. At least four major language families exist in India and each language family has its own structural framework. This structural framework unique to each language

family apparently poses a problem even in contemplating at one single formula for all Indian languages. Therefore it would perhaps be inappropriate and unrealistic to think of one single comprehensibility formula to be applied in the Indian context.

A cloze procedure in the context of measuring comprehensibility is not aimed at assessing the skills of individual examinees but rather meant to work out the average score of a sizable number of examinees which is used as an estimate or approximation of the text itself.

This cloze procedure is not merely made use of to assess overall linguistic competence but also used of for assessing narrower aspects like reading comprehension, text and subject comprehensibility, readability etc. While the overall average performance of the learners can be construed as the comprehensibility of text materials, the assessment of reading comprehension relates to the individual examinees. However, as an extension of reading comprehension, the average performance of the examinees in the reading comprehension tests can also be taken to supplement the findings of the cloze procedure intended for the overall assessment.

In addition to the assessment of the comprehensibility and overall language proficiency, cloze tests are used for several other purposes like educational, social, evaluative and pedagogical purposes. Interestingly, this procedure has also been used in the West as instructional technique. (John 1977, Sarin 1977). As pointed out earlier, cloze test has been considered as suitable for measuring various kinds of knowledge. Kohen (1980) argues that the cloze tests measure three types of knowledge viz., linguistic knowledge, textual knowledge and even the knowledge of the world.

The general presumption with reference to any instructional material is that it is a systematically graded material in terms of linguistic difficulty and thematic complexity. In other words, any good instructional material is expected to be graded in terms of the 'increasing difficulty' from the point of view of language structures as well as the content. Random deletion method is adopted if the aim

is to focus primarily on the content words and secondarily on the structure words.

Keeping in mind the level of the examinees, it would be reasonable and desirable to include an added dimension to the cloze test procedure in providing a mixed-up group of words against the passage from which the examinees are expected to pick up the words relevant to the deletion and complete them. While some prefer to provide these cues after the passage at the bottom, some others prefer to have them against the passage on the right hand side, the presumption being that it would be much easier for the students to quickly locate the right cue if the texts and the cues are presented horizontally rather than at the bottom of the passage.

Evaluation of language instructional material in MT/NL has to be viewed differently from L_2 and OT/L_f. The reason is that in the case of L_2 and OL/L_f instruction, the possibility of interference from the native language and other language backgrounds, if any, play a very significant role. As a result of such interference, various kinds of problems and difficulties in the learning of the target language could arise, which have to be very carefully taken into consideration in the development of the instructional packages for second and other/foreign language instruction. These aspects also need to be kept in view in the process of evaluating instructional material pertaining to L_2 and OT/L_f. This becomes easier and more purposeful if the target audience belongs to the same linguistic community and possesses comparable if not identical language backgrounds. However, if the target audience happens to be a heterogeneous combination, it will not be easier to take care of these aspects either in preparing the instructional material or in the process of evaluating such material.

6.8.2 Reading Comprehension Test

Reading comprehension tests are aimed at assessing the ability of the individuals in understanding and comprehending the given textual material. Such comprehension tests could also provide a clue or an idea about the conceptual abilities of the learners.

Therefore it is considered apt and necessary to adopt reading comprehension tests in addition to the cloze tests. The comprehension questions could be of various types: (i) Multiple-Choice, (ii) True or False/Yes or No, (iii) Matching, and (iv) Short answer and other open-ended types. Selection and choice of various item types depends upon the level of instruction and expressing learners' abilities of knowledge, content and skills.

6.8.3 Vocabulary Tests

The vocabulary tests primarily include selected content words in case of non-language subjects, the purpose being to assess the conceptual ability of the learners with reference to the specific disciplines like science, social studies and mathematics. On the other hand, they primarily include structure words besides some content words in case of language tests.

In addition to the three types of tests, some scholars have recommended another kind of vocabulary test involving dictation of about 10 words to the examinees asking them to write and consequently to read as written by them. It is apparently to see and assess to what extent they could take dictation without distorting the spelling system and also to see whether they could read as they had written (which might contain spelling mistakes) only taking into consideration the original stimulus i.e., as presented by the teacher during dictation. It is not very clear as to what extent this exercise would be useful and meaningful in the context of comprehensibility. It could, however, be adopted at the primary level.

While the above discussion basically lays more emphasis at the initial level of instruction, certain alterations and changes will be necessitated in the contents of the tools as the level of instruction varies. For instance, a dictation test is not necessary at higher levels of instruction whereas it is essential at the initial levels of instruction. Similarly, the standards of the texts chosen for purposes of comprehension tests and cloze tests will have to be commensurate with the level of instruction. This aspect however, is automatically

taken care of in choosing the text from the instructional material under evaluation.

6.9 Evaluation of Instructional Material in Subjects other than Language

As stated earlier, there is a clear cut and distinct difference in the processes involved in developing instructional materials for language instruction as opposed to the non-language subjects. The basic principles, however, are generally applicable to all instructional material irrespective of whether they are related to language instruction or subject instruction. With reference to the instructional material for subject teaching, it is to be taken note of that the content is more significant than the language/expression. It does not, however, mean that language or expression is totally ignored in this context. It is a well-established fact that content and language go together hand in hand and one has to be integrated with the other. The significant point that needs to be taken care of is that the ease or complexity of language used in subject instructional material has to be related to the language competence of the learners or their expected control over language with reference to language structures in particular and vocabulary in general. Needless to mention that vocabulary is very significant in the context of subject instructional material in that that the vocabulary specifically made use of has also to be built up among the learners. This makes it all the more necessary to gradually develop the discipline – specific vocabulary which has a direct bearing over the content. The principles of material preparation viz., selection, gradation, presentation and repetition are also important in the context of subject instructional material. The topics/themes related to subjects have to be carefully selected keeping in view the level of instruction and they have to be arranged and graded in a manner that there is a gradual continuum and uninterrupted development of the subject. As far as the presentational aspect is concerned, it is to be ensured that the theoretical and practical aspects of the subject are so integrated in the instructional material that the interest and motivation of the learners is maintained and such integration would help the learners in better conceptualization of the subject and the

theme. Thus, it may be noticed that the general principles of material preparation are applicable in the case of subject textbook as well. Therefore, in the process of evaluation of instructional material relating to subject specifics, it has to be more intensely examined from the point of view of content, giving due importance to the expression aspect. It is from this point of view that the suggested tool provided in the Annexure needs to be modified, amended, additions and deletions made depending upon the kind of instructional material evaluated.

Another significant point that has to be taken note of is that even in the case of language instructional material evaluation, the tool provided needs changes/modifications, additions and deletions depending upon whether the material evaluated pertains to the first language/mother tongue instruction, second language instruction or other language/foreign language instruction. Exercises provided in the lessons in the subject textbooks have to be examined from the point of view of coverage of content rather than the linguistic skills. Similarly, in the assessment of a learner's responses, significance is given to the content/theme rather than expression. This, however, will be dealt with in a little more in detail at a later stage.

6.10 Evaluation of Instructional Media

Instructional media could be several in nature. The basic and primary medium is the classroom wherein the teacher and the taught come directly in contact with each other and a direct face-to-face interaction takes place between the two. This facilitates solving the problems/difficulties of the learners instantly on the spot during the course of instruction. The other media include various types of audio-visual equipments including the latest innovation viz., computer. Looking back at the much older system of education when the technological gadgets and aids for instructional purposes were not available, even then an auxiliary and supplementary medium of laboratory and workshop instruction was also adopted for educational purposes particularly in the context of teaching natural and physical science subjects and to some extent, certain subjects classified as

humanities. While the theoretical aspects are discussed and taught in the classroom, the practical aspects are taken care of in the laboratories and workshops through actual demonstration of the scientific and technical processes. This could also be viewed as one kind of medium. With the advancement in the technological innovations for educational purposes, the audio-visual media also came into use actively in the field of education. These media are used for supplementing the classroom instruction except in the case of distance education/correspondence courses. Even in such cases, the contact programs become necessary to make up the gap arising from the non-availability of face-to-face direct interaction between the learners and the teacher. However, generally speaking, these media could be viewed as supplementary to the direct process of instruction. As the activities during course of instruction of any subject vary depending upon the medium adopted, the nature and kind of instructional material also differs based upon the kind and nature of medium chosen. Therefore, the evaluation of media of instruction has to be focused on the kind and nature of medium, the purpose, usage and the suitability and appropriateness of the material made use of including therein the various kinds of activities that would facilitate the learners to actively take part in the process of instruction. If it is a question of choice of one medium or the other, then the curriculum and syllabus remains the same, but the design of material differs. On the other hand, if it is a question of using various media as supplement to the direct process of instruction, the curriculum and syllabus differ and accordingly, the nature and kind of instructional material both from the point of view of delivering the subject and the activities of the learners.

The choice of medium of instruction depends upon, to a considerable extent, on the readiness of the learners to receive instruction through the medium chosen, the learners' mental maturity and exposure to various kinds of media, the age level and nature of instruction and the subject concerned. For instance, use of audio-visual media may not he as much desirable in the early stages of education as it could be in the slightly higher levels of instruction.

Efforts are, however, being made to make use of various media even at the initial levels of instruction. It is from the point of view of these aspects that efficacy and utility of the medium adopted for instruction has to be evaluated.

As in some other cases, the basic and primary tool for evaluating the media of instruction should be a checklist/rating scale. A word of caution in this context has to be mentioned in that that the expression "medium/media" in this context does not refer to the language of instruction but it refers to the process of instruction.

6.11 Evaluation of Instructional Methods

It is not infrequent in the field of education in India to talk about various methods of instruction labeled as science methods, social studies methods, mathematics methods, language methods etc. The expression "methods" in this context is apparently made use of with reference to the planning of instruction and the procedures adopted for instructional purposes. To be clearer, this has a bearing on the preparation of lesson plans indicating therein the aims and objectives of the day's lesson, a brief description of the content and procedures involved and the steps in which the instruction has to proceed with reference to the lesson chosen. But for this, the term "methods" does not make much sense when it is related with the subjects. It is, however, to be expected and acknowledged that each instructor will have his/her own techniques and strategies of instruction. "Methods" makes a lot of sense with reference to language instruction. Several language instructional methods are talked about out of which about 10 to 12 methods have become popular and are adopted depending upon the purpose of language instruction. It again has to be made clear that these "methods" are more significant in the context of L_2 and OT/L_f instruction. Therefore, to talk about methods of instruction, it is desirable to restrict it to the field of language education.

The popular language instructional methods include direct method, bilingual method, audio-lingual method, grammar method,

grammar-translation method, reading method etc. While the choice of one method or the other depends primarily on the purpose of instruction, it is again to be made very clear that in no one situation can anybody claim that a particular method alone is adopted for instruction. The reason is that the features of other instructional methods in some degree or the other creep in. However, each method has its own merits and demerits and limitations and each method is conceived of from the point of view of the purposes of instruction. For instance, "direct method" is adopted when the learners' community is small, the purpose of instruction is primarily to develop competence among the learners with reference to oral skills and the time available for instruction is relatively short. Besides, the chances of adopting other media like audio-visual etc. are less. Similarly, bilingual method is adopted if the target language happens to be a language other than the learner's native language or when the learners do not possess any control or knowledge of the target language. Similarly, the audio-lingual method is adopted for developing competence with reference to the oral skills. Reading method is adopted only for developing the reading skills, while grammar method and some other methods are adopted to develop both oral and written skills. It has to be made adequately clear that these methods figure in if the level of language instruction is initial or middle level. At higher levels of instruction, the method gets unified and assimilated into the adoption of target language itself, however, leaving the scope for the adoption of other languages to certain extent. These are the major aspects and considerations from which the evaluation of instructional methods in the context of language has to be considered.

In the case of methods with reference to non-language subjects, the evaluation has to be focused on the proper planning of lessons, making use of the equipment necessary for demonstration purposes, the teacher-learner activities and the like. Added to this, the manner in which the classroom instruction is supplemented and integrated with the activities in the workshops and laboratories has also to be

examined. In the case of methods evaluation also, the tools recommended for adoption include checklists, rating scales and observation schedules.

6.12 Instructor Evaluation

As mentioned in the earlier pages, each instructor possesses his own expertise, strategies and techniques in executing his job of teaching. Depending upon the efficiency and efficacy of instruction, different teachers are categorized as very good, good, average, poor etc. This primarily depends upon the manner of delivering the subject, the methods of presentation, the ways of clarifying the significant points in the subject delivered and the way in which opportunities are provided for periodical interaction of the learners with the instructors. It further depends upon the coherence, precision and cogency relating to the arrangement of ideas relating to the subject delivered, different viewpoints presented and relevant conclusions drawn besides the fluency of expression of the teachers. It has to be ensured that style, variety and standard of expression adopted by the instructor is in consonance with the learners' level. The instructors' efficiency also depends upon how he draws the attention of the learners, the manner in which he ensures the interest and motivation among the learners, active participation of the learners, non-discrimination among them and the techniques and strategies adopted for keeping the learners alert and attentive apart from establishing a good friendly rapport with the learners. Further, it also depends upon how the instructor provides ample opportunities for the learners to react, discuss and clarify different aspects of the subject taught without making the process a monotonous one-way communication. In other words, the process of instruction should not be either dictatorial or one sided. It is absolutely essential for the instructor to keep an open mind to accept and examine the viewpoints of the learners also relating to the subject taught. Once, the instruction is activity-based, efficiency automatically gets ensured. These aspects are equally important whether the instruction relates to

the language or non-language subjects like mathematics, natural and physical sciences, social sciences etc. These are the important aspects from which the instructor evaluation has to take place. In this case also, the tools normally adopted include observation schedules, checklists and rating scales.

A suggested tool for instructor evaluation in the context of language is presented hereunder:

- Does the teacher successfully create the situation with introductory remarks etc., before presenting the teaching point/item/concept?
- Does the teacher explain the difficult vocabulary items?
- Does the teacher reinforce the previous day's lesson before going on to a new item?
- Does the teacher explain the grammatical points involved clearly after introducing the lesson?
- Does the teacher ensure that he repeats the items afresh for each individual learner?
- Does he allow any of the learners to follow the other students (in drills)?
- Does he repeat the item again after each learner's repetition?
- Does he ever slow down from the normal speed and does he give unnatural pauses?
- Is the teacher partial to any of the learners?
- Does he neglect or ignore the better learners and concentrate more on the poorer ones?
- Does he clearly explain what he expects from the learner?
- Does he confine his expression and delivery to the vocabulary range of the learners?
- Does he introduce a lot of vocabulary items at once?
- Does the teacher hold the attention of the entire class?

a) Does he change the order of calling on the students for their responses?

b) Does he name the student after putting the question?

c) Does he answer the question himself if any learner fails to answer?

d) Does he repeat the same question?

e) Does he permit the learners to deviate from the topic?

f) Does he repeat the question if the learner has any difficulty?

- Does the teacher ensure that the students keep their books closed when the basic sentences are introduced by him?
- Does he make sure that no student repeats the same item more than thrice?
- Does he come back to the weaken students again?
- Does the teacher give an impression of ridiculing or discouraging any learner at any time?
- Does the teacher use drills which require the learners to use their vocabulary that are not in the material, but are within the range of the structures taught?
- Does he give adequate exercises as home assignments?
- Does he use appropriate visuals and charts (wherever possible) to explain the concept clearly?
- Does he use the blackboard to the extent necessary?
- Does he ensure the degree of learners' participation in the classroom?
- Is his emphasis in teaching skill-oriented rather than content-oriented?
- Is the teacher qualified to teach language?

In the case of evaluation of subject instructors, certain changes, additions and modifications in the tool suggested above have to be

carried out. Aspects like the adequate use of laboratory/workshop, integration of classroom instruction with the activities in the lab etc., opportunities provided for the learners to handle equipment and conducting experiments, writing up laboratory reports pertaining to experiments in the prescribed format and procedures, the efficiency and guidance provided for the learners in carrying out the practical experiment etc., have to be kept in view in evaluating the subject instructor.

CHAPTER

7

Evaluating Communication in Education

Communication plays a very significant role in any sphere of life and more so in the field of education. Irrespective of whether the matter under consideration relates to languages or the non-language subjects, it is essential to ensure an appropriate, precise and clear communication both from the instructors and the learners. Looking at it from the point of view of learner's performance, their expression in formal and informal situations has to be precise, definitive and clear and unambiguous. Further, the responses of the learners have to be to the point and should not be evasive or away from the main point. In the event of open-ended questions or paragraph questions or essays and the like, the arrangement of ideas is a very significant and important aspect. For a smooth reading, it becomes necessary that the ideas relating to the theme have to be logically and sequentially arranged so that there is a gradual development of the theme. Further, it is also necessary in the context of written responses that each main idea is dealt with in a separate paragraph wherein the main idea has to be followed by the related sub-ideas and the supporting arguments thereof. In addition, each paragraph must contain a brief conclusion regarding the idea dealt with in the paragraph. These brief conclusions at the end of each paragraph will be in addition to a comprehensive overall conclusion

provided at the end of the answers. It has also to be ensured that the answers begin with an introductory paragraph wherein a brief introduction and background related to the theme is presented.

As already stated time and again, language and content go hand in hand and it is a practical impossibility to conceive of them as watertight components. However, while the skills play a major role in the context of language, content plays a significant role in the context of non-language subjects. Expression, however, is important in both the cases. In the case of language, both grammatical and communicative competence are equally significant, while communicative competence is more significant and grammatical competence is relatively less significant in the case of non-language subjects. At least up to the end of the school level, the language aspect or the grammatical ability need not be given much importance in the case of content/subject-related aspects as in the context of language education. It could thus be concluded that while ensuring the effectiveness of communication, perfect linguistic competence need not be made rigid and compulsory in the context of non-language subject instruction. Such relaxation and flexibility do not find a place in the context of language instruction.

Evaluating communication, whether it is in the context of language or non-language subjects, becomes subjective in nature as it involves the rating and assessment of open-ended responses consisting of a small paragraph or a number of paragraphs in an essay or discourse. In order to minimize the subjectivity of assessment, it is desirable to have a multiple assessment of the same individual's responses by two or more examiners and consider the average of all as the final assessment. Even the individual examiner's assessment could be made a little objective in providing a number of criteria for assessing the open-ended responses. Such criteria could include the relevance and appropriateness of the ideas dealt with including the sub-ideas and supporting arguments, the level of discourse, the appropriateness of styles, vocabulary and sentence patterns. Further, the assessment could be related to illustrations and logical explanations contained in the responses.

As far as the linguistic proficiency is concerned, five levels of proficiency are talked about. They are:

1. Elementary proficiency
2. Limited working proficiency
3. Minimum professional proficiency
4. Complete professional proficiency
5. Native or bilingual proficiency

Elementary proficiency refers to the ability of the learners in providing small responses within the scope of his limited knowledge of the subject and skill. Such elementary proficiency is highly prone to contain various kinds of skill-related or content-related errors.

Limited working proficiency refers to the general ability of the learners to make use of linguistic skills confidently in general day-to-day life situations including the current events etc., which are non-technical in nature. The learners do not, however, possess the thorough control of the grammatical aspects of the language concerned.

Minimum professional proficiency refers to the learners' responses being fairly accurate from the points of view of structures, vocabulary and their relevance to the content. The degree of learners' comprehension will be reasonable enough and they could conduct themselves in a relatively effective manner in the midst of the native speakers of the target language.

Full professional proficiency relates to the learner's ability in fluently expressing himself with high precision of vocabulary. Occurrence of grammatical errors is a remote possibility. This also involves the learner's ability of interpretation of the subject concerned.

Native or bilingual proficiency in the context of language refers to the learner's ability being on par with that of the native speakers of the language. In such cases, the expression is accepted and appreciated from the points of view of choice of vocabulary, idiom, colloquialism and the cultural dimensions. This, however, does not

make any specific reference to the proficiency from the point of view of content. The implication is that the expression includes the content-related accuracy also.

Thus, the degrees of proficiency mentioned above could be observed in any field of instruction. However, the native-like or bilingual proficiency could be thought of primarily in the context of language education.

It would perhaps be relevant and reasonable to make a distinction between effective communication and communicability. The concept of communicability limits the scope of expression from the point of view of minimum adequacy of expression. In other words, the ideas or themes dealt with are presented in such a manner that they are not distorted or lost in the process of communication. Such expression may not, however, be 100% accurate from the point of view of grammaticality nor it may be appropriate from the point of view of the styles and registers. On the other hand, effective communication has to ensure the expression of 100% accuracy of content ensuring therewith the 100% relevance and appropriateness of styles, registers, vocabulary, sentence patterns and the like.

CHAPTER

8

Evaluating Learners' Performance

Evaluation and testing are commonly conceived of as applicable to the learners in any instructional program; without realizing the need and importance of evaluating several other aspects related to education which have been amply discussed in the earlier section of this book.

In addition to the misconception regarding evaluation and testing and its applicability to the learners' performance, another very serious misconception is that the learners' performance merely based on the outcome of tests and examinations is sufficient in itself for drawing inferences, value judgments etc., and for deciding upon the placement of learners. This has been a very powerful drawback in the examination system so far. The fact that mere tests and examinations are inadequate for a justified, realistic and reliable assessment of the learners' performance has been advocated even in the early sixties by scholars like Ebel[6] etc. However, the seriousness of this lacuna has not been properly realized and no remedial measures have been thought of and implemented to make the evaluation of learners' performance more dependable, realistic and reliable. To some extent, however, certain changes and modifications have taken place in the West in the context of evaluating learners' performance. Although it is claimed

that even in the Indian context some efforts have been made, they have not been effectively implemented.

Instruction and testing/measurement go hand in hand all through the instructional process. Such on-going testing/ measurement process during the course of instruction, to a very great extent, is informal based on which the instructors do try to make use of the feedback so obtained from time to time to improvise their process of instruction. While this informal ongoing process of testing, measurement/evaluation is made use of for the betterment of the instructional process, it has not been duly considered in the context of deciding upon the overall performance of the learners. Another factor that needs to be mentioned is that tests and examinations are a very common feature in the course of any instructional program: Again, the periodic tests conducted during the course of instruction are primarily used for diagnostic purposes and in trying to locate and find out the problems and difficulties of the learners and devising and developing remedial instructional material. Such periodic assessment has at best served as a word of caution and device to the poor performers that they have to improve upon their performance in future days to come. Instances are not infrequent wherein subject remedial instructional programs are held through tutorials, counselling etc., in certain contexts. However, generally speaking, the final tests/examinations alone have been considered in deciding upon the learners' performance at the terminal stage of instruction.

Educational psychologists have also argued time and again that it is highly improper to decide upon the fate of the learners merely based on the terminal examinations. The reasons put forward by them maintained that the psychological and mental state of the learners will not be consistent all through and that most learners are prone to become tense and nervous in a formal testing situation, that too, when they realize that such situations lead to deciding upon their final outcomes. The psychologists' views have also made clear that even during the course of instruction over an extended period of

time, the psychological bent of minds of the learners are prone to fluctuations and changes from time to time depending upon factors like their interest and motivation in the subject, time and duration of instruction, the environmental conditions under which the instruction takes place and several other personal factors which are idiosyncratic to each individual. The degree of readiness of the learners also contributes a lot to their performance in the testing situations. These views lead us to conclude that making outcomes of the terminal examinations the only basis for finally deciding the fate of the learners in any instructional program is inadequate, unsound, unreliable and unrealistic. This being the case, it becomes necessary for the decision makers in the field of education to suitably amend and reform the processes involved for finally deciding upon the degree and quality of the learners' performance and in drawing inferences and value judgments thereafter. It is in this context that the performance of the learners all through the instructional program from their entry point till the terminal point has to be taken into consideration. This fact opens up gates for considering a number of aspects relating to the activities of the learners in the classroom from the beginning to the end. It is not sufficient if the outcomes of the periodical tests are taken into consideration besides the performance in the terminal tests and examinations. Scholars like Genese and Upshur[7] and Johnson [8] have argued in extension of the views of Ebel etc. maintaining that the performance evaluation has to consist of evaluating with tests and without tests and a unified combination thereof could only form a realistic and dependable base for a meaningful and realistic performance evaluation. Ebel has talked extensively about relating to the classroom tests and all factors related to the test construction including the item preparation, estimating and interpreting and improving test reliability, improvising of test quality through item analysis, marks and marking systems etc. Thus, the stand taken by Ebel definitely focused on the inadequacy of tests and examinations for testing the final performance levels, but does not throw light on aspects which need to be taken care of for a more

realistic, valid and reliable overall performance evaluation. These missing factors have been very well realized and spelled out by Genese and Upshur and Johnson. Genese and Johnson have broadly classified the process of learners evaluation into two categories viz., (1) Evaluating without Tests and (2) Evaluating with Tests. Evaluating without tests includes atleast three aspects viz., (1) Classroom Observation (2) Portfolios and Conferences (3) Journals, Questionnaires and Interviews. Evaluating with tests includes various types of teacher-made and standardized tests and informal tests. Johnson has argued that in order to examine the performance and communication, one has to consider aspects like teacher-student interaction, community and school-based issues in communication and the like. Keeping these factors in view, it becomes necessary to conclude that all the factors and aspects mentioned by Genese, Upshur and Johnson and indirectly by Ebel earlier will have to be duly considered and appropriate weightages given to them in consolidating the overall performance of the learners and final conclusions and recommendations drawn.

According to Genese and Upshur, "Evaluation involves many different kinds of decisions: decisions about the placement of individual students in particular streams, levels, or courses of instruction; about ongoing instruction; about planning new units of instruction and revising units that have been used before; about textbooks or other materials; about students' homework; about instructional objectives and plans; and about many other aspects of teaching and learning"[9]. Keeping these innumerable purposes of evaluation in view and considering the arguments of different scholars presented above, it would be appropriate to consider all of them and plan a sound framework for the evaluation of learners' performance. As far as the classroom-based evaluation is concerned, Genese and Upshur have designed a very meaningful and sound framework presented below.

Input Factors

— Student needs and abilities	Instructional purposes	
— Time		
— Attitude		
— Resources	Instructional plans	
— Facilities		
— Support		
— Teacher abilities	Instructional practices	Outcomes

The above framework is equally applicable to any branch of education, be it language or other subjects like physical and natural sciences, mathematics, social sciences, humanities or any other subjects. The two processes of evaluation viz., evaluating without tests and evaluating with tests, if related to the above framework, would be meaningful to conceive of a sound, valid and reliable mechanism of performance evaluation. Genese and Upshur again have come out with a strategy for classroom-based evaluation which is diagrammatically presented hereunder:

Input Factors

— Student needs and abilities	Instructional purposes	
— Time		
— Attitude		
— Resources	Instructional plans	
— Facilities		
— Support		
— Teacher abilities	Instructional practices	Outcome

It is now apt, appropriate and relevant to discuss the two kinds of evaluation i.e., evaluating without tests and evaluating with tests hereunder.

8.1 Evaluating Without Tests

As stated earlier, evaluating without tests has three major criteria:

i) Observation in Classrooms,

ii) Portfolios and Conferences, and

iii) Journals, Questionnaires and Interviews.

8.1.1 Observation in Classrooms

Observation in Classrooms has to be ensured that it is continuing and ongoing from the very entry point to the terminal stages of instruction. Several scholars have worked on the formal methods of classroom observation, particularly with reference to second language education. Researchers have designed a formal procedure for classroom observation and devised a format under the Communicative Orientation of Language Teaching (COLT) observation scheme. The format may be seen in Annexure 2.

This format helps in real time observation in an instructional program by trained observers and would facilitate compiling detailed and complex information about the subject by the teachers and evaluators that is difficult for the participants themselves to elicit and understand. The classroom observation involves the evaluators/ teachers in carefully observing as to how the learners respond and how they make use of the instructional material, how they interact during the group work and how the interaction between the teachers and the taught takes place. It further provides necessary and relevant information for the instructors to clearly isolate the areas of difficulty for the learners and such areas which are not received by the learners. Thus, it thereby helps the instructors in designing proper remedial measures and materials for more effective learning. The classroom observations further assess the effectiveness of particular teaching strategies, identify the instructional activities and materials that the students enjoy and receive better.

Apart from the usefulness of classroom observation from the point of view of the overall assessment and evaluation of learners'

performance, such observation has relevance to drawing of inferences about teaching and learning for planning an instructional program. The regular and continuous classroom observation from time to time could help the teachers and material producers in deciding upon whether to retain the portions of material covered or to drop or modify for a better instructional process in future.

In order to make the optimal use of classroom observation, it is desirable to provide a manageable, systematic and effective strategy for organizing classroom observation. In this context, Genese and Upshur warned that "without a coherent strategy, teachers observance and inferences that they derive from them run the risk of being fragmented and disorganized and consequently ineffective"[10]. They argue that the classroom observation should be concerned with the instructional objectives. These objectives will have to be related to the activities that take place in the classroom with reference to both the teachers and the taught and necessary observations recorded. Such relation of the classroom observation with the instructional objectives will facilitate in making it manageable and systematic as it "defines and delimits what is to be observed". Following are the reasons for planning the process of classroom observation:

- To decide upon whether the learners are progressing in the instructional program in accordance with the syllabus and whether such progression is cumulative in nature.
- To identify the learners' difficulties and problems in receiving instruction which will help in improvising upon the methods and materials for better instructional purposes in future and for devising the remedial techniques and strategies for the present.
- To assess the degree and quality of achievements by the learners which will facilitate identifying specific future needs and plan appropriate instruction.
- To assess whether the learners' interest and motivation are sustained in receiving instruction and are useful from the point of view of the learners' needs, and their applicability in the real life situations.

There are several other reasons for planning a good strategy for classroom observation. As far as the regular classroom observation is concerned, it could be made more meaningful and reliable if it is related to the teacher-learner interaction, group interaction among the learners, and learners' interaction with outsiders in different contexts.

As appropriate tools are necessary for any systematic procedure, such tools become relevant and necessary in the context of classroom observation as well. In fact, a good and ideal record keeping relating to classroom observation helps the teachers in the following manner:

- It helps in keeping track of important information about student learning and effectiveness of information.
- It helps in forming sound impressions and learner achievement and progress.
- It helps in accurately identifying persistent difficulties experienced by the learners.
- It helps in reporting learner's performance and progress to other educational professionals, parents, evaluators etc.
- It helps in assigning grades and ranks as and when required.
- It helps in monitoring, evaluating and/or designing instructional plan.

Meaningful and effective evaluation requires the combination of several record-keeping techniques: some for daily recording and some for periodic recording; some focusing on the students and others on the instruction; and some that are general and some that are specific. Although several tools and instruments could be thought of, primarily following three tools are recommended to be actively made use of for classroom observation: (1) Anecdotal Records, (2) Checklists, and (3) Rating Scales.

8.1.1.1 Anecdotal Records

The anecdotal records provide day-to-day information and describe briefly the context in which the observation is made. They

contain the comments and recommendations of the observers/ teachers and the information so stored can be retrieved at any point of time. They could be structured or open-ended although they are particularly suited for open-ended recording of information. Such open-ended records are especially useful for recording observations and inferences that are unique and unexpected and that might be difficult to record using other methods. They are primarily used for internal consumption only. Normally, anecdotal records are adopted by the beginning level teaching or teachers with less experience. Once they gain sufficient experience in teaching, they could resort to other two methods of recording like checklists and rating scales.

8.1.1.2 Checklists and Rating Scales

Checklists and Rating Scales are similar in form and use. They consist of several items each describing a specific aspect required to be observed. While the anecdotal records are not pre-designed or pre-structured, checklists and rating scales are preplanned, and pre-structured clearly specifying the details of the aspects required to be observed. The categories and criteria for observation and assessment have to be well defined in precise, clear and unambiguous terms. Each item in the checklists and rating scales has to be limited to one and only one aspect and there should not be any overlap of any other aspect. The only difference between the checklist and the rating scale is that while the checklist provides the scope for observation like Yes/ No, the rating scales provide an additional facility to identify the degree of presence or absence of the aspect observed. The five or seven-point scale is normally adopted in such rating scales.

Samples of suggested checklists are presented hereunder. They are only suggestive in nature and require necessary changes, modifications, additions and deletions, depending upon what is assessed and the purpose for which the assessment is made.

The checklist for assessing language skills suggested by Genese and Upshur is given below:

It is primarily meant for assessing receptive and productive oral skills.

Classroom-based Evaluation in Second Language Education

The following checklists are designed to assess students' receptive and productive oral language skills. Assess only those items that are pertinent to your instructional objectives or are otherwise relevant. You can include additional aspects of oral language.

If you want more precise assessments, you could respond to each item using a rating scale based on criteria that are suitable for your purposes (for example, unsatisfactory, satisfactory, excellent).

Receptive Oral Skills	*Productive Oral Skills*
1. Understands simple directions	1. Pronounces vowel sounds correctly
2. Understands simple sentences	2. Pronounces consonant sounds well
3. Understands simple *Yes/No* questions	3. Pronounces blends correctly
4. Understands plurals appropriate to age	4. Uses word stress correctly
5. Understands vocabulary	5. Uses phrase/sentence stress well
6. Understands adjectives appropriate to age	6. Uses tone correctly
7. Understands several related sentences	7. Uses intonation correctly
8. Understands contractions, common shortened forms, and so forth	8. Gives one-word responses
9. Understands tense indicators	9. Produces simple sentences
10. Distinguishes tones and understands their meaning	10. Produces simple questions
11. Understands different intonation patterns	11. Gives simple directions
	12. Uses tense markers correctly
	13. Uses prepositions correctly
	14. Forms complex sentences
	15. Uses several continuous sentences
	16. Gives descriptions

12. Understands more complex directions	17. Uses vocabulary appropriate to age
13. Understands rapid speech	18. Speaks fluently
14. Understands language in Classroom situation	19. Uses classroom language easily
15. Understands language of peers	

The above checklists/rating scales are relevant in the context of language instruction. In the case of subject instruction, the above formats require a careful revision and a number of changes.

In the context of preparing checklists and rating scales, it is desirable to observe certain aspects which would facilitate working out more meaningful and useful formats. Following are the aspects that need to be considered in this context:

- Depending upon the purpose and the categories of observance intended to be made, the observers, needs or aspects of teaching and learning have to be clearly identified.
- In the process of working out the items to be included in the checklists/rating scales, such items must be included which reflect clearly observable aspects related to the instructional aims and objectives.
- To the extent possible, the categories suggested for inclusion in the checklists/rating scales should not overlap with each other and consequently redundancy should be avoided. In the event of overlapping or the items being redundant, the objectivity of check-lists/rating scales gets diluted thereby the reliability and validity go down in degree.
- It is desirable to prepare a brief description of each item included and such descriptions have to be clear, concise, unambiguous and meaningful.

- In case of items meant for using rating scales relevant and necessary descriptions for the response choices have to be worked out.
- Before finalizing the content and formats of checklists/rating scales, a pre-testing is desirable based on which necessary modifications and changes have to be effected in the formats and contents so as to ensure higher degree of reliability, validity, coverage etc.

8.1.2 Portfolios and Conferences

The second aspect of evaluating without tests is the use of portfolios and conferences. A portfolio is a meaningful and purposeful collection of each learner's work which demonstrates their activities, efforts, progress and achievement in the area of learning. The portfolios in the context of education include compositions, précis writing, creative writing and innovative efforts of the learners. They are particularly in the context of learners' achievement as they consist of regular time-to-time progress and achievements of the learners throughout the duration of learning. Making such portfolios as one of the bases for an overall evaluation of learners' performance is highly desirable as they are the ones which reflect the learners' performance throughout the learning process and without giving them a feel of formal evaluation. Formal testing situations, according to educational psychologists, might make the learners nervous and tense which results in their poor performance. Therefore, using portfolios as a basis for evaluation of learners' performance is very sound and objective. In addition to this, such portfolios, on the one hand, help the instructors in devising remedial measures and on the other hand, they help the learners realize and identify their problems and difficulties in the process of learning and motivate them for better learning in future. Portfolios are basically learner-oriented, collaborative and scholastic. The advantages and benefits of such portfolios could be from the points of view of what they provide and what comes out. Such advantages and benefits are given hereunder:

Benefits of Portfolios

Portfolios Provide

* A continuous, cumulative record of language development
* A holistic view of student learning
* Insights about progress of individual students
* Opportunities for collaborative assessment and goal-setting with students
* Tangible evidence of student learning to be shared with parents, other educators, and other students
* Opportunities to use metalanguage to talk about language

Portfolios Promote

* Student involvement in assessment
* Responsibility for self-assessment
* Interaction with teachers, parents, and students about learning
* Student ownership of and responsibility for their own learning
* Excitement about learning
* Students' ability to think critically about school work
* Collaborative, sharing class rooms

To be more specific, portfolios contain the learners' written expressions, details of books referred to, book reviews, audio recordings of their oral expressions etc. This, however, does not mean that each and every item throughout the course of instruction should be kept in the portfolios. They have to be restricted in number and frequency mainly from the practical and feasibility point of view. Only such items should be retained longer which, in the opinion of the teachers and instructors, will give a sound basis in the context of final evaluation of the learners. Others should be weeded out from time to time after taking necessary steps like working out the remedial measures, improving instructional strategies and techniques, ensuring the sustenance of interest and motivation of the learners and the like. An important point to be mentioned in this context is that

periodical reviews of such portfolios need to be made collectively by the learners and the teachers together. Apart from this, such review only by the instructors and evaluators would be required and would be useful for monitoring the learner's progress and that of instructors for further planning besides making an assessment and evaluation of the learners. In the context of reviewing the portfolios, Genese and Upsher suggest the following guidelines[12].

Guidelines for Reviewing Portfolios

Do's

* Build on what students have done and help them recognize what they have accomplished.
* Praise students for specific techniques or strategies they have used (in writing, for example) and point out the positive effects of those strategies.
* Listen and respond carefully to students' concerns about their progress or difficulties; offer reasonable suggestions in response to these concerns.
* In writing profiles, reinforce students' use appropriate terminology and metaliuguistic references.
* Discuss processes and strategies as well as products.
* Be realistic in setting goals.
* Be positive and supportive at all times.

Don'ts

* Focus on what has not been done or what is wrong with what has been done.
* Be judgmental about student work.
* Make too many suggestions.
* Take over discussion of students' work.

The conferences related to portfolios are primarily meant for reviewing. Thus, they help in the process of evaluation and normally

take place in the form of conversations or dialogues between the learners and the instructors. They could be of various types. For instance, they may take place between the teachers and the individual students or the teacher and groups of students or even the teacher and the entire class as such. While the portfolios are extensively made use of in the context of assessing learners' achievement, conferences are made use of for understanding the processes, techniques and strategies and approaches adopted by the learners during the course of instruction. These conferences facilitate a direct and face-to-face interaction between the learners and the teachers and provide an opportunity for on-the-spot remedial work, improvising instructional strategies and techniques and future planning. It has been argued that the conferences provide more advantages than the portfolios particularly from the point of view of the learners. Some such opportunities include their being self-reflective, being useful to each individual learning goal, in recognizing, appreciating and enjoying their accomplishments and achievements and in communicating role in the one-to-one conversation with their respective instructors about the school work etc. Conferences and Portfolios help the learners to improve themselves from the point of view of communication in general and from the point of view of the content and skills in particular. Conferences are made use of in grading and assessment and therefore the procedures adopted have to be made as much realistic, unbiased and uniformly applicable to each of the learners as possible. Some suggestions related to this are presented hereunder:

- The teacher should direct the conference and retain control of it at all times in order to ensure uniformity of the procedure.
- All students should participate to ensure the collection of the same kind of information to be applied to grading all students.
- The conference protocol should be the same for all students so that individual differences in student performance are not due to variations in protocol.
- Use assessment information from conferences along with information collected from other assessment procedures, such as testing.

- It is essential to keep detailed notes of conferences if they are to be used for grading if called upon to back up your grade with reference to information collected during the conference.

8.1.3 Journals, Questionnaires and Interviews

The third aspect in the context of evaluating without tests relates to the use of journals, questionnaires and interviews. Journals are otherwise called 'interactive diaries' or 'dialogue journals'. The journals are basically written in form and hence they focus around the ability of written expression of the learners wherein the integration of content and language is also included. They provide information regarding learning written skills and strategies, their experiences, process of learning, attitudes and feelings towards the learning program, their instructors and themselves, their interest, expectations and goals. Journals also help in multiplying the avenues and opportunities to the learners for functional communication with their teachers. Further, they enhance student involvement in and ownership of learning. The journal conversations include topic relevance, topic variety, elaboration of topic, meta-communication about reading and writing, audience awareness, awareness and using of print, creativity and independence in writing, grammar and language functions.

In order to make sure that the journals become more useful, interactive and conversational, no riders should be imposed on the learners and the highest possible degree of freedom of expression must be made available to the learners. In other words, the journals should not be highly structured. At the same time, the instructor should avoid direct evaluative feedback for the learners. The opinion and advantage of journals could also be accomplished through interviews and questionnaires. The difference lies in that that the journals are largely under the control of students while other two methods – questionnaires and interviews – are basically under the control of the instructors. Thus, the journals provide a freer hand to the learners compared to the freedom available to them in interviews and questionnaires.

Questionnaires and interviews are more formal and structured in nature in contrast with the journals. Therefore, they become more useful and meaningful when made use of for evaluating learners at higher stages of learning. For the reason that the structuring and formalization processes involved in developing and designing interviews and questionnaires, they consume more time and hence it is desirable to adopt them periodically and not as frequently as journals and other methods. To the extent possible, they must be made use of on completion of each unit of learning. The contents of interviews and questionnaires will also be similar to those contained in journals, checklists and rating scales. The difference lies in the manner of verbalization of items and their arrangement. Normally, questionnaires and interviews are made use of on completion of the learning process to understand about the achievement through a course of instruction. Besides, they are also used for getting feedback about the nature and kind of instruction, the learner's impressions and experiences in the course of instruction including therein their feelings and reactions about the content, organization, instructional materials, teachers' techniques and strategies etc. Therefore, they become more meaningful and useful in the context of evaluating aspects of education other than learners' performance.

The classroom observation, portfolios and conferences, journals, questionnaires and interviews have a significant role to play in the context of any instructional program as they cover almost all the strategies of learning from time to time throughout the course of instruction. Thus, they become meaningful and helpful in taking a consolidated view of learners' performance throughout the process of learning and not limited to one stage i.e., terminal stage. Further, they are more meaningful as the possible tension and nervousness do not figure in these cases as in the case of formal tests and examinations. Making use of all these as the bases for overall assessment and evaluation of learners' performance is in agreement with the views of the educational psychologists who vehemently argue that most learners are prone to become tense and nervous in the context of tests and examinations. Added to these, they provide a chance for assessing

the learners with reference to the coverage of instruction in depth and in detail; while the tests and examinations have a limited coverage as they include the selected aspects and contents of the syllabus and curriculum.

8.2 Self-evaluation Techniques

In addition to the three components of evaluation without tests discussed in the foregoing pages, a forth dimension of self-evaluation has also to be necessarily considered in the context of performance evaluation. This would facilitate the learners to locate, identify and understand the areas of difficulty with reference to the content and skills they are expected to achieve during the course of instruction. In order to facilitate this for the learners, some kind of instant reinforcement has to be built into either the teaching process or the instructional materials. To exemplify this aspect a little more in detail, the instructor during the course of instruction has to make sure that he provides the correct responses to the questions posed immediately following the learners' responses. This would facilitate the learners in comparing their responses with the correct responses provided by the instructor thereby making out the mistakes, lapses and errors, if any and could concentrate on those aspects for better learning. Apart from this, the learners would be in a position to compare their performances with those of the co-learners and develop a kind of competitive spirit whereby their process of learning could be made faster and more effective. The journals, portfolios and conferences would provide a helping hand in the learners realizing their own lacunae and drawbacks and in working towards remedying so as to relate their periodical achievements with the goals, aims and objectives of instruction thereby trained to understand the degree and extent to which they could achieve and the negative aspects, if any, could be made out. In this manner, the self-evaluation techniques have to be developed among the learners primarily from the point of view of realizing the lacunae, drawbacks, problems and difficulties, errors etc., by the learners themselves without being pointed out either by the instructor or by the co-learners. According to Educational Psychology, self-realization of the lacunae, drawbacks,

problems, difficulties, errors and the like by one's own self prevents the development of avoidable inferiority complexes. Such complexes are very prone to develop among the learners if the negative aspects in them are pointed out by others, be it their own instructors or the co-learners. At a slightly higher level of instruction, the self-evaluation techniques would also develop self-confidence among the learners so that they could get the clarifications from the instructors or examiners etc. as to why and how their performance has been rated low if such contingency arises. This would also facilitate the learners in trying to elicit the necessary information from the instructors regarding the ways and means for overcoming the areas of difficulties, mistakes and errors in the process of learning.

Another related aspect to self-evaluation techniques is the evaluation of co-learner's performance by their peers in the classroom. Such evaluation would facilitate better comparison of the achievements, problems and difficulties among the learners themselves besides developing the sense of positive competition among themselves in the race of learning. It is from these points of view that the concept of self-evaluation needs to be developed at least from the secondary stages of instruction if not from the very beginning of schooling.

The foregoing discussion has made it amply clear that for an overall comprehensive, realistic, valid and reliable performance evaluation, it is absolutely necessary to integrate the findings of evaluation without tests and with tests and comprehensive global views taken for drawing inferences about the final outcomes and placement of learners in any instructional program.

8.3 Performance Evaluation through Formal Testing

It has been made clear that for a realistic evaluation of learner performance, use of mere tests and examinations as the basis is inadequate. It has further been clarified that to make the evaluation of learner performance a realistic one, it is necessary to explore and exploit the non-testing methods of evaluation dealt with in the previous section and the outcomes thereof have to be integrated with

the outcomes of the formal tests and examinations. It is a very common phenomenon that in most contexts, tests and examinations are the only basis for evaluating learner performance. It is proposed to deal with the tests and examinations in detail in the present section.

Tests are one kind of method/process for collecting information and in this context the information is about the performance levels of the learners in an instructional program. Although, strictly speaking, a clear-cut evaluation should be made out, attempts to draw such a clear-cut distinction remain practically a herculean task. The reason is that the testing methods and non-testing methods of evaluation have something in common. For e.g. a reading conference with any learner could be precisely structured like a reading test. Similarly, the real delivery of a theme adopted in the non-testing methods of evaluation could be again structured as a composition test. It is in this way that there is proximity between the testing and non-testing methods of evaluation.

A test specifically relates to something like proficiency, intelligence etc. In the educational jargon, tests have definitive subject matter and content. Further, they elicit clearly observable behavior exhibited by the examinees/learners/test takers. They could be widely varied in nature in that that they could either deal with one task or may deal with a host of tasks within the same test. While a composition test assigns a single task of dealing with a given theme, a multiple-choice or an objective test consists of various types and kinds of test items that deal with a set of tasks. In such a situation, each item in the test deals with one particular task. Different test tasks represent different methods to elicit information regarding the learners' performance and all these taken together could form the basis for a realistic performance evaluation.

Tests and examinations result in assigning numerical scores or grades which represent attributes or characteristics to individual learners. To make sure that such scores or grades become useful and meaningful, it is necessary that they must have a frame of reference and such frame of reference is called "measurement". In this way, the tests become a form of measurement.

Broadly speaking, there are three aspects of tests, viz., (1) Content, (2) Method and (3) Measurement.

8.3.1 Test Content

The test content is precisely what the test is about. As mentioned above, it might be the general intelligence, linguistic proficiency or any subject related content; needless to state that the content of the educational tests is related to the academic subjects, knowledge and skills. In the case of language tests, the content primarily focuses on the skills, usage and grammar. In the case of non-language subject tests, the content refers to the facts and figures relating to the subject with which the test is concerned. In other words, a non-language subject test is focused on the subject matter and less on expression.

In the normal course, a test must contain the selected representative samples of almost all the areas covered in the process of instruction. No test, however, could cover each and every aspect of the subject dealt with. Therefore, in order to ensure a good quality test, it is necessary and absolutely essential for the test makers/instructors to carefully pick out the significant aspects of what is taught representing the entire spectrum of coverage before ensuring that such selected aspects represent a general overall knowledge and competence of the learners. Another significant point relating to test content is that no test content could precisely reflect either the content alone or the skills alone. The reason is that in any situation, expression and content go hand in hand and they cannot be totally isolated as complete entities in themselves. Content and expression are supplementary and complementary to each other. Therefore, the outcome of any test must rigidly reflect upon the performance confined either to a single subject or skill. Thus, it follows that the significance of test content reflects the individual performance with reference to the subject/content, knowledge and skills. However, depending upon the purpose of the test, inferences could be drawn about the test takers' competence relating to subject matter or skills or knowledge, taking into consideration the instructor's/examiner's knowledge and information about the backgrounds of the test takers.

8.3.2 Test Methods

Test methods are otherwise called the item types. Each item type or testing method calls for certain definitive kinds of skills or knowledge i.e., independent of the content itself. Experience shows that different item types could be made use of for testing the same skill or knowledge. However, the choice of item types/test methods depends upon the level of instruction, mental maturity of the learners and several related factors. Apart from these, the choice of item types/ test methods has to be made with reference to and in relation to the content with the related skill-oriented tasks. Inspite of exhaustive study and research, there has not been a common universal agreement about the choice of test methods in any context. It is for this reason that an objective test – be it a language test or a subject test – consists of various item types as normally, the total number of items in an objective test runs into a three digit number. Incidentally, various types of test items, their nature, and manner of construction, merits and demerits are dealt with in the later portion of this book.

8.3.3 Tests as Measurement

Measurement of educational achievement through the tests becomes difficult as a precise, realistic and reliable measurement necessitates that the tests must be sensitive to differences in degree within the same kind. This requires that the test items must measure the degree of a particular ability or attribute possessed by the examinee, but not merely the presence or otherwise of an ability or attribute. Further, realistic measurement demands that each test item has to measure one and only one attribute or quality as otherwise the measurement becomes impure or unrealistic.

8.4 Types of Measurement

Normally three types of measurement are talked about in the context of education. They are: (1) Norm-Referenced Measurement, (2) Domain-Referenced Measurement, and (3) Objective-Referenced Measurement.

8.4.1 Norm-Referenced Measurement

Norm-referenced measurement is the measurement carried out with reference to a known group. In other words, in the process of measurement of education, the frame of reference generally used is the knowledge and ability of known groups of people or explicit definitions of the knowledge or ability being measured. In preparing norm-referenced tests, it is necessary to examine each and every item included in the test as to whether the examinees in successful grades perform progressively better on each item as otherwise the item is not a good item for measuring the differences. Thus, it is an essential factor to ensure that the items contained in a norm-referenced test must result in variety and better degree of performance as the levels and grades go higher and higher.

8.4.2 Domain-Referenced Measurement

In the context of education, "domain" refers to a particular area of study like language, mathematics, physical and natural sciences, social studies and the like. The use of domains in test construction demand a common agreement on the nature and degree of knowledge and skills required to be included. It would be ideal if the domains are precisely and comprehensively described which facilitates the interpretation of test scores relative to a clearly identified domain or knowledge or skill or a combination thereof. On the other hand, norm-referenced testing provides interpretation of test scores relative to other examinees. In a domain referenced test, the domain chosen must be finite and known. A well described finite domain is quite nearer to a big chunk of tasks. This being the case, it is desirable that the tasks included in the tests must be representative of the whole domain relating to the subject in question.

8.4.3 Objective-Referenced Measurement

It is clear from the nomenclature itself that the items in an objective-referenced test have a direct bearing and reference to the aims and objectives of instruction. The difference between objective-referenced measurement and domain-referenced measurement is that

in the objective-referenced measurement, it is not necessary to cover the entire domain of the discipline. Thus, objective-referenced measurement could be restricted to a particular lesson or unit or a course of instruction. The choice of content areas and skills and integration thereof, if necessary, could be decided upon and worked out directly with reference to the aims and objectives of instruction. It follows that the objective-referenced measurement is a method very closely resembling the manner of instruction. Interpretation of tests scores in an objective-referenced measurement is made with reference to how many and to what extent the pre-determined instructional objectives have been achieved. Thus, the objective-referenced measurement is beneficial more for the syllabus makers, material developers and the instructors in comparison to the advantages for the learners.

8.5 Evaluating Learners' Abilities

Educational measurement and evaluation is concerned with all facets of education. However, performance evaluation is more frequently and actively carried out to assess the abilities of learners in an educational program. Such abilities could be wide ranging. They could be skill-oriented or general communication-oriented or content-oriented or a combination thereof. At the same time, instruments used for measuring and evaluating such abilities could be with reference to aptitude, achievement, proficiency, diagnosis besides overall intelligence etc. It is, therefore, proposed to discuss about various abilities in the following pages. Irrespective of the discipline in which measurement and evaluation are undertaken, various skills integrating with content form the basis. In this context, it is necessary to make it clear that measuring and evaluating all skills in a formal manner may not be warranted. They will, however, find a place in the on-going/continuous evaluation in some degree or the other. Such evaluation is informal as stated earlier.

8.6 Testing of Listening Skill

Listening is one of the receptive skills and this involves the oral understanding and comprehension of the subject dealt with.

Measuring and evaluating the listening skill finds more important place in the context of language instruction and more so in the case of second and other language instruction. It is no doubt true that listening skill has some place in the context of measuring and evaluating with reference to non-language subjects too. As far as the language instruction is concerned, the methods and measures adopted for the testing of language skills vary depending upon the level of instruction. In the early stages of second and other language instruction, it becomes necessary to ensure accurate reception of the oral communication in the target language. This involves accurate reception of the sound system, the words, the sentences and even running texts in the language concerned for good understanding and comprehension. It is a known fact that in certain situations, even a slight change in the quality of sound could result in the total distortion or loss of meaning. Therefore, the idiosyncratic features of the sound system in the target language have to be very carefully examined, understood and received by the learners. For instance, shortening or lengthening one vowel in a word could result in different meanings. For e.g. in Telugu, the word *nela* means 'month' whereas the lengthening of the initial vowel resulting in *ne:la* means `floor'. Similarly, in Kannada *odi* means `lap' whereas *o:di* means `having been defeated'. In the same manner, an e.g., from Hindi would further substantiate this aspect. *kəl* means 'tomorrow/ yesterday' while *ka:l* means `time'.

It could be clearly seen from these examples that abilities to listen and grasp accurately even at the sound levels is an essential requirement for good comprehension purposes. Apart from the features mentioned above, it is also possible that the target language (second language/foreign language) could contain certain typical and unique sounds which may be non-existing in the source language; for e.g., sound æ does not exist in Hindi while it not only exists in languages like Telugu, but is phonemic in nature. It may be clearly noticed in words like, *taTæku* 'palm leaf', *cæpa* 'fish', and *vellædu '(he)* went'. 'The non-existence of this sound in Hindi makes it difficult for the Hindi speakers learning languages like Telugu to

clearly perceive the sound in question besides making it difficult for them to pronounce it properly. Similarly, the sound ɛ exists in Hindi, whereas it is not found in Dravidian languages like Kannada, Telugu and Tamil. Therefore, for the Dravidian language speakers, it becomes difficult to perceive and pronounce words involving the sound ɛ like *ɛnək* 'mirror', *kɛsa* 'how'. The listening abilities in situations where the grasping of typical and unique sounds are concerned have to be developed among the learners through exercises pertaining to recognition and discrimination of sounds and through the words and sentences consisting of these sounds. Such exercises could be ear training, identification, recognition, discrimination and production. Attention is drawn for a detailed discussion regarding techniques of testing abilities to the author's book 'Evaluation in Language Education' (CIIL,1990).

The testing of listening abilities in the case of non-language subjects is also necessary though not as crucial and important as in the case of second and other language education. Abilities for listening comprehension in the context of non-language subjects has to be developed and in order to test such abilities, the mechanism made use of is through the application of vocabulary tests, listening comprehension tests etc., involving various item types. Mostly close items and to some extent open-ended items like the completion tests and short answer questions are used. These tests need to focus on the content more than the expression. However, proper understanding of the "registers" also needs to be ensured through the listening comprehension tests.

Techniques adopted for assessing listening abilities at the sound level could be adopted for testing these abilities with reference to the structure and vocabulary as well. Each language has its own structure and it is natural that structural differences do exist between the source language and the target languages. Such differences become the primary basis for adoption in preparing the listening tests in the target language. To cite an example, "equative sentences" are not existing in languages like Hindi, English, Malayalam, etc., whereas Dravidian languages like Telugu, Kannada and Tamil do have

equative sentences. This means that the grammatically acceptable sentences without the use of copula is possible in Telugu, Kannada and Tamil whereas it is not in the case of Hindi, English, Malayalam etc. For e.g. following are the equivalents in these languages to mean *'this is a book'*:

idi pustakam (Telugu)

idu pustaka (Kannada)

idu puttakam (Tamil)

It may be noticed from the above that complete grammatical expressions in the languages indicated is possible without using copula/verb. Such constructions are called "equative sentences" or "verbless sentences". On the other hand; the corresponding equivalents in Hindi, English and Malayalam are as follows:

yeh kitab hɛ (Hindi)

this is a book (English)

itu puttakam a:na (Malayalam)

The differences of the kind mentioned above could be many more which have to be taken care of for oral as well as written communication purposes. Consequently, testing in such areas also becomes necessary. As already stated, similar techniques as adopted in the case of sound system can be adopted at the structural level also in the context of language measurement and evaluation. As far as the comprehension test for listening skill is concerned, a carefully selected text from the language/subject is presented either through face-to-face interaction in the class room by the teacher or through the means of a recorded text in the laboratory.

The general recommendation is that such text must he read over to the examinees at least twice following which the comprehension questions are given. These questions in the case of language education mostly refer to vocabulary and structures while they primarily refer to the content in the case of non-language subject education.

8.7 Testing of Speaking Skill

Speaking is the other oral skill and this involves expression. As in the case of listening, the elements crucial with reference to speaking skill also relate to all the levels of language description viz., sound system, words, sentences and meaning. While all these are significant for the testing of speaking skill in the context of language education, testing with reference to vocabulary and registers is significant in the context of non-language subjects. The techniques adopted for testing of speaking skill in case of language education include mimicry, repetition, memorization and oral responses. The aspects included in these techniques could be at any and all levels of language description including vocabulary, styles and registers. Details can be seen from the author's hook "Evaluation in Language Education" referred to under Testing of Listening Skill.

8.8 Testing of Reading Skill

From the point of view of language education and language evaluation, most people believe that the written skills are more important than the spoken skills. This point is however controversial although the general consensus is in favor of assigning more importance to the written skills. As already stated in the earlier pages, in the normal course, an individual is exposed to reading and writing skills after entering the formal schooling system. In the case of language education, reading tests at the early stages of instruction involve the decipherment of printed letters and associating them with the corresponding pronunciation, making out different letters of the alphabets in isolation and in context, relating words to pictures and pictures to words, locating and identifying correct glosses for various vocabulary items, related synonyms and antonyms making out the analogies besides reading comprehension. While the vocabulary and comprehension tests are equally applicable in the context of language and non-language subject evaluation, the other tests including the visual discrimination tests are applicable in language education – be it MT/NL, SL or OL/LF. Discrimination tests could be many in

nature and they are: discrimination of letter among letters, letter among words, word among words, part of word among words, word in phrases, clauses, sentences and part of sentences in sentences. The vocabulary tests consist of the relation of a picture to the corresponding vocabulary item and vice-versa. Attention is drawn to the author's book "Evaluation in Language Education" for a detailed discussion.

Reading comprehension tests consist of a printed text selected from the concerned instructional material or material similar in nature and comparable in degree of difficulty. The examinees are given a chance to read the text twice after which a set of comprehension questions are provided. The examinees are expected to provide the answers based on the text given for reading. Such questions could be either close or open-ended limited to a free response consisting of a word, phrase or a single sentence or at best two sentences.

Before the presentation of the comprehension questions, the concerned passage for comprehension is either kept open or closed depending upon the levels and educational backgrounds of the examinees. Apart from the reading comprehension tests, cloze tests could also be adopted for reading comprehension.

Certain guidelines and suggestions regarding comprehension tests are as follows:

- In the initial stages, it is desirable to give the familiar passages to the learners for comprehension and at a later stage, unfamiliar passages could be given. In both the cases, it is advisable to include discourse patterns, if any, which may be of some acquaintance to the learner.
- A limited number of unfamiliar vocabulary items may be kept provided they do not play a crucial role in comprehending the major portions of the passages or in understanding the central idea of the passage.
- The contents of the passage should not be familiar to the learner particularly in the case of unfamiliar passages, even

though their knowledge is acquired from outside the class room.

- The length of the passages and the number of comprehension questions must be restricted to the optimum degree and the examinees should not get an impression that either the passage or the questions are beyond their scope.
- The writing of items shall, however, be in conformity with the basic procedures of item writing that have already been discussed.

Specialists in educational measurement and evaluation have talked about two kinds of tests in the context of testing reading skills. They are: (1) Power tests and (2) Speed tests

Power tests consist of a number of objective test items arranged in increasing order of difficulty and the examinees are expected to answer as many questions as possible in the prescribed period of time. While the measurement depends upon the accuracy of responses rather than the number of items answered, the evaluation is made based on both the accuracy of responses and number of items.

Speed tests consist of a large number of test items comparable in nature and degree of difficulty and the examinees are expected to answer as many questions as possible. The basis for measurement and evaluation is the number of accurate responses.

8.9 Testing of Writing Skill

Writing and reading skills have quite a lot in common. While reading refers to the receptive skill, writing refers to the productive skill. In the context of language education, aspects relating to the forms of letters including the sizes and shapes, writing down words, sentences etc., providing adequate and appropriate gaps and punctuation marks and the like play a significant role in writing. There are various kinds of writing tests which include pre-writing tests, partial or spot dictation, prepared dictation, paraphrased dictation and a dictation of unfamiliar material. Basically these tests

are relevant in the context of language education more than in the context of non-language subjects.

8.9.1 Pre-writing Tests

The first step for a second language learner in the case of mastering the writing skill is that he should be familiar with the graphemes of the target language. This becomes more important in the case of Indian languages each of which has a different writing system.

The basic procedure of testing the aspect of character formation is simple copying in which the learner tries to reproduce the written model as closely as possible and at a slightly later stage, quick copying is undertaken in the process of which the learner copies the material as fast as possible, while the legibility is maintained.

For the purpose of testing the learners' ability to form the written graphemes of the target language and their combinations within the morpheme, word and other higher levels, it might be adequate to test the legibility aspect rather than the exact reproduction of a written stimulus. Such stimulus may be a word or a sentence or a small paragraph.

A second way of testing this ability is to dictate words or phrases in the target language and ask the learner to write them down. In order to constantly maintain the desirable reinforcement, dictation tests must be corrected on the spot and given back to the learners immediately.

Dictation involves the relation of the written symbol to the spoken sound. Thus, while taking down dictation in the target language, the learner has to associate the sounds of the language to the corresponding written form. Dictation can be a strategy for testing the written skills at lower levels of language instruction. Scholars have argued that dictation cannot be more effective for testing of the writing skills in the case of advanced learners. This type of testing strategy requires the learner to possess accurate speech perception and its relation to the writing system. In spite of the

arguments in favor of and against dictation, it has its own merits. Most of the language teachers are convinced that the more dictation work is, the better the terminal achievement of the learners would be. Various types of dictation are possible.

8.9.2 Partial or Spot Dictation

In partial or spot dictation, each learner is given a copy of the passage in which function words or even only prefixes and endings are left out. While the passage is read by the examiner or the teacher, the student has to fill-in the blanks. Although such exercises require a lot of advance preparation, they have distinct advantages. The three major advantages are:

- They can be administered quickly and the scoring can be more objective.
- The examiner or the teacher can test only the problem areas.
- The learners need not waste time in writing words, phrases etc., that they have already mastered.

8.9.3 Prepared Dictation

Prepared dictation is nothing but dictating a familiar or a known passage or a conversation or a dialogue. The learner would have had the opportunity of either listening or reading the piece during the course of instruction. This kind of dictation is normally recommended in the initial stages of the introduction of writing skill.

If the teacher or the examiner wants to measure the learners' retention and familiarity with the patterns, each sentence should be read only once and at normal speed. Such exercises have proved to be more effective in the case of older students.

8.9.4 Paraphrased Dictation

In this type of exercise, the dictated material would be nothing but a gist of a dialogue or a selection that the learner is already familiar with. The dictated paraphrase will, however, contain only the familiar patterns and vocabulary that the student possesses, but the wording will not be exactly the same as contained in the original

dialogue or the selection. For this reason, the validity of this test and its results would be greater in assessing the learners' language achievement.

In this type of dictation, reading of each sentence is normally allowed once or twice at a normal speed.

8.9.5 Dictation of Unfamiliar Material

Dictation of a material unfamiliar to the learner is usually preferred to the other type of dictation at the intermediate and advanced levels and they are considered to be more reliable and valid to measure the learners' skill in dictation. Following procedures are adopted in giving such dictations.

First, the whole passage is read at normal speed. Then, the students are expected just to listen carefully and not to write. The passage is then read for a second time when each phrase is read at a time giving appropriate pauses during which the students are expected to write down what they have heard. The second dictation should be consistent and the teacher may read each phrase once or twice. After the dictation of the passage is over, it is read finally again at normal speed and the students are given a few minutes' time for any revision that might be felt necessary. The teacher must ensure that no part of the passage is read at the request of the learner.

In order to ensure consistency in reading and save the teacher's time etc., such dictations are recommended to be given through language laboratory where such facilities exist. In this case, the examiner must ensure that the students do not play the tape a second time.

CHAPTER

9

Types of Tests

Different kinds of tests are talked about in the field of education. Generally, four kinds of tests are talked about and they are:

1. Achievement Test
2. Proficiency Test
3. Aptitude Test
4. Diagnostic Test

Apart from these, some other tests are also talked about though less frequently in the field of education. Some of them are: (1) Intelligence Test (2) Creative Test, etc.

The nature and characteristics of these tests are dealt with in brief hereunder:

9.1 Achievement Test

The aim of achievement tests is to examine the extent to which the examinee has been successful in meeting the set aims and objectives and goals of instruction on completion of a program. The degree or extent of success will highlight the examinee's achievement with reference to the total coverage of the syllabus including therein the skills and content. The uses of achievement tests are manifold.

While on the one hand, they are primarily made use of for deciding upon the success or otherwise of the examinees, on the other, the findings are also made use of for effecting changes in the syllabus, instructional material, instructional methods or making any modifications therein for betterment in future. The limitations of achievement tests are that the findings will not be useful in predicting the future performance of the examinees. As the basis for achievement tests is clearly defined in terms of the syllabus, the aims and objectives and goals of instruction, the scope and extent of coverage in the achievement tests is clearly defined. Besides, being helpful in deciding upon the success or failure of the examinees, the findings would also help in placement of the examinees into different categories depending upon the degree and quantum of achievement. This would, in turn, facilitate in choosing appropriate course of study in future.

9.2 Proficiency Tests

Proficiency tests and achievement tests have much in common in that that both are used to make an assessment of the competence of the learners at the time of examinations. The difference is that the proficiency tests are not based on any clearly drawn up syllabus, instructional material, aims and objectives and instructional goals; while the achievement tests are based upon such well-designed factors. It could therefore be stated that the proficiency tests refer to the knowledge and competence of the examinees without reference to any framework, while the achievement tests are with reference to a clearly made out and pre-determined framework of instruction. Some standardized proficiency tests are TOEFL (Teaching of English as a Foreign Language), Michigan Test of English Proficiency, Cambridge English Proficiency Test, English Proficiency Test Battery, etc. Efforts are being made to develop such standardized proficiency tests in Indian languages in the recent past and hopefully they should be available for use in the near future.

9.3 Aptitude Tests

The name in itself is indicative of the nature and purpose of testing. Aptitude tests are used to understand the bent of mind of the examinees towards a subject of instruction. Such tests would be helpful in choosing the field of specialization of the examinees. Various kinds of aptitude tests are talked about which include (1) Scholastic Aptitude and (2) Vocational Aptitude Test.

9.3.1 Scholastic Aptitude Tests

Scholastic Aptitude Tests provide an idea as to how the examinees could perform in future courses of study and the possible degree of a success in them. The findings of scholastic aptitude tests could add or supplement the past academic performance of the examinees. The idea behind the scholastic aptitude tests is to see whether its findings contribute usefully to the predictive information already available.

9.3.2 Vocational Aptitude Tests

Vocational Aptitude Tests help in identifying the areas of interest of the examinees which, in turn, will help in locating and finding a suitable course of instruction that they could go through for prosecuting their own professional interests and benefits.

9.4 Diagnostic Tests

The name itself is indicative of the purpose. Diagnostic tests are made use of for locating and identifying the areas of difficulty/ problems of individual examinees during the course of instruction. This would facilitate in remedying their problems and difficulties in learning either in groups or individually by designing suitable remedial instructional material. Unlike the achievement and proficiency tests, diagnostic tests are administered at various stages of instruction in addition to the one at the entry level. The outcomes and findings of diagnostic tests could he useful not only from the learners' point of view, but from the points of view of the instructors, the material producers and the educational planners.

9.5 Intelligence Tests

Intelligence tests are administered individually on different examinees. Thus, they are individual-based and they provide information about the abilities of an individual which, in turn, can be made use of for making inter-comparison among the group of examinees. Intelligence tests can be said to be a measure of mental abilities of the examinees. Normally, such tests are made use of at higher levels like the masters and post-master's levels of education. The assessment of the intelligence tests is made by a group of proficient counsellors apparently to overcome the degree of subjectivity in assessment. The advantage in making the assessment of a group of counsellors and considering an average of the findings of all would facilitate reduction of the degree of individual bias though not completely. These tests are used in the context of selection and recruitment of personnel to different posts wherein a relative assessment of the interviewees becomes necessary. These intelligence tests could consist of both verbal and non-verbal communications besides the oral and written media. Thus, the coverage in the intelligence tests is global in the sense that they attempt to assess the overall intelligence, knowledge and experience of the examinees/ interviewees.

9.6 Creativity Tests

It is generally believed that the creative abilities among individuals are God given. At the same time, the nature and degree of such abilities will vary from individual to individual. On the one extreme, the creative abilities of an individual could be a cipher; while on the other, some individuals could possess a very high degree of creativity. Keeping in view the diversity and differences and the kind and nature of creative abilities among individuals, it follows that creativity tests do not measure the same attribute or attributes which are taken care of normally in the conventional intelligence tests. The unique feature of creativity tests is that they measure different and larger unrelated characteristics of people. The very concept of assessing divergent thinking and creativity demands that the

examinees answer by themselves without selecting or choosing a better answer from among the choices available. The content of such creativity tests could be eliciting information relating to a particular theme including therein the innovations and novel ideas of the individuals with practical devices besides the fluency, flexibility and originality exhibited by the examinees. This being the case, the measurement and assessment in the creative tests again becomes highly complex, difficult and subjective. The degree of subjectivity could be reduced by providing a set of criteria based on which the assessment could be made. Keeping in view the fact that the creative abilities of no two individuals could be identical, the inter-examinee comparison becomes very difficult if not impossible. Normally, the creative tests could be thought of either in the case of literary-oriented language tests or tests relating to innovative subjects like sciences in particular and some subjects in humanities in general.

Several other kinds of tests are also talked about. The scope of the present book does not permit discussing about all these tests.

9.7 Tests and Examinations

It is not infrequent in the field of educational evaluation and measurement to come across the expressions "tests" and "examinations". According to Davis, "the notion of tests conjures vague ideas of psychology and of intelligence, whereas an examination suggests the end of term multi-subject ordeal"[15]. Some specialists in the field of measurement and assessment have come out with the suggestion that tests are more objective, more reliable, more specific and more realistic than the examinations. Inspite of differences between the tests and the examinations pointed out by various scholars, most often than not, these two terms are used interchangeably. If one examines the structure of tests and examinations, it could be clearly seen that the tests are not as highly structured as the examinations are and the coverage in tests is limited while examinations have a vast coverage. Further, the examinations have one special advantage in that that the outcomes of examinations have a bearing on the curriculum and syllabus and the related aspects

of an educational program. Broadly speaking, therefore, one could say that examinations are normally meant for assessing the terminal performance of the learners whereas tests are made use of for learners' performance at different stages during the course of instruction. Another distinction between these two terms can be made in that that the examinations cover the entire spectrum of syllabus and content, while the tests cover one or some pre-determined portions of the syllabus and content.

CHAPTER

10

General Principles of Test Construction

10.1 Precautions for Construction of Good Language Tests

Any experimental design, framing a hypothesis etc., requires systematic procedure in order to ensure that the procedures adopted are appropriate and the goal is arrived at successfully. Similarly, the construction of tests also requires a systematic procedure and adoption of various steps which may broadly be categorized as follows:

1. Planning the test
2. Preparing the test items and instructions
3. Reviewing and revising the test items
4. Pre-testing the material and analyzing the results
5. Preparing the final form of the test

10.1.1 Planning the Test

The planning of a language test mainly involves the following factors:

a) Determination of the general course objectives.

b) Division of such objectives into their minute and sub-components.

c) Establishment of general design of the test including the time factor.

Division of the Course Objectives into their Components

A second pre-requisite in the planning of a good test is to clearly specify the learning outcomes in terms of the learner's terminal behavior and divide them into sub-components and define them as narrowly as possible so that the test items are focused on these outcomes.

10.1.2 Preparing the Test Items and Instructions

In order to have well-made test items for inclusion in a test, it would be necessary to prepare a larger number of test items which have to be tried out and reviewed to ensure that they are non-defective and perfect. Some items may, after review, be found to be defective which need to be discarded and should not be included in the test. Only after a thorough review and try-out of the test items, the required number of the perfect test items could be selected and retained for inclusion in the test.

While it is quite essential to ensure that each test item is perfect and non-defective in the sense that it is unambiguous in the task presented to the examinees and each item has one and only one correct answer, it is also equally essential to provide the instructions or the directions as to what is expected from the examinees taking the test. These instructions should be brief, simple and unambiguous. Wherever possible, example(s) should be provided so that the examinees do not misunderstand or misinterpret the instructions. To the extent possible, it is also desirable to ear-mark the time specified for each portion of the test and also the marks allotted to each portion.

10.1.3 Reviewing and Revising the Test Items

Once the test items are prepared, it is always desirable to get them reviewed by experts other than the item writer himself so that the items possessing even minor defects could be avoided and the test as a whole made as valid as possible.

10.1.4 Pre-Testing the Material

Once a tentative form of the entire test is framed consisting of various test items after review and revision, in order to make the test a standard, valid and reliable one, it is always recommended that the tentative form of the test is administered on a large sample of subjects of the same standard and kind as those for whom the test is aimed at. In other words, the tentative form of the test has to be tried out which may indicate the defective aspects which have to be used in revising and giving a final form for the test. It is only then that the construction of a good test could be ensured.

10.1.5 Assembling the Final Form of the Test

After the try-out is over, various test items included in the test under each category need to be graded in order of difficulty and re-arranged thereby giving the final form of the test.

It is only by subjecting the construction of language test to these five major steps that a good language test can be constructed and hence, though these factors appear to be trivial in nature, are very crucial and play a significant role in the construction of good tests.

CHAPTER

11

Types of Questions and Items Types

There are various kinds of questions and items very commonly used both in formal and informal testing and measurement situations. They can be broadly classified into two categories, viz., (1) Subjective Type and (2) Objective Type.

The generally adopted terminology in this context is that the subjective type of items are called 'questions' while the objective type of items are called 'items'. There is one more kind which falls between these two varieties. This is, in a way, an open-ended item (wherein the examinees have the freedom to express independently) in the assessment of which the objectivity is however, available. This means that while there is a scope for the individual's free expression in response to the question, the objectivity is ensured in that that the answer would be a very brief one and it could only be either correct or incorrect. The question types which are subjective in nature can again be sub classified into the following categories:

1. Essay/Composition
2. Paragraph Writing/Brief or Short answer (open-ended) question
3. Translation

4. Précis Writing
5. Composition.
6. Critical Writing

The objective type of items include

1. Multiple-Choice
2. Yes/No questions
3. Matching type
4. Fill-in the blank type
5. One-word answer questions

As far as the subjective questions are concerned, there is very little reliability from the point of view of assessment. The reason is that the answers consist of free expression on the part of the examinees without any restriction being imposed. Further, the answers run into a brief text which could be either a description or a narration or a critique and the like. It is natural that the ideas presented in such open-ended answers vary from individual to individual. This makes it difficult for any examiner to objectively measure and assess. On the other hand, the objective items have very high reliability as the correct answer to each of those items is very precise and clear and there is no chance for an alternative answer to be correct. Further, the free expression in these items is either nil or may at best consist of a word or a phrase or at the most a simple sentence. Thus, the degree of reliability in objective type of items is very high. The item that falls between these two extremes is called short-answer question. It demands a free expression on the part of the examinees running into one or two sentences. This being the case, certain degree of reliability can be expected from such questions as the accuracy of responses is practically definitive.

It may now be appropriate to discuss various kinds of questions and item types in brief hereunder.

11.1 Subjective Questions

11.1.1 Essay/Composition

There is very close proximity between the questions relating to the essays and those pertaining to composition. Essays refer to the content covered during the course of instruction besides its being available in the course material. Composition refers to a set of text-free items which would fall within the broad spectrum of the examinee's experience, exposure and knowledge. Thus, the essay type of questions is highly text-oriented and compositions are text-free. Another distinction from the point of view of the examinees' responses could be made out in that that the spectrum of coverage of content in the essay questions is quite definitive as it pertains to the coverage in the instructional material and the course of instruction. Definitive coverage cannot be expected in compositions as the expression in them relates to the individual's experience, exposure and knowledge. Thus, they are individual specific. This being the case, certain criteria and guidelines could be worked out for assessment and measurement of essay type questions so as to enable objectivity in some degree. It cannot, however, be made 100% objective as, at least from the point of view of usage of sentence patterns and vocabulary, the expression differs in quality and with reference to the styles etc. Thus, while the content will have a close resemblance of the spectrum of coverage during instruction, no such expectation can be made out of the styles, vocabulary and sentence structures. It follows that the assessment and measurement in the case of essays could be made objective from the point of view of content, but the expression can be measured only subjectively; on the other hand, assessment of composition writing remains subjective both from the point of view of content and expression. However, certain attempts and experiments are being carried out trying thereby to bring in certain degree of objectivity in measurement and assessment of compositions also. The attempts include identifying the major ideas relating to the theme, ease/difficulty of expression, observing the limitations pertaining to the vocabulary control and control of sentence patterns within the broad realm of the examinees' expected level of linguistic

competence. However, this aspect remains to be further explored and concrete bases arrived at.

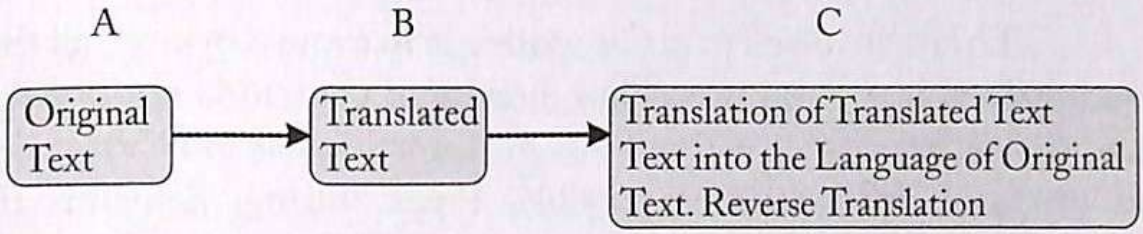
11.1.2 Paragraph Writing/Brief/Short Answer (Open Ended) Questions

Paragraph writing/Brief/Short answer (open-ended) questions are basically related to the content of the course of instruction. The expectation from these questions is to elicit answers pertaining to a restricted or limited area of the content providing therewith an opportunity for the examinee to express himself freely. These are, in a way, parallel to the essay type of questions. But, the difference is that the coverage of content is wider and broader in the case of essay type of questions whereas it is limited and restricted in the case of paragraph or short answer questions. The degree of subjectivity and objectivity with reference to measurement and assessment of responses relating to paragraph questions is also the same as those of the essay type of questions. Text-free paragraph questions are also possible and they are made use of at slightly higher levels of instruction. They refer to the themes and content areas beyond the scope of the coverage in the course of instruction, but within the experience and exposure of the examinees. Therefore, the kind of subjectivity that exists with reference to the composition will exist in this case also. But, the scope and coverage being limited in this case, some degree of objectivity can be expected from the point of view of assessment and measurement.

11.1.3 Translation

Translation is a kind of technique made use of in the context of language testing/measurement. This relates to transferring the expression and content from one language to the other. In the beginning levels of instruction, translation could be adopted as a testing method, in expecting the translation from the target language to the source language. At higher levels of instruction, however, it could be vice-versa. Measurement and assessment of translation is a little more objective compared to that of essay and composition questions. The reason is that while the content cannot be altered,

adoption of vocabulary and sentence patterns corresponds to those made use of in the text from which translation is expected. Thus, it may be noticed that the content and the expression are more controlled compared to essays and compositions. This facilitates ensuring certain degree of objectivity in assessment. The question still remains as to how good the translation is and how the degree of efficiency in translation can be assessed. The best mechanism for the assessment of degree of efficacy in translation is the adoption of 'reverse translation'. Reverse translation means that while the first translation exercise is made from target language to source language, the reverse translation takes place from the translation made into source language back to the target language. The reverse translation, thus made has to be compared with the original text and the degree of resemblance of such reverse translation with the original text determines the degree of efficacy of translation. This may be diagrammatically represented as follows:



Efficacy of Translation

Comparison between A and C

While the technique of reverse translation can be adopted with reference to limited extent of translations, there are practical constraints in adopting these in case of voluminous translations. It needs to be mentioned here that the translation as a testing mechanism is made use of from the secondary level of instruction onwards.

There are several aspects that need to be looked into regarding translation. Various types of translation are talked about wherein the prominent are:

1. Thematic translation and
2. Word-bound translation.

Thematic translation refers to the transfer of content from one language to the other without insisting upon the adoption of parallel styles of expression, vocabulary and sentence patters in the language translated into. This means that the concern is primarily about content and little concern is about expression. In the case of word-bound translation, the primary focus is laid on styles, vocabulary and sentence structures thereby insisting on a parallelism between those adopted in the original text and the translated text. This type of translation has a number of limitations and it could at best be made use of at lower levels of instruction in translations limited to small sentences or fractions thereof. There is little advantage and utility of word-bound translation at higher levels of instruction where content plays a prominent role. There are other varieties of translation viewed from the linguistic point of view and the scope of the present book does not permit discussing all those.

11.1.4 Précis Writing

The task involved in précis writing is to express in brief, all that is contained in a given text. Thus, brevity of content is not possible and the entire content without loss or distortion has to be expressed as precisely and briefly as possible. Précis writing demands the examinees to ensure that there is no loss of ideas contained in the original text. The scope available for précis and brevity could be applied to the supporting details and arguments in the original text in support of the ideas and sub-ideas. Further, the examinee has to ensure the same logical and sequential arrangement of ideas. At best, the précis arrived at may not contain some sub-ideas that are relatively less important. Freedom of expression for the examinee would no doubt be available for précis writing. This facilitates the examinees in choosing vocabulary and sentence patterns of their choice and in their control. They need not necessarily be the same as those used in the original text. Thus, it may be seen that while no loss of ideas can be acceptable in précis writing, freedom of linguistic expression could he granted to the examinees. Apart from these, the précis writing may also ignore the descriptions and criticisms within the original text. It follows that the assessment and measurement of

précis writing becomes objective from the point of view of content, while it remains subjective from the point of view of expression. Such exercises can be adopted from the secondary level of instruction onwards.

11.1.5 Composition

Broadly speaking, composition can be of three kinds, viz., (1) Guided composition (with linguistic cues or with pictorial cues), (2) Comprehension-cum-composition, and (3) Free composition.

Guided Composition involves the provision of hints and cues based on which a running text has to be developed. At the earlier stages of instruction, these cues could be either linguistic or pictorial. The linguistic cues involve making available content words relating to the theme and so also the pictures. In the very early stages of instruction, such cues are arranged in the manner in which the arrangement of ideas has to be ensured in developing composition. This makes the task easier for the examinees as the question of logical or sequential arrangement of ideas does not rest on the examinees' shoulders. The only expectation is that they have to make free expression based on the cues provided in the same serial order. The scope and range of sentence patterns and vocabulary remains limited.

At slightly higher levels of instruction, the cues so provided for the development of composition will be disorderly and the responsibility of arranging the ideas in a logical and sequential manner also rests on the examinees' shoulders. Thus, the task of the examinees becomes tougher. Guided composition can be relatively more reliable as the scope and coverage of content is pre-determined. Needless to say that the degree of reliability will be much more at the earlier levels of education and it will be relatively less at higher levels.

Comprehension-cum-composition is an extension of comprehension exercise. Comprehension involves the understanding of a given text and answering the questions correctly based upon the text. The composition aspect demands in addition to comprehension that the answers to the questions given have to be arranged and sequenced logically in such a manner that it yields free-flow of

expression of the theme dealt with. Therefore, the task of the examinees is two-fold in that that in addition to answering the questions, the answers have to be logically and sequentially arranged. This exercise of comprehension-cum-composition is relatively objective as the coverage of content is pre-determined and the flexibility of expression will be limited as the original text indicates the types of sentences and vocabulary to be made use of. Thus, the assessment and measurement in this case would be more objective compared to other two types of composition. The text chosen for this purpose could be either from the textual material or beyond the scope of the text. Depending upon the level of education, the choice of one kind or the other or a combination thereof can be preferred.

Free Composition provides complete freedom for the examinees both from the point of view of scope and coverage of content and from the point of view of expression. The theme, however, is provided. Each examinee will have his freedom to make use of the range of ideas depending upon his experience, exposure and knowledge. Thus, the assessment and measurement of free composition is highly subjective and relatively more difficult in comparison to other types of composition. This mechanism is adopted as a testing technique more usefully at higher levels of education rather than at the early stages.

11.1.6 Critical Writing

Critical Essay/Critical Writing is one of the important types of subjective questions that is adopted at higher levels of learning say from the graduation level onwards to assess the critical abilities of the examinees. In this case, a pre-determined concept is given to the examinees and they are asked to discuss different viewpoints of various scholars and examine them critically with a positive bent of mind. Further, in such critical examination, the examinees are expected to arrive at a reasonable and convincing and consolidated viewpoint about the theme/subject given. This exercise involves a detailed examination of merits and demerits, pros and cons strictly with a positive attitude avoiding the destructive/negative attitude.

A further expectation from the examinees could be that in addition to or apart from drawing up a consolidated view about the theme/ topic, they could even express their own viewpoint which might differ in some aspects or the other from the other viewpoints. In either case, the examinees have to provide necessary supporting details, reasons and arguments in respect of the conclusions arrived at.

Assessment and measurement of questions involving critical examination is relatively more difficult compared to the other kinds of subjective questions. The reason is that no two individuals can be expected to hold the same viewpoint about any topic/theme. Their judgments and attitudes depend upon their experience, exposure and knowledge. It is natural that the examiners will have their own viewpoints which may differ from the examinees' views. Where such a contradiction exists, the assessment is likely to be biased. With a view to reducing the degree of bias, it is desirable to have the assessment made by a group of examiners and the assessment of all the examiners is consolidated and final conclusion arrived at and relevant value judgments drawn. Such an exercise would no doubt reduce the degree of bias, but bias and subjectivity cannot be completely avoided. Examination of both views relating to content and the linguistic aspects of expression need to be considered in assessing the critical writing/critical essays. It would apparently be possible to adopt critical writing/critical essay as a mechanism of testing only at higher levels as abilities for critical examinations could be expected on the part of the examinees only after they attain mental maturity.

11.2 Objective Type of Test Items

Objective type items are very precise and objective from the point of view of assessment and measurement. In almost all cases, such items have pre-determined, brief and precise answers. Preparation of objective type of items demands very high skill and expertise unlike the subjective type of items. On the other hand, assessment and measurement of objective type items is very easy while it is very complex and difficult in the case of subjective type of items.

11.2.1 Multiple-Choice Items

Multiple-choice items contain a 'stem' and a number of 'options'. Stem refers to the question and the 'options' contain a number of answers to the question contained in the stem. These options contain one 'key' (correct answer) and 3 to 4 'distractors' which are incorrect, but closely resemble the key. The examinee has to select the 'key' among the options. The stem may be a direct question or an incomplete statement. A wide range of abilities could be tested and consequently the range of coverage of content and skills could also be wide enough. Constructing this type of items is very difficult as the test maker has to ensure close resemblance of the distractors with the key. This type of items involves a probability of guessing which depends upon the number of options.

11.2.2 Yes/No Questions

The *Yes/No* questions are similar in nature to the multiple-choice items. The difference is that the number of choices available in this case is two. Another difference is that the *Yes/No* type of items mostly involve statements relating to facts. They cannot, however, be incomplete sentences. The probability of guessing in this case is 50%. Preparation of this type of items is not difficult as this involves the selection of important statements relating to the content.

11.2.3 Matching Type

Matching type of items is very similar to the multiple-choice items. In this case, there are two lists of words which are called 'premises' and 'responses': The expectation from the examinees is that they have to relate each item under the 'premises' with the corresponding items in the 'responses'. Normally, the number of items in responses is more than those in the premises. The basis for matching has to be explained as the matching could be either between words or pairs of words wherein there would be a logical relationship and the like. In such cases, the kinds of logical relationship will have to be taken as basis for matching and hence the need for explaining the basis. The degree of guessing in this case is much less and they are not suitable for testing higher order abilities.

Constructing matching type of items is relatively easy compared to the multiple choice items.

11.2.4 Fill in the Blank Type

Fill in the blank type of items demands the examinees to fill up the blanks with suitable words in the isolated sentences given or in the running text wherein blanks are left. Normally, this type of items is made use of for testing the knowledge of the examinees with reference to content. However, they are also made use of in language education to test the examinees' linguistic competence. Depending upon what is aimed at – whether content or linguistic ability – the gaps or blanks left should relate to the content words or the structure words. It is very easy to construct and a very high degree of objectivity exists. At the earlier stages of education, cues in terms of a number of words at the end of the test could be presented whereas such cues need not be presented at higher levels of education. Scoring is also very easy besides ease in construction. Fill-in the blank type of items, in case of adopting a running text are precisely the same as 'cloze tests' discussed earlier.

11.2.5 One-Word Answer Questions

The one-word answer questions are simple questions. Complete sentences are not required for answering and it would be enough to give one word in response to each question. Construction of such questions is not difficult and scoring becomes easier and the assessment and measurement are highly objective in nature. Mostly, such one-word answer questions are relevant in the case of testing content rather than expression.

After having discussed the kinds of questions and item types, it would be relevant to discuss how the quality of test items could be tested. Various criteria adopted for the purpose may be seen from the following:

11.2.5.1 General Criteria

- Does the item measure an important learning outcome?
- Is the item focused to measure an important content area?

- Is the level of difficulty of the test item in consonance with the degree of achievement of the learner?
- Is the item likely to be accurately answered by the learners possessing higher abilities?
- Is the item likely to be answered wrongly by learners possessing lower abilities?
- Is the item independent or does it overlap with the other items?
- Is the item likely to provide a clue for the answers for any of the other items in the test?

In addition to the general criteria that may be adopted in respect of any type of test item, the following specific criteria may he adopted in case of the items of various types.

1. *Multiple-Choice Items*

   a) Is the stem precise and unambiguous?

   b) Is the stem a complete question in itself or does it require the examinee to read the alternatives provided to understand what is being asked?

   c) Is the context in which the question is being asked clear?

   d) Does the stem include anything that has to be repeated in all the choices?

   e) Are the choices given closer to each other in respect of
      (i) Content?
      (ii) Structure?

   f) Do the choices given include one and only one correct answer?

   g) Are the distractors provided intensely closer to the key?

   h) Is the language used in the items and instructions easy enough for the examinees to understand?

   i) Are the vocabulary items used in the items and instructions within the level of comprehension of the learner?

2. *Matching Type*

   a) Does the item include only homogeneous material in the premises?

   b) Is the number of responses enough so that the last of the premises can still have many options to choose from?

   c) Does the item specify the basis for matching, type of matching etc.?

3. *Simple Answer or Completion Type/Fill-in the Bank Type*

   a) Is the item constructed in such a way that only a single short answer is possible? (a word or a phrase or a clause or a simple sentence).

   b) Is the question direct and simple?

   c) Is the answer related to the main point in the statement?

   d) Does the item with the blank space make enough sense so that the learner knows what is expected of him?

   e) Are the directions given in clear and simple terms?

   f) Is the question likely to be answered or interpreted in the same way by the teachers/students/examiners?

   g) Is the difficulty level of the question appropriate to the degree of learner's achievement?

4. *True/False or Yes/No Type*

   a) Does the item include one and only one significant idea in each statement?

   b) Is the item precise enough to enable the examinee to judge whether it is true or false unequivocally?

   c) Is the language used in the statements simple enough?

   d) Does the item use negative statements sparingly and avoid double negatives?

CHAPTER

12

Scoring

The task of scoring is as significant and crucial as the test construction is. The techniques and methods of scoring differ from test to test depending upon the nature and kind and purpose of testing. Just as the mechanics of testing would be broadly divided into two categories viz., (1) Subjective and (2) Objective, scoring techniques also have to be looked at accordingly from the two points of view.

12.1 Scoring of Subjective Tests

There are two methods of scoring.

i) By assigning numerical scores.

ii) By assigning grades.

Assigning of numerical scores in respect of subjective questions requires a very high degree of precision, accuracy and relevance. Precision can be expected in the highest degree if the bases for assessment are very clearly spelled out without any room for diversity in thinking. Such precise and unalterable bases are extremely difficult to be worked upon and it remains only a theoretical possibility. This being the case, assigning of numerical scores in respect of subjective type of questions poses reservations about the validity, reliability etc.,

of assessment. Therefore, it is highly desirable and necessary to prefer assigning grades rather than numerical scores. Such assignment of grades has to be necessarily related with a fixed range between two numerical scores which are to be approximated. Such approximation and determination of' the ranges is possible in an effective manner only by experienced evaluators and testers. The possibility of difference of opinion between examiners assessing the same answers would be much less as the ranges pertaining to each grade will be and should be adequate enough to accommodate the differences of opinion. If, however, a large gap exists between different examiners' assessment, it has necessarily to be valued by a third examiner. It is only after having the assessment made by different examiner which falls within the fixed range of scores, an average could be taken and the relevant grade assigned. As already mentioned, the degree of subjectivity even in the case of subjective type of questions could be reduced by clearly and precisely providing a set of criteria and bases for assessment and measurement. But, for the possible reasons of differences and diversity with reference to the content and freedom of expression, subjectivity cannot be completely avoided and the assessment remains subjective beyond doubt.

It may not be out of place to take a look at the table relating to different kinds of validity presented by Cronbach in his 'Essentials of Psychological Testing'.

Predictive

Question asked	: Do test scores predict a certain important future performance?
Procedure	: Give test and use it to predict the outcome. Sometime later, obtain a measure of the outcome. Compare prediction with outcome.
Principle used	: Selection and classification.
Examples	: Admission test for Medical students compared with later marks.

Concurrent

Question asked : Do test scores permit an estimate of a certain present performance?

Procedure : Give test, obtain a direct measure of the other performance and compare the two.

Principle used : Tests intended as a substitute for a less convenient procedure.

Examples : Group mental test compared to individual test.

Content

Question asked : Does the test give a fair measure of performance on some important set of tasks?

Procedure : Compare the items logically to the content supposed to be measured.

Principle used : Achievement tests.

Examples : A test of shorthand ability is examined to see whether the content is typical of office correspondence.

Construct

Question asked : *How* can scores on this test be explained psychologically?

Procedure : Set up hypotheses. Test them experimentally by any suitable procedure.

Principle used : Tests used for description or in scientific research.

Examples : A test of art aptitude is studied to determine how largely scores depend on art training, experience in Western culture.

12.2 Scoring of Objective Tests

Scoring of objective tests is much easier while their construction is extremely difficult, laborious and time consuming.

Generally, each item in the objective tests is assigned one mark. Thus, total number of marks depends upon the total number of items in the test. In the normal course, while every correct answer gets one mark, every wrong answer gets zero each. The assignment of scores is generally made in this manner. There has been the concept of minus/ negative marking in the case of objective tests. The purpose behind negative marking is to dissuade and discourage the examinees from wild guessing. The reason is that in the absence of negative marking, most examinees, if not all, resort to guessing in spite of their being sure about their not being able to comprehend the questions and answer them. In this manner, even those examinees who do not possess any knowledge whatsoever merely take a chance to guess the correct answers. On the other hand, applying minus/negative marking will exercise a restraint on the examinee to prefer to guess the correct answers only where they feel that they know the subject, but they are, however, not sure about the exact and correct answers. Thus, negative marking would not provide complete freedom to make undesirable, unworthy and wild guesses. Therefore, this system could be made applicable at higher stages of learning or in such tests which are made use of for selections involving very senior and responsible positions. Such negative marking is, however, not recommended in normal educational situations except where a really tough competition has to be insisted upon.

12.3 Validity and Reliability

There are certain factors to be ensured in the context of educational testing and measurement. Validity and Reliability are two important characteristics. "Validity refers to the extent to which the results of a measuring instrument serve their intended use"[16]. Reliability is concerned with the consistency of results when the same measurement procedure is applied more than once. Such consistency provides a necessary but not a sufficient condition for obtaining validity and it indicates the degree to which test results are dependent *(Ibid)*.

With the advancement of innovations and research in the field of educational psychology, certain modifications and changes with reference to these concepts have been made. Validity in the present times is referred to as 'content validity'.

It is a known fact that testing and measurement are used for various kinds of judgment. Three kinds of validity are talked about in the context of educational testing. They are:

1. Content validity
2. Criterion-related validity
3. Construct validity

Content validity is demonstrated by showing how well the content of the test samples, the class of situations or subject matter about which conclusions and judgments are to be arrived at. This concept is specially significant and important in the context of achievement and proficiency testing. They are also significant in contexts like adjustment or social behavior based on observation in selected situations.

Criterion-related validity is demonstrated by comparing the test scores with one or more external variables considered to provide a direct measure of the characteristic or behavior in question. Such comparison may result in an expectancy table or a correlation relating the test scores to the criterion measure. Predictive uses of tests include long range forecast of one or more measures of academic achievement, prediction of vocational success and prediction of reaction to therapy. The criterion data collection could be either concurrently with the testing or after completion of the testing process. This depends upon whether the test is recommended for prediction or for assessment of the present status.

Construct validity is evaluated by investigating what qualities a test measures i.e., by determining the degree to which certain explanatory concepts or constructs account for the performance on the test. Examination of construct validity demands a combination of logical and empirical look. Construct validity is normally used when

the tester wants to increase his understanding of the logical qualities being measured by the test.

It is, however, to be taken note of that all the three concepts relating to validity are relevant and pertinent to all kinds of tests. It is the intended use of the test rather than its nature that determines what kind of evidence is required.

Criterion-related validity is made use of in the intelligence and scholastic aptitude tests to show the efficacy in predicting the degree of academic success. On the other hand, content validity is more significant in the achievement tests as already mentioned.

Reliability refers to the accuracy of measurement by a test. The degree of reliability can be determined by making use of parallel or identical tests on the same subjects. If the outcomes of parallel and identical tests are equal, identical or comparable, the existence of reliability can be ensured. In the absence of such equivalence between the outcomes of parallel tests, the reliability becomes questionable and consequently, the tests may have to be subjected to thorough revision.

Reliability is normally expressed by means of statistical methods involving correlation coefficient which is called 'reliability coefficient' or by means of 'standard error of measurement'.

12.4 Certain Considerations in Assembling and Scoring Tests

The task of test construction is very complex wherein high precision and care have to be exercised to make a good test. In the context of objective test, preparation of test items itself is a complex and difficult job. Added to this, assembling a good test is also complex and difficult. There are certain considerations that have to be taken care of in the context of assembling the test and scoring them. These considerations include the time factor, grading accuracy and time, test instructions, layout and format. It is not infrequent to notice that the test takers complain of shortage of time to complete the tests. This is a natural characteristic from such students who feel

and believe that a little more time would have been helpful. Their convictions and beliefs include that their responses could be correct in respect of many more questions and the manner of answering in case of subjective questions would be qualitatively better and that they could attempt many more questions particularly in the context of objective tests. 1t may be recalled that objective tests were categorized as power tests and speed tests. The feeling of shortage of time arises more in the context of speed tests. The reason is that in the speed test, the examinees are expected to answer as many questions as possible within a fixed period of time. It may also be recalled that in speed test, the difficulty level of the items will be the same. In the context of subjective tests where free expression is involved, most often than not, the examinees complain that the shortage of time resulted in their not being able to properly consolidate and express their ideas in a more logical and systematic manner. Further, many examinees normally complain that they had to scribble through in the exams, as a result of which either the aesthetics of their handwriting is adversely effected or at times their writing tended to be illegible. This being the case, in order to minimize the problems relating to inadequacy of time, it is necessary and desirable that the testers decide upon the time factor based on the time consumed by comparable group of examinees by the administration of a parallel test. At the same time, the distribution of the examinees must include highly intelligent, average intelligent and poor performers as well.

In the context of language tests, the complaint regarding the inadequacy of time relates to the examinees' feeling that the time factor did not permit them to make use of their entire linguistic knowledge. It means that for want of time, the examinees could not make use of the appropriate linguistic structures and vocabulary which resulted in poor performance. The remedy for this problem lies only in administering a parallel test to a comparable group of audience as a pilot measure and then deciding upon the duration of the test.

12.5 Grading Accuracy and Time

Scoring and grading has been dealt with in detail earlier. The scoring accuracy can be enhanced by increasing the number of tests and taking an average of the performances in different tests. At the same time, in the context of objective tests, it is desirable to include more number of limited response items compared to multiple-choice items as the later are less reliable than the former. The reason for the lesser reliability of the multiple-choice items is the opportunity for guessing that is available to the examinees. From the practical point of view, it may not be possible to administer more number of tests for the same purpose during the same time. Discussion regarding the accuracy and higher reliability of scoring and grading has been made in depth in the earlier chapters wherein various methods of assessment have been discussed. It may be recalled that for a more comprehensive, reliable, valid and dependable assessment, it is necessary to consider the outcomes of not only the formal tests but added to these, factors like classroom observation, portfolios and conferences, journals, questionnaires and interviews have also to be considered. It is only then that a more reliable and authentic assessment, measurement and evaluation can be expected.

12.6 Test Instructions

Providing suitable and relevant instructions for the examinees is as important as the test items are. The instructions must be very precise, clear, unambiguous and should not mislead the examinees. It is natural that different types of test items are contained particularly in an objective test. Therefore, added to the clear-cut instructions as to what is expected from the examinees etc., it is necessary and desirable to provide one example for each type of test items. Further, in the context of any objective test other than speed tests, the test maker has to necessarily ensure that the items included in the tests are graded in order of increasing difficulty. The reason for this is to make the test takers as comfortable as possible at the beginning itself in taking the test. This could automatically be ensured by making it

easier for the test takers to answer the questions at the beginning with relative ease and comfort.

Provision of instructions is equally important in the case of subjective tests also. It is generally noticed that instructions contained in subjective tests happen to be very general and unclear. To the extent possible in this case also, it is necessary to make the instructions explicit and clear. For instance, instead giving a general instruction like answer the following questions each not exceeding 350 words', it is necessary to briefly include what is expected of in such answers like 'examine the theme with reference to the viewpoints of several scholars and provide a convincing conclusion giving necessary supporting arguments thereof', etc., will be necessary. In the case of subject tests, it is desirable to provide a choice in respect of each question. Such choice could be built into each question or could be generally given by asking the examinees to answer three questions out of five or the like. In addition, the assessment and measurement of subjective tests has to be made by a group of examiners and an average of their assessments taken as discussed in detail earlier. Depending upon the theme, content and the nature of subject to be dealt with, it is also necessary to divide the test into two or three groups or parts.

12.7 Layout and Format

Formatting and layout is very important in the case of objective tests. Just as clear-cut, precise and unambiguous instructions provide a helping hand to the examinees to demonstrate their attainment of the instructional objectives, the layout would also provide such help. While the instructions and examples are directed towards preventing wrong guesses about the task requirements, on the other hand, the layout and format are directed towards removing restrictions.

One of the most frequently encountered problems in the teacher-made objective tests is a highly crowded page which is difficult to decipher and read. Such crowding of a page in a question paper could result in the test takers skipping lines in reading which could result in omitting of questions etc. In order to ensure such an

untoward thing does not happen, it is necessary to place the lines in the question paper in such a manner that the reading for the examinees would he smooth going and would not result in skipping lines. Similarly, appropriate spacing between one item and the other is also equally important. Similarly, the spacing between the instructions, the examples and the items is also a matter of serious concern. Another significant point with reference to the layout of the objective tests is that items similar in nature should be clubbed together and should be placed under the same set of instructions and example provided. As already stated, the test items have to be arranged in order of increasing difficulty in the case of power tests and all kinds of tests other than speed tests.

Genese and Upshur have provided certain suggestions/tips with reference to scoring of open-ended/subjective tests. They can be briefly mentioned as follows:

- The factors in qualities of good answers in general be identified and the examinees/students informed in detail about these. In addition to what is expected from the examinees, a brief description of the criteria can be included in the instructions which become a very sound base while assessing and scoring the test.
- The methodology of scoring has to be worked out fairly ahead. The examiners and decision makers have to decide whether the scoring method is holistic in nature or analytic. While holistic method of scoring helps in consolidating the performance, the analytic method of scoring helps in the detailed examination of the examinees' responses. In fact, the analytical scoring is more relevant in the context of subjective tests.
- The valuation/assessment of each of the answers will have to be made in the same serial order in which the questions have been attempted by the examinees. Such a systematic procedure is expected to make the scoring more efficient and objective as this would facilitate working with only one set of standards or scoring criteria at a time.

- In the case of subjective tests, it is desirable to re-read the answers a second time so as to make sure that the assessment is appropriate, just, convincing and reasonable. If this is done by a second examiner, it will be still better.
- As is clearly done in the case of mass testing and public examinations, it is necessary to be ensured that the examinees do not write their name and instead write their register/code numbers assigned to them.

There are several other considerations that have to be kept in mind in the context of test making, administering, assessing, scoring, etc., out of which the considerations discussed above are more crucial and significant.

CHAPTER

13

Interpretation of Test Scores

Interpretation of test scores is as important as assembling and scoring of tests. There are two considerations based on which the test scores could be interpreted. One is the frame of reference for the scores and the types of scores that are being interpreted. The other consideration is the manner in which the tests have been constructed.

13.1 Frames of Reference for Interpreting Test Scores

One way in which the test scores can be interpreted is by knowing the performance of other learners who have taken the same test that is under consideration. The average of the performance of all the test takers is taken as the point of reference in comparison with which the others' scores are interpreted. Those who get the same as the average scores are said to be at the average level of students in that particular grade. Similarly, those who score above the average are interpreted to be better performers and those who get less are interpreted to be poor performers. This kind of interpretation is called norm-referenced interpretation. In this context, it may be mentioned that fixed and standard norms are available in the case of standardized tests used for mass administration. Fixation of such norm is based upon what and how much is expected from the

examinees. From the point of view of instruction, the norm-referenced interpretation of scores no doubt helps in making a relative assessment of the learners with reference to the others in the group. But, it does not provide any information as to what exactly the knowledge and achievement of the learners are and what is lacking in them. This being the case, the domain-referenced tests are preferred more than the norm-referenced tests in any educational system. The domain-referenced tests reflect a well-defined domain of skills or knowledge and therefore it becomes possible to clearly state whether a given score represents the achievement or certain percentage of domain. Similarly, as the objective-referenced tests are based on the instructional aims and objectives, there exists a scope for the examiner to interpret a score as achievement/acquisition of certain percentage of what is aimed at.

As there is no clear cut procedure of assessing any or some of the specific objectives, mastery tests are made use of to determine whether a fixed level of achievement has been possible or not. Mastery tests are useful for the following purposes:

- Providing feedback for general instructional planning.
- Assessing student achievement in an area in which all students are expected to perform at some minimum level such as in a screening test.
- Determining whether all students have acquired knowledge or skill that are pre-requisite to undertaking new learning.
- Individualizing instruction in which each student is assessed at the end of each unit of instruction.

In respect of this, the open-ended tests facilitate assessing the average range of achievement among the learners in addition to facilitating and differentiating among students based on their achievement. The open-ended tests are used for

- Summative evaluation of units of instruction for purposes of general instructional planning,

- Summative evaluation of overall instruction at the end of a course with no intention that the feedback provided be used to modify instruction for the current students, and .
- Summative students' achievement of general instructional objectives at the end of the year or course for purposes of assigning grades.

Mastery tests lead to interpretation in terms of mastery of instructional objectives and consequently the effectiveness of instruction. On the other hand, open-ended tests end up in interpretations in terms of range of achievement and consequently it is a relative assessment of one student or the other in comparison to all others.

13.2 Interpreting Count and Reading Scores

This method relates to the assignment of innumerable scores or grades. The merits and demerits of assigning numerical scores and grades have been discussed very much in detail in the earlier pages. In any case, it is desirable to mention that ratings or assignment of grades can be based on highly sophisticated guidelines as otherwise they will be impressionistic and subjective. The question of resorting to the assignment of grades or ranges in terms of numerical scores arises in the case of subjective questions or open-ended testing.

❖ ❖ ❖

CHAPTER

14

Innovations and Explorations in Performance Evaluation

14.1 Open-Book and Take-Home Examinations and Assignments

Education system, in most of the Western countries and in particular in the United States, has unique features which are practically non-existent in the Indian education system. It is at the higher levels of education from undergraduate level onwards that the performance evaluation has several dimensions. The most common mechanism through which the learner evaluation is made is through the periodical and term-end examinations. Besides these examinations, the students are subjected to a host of exercises. They include open-book examinations, take-home examinations, assignments and term papers. No advance notice is given to the students about the open-book examinations. They are instantaneously held in the classroom and normally the subjective questioning methodology is made use of. The students are asked to answer the questions within a stipulated duration of time by treating the themes/topics given critically and creatively. This task demands that the examinees study and discuss various viewpoints of different people about the theme/subject, assess them with convincing facts and arguments and draw one's own conclusions after critically

examining all those. Such a task naturally demands that the students/ examinees must keep themselves abreast of the day-to-day developments about topics and disciplines of their study; as otherwise the kind of the task will not be possible for them to execute. In order to successfully attempt this exercise, the students are permitted to refer to the books and related material on the topic before attempting the answers. They could refer to the books in the departmental libraries. In certain cases, the students are divided into groups and the answers to the questions in the open-book examinations are expected to be developed by the groups concerned and a relative assessment is made group-wise. An added dimension to this is that the open-book examinations are made the basis for seminars wherein either the individual students or the groups are expected to make their presentations based upon the answers arrived at. The assessment is by a panel of experts and a consolidated score of' the assessment is taken for considering the degree of success or otherwise.

14.2 Take-Home Examinations

Take-home examinations share certain features with the open-book examinations. The tasks assigned in the take-home exams are more complex, time consuming and also demand the students/ examinees to come out with their innovations, discoveries, critiques and novel ideas. Further, the task demands a very systematic response including therein the aims and objectives, the scope and limitations, the methodology and procedure adopted, conclusions etc. Thus, tasks given as take-home examinations are mini-research projects in themselves. More often than not, the teachers/counsellors provide a general idea based on which the students are called upon to develop a project by themselves and work on it. This kind of tasks involves not only the inventory measures, but also demand that a detailed library and reference work is carried out. The time given for take-home examinations normally ranges from 3 to 5 days. The written responses are again assessed by a panel of experts and an average taken as the final score.

14.3 Assignments

Assignments are given periodically during the course of instruction. They very closely resemble the take-home examinations, but differ in that the assignments are more exhaustive treatments of the subject provided by the teacher/examiner. On the other hand, scope for determining the exact topic to work on is available for the examinees in the case of take-home examinations.

In the case of non-language subjects and in particular the disciplines like natural and physical sciences, psychology etc., project works are also assigned to the students. The tenure of such project works is the total duration of the semester/teaching session. The students are attached to one member of the faculty or the other who will be the supervisor and under whose guidance the project has to be carried out.

All these are integral and compulsory aspects of education in the West. Considering the outcomes, all these would no doubt be dependable and realistic. Therefore, incorporating these aspects apart from those mechanisms without tests discussed earlier would further strengthen the evaluation system in education.

❖ ❖ ❖

APPENDIX-1

Evaluation Sheets for Posting the Data

1. Graphemic/Orthographic Variations

1.1 Graphemic variation :

1.2 Diacritics :

1.3 Punctuation :

1.4 Letter design :

1.5 Orthographic/Printing errors :

1.6 Frequency and progression :

1.7 Overall assessment :

2. Morphemic/Lexical Variations

2.1 Colloquial to cultivated :

2.2 Neologisms :

2.3 Incidence of loan words :

2.4 Random/Competing spelling :

2.5 Compound vs. analytic :

2.6 Difficult words :

2.7 Spelling mistakes :

2.8 Use of wrong words or collocations :

2.9 Inadequate gapping between words:
(failure in continuous text, or *samhita)*

3. Syntactic/Syntagmatic Systems

3.1 Grammatical deviances :

3.2 Wrong conjunctions :

3.3 Semantic deviances :

3.4 Ambiguous phrases/expressions :

3.5 Sentence types :

4. Discourse/Textual Norms

4.1 Organization of discourse :

4.2 Paragraph organization :

4.3 Participant identification :

4.4 Unity and coherence :

4.5 Stylistic appropriateness :

4.6 Ideational failure (factual error) :

Learner Factors

(a) Ability to understand

(b) Aptitude study
(To be studied in the field)

(c) Educational achievement

— Book Identification :

— Lesson No. :

— Coverage :

— Range :

— Depth :

— Mode :

— Remarks :

Instructional Factors

Book Identification

(a) Self-sufficiency:

— Is there a teacher's handbook? *YES/NO*

— Is the teacher's handbook part of the text? *YES/NO*

— Brief comments on the effectiveness of the teacher's handbook:

(b) Adequacy in terms of

— Teaching points for grammar Adequate/Inadequate

— Diversified use of vocabulary Adequate/Inadequate

— Cues for active and passive vocabulary Adequate/Inadequate

— Instruction for paraphrasing Adequate/Inadequate

— Minimal pairs for listening skill Adequate/Inadequate

— For developing reading skill Adequate/Inadequate

— Questions for reading skill Adequate/Inadequate

— For speaking standard language Adequate/Inadequate

— Progression from learning to read to reading to learn Adequate/Inadequate

— Spelling

— Use of words and collocations Adequate/Inadequate

— Idioms and phrases Adequate/Inadequate

— Precise/expansion Adequate/Inadequate

— Instructions for essay writing Adequate/Inadequate

(c) Transfer

— Learner motivated to correlate themes to daily life

— Significant points brought out

— Learners motivated to think further

— Learners led to application of values

(d) Organization of language instruction

— Leads from simple to complex, known to unknown etc.

(e) Language styles

— Styles distinctions made

— Which style is dominant (specify)?

— Is there any effort to use the two styles?

(f) Identification of the learner

— Is it possible to identify the user/learner?

— Is the language responsive to the distinction of user/learner?

Presentational Factors

Book Identification

(i) Extrinsic Factors: *YES/NO*

— Size of the book _____. Is the size of the book appropriate for this level?

— Is the cover of the book aesthetically good/satisfactory/poor? *YES/NO*

— Is the cover's durability good/ satisfactory/poor? *YES/NO*

— Is the progression in the use of type-face satisfactory? *YES/NO*

— Are the colors used in the book relevant? Mention degree *High / Average / Low*

— Is there a balance between visuals and non-visuals? Mention degree *High / Average / Low*

(ii) Intrinsic Factors

(a) Are the titles suitable? *YES/NO*

— Are the titles interesting/motivating? *YES/NO*

— Give three examples of apt titles:

(b) Does the book have pictures? *YES/NO*

— Do the pictures reflect the theme of the lesson? *YES/NO*

— Give three examples of inappropriate pictures:

(c) Is the text of appropriate length? *YES/NO*

— Is the text suitably divided into paragraphs? *YES/NO*

(d) Are there any devices for recapitulations? *YES/NO*

— Are these devices suitably spaced? *YES/NO*

— Are these devices effective? *YES/NO*

— Give examples of inappropriate use of these devices

(e) Are there exercises in the book? *YES/NO*

— Are the exercises comprehensive? *YES/NO*

(f) Are any other linguistic devices used? (Specify) *YES/NO*

— If 'YES', are these devices effective?

PROFORMAS

(i) Aptitude Study

Instructions:

Read the given lesson. Indicate, for this lesson, on the three-point scale of High-Average-Low, the degree of the following factors. *(Put a ✓* in the appropriate box).

	High	Average	Low
1. Interesting	1.		
2. Exciting	2.		
3. Provoking	3.		
4. Novel	4.		
5. Activating	5.		
6. Suitable to level	6.		
7. Compatible to experience level	7.		

(Thus, if you find the lesson very interesting, it is *High* on the scale, and this should be indicated by putting ✓ against 1. under *High*.

(ii) Instructional Factors

[To be filled in by the teacher]

Instruction:

Please tick (✓) the appropriate box for *YES* or *NO*

1. Does the material in the text bring out the significant points? *YES* ☐ *NO* ☐
2. Do the textbooks motivate the learners to think further about their own experience? *YES* ☐ *NO* ☐
3. Do the books lead the learners towards applying moral/aesthetic values to other related situations? *YES* ☐ *NO* ☐

(iii) Presentational Factors

[Response is to be .sought from children]

Instructions:

Take three different *covers* of different textbooks, actually prescribed in different schools, and show them to the children and ask:

Which of these covers (I, II, III) you find

1.	Most attractive	I ☐	II ☐	III ☐
2.	So-so	I ☐	II ☐	III ☐
3.	Bad	I ☐	II ☐	III ☐

Tool for Fieldwork

Socio-cultural Factors

(a) Questionnaire about the school, to be filled-in by the Investigator.

Questionnaire about the student, to be filled-n by the Investigator.

(b) Questionnaire about the Teacher and the text, to be filled-in by the Teacher.

Composite Tool for Learner, Instructional Factors and Presentational Factors

(i) The School

[For each School to be investigated, fill in a separate proforma]

1. What is the name of the School?
2. Write the School's complete address
3. Mention the size (number of students in each: I ☐ II ☐ III ☐ of the three Classes – I, II and IV)
4. How many sections does each class have?: I ☐ II ☐ III ☐
5. Identify the developmental stage of the city. (Delete what is not applicable).
   i) Urban/Semi-urban/Rural
   ii) State capital / Division headquarter /District headquarter /Mofussil

iii) City of

Class A ☐

Class B ☐

Class C ☐

6. What is the management pattern of the School? (Delete what is not applicable). : Private Board of Trustees/ Individual/Government/ Autonomous Trust with Government nominee.

7. Identify the type of School. (Delete what is not applicable). : Government/Private/ Unaided/Aided/Private/ Public/Semi-Government (Zilla Parishad, Municipal).

8. (a) Is the School dominated by one linguistics group? : *YES* ☐ *NO* ☐

(b) Mention the dominated linguistic group/groups :

9. (a) Which language or languages are spoken in the neighborhood? (Mention the language/ languages). :

(b) Which language is dominant? (Write the name of the dominant language).

10. (a) Is the School dominated by one or more racial/ religious group or groups? : *YES* ☐ *NO* ☐

(b) Mention the dominant racial/religious group/groups. :

(ii) The Student

[To be filled in by the investigator in consultation with the Student]

1. Name of the Student:
2. His/Her class and section:
3. Name of his/her school:
4. His/Her mother tongue/home language:
5. His/Her age and sex: Age: ____ Sex: *M / F*
6. His/Her parent's occupation:
   (Write the choice from: Business/
   Service/ Teaching/Farming/
   Unskilled labor/Menial)
7. His/Her parent's education: *Father* *Mother*
   (Write the appropriate choice from
   Illiterate/Matriculate/Graduate/
   Postgraduate/Technical education).
8. His/Her religion and caste:
9. Which language/dialect is used by
   him/her at home with:
   a) Grandparents
   b) Parents
   c) Brothers
   d) Sisters
   e) Guests
   f) Servants
   g) Neighborhood
10. Which language does he/she use at school with
    a) Teachers
    b) Classmates

c) In common gathering

d) In the playground

11. Which language is used in education as medium:

a) Medium of instruction for target language (TL)

b) Medium of instruction for subjects

c) A language for explanation in the class?

12. To which language/languages is he/she: exposed? What is the language

a) of his/her neighborhood

b) of radio/TV seen by him/her

c) of films seen by him/her

d) of newspaper received at his/her home

e) of magazines at his/her home

f) of comics he/she reads

g) of story hooks he/she reads
(Put 'X' if not applicable]

13. Note on the verbal repertoire of the: speech community
(Enumerate the range of the languages, including dialects used by the speech community, immigrant speakers, if any. Also mention the dialect areas, speakers and languages covered by the textbook).

(iii) The Teacher

[This questionnaire is to be filled in by the Teacher]

1. Name :
2. Age :
3. Sex :
4. Mother tongue:
5. Which other languages do you know and use?:
   a) Understand :
   b) Speak :
   c) Read :
   d) Write :
6. Attitude to language :

   At the primary level, language has two functions – it is used both as a medium of instruction, in general, and as a subject.

Example:

As medium of instruction at the primary level, English is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful ☐
(Not suitable at all)

Please decide which alternative (i or ii or iii) you think is right, and then put a tick (✓) in the appropriate box. Thus, for example, if you think English is *'least useful'* (iii), you put a tick (✓) in box No. iii on the right.

Please answer the following questions:

6.1 (a) As medium of instruction at the primary level - Standard Hindi is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful ☐

(b) Urduised Hindi/Haryanvi is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful (Not suitable at all) ☐

(c) English is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful (Not suitable at all) ☐

6.2 As subject language i.e., language to teach science, social studies and literature

(a) Standard Hindi is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful ☐

(b) Urduised Hindi is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful ☐

(c) English is

(i) Most useful ☐

(ii) O.K. ☐

(iii) Least useful ☐

APPENDIX-2

COLT Part B
Communicative Features

Communicative Orientation of Language Teaching Observation Scheme

School_____Date of visit _____

Teacher_____Subject_____Coder_____

I. Teacher Verbal interaction

(a) Official task

(b) Target language

L_1

L_2

(c) Information gap

(i) Giving information:

- Predict
- Unpredict

(ii) Request information:

- Pseudo request
- Genuine request

(d) Sustained speech

Minimal

Sustained

(e) Reaction to form/message

Form

Message

(f) Incorporation of student utterances

Correction

Repetition

Paraphrase

Comment

Expansion

Clarify request

Elaborate request

II. Student Verbal interaction

(a) Official task

(b) Target language

L_1

L_2

(c) Information gap

(i) Giving information:

- Predict
- Unpredict

(ii) Request information:

- Pseudo request
- Genuine request

(d) Sustained speech

Minimal

Sustained

(e) Reaction to form/message

Form

Message

(f) Incorporation of student utterances

Correction

Repetition

Paraphrase

Comment

Expansion

Clarify request

Elaborate request

Figure 1. COLT observation scheme., Part B. [Reprinted with permission of the publisher, from N. Spada and M. Frolich (1995), The communicative orientation of language teaching (COLT) observation scheme: Coding conventions and applications, *App. 2. Sydney: National Centre for English Language Teaching and Research).*

References

1. **Caster V. Good,** *Ed.* "Dictionary of Education", New York: McGraw-Hill Book Company Inc., 1959. Pp.327-338.
2. **Carter V. Good,** *Ed.* "Dictionary of Education", New York: McGraw Hill Book Company Inc., 1959, Pp.327-338.
3. **Fred Genesee and John Upshur.** "Classroom Based Evaluation in Second Language Education", Cambridge University Press, 1996, P.6.
4. **Carter V. Good,** *Ed.* "Dictionary of Education", New York: McGraw Hill Book Company Inc., 1959, Pp.38.
5. **Paul L. Dresssel.** "Measurement and Evaluation of Instructional Objectives" in Readings. in *Measurement and Evaluation,* Ed. Norman E.Gronlund, MacMillan Company, New York, 1968, Pp. 11-17.
6. **Robert L. Ebel.** "Measuring Educational Achievement", Prenticehall International Inc., 1965.
7. **Fred Genesee and John A. Upshur.** "Classroom Based Evaluation in Second Language Education", Cambridge University Press, 1996.
8. **Caren E. Johnson.** "Understanding Communication in Second Language Classrooms", Cambridge University Press, 1995.
9. **Fred Genesee and John A. Upshur.** "Classroom Based Evaluation in Second Language Education", Cambridge University Press, 1996, P.3.
10. **Fred Genesee and John A. Upshur.** "Classroom Based Evaluation in Second Language Evaluation", Cambridge University Press, 1996, P.81.

11. **A. Cohen.** "Assessing Language Ability in the Classroom", Boston, MASS: Heinly and Heinly, 1994.

12. **Fred Genesee and John A. Upshur.** "Classroom Based Evaluation in Second Language Evaluation", Cambridge University Press, 1996, P.104.

13. **Peyton J. K. and L. Reed. Dialogue Journal Writing With Non-native Speakers:** A Hand Book for Teachers, Alexandria Va: TESOL, 1990 and Staton, J.R.R.W. Shuy, Peyton J. K. and L. Reed, Dialogue Journal Communication: Classroom Linguistics, Social and Cognitive Views, Norwood, N.J., Ablex, 1997

14. **Narasimharao, KVVL.** "Evaluation in Language Education", Mysore: CIIL, 1990.

15. **Davis, Alan.** "Testing Experimental Methods", P. 49.

16. **John W. French and William B. Michael.** In *Readings in Measurement and Evaluation,* Ed. Norman E. Gronlund, 1968, P.163

❖ ❖ ❖

Bibliography

Adams, G.S. and Torgerson, T.L. Measurement and Evaluation in Education, Psychology and Guidance. New York: Holt, Rinchart & Winston. 1964.

Ahmann, J. Stanley and Glock, Marvin D. Measuring and Evaluating Educational Achievement. New Jersey: Allyn & Bacon, 1971.

Ahmann. Testing Student Achievement and Aptitudes, New Delhi: Prentice-Hall, 1965.

Airasian, P. Classroom Assessment. New York: McGraw-Hill, 1991.

Airasian, P. Performance Assessment. In P.Airasian *(Ed.),* Classroom Assessment (Pp.251-306). New York: McGraw-Hill; 1991.

Airasian, P. Grading Pupil Performance. In P. Airasian *(Ed.),* Classroom Assessment (pp.307-353). New York: McGraw-Hill, 1991.

Allen, J.P.B. and S.Pit Corder, *(Ed.),* Techniques in Applied Linguistics, The Edinburg Course in Applied Linguistics. Volume 3, Oxford University Press, London, 1975.

Allen, J.P.B. and S. Pit Corder, *(Ed.).* Testing and Experimental Methods: The Edinburgh Course in Applied Linguistics, Volume-4, Oxford University Press, London, 1977.

American Council on Education. Educational Measurement, Washington: A.C.E., 1951.

Anastasi, Anne *(Ed.).* Testing Problems in Perspective, Washington, A.C.E., 1966.

Anastasi, Anne. Psychological Testing, New York: Macmillan, 1967.

Anastasi, Anne. "Some Current Developments in the Measurement and Interpretation of Test Validity". In Proceedings of the 1963 Invitational Conference on Testing Problems. New Jersey: E.T.S., 1964.

Anderson, Scarvia B. *et al.* Encyclopedia of Educational Evaluation, San Francisco: Jossey-Bass, 1975.

Bachman, L.F. Reliability. In L.F.Bachman *(Ed.),* Fundamental Considerations in Language Testing (Pp.160-235). Oxford: Oxford University Press. (1990).

Bachman, L.F. Validation. In L.F. Bachman *(Ed.),* Fundamental Considerations in Language Testing (Pp.236-295). Oxford University Press, 1990.

Belanoff, P. and M. Dickson. Portfolio : Process and Product. Portsmouth, N.H.:Heinemann, 1991.

Berk, R.A., *(Ed.).* Performance Assessment : Methods and Applications. Baltimore, Md.: The Johns Hopkins University Press, 1986.

Black, J.L. and Firm, F.E.S. New English Language Papers, London: John Murry, 1974.

Blanton, William and Fari, Roger. Reading Tests for the Secondary Grades: A Review and Evaluation; New York: International Reading Association, 1974.

Boer, D.L.De, *(Ed.).* Reading Diagnosis and Evaluation, New York: I.R.A., 1970.

Blook (Meriel) and Others. Objective Tests in English as a Foreign Language: Pupils Book 2, London: Macmillan, 1970.

Bloom, B.S. and Others. Taxonomy of Educational Objectives: Book 1 and 2, London: Longman, 1974 and 1971.

Bormuth, John R. On the Theory of Achievement Test Items, with an Appendix on the Linguistic Bases of the Theory of Writing

Items, by Peter Menzet, Chicago, University of Chicago Press, 1970.

Bracht, Glen H. and Others, *(Ed.)*. Perspectives in Educational and Psychological Measurement, 1972.

Bradfield, James H. and Moredock, H. Stewart. Measurement and Evaluation in Education, New York: MacMillan, 1957.

Carlson, S.B. Creative Classroom Testing. Princeton, N.J.: Educational Testing Service, 1985.

Carroll, John B. Fundamental Considerations in Testing for English Language Proficiency of Foreign Students. Washington D.C.: Centre for Applied Linguistics 1961.

Carroll, John B. and Sapon, S.M. Modern Language Aptitude Test, New York: The Psychological Corporation, 1959.

Carroll, John B. "The Psychology of Language Testing" in Language Testing Symposium, London: Oxford University Press, Ed. Alan Davies, 1968.

Carroll, John B. Notes on the Measurement of Achievement in Foreign Languages, (Mimeographed) August 1954.

Carroll, J.B. Testing Communicative Performance. Oxford: Pergamon Press, 1980.

Clark, John L.D. Foreign Language Testing : Theory and Practice, Philadelphia: Centre for Curriculum Development, 1972.

Corder, S. Pit. Introducing Applied Linguistics, Penguin Edition, U.S.A., 1973.

Cripwell, Kenneth R. What is a Cloze Test? How do I use it?, University of London, (Mimeographed) 1976.

Cronbach, Lee J. Essentials of Psychological Testing, New York; Harper & Two, 1970.

Dandekar, W.N. Techniques of Evaluation, Kolhapur : School and College Book Staff.

Daniels, J.C. and Diack, Hunter. Standard Reading Tests, London: Chatto & Windus,1958.

Davies, Alan. "Language Proficiency Testing and Syllabus", In Mr C.O. Bren (Ed)., ATESOL Testing in School Language Teaching: New Dimensions. Dublin: Dublin University Press, 1973.

Davies, Alan (Ed). Language Testing Symposium: Psycholinguistic Approach. London: O U P., 1970.

Davies, Alan. Testing and Experimental Methods.

Dave, R.H. and Patel, P.M. Educational Evaluation and Assessment, New Delhi: N.C.E.R.T., 1972.

deJong, J.H.A.L. and D.K.Stevenson. Individualizing the Assessment of Language Abilities. Clevedon, England: Multilingual Matters, 1990.

Dyer, Henry S. "Educational Measurement, Its Nature and Its Problem", in Evaluation in Social Studies, ed. by H.P. Beig, 1965.

Ebel, Rober L. Essentials of Educational Measurement, New Jersey: Prentice Hall, 1972,

Ebel, R.L. Measuring Educational Achievement, New Delhi: Prentice Hall, 1966.

Fradd, S.H., P.L. McGee and D.K. Wilen. Instructional Assessment: An Integrative Approach to Evaluating Students. Reading, Mass.: Addison-Wesley, 1994.

Frust, E. Constructing Evaluation Instruments. New York: Longmans, Green & Co., 1958.

Gerberich, J. Raymond. Specimen Objective Test Items: Guide to Achievement Test Construction, New York: Longmans, 1956.

Gerberich, J. Raymond. Specimen Objective Test Items: Evaluation in the Modern School, New York: David Mckay Co., 1963.

Godshalk, Swine Ford and Coffman. Studies of the Question Types in the Ceeb English Composition Test as Predictors of an Essay Criterion, New Jersey: E.T.S.

Gronlund, Norman E. Readings in Measurement and Evaluation, New York: Macmillan, 1966.

Gronlund, N. Constructing Achievement Tests, New Jersey: Prentice-Hall, 1968.

Halliday, M.A.K., *et al.* Linguistic Sciences and Language Teaching, Longmans Series, London: Longman's Linguistics Library, 1968.

Hambleton, R.K. and H.Swaminathan. Criterion-referenced Testing and Measurement: A Review of Technical Issues and Development. Review of Educational Research, 48, 1-47 (1978).

Harris, David P. "Linguistics of Language Testing" in Language Testing Symposium Ed. Alan Davies, Oxford University Press, 1970.

Heaton, J.B. Writing English Language Tests: A Practical Guide for Teachers of English as a Second Language, London: Longman's Group, 1975.

Hopkins, C.D. and I.A. Richards. Instructional Objectives. In C.D. Hopkins and L.A. Richards (Eds.), Classroom Measurement and Evaluation (Pp.39-52). Itasca, Ill.: F.E. Peacock, 1990.

Hopkins, C.D. and I.A. Richards. Using Observation in the Classroom. In C.D. Hopkins and I.A. Richards (Eds.), Classroom Measurement and Evaluation (Pp.69-99). Itasca, Ill.: F.E. Peacock, 1990.

Hopkins, C.D. and I.A. Richards. Marks and marking plans. In C.D. Hopkins and L.A.Richards (Eds.), Classroom Measurement and Evaluation (pp.404-436). Itasca, III.: F.E. peacock, 1990.

Ilg, F.L. and Ames, L.B. Improving the Classroom Test, New York: University Press, 1967.

Ingenkamp, Karlheinz, (Ed.). Development in Educational Testing: Proceedings of an International Conference held under the Olgis of the Padagogisches Zentrum Berlin, Vol.2. London: University Press, 1969.

Jacobs, H.L., S.A. Zinkgraf, D.K.Wormuth, V.F. Hartfiel and J.B. Hughey, Testing ESL Composition: A Practical Approach. Rowley, Mass.: Newbury House, 1981.

Jagger, A. and M.R.Smith-Burke, *(Eds).* Observing the Language Learner. Newark, Del.: International Reading Association, 1985.

Jennings, Lee B., "Classroom Translation : A Lesson Bugbear?", German Quarterly, Vol.XL, pp.518-529 (Sept. 1967).

Jones, R.L. and Spolsky B., (Ed). Testing Language Proficiency, Washington D.C., Centre for Applied Linguistics., 1975.

Jongsma, K. Questions and Answers: Portfolio Assessment. The Reading Teacher, 43,264.(1989)

Juola. A.F. Examination Skills and Techniques, Cliffs Notes, E.T.S., 1958.

Lado, Robert. Language Testing : Construction and Use of Foreign Language Tests (Teacher's Book), London: Longman's, 1967.

Larson, V.L. and N.L. Mckinley. Communicative assessment and Intervention Strategies for Adolescents. Eau Claire, Wis.: Thinking Publications, 1987.

Lazarus, M. Evaluating Educational Programs. Arlington, Va.: American Association of School Administrators, 1988.

Lewkowicz, J. and J. Moon. Evaluation: A Way of involving the Learner. In C.G. Alderson (Ed.), Evaluation (Pp.45-80). Oxford: Pergamon Press, 1985.

Lindeman, R.H. Educational Measurement, Bombay: Taraporewala, 1971. Lindvall, C.M. Testing and Evaluation, New York: Harcourt, Brace & World, 1961.

Lyons, John. An Introduction to Theoretical Linguistics, England: Penguin, 1971.

Lyons, John, *(Ed.)*. New Horizons in Linguistics, England: Penguin, 1973.

Lyons, John. and Wales Roger. Psycholinguistics Papers: The Proceeding of the 1966 Edinburgh Conference. University of Edinburgh, 1966.

Macintosh, H.G. and Morrison, H.G. Objective Testing, London: University Press. 1969.

Mackey, William F. Language Teaching Analysis, London: Longman's, Green & Co., Ltd., 1969.

Madsen, H.S. Techniques in Testing. New York: Oxford University Press, 1983.

Magnusson, David. Test Theory Reading, Massachusets: Addison Wesley, 1967.

Marshall, Habes. Essentials of Testing, London: Addison Wesley, 1972.

Menzel, Emil W. Suggestions for the use of New Type Tests in India, London: O.U.P., 1952.

Menzel, Emil W, Use of New Type Tests in India, London: O.U.P., 1961.

Menzer, F.W. Achievement Testing, Udaipur: Vidya Bhavan Society.

Murphy-O'Dwyer, I,. Diary studies as a Method for Evaluating Teacher Training. In C.J.Alderson (Ed.), Evaluation (pp,97-128). Oxford: Pergamon Press, 1945.

National Council of Educational Research and Training. Preparation and Evaluation of Text Books in English, New Delhi: NCERT, 1970.

National Council of Educational Research and Training. Preparation and Evaluation of Text Books in Mother Tongue : Principles and Procedures, New Delhi: NCERT, 1970.

Nunnaly, J.D. Educational Measurement and Evaluation, New York: McGraw-Hill, 1972.

Oller, J.W. "Close Tests of Second Language Proficiency and What They Measure", Language Learning, Vol.23, 1973.

Omaggio, A.C. Proficiency-Oriented Classroom Testing. Washington, D.C.: Center for Applied Linguistics, 1983.

Pattanayak, D.P. Aspects of Applied Linguistics, Bombay: Asia Publishing House, 1969.

Popham, W.J. Criterion-Referenced Measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Pickett, G.D. "A Comparison of Translation and Blank Filling as Testing Techniques", English Language Teaching, Vol.23 (1968).

Pilliner, A.E.G. Tests, Norm-Referenced and Criterion-Referenced, Mimeographed: Godfrey Thomson Unit for Academic Assessment, University of Edinburgh, 1973.

Pimsleur, Paul. Language Aptitude Battery, New York: Harcourt, Brace, & World, 1966.

Rea-Dickens, P. and K. Germane. Evaluation. Oxford: Oxford University Press, 1993.

Rea-Dickens, P. and K. Germaine. Evaluation. Oxford: Oxford University Press, 1992.

Rivers, Wilga. "Listening Comprehension", in Donoghue, Mildred R. (Ed.), Foreign Languages and the Schools: A Book of Readings, Dubugue, Iowa: Wm.C.Brown Company, 1967.

Robinson, Peter. "Basic Factors in the Choice, Composition and Adaptation of Second Language Test", in English Language Teaching, Vol.XVII (1967).

Schonell, Fred J. and Schonell, F. Eleanor. Diagnostic and Attainment Testing, including a Manual of Tests, Their Nature, Use, Recording and Interpretation, Edinburgh: Oliver & Boyd, 1963.

Srivastava, R.N. Evaluating Communicability in Village Settings, 2 Vols. New Delhi: UNICEF, 1978.

Stack, Edward M. Standards for Educational and Psychological Tests and Manuals, Washington: American Psychological Association, 1966.

Staton, J., R.W. Shuy, J.K. Peyton and L. Reed. Dialogue Journal Communication : Classroom, Linguistic, Social and Cognitive Views. Norwood, N.J.: Ablex, 1987.

Stiggins, R.J. and N.J. Bridgeford. The Ecology of Classroom Assessment. Journal of Educational Measurement, 22, 271-286. (1985)

Stone, Linton. Tests in English Structure and Usage, London: Macmillan, 1970.

Swineford, Godshalk and Coffman. The Measurement of Writing Ability New York: College of Entrance Examination Board, 1966.

Taylor, D. Teaching without Testing: Assessing the Complexity of Children's Literacy Learning. English Education, 22, 4-74. (1990)

Taylor, D. From a Child's Point of View: Alternate Approaches to Assessment. In J.A. Roderick (Ed.), Context-Responsive Approaches to Assessing Children's Language. Urbana, Ill.: National Council of Teachers of English. (1991)

Tierney, R.J., M.A. Carter and L.E. Desai. Portfolios and Self-Assessment by Students. In Portfolio Assessment in the Reading and Writing Classroom (Chapter 7). Norwood, Mass.: Christopher-Gordon, 1991.

Tyler, Leona E., *(Ed.).* Tests and Measurements, New Jersey: Prentice Hall, 1963.

Upshur, John A. "Language Proficiency Testing and the Contrastive Analysis Dilemma", Language Learning, 12/2, 1962.

Upshur, John A. "Productive Communicative Testing: Progress Report", in Perren, G.E. and J.L.M. Trimm (Eds.). Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics, Cambridge: Cambridge University Press, 1971.

Valdman. Trends in Languages Testing, New York: McGraw-Hill, 1966.

Vallette, Rebecca M. and Disick, Renees S. Modern Language Testing: A Hand Book, New York: Harcourt, Brace & World, 1967.

Vallette, Rebecca M. "Evaluation of Learning in a Second Language", in Bloom, Benjamin, S.J. Thomas Hastings and George T. Madams, (Eds.), Handbook of Formative and Summative Evaluation of Student Learning, New York: McGraw-Hill, 1971.

Wilkinson, A.M. and D. Atkinson. A Test of Listening Comprehension, Educational Review, 18/3, 1966.

Wood, Dorothy Atkins. Test Construction: Development and Interpretation of' Achievement Tests, Columbus Charles E. 1961.

Wrightstone, J., Wayan, F. and Others. Evaluation in Modem Education, New Delhi: Eurasia Publication, 1964.

❖ ❖ ❖

Index

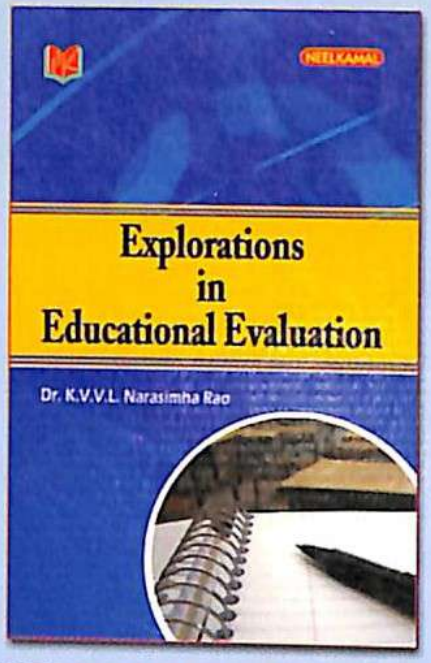
T

U

V

W

Y



ISBN: 978-81-8316-418-4

SACEM17

About the Book

Explorations in Educational Evaluation is an outcome of the author's experience in the fields of Language Education in general and Language Evaluation & Testing, Language Teaching in particular during the past 4 decades.

In spite of long years of research and innovations, in the field of Testing and Evaluation, Evaluation system in Education in the Indian context has been a matter of serious concern for Educationists, decision makers and policy formulators. Several inadequacies and lacunae exist in the existing system of evaluation. With a view to suggesting ways and means to fill up the gap to make measurement and evaluation in education and make it practical, just and realistic, the author has ventured to bring out this volume including therein aspects like performance evaluation, evaluation of methods, materials and media, programs, processes and product in the field of education in general and language education in particular. Comprehensibility of instructional material, integration of evaluation with and without tests, test construction, application and utility of questionnaires, checklists, rating scales, observation schedules in the context of Evaluation in Education are special features of the present volume. Issues, problems and lacunae relating to existing system of Education and Evaluation have been dealt with in detail suggesting therewith possible practical and reasonable solutions through innovations and explorations.

About the Author

KVVL Narsimha Rao, who holds a postgraduate degree in Applied Linguistics and a doctorate in Hindi, besides a certificate in German language has authored 24 books and published more than 70 researc papers on Linguistics & Applied Linguisti Educational Testing & Evaluation and Translat He has more than 30 years experience in Researc Guidance, teaching and directing programs in App Linguistics, Educational Testing & Evaluation, an Language Education and Educational Administration.

ISBN: 978-81-8316-418-4
9 788183 164184