ABOUT THE BOOK

Unit 1: This unit deals with a new take on Bloom's Taxonomy and Revised Bloom's Taxonomy enlightening readers about educational objectives and their importance in education.
Unit 2: This unit focusses on general concept of Measurement, Assessment and Evaluation in Education and tries to focus on their interrelationship.
Unit 3: This unit highlights the Steps, Construction and Standardisation of Achievement Tests and its importance in education.
Unit 4: This unit describes the concepts of Norm-Referenced Test and Criterion-Referenced Test and their differences in evaluation process
Unit 5: This unit emphasizes various Innovation in Evaluation in the form of Portfolio Assessment, Grading and Credit System, Open Book Exams, Question Bank
Unit 6: This unit deals with Reliability and Validity of an evaluation tool essential to judge its quality

ABOUT THE EDITORS

Dr. Sujit Samanta is B.Sc. [Chemistry (Hons.)] from University of Calcutta in the year 1997; M.Sc. (Chemistry) from University of Calcutta in the year 1999; Ph.D. from Jadavpur University in the year 2006; M.A. in Education from NSOU in the year 2009, M.Ed. from University of Calcutta in the year 2010. He has qualified both NET and GATE examination. Currently he is working in Vidyasagar Teachers' Training College, Keranitola, Midnapore, West Bengal as Associate Professor in Chemistry and has 17 years of teaching experience in Teacher Training College. He has several publications in national level journals and international level journals. He has authored book chapters in National publications in the topic of measurement and evaluation. He has written books for candidates undertaking School Service Examination. He has participated in many national and International Seminars/conferences.

Dr. Jayashri Roy is a teacher educator and author in the field of teacher education. She has done her Ph.D. in Geography from Burdwan University. She is working an Assistant Professor in Vidyasagar Teachers Training College, Medinipur, West Medinipur. She has published several papers in various journals and Books and presented many papers in national and international conferences/seminars/workshops related to geography and education . She has reviewed papers for various books and journals. She received her M.Ed. as well as M.A. in Education and M.A. in Women's and Gender Studies. She has done PG Certificate Course in Climate Change (IGNOU). She has acquired a Certificate Course on Capacity Building and Professional Development on Inclusive Education from NSOU and PG Diploma Course on Guidance and Counseling from WBUTTEPA and PG Diploma in Educational Management and Administration from IGNOU. She is a Life Member of the Indian Geographical Foundation, Calcutta; Institute for Social Development and Research, Ranchi; Geographical Development Research Institute, Gorakhpur and Indian Journal of Regional Science, Kolkata.

Dr. Ranita Banerjee is M.Sc. in Economics from University of Calcutta and she has a M.Ed. degree from University of Calcutta also. She has done her Ph.D. in Education from University of Calcutta. Currently she is working in Vidyasagar Teachers' Training College, Keranitola, Midnapore, West Bengal as Assistant Professor in Economics and has 14.5 years of teaching experience in Teacher Training Colleges. She has publications in few national level journals, UGC Care listed journals and Scopus Indexed Journals as well. She has authored book chapters in National and International Publications. She has participated in many national and International Seminars/conferences and has been in editorial board of journals. She has acted as Chairperson in NAAC sponsored seminar and reviewed papers for Scopus Indexed Journals.

KUNAL BOOKS
(Publisher & Distributors)
4648/21, First Floor, Ansari Road, Daryaganj
New Delhi-110 002 (India)
Ph.: 011-23275069, Mob.: 9811043697, 9868071411
E-mail: kunalbooks@gmail.com
Website: www.kunalbooks.com

₹1095/-

ISBN 978-81-967381-9-8

EDUCATION AND EVALUATION

Editors
Dr. Sujit Samanta
Dr. Jayashri Roy
Dr. Ranita Banerjee

KUNAL

EDUCATION AND EVALUATION

Editors
Dr. Sujit Samanta
Dr. Jayashri Roy
Dr. Ranita Banerjee

VIDYASAGAR TEACHERS' TRAINING COLLEGE
Midnapore, Paschim Medinipur
Govt. Aided: NAAC Accredited: B+
vttcollege.ac.in
[Affiliated to Baba Saheb Ambedkar Education University
(Erstwhile the West Bengal University of Teachers' Training, Education Planning and Administration)
and Recognised by N.C.T.E]

EDUCATION AND EVALUATION

EDUCATION AND EVALUATION

Editors

Dr. Sujit Samanta

Dr. Jayashri Roy

Dr. Ranita Banerjee

Patron

Dr. Monoranjan Bhowmik

KUNAL BOOKS

New Delhi-110002 (India)

KUNAL BOOKS
4648/21, 1st Floor, Ansari Road,
Daryaganj, New Delhi-110002.
Phones: 011-23275069, 9811043697
E-mail: *kunalbooks@gmail.com*
Website: *www.kunalbooks.com*

EDUCATION AND EVALUATION

First Published Oct, 2023
ISBN: 978-81-967381-9-8

Published in India by **Prem Singh Bisht** for **Kunal Books,** and printed at **Trident Enterprises**, **Noida, (U.P.).**

Patron

Dr. Monoranjan Bhowmik, Principal, Vidyasagar Teachers' Training College, Midnapore, Paschim Medinipur, WB

Editor –In –Chief

Dr. Jayanta Kumar Mete, Professor, Department of Education, University of Kalyani

Dr. Tapan Kumar Basantia, Professor, Department of Education, Central University of Jharkhand, Ranchi

Editorial Board

Dr. Sima Pal, Professor, Depart ment of Education, Assam University, Silchar

Dr. Partha Sarathi Mallik, Associate Professor and Head, School of Education, Gangadhar Meher, University, Odissa

Dr. Manisha Rani, Assistant Professor, School of Education, Mahatma Gandhi Central University, Motihari, Bihar

Editors of the Book

Dr. Sujit Samanta, Associate Professor,Vidyasagar Teachers' Training College, Midnapore, Paschim Medinipur, WB

Dr. Jayashri Roy, Assistant Professor, Vidyasagar Teachers' Training College, Midnapore, Paschim Medinipur, WB

Dr. Ranita Banerjee, Assistant Professor, Vidyasagar Teachers' Training College, Midnapore, Paschim Medinipur, WB

Vidyasagar Teachers' Training College
Recognised By NCTE Govt. Aided Affiliated to WBUTTEPA
Accredited By NAAC: Grade B+ (1st Cycle)

PRINCIPAL'S MESSAGE

VTTC is a premier teacher education institute in West Bengal whose sole mission is educational excellence. As the torchbearer of VTTC, I am devoted to keeping the flame of knowledge burning. My colleagues are also beside me in every academic endeavour with their effort to make it a success. This edited book is a feather to be added in the dazzling crown of VTTC and I wish the Editorial Board all success in undertaking such ventures in the future. This book brought together academicians from various training institutes who left a mark with their contributions. This book is a very essential addition to the existing body of knowledge, providing new visions to evaluation in the backdrop of New Education Policy 2020. I believe we are contributing significantly in the improvement of the teaching-learning process and evaluation processes in future. I am grateful to every stakeholder associated with this book and in the journey, we embarked upon.

Dr. Monoranjan Bhowmik
Principal

Vidyasagar Teachers' Training College
Recognised By NCTE Govt. Aided Affiliated to WBUTTEPA
Accredited By NAAC: Grade B+ (1st Cycle)

IQAC'S MESSAGE

"Dear Readers,

We are delighted to present this comprehensive book on education and evaluation, a crucial aspect of educational excellence. In our continuous pursuit of quality enhancement, the Internal Quality Assurance Cell (IQAC) is proud to support initiatives that promote knowledge, understanding, and innovation in education.

This book delves into the intricacies of assessment and evaluation methods, providing valuable insights for educators, researchers, and policymakers. As IQAC, we encourage the dissemination of best practices and cutting-edge research to foster a culture of continuous improvement in higher education.

We extend our gratitude to the authors for their scholarly contributions and to the readers for their interest in advancing the field of education. May this book inspire meaningful discussions, innovative approaches, and positive changes in assessment and evaluation practices.

Warm regards,

[Dr. Kishwar Badakhshan]

Co-ordinator, Internal Quality Assurance Cell (IQAC)"

PREFACE

The modern competitive educational arena demands quality of education in every stage of education. Adaptation to changes and upgradation of educational processes have become essential for improvement and progress. The education system needs to establish a strong feedback loop to ensure continuous assessment of every aspect of any program.

Measurement refers to the process by which the characteristics or dimensions of some physical object are determined. When used in the context of learning, it refers to the application of a standard scale or measuring instrument to an object, series of objects, phenomena, or conditions according to the practices adopted by persons skilled in the use of the instrument or scale. On the other hand, Evaluation is a complex and less understood term. Underlying the concept of evaluation is 'value'. It involves some process designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question.

This book is an effort to view evaluation and its tools, mechanisms, and processes in a new light. This book is important in explaining how and why evaluation in education can make a difference by finding weaknesses in every step of the education process and providing feedback. Various assessment scales are designed to test the proficiency level of learners, but choosing the correct scale for a situation and applying it properly makes all the difference.

This book is divided into five units, where the content is then presented in a simple and lucid language.

28th October, 2023

Dr. Sujit Samanta
Dr. Jayashri Roy
Dr. Ranita Banerjee

ACKNOWLEDGEMENT

We are thankful to all the members in the Editorial Board who have contributed their thoughts and criticisms in this book.

We are thankful to all the authors who have contributed their valuable ideas in completing this book.

We are thankful to the Principal of VTT College for his vision and endless motivation provided to the editorial board, which enabled us to complete this book.

We are thankful to all the faculties of Vidyasagar Teachers' Training College and the IQAC coordinator who have been with us in our journey. Thanks to the publisher, who has published this book in a very short time.

28th October, 2023

Dr. Sujit Samanta
Dr. Jayashri Roy
Dr. Ranita Banerjee

CONTENTS

Principal' s Message *vii*

IQAC' s Message *viii*

Preface *ix*

Acknowledgement *xi*

List of Contributors *xvii*

Unit :1-Bloom's Taxonomy

1. Exploring Taxonomies other than the Bloom's Taxonomy **3**
Mr. Mayukh Mazumdar

Unit 2: Measurement and Evaluation

2. Measurement and Evaluation: As Integral Part of Education **15**
Gitika Sharma

3. Importance of Educational Assessment and Measurement in Education **21**
Sonali Roy Chowdhury Ghosh

4. Evaluation: A Fundamental Parameter of Educational Institution **30**
Niladri Roy & Pinaki Barmon

5. Emergence of Integration of Technology in Teacher Education **34**
Partha Sarathi Nandi

6. Measurement, Assessment and Evaluation in Education **43**
Jaya Mandal

7. Evaluation: A Comprehensive Perspective **50**
Dr. Jayashri Roy

Unit: 3-Achievement Test

8. The Impact of Utilising the Smart Classroom Instructional Strategy on Academic Achievement **61**

Harun Al Rasid Mondal

9. Constructing and Standardizing Achievement Tests: A Comprehensive Guide **69**

Debashis Barman & Dr. Sanjib Kumar Roy

10. Steps For Construction & Standardization of Achievement Test **77**

Dr. Vandana Verma

Unit : 4-NRT and CRT

11. Norm Referenced Test and Criterion Referenced Test **91**

Bikshyatsib Sardar

12. Norm-Referenced Test and Criterion-Referenced Test **97**

Susmita Roy

Unit :5- Innovation in Evaluation

13. Unconventional Assessment: Conquering Open Book Examination **107**

Sonali Giri& Dr. Sunil Kumar Baskey

14. Effectiveness of Open Book Examination **121**

Dr. Shazia Hasnain

15. Learning to Learn: The Educational Value of Open Book Exams **128**

Dr. Tapas Karmakar

16. Qualitative Analysis on Short Answer Type Test in reference to MCQ Type Test **139**

Tania Bhadury & Prasanta Jana

17. Portfolio Assessment: An Effective Tool for Evaluating Learning **147**

Usha Rani Sahoo

18. The Impact of Continuous and Comprehensive Evaluation on Academic Achievement at Secondary Level Students in West Bengal **155**

Soma Sardar

19. Grading and Credit System: An In-Depth Analysis of Continuous Comprehensive Evaluation (CCE) in Education **163**

Asim Mahata

20. Grading and Credit system: Continuous Comprehensive Evaluation; An Overview **172**

Salu Das

21. Question Bank **181**

Dr. Ranita Banerjee

Unit :6-Reliability and Validity

22. Developing the Concept of Reliability **191**

Dr. Ratan Sarkar & Ms. Yogamaya Samanta

23. Reliability and Validity **206**

Sk Sanuar

24. Developing The Concept of Validity **213**

Dr. Ratan Sarkar & Ms. Yogamaya Samanta

25. Types of Reliability and Validity **231**

Bikshyatsib Sardar

Index ***240***

LIST OF CONTRIBUTORS

1. **Mr. Mayukh Mazumdar :** Assistant Professor, Department of Physics, Salesian College Siliguri (Autonomous**)**
2. **Harun Al Rasid Mondal :** M.Ed Student, Baba Saheb Ambedkar Education University
3. **Debashis Barman :** Research Scholar, Department of Education,Cooch Behar Panchanan Barma University.
4. **Dr. Sanjib Kumar Roy :** Professor, Department of Education,Cooch Behar Panchanan Barma University.
5. **Bikshyatsib Sardar:** PhD Scholar, Ravenshaw University, Cuttack
6. **Susmita Roy:** Former Student, Department of Education, University of Calcutta, Kolkata, West Bengal
7. **Sonali Giri :** Ph.D. Scholar, Department of Education, Rabindra Bharati University
8. **Dr. Sunil Kumar Baskey:** Assistant Professor, Department of Education Rabindra Bharati University
9. **Dr. Shazia Hasnain :** Assistant Professor, Department of Education, Aliah University, Kolkata
10. **Dr. Tapas Karmakar :** Assistant Professor, Calcutta Girls' B.T. College, Ballygunge, KolKata
11. **Partha Sarathi Nandi :** Assistant Professor, Idan Teachers Training College Paschim Nekra, Panskura, Purba Medinipur
12. **Tania Bhadury:** Research Scholar, Department of EducationRKDF University
13. **Prasanta Jana :** Assistant Professor, Subhas Chandra Basu B.Ed. Training College
14. **Gitika Sharma :** Assistant Professor, Department of Education Bodofa U.N. Brahma College, Dotma Kokrajhar Assam

15. **Usha Rani Sahoo :** Lecturer, Department of Education, Gandhi Mahavidyalaya, Rourkela, Odisha
16. **Soma Sard**ar: Assistant Professor, Department of Education, Kultali Dr. B.R. Ambedkar College
17. **Sonali Roy Chowdhury Ghosh:** Assistant Professor, Institute of Education, Haldia Purba Medinipur,West Bengal.
18. **Asim Mahata:** Assistant professor, Deshapran College of Teachers Education, Aurai, Contai, Purba Medinipur, Pin:
19. **Niladri Roy :** Asstt. Prof, B.Ed Section Shilbhadra Teachers Training Institute
20. **Pinaki Barman :** Asstt. Prof. B.Ed Section Dr. B.R. Ambedkar College
21. **Dr. Ratan Sarkar :** Head, Department of Education &Assistant Professor of Education, Department of Teachers'Training (B.Ed.), Prabhat Kumar College, Contai, Purba Medinipur, West Bengal
22. **Ms. Yogamaya Samanta :**State Aided College Teacher, Department of Teachers' Training (B.Ed.),Prabhat Kumar College, Contai, Purba Medinipur, West Bengal
23. **Sk Sanuar:** Ph.D. Research Scholar, Department of Education, Usha Martin University, Ranchi, Jharkhand, India.
24. **Jaya Mandal:** Former student, Department of Education, Calcutta University, Kolkata, West Bengal.
25. **Salu Das :** M.Ed. Student, Baba Saheb Ambedkar Education University (BSAEU) 25/2 and 25/3, Ballygunge Circular Road, Kolkata,West Bengal
26. **Dr. Vandana Verma :** Assistant Professor, Department of Education, Gindo Devi Mahila Mahavidyalya, Badaun, U.P.
27. **Dr. Ranita Banerjee :** Asst. Professor, Vidyasagar Teachers' Training College, Midnapore, Paschim Medinipur
28. **Dr. Jayashri Roy:** Asst Professor, Vidyasagar Teachers' Training College Midnapore, Paschim Medinipur

UNIT -1
Bloom's Taxonomy

1

Exploring Taxonomies other than the Bloom's Taxonomy

Mr. Mayukh Mazumdar

Abstract

The Bloom's Taxonomy and its revised version are the most widely used taxonomies of learning objectives. Since these have stood the test of time, it is a common sight that instructors often refrain from using any other taxonomy. Though there have been many other taxonomies developed in the recent past, this over-reliance on the 'classical' ones often keeps the instructors at bay from exploring the alternate ones. This article will be a modest attempt to briefly explore some of these other taxonomies and also check if some of them can complement the flaws of the Bloom's/ Revised Bloom's Taxonomy.

Keywords: *Bloom's taxonomy, revised Bloom's taxonomy, learning objectives, SOLO, new taxonomy*

Introduction

Learning/ educational objectives provide a framework for the knowledge and skills to be taught in a subject, thereby indicating things that are worth learning and type of thinking that is valued (Johnson et al., 2021). Classification of the types of objectives is an indispensable task for the teacher - it helps in framing the curriculum, setting evaluation standards, designing the lessons and in many allied aspects of education. This requires a credible framework for the categorization of the goals of any curriculum in terms of explicit and implicit cognitive abilities (Soozandehfar & Adeli, 2016) - called a Taxonomy of Learning/Educational Objectives. According to the dictionary.cambridge.org, taxonomy is a system for nomenclature and organization of things, pertinent to animals, into groups that share common qualities; it is a term borrowed from biological sciences. When used in the educational front, taxonomy of educational objectives acts as a scheme for classification of learning goals/

objectives and standards (Krathwohl, 2002). Over the years, many taxonomies have come up in the intellectual (cognitive), emotional (affective) and physical (psychomotor) domains of human development. In this article, however, only the first domain will be concerned with. The taxonomy of cognitive development that has been having a 'monopoly' over others (as far as the robustness and frequency of usage are concerned) is the Bloom's Taxonomy (BT) which was born in 1956. With very minimal yet significant modifications, a refined version of the same came up in 2001 which has been used in its place ever since. However, there have been many other taxonomies that often remain unheard and hence unexcavated due to the skewed fame of BT/RBT. In the following sections, some of these lesser-known taxonomies will be explored in brief.

A Brief Account of Cognitive Taxonomies

Benjamin Bloom and his research team in 1956 developed a framework for classification of objectives in the cognitive domain which is now widely known as the Bloom's Taxonomy (BT). The BT is hierarchical and has six-levels: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. A brief descriptor for each level is as follows:

Knowledge: recollection of discrete and specific bits of information, mainly core facts/ information in each knowledge field (Bloom et al., 1956, p.63).

Comprehension: understanding of the literatim encoded in a communication (Bloom et al., 1956, p.89).

Application: ability to use abstractions rightly for a given situation, when no mode of solution is specified (Bloom et al., 1956, p.120).

Analysis: breaking-down of a task into its constituent parts (which are organized) and finding the relationships between them (Bloom et al., 1956, p.144).

Synthesis: assembling connected parts together to form a whole - a structure that was non-existent before (Bloom et al., 1956, p.162).

Evaluation: making judgements about the values, purposes, ideas, methods, etc. by the use of criterion as yardsticks for accurate appraisal of the same (Bloom et al., 1956, 185).

BT is a cumulative succession of cognitive abilities, i.e. it underpins the idea that latter levels (say, synthesis) is a complex process involving the combination of all the subordinate levels (i.e. knowledge, comprehension, application and analysis). BT was a revolution and opened a new horizon of research in the field of educational studies; it was not long before its flaws were exposed (Amer, 2006). Some of these were even humbly accepted by

the creators themselves - an instance being the doubtful supremacy of the Synthesis level as the ultimate step of thinking or problem-solving (Bloom et al., 1956, p.185).

Many researchers have worked on modifying it since then. However, the only accepted correction to this BT is the Revised Bloom's Taxonomy (RBT) by an old student of Bloom, Dr. Lorin W Anderson and colleagues (Forehand, 2005 in Soozandehfar & Adeli, 2016) in 2001. Distinction between the content of thinking and the procedures used in problem-solving made this RBT two-dimensional from unidimensional (BT). On the surface, the *nouns* of each level of BT were replaced by *verbs*, and also the last two levels were interchanged: Knowledge>Remember, Comprehension>Understand, Application>Apply, Analysis>Analyze, Synthesis>Create and Evaluation> Evaluate (see Fig.1). Here, a learning objective comprises both of a noun/ noun phrase (subject matter content) and cognitive processes in form of a verb/ verb phrase (Krathwohl, 2002). For instance: "Students will be able to state the Law of

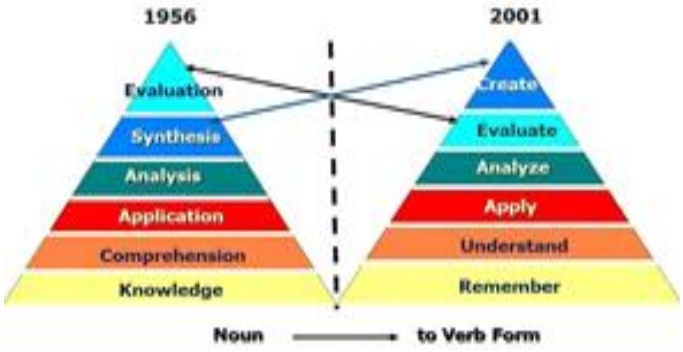


Fig.1: The (Original) Bloom's Taxonomy and its Revised Version

(https://thesecondprinciple.com/wp-content/uploads/2017/01/2001ChangesBlooms.jpg)

Conservation of Energy" is a learning objective statement containing the cognitive process (state) and the subject content (Law of Conservation of Energy). Whereas the BT emphasized on only six levels of cognitive processes, the RBT stressed more on the *19 subcategories* among six cognitive processes (Amer, 2006 & Krathwohl, 2002). However, the hierarchical nature is still persistent in this revised form.

Over time, many taxonomies have emerged as products of the diligence of cognitive psychologists, which are more relevant, research-based and accurate than the BT/RBT. Some of these are :

1. Wilson's Taxonomy for Mathematics
2. Structure of the Observed Learning Outcome (SOLO)
3. New Taxonomy by Marzano and Kendall

Each of these are hereby described with brevity.

1. Wilson's Taxonomy for Mathematics

Developed by James Wilson, this taxonomy is particularly concerned with teaching Mathematics and tackling relevant curriculum issues, effective instructional methods and evaluation, originally for secondary Mathematics (Lowry, 1973). It is often considered as an expanded model as Wilson classified the cognitive domain levels as an extension of the Bloom's taxonomy (Nayef et al., 2013). Four levels of cognitive domain objectives: Remembering, Understanding, Application and Analysis (Khuder, 1984 in Nayef et al., 2013) form the basis here.

Fig.2: The Levels of Wilson's Taxonomy

(https://www.researchgate.net/figure/Figure-3-Design-of-Wilsons-Taxonomy fig3 339427552)

The first level (Remembering) is the least level of expected skill of a student - recollection of previously learnt items (facts, terms and exercises). Second level (Understanding) is the capability to decipher ideas from verbal/ visual (symbols and signs) forms. Application is the power to choose and aptly use relevant rules, theories and principles to solve problems (familiar to what students already learned though not an exact replica). The highest level here (Analysis) embodies the last three levels of Bloom's taxonomy. Marked by novelty and creativity, the level includes discovery of mathematical experiences through solving non-familiar problems.

2.Structure of the Observed Learning Outcome (SOLO)

This was developed by Kevin Collis and John Biggs (1982) and is a framework that guides in describing the levels of ascending complexity with progress in the learning process. There are five levels in this taxonomy - Restructure (an incompetent level with unconnected bits of information), Uni-structural (understanding only with respect to only one relevant point of the task), Multiscriptual (understanding of several relevant and independent points but as a whole remains comprehended), Relational (comprehending the whole from the independent points and also appreciating their individual importance's in it) and Extended Abstract levels (connecting beyond the to-be-accomplished task and capability to transfer learning from one domain to another).

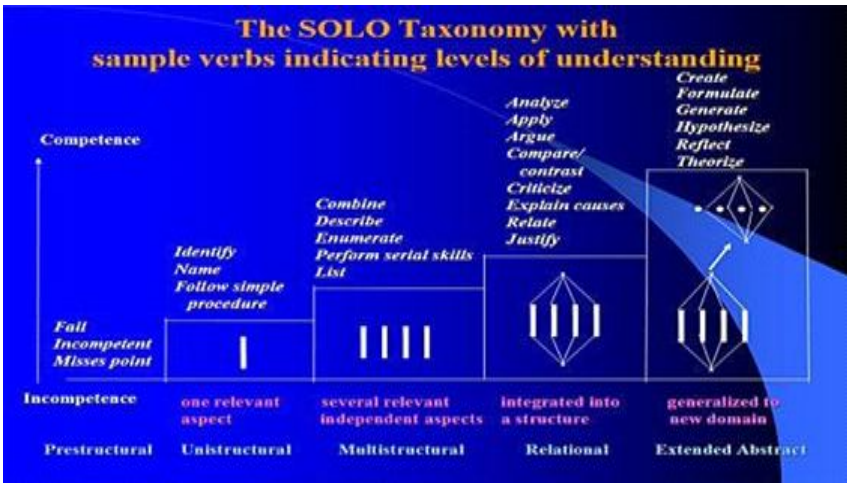


Fig.3: The Levels of SOLO Taxonomy

(https://www.johnbiggs.com.au/academic/solo-taxonomy/)

SOLO taxonomy can help frame learning objectives that guide students with unambiguous directions towards learning achievements and make lessons innovative by incorporating relevant didactic actions tuned with increased quality and complexity of thought (Somani, 2022).

3.New Taxonomy by Marzano and Kendall

This taxonomy was developed by Robert J. Marzano and John S. Kendall (2007). The taxonomy includes psychomotor procedures running parallel with mental procedures and information (Marzano & Kendall, 2007). Akin to the BT/RBT, this is also six-leveled but with three systems - the self-system, the metacognitive system and the cognitive system.

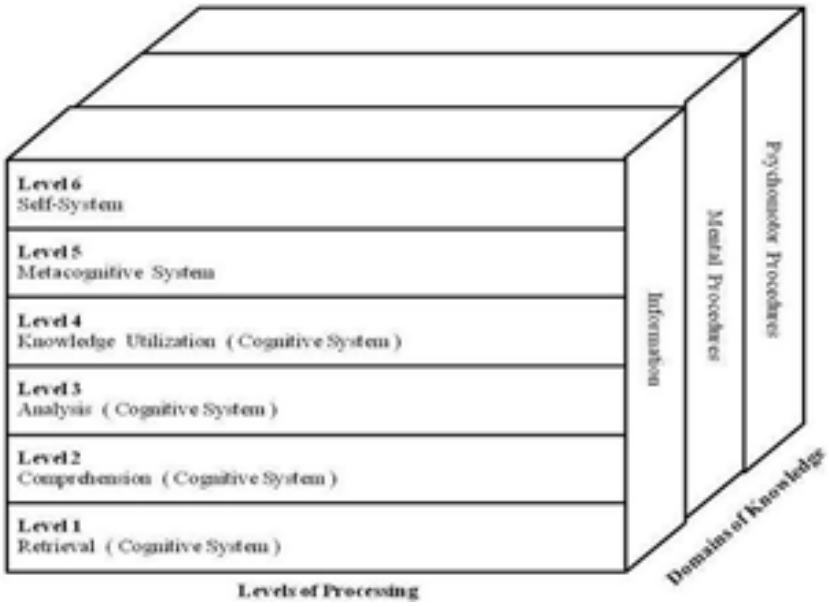


Fig.4: The Levels of the New Taxonomy

(https://www.google.com/search?client=tablet-android-xiaomi-rvo2b & amp; sca esv=559959589 & amp;hl=en& amp;sxsrf= AB5st BjUPgzl CCNwpmJfyk D0xevpfPeDEw:1692959541749 & a mp; q=marzano+new+ taxonomy & amp;tbm=isch & amp;source= lnms & amp;sa=X & amp;ved=2ahUKEwi Cmevgzfe AAxXja2w GHSlPASEQ0p QJegQIJ x AB & amp; biw = 1143 & amp; bih = 686 & amp; dpr =1.75 # imgrc=Ds DC6RVhvNJkm M & amp;imgdii=wR5xmYblEfkmmM)

Cognitive system embodies the first four levels: Retrieval (recollecting information), Comprehension (storage of knowledge in permanent memory through integration and symbolisation), Analysis (logical extension of knowledge by match-making, classification, error-analysis, etc.) and Knowledge Utilization (problem-solving, experimentations, investigations, etc.) (Johnson et al., 2021) while the remaining two systems form the fifth (Metacognitive system) and sixth (Self-system) levels.

Bloom's Taxonomy - An Overused Relic?

According to recent search results on Google Scholar (latest on 23rd August, 2023), the search term "Bloom's Taxonomy" bounteously returned 98,700 results whereas "Revised Bloom's Taxonomy" gave 41,900 results! But is something that worked in the 1950s still relevant in the current world?

A digital native will be tempted to say "No", and many educational researchers have questioned the 'omnipotence' of this taxonomy as well (following is a non-exhaustive list):

- Teachers of 2023 have better knowledge about the processes of teaching and learning, thanks to the intersections of Educational Studies with Psychology, than the 1950s. They know that teaching today has to embody various student-factors like personal beliefs, feelings, culture, environment and socio-economic background (Soozandehfar & Adeli, 2016). These were not accounted for in Bloom's taxonomy.
- The hierarchical nature of skills has researched evidence to be not always true (Marzano, 2006). For example: a student may know how to calculate divergence of a vector (applying) without knowing its physical significance (understanding).
- The revision by Anderson et al. (2001) was not very influential and just renaming of the levels and addition of some sub-levels - the taxonomy was still hierarchical.
- Not treating levels of 'Create' and 'Evaluate' on the same footing but keeping the former above the latter is unacceptable as both are equally complex and significant in problem-solving (Soozandehfar & Adeli, 2016).

Thus even though the RBT/ BT provides the teacher with practical and usable question-words, makes learning-designs streamlined and helps identify levels of cognitive complexities, it has some inherent defects and also does not comply with many demands of the modern-day pedagogies.

Potential Alternatives to BT

Having already toured some other taxonomies in the previous sections, this is the high time to test those as potential replacements of the BT/ RBT.

- Wilson's Taxonomy can be the best choice for dealing with the issues of Mathematics education as often the levels of BT can be misleading and create ambiguity in mathematical ideas (Khuder, 1984 in Nayef et al., 2013).
- SOLO Taxonomy is based on the process of understanding that learners use while responding to prompts and unlike BT integrates knowledge throughout the levels of the taxonomy (Somani, 2022).

- ❖ SOLO taxonomy is useful in the implementation of *constructive alignment* (Biggs, n.d.) where teaching and assessments of outcomes are aligned with the starting point - the intended learning outcomes. Such flexibility is missing in BT/RBT.
- ❖ Marzano and Kendall's Taxonomy is a representation of a model or a theory of human thought, unlike BT/RBT which is a framework (Marzano & Kendall, 2006, p.16). A theory or model has the ability to predict an incident but a framework necessarily does not. Hence, this new taxonomy can predict the invested attention and motivation that an individual will display if the individual's beliefs within the self-system is already understood (Marzano & Kendall, 2006, p.16).
- ❖ The higher status rendered to metacognitive and self-study systems above the cognitive system in this taxonomy enables framing strategies based on the research on the students' thought-processes. This makes it more learner-centric than BT/RBT.

Conclusion

Education will always be the perennial need for any progressing society at any timeline. For this, every component of the education system must be synchronous with the contemporary societal needs and global trends - an education system tainted with anachronic components will drive the national progress several decades backward. So, it is high time for the modern day teachers to free themselves from the shackles of antique taxonomies and not be hesitant about using research-based modern taxonomies that address the needs of the current Generation Z learners better. Selection of cognitive taxonomy must be guided by the teacher's knowledge about the following: type of target knowledge (factual/conceptual/procedural/metacognitive), complexity of the same and any discipline-specific outcome like 'historical understanding' (Weay et al., 2016). An eclectic mix of different taxonomies can also be an option worth considering.

References

Amer, A. (2006). Reflections on Bloom's revised taxonomy. *Electronic Journal of Research in Educational Psychology, 4*(1), 213-230. https://www.redalyc.org/pdf/2931/293123488010.pdf

Biggs, J. (n.d.). SOLO Taxonomy. *John Biggs.* https://www.johnbiggs.com.au/academic/solo-taxonomy/

Bloom, Benjamin S., and David R. Krathwohl. (1956). *Taxonomy of educational objectives: The classification of educational goals. Book 1, Cognitive domain.* Longmans, Green and Co. Ltd. https://repository.vnu.edu.vn/handle/VNU 123/89975

Johnson, C., Boon, H., & Dinan Thompson, M. (2022). Cognitive demands of the reformed Queensland physics, chemistry and biology syllabus: An analysis framed by the new taxonomy of educational objectives. *Research in Science Education*, *52*(5), 1603-1622. https://doi.org/10.1007/s11165-021-09988-4

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41(4)*, 212-218. https://www.tandfonline.com/doi/pdf/10.1207/s15430421tip4104_2

Lowry, William C. "9 Content Analysis of Mathematics Instructional Materials." *ASCD-17948* (1973): 65. https://files.eric.ed.gov/fulltext/ED077146.pdf#page=76

Marzano, R. J., & Kendall, J. S. (Eds.). (2006). *The new taxonomy of educational objectives*. Corwin Press.

Nayef, E. G., Yaacob, N. R. N., & Ismail, H. N. (2013). Taxonomies of educational objective domain. *International journal of academic research in business and social sciences*, *3*(9), 165. http://dx.doi.org/10.6007/IJARBSS/v3-i9/199

Somani, G. (2022, March 3). What Is SOLO Taxonomy? Importance And Levels Of SOLO Taxonomy. *MasterSoft*. https://www.iitms.co.in/blog/what-is-solo-taxonomy-important-and-levels.html

Soozandehfar, S.M.A & Adeli, M.R. (2016). A critical appraisal of Bloom's Taxonomy. *American Research Journal of English and Literature (ARJEL)*, *2*, 1-9. DOI:10.21694/2378-9026.16014

Weay, A. L., Masood, M., & Abdullah, S. H. (2016). Systematic review of revised Bloom Taxonomy, SOLO Taxonomy and Webb's Depth of Knowledge (DOK) in assessing students' historical understanding in learning history. Malaysian Journal of Higher Order Thinking Skills in Education, 3, 1-27. https://www.researchgate.net/profile/Ling-Wea Ang/publication/308982981 Malaysian Journal Of Higher Order T hinking Skills In Education S ystematic Review of Revised Bloom T axonomy SOLO Taxonomy and Webb's Depth of Knowledge DOK in Assessing Students' Historical Understandi/links/57fc864808ae6ce92eb2ba65/ andin.pdf

Yadav, R. (2023, August 4). Bloom's Taxonomy: An Outdated and Misleading Educational Relic-Yadav. *Royal Patiala*. https://royalpatiala.in/blooms-taxonomy-an-outdated-and-misleading-educational-relic-yadav/

Unit 2:
Measurement and Evaluation

2

Measurement and Evaluation: As Integral Part of Education

Gitika Sharma

Abstract

Measurement is a part and parcel of human life. From the very begining of human life measurement starts its function. Just after birth of a baby, doctors measure the height, weight, heart beat, temperature etc. to identify whether the baby is healthy or not. So, we cannot think about our life without measurement. Measurement gives description of any object in terms of numbers. It reflects the quantity aspect. It is the reliable source of evaluation. Because evaluation is done its qualitative judgement on the basis of measurement. In simple term evaluation means both quantitative and qualitative judgement regarding any object. Both measurement and evaluation contribute lots in the field of education.

Keywords: *Measurement, Evaluation, Education, Qualitative judgment, Quantitative judgment*

Introduction

Measurement is a common aspect of human life. From the very beginning of human life measurement starts its function. When human life starts to grow in mother's womb, measurement starts to influence. With the help of ultra sonography doctor measures the position of the fetus. Through the measurement doctor prescribes proper medicine, vitamins and any kind of protective measures to the pregnant mother. With the help of measurement doctor can even assume the expected delivery date of the baby also.

At present time measurement enters into all branches of science i.e. Physics, chemistry, Zoology, Biology, Social Science etc. It is also one essential part of Education and Psychology. In simple term measurement is the numerical

description of any object. It explains anything or object in terms of some unit like meter, kilometer, gram, kilogram, second, minute etc. Through measurement we can compare different objects. In the process of measurement we must have to fallow some established rules. Hence, number, unit and rule are the basic components of any measurement. That is why the definition of Campbell – "Measurement means assignment of numerals to object or events according to certain rules." becomes very popular (Bhatia, 1990). Form the different definitions put forward by various psychologists some common characteristics of measurement can be found out.-

- **Measurement is related with quantity aspects.**
- **The scope of measurement is narrow.**
- **Measurement is objective in nature.**
- **Measurement considers present situation only.**
- **Measurement is précised and scientific.**
- **Number and units are the most essential elements of measurement.**

Concept of Evaluation

Evaluation is a recent scientific concept. It is more comprehensive than measurement. In simple evaluation is a process of assessment or appraisal. It indicates both quantitative and qualitative judgment regarding any object or event. It is the act of placing value on something. James M. Bradfield said that – evaluation is the assignment of symbols to phenomenon, in order to characterise the worth or value of a phenomenon, usually with reference to some social, cultural or scientific standards (Goswami, 2016).

From the above definition it is evident that evolution is a symbolic expression of any event or object by which comparison can be done in terms of present and past. Again evaluation also shows the path far improvement. There is another one dimension of evaluation – the individual judgment. The judgment is done on the basis of some standard rules. The following characteristics of evaluation can express the concept of evaluation clearly-

- **Evaluation is both quantitative and qualitative judgment of any object.**
- **In evaluation past, present as well as future conditions are also considered.**
- **Evaluation is a continuous and comprehensive process.**
- **Evaluation is subjective in nature.**

- **The scope of evaluation is very vast.**
- **Evaluation not only determine whether some thing is good or bad, but it also gives importance on improvement of the situation.**

Relationship between measurement and evaluation

Both measurement and evaluation are closely connected. On the basis of the description given by measurement, evaluation judges its value. Proper evaluation is based on the proper measurement. For example a student's achievement in one particular year is measured with the help of different tests, like sessional test, practical examination, final examination etc. As well as his number of attendance is also counted. After that at the end of the academic session – the student's achievement is evaluated and place some grade. That means the raw informations are supplied by measurement for proper evaluation. If there is any mistake in the part of measurement, it automatically effect the evaluation process negatively. So, measurement is the reliable source of evaluation(Bhatia, 1990).

Though measurement and evaluation are closely connected, some common differences appear between them-

- **Measurement is quantitative,**

  But evaluation is both quantitative and qualitative.

- **The nature of measurement is objective. Because it based on logical number system**

But evaluation is subjective in nature. Because it is always influenced by individual judgment of the evaluator.

- **The scope of measurement is narrow.**

  But the scope of evaluation is wider than measurement.

- **The result of measurement is more accurate. Because it is done with same instruments.**

  But the result of evaluation may not be accurate due to the personal influence of the evaluator.

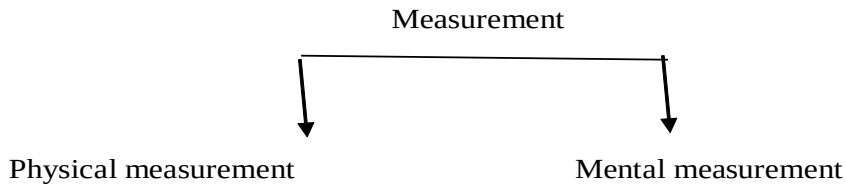
- **Measurement is scientific in nature.**

  But evaluation is philosophical in nature.

- **Measurement considers the present status of a particular event or object only.**

But evaluation is done by considering the past, present and future conditions of a particular thing or event.

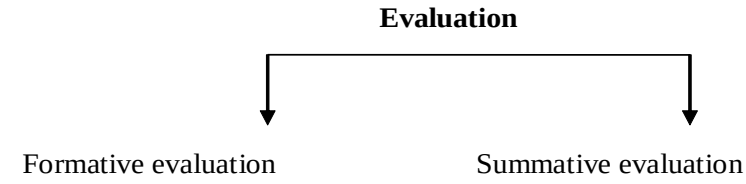
Types of Measurement and Evaluation



Physical measurement is concerned with size, height, weight, length, quantity etc. Physical measurement is done with the help of some instrument like – scale, weight, machine, compus etc. The value of such measurement is expressed with the help of some specific units. It is complete in itself and starts with a specific Zero point.

On the other hand psychological measurement is concerned with different mental processes. It can not directly measure mental characteristics of human being. It just measures a particular mental process by observing behavior of an individual in control environment. So, it is subjective in nature and has not any definite zero point. So, mental measurement is complex than physical measurement. The educational measurement is a part of mental measurement. Educational measurement is such a process through which success and failure of any education system can be measured. It indicates how much the educational goals are achieved.

Types of Evaluation



Formative evaluation is done during the class interaction. It is a part of teaching learning process. With the help of formative evaluation teacher is able to found out whether students able to understand the lesson or not. From the students feedback teacher gets opportunity to reform his teaching method if needed. For formative evaluation, teacher normally takes help of class work, question-answer, home work, unit test, assignment etc (Bhatia, 1990).

Summative evaluation is done at the end of one academic session. It is normally used to grade student's achievement and to certify them. External examination, presentation of project work, practical examination, viva-voc etc, are common means of summative evaluation.

Role of Measurement and Evaluation in Education

Both measurement and evaluation play very important role in all phases of teaching. The basic functions fulfilled by measurement and evaluation are – prediction, selection, classification, comparison, diagnosis etc. We can take help of measurement and evaluation to fulfill following needs-

- **Measurement and evaluation help teachers to understand their** students learning level. When the teacher is able to understand his student's ability, interest, attention level etc. he can adopt proper method of teaching, which leads effective teaching learning.
- Through measurement and evaluation teachers get necessary feedback, on the basis of which he can reform his instructional techniques (Asthana, 2016).
- Just as teachers measurement and evaluation also help the administrators in case of selection, classification and the grading of students. Through proper measurement and evaluation it is possible to select the right student for the right place.
- It is possible to find out whether the pre- determined objectives of education is achieved or not with the help of measurement and evaluation.
- Measurement and evaluation also help teachers to find out the strength and weakness of the students. After that teacher can arrange enrichment programme for the bright students as well as remedial classes for the needy ones.
- Measurement and evaluation is equally important for the student community also. The result of success can work as a motivational. The result of success can work as a motivational factor for all students which leads good study habit among them.
- It helps the administrators in maintaining the standard of admission, promotion and awarding of honour to the successful students.
- Any kind of reformation and reconstruction of total educative process is possible with the help of sound measurement and evaluation. Otherwise education system will be rigid, static and monotonous.

Conclusion

From the above discussion we conclude that without measurement and evaluation education will be just as a boat without sail. For effective functioning of one educational institution the measurement and evaluation process should be accurate. Through proper measurement and evaluation it possible to minimize

the rate of failure and achieve the goal of success (Asthana, 2016). Hence both measurement and evaluation can play the in-built monitor role in preparation of educational programme, assessment of its achievement level and improving its efficiency.

References

Asthana, B. (2016). Measurement and Evaluation in Psychology and Education. *Vinod Pustak mandir.*

Bhatia, K. (1990). Measurement and Evaluation in Education. *Prakash Brothers,* Ludhians.

Goswami, M. (2016). An Introduction to Educational and Evaluation. *Ashok Book Stall*, Panbazar, Guwahati- 1.

3

Importance of Educational Assessment and Measurement in Education

Sonali Roy Chowdhury Ghosh

Abstract

Measurement plays a crucial role in the learning system as it provides essential information about students' progress and understanding. It has significant administrative utility in education, surpassing evaluation in certain aspects. On the other hand, the evaluation process is invaluable in making necessary adjustments and enhancements to the learning system and instructional procedures.

Keywords: *Evaluation, Continuous Process, Remedial, Measurem*ent.

Introduction

Measurement is a vital aspect of the learning system as it plays a crucial role in assessing students' progress and understanding. Its administrative utility in education surpasses that of evaluation. Evaluation is valuable for making necessary improvements to the learning system and instructional procedures. Measurement is known for its precision and objective approach when compared to evaluation.

Evaluation

In everyday language, evaluation refers to a systematic process of determining and assessing the value, significance, and worth of a subject, using predefined criteria and standards. It is commonly used to evaluate the effectiveness of organizations, programs, designs, projects, or interventions and to aid decision-making. The main purpose of evaluation is not only to gain

insights into past or ongoing initiatives but also to facilitate reflection and identify opportunities for future improvements. Evaluations are conducted at the end of a specific period and can be applied to various domains, including the arts, criminal justice, non-profit organizations, government, healthcare, and human services (Shute, 2008).

Evaluation involves the structured interpretation and meaningful analysis of predicted or actual impacts of proposals or results. It assesses the original objectives and examines what was achieved, how it was accomplished, and whether the desired outcomes were met. Evaluations can be either formative, occurring during the development phase of a concept, project, or organization with the intention of enhancing its value and effectiveness, or summative, conducted after the completion of an action or project to draw lessons for future reference (Baker, 2008).

Since evaluation is based on theoretical frameworks, it varies according to the context, encompassing the theory, needs, purpose, and methodology of the evaluation process itself. It is essentially a critical and objective assessment of the extent to which a service or its components fulfill stated goals. Evaluations are studies that assist an audience in assessing the merit and worth of an object, service, or initiative (Cizek, 2010).

Importance of Evaluation

Evaluation offers a methodical approach for examining programs, practices, interventions, or initiatives to assess their effectiveness in achieving their objectives. Through evaluations, it becomes possible to identify successful aspects and areas that could be enhanced in a program or initiative.

The benefits of program evaluations include

- Demonstrating the impact of the program to potential funders
- Providing valuable insights for ongoing improvement efforts
- Seeking support to sustain the program
- Collecting valuable information on the approach that can be shared with others
- Helping assess whether the approach is suitable for replication in other locations with similar needs.

Educational Evaluation

Evaluation is an ongoing process that focuses more on a student's informal academic progress rather than their formal academic performance. It involves

understanding an individual's development concerning desired behavioral changes in the connection between their emotions, thoughts, and actions.

The concept of evaluation involves determining the value of something using a specific appraisal method. This blog will delve into educational evaluation, its significance, and the principles it entails (Hursh, 2005).

Educational evaluation includes standardized tests that assess a child's academic aptitude across various subjects. These assessments help identify whether a student is lagging behind consistently in all subject areas or if specific barriers hinder their performance in a particular subject.

Educational evaluators typically hold advanced degrees in education or psychology, and the evaluation process usually takes three to five hours to complete. Assessing the effectiveness of program interventions is an integral part of educational evaluation. In an educational context, these evaluations often pertain to learning areas (e.g., reading), behavioral, emotional, and social development (e.g., antibullying programs), or broader issues related to the entire school system, such as inclusive education (Boud, & Brew, 1995).

Characteristics of Educational Evaluation

1. **Continuous Process:** Educational evaluation is an ongoing and continuous process that goes hand in hand with the teaching-learning process. It involves regular assessments and feedback to monitor students' progress over time.
2. **Comprehensive:** Evaluation is comprehensive in nature as it encompasses all aspects that can be evaluated in the learning process. It assesses not only academic achievements but also social, emotional, and behavioral aspects of students.
3. **Child-Cantered:** Educational evaluation prioritizes the learning process over the teaching process. It focuses on understanding each child's individual needs, strengths, and weaknesses to facilitate personalized learning experiences.
4. **Remedial:** Evaluation provides feedback and comments on the results, which helps in identifying areas that need improvement. It is not a solution itself but rather a tool to guide remedial work and interventions to support students' learning and development.
5. **Cooperative Process:** Educational evaluation involves a cooperative effort among various stakeholders, including students, teachers, parents, and peer groups. Collaboration and communication among these parties contribute to effective evaluation practices.

6. **Effectiveness of Teaching Methods:** Evaluation assesses the effectiveness of teaching methods and instructional strategies used in the classroom. It helps identify which methods work best for different students and subject areas.
7. **Common Practice:** Educational evaluation is a common and integral practice for ensuring the proper mental and physical growth of students. It is an essential tool for tracking progress and making informed decisions in the educational setting.
8. **Multiple Aspects:** Evaluation is concerned with assessing the total personality of students. It considers various dimensions of a student's development, including cognitive, emotional, social, and physical aspects (Black, & Wiliam, 1998).

In conclusion, educational evaluation is a continuous and child-cantered process that involves multiple stakeholders and assesses various aspects of a student's growth and development. It is a fundamental practice in education, aiding in understanding student progress and tailoring educational approaches to meet individual needs effectively.

Importance of Educational Evaluation

In the teaching-learning process, educational evaluation is crucial since it serves a common goal.

1. **Diagnostic:** Educational evaluation is an ongoing and comprehensive process that helps teachers identify and understand the strengths and weaknesses of their students. Through evaluation, teachers can pinpoint problem areas and develop strategies to address them effectively.
2. **Remedial:** Once problem areas are identified through evaluation, remedial measures can be implemented to help students overcome their challenges. Teachers can provide individualized support and guidance to promote the desired changes in students' behavior and personality.
3. **Goal Clarity:** Educational evaluation plays a crucial role in clarifying the goals of education. It helps define the intended outcomes and changes in students' behavior and performance. By evaluating progress, teachers can demonstrate how learners' behavior has transformed over time.
4. **Guidance:** Effective evaluation provides teachers with valuable insights into their students' aptitude, interests, intelligence, and other relevant aspects. Armed with this information, teachers can offer personalized guidance and support to help students achieve their full potential.

5. **Classification Aid:** Evaluation allows teachers to classify students based on their abilities, intelligence, and interests. This classification assists in tailoring instructional approaches and providing appropriate challenges to students according to their individual needs.

Educational evaluation is an essential component of the teaching-learning process. It helps create a better understanding of students' strengths and weaknesses, guides instructional practices, and supports their overall growth and development. By embracing educational evaluation, educators can enhance the learning experience and foster positive changes in students' academic and personal lives.

Significance of Educational Evaluation

Educational evaluations serve the purpose of providing evidence-based insights into the potential improvement of educational outcomes through the implementation of intervention measures. Over time, the evaluation objectives have expanded alongside the parameters of educational assessment.

To better comprehend the different learning exams and evaluations available, you can identify the most suitable testing for your child and understand the reasons behind any learning difficulties or disparities they may be facing.

In certain cases, seeking assistance from a professional can be beneficial in determining whether your child requires an evaluation and the specific type of assessment that would be appropriate for them (Harlen, 2010).

Educational Measurement

Educational measurement refers to the utilization of educational assessments and the analysis of data, such as scores obtained from these assessments, to infer students' abilities and proficiencies. This field shares common ground with psychometrics. Educational measurement involves assigning numerical values to traits such as achievement, interests, attitudes, aptitudes, intelligence, and performance.

In the realm of education, when mathematical units are used to gauge a student's knowledge, it is known as educational measurement. Various aspects are measured under educational measurement, including academic achievement, intelligence, memory capabilities, personality, interests, and learning styles (Likert, 1932).

Educational measurement doesn't only pertain to students but also involves teachers, other educational staff, parents, and members of society who are connected to the educational process.

Measurement is defined as the process of assigning numerical values to objects based on accepted rules. In the context of education, it involves the teacher assigning numerical values to specific characteristics of students. Educational measurement can be described as assigning a number from a set of numbers to each element of a set of persons and things according to accepted rules.

Educational measurement holds significant importance in the field of education, providing several benefits as explained below:

1. For Students: Measurement helps students understand the extent of their academic knowledge, fostering motivation, self-confidence, and a competitive spirit. It also highlights their areas of improvement.
2. For Teachers: Educational measurement aids teachers in identifying deficiencies in the syllabus, teaching plans, and materials. This data enables them to make necessary adjustments to improve the teaching process.
3. For Parents and Family: Educational measurement informs parents about their child's educational progress, interests, abilities, and personality, allowing them to address any shortcomings effectively.
4. For Policy Makers: Data from educational measurement assists education policymakers in formulating and implementing effective education policies, including technology integration, career-oriented education, and opportunities for students.

Standard measurements, such as raw scores, standard scores, and percentile ranks, are used to determine the level of learning and the standard of the education system. The application of technology in education and its impact on social development and educational improvement are also considered within the measurement process. Overall, educational measurement is a crucial aspect of education that benefits students, teachers, parents, society, and policymakers alike.

Characteristic of Measurement

1. In educational measurement there is no absolute zero point

In educational measurement there is no absolute zero point. It is relative to some arbitrary standard. For example a student has secured 'O' in a test of mathematics. It does not mean that he has 'O' knowledge in mathematics. Because he may secured 30 in another test, which is easier than the first one. As the zero point is not fixed so we cannot say that a student with a score of '60' has doubled the knowledge of a student with a score of '30' (Ebel, 1961).

The units are not definite in educational measurement

In educational measurement the units are not definite, so we may not obtain the same value for every person. Because the test vary in their content and difficulty level. Therefore, one individual may per-form differently on different tests and different individuals may perform differently on one test.

It conveys a sense of infinity

It means we cannot measure the whole of an attribute of an individual. Generally, the scores obtained from a measurement are observed scores which contains measurement errors. So that true score is infinite and unknown.

It is a process of assigning symbols

Measurement is a process of assigning symbols to observations in some meaningful and consistent manner. In measurement generally we compare with certain standard unit or criteria which have an universal acceptability.

It cannot be measured directly

In case of educational measurement, we cannot measure for attribute directly. It is ob-served through behaviour.

It is a means to an end but not an end itself

The objective of educational measurement is not just to measure a particular attribute. Rather it is done to evaluate to what extent different objectives have been achieved.

Importance of Educational Measurement

Educational measurement is considered to be the largely need that is related to analysis of data elements, and it also focuses on conducting educational assessment of the tests. It was also identified that the major focus area is related to understanding the weak areas of the students. Measurement also helps in considering external assessment for the students, and it helps in understanding the cognitive areas of the individual. Educational measurement process is important for developing time, and it also ensures that the activities of the teaching professionals needs to be managed with data driven instructions (Biggs, 1999).

Conclusion

Measurement and evaluation are distinct processes that play complementary roles in assessing and improving performance in business and educational settings. Let’s summarize the key points:

- ❖ Measurement: It involves determining the characteristics and size of physical objects, such as speed, height, density, weight, and width, through a systematic procedure. However, measurements are limited to physical attributes and may not capture qualities that cannot be measured directly with instruments.
- ❖ Evaluation: Evaluation is a continuous process that combines quantitative and qualitative analysis based on value judgments. It goes beyond measurement and involves making judgments about the effectiveness of practices, methods, strategies, and tactics.
- ❖ Complementary Role: While measurement provides valuable data on physical attributes, evaluation complements this information by offering a broader assessment of performance, including non-physical aspects that are essential for decision-making and improvement (Shute, 2008).
- ❖ Tracking Development: Both measurement and evaluation serve as tools for monitoring and tracking the progress of individuals or groups in business or educational settings. They aid in understanding strengths, weaknesses, and areas that require improvement.
- ❖ Developing Solutions: The results obtained from both measurement and evaluation are used to develop prospective and efficient solutions to enhance key performance in business and educational establishments (Bean, 1996).

In conclusion, measurement and evaluation work together to provide a comprehensive understanding of performance in various settings. While measurement focuses on physical attributes, evaluation goes beyond to assess overall effectiveness and effectiveness of practices and strategies. Both processes are essential for making informed decisions and driving improvements in business and educational environments (Deno, 1985).

References

Baker, D. F. (2008). Peer assessment in small groups: A comparison of methods. *Journal of Management Education*, 32 (2), 183-209.

Bean, J. C. (1996). Engaging ideas: *The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. San Francisco: Jossey-Bass.

Biggs, J. (1999). *Teaching for quality learning at university: What the student does*. Bristol, PA: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, b5(1), 7-74.

Boud, D. & Brew, A. (1995). Developing a typology for learner self-assessment practices. *Research and Development in Higher Education*, 18, 130-135.

Cizek, G. J. (2010). "An Introduction to Formative Assessment." In *Handbook of Formative Assessment*, edited by H. L. Andrade and G. J. Cizek, 3–17. New York, NY: Routledge.

Deno, S. L. (1985). "Curriculum-based measurement: The emerging alternative." *Exceptional Children* 52: 219–32.

Ebel, R. (1961). "Must All Tests Be Valid?" *American Psychologist* 16 (640–647).

G. Stobart, & M. Montgomery (Eds.), *Developing teacher assessment* (pp. 29 – 52). Maidenhead: Open University Press.

Harlen, W. (2010). What is quality teacher assessment? In J. Gardner, W. Harlen, L. Hayward,

Hursh, D. (2005). "The Growth of High-Stakes Testing in the USA: Accountability, Markets, and the Decline in Educational Equality." *British Educational Research Journal* 31: 605–22.

Likert, R. (1932). "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22:5–55.

Nicol, D.J., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102-122.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35 (5), 535-550

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-b189.

4

Evaluation: A fundamental parameter of Educational Institution

Niladri Roy & Pinaki Barmon

Abstract

In education "evaluation" is a very modern concept. This process is not only modern but also continuous and comprehensive process. We all understand evaluation in education mainly as formative, summative, diagnostic, placement, etc. Even the concept is used. These concepts of evaluation are used in education to identify qualitative and quantitative aspects of students. But in educational institutions evaluation process is done not only for students to determine the quality of the institution but also to understand their position. The main objective of this paper is to discuss in detail the evaluation techniques used in the teaching-learning process in educational institutions, for students, and for the institution itself. I am mainly using an interpretive historical approach as a method based on secondary sources.

Keywords: *Evaluation, Teaching-learning process, Students, interpretive approach*

Introduction

Evaluation is one of the five key dimensions by which learning is defined. Evaluation system is one of the tools to determine what has happened in the external environment and internal environment of any educational institution and how far the results of those events affect the educational institution.

Evaluation in educational institutions mainly refers to both qualitative and quantitative evaluation. Because the evaluation system is a parameter for

determining the quality of educational institution and on the one hand, the evaluation system used in educational institutions is a standard for students, teacher and other people. In addition, this evaluation technique is a tool for determining the positional and quality standards of educational institutions.

It is true that the environment of an educational institution is an important tool for educational quality, but the evaluation system is the only tool to determine how conducive the environment is to learning (Kaur, 2013).

Objectives

The objectives of this study are as follows-

1. To emphasize the continuous and holistic evaluation system associated with the great profession of teaching.
2. To know the evaluation system linked to the traditional teaching-learning process of the educational institution.
3. To discuss the conventional critical evaluation method for the improvement of the quality of the educational institution.

Methodology-

This study is descriptive in nature. So basically, this study has been done based on secondary data Like -books, national and international journal, various articles and internet etc.

Discussion

Before the introduction of National curriculum Framework-2005, teachers were friend philosopher and guides to students. Besides, the main focal point of the education system was the teacher and the relationship between teacher and the student was that of father and son. Even the teaching-learning process in the classroom was traditional and the overall evaluation of a student was not so significant through this traditional education system. Therefore, when the National Curriculum Framework-2005 was introduced, one of the changes that took place in the teaching and learning process was the teaching method of teachers as well as the assessment system, evaluation system (Mohanty, 2002). Similarly in the evaluation system also various directions of continuous followed. In the field of evaluation, based on the peacock model, various indicators of preparatory evaluation have been proposed to be used, as well as five characteristics of periodic evaluation have also been reflected. In a word, with the aim of eliminating all the vaque ides of teacher evaluation related to students throughout the academic year, the range of free-thinking practice, the opportunity to practice various acquired skills throughout the entire academic year has been developed (Mishra, 2005).

We all know that Indian education system mainly prefers offline mode. That is why in the traditional teaching and learning process in the classroom, there is a face-to-face interaction between the teacher and student. Just as there are different aspects of teaching in the classroom, teachers can observe the students using different evaluation criteria in terms of different characteristics. A teacher is an assessment manager inside and outside the classroom, ha can observe and measure overall aspects of instruction, teacher interaction with students, classroom management etc. using evaluation criteria (Barua, 1971).

It is even true that evaluation is most powerful in the hands of the teacher to improve learning, as assessment is perhaps the most overused and least attended word. Therefore, the evaluation approaches used by teacher to measure student learning as follows-

1. According to phase of instruction

Areas of Function	Name of Evaluation	Function
Before instructional process	Placement evaluation	To know entry behaviour
During instructional process	Formative Evaluation	To know mastery in content
	Diagnostic Evaluation	To solve learning difficulties
After instructional process	Summative Evaluation	To certify the learner

2. According to Nature

1. Norm-Referenced Evaluation

It is a type of evaluation in which the performance of students is measured by comparing the relation with hypothetical average student.

2. In this evaluation, the performance of a student is measured against a predetermined learning standard. Those evaluations are extensively used in higher education.

Above these assessment approaches are directly used by teachers in an institution as evaluation parameters (Aggarwal, 2005).

Evaluation is an effective tool for teacher student evaluation as well as for improving the quality of educational institutions. Because the recognition and ranking of that institution depends on the quality of education is an educational institution. The following factors are identified as criteria for evaluation of a teacher in higher education institutions (Acharya, 1984).

1. Attendance and punctuality of the teacher.
2. Subject knowledge

3. Teaching methods used in presentation of the subject.
4. Class preparation
5. Transparency
6. Efficiency
7. Accessibility

Conclusion

The main objective of teaching is to bring a desired change in the student. The desired change in students represents the purpose of teaching or end of institution, while evaluation is the process of determining the extent, to which objectives are being achieved, the effectiveness of the learning experiences. Provided in the classroom, and how well the goals of education have been accomplished.

References

Acharya, S.C. (1984). *Pre-primary and Primary Education in Tripura and Cachar*, Development and problems, Ph.D Edu., GAU.

Aggarwal, D.D. (2005). *History and Development of Elementary Education in India*, vol 1 published by M/s Sarup & Sons Ansari Rd, Darya Ganj New Delhi.

Barua, A.P. (1971). *Wastage in Sibsagar and Gulaghat Sub Divisions-A Comparative study*, SIE, Assam.

Kaur, R. (2013). *Universalisation of Elementary of Education in India*. The policy perspective Educations confab vol2, No.4 PP 205-212.

Mishra, R.C. (2005). *Education Research and Development*, Kulbhushan Nangia, A.P.H. publishing corp 5, Ansari Rd. darya Ganj, New Delhi.

Mohanty, J. (2002). *Primary Education and Elementary Education*, Deep and Deep Publications Pvt. Ltd. F-159, Rajouri Garden, New Delhi.

5

Emergence of Integration of Technology in Teacher Education

Partha Sarathi Nandi

Abstract

Recently, technology has had a profoundly positive influence on education. Students and educators alike have gained access to a variety of previously inaccessible information and services due to the growing usage of devices like laptops, tablets, and smartphones. When we say that we've integrated technology into education, we're just referring to the ways in which we've used it to improve students' educational experiences. Adaptive learning software has allowed educators to design individualised curricula for their students. Students will be more engaged and motivated to learn if they are able to personalise their education to meet their own needs. By working together, educators are better able to learn about and implement cutting-edge methods of education, and they may also save time and energy when developing their own lessons. Students who have trouble hearing or seeing are now able to fully engage in classroom discussions and activities with the help of text-to-speech software and electronic magnifiers. Technology has revolutionised the educational landscape by allowing for more adaptability and individualization of instruction, as well as the possibility of enhanced collaboration and access to a plethora of materials. The TIM contains the five interrelated features of effective learning environments, which are as follows: active, collaborative, constructive, genuine, and goal-directed. These traits are linked to the five stages of technological acceptance and implementation (initiation, adoption, adaptation, infusion, and transformation). The ISTE Standards for Students now emphasise the need to encourage originality and foster effective

communication skills among future teachers. There is now a "Bring Your Own Device" policy in the field of teacher education.

Keywords: *Integration of technology , BYOD policy , ISTE standard , Flexibility in education*

Introduction

Many people think of the 21st century as the technological age. The function of technology in modern society is crucial. It's considered crucial to any nation's economic development. Without sufficient technical infrastructure, it is hard for a modern economy to thrive. This is because technology has greatly simplified and accelerated our work. The educational sector is only one area where technological advancements have had an effect.

The field of educational technology aims to improve education by investigating how to better analyse, design, create, execute, and assess educational environments, resources, individuals, and outcomes.

Many educators are pursuing master's degrees in educational technology in order to better integrate technology into their lessons, adapt to teaching in a blended or online environment, or explore ways to increase student engagement and achievement via the use of technological tools. In addition to developing these skills, students learn how to educate their peers on the need to integrate evidence-based strategies into their teaching (Anagun, Sengul .2018).

Importance of the Research

These qualities have been increasingly sought after, and educational technology has emerged as a primary driver of transformation within the classroom. Students are more self-motivated to study as a result of the increased access to knowledge made possible by technological advancements in education (Anagun, Sengul. 2018).

Technology's fourfold function in the field of education is as a curriculum component, instructional delivery method, pedagogical assistance, and device for improving the learning process as a whole. Both in the business world and in the classroom, education plays a key role. In the former, workers are encouraged to adopt new practises through formal education or training. In the latter, the goal of instruction is to pique pupils' interest. In any scenario, kids benefit from the use of technology since it enhances their ability to comprehend and remember material.

Educators' integration of technology into the classroom is often met with success. Some may argue that, despite the many benefits of incorporating technology into the classroom, not all of the repercussions are positive. However,

if technology is introduced into the classroom as part of noticeable, testable routines, the advantages quickly exceed the disadvantages. This is because of the potential distraction posed by the endless supply of information and entertainment accessible at any given time (Rabe-Hesketh, & Everitt, 2003).

Maintaining student interest

Participation and interaction are vital to the success of any educational programme. Technology's interactive nature keeps pupils interested, whether they're working alone or in groups.

❖ **Encourages pupils to use a variety of learning strategies**

It's important to remember that each youngster is unique in how they acquire information and how quickly they retain it. With the use of modern tools, educators may better meet the diverse requirements of their students. The use of technology allows students to learn at their own speed.

❖ **Life skill development for students**

The increased use of digital resources has spawned a new form of literacy. Every aspect of Microsoft Office and Google Drive is used on a daily basis by professionals in a wide range of disciplines, whether it's a budgeting spreadsheet, a presentation deck, or an attached file in an email. The more practise students have developing these skills, the better off they will be in the long run.

❖ **The Effect of New Technologies on Learning**

Information and communication technology (ICT) in the classroom has the potential to broaden student engagement and enhance teacher effectiveness. As Tinio (2002) says, ICT play an important role in the classroom by facilitating dialogue and group work among students.:

Active learning

The development of information and communication technology (ICT) has led to the digitization of examination processes, including data collection, analysis, and reporting. Because they can choose the pace at which they learn and how their knowledge can be applied to real-world problems, students are more involved in their education when utilising ICT.

Creative Learning

To create a tangible good or reach a particular pedagogical goal, information and communication technologies (ICT) encourage the manipulation of existing information as well as the development of one's own knowledge.

Integrative learning

As opposed to the more typical classroom setting, where focus is placed on one specific area, an integrated approach to teaching and learning is fostered through the use of ICT.

Evaluative learning

Interactive learning tools like computers and the internet put students at the centre of the process and provide them with immediate feedback on their progress. Rather than relying on memory and rote learning, students might instead explore new avenues of learning supported by constructivist ideas of education made possible by ICT (Palloff, and Pratt, 2013).

Research Objectives and Methods

1. To instruct students how to effectively employ ICT in their daily lives.
2. The goal is to use technology to foster more abstract thought and originality.
3. The goal is to provide students with a meaningful educational technology experience.
4. To spread awareness about online learning materials.
5. For the purpose of introducing pupils to the world of IT.
6. The goal is to learn how to put technology to good use.

Various approaches

We discuss some of the most important new methods that educators have developed in recent years, methods that every educator in the twenty-first century should be familiar with.

- ❖ The Classroom Flipped.
- ❖ Project-based education.
- ❖ Team-Based Instruction.
- ❖ Gamification.
- ❖ Studying through doing problems
- ❖ Theorising by Design
- ❖ Educating with the Mind in mind
- ❖ New Directions in Online Education
- ❖ Machine Learning and Artificial Intelligence

➢ **Future Prospects in E-Learning**

Artificial Intelligence and Machine Learning

The impact of AI and ML on eLearning is already being seen, and this is only expected to grow in the years to come. Learners may get individualised suggestions for courses and content, as well as immediate responses and help, through AI-powered tools. Algorithms developed through machine learning may also be used to help students fill in knowledge gaps and personalise their educational experiences.

Virtual And Augmented Reality

The use of VR and AR technology is likewise poised to change the face of online education. For the purpose of training, virtual reality and augmented reality may be utilised to generate lifelike simulations and settings. Virtual reality (VR) may be used by medical students to practise difficult surgical operations, while augmented reality (AR) can be used by engineering students to visualise and shape of 3D models.

Microlearning And Mobile Learning

Microlearning and mobile learning are two eLearning concepts that have taken off in big ways. Both microlearning and mobile learning enable students to access course materials from anywhere, at any time, utilising mobile devices such as smartphones and tablets. These tendencies are expected to keep growing in popularity as students' ability to focus decreases and their mobility increases.

Game-Based Learning and Gamification

The use of game theory, gamification and game-based learning aim to make education more interesting and fun. Teachers may motivate students to finish courses and reach their learning objectives by incorporating components like points, badges, and leaderboards. To take this a step further, "game-based learning" uses real video games as the instructional medium. Particularly useful for imparting abstract ideas and communication skill.

Social Learning And Collaboration

Ultimately, the importance of social learning and cooperation in eLearning is rising rapidly. Teachers can facilitate knowledge sharing and peer-based instruction by establishing virtual communities and group-based classrooms. Soft skills, such as communication and leadership, can benefit greatly from this method.

Synchronous and asynchronous

There are two main types of e-learning: synchronous and asynchronous. All of the participants in a synchronous learning session engage in their activities at the same time. Asynchronous learning, on the other hand, is done at the learner's own speed and allows for the sharing of ideas and information without relying on the presence of other learners at the same time. The term "synchronous learning" is used to describe the process of gaining knowledge alongside another person(s) at the same time. Face-to-face communication, live online tutoring and feedback, Skype calls, and online group work in chat rooms or virtual classrooms are all great examples. Synchronised learning encourages students to broaden their perspectives by forcing them to actively listen to and learn from their classmates while they work together. Many students find that their writing abilities and internet awareness flourish when they engage in synchronous learning. Virtual operating rooms are being considered for use in medical teaching at the graduate and postgraduate levels. Students with health issues or who have childcare duties might benefit from asynchronous learning. They can get their tasks done in a less stressful setting and with more leeway in regards to when they get them done (Bang, Eunjin & Luft, Julie, 2013).

Enhancing Online Events Through Technology

1. Flexibility

Webinar and conference attendees can benefit from eLearning's accessibility by participating from any location with an internet connection. This allows people to participate in activities they otherwise wouldn't be able to since they don't have to spend the time and money travelling to a physical site.

2. Accessibility

More individuals, including those with disabilities, will be able to take part in and enjoy online events if they are equipped with closed captioning, screen reader compatibility, and other forms of assistive technology.

3. Interactivity

The high level of involvement made possible by eLearning platforms may greatly enhance the quality of webinars and conferences. Real-time question and answer sessions, polling, and group discussions are all available to attendees.

4. Record-Keeping

The content of online events may be readily captured and preserved so that attendees can go back and revisit it at their convenience. This is especially helpful for events that may have intricate or technical content.

5. Cost-Effectiveness

Since there is no need for participants to visit a physical location or pay for the rental of a specific venue, eLearning has the potential to be a more cost-effective alternative to more conventional in-person events.

Ease of access and improved content quality are two ways in which eLearning may improve the webinar and online conference experience for attendees.

Value of the ISTE Standards for education

ISTE is most recognised for developing a set of guidelines for the use of technology in the classroom, including input from educators at all levels. The intent of these norms is to establish guidance for creating the best possible technological learning environment for these people.

In terms of how technology should be used in the classroom, educators and legislators have finally reached a consensus. Newly included in "The ISTE Standards for Students" are a focus on creative innovation, the cultivation of communication skills and collaborative work projects, critical thinking that leads to problem-solving and decision-making, and the acquisition of 21st-century digital citizenship skills. This is in addition to mastering the fundamentals of PC operation (Ayebo, Abraham & Assuah, 2017).

The goal of the ISTE Standards is to help teachers improve their own digital literacy and adopt teaching strategies that make the most of technology in the classroom. Create outstanding classroom instructors and school administration. Strengthen instructional strategies and institutional procedures. Raise the bar of professionalism in order to raise the quality of education. Create a strong bridge between academic institutions, business, and the general public.

Bring Your Own Device Policy

The concept of 'Bring Your Own Device' (BYOD) provides educators and students with the opportunity to experiment with new forms of instruction and learning as technology enters the classroom at an increasing rate. Finding low-cost methods to integrate technology into courses is a constant worry for educators. In this guide, we'll go through some of the tools and strategies your school may use to begin its transition to a BYOD policy.

Enabling pupils to bring their own mobile technological gadgets to school is known as "Bring Your Own Device" (or "BYOD" for short). Those students who have their own gadgets can use them in the classroom with the teacher's approval.

Enabling pupils to bring their own mobile technological gadgets to school is known as "Bring Your Own Device" (or "BYOD" for short). Those students who have their own gadgets can use them in the classroom with the teacher's approval(Ayebo, Abraham & Assuah, 2017).

Pros of BYOD Policy

Proponents of "bring your own device" in the classroom believe it encourages students to be more active learners. The use of modern technological tools in the classroom has been shown to increase students' motivation and retention of knowledge. The financial benefits of a BYOD policy are another major factor in favour of this initiative. It's not cheap to upgrade your technology, especially when you consider that it'll probably be obsolete in a few years (Anagun, Sengul , 2018).

Cons of BYOD policy

Those who disagree with BYOD acknowledge its benefits but argue that they are outweighed by its drawbacks. Many educators worry that allowing students to bring electronic devices to class would lead to more student distraction. However, pupils who are adept with technology will find ways to bypass filters and access the banned content.

Conclusion

The current research adds student involvement with technology in the classroom to the existing notion of student engagement. Because of its widespread usage and ubiquitous influence, educational technology engagement is the focus of this article, which builds on and refines existing measures to construct a new student engagement scale. Given the widespread use of these two technologies in the classroom, it is important to compare and contrast their effects on student motivation and performance. Finally, there are many promising new advancements on the horizon for eLearning. The future of eLearning is bright, with advances in artificial intelligence (AI), virtual reality (VR), microlearning (ML), gamification (gamification), social learning (SSL), and cooperation (C). These tendencies will likely continue to increase in popularity, shaping the future of education and training in an age of separated learners and decreasing attention spans. Education benefits from technology, but it also comes with potential risks. Both educators and their students should view this as a chance to help remove the barriers that have kept many from

realising their full potential. Therefore, it is important for all nations to upgrade their educational systems to better accommodate the use of technology (Ayebo, Abraham & Assuah, 2017).

References

Anagun, Sengul S. (2018). Skills and Managing Constructivist Learning Environment. *International Journal of Instruction*, Volume 11, No. 4, p.825-840

Ayebo, Abraham & Assuah, C. (2017). Exploring Teachers' Knowledge of Classroom Management and Control. *Malaysian Journal of Learning and Instruction*: Vol. 14 No. 1: 169-185

Bang, Eunjin & Luft, Julie A. (2013). Secondary Science Teachers Use of Technology in the Classroom During their First 5 Years. *Journal of Digital Learning in Teacher Education*. Volume 29, No. 4, pp.118-126

Palloff, R. M., and Pratt, K. (2013). Lessons From the Virtual Classroom: *The Realitiesof OnlineTeaching*. sanfrancisco, CA: John Wiley & Sons.

Rabe-Hesketh, S., and Everitt, B.(2003). *Handbook of Statistical Analyses Using Stata*. London, New York:CRC Press.

6

Measurement, Assessment and Evaluation in Education

Jaya Mandal

Abstract

As preceptors come more familiar with data- driven instruction, they're making opinions about what and how to educate grounded on information gathered from their scholars. In other words, preceptors first find out what their scholars know and do not know, and also determine how to bridge that gap. This helps preceptors to guide scholars in the right direction so that the pupil's capacities can be canalized in the right direction. Measuring, assessing, and assessing mean veritably different effects, and yet utmost of my scholars were unfit to adequately explain the difference. During the process of gathering information for effective planning and instruction, the terms dimension, assessment, and evaluation are frequently used interchangeably. These words, still, have significantly different meanings. Assessment and evaluation involve all functions of education and not just testing, measuring and testing. It goes beyond these effects and although it depends on dimension, the two terms aren't synonymous. In education, dimension and assessment is an important tool to assess pupil progress. In this content, we will learn about the generalities of dimension, assessment and evaluation in education, their meaning, description, purpose, types of assessment in education etc. Although nearly related, dimension, assessment and evaluation have different meanings and delineations. We've epitomized in this composition, a brief description about each of them.

Keywords: *Measurement, Assessment, Evaluation, Education Scales, Types*

Introduction:

Quality of education has become very important in today's competitive environment. There is definitely a need to adapt to change in the educational processes in order to improve. Educational measurement requires the establishment of a strong feedback loop with evaluation being a continuous process and not just left until the end of the program of study. Measurement refers to the process by which the attributes or dimensions of some physical object are determined. When used in the context of learning, it would refer to applying a standard scale or measuring device to an object, series of objects, events or conditions, according to practices accepted by those who are skilled in the use of the device or scale. On the other hand, evaluation is a complex and less understood term. Inherent in the idea of evaluation is 'value'. It involves engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question.

The process of measurement is an essential component of the learning system. Measurement has more administrative utility in education than evaluation. An evaluation process is useful in modifying and improving learning system and instructional procedure. The process of measurement is more precise and objective in approach, when compared to evaluation. Measurement is always done of a quality, attribute or variable of a thing or a person. Psychologists and educationists are mainly concerned with variables and attributes. The process of measurement converts the variables into variety which is used for drawing inferences. For example, intelligence is quantified in terms of IQ and achievement variable is measured in terms of scores.

Measurement

When used in relation to education, the word measurement is seldom distinguishable from other fields. To gauge the caliber or degree of a thing, a talent, or knowledge is all that it signifies. We measure with everyday tools from the physical world like tape measures, scales, and meters. These measurement techniques are kept to a standard and provide accurate findings. They accurately gather data for administrators and educators when used properly. Educational measurement is the process of estimating students' capabilities and capacities through the use of educational exams and the analysis of data, such as scores from those assessments.

A crucial step In the research process is measurement. It alludes to the procedure of allocating numbers to occasions, things, etc. in accordance with predetermined rules. It always has something to do with particular attributes or variables of an object. In education, measurement is more administratively

useful than assessment. It is the method through which scientists classify or quantify their findings. It offers the framework for evaluation.

"Measurement is the process of assigning signs to characterize the extent of phenomena as precisely as possible" – M. Brad Field.

"Measurement is defined as the assignment of a set of numbers to each person or set of objects according to certain established rules. The set of numbers depends on the nature of the property being measured and the type of measuring instrument used" – Professor Richard. H. Leadman.

Explain different levels of measurement

The following four levels or scales are typical for measurements:

1. **Nominal Scale or Classification Scale:** This is the most basic scale of measurement and it entails classifying or categorizing things into different groups or categories by using just numbers or symbols to identify or label them. For instance, to improve class management, a large class in a school might be divided into distinct categories as A, B, C, or Red, Blue, or Pink. Please take note that these characters are merely codes and do not represent dimensions.
2. **Ordinal scales or ranking scales:** These provide us with order, but they do not indicate how much or how little two rankings differ from one another. This order cannot be altered in any way. As an illustration, we categorize students into three groups: poor, average, and intelligent. However, we are unsure of how dissimilar these three groups are. Schools frequently use this scale to assign grades I, II, and III.
3. **Interval Scale:** The interval scale incorporates the prior scale's nominal and ordinal properties as well as an arbitrary zero point. In other words, there is no real or absolute zero. We can divide the scores on this scale into equal intervals, such as scores between 0 and 5, 5 and 10, 15, 20, 25, and 30, etc. Due to the similar class sizes (classes 0–5 have a size of 5, classes 5–10 have a size of 5, classes 15–20 have a size of 5, etc.), it is also known as an equal interval scale. In education and the social sciences, all sorts of measurement are typically done by interval scales because this measure has no absolute zero point and the current zero value is arbitrary. stats like mean, standard deviation, and product moment correlation, t-test and f-test can be used in interval scale.
4. **Ratio scale:** This is the greatest level of measurement and includes an absolute or true zero point in addition to the characteristics of nominal, ordinal, and interval scales. The ratio between any two scale

points on this scale is independent of the measuring unit. Scales that measure ratios include those for height, weight, kilogram, length, and vision. The behavioral sciences employ it less frequently than the physical sciences.

Assessment

Assessment is the process of gathering information on how well a student, school, classroom, or academic system performed in relation to a given set of standards, learning objectives, or curriculum. It includes information on student performance that is both quantitative and qualitative. Assessment is one of the main measurement methods in education. Tests, interviews, and behavior observation are among methods used by teachers to gather information. To ensure the validity and reliability of an assessment, it should be properly created and carried out. In other words, an assessment ought to be reliable and measure what it says it would measure (McMillan, 2004).

According to Brown (1990), the term "assessment" refers to a connected series that is used to identify a complicated trait of a person or group of people. This entails acquiring and analyzing data regarding the student's level of learning objective achievement. Additionally, assessments are used to pinpoint the problems and strengths of certain students so that teachers can offer individualized academic support, instructional programming, or social services.

Assessment can be defined as "a systematic approach to gathering information that can be used to make inferences about characteristics of people or things (AERA, et. al., 1999)". In the definition of evaluation, "a process of gathering evidence and making judgments about the results" is used. It is argued that measurement has a narrower definition than evaluation, but a wider definition than evaluation. The word "evaluation" originally meant "to sit by" or "to assist the judge," which illustrates that evaluation is a process of acquiring data and translating them into an understandable form; judging can be done on the basis of this assessment.

Four fundamental components make up the evaluation process: 1) Tracking advancement over time.2) To spur students on to study. 3) Evaluation of instructional strategies. 4) Student skill ranking in relation to full group evaluation.

Assessment, which is sometimes divided into three categories—assessment of learning, assessment for learning, and assessment as learning—is thought of as a component of the teaching-learning process. Let's go over each section in depth to help you grasp this.

Assessment of learning: It primarily focuses on students' performance in comparison to specified goals and standards. Summative assessment is another name for this. These tests are typically given by teachers to evaluate or rank their pupils at the conclusion of a unit, term, or semester. used to create learning pathways and future goals for students.

Assessment for learning: During the process of teaching and learning, teachers practice. Its primary goal is to facilitate students' learning and improve education. It happens during the classroom teaching and learning process. It is also known as "formative assessment" and is more common and typically unstructured.

Assessment as Learning: It generally focuses on how well students perform in relation to set objectives and benchmarks. Another word for this is summative evaluation. Typically, teachers use these assessments to rank or grade their students at the end of a unit, term, or semester. utilized to develop learning paths and long-term objectives for students (Agwagah, 1997).

Evaluation

It's critical to create valid and trustworthy tests that accurately quantify educational data. To effectively use data for education, however, evaluation of the obtained data is just as crucial.

Let's talk about evaluation, a notion that encompasses both measurement and evaluation. Evaluation is "a methodical procedure of gathering and analyzing data to ascertain whether goals have been, or are being, achieved" (Gay, 1991). Determination is then made as a result. Considering that "the purpose of assessment is to make a judgment about the quality or value of something," Ebel and Frisbie (1991) state that as a teacher, you should be aware of this. To determine a student's level of performance or achievement, assessments are used. Performance, however, is the main emphasis of your assessment. A performance's caliber or level. Evaluation could be characterized as a process of value judgment. It can also be used to describe the end result or product of a procedure. The phrase "measurement and evaluation are the means and evaluation is the end" might be used.

Using the data gathered from assessments is the process of assessment in education. The link between what was taught and what was intended for learning is assessed by teachers using this data. They assess the data to determine what pupils know and understand, how quickly and far they have advanced, and to compare their performance to that of other students (Darling-Hammond, 2006).

Characteristics of Evaluation

Every time you participate in an assessment process as a teacher, you should make sure it meets the following criteria, which are frequently referred to as "elements of a good evaluation".

Validity: A valid evaluation measures the behavior that is the subject of the objective(s) under consideration, testing what is intended to be tested.

Reliability: The constancy with which a question, test, or examination yields the same result under various but comparable circumstances is known as reliability. A trustworthy evaluation process independent of the evaluator's personal traits.

Practicability: Evaluation procedures should be reasonable, workable, and effective in terms of their price, duration, and simplicity of use.

Fairness: All students must receive fair evaluations. Accurately portraying the spectrum of desired behaviors as outlined by the course objectives will enable this.

Usefulness: Evaluation need to be beneficial to all students. Learners must be given access to evaluation feedback so they can demonstrate their existing strengths and deficiencies.

Objectives of Educational Assessment

The following are the goals of educational evaluation, according to Professor NM Downey.

Supplying data for grading, parent reporting, and student promotion.

- Assessing a single teaching strategy's efficacy or determining the relative merits of other approaches.
- To inspire learners.
- Choose students.
- Assessing the entire educational setting and outlining ways to enhance its many components.
- Collecting information for efficient educational and career advising.

Conclusion

Measurement and evaluation as veritable tools in driving the teaching-learning process towards the attainment of educational goals have given educators wider opportunities in improving students' performances and achievements. The cumulative result of the teaching-learning process or classroom activities is obtained through the instrumentality of measurement

and evaluation. The paper concludes that without measurement and evaluation, students' progress, achievement, abilities, aptitudes, competences, skills and efforts cannot be ascertained. Also, measurement and evaluation is not only relevant to the teaching-learning process but also serves administrative, guidance, research functions and other functions.

References

Agwagah, U.N.V. (1997). *Educational measurement and evaluation for colleges and universities.* Onisha: Cae publishers International Limited

Brown, D. H. (1990). *Language assessment: Principles and classroom practices.* London: Longman

Darling-Hammond, L. (2006). *Assessing teacher education: The usefulness of multiple measures for assessing program outcomes*. Journal of Teacher Education, 57(2), 120-138.7

McMillan, J. H. (2004). *Classroom assessment: Principles and practice for effective instruction*. New York, NY: Pearson Education.

7

Evaluation: A Comprehensive Perspective

Dr. Jayashri Roy

Abstract

Evaluation is an essential component of every teaching and learning program. Teaching and evaluation are not mutually exclusive. Indeed, teaching and learning are not possible without assessment. Testing is only a method of obtaining information about student behavior. The evaluation process is continuous and continuous. It helps students improve judgment, academic performance or achievement. Decisions should be made in all aspects of educational activities, so some assessment of teaching and learning abilities is important. Here are some ideas about assessment.

Keywords: *Evaluation, students, performance, teaching and learning*

Introduction

The process of evaluation happens in every aspect of life in one way or another. It is possible to distinguish between good and bad only through evaluation (Zimmerman, 2008 P.166-183). The evaluation process is very important for the general progress of the society. Assessment is the only way to determine a student's academic performance (Clark, 2012 p. 205–249). Therefore, goals and evaluation are closely related. When it comes to developing human resources, skills, motivation and knowledge, education is seen as an investment in people. Assessment contributes to curriculum development and enhances its effectiveness. An essential component of any learning program is assessment (Wiliam, 2011p.3-14). It enhances learning and instruction for both teachers and students (Andrade, H., & Brookhart, 2020 p. 357-372).

Evaluation is a constant process that never ends. It raises the standard of achievement and judgment among students. Since decisions are made in all areas of education, some assessment of teaching and learning abilities is necessary. (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46). It helps predict future performance in specific activities and provides suggestions for educational and professional choices. Evaluation is characterized as a broader term than measurement. It gives value judgment of numerical value (Clark, 2012 p. 205–249). It includes both tangible and intangible qualities. A student's overall personality development is evaluated as part of the comprehensive evaluation. It assesses students' growth in both academic and co-curricular domains.

Need and Importance of Evaluation

- Parents want to know how their children are doing academically so t he only way to determine the progress of students is through assessment (Crooks, 1988 p. 438-481).
- It helps in development of skills and abilities, motivation, good study habits, awareness of progress and timely feedback to students (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46).
- It helps in selection of students for particular courses, various professions and higher education (Wiliam, 2011p.3-14).
- A formative assessment helps in the selection and development of instructional strategies, protocols and techniques by educators (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46).
- Assessment helps in developing a better curriculum (Wiliam, 2011P.3-14).
- Enhances the skills of teachers and teaching techniques.
- Evaluation is able to identify deficiencies in students and school programs (Clark, 2012 p. 205–249).
- Through assessment it is possible to find out the potential skills and abilities of the students (Crooks, 1988 p. 438-481).
- Assessment helps classify students into different privileged groups.
- Assists in evaluating educators' skills to provide relevant learning experiences.
- Assessment helps in choosing a course of study and career (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46).

- Helps a student understand their learning rate and areas for improvement.
- In addition to students, parents, teachers and administrators can all benefit from the assessment process (Andrade, H., & Brookhart, 2020 p. 357-372).

Some key elements of Evaluation

- Systematic collection of evidence
- Its interpretation
- Judgement of value

Steps of Evaluation

- ***Identifying and defining the clear objectives***

Choosing what to assess or deciding on learning objectives is the first step in the assessment process. For example, what skills should a student acquire during a year of English study? When a learner learns his mother tongue, what kind of understanding must be developed? Until teachers identify and describe objectives, these questions will remain unanswered. (Clark, 2012 p. 205–249).

- ***Identifying and defining specific goals***

Learning is defined as the change in a desired direction of behavior. The teacher's first and foremost concern is what kind of learning outcomes can be expected from a student once he or she goes through the teaching-learning process. This specific goal will provide the direction of the teaching-learning process (Zimmerman, 2008 P.166-183).Evaluation planning and implementation should include determining evaluation feasibility, identifying stakeholders, and setting short- and long-term goals.

- ***Choose the appropriate assessment method***

Selecting the appropriate assessment technique is crucial. Just as a painter chooses the perfect brush for each stroke, so we must choose the right evaluation method.

- ***Appropriate educational programs should be planned***

The teacher must plan the learning activities of the students in the fourth step while considering the learning objectives.

- ***Assessment***

The teacher observes and evaluates the behavior of his students through tests in the fifth step. At this stage the evaluation process gains an additional dimension (Baas, Castilians, Vermeulen, Martens, & Segers, 2015 p. 33-46).

Here, the teacher will test the knowledge and skills the students have already acquired from their Syllabus in the classroom. The teacher may arrange practical tests, oral or written tests or both.

❖ ***Uses the result as feedback***

The last but not least, Important step in the evaluation process is using the results as feedback. If after testing his students, the teacher finds that the objectives have not been realized to a large extent so he will use the results to revise the objectives and organize the learning activities (Crooks, 1988 p. 438-481).

Education has degraded due to multiple reasons like

- Lack of proper infrastructure
- Poor pedagogy skills
- Lack of trained teachers
- Using theoretical approaches in imparting knowledge
- Poor assessment of the child's learning
- High competition among the schools (Wiliam, 2011, p.3-14)

Types of Evaluation:

Source: Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46; Gielen, Dochy, Onghena, Struyven, & Smeets, 2011 p. 729-735)

A. According to functions:

Placement Evaluation

A student's entry behavior is evaluated in accordance with the Placement Assessment Guidelines.

When admitting students to new courses of study, these evaluations are useful (Gielen, Dochy, Onghena, Struyven, & Smeets, 2011, p. 729-735).

- test of intelligence
- Observational Methods
- Exam for Entrance (Wiliam, 2011, p.3-14)

Formative Evaluation

Throughout the classroom formative assessment is used to monitor student learning progress. Its primary goal is to inform educators and students about their achievements and shortcomings in the classroom (Andrade, H., & Brookhart, 2020 p. 357-372).

A teacher can use formative assessment to regularly assess student progress. The instructor can change the learning objectives if necessary (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46, Clark, 2012 p. 205–249).

Characteristics of Formative Evaluation

- A crucial step in the learning process is formative evaluation.
- Formative Evaluation empowers student learning.
- It identifies the difficulties faced by a weak learner (Andrade, H., & Brookhart, 2020 p. 357-372)
- Results cannot be used for grading.
- It motivates students by letting students know how far they have come.
- It informs about the progress of the students so this evaluation motivates the students.
- It is a method of checking the effectiveness of evaluation. (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46).
- This evaluation does not take much time to complete.
- Feedback is considered a fundamental element of formative assessment. (Black & Williams, 1998; Crookes, 1988; Hattie & Timperley, 2007; Hattie, 2009).

- ❖ This is often an instructor-created test that teachers easily complete. (Clark, 2012 p. 205–249).
- ❖ This is usually a teacher-made test (Clark, 2012 p. 205–249).

Such as

Monthly tests

Class tests

Periodical assessment

Teacher's observation etc. (Wiliam, 2011p.3-14)

Disadvantages of Formative Evaluation

- ❖ *Time-consuming and resource-intensive*

Formative assessment is a time- and resource-intensive process that can be carried out on a daily, weekly, or monthly basis. This is because successful implementation of the plan necessitates regular data collection, research, reporting, and improvement.

- ➢ *Experts experienced with assessments*

Professionally qualified teachers who can analyze competency criteria and develop appropriate measures to assess student progress are needed to conduct formative assessment thoroughly. This type of education is currently lacking (Wiliam, 2011p.3-14)

Diagnostic Evaluation

During instruction, it is very important to determine the areas of weakness or difficulty of the students. This method tries to identify specific areas of weakness of a student within a particular course and helps in remediation (Crooks, 1988 p. 438-481)

B. According to Approaches:

Summative Evaluation

At the end of the course, a summative assessment is conducted to ascertain how much knowledge the students have gained after completing the predetermined objectives and syllabus. In fact, it is the assessment of students' performance at the end of a course. (Andrade, H., & Brookhart, 2020 p. 357-372).

Summative assessment is a method of grading students (Crooks, 1988 p. 438-481). It helps determine whether the learning objectives have been met.

Characteristics of Summative Assessment

- It is terminal in nature because it comes at the end of a course.
- It cannot be based on teacher observation alone.
- The results may be used for placement or grading purposes.
- Summative assessment is used to assess student learning at the end of a unit. (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015 p. 33-46)
- Assists students in developing a sense of personal responsibility
- It helps students connect what they learn in class to their real life and knowledge growth (Gielen, Dochy, Onghena, Struyven, & Smeets, 2011 p. 729-735).

Types of Summative Assessments

- Traditional school and university examination
- Final written test
- Test Created by Teachers
- Standardised examination
- Oral and Practical Examinations
- Research endeavours (Crooks, 1988 p. 438-481).

Disadvantages of Summative Evaluation

Summative assessment can cause anxiety in students. Anxiety, depression, and other negative emotions can arise from the pressure to do well on these assessments, and they can negatively impact students' learning and well-being (Dunn, Morgan, O'Reilly & Parry, 2004, p. 22-45).

C. Depending on the reference's nature:

Criterion-Referenced Test and Norm-Referenced Test

The two main approaches to assessment are norm-referenced and criterion-referenced (Dunn, Morgan, O'Reilly & Parry, 2004, p. 22). Generally, performance is evaluated using criterion-referenced testing. This type of test assesses a student's ability to behave in a standardized way. "Criterion-referenced test" was originally used by Glasser (1963) to refer to a performance continuum. Criterion-referenced assessments are believed to be more reliable, valid and transparent than norm-referenced assessments. Conventionally, norm-referenced assessment refers to the traditional category-based assignment of a number of measured characteristics (Baas, Castelijns, Vermeulen, Martens,

& Segers, 2015 p. 33-46). As described by Crooks (1988, p. 438-481), it shows the relationship between the measurement law and a group, a value, or an overall performance.

Conclusion

Evaluation is predicated on learning objectives; it helps determine whether those objectives are appropriate for a particular set of learners. Similarly, assessment evidences the value of learning experiences in the classroom even though it is predicated on them (Zimmerman, 2008 p.166-183). Assessment provides consistent feedback and validates the entire teaching-learning process. Assessment is needed at every step of the teaching-learning process (Andrade, H., & Brookhart, 2020, p. 357-372). It is essential to regularly assess the intellectual, emotional and physical development of students throughout their education and to conduct formative and diagnostic assessments to make prompt decisions about various remedial or enrichment programs. In order to categorize, grade, certify and promote students, summative assessment is needed (Dunn, Morgan, O'Reilly & Parry, 2004, p. 22-45). Assessments help instructors judge and make decisions about a student's academic career. (Crooks, 1988 p. 438-481).

To overcome these challenges and improve the quality of education the following issues can be assessed:

- ❖ The infrastructure of the organization should be maintained properly
- ❖ For a good result teachers need to improve their teaching skills.
- ❖ Teachers need to improve the quality of teachers
- ❖ Teaching can be done better if classrooms are equipped with modern teaching aids such as smart classrooms and digital course materials (Andrade, H., & Brookhart, 2020 p. 357-372).
- ❖ Peer learning is vital at every level which is essential to improve the quality of education (Clark, 2012 p. 205–249, Gielen, Dochy, Onghena, Struyven, & Smeets, 2011 P. 729-735).

References

Andrade, H., & Brookhart, S. (2020). Classroom assessment as the co-regulation oflearning. *Assessment in Education Principles Policy and Practice*, 27(4), 350–372.

Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2015). The relation between Assessment for Learning and elementary students' cognitive and metacognitive strategy use. *The British Journal of Educational Psychology*, 85(1),33–46. https://doi.org/10.1111/bjep.12058.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education Principles Policy and Practice*, 5(1), 7–74. https://doi.org/10.1080/0969595980050102.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24, 205–249. https://doi.org/10.1007/s10648-011-9191-6

Crooks, T. J., (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The Student Assessment Handbook*.London: Routledge Falmer.

Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education*, 36(6), 719–735. https://doi.org/10.1080/03075071003759037.

Glaser, B. G. (1963). Retreading Research Materials: The Use of Secondary Analysis by the Independent Researcher. *The American Behavioural Scientist*, 6, 1114.

Wiliam, D. (2011). *What is assessment for learning? Studies in Educational Evaluation*, 37, 3–14.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects.*American Educational Research Journal*, 45(1), 166–183. https://doi.org/10.3102/0002831207312909.

UNIT-3 :
Achievement

8

The Impact of Utilising the Smart Classroom Instructional Strategy on Academic Achievement

Harun Al Rasid Mondal

Abstract

The present study examines and contrasts the academic outcomes of pupils in a smart classroom setting with those in a traditional classroom setting. The present investigation is limited to students enrolled in the ninth grade at Habra High School. Students enrolled in flipped classes demonstrated superior academic performance compared to their counterparts in traditional settings. There is a significant gender gap in smart classrooms when it comes to pupils' academic performance. In terms of academic achievement, female students demonstrated superior performance compared to their male counterparts in the average classroom setting. The educational achievement of male students in the technologically enhanced classroom exceeds that of male learners in the traditional classroom setting. A significant discrepancy in academic attainment has been observed between male students in smart classrooms and those in conventional classes, with the former cohort demonstrating superior levels of achievement. Furthermore, it was discovered that.

Keywords: *Smart classroom instructional strategy, Academic achievement, Geography class*

Introduction

Learner-centred environment plays a valuable position in present education system (Hannifin and Land, 1997). It is a challenge to create such type of situation. Many professors and researchers advocate for the smart classroom

approach, in which instructors schedule time for students to work together on projects and use cutting-edge strategies for work done outside of class, such as watching digital video lectures before a lecture. (Bergmann and Sams, 2012). The smart classroom instructional strategy is a form of blended learning wherein students acquire knowledge of new content through online or offline video lectures, typically viewed at home. Consequently, what was traditionally assigned as homework is now completed during class time, allowing teachers to provide individualised guidance and engage in interactive discussions with students, rather than solely delivering lectures. This pedagogical approach is alternatively referred to as the backwards classroom, clever classroom, or reverse instruction. Smart classroom is joint venture of parents, students and teachers. Teacher should have mastery over the flipped teaching methodology and package. It is an innovative and exciting concept for teacher education. Because of this, students are able to participate actively in class discussions and other classroom activities. This prepares them for tasks requiring higher levels of thought. K-12 classrooms are the birthplace of the smart classroom instructional technique (Ash, 2012). Smart classroom researchers recommended that students treat taped lectures the same way they would treat assignments. A flipped curriculum was proposed by Prober and Khan (2013) for the field of medicine. They discovered that students preferred the SC model to the lecture format. Smart classrooms, as proposed by McLaughlin et al. (2013), are beneficial for students because they foster independence, growth, and participation. They found it more interesting and enjoyable, which improved their learning experience. They also saw an improvement in their schoolwork. Missidine, Fountain, Summers, and Gosselin (2013) examined the educational results of 589 students and found positive changes. In their research, Albert and Beatty (2014) found that students' academic performance improved in "Smart" classrooms. According to the research of Critz and Wright, students are taking more initiative in their education. According to research by Pierce and Fox (2012), 80% of pupils who used SC reported positive outcomes.. Students scored better in exams. Smart classroom model can deliver diversified benefits, spanning from effective teaching to active learning and also enhanced cognitive learning outcomes of students (Aidinopoules and Sampson, 2017), development of skills (Tanner and Scatt, 2015) as well as overall motivation (Sahin, Cavlazoglu, and Zeytuncu, 2015). Smart classroom model received a significant level of attention from both researchers and PR actioners, spanning a range various subjects including Mathematics (Katsa et al. 2016), Social studies (Aidinopoules and Sampson, 2017)., Humanities (Grossman, Figueir, and Mckinley, 2015). Majority of works has investigated the impact of Smart classroom model on a specific set of dimensions i.e. overall motivation, self-learning and cognitive learning outcomes. Hung (2015) studied that students

showed improvement in reading comprehension in English and improved significantly during the smart classroom. Research by Zappe et al. (2009) on a flipped undergraduate engineering course found that students felt the material had a favourable effect on their education. Results showed that students who were taught utilising the smart classroom strategy had higher levels of success than their standard classroom taught counterparts, according to research by Lai and Hu rang (2016).

Statement Of the Problem

The impact of implementing smart classroom teaching strategies on students' academic progress.

Objectives

1. In order to examine the impact of the smart classroom instructional technique on students' academic progress, a study was conducted.

Research Questions

1. Is there any significant difference in academic achievement scores of students in smart classroom and traditional classroom?
2. Is there any significant difference in academic achievement scores of male and female students in traditional classroom?
3. Is there any significant difference in academic achievement scores of male and female students in smart classroom?
4. Is there any significant difference in academic achievement scores of male students in smart classroom and traditional classroom?
5. Is there any significant difference in academic achievement scores of female students in smart classroom and traditional classroom?

Delimitations Of the Study

The scope of this study is limited to kids attending schools in the city of Habra only. The data was obtained exclusively from students enrolled in the ninth-grade Geography class.

Sample:

64 students served as a sample of the study. In present study, purposive sampling has been used for data collection. Data has been collected from Habra High School, Habra, N 24 Parganas. These 64 pupils were randomly assigned to either a control or an experimental group.. There were 32 students in each group. These 32 students were further categorized into 16 male and 16 female students.

Tools Used

Following tools were used for data collection:

1. Modules and video developed by the investigator
2. Academic achievement test developed by the investigator

Procedure

In present study, in traditional classroom method students were taught through lecture cum discussion method. Smart classroom instructional method composed of two phases.

1. FIRST PHASE: Students used offline you tube video and worksheets supplied by teacher before coming to class. Self-study was done by students. Thus they prepared their lesson themselves before coming to class.
2. SECOND PHASE: In class, students engaged in "high-order" activities, including discussing complex topics, solving problems, etc. The teacher created a positive learning environment and helped students go quickly through their coursework.

3. Statistical Techniques Employed

Descriptive statistics was used to analyze the data. For testing significance differences and effect on sample groups' T-test for small sample was used.

Table 1

Academic Achievement Scores of Students in Smart Classroom Model (Scm) and Traditional Classroom Model of Teaching

	M	**N**	**SD or s**	**SE_D**	**T**
Smart classroom students	16.31	32	**6.61**	0.52	4.29
Traditional classroom students	14.09	32	1.96		

It is observed from table 1 that T-ratio for difference in academic achievement between students of SC and TC was found to be 4.29. Entering table D with 62 df, we get entries 2.00 at 0.05 level of confidence and 2.66 at .01 level. Since T value reaches at 4.29 level of confidence. There exists a notable disparity in academic performance between kids in Smart classrooms

and those in traditional classrooms. The mean value of scores of smart classroom students is higher as compared to traditional classroom. One may make the case that students in virtual classrooms outperform their face-to-face peers on standardised tests.

Table 2

Academic Achievement Scores of Female and Male Students in Traditional Classroom

	M	**N**	**SD or s**	**SE_D**	**T**
Traditinal classroom female students	14.88	16	2.12	0.41	3.77
Traditional classroom male students	13.31	16	0.63		

Traditional classroom male students 13.31 16 0.63 It is observed from the table 2 that the t-ratio for difference in academic stress between female and male students of TC was found to be 14.88. Entering table D with 30 df, we get entries 2.04 at the 0.05 level of confidence and 2.75 at the .01 level. Since t value reaches at both levels of confidence. There exists a notable disparity in the levels of academic stress experienced by male and female students within a regular classroom setting. The average scores of female students are comparatively higher than those of male pupils. It can be argued that female students tend to outperform male pupils in terms of academic achievement scores within the context of typical classroom settings.

Table 3

Academic Achievement Scores of Female and Male Students in Smart Classroom

	M	**N**	**SD or s**	**SE_D**	**T**
Smart classroom female students	18.44	16	1.41	**0.5**	**8.47**
Smart classroom female students	14.19	16	1.42		

Table 3 shows that the t-ratio for the limited sample size for the gap in academic performance between female and male SC students was 8.47. Entering table D with 30 df, we get entries 2.04 at the 0.05 level of confidence and 2.75 at the .01 level. Since t value does not reach both levels of confidence. Therefore, the gender gap in smart classroom pupils' academic success is negligible. Female students, on average, have a better grade point average than male students. In the digital classroom, female pupils are able to outperform their male counterparts.

Table 4

Comparison of Male Students' Academic Performance in a Traditional Classroom with a Smart Classroom

	M	N	SD or s	SE_D	T
Smart classroom male students	14.88	16	2.12	**0.41**	**3.77**
Traditional classroom male students	13.31	16	0.63		

Table 4 shows that the t-ratio for the small sample indicating the academic success gap between male students in South Carolina was determined to be 3.77. We obtain entries 2.04 at the 0.05 level of confidence and 2.75 at the.01 level of confidence when we enter table D with 30 df. given that the t value is at both confidence levels. Therefore, there is a notable distinction between male students in Smart Classroom and Traditional Classroom academic accomplishment. In comparison to male students in traditional classrooms, the male smart classroom students' mean score is greater. It can be said that male students in smart classrooms perform better academically than male students in traditional classrooms.

Table 5

Academic Achievement Scores of Female Students in Smart Classroom Model (Scm) And Traditional Classroom (Tc)

	M	N	SD or s	SE_D	T
Smart classroom female students	18.44	16	2	**0.51**	**7.03**
Traditional classroom female students	14.88	16	2.12		

Table 5 shows that the sample size t-ratio for the gender gap in SC female students' academic performance is 0.89. Using 18 df, the values 2.10 and 2.88 are entered into table D at the 0.05 and.01 levels of confidence, respectively. Since the t-value is below the two- and five-percent confidence thresholds. Therefore, there is little to no difference between Smart classroom and Traditional classroom pupils' academic performance among female students. Female students in the smart classroom, on average, had greater test scores than their counterparts in the traditional classroom. Female students at South Carolina universities are more successful academically than their counterparts in more traditional educational settings.

Findings and Discussion of The Study

The academic performance of students in smart classrooms was shown to be superior to that of students in traditional classrooms. This observation aligns with the research conducted by Missidine, Fountain, Summers, and Gosselin (2013). In the conventional classroom, female pupils outperformed their male counterparts in terms of academic achievement. When compared to their male peers, female pupils demonstrated greater academic accomplishment in the smart classroom. Male students in the smart classroom outperformed male pupils in the traditional classroom in terms of academic performance. Female students in the smart classroom outperformed their regular classroom counterparts in terms of academic performance.

Implications

Although the study has been conducted on a small sample, it is giving some kind of trend with respect to Smart classroom instructional strategy viz. academic achievement. Findings of the present study suggest that in the Indian context, Smart classroom work better than traditional classroom. Academic

achievement was higher among female pupils than male students.There was higher academic achievement scores in female students than male students in the traditional classroom set up. But in Smart classroom model, academic achievement was enhanced in female students in comparison to male students. It could be inferred that Smart classroom works better in comparison to male students. For improved outcomes, this approach can be repeated with various samples, circumstances, variables, and practises..

References

Aidinopoulou, V., & Sampson, D. G. (2017). An action research study from implementing the smart classroom model in primary school history teaching and learning. *Journal of Educational Technology & Society,* 20(1), 237-247.

Albert,M., & Beatty, B. (2014). Flipping the classroom applications to curriculum redesign for an introduction to management course: impact on grades. *Journal of Education for Business*, 89, 419–424.

Ash, K. (2012). Educators view flipped model with a more critical eye. *Education Week*, S6–S7.

Bergmann, J., & Sams, A. (2009). Remixing chemistry class: two Colorado teachers make podcasts of their lectures to free up class time for hands-on activities. *Learning & Leading with Technology*, 36(4), 22–27.

Critz, C., & Wright, D. (2013). Using the smart classroom in graduate nursing education. *Nurse* Educator, 38(5), 210–213.

Hung, H. T. (2015). Flipping the classroom for English language learners to foster active learning. *Computer Assisted Language Learning*, 28(1), 81-96.

J. R. Young (2002). Hybrid teaching seeks to end the divide between traditional and online instruction, *Chronicle of Higher Education*. 48 (28), A33-A34.

McLaughlin, J., Roth, M., Glatt, D., Gharkholonarehe, N., Davidson, C., LaToya, G., et al. (2014). The smart classroom: a course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, 89(2), 236–243.

Missildine, K., Fountain, R., Summers, L., & Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *Journal of Nursing Education*, 52(10), 597–599.

Pierce, R., & Fox, J. (2012). Instructional design and assessment: podcasts and active learning exercises in a "smart classroom" model of a renal pharmacotherapy module. *American Journal of Pharmaceutical Education*, 76(10), 1–5.

Prober, C., & Khan, S. (2013). Medical education reimagined: a call to action.

Academic Medicine, 88, 1407–1410.

9

Constructing and Standardizing Achievement Tests: A Comprehensive Guide

Debashis Barman &Dr. Sanjib Kumar Roy

Abstract

Achievement tests serve as vital tools for assessing student knowledge and skills within specific subject areas. This abstract provides an overview of the key steps involved in the construction and standardization of achievement tests. Beginning with the design and blueprinting phase, the process involves meticulous item development, cognitive level specification, item review, and pretesting. Item analysis further refines the test, with attention to difficulty and discrimination indices. Following these steps, the standardized test is administered to a representative sample for score analysis, reliability assessment, and validity evaluation. Norming procedures yield standard scores that facilitate score interpretation. This comprehensive guide underscores the importance of content validity, construct validity, and criterion validity in establishing a test's validity claims. The iterative nature of test revision ensures continual improvement of the assessment tool. In conclusion, the construction and standardization of achievement tests demand a systematic approach, combining expertise in assessment methodologies and subject matter knowledge. This process culminates in the creation of a reliable and valid assessment instrument that accurately measures student achievement.

Keywords: *Achievement tests, Assessment, Standardization, Validity, Reliability, Item analysis.*

Introduction

Achievement tests play a crucial role in educational assessment by measuring a student's knowledge, skills, and abilities in a particular subject area. The process of constructing and standardizing achievement tests requires meticulous planning, rigorous methodology, and a deep understanding of the subject matter. This chapter provides a comprehensive guide to the step-by-step process of constructing and standardizing achievement tests. Achievement tests have become integral components of educational assessment systems, providing valuable insights into students' grasp of subject matter, cognitive skills, and educational progress. These tests serve as powerful tools for educators, policymakers, and researchers, aiding in the evaluation of curriculum effectiveness, instructional strategies, and overall student achievement. To ensure the accuracy, fairness, and validity of these assessments, the process of constructing and standardizing achievement tests demands meticulous planning, careful execution, and adherence to established psychometric principles (Hash, 1969).

Purpose and Significance

The primary purpose of achievement tests is to measure the knowledge and skills that students have acquired within a specific subject area or curriculum. Unlike aptitude tests, which gauge an individual's potential to learn, achievement tests focus on what has already been learned. Their results provide insights into the effectiveness of teaching methods, curriculum design, and educational policies. Achievement tests are widely employed at various levels of education, including primary, secondary, and tertiary education, as well as professional certification and licensure examinations.

Challenges and Considerations

Constructing and standardizing achievement tests presents several challenges that must be carefully addressed. Item development requires careful consideration of item quality, clarity, and alignment with learning objectives. The cognitive level specification ensures that the test comprehensively assesses various levels of cognitive complexity, from basic recall to higher-order thinking. Balancing content coverage and cognitive levels requires strategic planning to create a comprehensive test blueprint.

Another critical consideration is the avoidance of bias and cultural sensitivity in test items. The test should be accessible and fair to all students, regardless of their cultural or socio-economic background. Ensuring that the test content is reflective of the diversity within the student population enhances the test's validity and fairness (Hamilton, Stecher, & Klein, 2002).

Organization of the Chapter

This chapter aims to provide a comprehensive guide to the process of constructing and standardizing achievement tests. The subsequent sections delve into each step of the process, elucidating the methodologies, principles, and best practices involved:

1. Test Design and Blueprint

The foundation of a successful achievement test lies in the careful design of its blueprint. This blueprint outlines the content areas, the number of questions per content area, and the cognitive levels being assessed. A well-structured blueprint ensures that the test adequately covers the curriculum and assesses the desired skills.

- **Determine Objectives**: Define the purpose of the test, the subject area, and the specific learning objectives or content standards that the test will assess.
- **Create a Blueprint**: Design a test blueprint that outlines the content areas to be covered, the number of items per content area, and the cognitive levels to be assessed (e.g., recall, application, analysis) (Rudner, Conoley, & Plake, 1989).

2. Item Development

Item development involves creating the individual questions or items that comprise the test. These items should be clear, concise, and aligned with the learning objectives. There are various types of items, including multiple-choice, true-false, short-answer, and essay questions. Item writers must adhere to guidelines to maintain item quality, avoid biases, and ensure appropriate difficulty levels.

- **Write Clear Items**: Craft test items that are clear, concise, and aligned with the learning objectives. Ensure that each item has a single correct answer and is free from bias or ambiguity.
- **Use Different Item Types**: Develop a variety of item types (e.g., multiple-choice, true-false, short answer, essay) that reflect the intended cognitive levels and assess different skills (Elwick, 2021).

3. Cognitive Level Specification

Each item on the achievement test should target a specific cognitive level. Bloom's Taxonomy provides a framework for categorizing cognitive levels, ranging from simple recall of information to higher-order thinking skills such as analysis and synthesis. A balanced distribution of items across different cognitive levels ensures a comprehensive assessment (Rudner, Conoley, & Plake, 1989).

- **Mapping Learning Objectives:** Start by identifying the specific learning objectives you want to assess. These objectives should be clear and aligned with the curriculum.

- **Matching with Bloom's Levels:** Determine the appropriate cognitive level for each learning objective. For instance, if the objective involves recalling facts, it aligns with the knowledge level. If it requires analysing information, it aligns with the analysis level.

- **Creating Test Items:** Craft test items that correspond to the chosen cognitive levels. Ensure that each item challenges students to demonstrate the intended thinking skills.

4. Item Review and Revision

Prior to finalizing the test, a thorough review of items is essential. Experts in the subject matter and assessment methodologies should review the items for accuracy, clarity, and alignment with the content and cognitive levels. Feedback from the review process informs necessary revisions.

- **Expert Review**: Have subject matter experts review test items for accuracy, clarity, and alignment with the learning objectives and cognitive levels.

- **Revise as Needed**: Incorporate feedback from the review process to revise and refine test items for improved quality and relevance (Elwick, 2021).

5. Pretesting

Conducting a pre-test involves administering the test to a small sample of the target population. Pretesting helps identify potential issues with items, such as ambiguous wording or unexpected difficulty levels. Analysis of pre-test data allows for item refinement and the calculation of item statistics.

- **Administer Pre-test**: Administer the test to a small sample of students who are similar to the target population. Collect data on item performance, time taken, and student feedback.

- **Analyse Pre-test Data**: Analyse pre-test data to identify problematic items, confusing wording, or unexpected difficulty levels.

6. Item Analysis

Item analysis involves statistical evaluation of each item's performance based on data from the actual test administration. Common item statistics include difficulty index (proportion of students who answered correctly), discrimination index (item's ability to distinguish between high and low

achievers), and item-total correlation. Items with poor statistics may be revised or removed (Rudner, Conoley, & Plake, 1989).

- **Calculate Difficulty Index**: Calculate the proportion of students who answered each item correctly (difficulty index) to identify overly easy or difficult items.
- **Compute Discrimination Index**: Analyze how well each item discriminates between high and low achievers by comparing item performance with total test scores.

7. Test Assembly

With refined items, the test is assembled according to the predetermined blueprint. The order of items, the balance between content areas and cognitive levels, and the overall length of the test are carefully considered during this stage. The process of test assembly,

- **Blueprint Adherence:** Ensure that the test assembly aligns precisely with the test blueprint. This involves distributing items according to content areas and cognitive levels.
- **Item Selection:** Review all the developed items and select those that best represent the intended content and cognitive levels. Ensure that each item contributes meaningfully to the assessment.
- **Cognitive Level Balance:** Strive for an equitable distribution of items across different cognitive levels. This ensures that the test measures a range of thinking skills and avoids skewing towards specific abilities (Rudner, Conoley, & Plake, 1989)..
- **Item Ordering:** Organize the items logically and coherently. Typically, tests start with easier items and progress to more challenging ones. However, this order can vary based on the test's purpose and design (Elwick, 2021).

8. Standardization Sample Selection

The standardization sample consists of a representative group of individuals who will take the test for the purpose of establishing norms. This sample should reflect the demographic characteristics of the target population. A larger and more diverse sample ensures more robust and generalizable results. The process of standardization sample selection,

- **Define Target Population:** Clearly identify the population for which the test is intended (e.g., students of a particular grade level or subject).

- **Sample Size:** Determine an appropriate sample size that provides statistical validity without being excessively large or burdensome to administer.
- **Random Sampling:** Use random sampling techniques to ensure that each individual in the target population has an equal chance of being selected for the sample.
- **Demographic Representation:** Strive to include a diverse group of participants that reflects the demographic characteristics of the target population, including factors such as gender, ethnicity, socioeconomic status, and geographic location.
- **Informed Consent:** Obtain informed consent from participants or their guardians if the sample includes minors (Elwick, 2021).

9. Test Administration

Test administration is a critical phase that ensures the reliability and validity of an achievement test. By adhering to standardized procedures, educators and assessment professionals guarantee consistent test-taking conditions and contribute to the overall quality of the assessment. A well-administered test provides accurate and meaningful data that can be used for subsequent analyses and score interpretation (Saraf, 1968).

10. Score Analysis

Score analysis involves examining the performance of the standardization sample on the test. This includes calculating descriptive statistics such as mean, median, and standard deviation. These statistics provide insights into the overall performance of the group and aid in score interpretation.

11. Reliability Analysis

Reliability is a fundamental aspect of the quality of an achievement test. It refers to the extent to which the test consistently measures the same construct over multiple administrations or under varying conditions. In other words, a reliable test should produce consistent results for the same group of examinees. There are several methods for assessing the reliability of an achievement test, each offering insights into the stability and consistency of the test scores (Rudner, Conoley, & Plake, 1989).

Reliability analysis assesses the consistency and stability of an achievement test's scores. It's crucial to ensure that the test provides dependable and accurate results. There are different methods to measure reliability (Elwick, 2021).

- **Test-Retest Reliability:** Administer the same test to the same group twice. A high correlation between scores indicates good reliability.
- **Split-Half Reliability:** Divide the test into two halves and compare scores. A strong correlation suggests internal consistency.
- **Cronbach's Alpha:** Measures internal consistency by assessing how closely items are related. A higher value implies better reliability.

Inter-Rater Reliability: Ensures consistent scoring for constructed-response items. Kappa coefficient or intraclass correlation are used.

Reliability coefficients range from 0 to 1. Higher values indicate better reliability. However, an acceptable level depends on the context and purpose of the test. High reliability is crucial for test validity and enhances the credibility of the scores. It's a key factor in making accurate inferences about student performance (Elwick, 2021).

12. Validity Assessment

Validity refers to the extent to which the test measures what it is intended to measure. Various types of validity evidence, such as content validity, criterion validity, and construct validity, are collected to support the test's validity claims. This step ensures that the test accurately reflects the underlying construct.

- **Content Validity**: Ensure that the test content adequately covers the intended learning objectives and cognitive levels.
- **Criterion Validity**: Correlate test scores with external criteria (e.g., other established tests or measures) to determine the test's ability to predict real-world performance.
- **Construct Validity**: Analyse the relationship between test scores and related constructs to establish the test's validity.

13. Norming and Standard Score Creation

Norming involves establishing the normative reference group and creating standard scores. Standard scores, such as percentile ranks, z-scores, and stanines, allow for meaningful score interpretation by comparing an individual's performance to the reference group.

14. Test Revision

Based on the findings from reliability, validity, and norming analyses, the test may undergo further revision to improve its psychometric properties. This iterative process ensures that the test is fair, accurate, and reliable for its intended purpose (Creswell, 2012).

Conclusion

The construction and standardization of achievement tests are intricate processes aimed at creating reliable and valid assessment tools. These steps, including test design, item development, analysis, and standard score creation, ensure a comprehensive and fair evaluation of students' knowledge and skills. By following these procedures, educators and assessment experts contribute to accurate measurements, informed decision-making, and educational growth. Continuous evaluation and refinement of tests are essential for adapting to changing educational needs and maintaining assessment quality. Overall, constructing and standardizing achievement tests supports equitable and meaningful education outcomes.

References

Creswell, J.W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.). Boston: Pearson.*

Elwick, J. (2021). *Making a grade: Victorian examinations and the rise of standardized testing*. University of Toronto Press.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education. Rand Corporation.*

Hash, J. A. (1969). *The construction of a standardized achievement test in forestry*.

National Research Council, Division of Behavioral and Social Sciences and Education, Center for Education, Board on Testing and Assessment, & Committee on the Foundations of Assessment. (2001*). Knowing what students know: The science and design of educational assessment.* National Academies Press.

Rudner, L. M., Conoley, J. C., & Plake, B. S. (1989). *Understanding achievement tests: A guide*

Saraf, S. N. (1968). *Construction & standardisation of an achievement test in Indian history*.

10

Steps For Construction & Standardization of Achievement Test

Dr. Vandana Verma

Abstract

The development of a well-constructed and standardized achievement test is a meticulous and essential process in the realm of educational assessment. Achievement tests serve as powerful tools to measure the knowledge, skills, and competencies that individuals have acquired in specific subject areas or domains. These tests play a pivotal role in evaluating the effectiveness of educational programs, guiding instructional decisions, and providing valuable insights into the learning progress of individuals This abstract outlines the systematic steps involved in creating a valid, reliable, and fair achievement test. Beginning with the definition of the test's purpose and content domain, the process proceeds through item generation, review, and selection, followed by pilot testing and subsequent item analysis. The test is then refined, administered to a representative sample, and its reliability and validity are assessed. Norms and standard scores are established through standardization, culminating in the production of informative score reports. This abstract provides a concise overview of the comprehensive process that ensures the quality and effectiveness of achievement tests, ultimately contributing to informed educational decisions and improved learning outcomes.

Keywords*: Achievement test, Standardization, Reliability, Validity and Norms.*

Introduction

One instrument that we, as educators, use to evaluate a student's level of

knowledge or proficiency in a given subject is an achievement test. It is usually given to assess the effectiveness of the instruction and the degree of learning that has occurred after a period of instruction.

An assessment tool known as an achievement test is used to gauge a person's level of knowledge, proficiency, or aptitude in a given subject. Achievement tests can be used in many different contexts, including employee training programs, schools, and certification programs. They can be used to evaluate a student's academic performance, job-related skills, or competency in a particular field.

Meaning of Achievement Test

Achievement is defined as one's proficiency, accomplishments, and level of learning. It has a direct bearing on how students develop and grow in learning environments. An achievement test evaluates knowledge or skills that have been developed. Standardized tests are the most popular kind of achievement tests. They are designed to assess the abilities and knowledge that students have acquired at a particular grade level, typically through organized instruction like classroom instruction or training. It is a crucial instrument for evaluating schools and plays a major role in tracking students' academic and instructional advancement. Exams ought to provide a precise representation of the knowledge and abilities of students in the field or subject matter being assessed. Precise achievement data are crucial for both program evaluation and curriculum and instruction planning.

Objectives

To Identify and explain reasons for performing tests. Understand testing terminology to communicate clearly with students and colleagues.

To Evaluate a test's validity and reliability. Select appropriate tests.

To Administer test protocols properly and safely.

Functions of Test

- ❖ It offers a foundation for advancement to the subsequent grade.
- ❖ to determine each student's current standing in a variety of academic areas.
- ❖ It facilitates the decision of where to place the students in a given section.
- ❖ to inspire the pupils prior to starting a new task.
- ❖ to be aware of how well a student is doing both in clinical and theoretical settings.

- ❖ to reveal students' problems so the teacher can assist them in solving them.

Characteristics of a Good Test

- ❖ It can be tried out and selected on the basis of its difficulty level and discriminating power.
- ❖ Directly related to the educational objectives.
- ❖ It should possess description of measure behavior in realistic and practical terms.

Contains a sufficient number of test items for each measured behavior; concerned with important and useful matter; comprehensive, brief, precise and clear.

- ❖ It should be divided into different knowledge and skills according to behavior to be measured.
- ❖ Standardized the items and made instructions clear so that different users can utilize it.
- ❖ Rules and norms have to be developed so that various age groups can use at various levels. It provides equivalent and comparable forms of the test.
- ❖ A test manual has to be prepared, which can act as a guide for administering and scoring.

Purpose of Achievement Test

- ❖ Diagnose strength and weaknesses
- ❖ Assign Grades
- ❖ Achieve Certification or Promotion
- ❖ Advanced Placement/College Credit Exams
- ❖ Curriculum Evaluation
- ❖ Accountability
- ❖ Informational Purposes
- ❖ Construction of an Achievement Test

Achievement Test

Major steps involved in the construction of achievement test

1. Planning of test
2. Preparation of a design for the test

3. Preparation of the blue print

4. Writing of items

5. Preparation of the scoring key and marking scheme

6. Preparation of question-wise analysis

1. Planning of Test

1.1 Instructional Objectives of the Test

The first and the most important step in planning a test is to identify the instructional objectives. Each subject has a different set of instructional objectives. In the subjects of Science, Social Sciences, and Mathematics the major objectives are categorised as knowledge, understanding, application and skill, while in Languages the major objectives are categorised as knowledge, comprehension and expression. Knowledge objective is considered to be the lowest level of learning whereas understanding, application of knowledge in sciences or behavioural sciences are considered higher levels of learning.

Determine the maximum time and maximum marks

2. Preparation of a design for the test

Important factors to be considered in design for the test are:

2.1 Weightage to objective

2.2 Weightage to content

2.3 Weightage to form of questions

2.4 Weightage to difficulty level.

2.1 Weightage to Objectives

This indicates what objectives are to be tested and what weightage to be given to each objectives.

S. N.	Objectives	Marks	Percentage
1	Knowledge	3	12
2	Understanding	2	8
3	Application	6	24
4	Analysis	8	32
5	Synthesis	4	16
6	Evaluation	2	8
	Total	25	16

2.2 Weightage to Content

This indicates the various aspects of the content to be tested and weightage to be given to these different aspects.

2.3 Weightage to form of questions

This indicates the form to be included in the test and the weightage to be given for each form of Question.

S. N.	Content	Marks	Percentage
1	Subtopic 1	15	60
2	Subtopic 2	10	40
	Total	25	100

2.4 Weightage to Difficulty Level

This indicates the total mark and weightage to be given to different questions.

S. N.	Form of Questions	Number of Questions	Marks	Percentage
1	Objective Type	14	7	28
2	Short Answer Type	7	14	56
3	Long Answer Type	1	4	16
	Total	22	25	100

S. N.	Form of Questions	Number of Questions	Marks	Percentage
1	Objective Type	14	7	28
2	Short Answer Type	7	14	56
3	Long Answer Type	1	4	16
	Total	22	25	100

3. Preparation of the Blueprint

Blueprint is a three dimensional chart giving the placement of the objectives, content and form of questions .

4. Writing of Items

The paper setter follows the blue print when writing items. When drafting the items, the degree of difficulty must be taken into account. It should also verify that all of the questions can be addressed in the time given. It's best to list the questions according to increasing levels of difficulty.

5. Preparation of the Blueprint

Objectives	Knowledge			Understanding			Application			Analysis			Synthesis			Evaluation			Total
Form of Questions Content	O	SA	E	O	SA	E	O	SA	E	O	SA	E	O	SA	E	O	SA	E	
Subtopic 1	2(4)			1(2)			2(4)	2(1)			4(2)			2(1)			2(1)		15
Subtopic 2	1(2)			1(2)				2(1)				4(2)		2(1)					10
Total Marks	3	0	0	2	0	0	0	0	0	0	4	4	0	4	0	0	2	0	25
Grand Total	3			2			6			8			4			2			

Note: O-Objective Type. SA-Short Answer Type. E-Essay Type

6. Preparation of the scoring key and marking scheme

In the case of objective type items where the answers are in the form of some letters or other symbols a scoring key is prepared.

Scoring Key

Question No.	Answer	Marks
1	A	1/2
2	C	1/2
3	B	1/2
4	A	1/2
5	D	1/2

In the case of short answer and essay type questions, the marking scheme is prepared.

In preparing marking scheme the examiner has to list out the value points to be credited and fix up the mark to be given to each value point.

Marking Scheme

The "Marking Scheme" preparation is the final step. The marking system aids in avoiding inconsistent decision-making. The possible answers to the test items are organized in the marking scheme. The different response value points are graded, and the marks that are permitted for each value point are noted. The marking scheme guarantees impartiality in assessment and removes score discrepancies that might arise from the peculiarities of the assessor. Naturally, the scoring key—which is prepared with regard to objective-type questions—is included in the marking scheme. Let's go over this in more detail. Apart from the quality of the question paper, reliability of assessment, to a great extent, depends on the degree of consistency of scores assigned to the students by different examiners or by the same examiner on two different occasions. Thus, variation can occur because of any one of two different reasons:

1) As a result of the same examiner's inconsistent application of standards when evaluating various answer scripts.

ii) As a result of various examiners applying various standards of judgment.

iii) An examiner is considered consistent in their marking if they consistently assign the same grade or marks to an answer script after exposing it to them multiple times. Because of this, it could be claimed that the evaluation they completed was more accurate and consistent than that of the other examiner, whose case had a larger variance in the marks awarded.

Highlights of a good marking scheme

1) It is a three column statement showing serial number of the questions, their expected outline answers and the marks allotted to each value point under them.

2) In respect of long answer or essay type questions, the expected outline answers should:

i) be complete and cover all possible or major areas as demanded by the questions

ii) clearly indicate each expected point or the parts under the outlined major areas iii) provide direction as to whether all points will count towards a complete or correct answer or a set of points will be adequate enough for full credit (All this should be clearly reflected), and

iv) indicate marks for each expected point. Marks so distributed over expected points or their sets should be equal to the total marks assigned for a question.

3) In respect of short answer questions, a complete answer may be provided with its break-ups where ever necessary along with the break-up of marks.

4) Out of the total marks assigned for a question, cach point so explained may be assigned marks according to their significance in the answer.

5) In some situations, apart from the content, other qualities of answer may also matter significantly, particularly in long answer or essay type questions. These could be logical approach, coherence, lucidity of expression, the style of presentation etc. Some marks may also be set apart for such overall quality of answer which cannot be usually covered in cnumeration of the content points.

6) The scheme of marking needs to be comprehensive enough not to leave any point unexpected and thus should provide clear guidelines in respect of the break-up of marks over different points or parts of the answer.

7) If a question entails some other points beyond one's expectation, a provision may also be made to take them into account and suitably reward them.

6. Question-wise Analysis

Question-wise analysis is the final and sixth step. This kind of exercise aids in the paper setter's assurance that the question paper is balanced. During the analysis of questions. The study examines every query using a range of criteria from the blueprint.

Pilot the test

It is best to pilot the test on a small sample of people who are representative of the target population before distributing it to the entire population.

This will assist in determining whether there are any issues with the test items, like imprecise directions or confusing questions. The test items are improved and the test's overall quality is raised with the help of the pilot's feedback.

Administer the test

Once we have constructed and piloted the test, it is ready for administration to the target population. It's essential to ensure that the testing environment is conducive to concentration and focus.

It is important to provide clear instructions to ensure that participants understand what we actualy expect from them.

Test security

Maintain the security and confidentiality of the test to prevent cheating or other forms of test compromise. This might involve using secure testing environments, monitoring test takers during the testing process, or implementing other security measures

Check the validity and reliability

Verify the test's validity and reliability. While reliability refers to the consistency of test results over time or across different test takers, validity refers to the degree to which the test measures what it is intended to measure. Validity and reliability are crucial factors to take into account to guarantee the test's overall quality and utility.

Conclusion

It's time to score the test once we are able to administer it. Either a manual process or automated scoring systems can be used. For scoring to be impartial and consistent, the scoring key needs to be created in advance. To ascertain the degree to which the test objectives were fulfilled and the general efficacy of the instruction, an analysis of the scores is necessary.

References

Anastasi, A. (1982) *Psychological Testing*. 5th Ed., New York: Mac Millan Publishing Company,135.

Burton, S.J., Sudweeks, R.R., Merrill, PF & Wood, B. (1991) How to prepare better multiple choice test items: *Guidelines for university faculty: Brigham Young University Testing Services and The Department of Instructional Scienc*e. Available at https://testing.byu.edu/handbooks/betteritems.pdf

Craighead, W.E. & Charles, B.N. (2001). *The Corsini Encyclopedia of Psychology and Behavioral* Science New York: Wiley and Sons

Ebel, Robert L., (1966): *Measuring Educational Achievement, Prentice-Hall of India*, New Delhi.

Harper, A. Edwin, J. and Harper, Erika S., (1992): Preparing Objective Examinations. *A Handbook for Teachers, Students and Examiners*, Prentice-Hall of India, New Delhi.

Heale, R. & Twycross, A. (2015). *Validity and Reliability in quantitative studies Evidence-Based Learning*, 18(3), 66-67 Available at https://ebn.bmj.com/content/ ebnurs/18/3/66.full pdf

Kelley, T.L. (1939). *The selection of Upper and lower groups for the validation of test items. Journal of Educational Psychology*, 30(1), 17-24. Available at https://www.researchgate.net/publication/232494422 The selection of upper and lower groups for thevalid ation of test Items

Popham, W. James, (1990*): Modern Educational Measurement*: Pracatitioners Perspective. Prentice-Hall, USA. Remmers, H.H et al., (1967): *A Practical Introduction to Measurement and Evaluation.* Universal Bookstall, Delhi.

Robson, C. (2011). Real World Research: *A Resource for Users of Social Research Methods in Applied Settings*, (2nd Ed.). Sussex, A. John Wiley and Sons Ltd.

Shelat, NA (2015). A study of the factors affecting academic achievement of Bangladesh primary school children of Dhaka city. *Unpublished Ph.D. thesis, Maharajaraja Sayajirao University, Baroda.* https://old.amu.ac.in/emp/studym/99997033.pdf

UNIT - 4:
NRT AND CRT

11

Norm Referenced Test and Criterion Referenced Test

Bikshyatsib Sardar

Abstract

The concept of testing students' knowledge and skills is pivotal in various domains, ranging from education to professional development. This book chapter delves into a detailed exploration of the two primary testing paradigms, norm-referenced and criterion-referenced testing. With a comprehensive analysis of their meanings, definitions, distinguishing characteristics, illustrative examples, utility, advantages, and limitations, this chapter provides a holistic understanding of the role these methodologies play in evaluating students' performance. The chapter commences by clarifying the foundational concepts of norm-referenced and criterion-referenced testing. To enable a thorough comprehension, the distinctive characteristics of both testing paradigms are meticulously elucidated. Real-world examples drawn from educational contexts are presented to illustrate the practical applications of these methodologies. The chapter delves into the varied uses of each approach, highlighting how a norm-referenced testing aid in comparative benchmarking and criterion-referenced testing informs the mastery of targeted skills. Moreover, the advantages and limitations inherent in norm-referenced and criterion-referenced testing are critically examined. In essence, this book chapter provides an insightful journey into the realms of norm-referenced and criterion-referenced testing. By revealing their meanings, characteristics, illustrative examples, applications, advantages, and limitations, the chapter equips readers with a comprehensive toolkit for informed decision-making in designing effective evaluation strategies. This understanding not only enriches evaluation practices but also contributes to the enhancement of learning experiences across diverse domains.

Keywords*: Testing Paradigm, Norm-Referenced Test, Criterion*-Referenced Test and Evaluation.

Introduction

Educational assessment plays a pivotal role in evaluating student learning and guiding instructional strategies. Among the various assessment approaches, two distinct but essential approaches stand out: Norm Referenced Test (NRT) and Criterion Referenced Test (CRT). These approaches, often referred to as the typology of the nature of reference, hold the key to understanding how we measure and evaluate student performance in educational contexts. While there are certain similarities between NRT and CRT, their fundamental differences lie in the construction and interpretation of test scores. This paper attempts to explore the unique characteristics and applications of these assessment methods, offering educators, researchers, and assessment professionals' valuable insights into their use within the realm of education.

Norm Referenced Test (NRT)

Norm-Referenced Tests are made to rank each student according to their performance in comparison to a larger group of fellow learners. These tests measures sample skills and knowledge from diverse sources, including textbooks, syllabi, and expert curriculum judgements. The test items are intentionally varied in difficulty to distinguish between low and high achievers. NRT (Norm-Referenced Test) results are typically expressed as either a percentile rank or a score equivalent to a specific grade level. These results are assessed by comparing an individual's performance to the performance of a larger group of test-takers. The primary purpose of NRT is to understand each student's achievement level within a specific class or group. To interpret a student's position in the class, the teacher considers the overall performance of the entire class in the subject being assessed, which is known as the norm or norm group. Some of the definitions of NRT are:

"A norm-referenced test is designed with the specific intention of providing an assessment of an individual's performance that can be understood within the context of their relative position among recognized peer groups." — Adapted from N.R. Gronlund (2007)

"A NRT is employed to measure a student's progress in attaining specific learning goals and to compare their level of achievement to that of both their peers and established norm groups." — Adapted from Bormuth (1970)

Examples of NRT

Biki student of Class –IX has secured 78 out of 80 marks in Mathematics of 1st summative test. His individual score says nothing about his position in class in Mathematics. Thus, the teacher will compare his performance with class Mean and SD. Then the teacher will calculate percentile rank of Biki i.e. *4th position in Mathematics*. Let us examine other examples of NRT below:

- ❖ Biki stood 3rd position in the Class-X.
- ❖ Roki score in Sciences is above average.
- ❖ Suru secured 2nd position in NEET examination.

If you examine each of the above examples, the scores secured by the learners have not been reflected. But, their performance in the examination has been interpreted in terms of a norm group which may be a *class*, *school*, *locality*, a *state* or a *nation*.

Main Characteristics of a Norm-Referenced Testing

- ❖ It is administered after deliberate instruction by the teacher.
- ❖ Its main objective is to measure students' achievement curricular subjects.
- ❖ It is prepared for particular grade.
- ❖ It is presented in the form of a percentile rank, a normalized standard score, a linear standard score or a grade equivalent score.
- ❖ It classifies achievement of the students below average, average and above average for a given class or grade.

Use of Norm Referenced Test

- ❖ To select the best of the learners in a group.
- ❖ To identify relative performance of a particular learner in a group.
- ❖ NRT scores can be employed for certifying or grading students.
- ❖ It can also be used for guiding students' placement into higher education programs and facilitating their entry into diverse career fields.

Limitation of Norm-Reference Test

- ❖ Norm referenced tests promote unhealthy competition among learners as they try to excel each other. Further, these tests adversely affect the self-concept of low scoring learners.
- ❖ Test items which are answered correctly by most learners are not included in these tests because of their inadequate contribution to response variation.

Criterion Referenced Test

CRT is meant to measure the achievement of an examinee in any subject with reference to some fixed criterion/domain rather than the performance of other examinee. A Criterion-Referenced Test (CRT) assesses a student's performance based on specific standards or criteria, rather than comparing them to their peers, as is the case with Norm-Referenced Tests. In CRT, students' performance is measured against a predetermined benchmark known as the criterion. This approach allows us to evaluate an individual examinee's

performance independently, without reference to others. The primary goal is to determine whether each student has successfully acquired particular skills or grasped specific concepts outlined in the curriculum. These skills are typically identified and defined by teachers and curriculum experts. In CRTs, the test items are designed to be of comparable difficulty for the targeted skills. Each student's performance is compared against a set standard of acceptable achievement, making it unnecessary to consider how other examinees have fared. This ensures that the focus remains on individual mastery of the prescribed skills and concepts. N.E. Gronlund (2007) defines CRT as "A test is intentionally created to yield a performance measure that can be understood within the context of a clearly defined and limited domain of learning tasks."

Examples of CRT

- ❖ During a spelling test, Biki spelled 9 out of 10 words correctly, indicating proficiency in spelling.
- ❖ In a math quiz, Dipti solved 7 out of 8 problems accurately, demonstrating a strong understanding of the math concepts covered.
- ❖ On a reading comprehension exercise, Roki correctly answered 12 out of 15 questions, showcasing good comprehension skills.

These examples illustrate how CRT measures an individual student's performance against specific criteria or standards within a particular subject or skill area.

Uses of Criterion-Referenced Test

CRT can be used to:

- ❖ Identify those learners who have acquired mastery on intended learning outcome in a subject (with reference to acquisition of knowledge and skill).
- ❖ Find degree of mastery a learner has achieved in relation to a specific concept.
- ❖ Tracking learning progress and find out areas of difficulty among students.
- ❖ Determine the minimum level of learning attainment and the level of mastery attained by students in the context of any given learning task.

Limitations of CRT

- ❖ CRT tells only whether a learner has reached proficiency level in a given task or not. It does not say what are the strength areas of the performance or weakness of performance.

- ❖ CRT doesn't allow for comparing the performance of examinees in a specific location with national norms.
- ❖ There is probability that the test scores of individual students may manifest incorrect standards if the standards are too challenging, too easy or loosely verbalized.
- ❖ The development process is difficult and time-consuming.

In the light of uses and limitations of NRT and CRT, we may conclude that both have an important place in educational testing. Both these tests can serve basic functions of evaluation in classroom.

Difference between NRT and CRT

The following differences between NRT and CRT are adapted from Popham, J. W. (1975). Educational Evaluation.

Dimensions NRT CRT

Purpose To rank and compare each student's performance in relation to their peers in a broad knowledge domains. To determine whether each student has achieved specific concepts or skills.

Dimensions	NRT	CRT
Purpose	To rank and compare each student's performance in relation to their peers in a broad knowledge domains.	To determine whether each student has achieved specific concepts or skills.
Content	Assesses a wide range of skills drawn from diverse textbooks, syllabi, and the expertise of curriculum specialists.	Assesses particular skills that constitute a specified curriculum, as determined by educators and curriculum experts.
Item characteristics	Items vary in difficulty level. Items that distinguish between low and high performers are chosen.	The items used to assess any specific ability are all of the same difficulty level.
Interpretation of scores	Test scores are often expressed as percentiles, stanines, standard scores, or grade equivalents, allowing test-takers to be compared to a norming group.	It focus on whether a test-taker has met a predefined standard or criterion, often indicating mastery or non-mastery.

Conclusion

Norm Referenced Test and Criterion Referenced Test represent two fundamentally different approaches to educational assessment, each offering distinct advantages and serving unique purposes in the realm of education. NRT, as its name implies, involves comparing a student's performance to that of a larger group, often their peers, and is primarily used for ranking and categorizing individuals. This approach provides a relative perspective on a student's achievement and typically yields scores like percentiles or grade-equivalent scores. On the other hand, CRT focuses on assessing whether a student has attained specific, predetermined learning objectives or skills. It is not concerned with how a student compares to others but rather whether they have mastered the defined criteria. CRT scores are absolute in nature, reflecting a student's proficiency in the subject matter. Both NRT and CRT have their strengths and limitations, and the choice between them should align with the assessment's goals and the information educators seek to glean. Ultimately, understanding these assessment approaches is vital for educators to make right decisions about how to best evaluate and support student learning effectively.

References

Darwesh, A. J. A., & Yahya, E. F. (2009). Norm-Referenced Tests vs Criterion-Referenced Tests. *Al-Adab Journal*, 1(89), 124-139.

Gronlund, N. E., & Linn, R. (1990). Measurement and Evaluation in Teaching (6th Ed.). *Macmillan Publication.*

IGNOU. (2000). Educational Evaluation (ES-333, B.Ed.), Block-1, Unit-2 '*Approaches to Evaluation*' (pp. 20-33). New Delhi: IGNOU.

Popham, J. W. (1975). *Educational Evaluation. Prentice-Hall International.*

Popham, J. W., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6(1), 1–9. https://doi.org/10.1111/j.1745-3984.1969.tb00654.x

12

Norm-Referenced Test and Criterion-Referenced Test

Susmita Roy

Abstract

The methodical, scientific process of assessing how well students have met learning objectives is called evaluation. "It is now agreed that evaluation, a continuous process, forms an integral part of the total system of education," states The Kothari Commission (1964–1966). There exist various test types that are utilized to achieve the ideal evaluation or value judgment. Among the tests that teachers use to assess the quality of the student's learning experience through teaching-learning activities are the Norm-Reference Test (NRT) and the Criterion Reference Test (CRT). When the results of an assessment are analyzed and judged against a specific standard to measure and evaluate one or more students' performance, such assessment is called Norm-Based assessment. On the other hand Criterion Based assessment in general, any type of test has several objectives. Tests are determined by the direction of objectives.

Keywords : *Norm-Reference Test, Criterion-Reference Test, Evaluation , Systematic.*

Introduction

Teachers use to measure the quality of the learning experience that students gain through teaching-learning activities are Norm-Reference Test (NRT) and Criterion Reference Test (CRT). These two aspirations are briefly discussed below:

Norm-Referenced Test or NRT

Comparing a test taker's score to that of other test takers or groups is a method known as the "Norm-reference test" that can be used to meaningfully interpret test results. In essence, it's a competitive assessment method where information is given based on how well a student performs in relation to other students. It isn't evaluated based on how well a certain student has understood the text's material or how well he has done so in relation to predefined goals. A group number is contrasted with the common value or bow. A student's relative standing within the group is referred to as their norm (Brown, F. 1976).

The main purpose of NRT is to identify a student's specific position within his or her group without using any predetermined criteria. But the norm used in this marking is determined in various ways—sometimes the average of marks obtained by all the students in the group, sometimes the absolute standard of aspiration. The results of such assessment are used to mark the students' position or rank in the group such as first, second, third etc. or to indicate their grade or level. Students' proficiency is thus demonstrated at the regional or national level. Such assessment is conducted in entrance exams for various courses or for promotion in the workplace (Alderson, JC, Clapham, D & Wall, D. 1995).

For norm reference assessment, all of our exams—in class, in public, and on standardized tests—are utilized. When a class member is asked, "Who is the best student?" Who started first? Who received the lowest grade? Then, assessments based on norms are created to address those queries. Additionally, we assess an individual's performance in relation to that of the group as a whole. For instance, competitive exams like the UGC NET examination, aptitude tests, and IQ tests (Corbett, H.D. & Wilson, B.L. 1991).

Advantages

Standardized Results

Norm-reference exams are uniformly and consistently administered using a standardized format. It makes sure that every test-taker receives the same guidelines and conditions, which lessens the chance that any unrelated factors could affect their results. Because test takers' scores are determined by identical standards, it also makes it possible to compare them with one another (Corbett, H.D. & Wilson, B.L. 1991).

Relative Performance

Rather than measuring absolute performance, NRTs are used to measure relative performance. In this context, test takers are evaluated not just on their scores but also on how well they perform relative to their peers in fields like education and employment.

Large Sample Size

Because NRTs are made to be given to big groups of people, they can yield a wealth of data that can help guide future research and policy choices. Additionally, by using this data, trends and patterns in performance can be found, which can help identify areas that might benefit from more resources or support.

Disadvantages

The Norm-Referenced Test has the following disadvantages,

Limited scope

Norm-reference tests may not give a complete picture of a person's abilities because they are meant to measure a particular set of skills or knowledge. This can be troublesome in domains like education, where a more comprehensive method of assessment might be required.

One-Size-Fits-All Approach

NRTs are not adapted to each person's needs or abilities because they are meant to be given to large groups of people. This test may be too simple or too complex for some test takers, which could have an impact on their scores and the validity of the findings.

Reliance on Norm Group

The dependability of norm-referenced tests is contingent upon the norm group they are grounded in. The results could be deemed biased or inaccurate if the norm group is not representative of the population under investigation. Additionally, norm groups may become out of date over time, which may have an impact on how accurate test results are (Corbett, H.D. & Wilson, B.L. 1991).

Lack of Contextual Information

Contextual details such as a test taker's history, experience, or motivation are not provided by NRT. This outcome can be very important in figuring out why a student did not do well on a given test.

Criterion Referenced Test

Comparing students' results in terms of a set of criteria is a second method for interpreting test results meaningfully. We refer to this process as the Criterion Reference Test. The first person to clarify this was Professor Robert Glasser. He said that using the Criterion Reference Test, one could determine the general caliber of that specific student's proficiency. The Criterion Referenced Test has a clearly defined standard of proficiency that serves as the basis for determining the student's overall level of proficiency.

The Popham and Husek (1969) definition is stated. "A CRT is one that is used to determine the level of success of learner." In schools, CRT are used to assess particular skills and knowledge that students have probably acquired. This establishes the degree to which they have mastered a standard. With CRT, educators can evaluate how they can support students' development in particular areas.

Advantages

The Criterion-Referenced Test has the following advantages,

Specificity

In CRT there is a definite goal that's being evaluated. A single item corresponds to a particular aim. There is no ambiguity about what is being tested on and it also makes the test reliable.

1. Topic Expertise

Pupils who monitor their progress on Criterion-Referenced Tests perform well academically. Test items can be tailored to match particular needs, and CRT can be used to gauge students' comprehension and knowledge of the subject. When it comes to giving indicative reports of students' overall progress, CRT makes very helpful measurements possible.

2.Controlled on a Local Level

Most of the time, Criterion-Referenced Test happen in the classroom, so the teacher can quickly determine whether the students have met the standard and plan for next course of action (Brown, DH. 2003).

3. Revision

This aspiration offers the chance to make the required adjustments, improvements, and modifications to the way text content is presented as well as instructional strategies and tactics. Even when using teaching lights, the instructor takes the required precautions. CRTs actively participate in identifying the learning objectives that a student has not met. After that, the instructor acts appropriately to accomplish those objectives (Bond, L. 1996).

4. Individuality

In Criterion-Referenced Test, students are assessed independently. Their performance is not compared to that of another students.

Disadvantages

The Criterion-Referenced Test has the following disadvantages,

1. Lack of Generalization

In Criterion-Referenced Test the results can not be generalized in a broad area. Because in this test individual is focusing on himself or herself to achieve the criteria. So results cannot be generalized beyond a certain point (Anastasi, A. 1988).

2. Expensive

Here some tests can be time consuming and expensive. So this is one of the disadvantages of CRT.

3. Prior Access

Students may take undue advantage by gaining prior access.

4. Lack of Efficient Leadership

Ineffective leadership can lead to issues. Therefore, CRT requires effective leadership and teamwork. For example, a school may not be able to create learner-centered assessments if they are creating assessments for special education students without the assistance of professionals with the necessary training (Bachman, LF. 1995).

Difference between Norm-Referenced Test and Criterion-Referenced Test

NORM-REFERENCED TEST	CRITERION-REFERENCED TEST
1. In NRT the objective is to measure students' performance and compare them with other students in the classroom.	1. In CRT the objective is to provide a picture of overall development by setting standards of student proficiency with predetermined criteria.
2.The plan of study is adopted according to the general description of the subject matter.	2. Aspirational planning is undertaken accordingly by identifying the desired behaviors of students in terms of instructional objectives.
3. The difficulty level of posts is usually between 0.3 and 0.7.	3. The level of difficulty depends on the nature of the learning task.
4. In NRT, type of interpretation is relative (percentile or grade equivalent).	4. In CRT, type of interpretation is absolute (percentage).
5. The result is declared with the help of original marks, percentile, standard marks etc.	5. Posts are selected according to the degree of difficulty and distinction of the post.
6. NRT usually focus on summative evaluation.	6. CRT focus on formative, diagnostic evaluation.
7. For instance, Aptitude test, IQ test, Competitive exam etc.	7. For instance, Class test, Competency based test, Driving tests etc.
8. It affects others performance.	8. It doesn't affect others performance
9. Here the judgement works and how well the teaching learning objectives are followed in the aspirational position.	9. Here the focus is student learning rather than scoring well.
10. Students have little or no idea what content to expect in questions.	10. Students know exactly what content to expect in test questions.

Conclusion

Based on the results of how well the student or aspirant has achieved these objectives, educational measurement or evaluation can be used to know or understand the actual standing of the student and based on that to provide future direction or advice to the student. The objective of the study is to define norm-referenced tests and criterion-referenced tests and to find out the difference between them.

References

Alderson, JC, Clapham, D & Wall, D. (1995). *Language Test Construction and Evaluation*. New York: Cambridge University Anastasi, A. (1988). *Psychological Testing*. *New York, New York*: MacMillan Publishing Company.

Bachman, LF. (1995). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bond, L. (1996). *Norm and Criterion-referenced Testing: Practical Assessment: Research &* Evaluation (O). (http://www.ericae.net/pare/getvn.asp?v=5&n=2 (Accessed 10 November 2014).

Brown, D.H. (2003). Language Assessment: *Principles and Classroom Practices*. San Francisco: Longman.

Brown, F. (1976). *Principles of Educational and Psychological Testing*. 2nd edition. New York: Holt, Rinehart and Winston.

Corbett, H.D. & Wilson, B.L. (1991). *Testing, Reform and Rebellion. Norwood, New Jersey: Ablex* Publishing Company.

Popham, J. W., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6(1), 1–9. https://doi.org/10.1111/j.1745-3984.1969.tb00654.xRomberg,

Wilson, L. & Mamphono Khaketla (1991). "The Alignment of Six Standardized Tests with NCTM Standards", *an unpublished paper, University of Wisconsin Madison.* In Jean Kerr Stenmark (ed; 1991). Mathematics.

Unit: 5
Innovation in Evaluation

13

Unconventional Assessment: Conquering Open Book Examination

Sonali Giri & Dr. Sunil Kumar Baskey

Abstract

Traditional closed-book examinations have been the norm for centuries, testing students' knowledge and understanding of course material within a fixed time frame. At present, a revolutionary approach known as the Open Book Examination has gained significant attention and adoption in various educational institutions. The OBE differs fundamentally from its conventional counterpart in that it allows students to refer to course materials, textbooks, and notes during the examination. The purpose of this paper is to explore the concept of Open Book Examinations in greater details like process, rules-regulations, delving into its benefits and drawbacks, as well as examining its impact on student performance and learning outcomes. This paper will consider the practical implications of implementing OBEs and their potential to address the limitations of traditional examinations, as well as the concerns raised by skeptics. To provide a comprehensive analysis, a review of existing literature on Open Book Examinations will be conducted, examining studies and academic papers from reputable sources. Interviews with educators and students who have experienced OBEs will be included to gain valuable insights into their perspectives and experiences. It contributes to the ongoing conversation about the evolution of assessment methodologies and the potential transformation of traditional examination practices. By exploring the OBE paradigm, this paper wants to seek uncover new possibilities for enhancing students' learning experiences and promoting a more comprehensive evaluation of their knowledge and skills.

***Keywords**: Unconventional Assessment, Traditional Examination, Open Book Examination, Students' Learning Experience and Role of Teachers.*

Introduction

Open book examination is a unique academic challenge that requires a distinct set of strategies and skills. Unlike traditional closed-book examinations, where students are expected to rely solely on their memorized knowledge, open book examinations allow students to access their study materials during the test. However, this apparent advantage can quickly become a double-edged sword if not approached strategically. An open book examination is a type of assessment in which students are permitted to use reference materials, such as textbooks, notes, or other authorized resources, during the examination. Unlike traditional closed-book examinations, where students must rely solely on their memory and understanding of the subject matter, open book examinations provide test-takers with access to external information to aid them in answering questions or solving problems.

The concept behind open book examinations is to assess a student's ability to apply their knowledge, critical thinking skills, and problem-solving abilities rather than their rote memorization. These examinations are designed to test a student's comprehension, analysis, and synthesis of the subject matter, as well as their ability to locate and utilize relevant information effectively.

Open book examinations have a history that spans several decades, and their origins can be traced back to innovative educational practices and pedagogical philosophies. While there is not a single specific reference point for the inception of open book examinations, various developments and educational thinkers have contributed to their evolution.

One of the early influences on open book examinations can be attributed to the progressive education movement in the early 20th century. Figures like John Dewey emphasized learning by doing and critical thinking over rote memorization. Dewey's ideas laid the foundation for the broader shift towards student-centered learning and assessments that emphasize understanding and application of knowledge.

Another important influence on open book examinations can be found in the work of Ralph Tyler, an influential educational psychologist. Tyler's work on curriculum and assessment stressed the importance of aligning assessment methods with educational objectives. This perspective encouraged educators to design assessments that measured higher-order thinking skills rather than mere recall of facts.

Objectives of the Study

The study is mainly focused on following objectives.

1. To study about the concept of unconventional assessment.

2. To study about the features and types of open book examination.
3. To study the differences between traditional examination and open book examination.
4. To study about the learning experiences and role of teachers in open book examination.
5. To study about the benefits and challenges of open book examination.

Concept of Unconventional Assessment

To know the about unconventional assessment, at first, we have to know about the traditional or conventional assessment. Traditional assessment refers to the conventional methods and practices used to evaluate students' knowledge, skills, and abilities in educational settings. These assessments have been widely employed for many years and often involve standardized tests, written exams, quizzes, and other structured assessments. Generally, it emphasized to assess the memory-based knowledge.

Unconventional assessment, a transformative paradigm in education, represents a departure from traditional testing methods, aiming to assess skills, knowledge, and capabilities in innovative ways. Unlike traditional examinations, unconventional assessments focus on real-world applications, encouraging critical thinking, problem-solving, and creativity. These methods, such as project-based evaluations, portfolios, and simulations, offer a more comprehensive understanding of an individual's abilities.

One key advantage of unconventional assessments lies in their ability to capture practical skills and talents that conventional exams often miss. For instance, in project-based assessments, Open book examinations, students tackle real-life problems, honing their analytical skills and collaborative abilities. Portfolios, on the other hand, provide a holistic view of a person's progress, displaying a journey of learning rather than just a snapshot of knowledge at a single point in time.

Moreover, these assessments promote personalized learning. They recognize that each student learns differently and excels in various areas, allowing for tailored evaluation methods. Peer assessments foster teamwork and communication skills, emphasizing the importance of interpersonal abilities in addition to academic knowledge.

Furthermore, unconventional assessments prepare individuals for the complexities of the modern world. They mirror professional scenarios, where problem-solving, adaptability, and creativity are highly valued. As industries evolve, these assessments ensure that students are equipped with skills beyond rote memorization, preparing them for diverse challenges in the workforce.

However, implementing unconventional assessments requires thoughtful planning and innovative teaching methods. Educators must design assessments that are meaningful, engaging, and aligned with learning objectives. Additionally, they must ensure fairness and consistency in evaluation while encouraging creativity and individuality. Open Book Examination becomes one of the popular unconventional assessments.

Characteristics of open book examination

Open-book examination is a type of assessment in which students are allowed to refer to textbooks, notes, or other reference materials during the examination. Unlike traditional closed-book exams, where students are expected to rely solely on their memory and understanding of the subject matter, open-book examinations test a student's ability to apply and synthesize information from various sources effectively. The important features of open book examinations are follows:

1. **Resource Access:** Students are typically allowed to bring specific reference materials, such as textbooks, class notes, and approved online resources, into the exam room or access them digitally during an online assessment.
2. **Complex Questions:** Open book examinations often feature more complex and application-based questions that require a deeper understanding of the subject matter. These questions may require students to apply concepts to real-world scenarios or analyze and synthesize information from multiple sources.
3. **Time Management:** Although students have access to reference materials, open book examinations are often timed, which requires effective time management skills. Students must balance the need to search for information with the need to answer questions efficiently.
4. **Critical Thinking**: Success in open book examinations depends on a student's ability to critically evaluate and select the most relevant information from their resources to address the specific requirements of each question.
5. **Organization**: Effective organization of reference materials is essential. Students need to know where to find relevant information quickly, which can be achieved through well-structured notes or bookmarks in textbooks and digital resources.
6. **Ethical Considerations**: Students are typically expected to adhere to ethical guidelines during open book examinations, such as not collaborating with others or using unauthorized resources. Honesty and academic integrity are paramount.

7. Open book examinations are commonly used in courses that **emphasize problem-solving, analysis, and the application of knowledge** rather than mere memorization. While they provide students with a safety net of reference materials, they also challenge them to think critically, make connections between concepts, and demonstrate a deep understanding of the subject matter.

Types of Open Book Examination

Open book examinations come in various formats to assess different aspects of a student's knowledge and skills. Here are some common types of open book examinations:

❖ **Traditional Open Book Examination:**

In this type of open book exam, students are allowed to bring their textbooks, class notes, and other relevant materials into the exam room. They can use these resources to answer questions on the exam paper.

❖ **Take-Home Open Book Examination:**

This type of open book exam is typically completed outside of the classroom environment. Students receive the exam questions and materials, and they are given a specified amount of time (usually more than a traditional exam) to complete and submit their responses, often within a day or more.

❖ **Open Book Online Examination:**

With the rise of online learning, open book examinations can also be administered through digital platforms. Students can access their resources electronically while taking the exam online.

❖ **Hybrid or Partial Open Book Examination**:

In this format, students are allowed to bring certain specified materials, such as a formula sheet or a set of notes, while other resources are prohibited. This type combines elements of both open book and closed book examinations.

❖ **Scenario-Based Open Book Examination:**

Students are provided with a complex scenario or problem, and they are required to use their resources to analyze the situation and develop a solution or response. These examinations often emphasize critical thinking and problem-solving skills.

Where the Open Book Examination differs from Traditional Examination

❖ **Resource Access:**

According to (Hassel, 2018),in an open book exam, students are explicitly allowed to bring textbooks, notes, and other reference materials into the exam

room or access them digitally during online assessments. In traditional examination, students are not allowed to bring such types of reference materials.

❖ **Question Types:**

When the questions come about the question types of Open book examinations. Paulson (2003) said that the open book examination often features complex, application-based questions that require students to analyze, synthesize, and apply knowledge to real-world scenarios. These types of questions generally avoid to know the memory-based answers or data-based answers. In the other hand traditional examinations include a mix of memory-based questions and those that require critical thinking, but the emphasis is often on assessing a student's recall of facts and concepts.

❖ **Time Management**

Open Book Examination Provides more time flexibility, as students can look up information, which might lead to longer completion times but in traditional Examination, it Emphasizes time management skills, as students need to answer questions within a limited timeframe.

❖ **Critical Thinking**

Open Book Examination emphasizes critical thinking skills, as students need to analyze and apply information from various sources but traditional Examination focuses on recalling information and demonstrating knowledge.

❖ **Stress Levels**

Open Book Examinations generally, reduces stress, as students have access to reference materials, allowing them to focus more on understanding concepts. But in the other hand traditional examination tends to create higher stress levels due to the pressure of recalling information.

❖ **Cheating and Plagiarism**

Open Book Examinations are more challenging to prevent cheating, as students can collaborate or copy information without detection. But the traditional examinations are easier to monitor and prevent cheating, although it still occurs through various means.

❖ **Learning Experiences**

Open Book Examination promotes deeper understanding and application of knowledge. But the Traditional Examination mostly encourages rote learning and memorization.

Focus Area of Learning Experiences in Open Book Examination

Open book examinations can provide a unique learning experience for students, encouraging a deeper understanding of the subject matter and the

development of critical thinking skills. Here are some key points to consider regarding students' learning experiences in open book examinations:

1. **Enhanced Critical Thinking**: Open book examinations require students to think critically and apply their knowledge rather than simply memorizing facts. They must analyze the information at hand and decide which references are relevant to answer the questions effectively.
2. **Promotes Independent Learning**: Open book examinations encourage students to become more self-reliant learners. They must prepare by organizing their materials, making study aids, and understanding the subject thoroughly, which fosters independent learning skills.
3. **Emphasis on Resource Management**: Students learn to manage their resources effectively, selecting and using references efficiently during the examination. This skill can be valuable in real-world situations where professionals need to find and apply information quickly.
4. **Application of Concepts**: Open book examinations assess a student's ability to apply concepts to real-world scenarios. Students must not only find the relevant information but also apply it correctly to solve problems or answer questions.
5. **Deeper Understanding**: Preparing for open book examinations often leads to a deeper understanding of the subject matter. Students need to understand the material well enough to know which resources to use and how to apply them.
6. **Less Stressful**: Some students find open book examinations less stressful than closed book examinations because they have access to their notes and textbooks. This can reduce test anxiety, allowing them to perform better.
7. **Real-World Relevance**: Open book examinations mimic real-world scenarios where professionals often have access to reference materials. Students can see the relevance of their learning to future careers.
8. **Improved Note-Taking Skills**: Preparing for open book examinations can enhance students' note-taking skills. They learn to create effective study aids, such as summary sheets, concept maps, and annotated references.
9. **Time Management**: Open book examinations require effective time management during the test. Students must allocate their time wisely, balancing between searching for information and answering questions.

10. **Cautions About Over-Reliance**: One potential downside is that some students may become overly reliant on their reference materials, neglecting a deeper understanding of the subject. To mitigate this, educators can design questions that require critical thinking rather than simple information retrieval.
11. **Academic Integrity**: Educators need to emphasize the importance of academic integrity during open book examinations, encouraging students to cite sources properly and avoid plagiarism.
12. **Continuous Learning**: Open book examinations promote continuous learning. Students often find that the materials they create for these examinations can serve as valuable study aids for future courses or professional contexts.

open book examinations can provide a rich learning experience for students by promoting critical thinking, independent learning, and the application of knowledge in real-world contexts. However, educators must strike a balance to ensure that students do not rely solely on their reference materials and emphasize the importance of ethical practices in academic work.

Role of Teachers in Open Book Examination

Teachers play a crucial role in the success and effectiveness of open book examinations. Their responsibilities extend beyond designing the test questions to check the answer scripts. Few important aspects of the role of teachers in open book examinations are following:

1. **Question Design**: Teachers must create questions that require higher-order thinking skills. These questions should go beyond rote memorization and encourage students to analyze, synthesize, and apply the knowledge they have gained. Questions should be clear, concise, and relevant to the learning objectives.
2. **Providing Guidelines**: Teachers should provide clear guidelines and instructions on how the open book exam will be conducted. This includes specifying which reference materials are allowed, which are not and how students should cite sources, and if there are any time limits and other regulations.
3. **Selecting Resources**: Depending on the nature of the examination, teachers may need to select or approve specific reference materials that students can use during the test. This ensures that students have access to appropriate resources and that the exam remains fair.
4. **Monitoring Integrity**: Teachers should take measures to prevent academic dishonesty. This includes educating students about proper

citation practices, monitoring the exam environment (if it is an in-person examination), and using plagiarism-detection software for submitted work.

5. **Feedback and Support**: After the examination, teachers should provide feedback on students' performance. This feedback can help students understand their strengths and areas for improvement. Teachers can also offer support and clarification on any misconceptions that may have arisen during the examination.
6. **Promoting Learning**: Open book examinations should be viewed as a learning opportunity. Teachers can encourage students to approach these examinations as a chance to deepen their understanding of the subject matter. They can provide guidance on how to prepare effectively, such as creating study aids and organizing reference materials.
7. **Monitoring Time Management**: Teachers should ensure that students are managing their time effectively during the examination. This might involve setting time limits for different sections or reminding students of time remaining.
8. **Flexibility and Adaptation**: Teachers should be prepared to adapt to unexpected situations. For instance, if students encounter technical difficulties during an online open book exam, teachers should have contingency plans in place.
9. **Reflecting on Assessment**: After the open book examination, teachers should reflect on its effectiveness as an assessment method. They can consider whether the questions aligned with the learning objectives, whether the resources provided were helpful, and if any adjustments are needed for future assessments.
10. **Continuous Improvement**: Teachers can use feedback from students and their own observations to continuously improve the open book examination process. This might involve refining the guidelines, selecting more relevant reference materials, or modifying the types of questions asked.
11. **Supporting Student Well-Being:** Recognizing that open book examinations can be less stressful for some students, teachers can offer support for students who may still experience anxiety or other challenges related to assessment.

In essence, teachers play a multifaceted role in open book examinations. They are responsible for designing meaningful assessments, ensuring fairness

and academic integrity, providing guidance and support, and using these assessments as tools for promoting learning and growth among their students.

Benefits of Open Book examinations

- **Encourages Deeper Understanding**: According to Bain, K. (2004), Open book examinations promote a deeper understanding of the subject matter as students are encouraged to analyze and apply knowledge rather than rely on rote memorization.
- **Enhances Critical Thinking:** By requiring students to access and apply information from various sources, open book examinations foster critical thinking and problem-solving skills.
- **Reflects Real-World Scenarios**: Open book examinations align with real-world situations where professionals have access to references and resources, preparing students for practical challenges.
- **Reduces Test Anxiety**: Students often experience less test anxiety in open book examinations, as they know they can refer to their notes and textbooks. This can create a more relaxed testing environment.
- **Promotes Responsible Resource Management:** Open book examinations teach students how to manage and use resources effectively, a valuable skill for lifelong learning and professional success.
- **Focuses on Application:** Open book examinations emphasize the application of knowledge to real-world scenarios, ensuring that students can use what they have learned in practical contexts.
- **Provides Immediate Feedback:** Open book examinations often allow educators to provide more insightful feedback, as questions can be designed to assess comprehension and application rather than simple recall.
- **Supports Diverse Learning Styles**: Open book examinations can accommodate a variety of learning styles and abilities, allowing students with different strengths to demonstrate their understanding effectively.
- **Enhances Time Management Skills:** Students must manage their time effectively during open book examinations, balancing the need to locate information with completing the exam within a specified time frame.
- **Encourages Academic Integrity:** Trusting students with open book examinations encourages responsible academic behavior, promoting ethical conduct in learning.

These benefits underscore the value of open book examinations in promoting meaningful learning experiences and preparing students for the complexities of education and professional life.

Challenges of Open book Exam

Open book examinations offer several benefits, as discussed earlier, but they also come with their set of challenges. These challenges are important to consider when implementing open book assessments. Here are some common challenges of open book examinations:

1.Risk of Over-Reliance on Resources:

Students may become overly reliant on their resources and fail to develop a deep understanding of the material.

2. Academic Integrity Concerns:

Ensuring academic integrity can be challenging, as there is a risk of students plagiarizing or collaborating during open book examinations.

3. Complex Question:

Crafting questions that require higher-order thinking and prevent students from simply looking up answers can be difficult.

4.Time Management Issues

Students struggle with time management during open book examinations, spending too much time searching for information and not enough time answering questions.

5. Difficulty in Grading:

Grading in open book examinations can be time-consuming, as it often involves assessing the quality of reasoning and application of knowledge.

6. Lack of Standardization

Ensuring consistency in open book examinations across different instructors or courses can be challenging, leading to potential disparities in assessment difficulty.

Resource Overload:

Students may bring an excessive number of resources into the exam, making it difficult to locate necessary information quickly.

7. Test Anxiety for Some Student:

While open book examinations can reduce test anxiety for some, they may increase anxiety for others who feel overwhelmed by the volume of resources available.

8. Logistical Challenges for Large Classes:

Managing and proctoring open book examinations, especially in large classes, can be logistically challenging.

9. Limited Skill Assessment:

Depending solely on open book examinations may limit the assessment of certain skills, such as memory and quick recall, which are necessary in some contexts.

It's important for educators to be aware of these challenges and address them through thoughtful design, clear guidelines, and a focus on assessing not just what students know but also how they apply their knowledge in meaningful ways.

Conclusion

Traditional closed-book examinations have long been the hallmark of assessing student knowledge and understanding. However, the landscape of education is evolving, and unconventional assessment methods are gaining traction. One such method is the open-book exam, which challenges students to demonstrate critical thinking and problem-solving skills in real-world contexts.

Open-book examinations represent a significant departure from the rote memorization often associated with closed-book tests. Instead of relying solely on memory, students are encouraged to delve into course materials, textbooks, and other resources. This approach promotes a deeper comprehension of the subject matter by emphasizing comprehension and application over memorization. A fundamental advantage of open-book examinations is their ability to foster higher-order thinking skills. Students are not merely expected to regurgitate information; they are tasked with analyzing, synthesizing, and evaluating information to address complex questions. This mirrors the demands of modern workplaces, where employees are expected to apply their knowledge creatively to solve multifaceted problems.Moreover, open-book examinations align with the concept of lifelong learning, acknowledging that knowledge is continually evolving. These assessments teach students not just what to learn but how to learn. This skill is invaluable in an era when adaptability and continuous self-improvement are essential in both academic and professional pursuits.

To maximize the effectiveness of open-book examinations, careful design is crucial. Thought-provoking questions should require students to think critically and apply concepts from their resources effectively. Time constraints can also be used judiciously to ensure students adequately prepare and do not become overly reliant on resource access during the assessment.

Research has shown that open-book examinations can lead to better long-term retention of knowledge compared to closed-book examinations (Smith & Smith, 2010). Moreover, these assessments have been found to encourage deeper engagement with course materials, promoting better understanding (Martinez & Munday, 2016).

In conclusion, unconventional assessments like open-book examinations represent a progressive approach to education. They acknowledge the changing nature of knowledge and emphasize the importance of practical application in the real world. By conquering the open-book exam, students not only demonstrate mastery of content but also their ability to navigate the vast sea of information, a skill that is increasingly critical in the dynamic landscape of the 21st century.

References

Angelo, T. A., & Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers*. Jossey-Bass.

Bain, K. (2004). *What the Best College Teachers Do*. Harvard University Press.

Banta, T. W., & Palomba, C. A. (2014). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. Jossey-Bass.

Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University*. McGraw-Hill Education.

Brookhart, Susan M. (2004). *Educational Assessment Knowledge and Skills for Teachers*. Educational Measurement: Issues and Practice, vol. 23, no. 1, pp. 23-27.

Brown, G. A., Bull, J., & Pendlebury, M. (1997). *Assessing Student Learning in Higher Education*. Routledge.Dewey, J. (1938). *Experience and Education*. Kappa Delta Pi.

Martinez, M. E., & Munday, K. C. (2016). *The impact of open-book testing on student learning in a large introductory biology course*. Journal of College Science Teaching, pp. 14-21.

McCsabe, D. L., Butterfield, K. D., & Treviño, L. K. (2006). *Academic Integrity in Honor Code and Non-Honor Code Environments: A Qualitative Investigation*. The Journal of Higher Education, 234-259.

McKechnie, W. J., & Svinicki, M. (2014). *McKechnie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers*. Cengage Learning.

Millis, B. J., & Cottell, P. G. (1998). *Cooperative Learning for Higher Education Faculty*. American Council on Education

Nilson, L. B. (2016). *Teaching at Its Best: A Research-Based Resource for College Instructors*. Jossey-Bass.

Nitsko, A. J., & Brookhart, S. M. (2014). *Educational Assessment of Students*. Pearson.
Popham, W. J. (2008). *Transformative Assessment*. ASCD.

Popham, W. James. (2002). *Modern Educational Measurement: Practical Guidelines for Educational Leaders*. Allyn & Bacons

Sambell, K., & McDowell, L. (1998). *The construction of the hidden curriculum: messages and meanings in the assessment of student learning*. Assessment & Evaluation in Higher Education, 391-402.

Schulte, M. (2009). *The Student EQ Edge: Emotional Intelligence and Your Academic and Personal Success*. Jossey-Bass.

Smith, D. C., & Smith, R. D. (2010). *The effects of open-book, closed-book, and cheat sheet examinations on student achievement in a computer science course*. Journal of Educational Computing Research, pp. 349-365.

Sutherland-Smith, W. (2008). *Plagiarism, the Internet, and Student Learning: Improving Academic Integrity*. Routledge.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.

14

Effectiveness of Open Book Examination

Dr. Shazia Hasnain

Abstract

Examinations have historically been the primary method of assessment in schools and institutions around the world. Examinations held in a traditional way have been criticised for their emphasis on rote learning, giving the least value to the critical understanding of the content. Marks and grade-oriented examinations have led to many students being afraid of exams as their concern becomes that performing poorly will cause them to embarrass themselves in front of family and friends. Open book examinations offer a fresh perspective on exams. An open book test allows students to consult their textbooks, class notes, or any other authorised source of information while responding to questions. It may also imply that the question is given to the students so that they can synthesise, compare, and assess knowledge. In this kind of test, the focus is on comprehension of basic concepts which calls for a thorough understanding of the subject matter. The current study aims to review and analyse the previous works on the effectiveness of open book examination.

Keywords: *open book, examination, effectiveness, review*

Introduction

The education system followed in India is examination oriented where it is compulsory for the students to pass the written examination to move to the next grade or semester. The students are expected to learn the content and write it in exams. Even the questions are such that require "recalling of facts" and answers are expected in "bookish language" (Ashri & Sahoo, 2021). If the questions are not all comprehensive, not adhering to Bloom's taxonomy then the examination becomes restricted to examining only the knowledge of

the students without checking their application and skills (Kaur, 2020). This kind of examination helps in identifying whether students are well aware of the content taught, but most of the time it encourages rote memorisation which in turn leads to stress in students. Even the National Curriculum Framework (2005, pg.71) said that the "termevaluation is associated with examination, stress, and anxiety." To reduce exam-related stress, NCF (2005) advised that scientists and educators collaborate to develop innovative methods of testing pupils. Recently, the National Education Policy (2020) has emphasized the need to change the examination system so that assessment is formative and "more competency-based, promotes learning and development for our students, and tests higher- order skills, such as analysis, critical thinking, and conceptual clarity." Considering the need for a change in the evaluation technique of students it is important to try out different evaluation techniques one such being open book examination (OBE).

Open book examination

A test that is open book allows students to respond to questions using any authorized source of knowledge, including their textbooks, class notes, or other notes. This distinguishes it from conventional examination procedures, which require pupils to write in the examination room without any materials. As it can assess students' high-level abilities such as thinking, problem-solving, and reasoning, OBE is increasingly becoming a crucial evaluation tool in education, particularly at the university level (Feller, 1994; Ioannidou, 1997 as cited in Mahmoudzadeh-Sagheb et al., 2015).

OBE improves the learning environment and aids students in comprehending and answering questions more effectively (Brightwell et al., 2004). Students study a variety of sources while getting ready for OBEs, including textbooks, classroom notes, online blogs, etc. As a result, they are able to connect the ideas and are inspired to learn in a more original way rather than through "rote learning" (Theophilides & Koutselini, 2000). Additionally, it accurately portrays real-world issues and the likely solutions that students will have to come up with (Brightwell et al.,2004). Open book test does not imply that it is very easy and does not require preparation. On the contrary, preparation is the key to the success of open book tests. With proper preparation and organized notes, it will be easier to analyse compare, or synthesize because the examination will have a time limit. Improper preparation can lead to a waste of time in looking for the materials. According to NCF (2005) devising an effective open book examination is a challenging task that needs to be taken up at all levels of school.

Open-book exams are one of the tools increasingly offered today in order to reduce test anxiety and to help students demonstrate their knowledge and abilities and correspondingly, have a more positive attitude towards exams.

The present study reviews previous works on open book examination and analyses how effective its results were in different countries and on different students.

Objectives of the study

1. To review and analyse the previous works on open book examinations.
2. To analyse the effectiveness of the open book examinations.

Review of Literature

There have been ample works on open book examinations some of which has been reviewed in this study. Empirical studies that were accessible were reviewed.

Swart & Sutherland (2014) conducted the study on senior electrical engineering students in Africa. Students from two different universities who were enrolled in distance learning and conventional residential classes took open- and closed-book exams. It was time lag research that was conducted from 2003 to 2012. Regular students performed better on open book tests than on closed book tests, according to a comparison of the two types of exams' results. However, this could not be stated for the pupils who studied remotely. With the aid of a questionnaire, the study discovered how the students felt about open-book exams. After doing a qualitative analysis, the students provided favourable feedback on open book exams, highlighting how it lessens stress and memorization.

Ashri and Sahoo (2021) evaluated the performance of commerce students in open- and closed-book exams. The survey was carried out among students at Delhi University in India in 2020, a year when the education system switched to online platforms because of the Covid 19 pandemic. The same exam was given in closed book mode during the study, then two days later it was given online in open book mode, with access to the internet and study materials. Open book exams yielded much better results than closed book tests. The authors of this study came to the conclusion that open book examination implementation in India is difficult and calls for a carefully thought-out curriculum.

Green et al. (2016) investigated the effectiveness of open-book versus closed-book tests. It was discovered that students fared better in open book exams than in closed book exams during a study of an introductory managerial accounting course for undergraduates in the United States. Pre-test quizzes with open and closed books, as well as a significant open book exam, were all included in the testing procedures used over the course of the semester. Open book pre-quizzes helped students prepare for the final test, and all students

agreed that open book exams should be used in the classroom. Additionally, they admitted that it improved their learning and that they did not understudy as they did not believe the open book test would be simple.In an open book exam, Das (2017) looked at how eighth-grade students performed in language classes and whether open and closed book test affected their exam anxiety. The research was carried out in Gujarat, India. The average student performance on the language exams (English, Hindi, and Gujarati) administered in open-book and closed-book forms was the same. Even the students taking tests in either mode experienced the same level of anxiety.

Rummer et al (2019) conducted a study on open and closed book testing on university students in Germany. The exam was conducted on cognitive psychology. It was an experimental study spanning eight weeks and it concluded that closed book testing was more effective than open book exams. Retrieval practice effects have been used to explain why closed book exams resulted in greater performance.

The study on Pakistani student teachers enrolled in a teacher education program was carried out by Khalid et al. in 2022. The preferences of fifteen student teachers for the open and closed book exams they had already taken were discovered using a structured interview. The participants praised the open book exam in their interview and stressed that it encouraged higher-order thinking rather than just memorizing facts. They claimed that considering OBE demands thinking and knowledge application, these exams shouldn't be viewed as simple.

Additionally, they noted that not all disciplines in higher education can benefit from open book exams because some require memorizing.

Discussion

The studies reviewed were mostly concerned with the way students performed in open book examination compared to closed book examination **(**Ashri & Sahoo, 2021; Das 2017; Green et al., 2016; Rummer et al., 2019) while some studies also analysed the perception and experience of students regarding open book examination (Khalid et al., 2022, Swart & Sutherland, 2014; Green et al., 2016). Most of the studies reported positive effects of open book exams. Also, students in the reviewed studies showed a preference for open book tests as they promote higher-order thinking and reduce anxiety. Notably, the students who gave their perceptions were mature college and university students so their experience with open book examinations can be considered to be important.

Of all the studies reviewed only one was conducted on school students (Das, 2017) while the rest were conducted on university and college students.

It is however a matter of further investigation if open book tests can be held for school students and whether open book questions can adequately assess all the subjects. Das (2017) conducted the study onlanguage subjects at the school level and the open book exam did not produce effective results. The rest of the reviewed studies on open book exams had university and college students belonging to different disciplines such as engineering (Swart & Sutherland, 2014), commerce (Ashri & Sahoo, 2021), accounting course (Green et al., 2019), psychology (Rummer et al, 2019), student teachers (Khalid et al, 2022). Research with open book tests have been conducted mostly on mature students. As Doghonadze & Demir (2013) pointed out that it is more suitable for university students.

Without adequate practice beforehand, students may not know how to use the limited exam time in searching for the answers in the resources that they will have with them.

However, Rummer et al. (2019) conducted both open and closed book examinations on university students and found closed book exams effective and the reason given was that students were used to retrieval practice. Thus, it can be said that the effectiveness of open book examination may depend on the subject and also the way the students are trained to answer questions.

Notably, the perception of teachers is not much explored in any of the reviewed studies. Teachers play an important role in conducting the exam and preparing the students on how to answer the questions with all the resources at hand. Even the onus of unbiased evaluation in open book exams lies on the teacher (Doghonadze & Demir, 2013). Therefore, their point of view on open book examination and the challenges that they may face in preparing the students for such exams must be investigated.

Advantages of implementing Open book examinations

From the literature reviewed the benefits of open book examination can be summarized as follows-

1. Open book examination encourages critical thinking and analysis.
2. Students' anxiety and stress are reduced as they have the necessary notes with them.
3. It can be conducted in online mode as well, and this can save time, paper, and resources.
4. It may aid in reducing unethical practices in the examination hall such as cheating.
5. Students may find open book examination interesting as it breaks the monotony of conventional examinations.

6. It will deter students from rote memorization.
7. Students may develop effective study techniques.
8. Open book examination is learner centred.

Conclusion

To sum up, the present study demonstrates that open book exams are employed at many universities and colleges around the world in a variety of fields and that they primarily yield excellent outcomes. Open book examination is well suited for subjects where critical and creative thinking is to be evaluated (Dhelawala & Saiyad, 2018). However, this does not imply that closed book exams must be completely avoided. A combination of the two may be useful for assessing the students and lowering their exam anxiety. As Das (2017) suggested constructivist classroom may be useful in developing students' analytical thinking abilities that will support open book exams. In addition, teachers need training and preparation before implementing the open book examination system as they need to frame questions that will evaluate the higher thinking skills of the students. Students must also receive instruction on how to respond in open-book exams. It cannot be simply copying from the notes and materials they will have with them. Teachers should practice the standard questions throughout class time to help students feel comfortable taking the open book format exams (Gray, 1994). Also, teachers have to be eager to adopt innovative methods of instruction and evaluation (Ashri & Sahoo, 2021; Green et al.,2016).

References

Ashri, D., & Sahoo, B. P. (2021). Open book examination and higher education during COVID-19: Case of University of Delhi. *Journal of Educational Technology Systems*, *50*(1), 73-86.

Brightwell, R., Daniel, J. H., & Stewart, A. (2004). Evaluation: Is an open book examination easier?. *Bioscience Education*, *3*(1), 1-10.

Das, J. (2017). A study on the open book examination in terms of achievement in language subjects and examination anxiety of standard VIII students. *International Journal of Research and Review*, *4*(5), 46-54.

Dhelawala, Z., and Saiyad, A. (2018) To Study the Impact of the Open Book Examination in Terms of the Achievement of Standard VIII Students in different subjects i.e. English, Science, S.S., Mathematics, Hindi, and Gujarati. *BPAS Journal of Education, 1 (1), 12-26*

Doghonadze, N., & Demir, H. (2013). Critical analysis of open-book exams for university students. In *ICERI2013 Proceedings* (pp. 4851-4857). IATED.

Feller M. (1994). Open-book testing and education for the future. *Studies in Educational Evaluation*. 20(2):235-8.

Govt. of India (2020). National Education Policy 2020.*https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_ 0.pdf*

Gray, T. G. (1994). Open book examinations. *Biochemical Education, 22*(3), 122-125. Green, S.

G., Ferrante, C. J., & Heppard, K. A. (2016). Using Open-Book Exams to Enhance Student Learning, Performance, and Motivation. *The Journal of Effective Teaching, 16*(1), 19-35

Ioannidou MK.(1997) Testing and life-long learning: Open-book and closed-book examination in a university course, *Studies in Educational Evaluation*, 23(2):131–9.

Kaur, S.(2020). Examination System in India: Prevailing Defects and Suggestions for Improvement. *Research Guru: Online Journal of Multidisciplinary Subjects, 13* (4). 190-192

Khalid, S., & Saeed, S., Sami, A., Alvi, G. (2022). Exploring the Students' Preferences for Open and Closed Book Examination to Assess Learning in Pakistan. *Indian Journal of Economics and Business, 21*(1).

Mahmoudzadeh-Sagheb, H., Heidari, Z., & Mohammadi, M. (2015). A Survey of the Students' Perspectives of Open-Book Examinations in the Histology/ Embryology Course. *Journal of Medical Education, 14*(1), 24-30.

National Council of Educational Research &Training NCERT, 2007. "National Curriculum Framework 2005," Working Papers id:1138, eSocialSciences. https://ncert.nic.in/pdf/nc-framework/nf2005-english.pdf

Rummer, R., Schweppe, J., & Schwede, A. (2019). Open-book versus closed-book tests in university classes: A field experiment. *Frontiers in psychology, 10*, 463.

Swart, A. J., & Sutherland, T. (2014). Student perspectives of open book versus closed book examinations a case study in satellite communication. *Int J Eng Educ, 30*(1), 210-7.

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open- book examination: A comparative analysis. *Educational Research and Evaluation, 6*(4), 379-393.

Websites

https://lsc.cornell.edu/how-to-study/studying-for-and-taking-exams/open-book-exams/

https://timesofindia.indiatimes.com/education/news/open-book-examination-what-does-it- mean/articleshow/76112795.cms

15

Learning to Learn: The Educational Value of Open Book Exams

Dr. Tapas Karmakar

Abstract

Open book exams have emerged as an innovative assessment method, offering a unique approach to evaluating students' knowledge and skills. Unlike traditional closed book exams that rely heavily on rote memorization, open book exams prioritize critical thinking, problem-solving, and comprehension. This abstract explores the educational value of open book exams and how they contribute to fostering a deeper understanding of the subject matter.The advantages of open book exams lie in their ability to encourage active learning and engagement among students. By providing access to reference materials during the assessment, students are motivated to explore and delve into the topic beyond surface-level memorization. As a result, they develop practical research and information retrieval skills, learning how to locate and utilize credible sources efficiently.Open book exams also shift the focus from mere memorization to the application of knowledge in real-world scenarios. Students are prompted to think critically, analyze information, and synthesize ideas to solve complex problems. This fosters a deeper appreciation for the subject matter and enhances students' ability to adapt their knowledge to practical situations.

To succeed in open book exams, students must develop effective study habits and time management skills. Organizing study materials and creating comprehensive notes become crucial for quick and efficient reference during the exam. Moreover, practicing with mock open book exams allows students to simulate real test conditions, identify areas of weakness, and improve their

overall preparedness.Despite their educational value, open book exams present challenges that students must navigate. Over-reliance on reference materials can hinder independent thinking and problem-solving. Therefore, striking a balance between using resources and relying on prior knowledge is essential to derive maximum benefit from open book assessments. Additionally, managing stress and time pressure during the exam is crucial to maintaining focus and composure.The impact of open book exams extends beyond academic settings. They instill self-directed learning skills, empowering students to become lifelong learners. By bridging the gap between theoretical knowledge and real-world application, open book exams prepare students for the challenges they may encounter in their future careers. In conclusion, open book exams offer a valuable educational experience, promoting critical thinking, practical research skills, and a deeper understanding of subjects. These assessments go beyond the limitations of closed book exams and encourage students to embrace continuous learning. By nurturing their abilities to think independently, analyze information, and apply knowledge, open book exams empower students to thrive in both academic and professional realms.

Keywords: *Educational Value, Open book exams, Memorization, Comprehensive, Prior Knowledge, Real-world application*

Introduction

Open book exams have emerged as a modern and innovative approach to assessing students' knowledge and skills in the realm of education. Unlike traditional closed book assessments, open book exams permit students access to reference materials during the examination process (Hockings, 2010).

Open book exams refer to a form of assessment where students are permitted to refer to textbooks, notes, and other reference materials during the examination. Unlike traditional closed book exams that necessitate memorization of facts and concepts, open book assessments allow students to access external resources, thus providing a more practical and applicable evaluation of their understanding.

The primary educational purpose of open book exams is to encourage deeper learning and comprehension among students. By allowing access to reference materials, these assessments shift the focus from rote memorization to critical thinking and problem-solving. Students are presented with the opportunity to explore the subject matter in greater depth, leading to a more profound understanding of the concepts and principles.

Moreover, open book exams are designed to mirror real-life scenarios where individuals are not expected to recall information from memory but rather apply their knowledge in practical situations. In professional settings,

employees frequently consult reference materials or digital resources to solve complex challenges, making open book assessments more relevant and applicable to real-world contexts (Price, Handley, Millar, & O'Donovan, 2010).

One of the key benefits of open book exams is the emphasis they place on cultivating critical thinking and problem-solving skills. Instead of relying solely on memorized information, students are encouraged to analyze and synthesize data from various sources to arrive at well-reasoned conclusions. These skills are highly valued in academia and the workforce, where individuals are expected to tackle complex issues, evaluate evidence, and propose innovative solutions.

Through open book exams, students learn to assess the credibility and relevance of information presented in reference materials, developing their abilities to discern reliable sources from unreliable ones. Furthermore, this evaluation method nurtures creativity, as students are prompted to connect disparate ideas and concepts from different sources to arrive at comprehensive answers.

Open book exams serve as a platform for promoting deeper learning, critical thinking, and problem-solving skills among students. By allowing access to reference materials, these assessments encourage a shift from rote memorization to understanding and applying knowledge. The educational value of open book exams lies in their ability to mirror real-world scenarios and prepare students for challenges beyond the classroom. As we delve into the following sections, we will explore the advantages, effective strategies, and lifelong learning benefits that open book exams offer to students and future professionals (Subramanian, 2019).

❖ Advantages of Open Book Exams

A. Encouraging Deeper Understanding of the Subject Matter:

Promoting Active Learning and Engagement: Open book exams motivate students to actively participate in their learning process. Instead of passively memorizing information, students are encouraged to interact with the subject matter, analyze concepts, and seek a comprehensive understanding of the material. This engagement fosters a deeper connection to the topics at hand, resulting in a more profound grasp of the subject matter.

Fostering Curiosity and Exploration: By providing access to reference materials, open book exams ignite students' curiosity to explore beyond the prescribed textbooks. The freedom to consult additional resources encourages learners to delve into related topics, pursue supplementary knowledge, and gain diverse perspectives. This exploration enriches their understanding and broadens their horizons (Price, Handley, Millar, & O'Donovan, 2010).

B. Emphasizing Comprehension Over Memorization:

Encouraging Students to Think Beyond Rote Memorization: Closed book exams often prioritize memorizing facts and formulas, which can lead to surface-level learning and limited retention. In contrast, open book exams place emphasis on comprehension. Students are encouraged to comprehend the underlying principles and connections between concepts, promoting a deeper and more meaningful understanding of the subject matter.

Focusing on Applying Knowledge to Real-World Scenarios: Open book exams mirror real-life situations where individuals have access to resources to solve problems. By encouraging students to apply their knowledge to practical scenarios, these assessments prepare them for challenges they may encounter beyond the classroom. This practical application enhances their ability to adapt their learning to real-world contexts (Rathi, & Baveja., 2021).

C. Developing Practical Research and Information Retrieval Skills

Learning to Locate Relevant Information Efficiently: Open book exams necessitate efficient information retrieval. Students must navigate through reference materials to find the most relevant information that addresses the exam questions. This skill of quickly locating pertinent data is valuable in academic and professional settings, where information is abundant and time is of the essence.

Assessing the Credibility and Reliability of Sources: In open book exams, students learn to critically evaluate the credibility of various sources they encounter. By distinguishing reliable and reputable resources from unreliable ones, students enhance their research skills and develop a discerning approach to information gathering. This skillset equips them to make informed decisions and draw evidence-based conclusions in their academic pursuits and future careers (Rathi, & Baveja., 2021.

❖ Building Effective Study Habits for Open Book Exams

A. Organizing Study Materials and Resources:

Creating Comprehensive Notes and Summaries: Preparing for open book exams begins with creating well-organized and detailed notes. While students have access to reference materials during the exam, having concise and comprehensive notes can significantly aid in quick information retrieval. Summarizing key concepts, formulas, and explanations allows for easier navigation and saves valuable time during the examination.

Structuring Resources for Easy Reference During Exams: To optimize efficiency during the exam, students should organize their reference materials in a systematic manner. Tabbing textbooks, color-coding sections, or using

sticky notes can help mark important pages or information. An easily accessible table of contents for digital resources can also streamline the process of finding relevant data swiftly.

B. Understanding the Syllabus and Learning Objectives

Aligning Study Efforts with Exam Expectations: Understanding the exam's format and content is vital in structuring one's study plan effectively. Students should thoroughly review the syllabus and exam guidelines to identify the topics that will be assessed. Aligning study efforts with the exam expectations ensures that valuable time is spent on relevant areas of the curriculum.

Identifying Key Concepts and Topics to Focus On: While open book exams allow access to resources, time constraints necessitate prioritization. Identifying the most critical concepts and topics helps students allocate their study time wisely. Focusing on understanding these core elements deeply equips students to answer questions confidently and efficiently during the exam.

C. Practicing with Mock Open Book Exams:

Simulating Exam Conditions to Enhance Preparedness: Conducting mock open book exams under timed conditions simulates the pressure and environment of the actual assessment. This practice helps students acclimate to accessing resources swiftly, managing time effectively, and applying knowledge in a time-sensitive setting. Regular mock exams build confidence and reduce anxiety when facing the real assessment (Rathi, & Baveja., 2021).

Identifying Weaknesses and Areas for Improvement: Analyzing performance in mock exams reveals areas of weakness or gaps in knowledge. By reviewing answers and understanding errors, students can focus their efforts on improving those specific areas. Seeking feedback from instructors or peers can also provide valuable insights for enhancing overall preparedness (Sharma, 2017).

Building effective study habits for open book exams requires a strategic and disciplined approach. Organizing study materials, understanding the exam's expectations, and engaging in regular practice are essential components of effective preparation. By developing a systematic study plan and honing critical skills like quick information retrieval and time management, students can maximize the benefits of open book exams and perform at their best during these assessments (Subramanian, 2019).

❖ Strategies for Success in Open Book Exams

A. Time Management During the Exam:

Setting Time Limits for Different Sections: In an open book exam, time management is crucial to ensure that all questions are attempted within the allocated time frame. Students should divide the total exam duration by the number of questions or sections to determine how much time to spend on each. Adhering to these time limits helps prevent spending too much time on a single question, allowing adequate time for other sections (Sharma, 2017).

Prioritizing Questions Based on Point Values and Complexity: Not all questions in an open book exam carry the same weightage. It is essential to identify questions with higher point values and complex problem-solving requirements. By addressing high-scoring questions first, students secure crucial marks and gain confidence to tackle other questions effectively.

B. Efficient Use of Reference Materials

Familiarizing with Permitted Resources Beforehand: Understanding the allowed reference materials is vital for efficient use during the exam. Students should review the guidelines provided by instructors and clarify any doubts about the resources permitted. Familiarity with the reference materials beforehand enables quick navigation during the exam (Sharma, 2017).

Annotating and Marking Essential Sections for Quick Access: During the preparation phase, students should annotate their textbooks and reference materials with essential notes and markers. Highlighting key sections and jotting down quick reminders help in locating relevant information swiftly during the exam, saving precious time and reducing confusion (Rathi, & Baveja., 2021).

C. Developing Critical Thinking and Analytical Skills:

Understanding Underlying Concepts and Principles: Open book exams require more than mere regurgitation of information. Students must grasp the fundamental concepts and principles behind the subject matter. Developing a deep understanding of the material empowers them to apply knowledge to various scenarios and answer questions with more insight.

Applying Knowledge to Solve Complex Problems: In open book exams, students encounter questions that necessitate critical analysis and problem-solving. Practicing with complex problems and case studies during preparation enhances their ability to think critically and devise innovative solutions.

D. Avoiding Plagiarism and Maintaining Academic Integrity

Properly Citing Sources and References: When incorporating information from external sources in answers, students must provide proper

citations and references. Adhering to the appropriate citation style (e.g., APA, MLA) ensures academic integrity and acknowledges the contributions of other authors.

Paraphrasing and Attributing Ideas Appropriately: To avoid plagiarism, students should practice paraphrasing ideas and concepts in their own words, giving due credit to the original sources. Proper attribution of ideas showcases ethical academic conduct and demonstrates an in-depth understanding of the subject matter (Sharma, 2017).

Succeeding in open book exams requires a combination of effective time management, efficient use of reference materials, honing critical thinking skills, and maintaining academic integrity. By following these strategies, students can approach open book exams with confidence, showcase their comprehensive understanding of the material, and excel in their assessments.

❖ Overcoming Challenges in Open Book Exams

A. Avoiding Over-Reliance on Resources:

Balancing the Use of Reference Materials and Prior Knowledge: One of the main challenges in open book exams is striking a balance between utilizing available resources and relying on prior knowledge. Students must avoid excessive dependence on reference materials, as it may lead to superficial answers and hinder critical thinking. By consolidating their understanding of key concepts through thorough preparation, students can leverage reference materials effectively to support their arguments and reinforce their responses.

Recognizing When to Refer to Resources and When to Think Independently: While open book exams grant access to resources, not every question requires external information. Students should be discerning in identifying questions that can be answered confidently based on their knowledge. This approach prevents the loss of valuable time searching for information that may not significantly contribute to their response.

B. Managing Stress and Time Pressure

Coping with the Pressure of Time Constraints: Open book exams may instill a false sense of security due to the availability of resources. However, time constraints can still be challenging. Students should remain mindful of the allocated time for each question and practice staying focused while under time pressure. Developing a time management strategy during mock exams helps build the skills necessary to handle the pressure effectively.

Staying Focused and Composed During the Exam: To combat stress, students must maintain their focus throughout the exam. Taking a few deep

breaths before starting and periodically during the assessment can help alleviate nervousness and enhance concentration. Staying composed enables students to think clearly, make rational decisions, and avoid mistakes that may result from panic(Nicol, & Macfarlane-Dick, 2006).

Open book exams come with their unique set of challenges, including avoiding over-reliance on resources and managing stress and time pressure. By developing a balanced approach to using reference materials and honing critical thinking skills, students can effectively navigate open book exams. Furthermore, coping with time constraints and maintaining composure during the assessment can lead to more accurate and well-reasoned responses. With adequate preparation and the adoption of effective coping mechanisms, students can overcome these challenges and maximize the benefits of open book exams to showcase their comprehensive understanding of the subject matter (Gikandi, Morrow, & Davis, 2011).

❖ The Role of Open Book Exams in Lifelong Learning

A. Encouraging Self-Directed Learning:

Developing Skills for Continuous Learning Beyond the Classroom: Open book exams promote self-directed learning, a critical skill in lifelong learning. When students have access to resources during assessments, they learn how to navigate through various references effectively. This process instills in them the ability to seek knowledge beyond traditional classroom settings, fostering a sense of autonomy in their learning journey (Pandey, 2020).

Promoting Curiosity and Seeking Knowledge Proactively: With the freedom to explore reference materials during exams, students become more curious and proactive in seeking knowledge. They are encouraged to delve deeper into subjects of interest, inquire beyond the prescribed curriculum, and pursue supplementary information. This curiosity-driven approach enriches their understanding and nurtures a lifelong passion for learning (Nicol, & Macfarlane-Dick, 2006).

B. Preparing Students for Real-World Problem-Solving

Bridging the Gap Between Theoretical Knowledge and Practical Application: Open book exams mirror real-life scenarios where individuals have access to information to solve problems. By requiring students to apply their knowledge and critical thinking skills to practical situations, these assessments bridge the gap between theoretical concepts and real-world application. This prepares students to face challenges in their future endeavors with confidence and adaptability (Subramanian, 2019).

Equipping Students with Research and Problem-Solving Skills Valued in Various Professions: In professional settings, employees often encounter complex problems that demand in-depth research and analysis. Open book exams develop students' research skills, teaching them to efficiently gather and evaluate information from multiple sources. Additionally, problem-solving exercises in open book exams nurture students' ability to think critically, make informed decisions, and propose innovative solutions—a highly sought-after skillset valued across various industries (Rathi, & Baveja., 2021).

Open book exams play a vital role in fostering lifelong learning skills and preparing students for real-world challenges. By encouraging self-directed learning and promoting curiosity, these assessments empower students to continue their educational journey beyond formal classrooms. Furthermore, the emphasis on applying knowledge to practical scenarios equips students with problem-solving and research skills that are highly valued in their future professions. As open book exams continue to evolve as a progressive assessment method, they become instrumental in nurturing well-rounded and adaptable individuals who are lifelong learners in every sense of the term (Biggs, & Tang, 2011).

Conclusion

Open book exams offer significant educational value that goes beyond conventional closed book assessments. Through this progressive evaluation method, students are encouraged to develop a deeper understanding of the subject matter, promoting active learning, critical thinking, and problem-solving skills. The freedom to access reference materials empowers learners to explore beyond the prescribed curriculum, fostering curiosity and self-directed learning. By bridging the gap between theoretical knowledge and real-world application, open book exams prepare students for the challenges they may encounter in their future academic pursuits and professional careers (Biggs, & Tang, 2011).

Embracing open book assessments provides students with valuable learning opportunities. It challenges them to think beyond mere memorization and instead focus on comprehending concepts, principles, and their practical implications. Encouraging students to engage actively with resources during exams cultivates a lifelong passion for learning, inspiring them to seek knowledge proactively and autonomously even outside the classroom (Subramanian, 2019).

Furthermore, open book exams underscore the importance of continuous learning and critical thinking in academic and professional growth. Navigating through reference materials during assessments fosters research skills and the ability to evaluate information critically. By applying knowledge to solve complex problems, students develop a versatile skillset that is highly valued in various professions (Bhatia, & Singh, 2019).

As students progress through their educational journey and beyond, open book exams serve as a gateway to a more enlightened and empowered learning experience. They provide a platform for students to showcase not just their memorization abilities but also their capacity to comprehend, analyze, and apply knowledge effectively. By mastering the art of open book exams, students unlock the path to becoming adaptable, resourceful, and lifelong learners who thrive in the ever-evolving landscape of education and the professional world (Rathi, & Baveja., 2021).

References

Bhatia, M., & Singh, S. (2019). Exploring the Impact of Open Book Examinations on Student Learning: A Case Study of Engineering Courses. *Journal of Higher Education Policy and Management*, 41(5), 566-580.

Biggs, J. B., & Tang, C. (2011). *Teaching for Quality Learning at University*: What the Student Does (4th ed.). McGraw-Hill Education.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online Formative Assessment in Higher Education: *A Review of the Literature. Computers & Education*, 57(4), 2333-2351.

Ghosh, A., & Chakraborty, S. (2018). Open Book Examinations: Perceptions and Practices in Indian Higher Education. *Journal of Studies in Education*, 8(2), 1-17.

Hockings, C. (2010). Inclusive Learning and Teaching in Higher Education: *A Synthesis of Research. Higher Education Academy*.

Nicol, D. J., & Macfarlane-Dick, D. (2006). *Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice*. Studies in Higher Education, 31(2), 199-218.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All That Effort, But What Is the Effect? *Assessment & Evaluation in Higher Education*, 35(3), 277-289.

Pandey, R. (2020). Open Book Examination: An Innovative Approach to Student Learning. *Journal of Educational Technology in Higher Education*, 17, 18.

Rust, C. (2002). *The Impact of Assessment on Student Learning: How Can the Research Literature Practically Help to Inform the Development of Departmental Assessment Strategies and Learner-Centred Assessment Cultures*? Active Learning in Higher Education, 3(2), 145-158.

Sharma, P. (2017). Promoting Higher-Order Thinking Skills Through Open Book Examination. *Journal of Indian Education*, 43(2), 19-34.

Subramanian, N. (2019). The Pedagogical Implications of Open Book Examination: A Study of Indian Higher Education. *Indian Journal of Higher Education*, 11(3), 58-71.

Rathi, R., & Baveja., S. (2021). Open Book Examinations: A Paradigm Shift in Assessment. *International Journal of Innovative Research and Practices*, 9(1), 23-27

Website

https://www.mizanurrmizan.info/what-is-open-book-exam-definition-importance-benefits-of-open-book-examination-and-its-challenges https://assiniboine.net/sites/default/files/documents/2020 03/how to create open book exams.pdf

16

Qualitative Analysis on Short Answer Type Test in reference to MCQ Type Test

Tania Bhadury & Prasanta Jana

Abstract

This article presents analysis on short answer type test in reference to MCQ type test in examination. Our memory and application skills are put to the test on short-answer questions. However one might be permitted to utilize punctuation marks and instances, it often requires students to compose a few paragraphs. Being able to provide straightforward responses to questions is key to performing effectively in short-answer tests. Students are required to respond to a given topic in writing on a short-answer test in a succinct and detailed manner, typically using whole sentences. These tests measure their capacity for synthesizing and using the knowledge they gained from the lessons, surveys, and conversations. In comparison with multiple choice questions (MCQs), short answer questions (SAQs) are intended to obtain responses directly from pupils. SAQs may be utilized to assess skills and comprehension, facilitate interaction with academic work or certain areas of investigation, and promote a kind of continuous education. In this article, the researcher discusses that in classroom and for exams also short answer type test is more effective and valuable to assess pupils. Evaluation procedure becomes more effective and valuable through short answer type tests.

Keywords: *Short answer type test, MCQ, strategies of SAQ and MCQ, design of SAQ and MCQ, student & teacher's review*

Introduction

Regarding the many item types or forms of testing that might be utilized, assessment has an empirical basis. The types of questions include true-false, fill-in-the-blank, multiple-choice, fill-in-the-blank, short answer, problem solving, and essay questions. There are more different kinds of questions than MC and TF questions, such as constructed response (CR) questions and open-ended questions that call for the creation of a rubric. It is important to consider if these two primary forms offer distinct metrics for evaluation when choosing the kinds of items for measuring instructional objectives (Aggarwal, 1994).In terms of the kind of response that may be provided, a closed-ended question differs from an open-ended inquiry. The second permits pupils to create their own replies to create a vast range of solutions, whereas the first only offers a small number of possibilities to pick from. Another distinction is that while open-ended questions allow for a wider range of replies, rating the responses to closed-ended questions is significantly quicker than rating the responses to open-ended questions (Parmenter, 2009).

Short-answer questions are open-ended and force pupils to come up with an answer on their own rather than choose one from a list of options. They are frequently used in exams to gauge students' fundamental knowledge and comprehension of a subject before more in-depth evaluation questions. In order to increase their validity, reliability, and sample capability, test sheets are increasingly composed of short answers (Aggarwal ,1994).

As an objective assessment method, multiple-choice exams have several benefits, such as being straightforward to score, covering a wide variety of subject matter, and serving both analytic and constructive functions. They could, however, just test lower-order thinking abilities and promote one-right-answer thinking. MCQs could encourage a superficial approach to study and prevent pupils from showcasing their breadth of knowledge. Students who are test-oriented may prefer MCQs since they often contain all the information needed for a high score. Additionally, multiple-choice exam performance could provide teachers with inaccurate data (Kothari, & Garg , 2019).

Method

This is a qualitative based descriptive study, where the researcher discusses about the merits, demerits and design of SAQ & MCQ. The study's objective was to compare the effectiveness of MCQs and SAQs.

Short-answer questions

Short-answer questions fall under the category of built-in responses or open-ended questions, which demand responses from test takers. In actual use, they can take the form of word replacement or text accomplishment. The

responses could only need one word, a couple of sentences, or to be bulleted. They are frequently employed for measuring low levels of cognition (Chan, 2009). Creating short-answer questions and grading the responses is simple (Sam, Hameed, Harris, & Meeran, 2016).

Even though they come in a variety of formats, short-answer questions often demand precise replies and have the following characteristics:

1. Most people read and respond to them in about five minutes, and many just need a single one.
2. Short answers allow for a wider range of topics.
3. They often score with more objectivity.
4. More trustworthy and dependable than essay-based tests.
5. Unlike with objective inquiries, where the answer is predetermined, the student provides the response.
6. Accurate and detailed in terms of the breadth and depth of the responses.

Merits of Short Answer Type Test

1. Compared to lengthy response questions, the objectivity of scoring may be better maintained with short answer questions.
2. In addition, the question-setter may formulate more of these in the same amount of time as long-answer questions. As a result, the subject (course) may be covered more thoroughly.
3. They are more trustworthy than inquiries requiring lengthy responses.
4. The pupils are less likely to hazard a guess.
5. Governance and processing are simple.
6. It strikes a balance between the written work and the testing elements' target forms.

Demerits of Short Answer Type Test

The short-answer exam items don't have any such substantial drawbacks. Scoring may be impacted by penmanship, speech, emotion, and response organization. Long-answer types of examinations should be avoided in favor of short-answer types of tests (Mangal , & Mangal , 2018).

Design

In contrast to multiple-choice questions (MCQs), short-answer questions (SAQs) are intended to provoke quick responses from pupils. SAQs may be used to assess knowledge and comprehension, facilitate interaction with

scholarly literature or a specific case study, and promote a kind of learning that is progressive. They are applicable to both formative and summative evaluations. SAQs can be written as lists of answers, short descriptive or qualitative single-sentence responses, diagrams or graphs with justifications, or even by adding missing words to a phrase. As the name implies, the response is frequently succinct (Weimer, 2015).

The way to design good questions are:

- ❖ State the inquiries in plain words.
- ❖ Make sure each question has just one valid response, and specify how the response should be given.
- ❖ Initial inquiries are preferable to sentence completions.
- ❖ Be specific for mathematical inquiries about the markings for both the procedure and the item, as well as if components are included in the solution.
- ❖ Be ready to embrace more solutions, a few of which one might not have expected.

Multiple-Choice Question

The three components of a multiple-choice question (MCQ) consist of a stem-wise list of potential responses and a key. While it may seem like these questions demand memorization, they really stimulate a deeper understanding of the course material. Questions should demand analysis in order to promote analytical thinking, like a situational analysis preceding MCQs that urge students to make decisions or use certain strategies. Generating options with a high level of discrimination can also help with MCQs that assess students' capacity for critical thought (Kothari , & Garg, 2019).

Advantages

1. Either manually or digitally, scoring is fast and easy.
2. May be designed to assess a variety of higher-order cognitive abilities.
3. May engulf several topics on a single test and be completed in class time.

Disadvantages

1. Give untrained pupils the chance to hazard a guess; if they guessed correctly, they would receive a score for the knowledge they lacked.
2. Expose pupils to false information that can skew their future interpretations of the material.

3. Spend some time and effort crafting them (particularly effective ones).

General strategies

a) Write a couple of questions following the lesson when the educational information is still vivid in our memory. Multiple-choice questions are hard to answer and take time to produce.

b) By telling pupils to choose the best response instead of the accurate response, it's acknowledging the possibility that the interruptions may contain some truth.

c) When instructors speak in everyday language, pupils are more inclined to reject arguments as false.

d) Avert providing linguistic linkage hints from the key's base. Pupils are more inclined to select the right response if the key contains words that are quite close to those in the stem.

e) Avert tricky inquiries. Questions ought to be written so that learners who are familiar with the subject may determine the right response.

f) Don't use nasty language.

Design

- ❖ Consider if the pupils could respond to the question without considering the possible answers. If so, the base is suitable.
- ❖ Include all pertinent information at the base of the stem.
- ❖ Omit extraneous language and unrelated details from the starting point.

There are several programs that can analyze the reliability and validity of MCQ test outcomes. Test the questions before using them in summative assessments so that formative testing can create the necessary data. In addition to asking students to respond, we may also inquire about their level of belief in the response they are providing. This decreases guesswork and gives the student information about how much they have understood or comprehended so far (Gay, 1980).

Student and staff experience

A) Benefits

SAQ:

Students: They are frequently less apprehensive than many other testing techniques since they are more accustomed to the process.

Staff: Short-answer questions may be marked quickly and by a variety of assessors as long as the questions are written such that the assessors can take into account all possible replies. With AI, feedback creation is supported. Additionally, they are rather simple to configure.

MCQ:

Students: excellent to allow for self-evaluation, especially online where remarks are immediate.

Staff: are simple to mark, can be put into a database of questions that can be used again, and are an effective way to assess many pupils (Chan , 2009).

B) Challenges

SAQ:

Tests reduce proficiency in learning and could promote superficial strategies. The complexity of the question format increases to test more advanced skills. The use of shorter articles alongside larger ones makes time management essential. Open-ended questions can result in lengthy responses if they are not handled correctly; therefore, it is crucial to have established standards for the types of answers.

MCQ:

Writing questions to assess higher-order learning can be difficult since the framework of the questions often becomes more complicated than the material being evaluated (see Question Pro in the list of useful assets below). While completing a summative multiple-choice exam, students should prepare so that their awareness of the subject matter is examined rather than their comprehension of the question format.

C) Workload

Students

Learners may need more time to fully utilize the suggestions than it takes them to respond to the questions, though this is certainly one of their greatest assets.

For staff

Designing multiple-choice questions takes a lot of work, but after a solid topic collection can be built, using questions and grading them take very little time.

Discussion:

The study contrasts multiple-choice and short-answer questions and offers

fresh data not seen in other studies. Small sample numbers, low-stakes situations, and an approach where the identical question was put forward in both forms were all employed in prior studies. The current study employed randomization to gather data from a sizable sample of test takers in pressure-filled circumstances. Additionally, it assessed the variance in reaction times between the two formats, giving rise to a more thorough knowledge of how item format affects response times (Budiyono, 2018).

Conclusion

Contrary to multiple-choice questions (MCQs), short-answer questions (SAQs) are intended for obtaining immediate responses from pupils. SAQs may be used to assess knowledge and comprehension, facilitate interaction with scholarly literature or a specific case study, and promote a kind of learning that is progressive. With MCQs, accurate answers are frequently prioritized over critical thinking abilities. Being succinct is essential while answering SAQS, and many questions have a set minimum amount of thoughts or facts you must provide in order to receive all possible points. Our collective understanding of how to create good SAQs will probably grow with increased use and research (Bridgeman, Morgan. 1996).

At the same time, we will presumably learn crucial information on the timing of SAQs' assessment improvement. It will be challenging to create successful SAQs in some subject areas, just as it would be with MCQs. It's possible that the two forms work best together in this regard. It is also evident that certain evaluation contexts may benefit from the relative difficulty of SAQs, while other contexts may find this attribute to be a drawback. Although the greater reaction time for SAQs is unquestionably a drawback, other advantages could make up for it. The issue for test designers will be figuring out how to effectively take advantage of those benefits to enhance assessment when the possibility of automated scoring SAQs emerges (Baghaei, & Amrahi, 2011).

References

Aggarwal J.C., (1994), Essentials of Educational Psychology, *Vikas Publishing House* PVT LTD, Delhi – 110032

Baghaei, P. & 1 & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple-choice items. *Psychological Test and Assessment Modeling, Volume* 53, 2011 (2), 192-211. Retrieved from https:/ /pdfs.semanticscholar.org/9991/fb47cc50ff7f1d9a694d5ec72b9b3 4f26f69.pdf

Bridgeman, B; Morgan, R. (1996). Success in college for students with discrepancies between performance on multiple-choice and essay tests. *Journal of Educational Psychology*, Vol 88(2), Jun 1996, 333-340. Retrieved from http:// dx.doi.org/10.1037/0022-0663.88.2.333

Budiyono, B. (2018). Multiple Choice Questions (MCQs) vs. Short Answer Questions (SAQs) for *Inferential Comprehension*. Retrieved fromhttps://ojs.unpkediri.ac.id/index.php/inggris/article/download/12506/1012/

https://ojs.unpkediri.ac.id/index.php/inggris/article/download/12506/1012/

Chan C. (2009) *Assessment: Short Answer Questions, Assessment Resources @*HKU, University of Hong Kong. Retrieved from http://ar.cetl.hku.hk/am_saq.htm

Gay, L. R. (1980). The Comparative Effects of Multiple-Choice versus Short-Answer Tests on Retention. *Journal of Educational Measurement.* Vol. 17, No. 1 (Spring, 1980), pp. 45-50. Retrieved from https://www.jstor.org/stable/1434671?seq=1#page scan tab contents

Kothari C.R., Garg G., (2019), Research Methodology Methods and Techniques, New Age *International Publishers*, New Delhi – 110002

Mangal S.K., Mangal S., (2018*), Research Methodology in Behavioural Sciences*, PHI Learning Private Limited, Delhi – 110092

Mangal S.K., (2017), *Statistics in Psychology and Education,* PHI Learning Private Limited, Delhi – 110092

Mullen, K & Schultz, M. (2012) Short Answer versus Multiple Choice Examination Questions for First Year Chemistry. *International Journal of Innovation in Science and Mathematics Education IJISME Vol* 20, No 3 (2012). Retrieved from https://openjournals.library.sydney.edu.au/index.php/CAL/article/view/6441/7096

Parmenter, D. A. (2009). *Essay versus multiple-choice: student preferences and the underlying* rationale with implications for test construction. Retrieved fromhttp://www.freepatentsonline.com/article/Academy-EducationalLeadership-Journal/208747754.html

Ramraje, S. N & Sable, P. L. (2011). *Comparison of the Effect of Post-Instruction Multiple-Choice and Short Answer Tests on Delayed Retention Learning. Retrieved from* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3562952/pdf/AMJ-04332.pdf

Sam, A. H., Hameed, S., Harris, J. & Meeran, K. (2016). *Validity of very short answer versus single* best answer questions for undergraduate assessment. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5064885/pdf/12909_2016_Article_793.pdf

Tychonievich, L. (2012). *Multiple choice or short answer? The intellectual impact of question format. Retrieved from* http://www.cs.virginia.edu/~lat7h/blog/posts/206.html

17

Portfolio Assessment: An Effective Tool for Evaluating Learning

Usha Rani Sahoo

Abstract:

Portfolio is a systematic purposeful collection of students work that can serve a variety of instructional and assessment purposes. Unlike other standalone assessments, portfolios include multiple artifacts and sources of information such as students' essays, creative short story, artwork, lab reports, book review etc., which reflects the whole picture of what a student know and can do. Portfolio is not an assessment item rather it is a source or basis for assessment of learning and instruction. It requires a scoring guide or rubrics to score. There are two main types of portfolios, 'Instructional' or 'working portfolios' and 'Showcase portfolios. Being formative in nature, instructional or working portfolios demonstrate for monitoring progress of the students over a period including the unfinished and finished exemplary whereas the showcase portfolio is summative in nature that displays the current accomplishments and mastery at the end of a course or year. Versatility and multi-dimensional display of contents make it a useful tool for students' assessment. The purpose of preparing a portfolio is to encourage students to become more reflective about their own work and boost their self-confidence and self-appraisal while demonstrating their abilities and achievements. It can be considered as a mean to communicate the progress of students to parents as well. Besides these potential strengths, portfolios have its weaknesses for which it has been kept outside of the domain of conventional assessment tools. Lack of generalizability and common bases of comparability may result in unbiased and non- consistent scoring. It can be time consuming to assemble and needs constructive feedback for students' development.

Keywords: *multiple artifacts, instructional portfolio, showcase portfolio, self-appraisal, constructive feedback*

Introduction:

Portfolio usually refers to a collection of assets or records showcasing the best work. It may contain the collection of art works related to an artist portfolio, investment and financial statements related to investors portfolio, best photographs related to a model portfolio etc. From educational point of view the portfolio is a purposeful collection of students work showcasing the students' effort, their progression, and achievements by selected performances in one or more areas of the curriculum. (Paulorganized collection of evidence used by the teacher and student to monitor the growth of the student's knowledge, skills, and attitudes in a specific subject area. The growing needs to look at actual sample of students writing for assessment and evaluation, introduced the use of portfolio assessment Pat Belanoff and Peter Elbow first introduced portfolio assessment in place of writing exit exam at Stony Brook University of New York in 1983. Since then, it has been increasing frequency as adopted by individual classroom teachers for instruction and assessment of students' progress. (Williams,NA) son, Paulson, Meyer, 1991) Therefore, a portfolio assessment refers to a systematic and

What qualifies as a portfolio of student work?

Any random collection of pieces of students works or their observations in a folder does not qualify as a portfolio. An actual portfolio differs from the collections of all the documents or photos of achievements placed together. A portfolio contains several multifaceted students' best efforts selected by keeping in mind to serve a particular purpose. Instead of containing all the works of a student, it may contain examples of best work from several categories like teacher notes, reading log, video tape of group projects, audio tape of oral reading, journal writing samples, a book review, a letter to a friend, a creative story writing, a persuasive essay, awards in mathematical quizzes, calligraphy writing, students' self-reflection and so forth. Therefore, it is sometimes described as portraits of a person's accomplishments. It is usually a self-portrait but often has structured from the guidance and feedback from teachers and peers.

Contents of Portfolio:

The portfolio for its effective use, may contains the following artifacts (Epon, R. 2013).

Type of work	Examples
Recognized work of students	Achievement of any awards or prize got in Local, State or National level
Individual Student record of participation	Class representative or president in school
Group work activity	Group project work or team activity
Extracurricular Activities	Excelled in any extracurricular activity
Religious Activity	Work in any religious institution or active in any spiritual centre
Regular Maintaining records	Daily Journals, book keeping that worth acknowledgeable

Purpose of Portfolios

The first question that come to the mind while designing a portfolio is "What is the purpose of the portfolio" Because without a clear understanding of it, the portfolio may be indistinguishable from unorganized samples of work. Portfolios can be intended to achieve the following purposes,

- To display evidence about student progress or growth in any field of education
- To provide evidence for certifying student achievement
- To improve the self-evaluation skill of students as an instructional goal.
- To assess students' work over a span of time.

Importance of Portfolio Assessment

There are different types of portfolios based on its varied purpose. The following are the importance of using a portfolio assessment (Education Reform, 2016):

- Students' portfolios can be used in evaluating effectively their learning progress and achievement.
- Portfolios can help teachers to collaborate and reflect on the growth, progress, or improvement of learners over time.

- ❖ Instead of testing the learned specific knowledge, portfolios help in determining the learners' application skill of their knowledge and understanding to complex problem.
- ❖ Portfolios are greater tools to be used as a communication medium between teachers and parents by demonstrating concrete examples of students' progress.
- ❖ Portfolios help students to build positive attitudes and habits by taking ownership of their own learning.

Types of Portfolios

Based on different purposes of portfolio construction, there are different types of portfolios. As per the intended purpose the portfolio contains different materials. (Gronlund & Linn, 2000)

a) Working portfolio vs Finished Portfolio

Working portfolio are often used to document daily activities and mostly useful for formative evaluation. It is mostly used when students are encouraged to independently express themselves and to get formative feedback over their work.

As the name suggests, the finished portfolio implies the collection of completed works for a particular audience. It is mostly concerned with summative evaluation. It can be used by prospective employer.**b)**
Current Accomplishment vs Developmental Portfolio When the focus of a portfolio is on accomplishment covering a relatively brief period its termed as current accomplished portfolio. The unfinished works are excluded from this portfolio.

Whereas the Developmental portfolio presents the progress or growth of students' performance. The time frame is generally longer as compared to accomplished portfolio. It is used in identifying strength and weakness of the students and keep a track of the development process.

c) Showcase portfolio vs Documentation Portfolio

The name Showcase Portfolio itself express the inclusion of best work of students to display. The goal of this type portfolio is for students to learn to identify work that best demonstrate what they know and can do with specific purpose and audience in mind.

Documentation Portfolio on the other hand, usually intended to provide evidences of the achievement of instructional objectives. This is why the scope of inclusion is broad in this as compared to showcase portfolio.

Potential Strengths and Weakness of Portfolios

The portfolios are believed to have numerous potential strengths as its prepared by students itself. Those are (McDonald, 2012),

a) Students will attain autonomy by getting opportunity to control their own work.

b) It will develop an experience of individuality among students.

c) It can make students life disciplined by developing record keeping skills and time management skill among them.

d) It makes the assessment procedure flexible by allowing a broader range of documents or showcasing of varied achievement from different spheres.

e) While summarizing their experience and excellence, it will develop self-efficacy and self- confidence among the students.

f) Students will be motivated towards success while reflecting on their own work.

Although the potential strengths of portfolios make them attractive, it also has weaknesses. Those are:

a) Portfolios usually are very time consuming to assemble.

b) Assessment of portfolios are time taking for teachers.

c) Considerable thoughts, preparation and experience are needed to ensure the portfolios justify its making and evaluation.d) Difficulty of assigning clear scoring criteria for a portfolio because of its inclusion of diverse materials.

e) Careful planning is required to prepare a portfolio aligning to different purposes.

f) Lack of generalizability and common bases of comparability may result in unbiased and non-consistent scoring.

Models of Portfolio Assessment

As per the different purposes of portfolio preparation, its assessment leads to large variation in the implementation. There are two types of portfolio assessment model, an Accountability Model, and a Classroom Based Model.

Two-models of Portfolio Assessment

Accountability Model	Classroom Based Model
Provide valid and reliable data to determine program effectiveness at the school level	Purpose: Improve classroom instruction
Low teacher involvement	High teacher involvement
Centralized content selection	Teacher selected portfolio contents
District or schoolwide objectives determined by the program	Mastery objectives individualized to student/classroom needs
Use specific rubrics districtwide to determine the mastery level of each student	Teachers use their professional judgment to determine the mastery level of each student
Portfolio contents are mandated	Teacher/student decide what to put into the portfolio
Use scoring rules or second raters to improve scoring consistency across teachers	Scoring done by individual teacher(s)
Process monitoring is important	Monitoring is not needed
Score summarization and school level aggregation are needed	Score summarization and school level aggregation are not needed
Potential for high level of technical quality (reliability and validity	Potential for low level of technical quality (reliability and validity)

(Source: Joyce, B. and Yang, H., 1995)

Designing an Effective Portfolio

Although the physical form of portfolio is secondary to the purpose, there is a consideration of the design. Evaluation criteria of Portfolio

Goal setting	The objective or purpose should be vividly and clearly stated to determine the goal and provide direction in preparing a portfolio
Selecting	It must be relevant to the goal of instruction set. As per the criteria different works should be selected
Presenting	The synchronized presentation of the selected documents should collaboratively reflect the broader picture intended towards the attainment of the goal.
Reflecting	The student's own reflection and peer reflection should be included for its evaluation
Evaluating	The evaluation criteria should be mentioned for its assessment by teachers or self

Specification of evaluation criteria enhances the fairness. The students need to know what is expected from them and how their work will be evaluated for being consisted with their work. Therefore, it is needed to prepare an unbiased scoring criterion. The portfolio can be well evaluated with clearly specified rating scale or vividly mentioned rubrics. Analytic scoring rubrics can be used on individual portfolio for formative evaluation and Holistic scoring rubric is appropriate for summative evaluation (Gronlund & Linn, 2000). With proper mentioned evaluating tools, the students can unbiasedly assess themselves and realizing their own strength and

Conclusion

Learning assessment is an essential part of classroom instruction. With the shift of focus from teacher centred educational environment to students centred, it becomes prominent to include students in their own assessment. In this regard portfolio assessment can be regarded as an important tool because of its dynamic inclusion of students work. With varied purposes students' portfolio assessment can be considered as an influential tool for todays scenario to include students, teachers, parents, and peers for overall growth of learning. weakness can contribute to their own progressIt contains the following procedures,

References:

Barton, J, and A. Collins, (1997). *Portfolio assessment: A handbook for educators.* Menlo Park, CA: Addison-Wesley.

Benoit, J. & Yang, H. (1995). A Redefinition of Portfolio Assessment Based Upon Purpose: Findings and Implications from a Large-Scale Program. *American Educational Research Association, San Francisco, California.* https://files.eric.ed.gov/fulltext/ED388680.pdf

Epon, R. (2013). Portfolio Assessment. https://www.slideshare.net/rhandi20/portfolio- assessment-26375689

Gronlund, N. E., & Linn, R. L. (2000). *Measurement and Assessment in Teaching* (8th ed.).Merrill Prentice Hall.

McDonald, B. (2012). Portfolio assessment: direct from the classroom. *Assessment & Evaluation in Higher Education, 37(*3*)*, 335-347.

Paulson, F.L., P.R. Paulson, and C.A. Meyer. 1991. What makes a portfolio a portfolio? *Educational Leadership, 58*(5), 60–63.

Portfolio.(2016, August 02). *The Glossary of Education Reform.* https://www.edglossary.org/portfolio/

Williams, N. A. (n.d.). A Brief History of Portfolios. https://www.nicoleannwilliams.com/a-brief-history-of portfolios.html#:~:text=Why % 20 did %20 Portfolios %20 Come % 20 into, respond % 20 to % 20 these % 20 growing % 20 needs.

18

The Impact of Continuous and Comprehensive Evaluation on Academic Achievement at Secondary Level Students in West Bengal

Soma Sardar

Abstract

CCE (Continuous and Comprehensive Evaluation) is the new evaluation system in India. It has started to change the Indian education scenario. CCE is considered as a new innovative concept in the teaching learning process. CCE refers to a system of school based assessment that covers all aspect of student's growth and development. The evaluation system is divided into two parts, namely Formative and Summative. The West Bengal Board of secondary Education has introduced CCE in all Government, Government sponsored and Government Aided School in 2007-2008.The format of CCE has been changed by WBBSE in the academic year 2013 for the students of class V-VIII and at secondary level in 2015 .In this evaluation system two types of assessments are framed- Formative and Summative. The present study attempts to find out the impact of CCE on Academic Achievement at secondary level student. The Researcher prepared questionnaire for the purpose of the study. Purposive sampling method was used. There were 10 schools of urban and rural areas from Nadia, Murshidabad and South 24 parganas Of West Bengal. Data of 75 students was collected. Students were taken from class IX and X. Simple percentage method was followed. The results showed that the students have improved a lot. They have done better performance in previous studies. After some modification it is hoped that the new evaluation system will play a more important role in education in the future.

Keywords: *Continuous and Comprehensive Evaluation System,, Formative, SummativeSecondary level.*

Introduction

"Education aims at making children capable of becoming responsible, productive and useful members of society, knowledge, skills and attitude are built through learning experiences and opportunities created for learners in school. It is in the classroom that learner can analyze and evaluate their experiences, learner to doubt to question to investigate and to think independently."(CBSE manual, 2009. P-1)

Education plays an important role in the development of the nation. The aim of education is to make children independent, responsible, being good human and useful members of society. In teaching learning process evaluation is a good tool to judge the students capability.

Evaluation is an important and integral part of education. The term evaluation is a newly introduced term .According to Tyler the definition of evaluation is:"The process of determining to what extent the educational objectives are being realized."(Tyler, 1950, p -69) The main purposes of evaluation is to help the all-round development of learners both scholastic and non scholastic area.

After many reforms ,the new form of evaluation CCE is consider as one of the major steps taken in this regard to improvement the quality of school education. RTE Act 2009 section 29 ,the School Education Department Government of West Bengal decided to implement revised CCE in all Government ,Government sponsored and Government aided school from class V to class VIII from 2013 (139/29.01.2013). The Expert committee on School Education was assigned the role to design and develop a framework for implementation of CCE in West Bengal. The Peacock Model for CCE was developed by the committee and was ratified by the West Bengal Board of Secondary Education. In case of class IX & X the earlier system of CCE (which was introduced from the academic year 2008-2009) was followed. The new evaluation framework CCE has introduced by WBBSE for the students of class IX (22.01.2015) and same for the students of class X from the academic year 2016. In the new evaluation framework the existing pattern of Oral Examination was replaced by Internal Formative Evaluation which work as a natural follow up the CCE at the elementary level .The IFE pattern was developed by the expert committee and was introduced by the Board in the affiliated schools from academic session beginning in January 2015for class IX and 2016 for class X. The restructured revised evaluation framework was approved by West Bengal of Board of Secondary Education (WBBSE)

The term Continuous has two components known as continual and Periodicity. The Continual Process which means from the beginning and during the instructional process where as periodicity refers frequency / term of the

assessment. The term Comprehensive covers Scholastic areas, Co – Scholastic activities. The scholastic components include subject specific areas such as Academic .The co – scholastic components include life skills, attitudes & values. Evaluation is two types under CCE named as Formative and Summative assessment. The formative test will be conducted three (3) times in an academic year. It is carried throughout the years by the teachers with the help of six (6) methods. These are Survey Report, Nature study, Case study, Creative writing, Model making, open text book evaluation. Out of these six methods ant three are to be in an academic year. In a particular class one method is to be applied for one term. There should not be any repetition of a particular modality for a particular class in an academic year. In an academic calendar there will be three summative evaluations.

Summative Evaluation

First Term		Second Term		Third Term	
Full Marks	Specific Period	Full Marks	Specific Period	Full Marks	Specific Period
280(40*7)	April 1 to April 10	280(40*7)	August 1 to August 10	630(90*7)	December 1 to December 15

Formative Evaluation

First Term		Second Term		Third Term	
Full Marks	Specific Period	Full Marks	Specific Period	Full Marks	Specific Period
70(10*7)	January to April	70(10*7)	May to August	70(10*7)	September to December

GRADING SYSTEM

Marks Scale	Grade	Remarks
90-100	AA	Outstanding
80-89	A^+	Excellent
60-79	A	Very good
45-59	B^+	Good
35-44	B	Satisfactory
25-34	C	Marginal
Below 25	D	Disqualified

Objectives of The Study:

- To study the concept of the students towards CCE
- To study the concept of the students towards Summative and Formative Evaluation
- To study the impact of CCE on academic achievement
- To study the effectiveness of CCE

Delimitation Of the Study

- Study is delimited to Govt. Schools and Govt. sponsored schools only.
- Study is delimited to IX & X standard students only.

Methodology:

- **Sample of the study**: Purposive sampling procedure was followed. The sample was taken from different school situated in different districts viz. –Nadia, South 24 parganas. A sample of 75 students was taken for the study.
- **Research Methodology**: Survey method was used for this study.
- **Tool of the study:** A self made questionnaire was made by researcher for this study. Questionnaire has 27 questions on CCE. Questions are based on these points.

- Concept of CCE
- Term of CCE
- Concept of IFA
- Scholastic & co scholastic activities
- Mental health
- Academic performance
- Infrastructure of school
- Implementation of CCE

- **Statistical method used:** Simple percentage method was used in this study.
- **Procedure of the study**: Researcher made Google form for collecting data.

ITEMS	STUDENTS RESPONSE (%)		
	Y	N	NR
1. Do you know anything about CCE?	73.33	13.33	13.33
2.Do you attend school everyday	66.7	21.33	12
3. Are teaching aids used in your classroom?	70.66	10.66	18.66
4. Are co –curricular activities organized properly in school ?	68	18.76	13.33
5. Do you think co –curricular activities are essential in school?	80	8	12
6. Does the school have the necessary arrangements for co –curricular activities?	69.33	18.7	12
7. Do you have any idea about IFA(Internal Formative Assessment)	62.66	25.33	12
8. Is IFA useful for society?	73.33	12	14.7
9. Do you prepare any model in formative assessment?	69.33	17.33	13.33
10. Does it take a lot time to create or make or write project?	60	25.33	14.7
11. Do you easily get the necessary things for the project?	61.33	22.7	16
12. Does creating a project reduce your reading time?	54.7	29.33	16
13. Do you get help from your parents in project work?	68	17.33	14.7
14. Are nature study, survey, creative writing, model making (any 3) used in IFA in your school?	69.33	14.7	16
15. Are one formative assessment conducted in an academic year?	46.7	34.7	18.7
16. Are three formative assessments conducted in an academic year?	65.33	16	18.7
17. Is the formative assessment conducted before the summative evaluation?	65.33	17.33	17.33
18. Do you like making project?	72	12	16
19. Do you feel scared during the test?	46.7	38.7	14.6
20. Do you feel stress, anxiety before examination?	58.7	26.7	14.6
21. do you think only subject knowledge is evaluated?	42.7	42.7	14.6
22. Do you feel any effect of CCE on your performance?	61.33	22.7	16

23. Are you happy with CCE?	69.3	13.33	17.33
24. Is the current academic system up to date?	56	28	16
25. Do you Identify your weakness?	70.7	14.66	14.66
26. Do you enjoy evaluation?	72	10.7	17.33
27. Are your results better than before?	73.33	12	14.7

Findings and Discussion

- Students gave their feedback based on the question "Do you know anything about CCE?" asked in the questionnaire. Result shows that 73.33% students know about it whereas 13.33% students don't know about CCE. 13.33% students did not give their answer.
- Most of the students attend school regularly.
- Most of the students agree that teaching aids are used in their classroom.
- Most of the students said that their school has the necessary arrangements to organize co-curricular activities. They think it is necessary in life. It is organized properly in school.
- It was found that regarding the question" Do you have any idea about IFA?"62.66% students responded with Yes and 25.33% with No whereas 12% students did not give their feedback.
- To the question "Is IFA useful for society?" 73.3% of the student respondent with Yes and 12% with No whereas 14.7% did not give any response.
- Students gave their feedback based on the question "Do you prepare any model in formative assessment?" 69.3% students responded with Yes and 17.33 with No.13.33% students did not give their answer.
- Students gave their feedback on the question "Are nature study, survey, creative writing model making (any 3) used in your school?" .Results shows that 69.3% students agree that different methods in formative evaluation are used in their school.
- 65.3% students revealed that 3 Formative Assessment is conducted in an academic year.46.7% students revealed that one Formative Assessment is conducted in an academic year.
- Most of the students agree that they feel scared during the test.

- To the question "Do you feel stress, anxiety before examination?"58.7% of the students responded with Yes and 26.7% of the students with No and 14.6% of the students with no response.
- To the Question "Do you think only subject knowledge is evaluated?" students gave their feedback. Result shows that 42.7% students realize that only subject knowledge is not evaluated in their school .Whereas 42.7% students think only subject knowledge is evaluated in their school.
- Most of the students said that CCE has an effect on their performance. They happy with CCE. Majority of the students agree that their results are now much better than before.
- 56% of the students said that current academic system is up to date. But28% students responded with No. Rest of the students did not give any answer.
- To the question "Do you identify your weakness?" students gave their feedback .70.7% students responded with Yes and 14.6% with No whereas14.7% of the students did not give any answer.
- Most of the students said that they enjoy evaluation.

Main findings of the study

Students know about the CCE. They are happy with new evaluation system. They enjoy the evaluation system. After the implementation of new evaluation system their interest in reading has increased. The rate of absenteeism has decreased. They all agree that new evaluation system is up to date .But it needs some changes. The results have been much better than before. The results of the back bencher students have undergone an unprecedented change. Every student's academic performance has improved.

Conclusion

Education is an important and integral part of education system .We can not avoid this or cannot abolish this process but we can make it better through the implementation of the CCE. CCE is an effective tool to improve the quality of teaching learning process in the school. The evaluation system helps the teacher to take necessary corrective measures for the further empowerment of child .CCE has taken a step forward in being a part of the historical examination reforms.CCE is decrease the workload ,exam fear , reading process and the results are automatically better than before. It also helps the students who are not good in academics to show their talent in other fields such as arts, humanities sports, music and athletics and also help to motivate the students

who have a thirst of knowledge. CCE has facilitate children to achieve students best performance at secondary level which is the terminal point for their future career .The country which is on the threshold of being the next economic superpower requires the confident and educated youngsters. So we should prepare our children for the future life with the help of CCE. The new education system is not Aladins Surprise lamp. This is the just like guide. So criticism is normal. There must be some mistakes .These need to be modified refinement.

References

Aggarwal J.C (2018) *Essentials of Examination System Evaluation Tests and Measurement. New Delhi: Vi*kash Publishing House Pvt. Ltd.

CBSE (2009) *Teachers Manual on CCE .New* Delhi: CBSE

CBSE *Continuous and Comprehensive Evaluation*(CCE) Report 2014.CBSE, India

NCERT (2005), *National Curriculum Framework-2005,* New Delhi

NCERT (2000), *National Curriculum Framework for School Education*, NCERT, New Delhi

Saluja Anshul (2016),New Dimensions in Evaluation : Teachers Perspective on CCE and its implementation, International Journal of Science Technology and Management, Vol No 5,Issue 8

Sing,N & Pany,S.(2016), Continuous and comprehensive evaluation : a paradigm shift in evaluation. *Online Interdisciplinary Research Journal* ,VI(Spl. Issue), 139-147

Thakur, U. R. (2016), A study on the perception of the students and teachers towards continuous and comprehensive evaluation at secondary school level under Dibrugarh district of Assam, *International Journal of Humanities & Social Science Studies*, II(IV),165-169

Tyler, R.W. (1950). *Basic principles of curriculum and instruction.* Chicago: University of Chicago Press.

19

Grading and Credit System: An In-Depth Analysis of Continuous Comprehensive Evaluation (CCE) in Education

Asim Mahata

Abstract

The adoption of the Continuous Comprehensive Evaluation (CCE) framework within the grading and credit system has ignited significant discourse and debate within the realm of education. This comprehensive analysis delves into the multifaceted dimensions of CCE, scrutinizing its conceptual foundations, implementation challenges, and potential implications for educational pedagogy and student learning outcomes.Drawing from a wide array of scholarly literature, this research interrogates the origins of CCE as a paradigm shift from conventional summative assessment methods to a more holistic, formative evaluation approach. The underlying philosophy of CCE, emphasizing skill development, individualized learning trajectories, and holistic assessment, reflects a departure from the rigid assessment paradigms of the past.This analysis critically evaluates the impact of CCE on both educators and students. It examines how CCE influences pedagogical strategies, fostering a shift towards learner-cantered teaching and greater focus on formative feedback. Simultaneously, it explores the psychological implications of continuous evaluation on students, investigating its influence on motivation, self-esteem, and the development of a growth mindset. This in-depth analysis navigates the intricate landscape of Continuous Comprehensive Evaluation within the grading and credit system, unravelling its potential to reshape educational paradigms. Through a critical lens, this research contributes to a

nuanced understanding of the challenges and opportunities inherent in the implementation of CCE, paving the way for informed policy decisions and pedagogical advancements in the realm of education.

Keywords: *CCE, Grading and Credit System, Education Assessment, Formative Assessment, Pedagogical Shift, Learning Outcomes.*

Introduction

The Latin word "gradus," which meaning "step," is where the English term "grade" originates. Grading is the practise of classifying topics or students according to predetermined criteria with the goal of reducing misclassification. Based on their performance and competency, students are grouped into a few ability categories for grading. It entails the employment of a set of specialised symbols or numbers, the meaning of which should be understood by all parties involved—students, instructors, parents, and other stakeholders—in a clear and consistent manner. For the aim of meaningful and exact communication, the grading symbols must represent the same thing to every user. The topic matter, level of difficulty of the exam, and various abilities—intellectual, physical, emotional, social, and so forth—that are evaluated using assessment instruments are only a few of the factors that affect the grading process. Accuracy is also necessary for the evaluation programme *(*Linn, Robert & Norman E. Gronland, 2000*)*.

A school-based assessment, or CCE, examines every facet of educational programmes that are relevant to a child's growth. It places special emphasis on two goals: thorough assessment of learning outcomes and continuous evaluation. It encompasses the cognitive, emotional, and psychomotor areas of learning. It approaches assessment as a process of growth. The assessment of cognitive skills including knowledge, comprehension, application, etc. is related to the cognitive domain. In the emotional domain, evaluation refers to the assessment of characteristics including motivations, interests, attitudes, and other personality traits. Assessing learners' hand-use abilities is part of the psychomotor domain of evaluation (e.g. in handwriting, building and projects). Students' performance in both academic and extracurricular activities is evaluated in CCE. Through the assessment of students' performance in both kinds of activities, CCE seeks to lessen the burden that students have in the classroom while also enhancing their general talents and skills. First, let's clarify the two concepts used in CCE: comprehensive and continuous.

Statement of the Problem: Grading is the practise of classifying topics or students according to predetermined criteria with the goal of reducing misclassification. Based on their performance and competency, students are grouped into a few ability categories for grading. It entails the employment of

a set of specialised symbols or numbers, the meaning of which should be understood by all parties involved—students, instructors, parents, and other stakeholders—in a clear and consistent manner. For the aim of meaningful and exact communication, the grading symbols must represent the same thing to every user. The nature of the subject matter, the level of difficulty of the question paper, and the many abilities—intellectual, physical, emotional, social, and so forth—that are evaluated using evaluation tools are just a few of the factors that affect the grading process. Accuracy is necessary in the assessment programme. Therefore the researcher has decided to fix the problem stated as "Grading and Credit System: An In-Depth Analysis of Continuous Comprehensive Evaluation (Cce) In Education".

Objectives: The present study has been carried out to study the types of grading in the CCE. It also focuses on the advantages and disadvantages of grading system in CCE.

❖ Discussion

Types of Grading

1. Direct Grading: In direct grading, the examinees' performance is evaluated qualitatively. It might take the form of the anticipated behaviour for the answer to the question or activity that has to be graded, or it could be determined by using specific indicators presented as certain behaviours based on the examiner's view that is then immediately translated into letter grades. The evaluation of both cognitive and non-cognitive learning outcomes may be done successfully using this technique. It is recommended that significant non-cognitive aspects be listed according to stage, assessed, and reported in terms of letter grades independently. Examiners must classify responses and provide a quality score based on the question's scope and the weight assigned by the question paper in order to evaluate cognitive learning outcomes. For weighting, numerical numbers are often used. Multiplying the grade point by the weight age yields the answer's weighted grade points. "The Grade Point Average (GPA) for the paper is then calculated by dividing the total weighted grade points by the total weightage. The cumulative grade point average might be calculated by averaging the grades of many papers. The following table depicts the assignment of grades" *(*Dandekar W. N., 2004*)*,

Five Point Scale:

Q.No	Grades awarded	Grade point	Weightage	Weighted grade points
1	A	5	1	5
2	C	3	2	6
3	B	4	3	12
4	D	2	4	8
5	A	5	5	25
			Sum =15	Sum=56

The entire weighted grade point divided by the sum of the weights determines the paper's grade point average. This time, 56/15 = 3.73. This GPA may be transformed into an appropriate grade point system to determine the paper's grade.

1. Indirect grading: In this method, the answers get their regular marks. These scores are then translated into grades. "Norm-referenced (relative) and criterion-referenced (absolute) grading are the two basic methods used for this". The most appropriate grading scheme will vary depending on the application. To determine which students in a group are the best, norm-referenced grading may be used. Criterion-referenced grading is the ideal method if the goal is to show what each student has accomplished.

2. (a) Grading according to criteria

When determining a student's grade on an absolute basis, the authority compares their performance to predetermined criteria, ideally with regard to predetermined performance requirements and for the benefit of the school or board. These standards might relate to the level of proficiency that students should reach and could be expressed as (a) activities that need to be completed or (b) the percentage of right answers that need to be scored on an exam that is intended to gauge a precisely defined set of learning tasks. Instead of assigning letter grades based on a relative performance criterion, a standard-based system bases its decisions on an absolute norm. Every student will get good scores if they show a high degree of mastery in line with the defined performance requirements. "In this case, a student's grade is determined by their level of performance or success, independent of the performance or achievement of

other students. A percentage of marks may be used as an absolute benchmark for mastery learning"(Rao V. K., R. S. Reddy, 2007*)*, The grading system has several benefits, some of which are stated here.

The assignment of grades can be based on the range of marks as shown below.

Marks - Range	**Grade**	**Grade point**
91 - 100	A1	10.00
81 - 90	A2	9.0
71 - 80	B1	8.0
61 - 70	B2	7.0
51 - 60	C1	6.0
41 - 50	C2	5.0
33 - 40	D	4.0
21 - 32	E1	
00 - 20	E2	

When using this grading method, teachers should be somewhat aware of the expected degree of performance or success. Instructors must to be familiar with the test's difficulty level as well as each item on it.

3. (B) Relative grading with norm referenced when assigning a non-referenced grade, the instructor contrasts the pupil's work with that of a reference group, which is usually the student's peers. "Instead of an absolute accomplishment criterion, the grade is decided by the student's relative standing within the whole group under this method. Both the individual student's performance and the group's performance have an impact on the grade since the grading system is based on relative performance" *(*Rao V. K., R. S. Reddy, 2007*)* . In order to award relative grades, pupils are ranked according to their overall accomplishment, and letter grades are assigned based on each student's position within the group. This ranking may be based on the cumulative distributions of many classroom groups enrolled in the same course, or it may be restricted to a single classroom group.

The most popular use of this methodology is normal curve grading. Two distinct methods are used for grading on the normal curve. Prefixed interval approach and prefixed percentage approach are these. A grading scale with odd numbers is often utilised. When using the prefixed percentage technique, the proportion of each grade that will be assigned must be decided upon before letter grades are allocated. A five-point grading system, for instance, must have an equal proportion of As, Fs, and Bs and Ds on the normal curve.

How Grades are produced: Any effort to characterise grading systems without taking into account other elements of this network must inevitably be incomplete as grading systems are just one part of an interconnected network of educational activities. The methods by which grades are first determined—the classroom exam—may be the most significant of these processes. Of course, there are entirely formal distinctions here as well; for instance, between multiple-choice and essay exams, or between in-class and take-home assignments or questionnaires. In addition, the quality of the test items themselves should be considered, taking into account not just their substance but also their clarity and, in the case of multiple choice exams, their distracters. Any analysis of the relationship between grading and testing must address the relationship between one or both of these processes and instruction and learning. Testing and learning have a very clear link, particularly when exams are designed to support or encourage learning as well as to assess student progress. As a result, creating a question or activity in the classroom that demands reading comprehension from the students before they can respond to it or finish it is simple. "While the specific testing and grading plan used by the instructor does inform the student about what constitutes relevant knowledge and what attitude he or she holds towards precise evaluation and academic competition, teaching would seem to be somewhat further removed from issues of testing and grading"*(* Sharma R. A, 2004*),* Educational experts have distinguished between pupils who are learning focused and those who are grade driven. Pupils are not exempt from testing and grading methods. This contrast implies that for some students, the classroom is a place where they experience and appreciate learning for its own sake, even though it is undoubtedly much too one-dimensional. For other pupils, on the other hand, the classroom serves as a testing ground where passing becomes more crucial than really understanding the material. When asked how they become fixated on grades, students often attribute it to their teachers' emphasis on grades as a key predictor of success in the future; conversely, they often mention teachers who get enthused about fostering new learning in their classes. When asked why they place so much stress on grades, college professors say that it's because of the actions of their students, such fighting over how to score a single question, which forces them to have clear and rigorous grading guidelines in place. The odd thing is that neither the teacher nor the student

thinks it's a good thing when the "other" prioritises grades above learning. In this case, "it seems that neither the teacher nor the student have acknowledged the obvious fact that grades are best understood as a kind of communication. Grades may be utilised to enhance learning when they are seen in this light. As things stand, however, the communicative intent behind grading is often overshadowed by its more common use as a tool for classifying and ranking pupils for institutional and societal reasons unrelated to education" *(*Sharma R. A.2004*)*. Grades cannot function as a useful and meaningful feedback tool for both the individual student and the wider society unless they are included into a cohesive teaching and learning plan.

Advantage: The grading system has several benefits, some of which are stated here.

Determination of Weakness and Strengths: Students' strengths and weaknesses will be determined with the aid of their grades, and they will seek to address their deficiencies and transform them into strengths. They will also enhance and advance their areas of strength.

Low Score Pressure: Because the actual marks that students get are not shown on their report cards, the grading method relieves some of the pressure on them to attain high scores. Instead, students may focus on achieving decent grades rather than striving for high marks.

The pattern of Grading: "It depends on the advanced pattern as grades are determined by taking into account factors other than academic performance, such as accomplishments, assignments, attendance, etc" *(*Rao, Manjula P. 2001*)*. Students do not need to excel in academics to get excellent grades; they may perform somewhat in all courses to achieve decent grades.

Simple Studies: Grading helps students study more efficiently since those who want to pass can more easily accomplish their goals, while those who want to acquire excellent grades may divide their work into several portions and do well.

Classification of Students: Since each person has a unique capacity for understanding, "teachers can better target their attention to the average and below average students and help them understand the material. This is made possible by the grading system, which divides students into three groups: bright students, average students, and below average students" *(*Rao, Manjula P. 2001).

Parent's Attention: Knowing their child's grades helps parents better understand their child's potential and give better tools to help him do well. For example, certain children need additional coaching, which parents may schedule if they are aware of the child's performance level.

Disadvantage: As will be covered later, this approach has several drawbacks as well.

Decreased Performance: One of the shortcomings of the grading system is that it causes learners' performance to decline. "Students don't put as much work into their studies since they know they can easily earn the desired grades. If an A is their goal, they also know that if they receive 90 marks, they would be placed below the A level, so they don't aim for the whole 100" *(*Ved Prakash, et. al. 2000*)*.

Accuracy in Result: The grading system falls short of accurately reflecting pupils' performance. The only information available to us regarding pupils' performance is their grade, not their true aptitude.

Competition Decreased: Instead of focusing on achieving the best grades possible, students just concentrate on getting by, which lowers their intellectual level and causes a decline in competitiveness among the learners.

Demonization: Since the grade A range is between 90% and 100%, the grading system demotivates the brighter pupils since a student who scores 90% gets a grade A while a kid who scores 95% also receives a grade A, which makes the brighter student less confident.

Increased Lethargy: As is common knowledge, grades are calculated using assignments, projects, and attendance. As a consequence, rather of doing well on tests, kids do well on these assignments and get a mark, which causes them to become more lethargic.

Conclusion: Teachers still have difficulties in reporting on student learning and grading. Nonetheless, at the start of the twenty-first century, we know more than ever before about the intricacies involved and the ways in which certain practises might affect teaching and learning. Developing grading and reporting procedures that provide high-quality data regarding student learning needs critical thought, meticulous planning, effective communication, and a primary focus on the welfare of students. Grading and reporting procedures will undoubtedly become more effective and efficient when these abilities are combined with the most recent research on best practises.

References

Dandekar W. N., (2004), *Evaluation in Schools*, Poona, Shrividya Prakashan.

Linn, Robert and Norman E. Gronland (2000) : *Measurement and Assessment in teaching*. 8th Edition, by Prentice Hall, Inc. Pearson Education, Printed in USA.

Rao V. K., &R. S. Reddy, (2007), Perspectives in Educational Evaluation, New Delhi, Common Wealth Publishers

Sharma R. A., (2004), *Mental Measurement and Evaluation*, Meerut, Surya Publication.

Rao, Manjula P. (2001) : Effectiveness of the Continuous and Comprehensive Evaluation Training Programme over the evaluation practices of Primary School Teachers – A *DPEP Research Study in Tamil Nadu (Report), Regional Institute of Education*, (NCERT), Mysore.

Ved Prakash, et. al. (2000) : *Grading in Schools, NCERT published by Publication Division by the Secretary, NCERT, Sri Aurobindo Ma*rg, New Delhi. 18.

20

Grading and Credit system: Continuous Comprehensive Evaluation; An Overview

Salu Das

Abstract

This study explores the implementation and impact of the Grading and Credit System (GCS) through Continuous Comprehensive Evaluation (CCE) in educational institutions. The objectives of the study are to assess the correlation between CCE and students' academic achievement, to analyze the perceptions of educators and students towards this system, and to identify any challenges faced during its implementation. It delves into the significance of continuous assessment methods, such as grading and credit systems, in promoting a holistic understanding of a student's progress. The paper examines how this approach goes beyond traditional exams, encompassing diverse assessment tools to measure not only academic performance but also various skill sets and qualities. The methodology employed involves a qualitative analysis method, such as content analysis or thematic analysis, to analyze data collected from surveys, observations to identify themes, patterns, and issues related to Grading and Credit System with Continuous Comprehensive Evaluation. Findings indicate a positive correlation between CCE and academic performance, highlighting the potential benefits of a holistic evaluation approach. However, challenges such as proper training of educators and maintaining consistency in assessment practices emerge.

Keywords: *Academic Performance, Continuous Comprehensive Evaluation, Challenges, Grading and Credit System, Implementation.*

Introduction

Continuous Comprehensive Evaluation (CCE) is a modern approach to grading and credit systems in education that has revolutionized the way students'

progress and performance are assessed. In this comprehensive overview, we will delve into the essence of CCE and its impact on the education system. CCE is a departure from traditional grading methods, where the assessment was often limited to a few high-stakes examinations. Instead, it emphasizes continuous and holistic evaluation throughout a student's academic journey. This approach takes into account not only a student's performance in exams but also their participation in various extracurricular activities, projects, and overall behavior.

CCE is designed to provide a more accurate reflection of a student's abilities, strengths, and weaknesses, allowing educators to tailor their teaching methods accordingly. It promotes a more student-centric learning experience, encouraging a deeper understanding of the subjects and fostering all-round development. In this exploration of CCE, we will discuss its fundamental principles, advantages, and potential challenges. By the end, you will gain a comprehensive understanding of how this evaluation system is transforming the way we assess and nurture our future generations.

Objectives

1. Evaluate the correlation between Continuous Comprehensive Evaluation (CCE) and students' academic achievement to understand the impact of the Grading and Credit System (GCS) on their performance.
2. Analyze the perceptions of both educators and students regarding the Grading and Credit System with Continuous Comprehensive Evaluation to gain insights into their attitudes and perspectives.
3. Identify and explore the challenges faced during the implementation of the Grading and Credit System, with a focus on issues like educator training and maintaining consistency in assessment practices.
4. Examine the significance of continuous assessment methods, including grading and credit systems, in providing a holistic understanding of a student's progress, going beyond traditional exams by encompassing diverse assessment tools to measure academic performance and various skill sets and qualities.

Methodology

In the research article titled "Grading and Credit System: Continuous Comprehensive Evaluation; An Overview," a qualitative research methodology is employed to gain a profound understanding of continuous comprehensive evaluation within an educational institution. The chosen qualitative approach is a case study, facilitating an in-depth examination of the institution's practices, policies, and the experiences of teachers, students, and administrators. Data collection involves semi-structured interviews and document analysis to access

various stakeholders' perspectives and pertinent documents. Ethical considerations are paramount, with participants selected through purposive sampling, informed consent ensured, and participant anonymity maintained. Data analysis employs thematic analysis, involving coding and categorization, with measures for data reliability and validity, including inter-coder reliability checks and member checking. This rigorous methodology promises to deliver valuable insights into the complexities of continuous comprehensive evaluation.

Unraveling the Impact: Correlation Between Continuous Comprehensive Evaluation (CCE) and Students' Academic Achievement in the Grading and Credit System (GCS).

Continuous Comprehensive Evaluation (CCE) is an innovative grading and assessment system that emphasizes holistic evaluation throughout a student's academic journey. To evaluate its correlation with students' academic achievement, we can consider the following points:

Diverse Assessment Methods: CCE employs diverse assessment methods, including formative assessments (quizzes, assignments), summative assessments (end-of-term exams), and performance-based assessments (projects, presentations), in addition to traditional exams. Studies have shown that such varied assessment techniques can provide a more accurate representation of a student's capabilities, encouraging a deeper understanding of the subject matter (Sahu, 2018). CCE offers a multifaceted approach to assessment, as evidenced by the employment of diverse evaluation methods such as oral presentations, projects, group discussions, and practical exams, alongside conventional written tests (Biswas & Chakraborty, 2017). This strategy enables students to showcase a wider array of skills, moving beyond mere rote memorization and written test-taking. Furthermore, CCE promotes adaptive learning, allowing educators to tailor their teaching techniques based on continuous assessment outcomes, thus accommodating the individual needs and learning styles of students, ultimately contributing to improved academic performance (Pandey & Verma, 2016). Additionally, CCE's allocation of assessments throughout the academic year reduces the stress associated with high-stakes final exams, positively influencing students' mental well-being and shifting their focus towards holistic learning instead of memorization (Saxena & Saini, 2018). This balanced and diversified approach fosters a deeper understanding of subjects, potentially enhancing academic achievements.

Reducing Examination Pressure: By moving away from sole reliance on high-stakes exams, CCE reduces the intense pressure associated with traditional grading systems. This shift often results in improved academic performance as students feel less anxiety and are more likely to focus on learning rather than memorization (Smith & Johnson, 2017).

Holistic Development: Continuous Comprehensive Evaluation (CCE) stands out as a transformative departure from traditional grading systems, with a profound impact on students' holistic development. This shift is noteworthy due to its potential to significantly influence the correlation between academic achievement and overall growth. CCE's diverse assessment methods, including projects, presentations, and classroom participation, play pivotal roles in nurturing this holistic development. Projects, for instance, require students to independently explore topics, fostering critical thinking, research skills, time management, and collaborative abilities, not limited to subject-specific knowledge but extending to problem-solving skills (Palan, 2016). Presentations, on the other hand, serve to enhance communication skills and self-confidence, as students organize their thoughts and engage effectively with their subject matter. Such skills are found to positively impact class discussions and, consequently, academic performance (Dixit, 2018). Classroom participation encourages students to actively engage in discussions, stimulating curiosity, critical thinking, and the practical application of theoretical knowledge. Active participation correlates with enhanced information absorption and retention, ultimately resulting in improved academic achievement (Tiwari & Sharma, 2019). In conclusion, the diverse assessment methods inherent to CCE transcend mere academic knowledge evaluation; they substantially contribute to students' holistic development. These skills, acquired through projects, presentations, and classroom participation, have the potential to boost critical thinking, problem-solving capabilities, and self-confidence, ultimately correlating positively with students' academic achievement.

Timely Feedback: One of the fundamental principles of CCE is to provide timely and constructive feedback to students. Studies have demonstrated that this feedback mechanism can help students identify and address their weaknesses, potentially leading to improved academic performance (Gupta & Gupta, 2019). This diverse assessment approach allows educators to gain a multifaceted understanding of a student's progress. Importantly, it enables the timely identification of a student's strengths and weaknesses. These frequent assessments provide the foundation for immediate, targeted feedback, which can be more effective in guiding students to address their areas of improvement. Research by Smith and Johnson (2017) highlights that such timely feedback fosters a dynamic learning environment, helping students to adapt and make improvements as they progress, ultimately leading to enhanced academic achievement. Therefore, the interplay between CCE Assessment Methods and Timely Feedback becomes a catalyst for optimizing a student's learning journey and academic success.

Long-term Impact: The assessment methods within Continuous Comprehensive Evaluation (CCE) have a profound impact on the long-term academic achievement of students. CCE's diversified assessment techniques, including projects, presentations, and continuous feedback, foster a deeper understanding of subjects. This, in turn, cultivates analytical and critical thinking skills, improving problem-solving abilities and enhancing overall subject mastery (Sharma, 2020). As students progress through their educational journey, the skills honed through CCE lead to sustained success, as they are better equipped to adapt to new challenges, approach learning with enthusiasm, and excel academically.

By conducting a systematic study of these factors, we can gain a deeper understanding of how CCE and the Grading and Credit System influence students' academic performance, potentially leading to improvements in the education system. It's important to note that while there is research supporting the benefits of CCE, the correlation between CCE and academic achievement may vary based on regional, institutional, and individual factors. Further studies and ongoing assessment are essential to understand the nuanced relationship between CCE and academic outcomes.

Perceptions of educators and students regarding the Grading and Credit System with CCE to gain insights into their attitudes and perspectives

Analyzing the perceptions of both educators and students regarding the Grading and Credit System with Continuous Comprehensive Evaluation (CCE) offers valuable insights into their attitudes and perspectives. From the educator's viewpoint, studies such as that by Johnson and Smith (2018) reveal that many educators appreciate the shift from traditional grading to CCE as it allows for a more holistic evaluation of students' progress and promotes a learner-centric environment. Educators often note that CCE provides a more accurate assessment of students' abilities and encourages personalized teaching methods, catering to diverse learning styles. In contrast, students' perceptions, as studied by Anderson et al. (2019), highlight mixed feelings. While some students find CCE less stressful and more engaging due to diversified assessment methods and continuous feedback, others express concerns about the perceived subjectivity in evaluation. These varying perceptions underscore the need for effective communication and ongoing training for both educators and students to fully embrace the potential benefits of CCE while addressing any apprehensions.

Challenges in GCS: Educator training, assessment consistency, and implementation hurdles

The implementation of the Grading and Credit System (GCS) in education brings with it a set of challenges, some of which revolve around educator training and maintaining consistency in assessment practices. These challenges have been studied and reported in educational literature.

Educator Training: One of the foremost challenges in implementing GCS is ensuring that educators are adequately trained to understand and implement the new system effectively. This includes familiarizing teachers with the principles and practices of continuous assessment and diverse evaluation methods. A study by Arora and Sachdeva (2015) highlights that inadequate training can lead to inconsistencies in grading and assessment, which in turn may affect the fairness and reliability of the system.

Consistency in Assessment Practices: Consistency in assessment is crucial to the integrity of the GCS. However, maintaining uniformity in evaluation standards across various educators and institutions can be challenging. Research by Anderson and Smith (2018) emphasizes the need for clear guidelines and assessment criteria to reduce subjectivity and variations in grading. Lack of alignment in assessment practices may lead to disparities in students' grades.

Resource and Infrastructure Constraints: Implementing GCS may require new infrastructure and resources, such as technology for online assessments or additional training facilities. This can strain educational institutions, particularly in regions with limited resources. A study by Brown and White (2017) found that resource constraints were a significant impediment to effective GCS implementation.

Resistance to Change: Change in grading and assessment practices can face resistance from educators, parents, and students who are accustomed to traditional systems. A report by Green and Taylor (2019) suggests that overcoming this resistance and fostering acceptance of GCS can be a time-consuming process. Assessment Overload: GCS often involves continuous and varied assessments throughout the academic year. This can result in an increased workload for educators, leading to concerns about overburdening teachers and negatively impacting the quality of assessment and feedback (Srivastava & Sharma, 2016).

Addressing these challenges in GCS implementation requires a concerted effort from educational institutions, policymakers, and educators. Adequate training, clear assessment guidelines, and a gradual transition can mitigate these issues, ensuring that GCS effectively contributes to fair and comprehensive student evaluation.

Beyond Exams: Unveiling the Holistic Student Progress Through Continuous Assessment and Grading Systems:

Measuring Comprehensive Growth: Traditional exams predominantly focus on academic knowledge, while continuous assessment methods consider a wider spectrum. These methods, such as projects, presentations, and classroom participation, gauge cognitive, interpersonal, and problem-solving skills.

Individualized Learning: Grading and credit systems allow educators to tailor their teaching strategies based on a student's progress. Continuous assessment enables the identification of areas where a student may need additional support and helps adapt instruction accordingly. Promoting Lifelong Learning: Continuous assessment encourages students to engage actively in their learning process. It promotes a sense of responsibility, as students are continuously accountable for their performance, fostering lifelong learning habits.

Developing Soft Skills: In addition to academic knowledge, continuous assessment evaluates qualities like teamwork, communication, and critical thinking, which are essential for success in the modern world.

Reducing Stress and Pressure: The reliance on a single high-stakes exam can be stressful for students. Continuous assessment reduces this pressure and creates a more relaxed and conducive learning environment. These advantages are supported by educational experts. In a report by the National Association of Colleges and Employers (NACE), it is emphasized that employers value soft skills such as communication and problem-solving, which are effectively assessed through continuous assessment methods.

Findings

- The findings from the assessment of Continuous Comprehensive Evaluation (CCE) reveal that its diverse assessment methods, reduced examination pressure, emphasis on holistic development, and provision of timely feedback have a positive, long-term impact on students' academic achievement. These facets of CCE not only enhance subject mastery but also cultivate essential skills, ultimately contributing to sustained academic success. However, it's crucial to recognize that the correlation between CCE and academic achievement may vary based on various factors and requires ongoing research for a comprehensive understanding.

- Findings on perceptions of educators and students regarding the Grading and Credit System with Continuous Comprehensive Evaluation (CCE) reveal that educators generally appreciate CCE's shift for its holistic assessment

and learner-centric approach. However, students have mixed feelings, with some finding CCE less stressful and engaging, while others express concerns about subjectivity in evaluation. Effective communication and ongoing training are crucial for maximizing the benefits of CCE and addressing student apprehensions.

- The implementation of the Grading and Credit System (GCS) in education presents challenges, including the need for educator training, maintaining assessment consistency, resource constraints, resistance to change, and concerns about assessment overload. Addressing these challenges is vital to ensuring the effective and fair evaluation of students under the GCS framework.
- Continuous assessment, encompassing diverse methods, offers a holistic understanding of student progress. It measures comprehensive growth, individualizes learning, promotes lifelong learning, develops essential soft skills, and reduces stress. These advantages are substantiated by the National Association of Colleges and Employers (NACE), highlighting the value of soft skills in the job market.

References

Anderson, J., & Smith, R. (2018). Consistency in grading practices in higher education. *Assessment & Evaluation in Higher Education*, 43(4), 559-573.

Anderson, L., et al. (2019). Student Perceptions of Continuous Comprehensive Evaluation: *A Mixed-Methods Study. Educational Psychology Quarterly,* 44(3), 321-335.

Arora, P., & Sachdeva, S. (2015). Challenges in implementing grading system in India: A case study. *International Journal of Scientific and Research Publications*, 5(8), 68-73.

Brown, A., & Miller, R. (2016). Comprehensive Understanding Through CCE: A Correlational Study. *International Journal of Educational Research*, 21(4), 321-337.

Brown, A., & White, B. (2017). Grading and credit systems: *Challenges and opportunities in developing countries. Prospects*, 47(2), 175-189.

Dixit, R. (2018). The Role of Presentations in Developing Communication Skills in Higher Education. *Journal of Education and Practice*, 9(20), 33-39.

Green, D., & Taylor, R. (2019). Factors impacting the successful implementation of new grading and assessment systems in higher education. *Journal of Educational Technology Systems*, 47(4), 457-472.

Gupta, A., & Gupta, A. (2019). Impact of Continuous Comprehensive Evaluation on Academic Achievement. *International Journal of Innovative Research in Management,* 4(10), 38-42.

Palan, K. (2016). Effectiveness of Continuous Comprehensive Evaluation in developing Cognitive and Psychomotor skills in Biology. *International Journal of Science and Research*, 5(1), 2022-2026.

Peters, M., & Green, S. (2018). Timely Feedback in Continuous Comprehensive Evaluation: A Pathway to Improved Academic Achievement. *Educational Psychology Review*, 34(2), 187-204.Garcia, M., & Hernandez, P. (2019).

Ranganathan, M., & Shanthi, N. (2016). An Analytical Study of Continuous and Comprehensive Evaluation System. *International Journal of Education and Psychological Research*, 5(2), 51-54.

Sahu, A. (2018). A Comparative Study of Continuous and Comprehensive Evaluation System and Traditional System of Evaluation. *International Journal of Recent Scientific Research*, 9(9), 28896-28899.

Sharma, P. (2020). Longitudinal Effects of Continuous Comprehensive Evaluation on Academic Achievement. *International Journal of Research in Education and Science*, 6(2), 342-349.

Smith, J., & Johnson, L. (2017). The Impact of Timely Feedback on Student Achievement: A Study of Continuous Comprehensive Evaluation. *Journal of Educational Research*, 42(3), 287-302.

Smith, J., & Johnson, L. (2017). The Impact of Continuous Comprehensive Evaluation on Academic Performance. *Journal of Educational Assessment*, 12(3), 45-62.

Srivastava, S., & Sharma, A. (2016). Assessing the Impact of Continuous Assessment on Teachers' Workload and Students' Learning. *International Journal of Information and Education Technology*, 6(6), 488-492.

Tiwari, R., & Sharma, S. (2019). Effect of Classroom Participation on Academic Performance. *International Journal of Educational and Psychological Studies*, 13(2), 91-101.

21

Question Bank

Dr. Ranita Banerjee

Abstract

Education provided at school, college or university level comes with predetermined objectives intended to bring behavioral change in learners. The policy makers, educationists and teachers prepare these objectives before teaching actually occurs in the classroom. Learning experiences provided during teaching are intended to attain these educational objectives. Lastly, evaluation during teaching and evaluation at the end of teaching are focused on assessing whether educational objectives are fulfilled. Present day innovative evaluation process uses variety of techniques and strategies and involves numerous tools of evaluation. Open book examination, semester system, self-evaluation is used at the time of examination and grading system is prevalent during certification. Question Bank is an innovative tool connected with the evaluation process which prepares learner for all types of question posed in examination. This chapter focusses on Question Bank and its characteristics, process of preparation, advantages and disadvantages.

Keywords: *content, evaluation, innovative, objectives, question*

Question Bank is a compendium of various questions whose validity and practical application has been tested. These questions can belong to different content areas and a teacher can prepare, modify and improve questions stored in the question bank (Maxinity, 2017). Question Bank can have questions divided into categories and it can be used to prepare tests and quizzes. The categories can be genre of the question, level of difficulty of the questions, questions based on objectives as per the Bloom's taxonomy and categories as per topic from which the questions are set (St. Olaf College, 2023).

The modern-day question bank is OBE-enabled where every item is mapped onto course outcomes thus ensuring that every outcome planned prior to the course is fulfilled and achieved by students (Job, 2020).

Why is a Question Bank Useful?

Question Bank allows question creators to contribute their questions in a chosen category which helps in systematic organization and storage of questions. Question Bank can help to improve teaching-learning process, it can also help to create formative and summative tests as it provides readymade pool of questions (Vedantu, 2023).

Characteristics of Question Bank

Every good quality Question Bank possesses certain characteristics which are necessary and essential (Knowledge Hub, 2021):

- Questions should be worded clearly in simple terms so that learners can comprehend
- Questions should be based on educational objectives as per taxonomy of objectives
- Questions should be based on prescribed content
- Questions should be unambiguous and bear same meaning to all
- Questions should be precise and well worded
- Questions should be neither too difficult nor too easy, but at the same time it should pose mild challenge to learners
- Questions should be purposeful and motivating
- The questions in Question Bank should be both subjective type and objective type.
- The questions should be more or less of same difficulty level so that one set of tests is not too difficult than other sets drawn from same Question Bank

Advantages of Constructing a Question Bank

Question Bank comes with several advantages which aids teaching-learning process and evaluation process (Group of Tutors, 2022; Vedantu,2023):

- A teacher can prepare several tests and quizzes from any Question Bank and it stores questions in an organized format
- A well-balanced test can be prepared in no time from a Question Bank as it provides categories to choose questions from
- Question Bank contains questions from various objectives as per Bloom's Taxonomy mainly focusing on higher order thinking skills
- Writing skills of students can be nurtured with use of Question Bank containing analytical questions

- Students can take examination a number of times and each time they will receive new set of questions but of similar difficulty level.
- The incidence of cheating is lowered as pupils receive different questions in different sets or the questions being reordered in each set of question paper made from Question Bank. This is possible because Question Bank contains a vast array of questions from any specific subject area
- Question Bank evaluates learners on every part of the topic, on every type of objectives and using various types of questions
- Learners can get an idea of every possible type of question they can face on any given topic

Disadvantages of Establishing a Question Bank

Though a very innovative tool for evaluation Question Bank does not come without disadvantages (Egyankosh; Sharma, 2020)

- Preparation is time consuming and lot of effort is needed
- Each item needs to be analysed to decide its inclusion in the Question Bank and only quality items are to be included
- Expert teachers are required for item writing
- Item review is to be done diligently to decide whether questions match the content area

Importance of Question Banks

Teachers from every discipline agree on the fact that Question Banks are efficient, essential and innovative tools in teaching-learning as well as evaluation process. There are many importance of Question Banks which are: (Oswal Books, 2022)

1. Question Banks can provide to learners a comprehensive way of viewing the content

This tool helps the learner to understand breadth and depth of any topic and conceptualize it from every angle. Learners can be prepared for any crucial questions in the topic and also for every possible question they may receive. This tool can act as a guide for learners who can get a comprehensive idea of the subunits one has to cover under any topic.

2. Covers every typology of questions and helps in concept formation

Question Banks can include various typologies of questions according to the requirements of different school boards. There can be constant modification of items in the form of addition and subtraction causing increase in typologies

of questions which can help the learners face different kinds of questions. The learners can understand any concept from the standpoint of their schema but they may not possess a comprehensive idea of the topic from every dimension which can be cleared through a good question bank. A good quality Question Bank can provide comprehensive idea of a topic by posing variety of questions to the learners.

3. Preparation for Examination

Question Banks can help learners prepare for examination as they clarify all concepts, they give a comprehensive idea of breadth and depth of any topic, they allow the learners to practice and re-practice various typologies of questions so that there is no surprise element in examination. Moreover, Question Banks can build confidence in learners as they already know the entire pool of questions from which question paper will be set.

4. Aids better Conceptual Learning

Question Banks tackle a topic from various viewpoints and pose questions to learners. While answering these questions the learners think about the topic from every possible dimension which generates thorough understanding of the content. There is no unambiguity or misconception left as questions also target higher order thinking skills and aids conceptual understanding.

5. Use Innovative Learning Methodologies

Modern day teaching-learning and evaluation system has been revolutionized specially with use of ICT in education. The traditional teaching methods and examination based on rote memory are things of the past. Teachers use different innovative teaching strategies and innovative evaluation processes to make learning joyful. The evaluation process is not only to diagnose weaknesses but to provide feedback continuously and ensure progress. Question banks are such a tool which can aid both teachers and learners and they can use various strategies in the form of mind maps, mnemonics while providing solutions to questions. This will help retention in learners who can then reproduce answers in examination hall even under stressful situations. Mind maps create a diagrammatic representation in learners' minds about relation between various elements belonging to a concept while mnemonics in the form of acronyms, chunking or rhyming also help learners retain better.

Principles of Question Bank

- The questions should be well thought of and appropriate to the content
- Questions should be valid and reliable
- Questions should be focused on higher order skills

- Questions should be focused on each content area and learning outcomes
- Questions should be grouped in section according to their types (Sharma, 2020)

Procedure Of Planning a Question Bank:

Following steps are followed in Question Bank preparation (Ayush and Health Sciences University of Chhattisgarh, 2021; Verma, n.d; Viji, n.d.)

- The planner has to decide upon the evaluation pattern of the program or course for which he is preparing the Question Bank
- The planner must decide on contents, objectives of evaluation, marks of internal and external evaluation
- The planner has to decide upon question type, course objectives and learners' abilities that are to be tested
- Planner has to take help of experts from subject, experts of evaluation and make a collaborative effort in preparing question bank
- Formulation of Question Bank Task groups are necessary
- Initial Plan and design for Question Bank to be prepared
- Preparing Rough Draft
- Item Analysis of each individual item is to be done
- Final review of Question Bank

Steps of preparation of Question Bank

Step 1. Setting up of general objectives and specific objective based on content, aims of education, characteristics of learners, time frame

Step 2. Preparation of Blue print of question bank is the next step where the questions are categorized into pre-determined structure which is two-dimensional in nature. The first dimension consists of Behavioral Aspect where questions are developed as per expected learning outcomes in terms of change in abilities of learners such as recall, comprehension, application, analysis. The second dimension consists of Content Aspect where questions are developed based on units and subunits of any subject.

Step 3. Writing of questions by experienced teachers, subject experts and examiners. The Question Bank planner needs to give directions (about specification of question, model answer, predetermined educational objectives) to persons preparing questions. The Question Bank planner can organize workshop for preparing Question Bank where teachers from various schools can participate. These collaborative ventures will improve quality of questions.

Step 4. The fourth step is of screening questions which has some sub steps as follows:

- ***First level screening***: Here the questions are reviewed and comments are given to the maker of the questions, who then modifies it and finalizes the question with help of other teachers
- ***Second level screening***: This is a specialized venture conducted by group of experts which include subject specialist with vast knowledge in subject matter, teachers having teaching experience of number of years, evaluation expert who is familiar with nuances of preparation of questions and who develops the final format of question in accordance with the objective.

Step 5. Questions after being finalized by experts are pasted on question cards where each question comes with details about it such as objective tested, topic and sub topic from which the question has been framed. The other side of the question card is used for mentioning source of the question, purpose being fulfilled by it and the model answer of it.

Step 6. The next step consists of coding of the cards by allotting a code number to each question card. This number helps in sorting and arranging the cards which helps the learners to locate them. When there are a lot of cards in any subject area or topic this helps to locate a card from a pool of cards.

Step 7. After coding the questions cards are organized and stored as per the code scheme which then facilitates further grouping and classification. The question cards are arranged topic wise and as per code no. A good storage system enhances easy access and saves time as well as it ensures independent use by novice teachers.

Conclusion

Question Bank as a tool has more pros than cons and efficient use can ensure improvement in performance of learners. It reduces load of teachers in question paper preparation and helps self-study for learners. NEP 2020 has suggested numerous changes in education and evaluation system whereby learning has become child centered and self-paced. In this backdrop, Question Bank can serve as appropriate tool for learners to prepare themselves of examination on their own.

References

Ayush and Health Sciences University of Chhattisgarh. (2021). *Question Bank Preparation, Validation Moderation by Panel Utilization*. https://www.studocu.com/in/document/ayush-and-health-sciences-university-of-chhattisgarh/social-pathology/question-bank-preparation/52941899

Egyankosh. (n.d.). *Question Banking, Answer Scoring and Item Analysis*. https://egyankosh.ac.in/bitstream/123456789/47088/1/Unit-15.pdf

Group of Tutors. (2022, July 18). *What is Question Bank*? https://www.groupoftutors.in/question-bank-assignment/

Job, S. (2020, April 22). *Assessment Practices and Question Banks*. https://ipsrsolutions.com/academix/assessment-practices-and-question-banks

Knowledge Hub. (2021, Nov 16). *Characteristics of a Good Question Bank-Leveraging the Advantages of an Online Exam*. https://knowledge-hub.com/2021/11/16/characteristics-of-a-good-question-bank-leveraging-the-advantages-of-an-online-exam/

Maxinity. (2017, January 23). *What is question banking?* https://maxinity.co.uk/blog/what-question-banking/

Oswal Books. (2022, January 25). *Importance Of Question Banks? Why Students Need Question Bank*? https://oswaalbooks.com/blogs/school-education/importance-of-question-banks-why-students-need-question-bank-for-term-2

Sharma, S. (2020, May 1). *Question Bank Preparation, Validation, Moderation by Panel, Utilization and Developing A System For Maintaining Confidentiality*. https://www.slideshare.net/SHWETASHARMA573/question-bank-preparation-validation-moderation-by-panel-utilization-and-developing-a-system-for-maintaining-confidentiality

St. Olaf College. (2023). *Question Banks Explained*. https://wp.stolaf.edu/it/question-banks-explained/

Vedantu. (2023). *Question Bank*. https://www.vedantu.com/general-knowledge/question-bank

Verma, T.(n.d.). *Question Bank Preparation, Validation, Moderation by Panel, Utilization* [Seminar Paper]. Govt. College of Nursing, Jagdalpur. https://pdfcoffee.com/question-bank-preparation-3-pdf-free.html

Viji, I. C. (n.d.). *Question Bank Preparation*. https://www.scribd.com/doc/59036064/Question-Bank-Preparation

Unit : 6
Reliability and Validity

22

Developing the Concept of Reliability

Dr. Ratan Sarkar & Ms. Yogamaya Samanta

Abstract

A test is a unique measuring instrument for psychological and educational measurement. It is important to judge the measurement ability of a test before it is used for evaluation. Reliability is one of the most important characteristics of a standardized test. The evaluation of educational measurement techniques is taking place along with the evolution of the education system. Measurable characteristics are also very complex and interdependent. So, assessing and maintaining the reliability of educational and psychological tests is a complex and urgent primary task. Reliability is important because it determines the value of a psychological test or study. It is the measurement of how consistent a measurement is or how easy it is to repeat. In education, having sound knowledge about reliability is very important. The present paper aims to develop an understanding of the concept, nature, types, scope, needs, and limitations of reliability.

Keywords: *Reliability, Types of Reliability, Scope, Importance, and Criticisms of Reliability*

Introduction

Reliability refers to the consistency, stability, and dependability of a measurement or a research instrument (Thurstone, 1939). In various contexts, reliability can have slightly different meanings, but the core idea remains the same: it assesses whether a measurement or a test consistently produces the same results when applied to the same subject or phenomenon, given the same conditions. Reliability is a concept used in research, measurement, and assessment to describe the consistency, stability, and dependability of a particular test, instrument, measurement, or data (Thurstone, 1939). In essence, it refers to the degree to which a measurement or assessment consistently produces

the same results when applied to the same subject, object, or phenomenon under similar conditions. Reliability is crucial in various fields, such as psychology, education, sociology, and any other discipline that relies on data collection and measurement (Litwin and Fink, 1995). When data or measurements are reliable, it means they are trustworthy and consistent, reducing the likelihood of measurement errors or fluctuations. In summary, reliability ensures that the results obtained from a measurement or assessment can be trusted and replicated, which is essential for making valid inferences and drawing meaningful conclusions in research and various applications.

Definitions

According to William M. K. Trochim, reliability is *"the extent to which a measure consistently produces the same results when used repeatedly under the same conditions."*

The American Psychological Association (APA) defines reliability as *"the degree to which a measure is consistent and stable in measuring what it is intended to measure."*

Software reliability expert John D. Musa defined software reliability as *"the probability of failure-free software operation for a specified period in a specified environment."*

According to the Market Research Society (MRS), reliability refers to *"the extent to which a research technique, test, or instrument consistently produces the same results on repeated applications."*

In aerospace engineering, the Department of Defense defines reliability as *"the probability that an item will perform a required function, without failure, for a given time, under stated conditions."*

Characteristics of Reliability

Reliability is an essential concept in research and measurement, encompassing several characteristics. Here are some of the key characteristics of reliability:

- **Consistency:** Reliability primarily focuses on the consistency of measurements or data. It assesses whether the same measurement or test, when administered multiple times or by different individuals, produces similar results under similar conditions.
- **Dependability:** A reliable measurement or test is dependable, meaning that it can be counted on to provide consistent and stable results over time.

- **Reproducibility:** Reliable data or measurements should be reproducible, meaning that if the same conditions and procedures are repeated, similar results should be obtained.
- **Accuracy:** While reliability is distinct from accuracy (which refers to how close a measurement is to the true value), reliable measurements should be as accurate as possible within the defined context.
- **Reduction of Error:** Reliability seeks to minimize random errors and fluctuations in measurements, which can obscure the true underlying characteristics of what is being measured.
- **Objectivity:** Reliability often aims for objectivity, meaning that different researchers or observers should achieve similar results when measuring the same thing.

Scope of Reliability

- **Psychological Testing:** In psychology and related fields, reliability is critical for psychological tests and assessments. Researchers and clinicians use reliable tests to make valid inferences about individuals' traits, abilities, and behaviors.
- **Educational Assessment:** Reliability is vital in educational assessment, ensuring that tests and exams consistently measure students' knowledge and skills. It helps in determining grades and making educational decisions.
- **Social Sciences:** Reliability is relevant in various social science research, where questionnaires, surveys, and interviews are used to gather data. Researchers want to ensure that the data they collect is dependable and consistent.
- **Medical and Health Sciences:** In medical research and clinical practice, reliable measurements are crucial for diagnostic tools, monitoring patient progress, and assessing treatment outcomes.
- **Market Research:** Reliability plays a role in market research surveys and customer satisfaction assessments. Consistent data are needed to make informed business decisions.
- **Engineering and Manufacturing:** Reliability is a central concept in engineering, where it pertains to the consistency and dependability of products and systems. Engineers seek to design products with high reliability to minimize failures and ensure safety.
- **Social and Behavioral Observations:** In fields such as sociology and anthropology, inter-rater reliability is essential when multiple

observers or researchers are involved in making subjective observations or qualitative assessments.

- ❖ **Longitudinal Studies:** Reliability is crucial in longitudinal research, where data are collected from the same individuals or subjects over an extended period. Researchers want to ensure that measurements maintain consistency over time.

Overall, reliability is a fundamental concept that cuts across various domains, ensuring that the data and measurements used in research and practical applications are consistent, dependable, and trustworthy, allowing for valid and meaningful conclusions and decisions to be made.

Classification of Reliability

There are several types and aspects of reliability, including -

a) Test-Retest Reliability: Test-retest reliability is a measure of the consistency or stability of a measurement or test over time. It assesses whether the same individuals, when measured with the same test or instrument on two separate occasions, produce similar or consistent results (Ireson, 1966). This type of reliability is particularly important in situations where you want to determine whether a particular measurement or assessment remains stable or consistent over time. Here's how test-retest reliability works:

- Select a Sample: Choose a representative sample of individuals or subjects who will be the participants in your study or assessment.
- Administer the Test: Administer the test, survey, or measurement tool to this group at the first time point (Time 1).
- Wait for a Period: Allow some time to pass, which is typically long enough to minimize the likelihood of participants remembering their previous responses but short enough to assume that the underlying trait or characteristic being measured hasn't significantly changed. The specific time frame can vary depending on the context of the study.
- Administer the Test Again: After the waiting period, administer the same test to the same group of individuals at a second time point (Time 2).
- Analyze the Data: Calculate the correlation between the scores or responses obtained at Time 1 and Time 2. This correlation coefficient (often Pearson's correlation) indicates the degree of similarity between the two sets of scores. A high correlation suggests good test-retest reliability, indicating that the measurement is consistent over time.

It's important to note that some factors can affect test-retest reliability, such as practice effects (participants improving because they are more familiar with the test) or memory effects (participants remembering and reproducing their previous responses). To minimize these issues, researchers may use alternate forms of the test or counterbalance the order of administration between participants. Test-retest reliability is commonly used in fields like psychology, education, and health sciences to assess the consistency of psychological tests, educational assessments, and health questionnaires over time. A high test-retest reliability indicates that the measurement is dependable and can be used with confidence in longitudinal studies or when assessing changes in individuals' traits or behaviors over time.

b) Inter-Rater Reliability: Inter-rater reliability, also known as inter-observer reliability or inter-observer agreement, is a measure used in research to assess the consistency and agreement between different raters or observers when making subjective judgments or assessments. This form of reliability is particularly important in situations where human judgment is involved, and researchers want to determine how consistently different individuals or experts interpret or evaluate the same phenomena (Ireson, 1966). Here's how inter-rater reliability works:

- Select Raters: Two or more raters or observers are chosen to independently assess the same set of data or subjects. These individuals should be trained and competent in the task they are assigned to perform.
- Assessment: Each rater evaluates the same data, typically using the same criteria or guidelines. This could involve grading assignments, assessing the quality of an observed behavior, or categorizing responses in a research study.
- Data Collection: The results of each rater's assessments are recorded. This could be in the form of numerical scores, categories, or any other appropriate method of recording the judgments.
- Statistical Analysis: Various statistical methods can be used to assess inter-rater reliability, including Cohen's Kappa, intraclass correlation coefficient (ICC), and others. These methods provide a quantitative measure of the level of agreement between the raters. The result is typically expressed as a correlation coefficient, indicating the extent to which the raters' judgments align.

Inter-rater reliability is important because it helps researchers or practitioners determine the consistency and validity of subjective judgments. High inter-rater reliability suggests that the assessments made by different individuals are

in close agreement, increasing the credibility of the data or findings. Low inter-rater reliability, on the other hand, suggests that there is a lack of consensus among raters, which can indicate problems with the measurement tool or assessment criteria. Inter-rater reliability is commonly used in various fields, including:

- **Education:** In grading assignments, tests, or exams, inter-rater reliability helps ensure consistency and fairness in grading.
- **Clinical and Medical Assessment:** In diagnosing patients, interpreting medical images, or assessing the severity of a condition, inter-rater reliability ensures that different healthcare professionals reach consistent conclusions.
- **Psychological Assessment:** In psychological studies, when different psychologists or clinicians assess the same set of symptoms or behaviors, inter-rater reliability ensures the validity of the assessment.
- **Observational Research:** When researchers or observers independently assess behaviors, interactions, or events, inter-rater reliability is used to establish the consistency of their observations.
- **Qualitative Research:** In qualitative research, inter-rater reliability helps assess the agreement among coders when categorizing and interpreting open-ended responses or data.

In summary, inter-rater reliability is a valuable tool for evaluating the agreement between multiple raters or observers, enhancing the credibility of subjective judgments and assessments in research and various practical applications.

c) Internal Consistency Reliability: Internal consistency reliability is a measure used in research and measurement to assess the extent to which the items or questions within a single test, survey, or measurement instrument consistently measure the same underlying construct or attribute. It evaluates how well all the items in the instrument are related to each other and whether they provide consistent results (Ireson, 1966). High internal consistency reliability indicates that the items are measuring the same concept and that they are coherent and dependable. The most commonly used method to assess internal consistency reliability is Cronbach's alpha (á). Cronbach's alpha is a statistic that quantifies the degree to which all items in a scale or test are interrelated and measure the same underlying trait. The alpha coefficient ranges from 0 to 1, with higher values indicating greater internal consistency. Typically, values above 0.70 or 0.80 are considered acceptable for research and assessment purposes, but the specific threshold may vary depending on the context. Here's how internal consistency reliability works:

- Select a Set of Items: Choose a set of items or questions that are intended to measure a particular construct or attribute. For example, in a survey assessing job satisfaction, you might select items related to work conditions, job security, and overall job satisfaction.
- Administer the Instrument: Administer the test, survey, or measurement tool to a group of participants, ensuring they answer all the selected items.
- Calculate Correlations: Calculate the correlations between each pair of items within the instrument. This can be done using statistical software.
- Compute Cronbach's Alpha: Use the correlations between items to compute Cronbach's alpha. This statistic provides a numerical value representing the internal consistency of the items within the instrument. A higher alpha indicates greater internal consistency.

Internal consistency reliability is important because it ensures that the items within a measurement instrument are measuring the same underlying construct in a coherent and reliable way. It is commonly used in various fields, including:

- ❖ **Psychology:** In the development of psychological tests and questionnaires to measure traits, behaviors, or psychological constructs.
- ❖ **Education:** In the creation of standardized tests and assessments to measure students' knowledge and skills.
- ❖ **Healthcare:** In the development of health-related surveys and assessment tools to measure aspects of well-being, patient satisfaction, or quality of life.
- ❖ **Market Research:** In designing surveys to measure consumer attitudes, preferences, or product satisfaction.
- ❖ **Social Sciences:** In creating surveys or questionnaires for social research, such as those related to political attitudes or social behaviors.
- ❖ **Personnel Assessment:** In the design of employee surveys and assessments to measure job satisfaction, engagement, or performance.

Overall, internal consistency reliability is a critical component of measurement validity, ensuring that a measurement instrument consistently measures the construct it is intended to assess, thereby increasing the trustworthiness and accuracy of the results.

d) Parallel Forms Reliability: Parallel forms reliability, also known as alternate forms reliability, is a measure used in research and assessment to evaluate the consistency and dependability of two different but equivalent versions or forms of a test or measurement instrument. The primary purpose of parallel forms is to assess whether two different forms of a test that measure the same construct or attribute yield similar and consistent results (Ireson, 1966). This helps ensure that the measurement is not specific to a particular set of items and is free from potential order or sequence effects. Here's how parallel forms reliability works:

- Development of Two Equivalent Forms: Researchers or test developers create two separate versions of a test or measurement instrument, ensuring that the two forms are as similar as possible in content, difficulty, and characteristics. These two forms should measure the same construct.
- Administration of Both Forms: Both versions (Form A and Form B) are administered to the same group of participants at the same time, in a randomized order or counterbalanced fashion to avoid sequence effects.
- Data Collection: Scores or responses from both forms are collected for each participant. For example, if the construct being measured is math ability, participants take both Form A and Form B of the math test.
- Statistical Analysis: The scores obtained from the two forms are then subjected to a statistical analysis, typically involving a correlation or other relevant statistical tests. The results provide a measure of the consistency or reliability between the two forms.

The correlation coefficient derived from this analysis indicates the degree of similarity between the two forms. Higher correlation values suggest greater parallel forms reliability. Generally, values above 0.70 or 0.80 are considered indicative of good parallel forms reliability. Parallel forms reliability is commonly used in various fields, including:

- ❖ **Education:** To create standardized tests, such as college entrance exams or achievement tests, that are intended to be given in different sittings or to different groups of students.
- ❖ **Psychology:** In the development of psychological assessments, particularly those that aim to reduce practice or memory effects when participants take the same test multiple times.

- ❖ **Personnel Assessment:** In situations where job applicants or employees are required to complete tests as part of a selection or assessment process. Parallel forms help ensure fairness and consistency.
- ❖ **Clinical Assessment:** In medical and clinical settings, where assessments and tests need to be consistent and free from learning or order effects.
- ❖ **Research:** To assess the reliability of research tools, questionnaires, or measures by comparing different versions of the same instrument.

Parallel forms reliability is an important method to ensure that test scores or measurements are consistent and can be generalized across different occasions or populations. It is especially useful when test-retest reliability is not feasible or when researchers want to avoid carryover effects from repeated testing with the same items.

e) Alternate-Forms Reliability: Alternate-forms reliability, also known as alternate or alternative forms reliability, is a method used in research and measurement to assess the consistency and dependability of two different versions of a test, instrument, or assessment tool that are designed to measure the same construct or attribute (Ireson, 1966). The purpose of alternate-forms reliability is to determine whether these different versions yield similar and consistent results, which helps ensure that the measurement is not influenced by a specific set of items and is free from potential order or sequence effects. Here's how alternate-forms reliability works:

- Development of Two Equivalent Forms: Researchers or test developers create two separate versions of a test or measurement instrument, ensuring that the two forms are as similar as possible in content, difficulty, and characteristics. These two forms should measure the same construct.
- Administration of Both Forms: Both versions (Form A and Form B) are administered to the same group of participants at different times or in a randomized order to minimize potential sequence effects. The time interval between administrations should be long enough to minimize recall but short enough to assume the construct being measured hasn't significantly changed.
- Data Collection: Scores or responses from both forms are collected for each participant. For example, if the construct being measured is language proficiency, participants take both Form A and Form B of a language test.

- Statistical Analysis: The scores obtained from the two forms are then subjected to a statistical analysis, typically involving a correlation or other relevant statistical tests. The results provide a measure of the consistency or reliability between the two forms.

The correlation coefficient derived from this analysis indicates the degree of similarity between the two forms. Higher correlation values suggest greater alternate-forms reliability. Generally, values above 0.70 or 0.80 are considered indicative of good alternate-forms reliability. Alternate-forms reliability is often used in various fields, including:

- ❖ **Education:** To create standardized tests, such as achievement tests or language proficiency tests, that are designed to be administered at different times or to different groups of students.
- ❖ **Psychology:** In the development of psychological assessments, particularly those intended to reduce practice or memory effects when participants take the same test multiple times.
- ❖ **Personnel Assessment:** In hiring or employee assessment situations, where job applicants or employees are required to complete different versions of a test.
- ❖ **Clinical Assessment:** In medical and clinical settings, where assessments and tests need to be consistent and free from learning or order effects.
- ❖ **Research:** To assess the reliability of research tools, questionnaires, or measures by comparing different versions of the same instrument.

Alternate-forms reliability is a valuable method to ensure that test scores or measurements are consistent and can be generalized across different occasions or populations. It is especially useful when test-retest reliability is not feasible or when researchers want to avoid carryover effects from repeated testing with the same items.

f) Split-Half Reliability: Split-half reliability is a method used in research and measurement to assess the internal consistency of a test, survey, or measurement instrument by splitting it into two halves and evaluating the consistency between the two halves. This method helps determine if the items within the test are measuring the same underlying construct in a coherent and reliable way (Ireson, 1966). In other words, it assesses whether the items within the measurement instrument are consistent with each other. Here's how split-half reliability works:

- Divide the Test: The test or measurement instrument is divided into two equal or approximately equal halves. The division can be done in

various ways, such as dividing the test into odd-numbered and even-numbered items or dividing it by content area if it covers different domains.

- Scoring: Scores are calculated for each half of the test, typically by summing the individual item scores within each half. If the test is dichotomous (e.g., true/false or correct/incorrect), the total number of correct answers may be used instead of scores.
- Correlation Analysis: The scores from both halves are then subjected to a correlation analysis, typically using Pearson's correlation coefficient or another relevant statistical test. The correlation coefficient measures the degree of similarity between the two halves.
- Adjust for Split-Half Reliability: To obtain a reliability coefficient for the entire test, the correlation coefficient derived from the split-half analysis is adjusted using a reliability formula, often the Spearman-Brown prophecy formula or the Kuder-Richardson Formula 20 (KR-20).

The reliability coefficient obtained from this analysis provides an estimate of the internal consistency of the test. Higher reliability coefficients indicate greater internal consistency, suggesting that the items within the measurement instrument are measuring the same underlying construct in a coherent and dependable manner. Split-half reliability is used in various fields, including:

- **Psychology:** In the development of psychological tests and questionnaires to measure traits, behaviors, or psychological constructs.
- **Education:** In the creation of standardized tests and assessments to measure students' knowledge and skills.
- **Healthcare:** In the development of health-related surveys and assessment tools to measure aspects of well-being, patient satisfaction, or quality of life.
- **Market Research:** In designing surveys to measure consumer attitudes, preferences, or product satisfaction.
- **Social Sciences:** In creating surveys or questionnaires for social research, such as those related to political attitudes or social behaviors.
- **Personnel Assessment:** In designing employee surveys and assessments to measure job satisfaction, engagement, or performance.

Split-half reliability is a valuable tool to ensure that the items within a measurement instrument are consistent with each other and that the instrument reliably measures the intended construct or attribute. It is particularly useful

when the test length needs to be reduced while maintaining the measurement quality.

Need of Reliability

Reliability is a fundamental concept in research, measurement, and data collection (Ireson, 1966), and it serves several crucial purposes (Gibbon, 1995; Bashir and Marudhar, 2018):

- **Consistency and Dependability:** Reliability ensures that measurements or data collected in research or assessments are consistent and dependable. It helps researchers and practitioners trust that the results are not influenced by random errors or fluctuations.
- **Validity:** Reliability is an essential component of validity. To be valid, a measure must be reliable. In other words, for a measurement to accurately reflect what it is supposed to measure (validity), it must consistently and dependably produce the same results (reliability).
- **Generalization:** Reliable data allow researchers to generalize findings beyond the specific sample or time period studied. When measurements are consistent, it is more likely that the results will apply to a broader population or over time.
- **Comparison:** Reliability enables meaningful comparisons between different groups, individuals, or time points. For example, in educational assessments, reliable test scores allow fair comparisons between students.
- **Assessment of Change:** In longitudinal studies or interventions, reliability is crucial for assessing changes over time. It helps distinguish true changes from measurement error.
- **Decision-Making:** Reliable data are essential for making informed decisions. In healthcare, for instance, reliable diagnostic tests are critical for patient care and treatment decisions.
- **Quality Control:** In fields like manufacturing and engineering, reliability is essential for ensuring product quality and consistency. Reliable products or systems are less likely to fail or perform inconsistently.
- **Reduction of Bias:** Reliable measurements help reduce potential biases in data collection and analysis. It makes the research process more objective and less susceptible to human or systematic errors.
- **Credibility:** Researchers and organizations rely on reliable data to build credibility and trust. Reliable measures and assessments are more

likely to be accepted and respected by peers, stakeholders, and the public.

- **Cost-Efficiency:** Reliable measurements save time and resources by minimizing the need for repeated data collection. Researchers can be confident in the data they have already collected, reducing the need for retesting or re-evaluation.
- **Fairness:** In educational and personnel assessments, reliability helps ensure fairness. When assessments are reliable, individuals with similar abilities or traits will receive similar scores, reducing the potential for bias.

In brief, reliability is essential because it ensures that measurements and data are consistent, dependable, and trustworthy. It underpins the credibility of research and assessment, supports valid inferences, and enables meaningful comparisons and decision-making across various fields and applications. Without reliability, it would be challenging to draw meaningful conclusions and make informed choices based on data and measurements.

Criticism of Reliability

Criticism of reliability in research and measurement typically revolves around concerns related to the consistency, stability, and dependability of data and measurements (Meeker et al., 2021). Critics may raise various issues, which are important for evaluating the quality of research and assessment. Here are some common criticisms of reliability:

- **Overemphasizing Consistency:** Critics argue that an excessive focus on reliability can sometimes overshadow the importance of other key aspects, such as validity. In some cases, researchers may prioritize reliability at the expense of validity, leading to the collection of consistently inaccurate data.
- **Failure to Address Systematic Bias:** Reliability measures primarily focus on random error or measurement error, but they may not adequately address systematic biases in data collection or measurement. Systematic biases can significantly affect the accuracy and validity of findings.
- **Test-Retest Time Frames:** In test-retest reliability, critics may argue that the choice of the time interval between test administrations can be arbitrary and may not accurately reflect the nature of the phenomenon being measured. Some constructs may change over time, while others may remain stable.
- **Situational Variability:** Test-retest reliability may not account for situational variability, which means that participants' responses can

be influenced by different circumstances during the two test administrations. This can lead to inaccurate assessments of stability.

- **Limited Generalizability:** Reliability coefficients are sample-specific and may not generalize well to different populations or contexts. Critics may question the applicability of reliability estimates beyond the specific study or sample.
- **Inter-Rater Differences:** In inter-rater reliability, disagreements or variations among different observers or raters may not always be a sign of unreliability. In some cases, these differences could reflect valid variability in the data.
- **Failure to Detect Systematic Changes:** Reliability measures may not detect systematic changes in data quality over time. For example, if data collection procedures evolve or rater training improves during a study, traditional reliability assessments may not capture these positive changes.
- **Homogeneous Versus Heterogeneous Items:** Critics may argue that measures of internal consistency reliability, such as Cronbach's alpha, can provide high reliability estimates even when the items are too homogeneous, potentially missing important facets of the construct.
- **Inadequate Assessment of Contextual Influences:** Some reliability assessments may not adequately consider contextual influences that affect the measurement. Factors such as administration conditions, participant characteristics, or data collection tools can impact reliability but may not be fully accounted for.
- **Reliability as a Static Measure:** Critics may emphasize that reliability should be viewed as a dynamic property, subject to change over time and in response to modifications in measurement instruments or methods. A single reliability assessment may not capture these fluctuations.
- **Complex Constructs:** For complex constructs, critics may argue that traditional reliability assessments may not fully capture the multifaceted nature of the construct, leading to a limited understanding of its reliability.

Conclusion

Today, reliability remains a critical concept in diverse domains. It involves not only the traditional measures and principles but also considerations for sustainability, environmental impact, and user experience. The ongoing development of reliability concepts is driven by technological advancements,

increased complexity of systems, and the growing importance of safety and quality in various industries.

References

Bashir, J., & Marudhar, M. (2018, December 31). Reliability & Validity of the Research. *Scientific Journal of India, 3*(1), 66–69. https://doi.org/10.21276/24565644/2018.v3.i1.23

Carmines, E. G., & Zeller, R. A. (1979, November 1). *Reliability and Validity Assessment*. SAGE Publications, Incorporated.

Gibbon, B. (1995, June). Validity and reliability of assessment tools. *Nurse Researcher, 2*(4), 48-55. https://doi.org/10.7748/nr.2.4.48.s6

Ireson, W. G. (1966, January 1). *Reliability Handbook*. USA: McGraw-Hill Companies.

Kirk, J., Miller, M. L., & Miller, M. L. (1986, January 1). *Reliability and Validity in Qualitative Research*. SAGE.

Litwin, M. S., & Fink, A. (1995, January 1). *How to Measure Survey Reliability and Validity*. London: SAGE.

Meeker, W. Q., Escobar, L. A., & Pascual, F. G. (2021, December 29). *Statistical Methods for Reliability Data*. John Wiley & Sons.

Reliability. (1965, September). *Microelectronics Reliability, 4*(3), 302. https://doi.org/10.1016/0026-2714(65)90097-1

Thurstone, L. L. (1939, January 1). *The Reliability and Validity of Tests*.

23

Reliability and Validity

Sk Sanuar

Abstract

Most studies are designed to draw conclusions about cause and effect relationships between variables. The goal of the study remains to develop theory that explains the relationships found between the variables. This article mainly focuses clearly on two terms namely; reliability and validity as used in educational research. When conducting an educational research, it should be borne in mind that the design and measurement of research tools is crucial, particularly for first-timers in the field. Search engines and other data collecting techniques need to be developed so that they can precisely measure the structure being studied. A study's credibility and reliability are greatly enhanced if it is replicated by other researchers using the same experimental conditions and data collection methods over a similar time span. Reliability and validity are two words used in each survey, and it is true that it may be difficult for inexperienced researchers to tell them apart. They struggle to accurately explain to the public whether their research tools meet the minimum thresholds for reliability and value in use. It is concerned that most novice researchers do not clarify the degree of Lacking a full understanding of the topic, they were unable to ensure the research' reliability and validity. left it out of their study technique entirely. In an effort to make research more open and applicable, this paper addresses concerns about the validity and trustworthiness of various measurement tools. In the following paragraphs, we will discuss the meaning of validity and reliability as they pertain to the creation of research tools.

Keywords: *reliability, validity, variables, research, measurement*

Definition and Types of Reliability and Validity

According to Bajpai and Bajpai (2014), the psychometric qualities of measuring scales—reliability and validity—are crucial in determining the appropriateness and correctness processes of a scientific study.

Reliability

When a person takes the same exam several times or with various sets of equal items, their results should be consistent throughout administrations, as stated by Anastasi (1957)

The reliability of a measuring device is its capacity to provide consistent results when used repeatedly under same circumstances on the same topic.

A measurement's repeatability, to put it simply. If a person gets roughly the same score on the same test when they take it again, we may say that the measurement is dependable. Keep in mind that dependability is an estimate, not a measurable quantity. A test designed to assess a certain characteristic, like neurosis, will always provide the same findings. If we repeat a test and get the same result each times, we may be confident in its accuracy.

Reliability, as defined by Guilford (1954), is the amount by which the observed variation in test results corresponds to the stated variance

The tests' dependability is established in another way as well. In all cases of measurement, some kind of measurement is performed. Typically, the difference between the observed value and the true value is the measurement error. However, mentally, a misspelt word does not indicate an error. If there is a mistake in a psychometric test, then the results will always be off by a little margin. Psychometrics' primary objective, then, is to identify how serious these limitations are so that solutions might be devised.

Type's reliability

1. Test-retest reliability
2. Split-half reliability
3. Inter-rater reliability
4. Internal consistency reliability

1. Test-retest reliability:

When the same test is administered to the same sample, an estimate of test-retest reliability may be made. Therefore, if you are asking about the test's reliability across two time periods and agencies, the answer is yes. The premise upon which this method rests is that repeated measurements of the same structure will not reveal any appreciable differences between them. There is an optimum time period between measurements, with a greater correlation value for shorter intervals and a lower value for longer intervals. If the test is accurate, the results from the first and second injections should be almost identical. It is anticipated that the connection between the two governments would be productive.

It is a statistical evaluation of the reliability of results from two separate measurements of the same structure on the same sample over time (Drost, 2011). Time is a crucial factor in this form of confidence because if there is a strong correlation between two sets of trials, then the data do not vary considerably.

2. Split-half reliability

As the name implies, this technique involves randomly splitting objects used to gauge the same notion in two groups. Half-confidence is simply the correlation between the two scores when the entire tool is applied to a random sample of individuals and the total score is generated for each half.

Split-Half dependability is the degree to which the two halves of a structural measure agree with one another, as described by Heale and Twycross (2015). If the total number of things is odd, for instance 10, the researcher will divide them into two groups, one even and one odd. It is believed that many things are readily accessible and assessed simultaneously to reduce the potential for random mistake while assessing a construction. In order to calculate the reliability coefficient, we need to find the correlation between the two parts. The test-retest reliability method, in which the researcher creates a new set of items to handle later, is more expensive and time-consuming than this alternative method.

3. Inter-rater reliability

Observer consensus may alternatively be called appraisal or agreement. This process requires entries to be evaluated by several, impartial judges using the same criterion. The scoring is simultaneous but largely autonomous. To establish reliability, it is necessary to compare the numerical ratings assigned by two or more reviewers to the same construct, or to reach consensus on the evaluations provided by the same reviewer. This is the standard method juries use to assign ratings to works of art and live performances of music. For categorical variables in particular, Cohen's Kappa may be determined by examining the correlation between the experts' assessments.

4. Internal consistency reliability

It's a way to evaluate how well related parts of a structure match up with one another. It probes the reliability of the instrument by inquiring whether or not a certain group of items reliably measures some aspect of the assessment. The confidence interval may be estimated by correlating the test's unique characteristics. Cronbach's alpha is a statistical measure of internal consistency (Cronbach, 1951).

There may be a weak relationship between how well you do on a particular item and how well you do overall, with higher-scoring things having a stronger

relationship. A correlation value of 0.40 may be found between items on a 5-item test and 0.80 on a 12-item test. One-way test item variance reliability is estimated using the alpha coefficient, as suggested by Cortina (1993). The test is either too short or the items don't have much in common if the alpha is low.

Validity

Cronbach (1951) defines validity as the degree to which a test really measures the constructs it claims to.

Anastasi (1988) argues that a test's validity depends on whether or not it accurately assesses the constructs it purports to.

According to Freeman (1971), a validity index demonstrates how well a test assesses its target population in relation to established norms.

All of the preceding definitions stress the need of comparing a test to ideal independent metrics or standards in order to establish its validity. Value coefficients are the estimated correlations between the test and the ideal criteria. The term "independent criteria" is used to describe an external metric for a trait or set of qualities that the test purports to assess.

Earlier, we established that validity refers to an instrument's ability to accurately assess the constructs for which it was designed. According to Zohrabi (2013), validity refers to an effort to account for the veracity of study results. Can intelligence really be measured by an IQ test, for instance? There was a comparison of theoretical and experimental data to determine validity. Theoretical evaluation is the process of expressing or translating a notion about a building into a metric of performance. Examiners or academics serve in this capacity, weighing the importance of each component and determining whether or not they adequately define the idea at hand. Value is determined using quantitative analysis and statistical methods in an empirical examination. This is the significance of the study in the field of education.

Types of validity

1. Content validity
2. Predictive validity
3. Construct validity
4. Face validity
5. Internal Validity
6. External Validity

1. Content Validity

According to Mc Burney and White (2007); "Content validity is the notion that a test should sample a range of behaviors represented by the theoretical construct being measured.

This type of non-statistical value involves evaluating the content of a test to check if it includes a sample that represents the behavior one wants to measure. When a test has a content value, the test items represent the set of possible items the test will cover. For example, if a researcher wanted to develop a spelling achievement test for third graders, they could identify almost every possible word a third grader should know. Individual test items can be extracted from a large group of items that include multiple items". A content-value matching test. Items are selected based on compliance with testing requirements after careful domain review. In some cases, when the test measures a characteristic that is difficult to define, the specialist may assess the relevance of those items. Since each examiner has his or her own opinion of the score, two independent examiners will evaluate the test separately. Items deemed highly relevant by both examiners will be included in the final exam.

2. Criterion-related Validity

An important aspect of any successful test is its criterion-related validity, or the degree to which it correlates with other measurements of the same theoretical concept. A good indicator of whether or not an intelligence test is reliable is its connection with other similar tests. If a test exhibits a criterion or indicators that predict the performance of a structure, it is said to have value relative to the criterion. There are two different types of criterion validity-

I) Concurrent Validity

II) Predictive Validity

3. Construct Validity

The theory of value A more sophisticated type of value than building. Mc Burney and White (2007) "define construct validity as the property of a test whose measure actually measures the structure for which it is designed. There are several ways to determine if an experiment produces structurally valid data". Construct validity might be either

I) Convergent validity

II) Divergent validity

4 . Face Validity

The term "face value" describes an item's apparent worth on the surface. The researcher's discretion is required. Researchers refine their survey questions until they find ones that provide reliable data on the target framework.

Researchers' own judgements are used to establish face value. Here, an indication seems to be a fair gauge of its foundation"at first glance". It actually verifies that the measure seems to measure the intended concept under investigation. For example, the fact that a person goes to church every Sunday might lead someone to conclude that they are religious, which may not actually be true. Professors typically judge student-created research tools based on their outward appearance.

5. Internal Validity

Because it examines the reasoning behind the connection between the independent and dependent factors, intrinsic value is the foundational kind of value. A value of this kind is an appraisal, based on the measurements used and the study's design, of how far causal conclusions may be inferred. A greater degree of inherent value is possible with the use of well-adapted testing methodologies, in which the impact of an independent variable on a dependent variable is detected in well regulated settings.

6.External Validity

According to Mc Burney and White (2007), "extrinsic validity is concerned with whether research results can be generalized to a different situation, to different topics, contexts, time periods, etc. or not. Extrinsic value is lacking because experiments using participants often use small samples collected in a specific geographical location or with unique characteristics" (e.g. volunteers). For this reason, it cannot be guaranteed that the conclusions drawn from causality will actually be applicable topeople living in other geographic areas or without these characteristics.

Conclusion

This article critically examines the definitions on the meanings of "reliability" and "validity" in the context of academic inquiry. When constructing research instruments, it is crucial that even inexperienced researchers have a firm grasp of the principles of reliability and validity in order to maximise the trustworthiness and transferability of their findings. findings from a study. "The types of confidence identified include: Retest reliability, split confidence, inter-reviewer confidence, and internally consistent reliability. The function of reliability in a study is to ensure that the observed score closely matches the actual score obtained by minimizing measurement error"(Bollen, K. A. 1989).

Several other kinds of values, including "face value" "content value,""convergent value, "discriminant value" and "criterion value," have been studied. A research instrument must be dependable in order to be valid, yet reliability is independent of validity. The idea and validity of the research used to interpret the tests are crucial.

In psychometric testing, "reliability refers to the consistent property of measurements. There are different types of reliability. Pearson's time-product correlation coefficient can be used to assess the consistency of psychometric test results". Test-retest reliability describes this kind of trust. Using a large, diverse sample, we correlated the numerical scores of two identical forms delivered in a counterbalanced way to determine the reliability of alternative forms. Internally consistent approaches to reliability include halved confidence, where scores between tests are correlated, and alpha, which can be thought of as the average of all tests. All possible halving coefficients. For tests that require the judge's judgment to score, reliability between scores is needed. Calculating core bin reliability is simple. Two or more raters assess the same sample of tests separately and then compare their results using a co relational analysis. A test's validity depends on how well it assesses the constructs it sets out to examine. "A test is valid to the extent that the inferences drawn from it are relevant, eaningful, and useful. There are different types of values: content validity. Reliability and validity (external and internal) determine the degree to which the question, task, or test items represent the overall behavior of the test. vi for which the test is designed to be sampled". (Drost, E., A. 2011). A test has surface value if it appears valid to the user, the examiner, and especially the tester. "Criterion-related value is demonstrated when the test is effective in predicting performance against an appropriate outcome measure. A survey is intrinsically valid if a true cause-and-effect relationship exists between the independent and dependent variables. Confounding occurs when it is not possible to evaluate the effects of two independent variables in an experiment separately. Extrinsic validity relates to whether the study results can be generalized to a different situation, different themes, settings, time, etc. Aspects of the intrinsic value of an experiment include events outside the laboratory, maturation, experimental effects, regression effects, selectivity, and mortality. Threats to external validity include problems that arise from generalizing to other topics, times, or contexts" (Mohajan, H. 2017). Tester bias can be reduced by preventing the test from knowing the conditions of the test or its purpose, and by standardizing the process as much as possible.

References

Babour, R. S. (1998). Mixing qualitative methods: Quality assurance or qualitativequagmire? *Qualitative Health Research,* 8(3), 352-361

Bajpai, S. R., & Bajpai, R. C. (2014). Goodness of Measurement: Reliability and Validity. *International Journal of Medical Science and Public Health,* 3(2), 112-115.

Bogdan, R. C. & Biklen, S. K. (1998). *Qualitative research in education: An introduction to theory and methods* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Bollen, K. A. (1989). *Structural Equations with Latent Variables* (pp. 179-225). John Wiley & Sons.

Campbell, D.T. and Fiske, D.W. (1959). *Convergent and discriminant validation by the multitrait-multimethod matrix.* Psychological Bulletin, 56, 81-105.

Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78 (1), 98-104.

Creswell, J. W. & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-131.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika,16(3), 297-334.

Drost, E., A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38 (1), 105-124.

Fiske, Donald W. (1982). *Convergent -Discriminant Validation in Measurements and Research Strategies.* In Brinberg, D. and Kidder, L. H., (Eds), Forms of Validity in Research, pp. 77-93.

Heale, R., & Twycross, A. (2015). *Validity and Reliability in Quantitative Studies.* Evidence Based Nurs, 18(4), 66-67.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative MarketResearch*, 3(3), 118- 126.

Mohajan, H. (2017). Two criteria for good measurements in research: validity and reliability. *Annals of Spiru Haret University. Economic Series*, 17(4), 59-82.

Nunnally, J. C. (1978). Psychometric Theory. *McGraw-Hill Book Company*, pp. 86-113, 190- 255.

Rosenthal, R. and Rosnow, R. L. (1991). Essentials of Behavioral Research: Methodsand Data Analysis. Second Edition. *McGraw-Hill Publishing Company*, pp. 46-65.

24

Developing The Concept of Validity

Dr. Ratan Sarkar & Ms. Yogamaya Samanta

Abstract

Validity is one of the most important features in the evaluation of any measurement instrument. Measurement errors damage the function of scores and significant results. Validity is essentially ensuring that the data are sound and replicable and that the results are accurate. It acts as a pool for the purpose of the research and helps the researcher choose data to quantify that purpose. A measuring instrument is valid if it measures what it intends to measure. Validity ensures the integrity, quality, and truthfulness of a measure. Therefore, having good knowledge of validity is essential. The present paper aims to develop an understanding of the concept, nature, types, scope, needs, and limitations of validity.

Keywords: *Validity, Concepts, Types, Needs, Scope, Application and Criticism.*

Introduction

The word 'valid' is derived from the Latin words 'Validus', which means strong. Understanding validity means the understanding of measurement and evaluation. It gives the answer of the question," How good is this measure?" It is based on the strength of a collection of different types of evidence. It helps to determine what types of tests are use, what method will be used. It helps to increase transparency and decrease the bias or errors in research (Dogramaci, 2010). Validity helps researchers, practitioners, and decision-makers draw accurate and meaningful conclusions from their data, leading to better-informed choices and actions (Werdelin, 1980).

Validity refers to the extent to which a test or measurement accurately measures what it is intended to measure. It is a crucial concept in research

and assessment, as it ensures that the results obtained from a test or measurement are meaningful and reliable. In the context of research, validity refers to the degree to which a study accurately measures or represents the construct or concept it is investigating (Messick, 1995). It is concerned with whether the study methods and instruments are appropriate and capable of capturing the intended variables or concepts. It is important to establish and demonstrate validity to ensure that the results obtained from a study or assessment are meaningful, reliable, and generalizable. Validity provides confidence that the measurements or tests are accurately capturing the intended concepts and can be used to draw meaningful conclusions or make informed decisions based on the results.

Definitions

According to Cook and Campbell (1979) *"Validity is the best approximation to the truth or falsity of propositions"*.

According to Messick (1989) *"Validity always refers to the degree to which empirical evidences and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores"*.

According to Bond (2003) *"Validity is foremost on the mind of those developing measures and that genuine scientific measurement is foremost in the minds of those who seek valid outcomes from assessment"*.

Nature and Characteristics of Validity

When we talk about validity, we're talking about whether or not our measures are actually picking up the contrast that we think they are. The nature and characteristics of validity are discussed below-

- **Accuracy:** Validity is concerned with the accuracy of the measurement or assessment. It addresses the extent to which the data or results reflect the true attributes or characteristics of what is being measured.
- **Appropriateness:** Validity considers whether the measurement instrument, questions, or methods are appropriate for the specific research or assessment goals. It should align with the construct or concept under investigation.
- **Relevance:** A valid measurement should be relevant to the research objectives. It should provide information that is meaningful and valuable for the study's purpose.
- **Meaningfulness:** Validity ensures that the data or results have meaning and significance. It helps ensure that the measurements are not arbitrary but have interpretability and relevance.

- **Generalizability:** Validity is concerned with the extent to which the findings can be generalized to a broader population, context, or time frame. Valid measures should not be overly specific.
- **External Validity:** External validity is a subset of validity that focuses on the generalizability of research findings to real-world situations. It asks whether the findings can be applied to practical contexts beyond the research setting.
- **Internal Validity:** Internal validity is a subset of validity that assesses the extent to which the research design, methods, and measurements are free from confounding factors or sources of bias that could lead to incorrect conclusions.
- **Content Validity:** Content validity concerns whether the measurement instrument adequately covers the entire domain of the construct under investigation. It examines the representativeness of the items or questions.

Scope of Validity

- **Psychological Testing:** In psychological assessments, validity is critical to ensure that tests and questionnaires accurately measure traits, behaviors, or psychological constructs. Validity is used to assess whether the instruments truly capture the intended psychological attributes.
- **Educational Assessment:** In education, validity is essential to develop standardized tests and assessments that accurately measure students' knowledge, skills, and abilities. It helps determine whether the assessments are appropriate for their intended purpose.
- **Healthcare and Clinical Diagnosis:** Validity is crucial in healthcare and clinical settings to ensure that diagnostic tools and assessments accurately identify medical conditions or predict patient outcomes. Valid measures are critical for patient care.
- **Market Research:** In market research, validity is necessary to design surveys and questionnaires that accurately capture consumer attitudes, preferences, purchase intentions, and other relevant factors. Valid measures help businesses make informed decisions.
- **Social Sciences:** Validity is relevant in various social science research, such as studies on political attitudes, social behaviors, or public opinion. It ensures that the measurements accurately reflect the concepts under study.

- **Personnel Assessment:** In personnel assessments, including employee performance evaluations and job selection, validity is crucial to determine whether assessments accurately predict job performance or other relevant criteria.
- **Experimental Research:** In experimental research, internal validity is a primary concern, ensuring that the design and methods provide accurate information about cause-and-effect relationships.
- **Cross-Cultural Research:** Validity is essential in cross-cultural research to ensure that measurement instruments are culturally appropriate and can be used to make valid comparisons across different cultures.
- **Measurement Validation:** In psychometrics and measurement science, validity is a central concept used to validate measurement instruments, including questionnaires, surveys, and assessments.

The scope of validity extends to a wide range of fields and applications, as it is fundamental to ensuring that data and measurements are accurate, relevant, and meaningful for the intended purposes. Validity helps researchers, practitioners, and decision-makers draw accurate and meaningful conclusions from their data, leading to better-informed choices and actions.

Types of Validity

There are several types of validity, each addressing different facets of measurement accuracy and appropriateness. Some key types of validity are explained below:

a) Content Validity: Content validity is a type of validity used to evaluate the extent to which a measurement instrument, such as a test, questionnaire, or survey, adequately and comprehensively covers the domain or content it is intended to measure. In other words, content validity assesses whether the items, questions, or tasks within the instrument are representative of the full range of the construct or trait being studied (Benson, 1981; Cook and Campbell, 1979). Key points about content validity:

- **Content Domain:** It involves an examination of the content domain, which is the complete set of items or elements that make up the construct or trait. Content validity assesses whether the measurement instrument adequately samples from this content domain.
- **Subjective Judgment:** Content validity often relies on expert judgment to determine whether the items or questions are relevant, appropriate, and representative of the construct. Experts may include subject matter experts, scholars, practitioners, or individuals with relevant expertise in the field.

- **Face Validity:** While content validity is a more rigorous assessment, it is related to face validity. Face validity refers to whether the measurement instrument appears, on the surface, to be measuring what it claims to measure. Content validity delves deeper into the content and the expert judgments about that content.
- **Item Review:** In the process of evaluating content validity, experts review each item or question in the measurement instrument. They assess the clarity, relevance, and coverage of the items. Items are retained, modified, or eliminated based on this review.
- **Sampling of Content:** Content validity ensures that the measurement tool samples a representative and meaningful selection of content from the construct. If the content is too narrowly defined or limited, the validity may be compromised.
- **Content Representativeness:** Content validity aims to establish that the items or questions reflect the breadth and depth of the construct. It should capture the various aspects, components, or dimensions of the construct.
- **Qualitative and Quantitative Methods:** Content validity can be assessed using both qualitative and quantitative methods. Qualitatively, experts provide judgments and feedback. Quantitatively, statistical methods, such as content validity ratio (CVR) and content validity index (CVI), can be used to quantify content validity.
- **Revisions and Iterations:** The process of establishing content validity often involves iterations of item development and expert review. The measurement instrument is refined based on expert feedback until a satisfactory level of content validity is achieved.
- **Field Testing:** After content validity is established, the measurement instrument may be field-tested with a sample of the intended population to ensure that it performs as expected.

Content validity is particularly relevant in the development of assessments and questionnaires for research, education, and other fields where accurate and representative measurement is crucial. Ensuring content validity is an essential step in building confidence that a measurement tool accurately captures the concept or construct it is intended to measure.

b) Criterion-Related Validity: Criterion-related validity is a type of validity used to assess the extent to which a measurement instrument, such as a test, questionnaire, or assessment tool, can predict or correlate with an external criterion or outcome that is theoretically related to the construct being

measured. It examines the ability of the measurement instrument to accurately predict or relate to a particular criterion, which could be another test, a behavioural outcome, a real-world performance, or some other measurable variable (Messick, 1995). There are two subtypes of criterion-related validity:

- **Concurrent Validity:** Concurrent validity assesses the degree to which the measurement instrument correlates with a criterion that is assessed simultaneously or at the same time. In other words, it examines the ability of the instrument to predict an external criterion that is measured concurrently.

  Example: If a new intelligence test is developed, concurrent validity might be assessed by administering the new test and an established intelligence test to the same group of participants. The correlation between the scores of the new test and the established test measures concurrent validity.

- **Predictive Validity:** Predictive validity assesses the extent to which the measurement instrument can predict a future criterion or outcome. It examines whether the instrument can accurately forecast a criterion that occurs at some point in the future.

  Example: In personnel selection, predictive validity might be evaluated by administering a pre-employment test to job applicants and then tracking their job performance over a specific period. The degree to which the test scores predict future job performance measures predictive validity.

Key points about criterion-related validity:

- **External Criterion:** Criterion-related validity relies on the existence of an external criterion, which is an independent measure or outcome that is theoretically related to the construct being measured. The criterion is used to validate the measurement instrument.
- **Correlation Analysis:** Criterion-related validity is typically assessed using correlation analysis, where the correlation coefficient (e.g., Pearson's correlation coefficient) is calculated to determine the strength and direction of the relationship between the measurement instrument and the criterion.
- **Thresholds:** Researchers may establish specific thresholds or cutoff points for the measurement instrument to determine whether it accurately predicts or correlates with the criterion. These thresholds can be used for decision-making in various contexts, such as hiring or diagnosis.

- **Real-World Applications:** Criterion-related validity is particularly relevant in fields such as education, psychology, healthcare, and personnel assessment. It helps determine whether an assessment tool is valid for making predictions or decisions in real-world situations.
- **Longitudinal Studies:** Predictive validity often requires longitudinal studies or tracking participants over time to observe whether the instrument's scores are predictive of future outcomes or behaviors.

Criterion-related validity is an essential aspect of validity, especially when researchers aim to assess the practical utility of a measurement instrument in making predictions or decisions. It helps ensure that the instrument has real-world applications and accurately relates to relevant external criteria.

c) Construct Validity: Construct validity is a type of validity used to assess the extent to which a measurement instrument, such as a test, questionnaire, or assessment tool, accurately measures the underlying theoretical construct or concept it is intended to measure. This type of validity is concerned with the theoretical meaning of the construct and the extent to which the measurement aligns with that theory. It goes beyond face validity (how well the instrument appears to measure the construct on the surface) and content validity (whether the instrument covers the entire content domain) by examining the underlying construct's structure, relationships, and theoretical implications (Bond, 2004).

Key points about construct validity:

- **Theoretical Framework:** Construct validity is based on a theoretical framework or model that defines and describes the construct of interest. The measurement instrument should align with this theoretical framework.
- **Hypothesis Testing:** Construct validity often involves testing hypotheses about the relationships between the measurement instrument and other variables or constructs that should theoretically be related. These hypotheses are based on the theoretical understanding of the construct.
- **Convergent and Discriminant Validity:** Researchers assess construct validity through measures of convergent and discriminant validity. Convergent validity evaluates whether the measurement instrument correlates positively with other measures that should theoretically be related to the same construct. Discriminant validity assesses the extent to which the measurement tool correlates negatively or not at all with measures that should theoretically be unrelated to the construct.

- **Factor Analysis:** Factor analysis is a common statistical technique used to examine the underlying structure of constructs. It helps identify whether the items or questions in the measurement instrument load on the same factor, supporting the idea that they measure the same construct.
- **Theoretical Development:** Theoretical development of the construct involves defining its key dimensions, attributes, and relationships. Researchers use this theoretical understanding to create a measurement instrument that accurately reflects the construct.
- **Empirical Testing:** Construct validity is empirically tested through statistical analyses, such as factor analysis, regression analysis, or structural equation modelling. These analyses help confirm whether the measurement instrument aligns with the theoretical construct.
- **Criterion-Related Validity:** In some cases, construct validity can be evaluated by examining the measurement instrument's relationships with external criteria. This includes predictive and concurrent validity, where the measurement is expected to predict or correlate with specific criteria based on the theoretical construct.
- **Incremental Validity:** Researchers may assess whether the measurement instrument provides additional information or predictive power beyond existing tools or measures. Incremental validity evaluates whether the new instrument contributes valuable information to an assessment.

Construct validity is crucial in ensuring that a measurement instrument accurately reflects the underlying construct and that it is not influenced by irrelevant factors or constructs. It is particularly relevant in psychology, social sciences, and other fields where researchers aim to measure abstract or theoretical constructs (Litwin and Fink, 1995). Construct validity provides evidence that the measurement instrument is capturing the intended psychological, social, or educational attribute.

d) Convergent and Discriminant Validity: Convergent and discriminant validity are two important components of construct validity used to assess the accuracy of a measurement instrument in relation to other constructs. They help determine whether the instrument behaves as expected in relation to both related and unrelated constructs. Here's an explanation of each:

- **Convergent Validity:** Convergent validity assesses the degree to which a measurement instrument correlates positively with other measures or variables that theoretically should be related to the same

construct. In other words, it examines whether different measures of the same construct produce similar results. Key points about convergent validity:

- It involves comparing the measurement instrument with other measures or variables that are conceptually or theoretically related to the same construct.
- High convergent validity is demonstrated when the measurement instrument shows strong, positive correlations with these related measures. This indicates that the instrument is measuring the same underlying construct as the related measures.
- Convergent validity provides evidence that the measurement instrument accurately reflects the specific construct it is intended to measure.

For example, in a study of self-esteem, a self-report questionnaire measuring self-esteem should demonstrate convergent validity by showing strong positive correlations with other measures of self-esteem.

❖ **Discriminant Validity:** Discriminant validity assesses the extent to which a measurement instrument correlates negatively or not at all with measures or variables that should theoretically be unrelated to the construct of interest. It helps ensure that the instrument does not produce spurious or inappropriate relationships with unrelated constructs. Key points about discriminant validity:

- It involves comparing the measurement instrument with measures or variables that are conceptually or theoretically unrelated to the construct under study.
- High discriminant validity is demonstrated when the measurement instrument shows weak or no correlations with these unrelated measures. This indicates that the instrument is not measuring unrelated constructs.
- Discriminant validity is important for ensuring that the measurement instrument is specific to the construct of interest and is not influenced by irrelevant factors.
- For example, in a study of job satisfaction, a self-report questionnaire measuring job satisfaction should demonstrate discriminant validity by showing weak or no correlations with unrelated variables, such as dietary habits or weather conditions.

Both convergent and discriminant validity are critical for establishing the construct validity of a measurement instrument. They provide evidence that the instrument accurately reflects the intended construct and does not produce

misleading or erroneous relationships with related or unrelated variables. Researchers often use correlation coefficients, factor analysis, and statistical tests to assess convergent and discriminant validity.

e) Face Validity: Face validity is a type of validity that assesses whether a measurement instrument, such as a test, questionnaire, or assessment tool, appears, on the surface, to measure what it is intended to measure. In other words, it examines whether the instrument, based on a superficial examination, seems to be a valid tool for measuring the concept or construct it is designed to assess. Key points about face validity:

- **Superficial Appearance:** Face validity is concerned with the superficial or initial impression that the measurement instrument gives to individuals who use or view it. It assesses whether the instrument "looks like" it is measuring what it claims to measure.
- **Subjective Judgment:** Face validity is subjective and relies on the judgment of individuals who are not necessarily experts in the field. It is often based on the opinions of laypersons or those who are not trained in the construct being measured.
- **No Formal Analysis:** Unlike other forms of validity, such as content validity or construct validity, face validity does not involve a systematic or formal analysis of the instrument. Instead, it is more of an intuitive or common-sense assessment.
- **Limited as a Sole Measure:** Face validity is limited as the sole measure of validity because it is based on appearances and subjective opinions. It does not provide strong empirical evidence that the instrument truly measures the intended construct.
- **Use in Questionnaires:** In some cases, researchers intentionally design measurement instruments to have face validity, especially in surveys and questionnaires. This approach can make participants more willing to participate in the study if they believe the questions are relevant and appropriate.
- **Feedback from Participants:** Researchers may seek feedback from participants or stakeholders to assess the face validity of an instrument. This feedback can help identify any items or questions that may be confusing, irrelevant, or inappropriate from the participants' perspective.
- **Enhancing Acceptance:** While face validity alone does not provide strong evidence of validity, it can enhance the acceptance and engagement of participants with the measurement instrument. If

participants perceive that the instrument is relevant and meaningful, they are more likely to respond accurately.

It's important to note that face validity is not a strong indicator of the actual validity of a measurement instrument. It is primarily used to assess how the instrument is perceived by individuals who encounter it. While face validity can be useful in some situations, researchers typically rely on more rigorous forms of validity assessment, such as content validity, construct validity, and criterion-related validity, to provide empirical evidence of the instrument's validity.

f) Ecological Validity: Ecological validity is a concept used in research and psychology to assess the extent to which the findings and results obtained in a controlled or experimental setting can be generalized to real-world, everyday situations or "ecological" environments. It examines whether the behaviors, actions, or phenomena observed in a research study accurately reflect what individuals would do in their natural, uncontrolled settings. Key points about ecological validity:

- **Controlled vs. Ecological Settings:** Ecological validity considers the potential differences between controlled or laboratory settings and the complex, dynamic, and often unpredictable settings of the real world. In a controlled setting, researchers have more control over variables, but the trade-off may be a lack of ecological validity.
- **Generalization to Real Life:** Researchers are concerned with the degree to which the findings of a study can be generalized to real-life situations. High ecological validity means that the results are likely to apply to everyday life, while low ecological validity suggests the findings may not transfer well to real-world contexts.
- **External Validity:** Ecological validity is often related to external validity, which is a broader concept that encompasses the generalizability of research findings beyond the specific study or experimental conditions.
- **Natural Behavior and Context:** High ecological validity implies that the behavior of participants in a study is more natural and representative of how they would behave in their everyday lives. It also considers the context in which these behaviors occur.
- **Situational Factors:** Ecological validity takes into account situational factors that may influence behavior and decision-making in real-life situations. These factors can include social norms, cultural influences, environmental conditions, and the presence of distractions.
- **Field Studies and Observations:** Researchers may use field studies, naturalistic observations, or experiments conducted in real-world

settings to enhance ecological validity. These methods aim to capture behavior as it naturally occurs.

- **Trade-Off with Control:** There is often a trade-off between control and ecological validity. Studies conducted in more controlled settings may lack ecological validity, while studies conducted in more ecologically valid settings may have less experimental control.
- **Balance:** Researchers must strike a balance between experimental control and ecological validity based on their research goals. Some studies may prioritize internal validity (the accuracy of cause-and-effect relationships) over ecological validity, while others may focus on generalizability to real-life situations.
- **Practical Applications:** Ecological validity is particularly important in fields such as clinical psychology, education, and usability testing, where the practical application of research findings in real-world settings is essential.
- **Replication and External Validation:** Researchers may use replication studies or external validation in different settings to assess the ecological validity of their findings and ensure they hold up across various conditions.

Assessing ecological validity is crucial for understanding how research findings apply to the real world. Researchers must consider the balance between internal and external validity to ensure that their findings are both scientifically sound and practically relevant.

g) Incremental Validity: Incremental validity is a concept used in research and measurement to assess the degree to which a new measurement instrument, variable, or assessment tool provides additional and unique information beyond what is already provided by existing measures or variables. In other words, it evaluates whether the new instrument or variable contributes value in predicting or explaining an outcome above and beyond what is accounted for by other established measures. Key points about incremental validity:

- **Additional Predictive Power:** Incremental validity is concerned with whether a new measurement or variable can improve the prediction or explanation of an outcome compared to existing measures alone. It examines whether the new addition offers additional and unique information.
- **Regression Analysis:** Incremental validity is often assessed using regression analysis. In this approach, researchers examine how well a combination of existing measures and the new measure predict or

explain an outcome, and they compare this to the prediction or explanation based solely on the existing measures.

- **Decision-Making:** Incremental validity is relevant in situations where decisions or predictions are made based on multiple variables. It helps determine whether including the new measure enhances the quality and accuracy of those decisions.
- **Overcoming Limitations:** Sometimes, existing measures may have limitations or may not fully capture the complexity of a construct. Incremental validity allows researchers to test whether the addition of a new measure can address these limitations.
- **Avoiding Redundancy:** Incremental validity ensures that the new measure does not merely duplicate the information provided by existing measures, which would lead to redundancy. Instead, it should provide unique insights or information.
- **Practical Applications:** Incremental validity is important in various fields, including personnel assessment (e.g., job performance prediction), clinical psychology (e.g., diagnosis and treatment prediction), and educational assessment (e.g., predicting student success). In these contexts, it is crucial to identify measures that add value to decision-making.
- **Holistic Assessment:** Incremental validity supports a more holistic approach to assessment by considering the collective contribution of multiple variables in making decisions or predictions.
- **Research Goals:** Researchers should consider their research goals and the specific context in which incremental validity is being assessed. In some cases, it may be more important to prioritize incremental validity, while in others, existing measures may suffice.
- **Robustness Testing:** To assess incremental validity, researchers may use techniques like hierarchical regression analysis, stepwise regression, or cross-validation to ensure the findings are robust and generalize to other samples or conditions.

Assessing incremental validity is essential for determining whether a new measurement or variable provides valuable information for decision-making or prediction. It helps avoid the inclusion of unnecessary or redundant measures and ensures that assessments are efficient and effective in achieving their intended goals.

Needs and Importance of Validity

The concept of validity is fundamental in research and measurement, and

its importance lies in ensuring that data and measurements are accurate, meaningful, and appropriate for the specific context and purpose. The key reasons of why validity is crucial are mentioned below:

- **Accurate Inferences:** Validity is essential for making accurate inferences based on data. It ensures that the measurements or assessments reflect the true attributes or characteristics of the concepts, traits, or constructs being studied. Without validity, researchers may draw incorrect conclusions.
- **Research Credibility:** Validity enhances the credibility and trustworthiness of research. When the measurements are valid, the findings are more likely to be respected and accepted by peers, stakeholders, and the broader scientific community.
- **Measurement Quality:** Validity is a critical component of measurement quality. It helps ensure that the data collected is of high quality, making it suitable for drawing meaningful conclusions and making informed decisions.
- **Appropriate Decision-Making:** In various fields such as education, healthcare, and personnel assessment, validity is essential for making appropriate decisions. For example, in personnel selection, valid assessments help identify candidates who are more likely to perform well on the job.
- **Fairness:** Validity contributes to fairness in assessments and evaluations. It ensures that individuals are evaluated using measures that accurately reflect their abilities, knowledge, or traits, reducing the potential for bias.
- **Generalizability:** Validity allows for the generalization of research findings beyond the specific study sample and context. When measurements are valid, it is more likely that the results can be applied to broader populations, settings, or time frames.
- **Avoiding Misclassification:** Validity helps prevent the misclassification of individuals or subjects based on incorrect or inadequate measurements. For instance, in medical diagnosis, valid tests are crucial for accurate patient classification.
- **Scientific Progress:** Validity contributes to the advancement of scientific knowledge by ensuring that research findings are based on sound and accurate measurements. It promotes the development of more robust theories and models.

- **Legal and Ethical Implications:** In some cases, the validity of assessments has legal and ethical implications. For example, in legal proceedings, the use of valid psychological assessments is essential for making decisions about an individual's mental state.
- **Educational Accountability:** In education, validity is important for educational accountability. Valid assessments are necessary for making fair and meaningful judgments about students' knowledge and skills.
- **Resource Allocation:** In business and organizational settings, validity in assessments can influence resource allocation, such as training and development investments or staffing decisions.
- **Quality Control:** In manufacturing and engineering, validity is essential for quality control. Valid measurements ensure that products or systems meet defined standards and specifications.

In summary, the need for validity is pervasive across various fields and applications, and it is critical for ensuring that research findings and decisions are based on accurate, meaningful, and relevant data. Without validity, the integrity and credibility of research and assessment are compromised, potentially leading to incorrect conclusions and actions. Validity, therefore, is a cornerstone of scientific and practical endeavours.

Criticisms of Validity

Criticism of validity in research and measurement often stems from concerns about the accuracy, appropriateness, or generalizability of the measurements or assessments. While validity is essential for ensuring that data accurately reflect the concepts or constructs of interest, there are several criticisms and challenges related to validity:

- **Subjectivity and Expert Judgment:** Some validity assessments, especially content validity, rely on expert judgment to evaluate the relevance and representativeness of items or questions. Critics argue that these judgments can be subjective and vary among experts.
- **Changing Nature of Constructs:** Constructs or concepts can change over time, making it challenging to establish the validity of measurements that were designed to assess those constructs in the past. This is particularly relevant in fields where societal norms and understanding of concepts evolve.
- **Difficulty in Establishing Absolute Validity:** Establishing absolute validity is challenging. Researchers often aim to accumulate evidence to support validity rather than definitively prove it. Critics may argue that this leaves room for uncertainty.

- **Lack of a Gold Standard:** In some cases, there may be no gold standard or perfect criterion to use for criterion-related validity assessments. This can make it difficult to determine whether a measurement tool is accurately predicting a criterion.
- **Cross-Cultural Validity:** Validity may vary across different cultural or linguistic groups. Critics point out that a measurement instrument that is valid in one culture or language may not be valid in another, raising concerns about the generalizability of findings.
- **Overemphasis on Predictive Validity:** In some cases, predictive validity is overemphasized, especially in personnel assessments. Critics argue that emphasizing the ability to predict future outcomes may neglect the construct validity aspect and the theoretical underpinnings of the measurement.
- **Differential Item Functioning (DIF):** DIF occurs when the measurement tool functions differently for different subgroups. Critics may highlight concerns about DIF, as it can lead to the invalid comparison of groups or individuals.
- **Ecological Validity Concerns:** Some critics argue that research studies conducted under controlled laboratory conditions may lack ecological validity, meaning the findings may not accurately reflect real-world behavior or experiences.
- **Use of Artificial Scenarios:** In psychological testing, some assessments use artificial scenarios or hypothetical constructs. Critics may question the ecological and practical validity of these measurements.
- **Cultural Bias in Validity Evidence:** Validity evidence may be biased toward specific cultural or demographic groups. Critics argue that the underrepresentation of diverse populations can affect the validity of assessments for those groups.
- **Criterion-Related Validity Limitations:** In predictive and concurrent validity assessments, there may be limitations in the selection of criteria or the timing of data collection. Critics may argue that these factors can affect the assessment's validity.
- **Validity as a Continuous Process:** Critics may emphasize that validity is not a one-time determination but an ongoing and evolving process. Ensuring validity requires continuous examination and adaptation.

It's important to acknowledge these criticisms as they highlight the need for careful consideration and transparency when assessing and reporting validity

in research and measurement. Researchers must address these concerns to enhance the quality and relevance of their validity assessments. Validity should be evaluated in a rigorous and comprehensive manner, considering different aspects of measurement quality.

Conclusion

Validity should be evaluated in a rigorous and comprehensive manner, considering different aspects of measurement quality. The maximum level of validity is equal to the square root of reliability. Assessing and maintaining the validity of various educational, social, and psychological tests is a complex but urgent fundamental need. Validity extends to a wide range of fields and applications, as it is fundamental to ensuring that data and measurements are accurate, relevant, and meaningful for the intended purposes.

References

Bond, Trevor. (2004). *Validity and assessment: a Rasch measurement perspective. Metodologia de las Ciencias del Comportamiento, 5* (2), 179-194.

Benson, J. (1981, October). A Redefinition of Content Validity. *Educational and Psychological Measurement, 41*(3), 793-802. https://doi.org/10.11770013164481 04100320

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings.* Chicago, IL: Rand McNally.

Dogramaci, S. (2010, August 23). *Knowledge of Validity*. *Noûs*, *44*(3), 403–432. https://doi.org/10.1111/j.1468-0068.2010.00746.x

Litwin, M. S., & Fink, A. (1995, January 1). *How to Measure Survey Reliability and Validity*. SAGE.Lissitz, R. W. (2009, November 1). *The Concept of Validity*. IAP.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into scoremeaning. *American Psychologist*, 50, 9, 741 - 749.

Werdelin, I. (1980, January 1). *The Concept of Validity*.

Wainer, H., & Braun, H. I. (2013, July 4). *Test Validity*. UK: Routledge.

25

Types of Reliability and Validity

Bikshyatsib Sardar

Abstract

This book chapter delves into the pivotal dimensions of reliability and validity within the realm of assessment and evaluation. Encompassing foundational concepts, defining characteristics, different types, real-world limitations, and the intricate influences that shape these constructs, the chapter lays a comprehensive framework for researchers and practitioners alike. Beginning with reliability, the chapter examines the defining traits that underpin reliable measurements and subsequently explores several types of reliability like; Test-Retest Reliability, Parallel Forms Reliability, Split-half Method, Inter-Rater Reliability and Richardson Method. Further this chapter critically examines the inherent limitations of each type of reliability. It subsequently investigates the factors influencing reliability coefficients, encompassing variables such as test length, participant attributes, and contextual variations, etc. Shifting to validity, the chapter dissects the fundamental characteristics of valid measurements and then embarks on an intricate exploration of six distinct types of validity like; Content Validity, Construct Validity, Criterion Validity, Concurrent Validity, Predictive Validity and Face Validity. Each validity type is illustrated with examples, coupled with a meticulous analysis of their inherent limitations. The chapter culminates by unravelling the intricate influences that impact the validity of measurements within the dynamic landscape of assessment and evaluation. In essence, this book chapter crafts a comprehensive tapestry of reliability and validity within the realm of assessment and evaluation. This profound comprehension serves as bedrock for upholding assessment integrity, cultivating credible and impactful outcomes within the realm of education and evaluation.

Keywords: *Reliability, Validity, Assessment, Evaluation and Measurement.*

1.0 Introduction

In the realm of education and evaluation, the quest for accurate and meaningful information is paramount. Whether in the classroom or within the broader educational system, the ability to measure learning outcomes, assess instructional effectiveness, and make informed decisions hinges upon two critical principles: reliability and validity. Together, reliability and validity form the bedrock of credible educational research, data-driven decision-making and fair evaluations. By employing assessments and measurements that are both reliable and valid, educators, administrators, policymakers, and researchers can foster an educational landscape grounded in accuracy, transparency, and integrity. Throughout this chapter, we will delve into the intricacies of reliability and validity in the context of education and evaluation. We will explore the different forms of reliability, the methods to enhance it, and the importance of consistent measurement practices. Likewise, we will examine the diverse dimensions of validity, the methods to establish it, and its role in promoting fair and effective assessments. Emphasizing the significance of these principles, we aim to equip educators and evaluators with the knowledge and tools to elevate the quality of educational practices and ensure that our efforts towards learning and improvement are built upon a firm and trustworthy foundation.

2.0 Reliability

It refers to the stability and consistency of measurement and assessment. It ensures that when a particular tool or instrument is used multiple times or by different evaluators, it yields consistent results. Imagine a teacher grading a student's essay and arriving at the same score if the essay were to be re-graded later or evaluated by another instructor. Reliability instills confidence in the data collected, offering a stable foundation upon which educators and evaluators can base their judgments. It ensures that the results of an assessment or evaluation remain consistent over time and across different situations. In education, reliable measurements mean that when we test the same students or people more than once, we get similar results each time.

3.0 Types of Reliability

Types of Reliability refer to various methods used to assess the consistency and stability of measurements obtained from a tools, such as questionnaires, tests, or scales. Ensuring high reliability is crucial in any type of measurement, as it enhances the trustworthiness and repeatability of measurement results. Various methods of establishing reliability of a test are described below.

3.1 Test-Retest Reliability

This reliability method examines how consistent a measurement is over

time. It is one of the simplest ways to check a test's reliability. To establish it, the same test is given to the same group of people on two separate occasions, with a gap in between. The scores from both occasions are then compared to see how closely they match. It involves administering the same test twice, with a time gap of at least 15 days and no more than six months. After collecting scores from both instances, correlation methods like Spearman's Rank Difference or Pearson's Product Moment Correlation are used to calculate a correlation coefficient. A high test-retest reliability coefficient indicates that the tool consistently provides reliable results over time.

3.1.1 Limitations of this method

- Administering the same test two times to a group may lead to a carry-over effect, where participants remember previous items.
- Maintaining an optimal time gap between tests is crucial. Too short a gap intensifies carry-over effects, while an excessively long gap may be influenced by participants' maturation.
- This method is susceptible to errors, including memory, carry-over, practice, and participant maturation effects.

3.2 Parallel Forms Reliability

Parallel forms reliability, also referred to as equivalent forms reliability, evaluates the consistency between two distinct versions of a single test, both designed to assess the same concept or construct. This method serves as an alternative to test-retest reliability, addressing some of the potential sources of error inherent in the test-retest method. In this method, two sets of parallel tests are meticulously developed to ensure equivalence in various aspects such as objectives, item types, content, quantities, and level of difficulty, time constraints, discrimination values, and conditions of administration. Both versions are then administered to a group of participants, and the obtain scores are subjected to correlation analysis, employing methods such as Spearman's or Karl Pearson's, to calculate the correlation coefficient between the two versions. A strong parallel forms reliability coefficient signifies that both versions produce similar outcomes, making them suitable for interchangeable use. The primary goal of this approach is to create two test forms that are essentially equivalent in their measurement of the intended construct.

3.2.1 Limitations of this Method:

- Developing multiple alternative test forms for a given variable can prove to be a highly challenging task.

- ❖ Generating a substantial number of items that effectively capture the same underlying construct can be problematic.
- ❖ Ensuring that two alternate forms of a test are truly parallel measures may also present significant difficulties, and in some cases, this goal may be unattainable.

3.3 Internal Consistency Reliability

Internal consistency reliability assesses the homogeneity of a test, examining the consistency among items within a single test designed to measure the same construct. When all items in the test effectively measure the same function, it is considered a homogeneous test, leading to a high level of internal consistency reliability. Two commonly used methods for estimating internal consistency reliability are the Split-half method and the Rational equivalence method.

3.3.1 Split-half method

The split-half method effectively reduces many of the errors associated with the test-retest approach since it eliminates the need for conducting the test twice with the same group. In this method, the test is administered only once to a single group of participants. After collecting their scores, the test is divided into two equal parts, often by distinguishing between even items and odd items, or by employing other techniques for splitting it into two parts. The scores for the odd items and the even items are then treated separately, and appropriate statistical methods are applied to assess the reliability of each half's scores. Often, the Spearman-Brown formula is employed to calculate the overall reliability of the full test. This method offers a practical means of assessing reliability without the need for multiple test administrations.

3.3.1.1 Limitations of this method:

- ❖ A critical consideration is the method of dividing the whole test into two equal parts, which can take various forms such as odd-even or segmenting by item number, first fifty items and next percentage, etc.
- ❖ Correlation between test parts and overall reliability varies with item-split techniques.

3.3.2Rational equivalence method

The rational equivalence method is a statistical approach used to determine the reliability of a test. Kuder and Richardson introduced two renowned formulas, namely Kuder and Richardson (KR) Formula 20 and KR Formula 21. Lee J. Cronbach referred to this method as the Coefficient of Internal

Consistency. Notably, it represents an enhancement over earlier reliability assessment methods, notably reducing errors in reliability calculations. The rational equivalence method extends from the principles of the split-half method. In Richardson and Kuder formulas, there is an underlying assumption that all test items share equal difficulty, although not a requirement for the same individuals to answer each item correctly. The level of item difficulty depends on an individual's familiarity with the item, categorizing it as easy if known and challenging if unfamiliar. This approach relies on the assumption of test homogeneity.

3.3.2.1 Limitations of this method

- ❖ Reliability index calculation overlooks item difficulty variation.
- ❖ The reliability coefficient is an approximation, typically lower than the precise reliability coefficient.
- ❖ Inapplicable for power tests or heterogeneous tests.
- ❖ Various KR formulas yield varying reliability coefficients.

3.4 Inter-Rater Reliability

Inter-rater reliability, sometimes known as inter-observer reliability, quantifies the degree of consistency and concordance among various raters or observers who employ the same assessment instrument to assess a specific phenomenon. It assesses the degree to which multiple individuals, who may have subjectivity in their judgments, provide consistent and similar assessments or ratings. Inter-rater reliability is crucial in research, clinical assessments, and various fields where human judgment is involved to ensure that the results or evaluations are consistent and dependable across different raters or observers. Statistical techniques, such as Cohen's kappa or intra-class correlation coefficients, are commonly used to quantify inter-rater reliability.

3.4.1 Limitations of this Method

i. Inter-rater reliability can be compromised by the personal biases and subjectivity of individual raters, leading to inconsistent assessments.

ii. In the absence of clear and standardized criteria, raters may interpret and apply assessment guidelines differently, resulting in discrepancies in their ratings.

iii. Differences in the level of training and experience among raters can impact the reliability of their assessments, with less experienced raters more likely to introduce inconsistency.

4.0 Validity

Validity addresses the fundamental inquiry of whether an assessment genuinely measures what it asserts to measure. It delves into the precision and appropriateness of an evaluation, ensuring that it aligns with the intended learning objectives. For instance, when a mathematics test is crafted to evaluate a student's algebraic proficiency, the test's validity relies on its capability to precisely assess that specific knowledge, skills, or attributes, without interference from unrelated factors. In the realm of education, validity pertains to the degree to which an assessment accurately and truthfully measures its intended target. In essence, a valid assessment effectively captures the skills, knowledge, attitudes or attributes it is expressly designed to evaluate.

5. Types of Validity

Different type of validity in education and evaluation serves as a distinct approach or criterion for assessing how accurately and appropriately an assessment instrument measures the intended construct, concept, or phenomenon. These types help ensure that assessments provide reliable and meaningful information for making informed decisions in educational and evaluative contexts. Various types of validity are explained below.

5.1 Face Validity

Face validity, also known as validation by face, serves as the initial step in assessing the validity of a test. While not extensively utilized due to its limited scope, face validity offers a preliminary understanding of a test's validity. In this approach, the test's appearance, its intended purpose, the objectives it aims to measure, the dimensions it covers, and the language used are considered to determine its face validity. It represents the most basic level of evaluating a test's validity and is typically employed when time constraints hinder the use of more accurate validity determining methods. Before moving to other validity determining methods, a judgment is made regarding whether the test possesses face validity. If it lacks face validity, it is uncommon to proceed with other validity determination methods.

5.2 Content Validity

This method is also known as rational or logical validity. It involves a comprehensive assessment of both the format and content of the test to make a determination regarding its validity. As described by Anastasi (1968), content validity entails a systematic scrutiny of the test content with the aim of ascertaining whether it adequately covers a representative sample of the behavioural domain it intends to measure. Content validity seeks to align the test items with the instructional objectives, making it particularly pertinent for

achievement tests. To assess the content validity of a test, it is common practice to evaluate the items in accordance with the blueprint or table of specifications upon which the test was originally developed.

5.3 Construct Validity

This method is also known as trait or psychological validity, involves an assessment of test scores in terms of a specific construct or theoretical concept. For instance, in the context of a student's achievement, this construct could encompass elements like attitude, practice, interest, motivation, aptitude, intelligence, and more. Construct validity refers to the degree to which a test can be regarded as a valid measure of a psychological variable, trait or theoretical construct. To establish construct validity, an in-depth analysis is conducted to explore the relationships between the variables linked to the test that collectively contribute to the intended construct. This process helps determine whether the test effectively measures the underlying concept it aims to assess. For example, it is theoretically expected that intelligence and achievement are positively correlated. To determine the construct validity of an intelligence test, an analysis is performed by correlating the test scores with the students' scores on an achievement test. This measurement is grounded on the assumption that students who perform strongly on the intelligence test should likewise excel on the achievement test, aligning with the positive correlation suggested by theory. If the correlation turns out to be negative, it suggests that the intelligence test lacks construct validity.

5.4 Criterion Validity

It is also known as instrumental validity. This method is use to validate a measurement by comparing it with another established process or method that has already been confirmed as valid. This type of validity always use an "external" criterion, which can either, be concurrent or predictive in nature. When we use an external criterion that relates to concurrent information, we refer to it as concurrent validity. Likewise, when the criterion is linked to future performance, it is termed predictive validity. Imagine a new standardized test designed to assess students' mathematical problem-solving skills. To validate this test, a criterion-related approach is employed. We compare the scores obtained from this new test to the scores of a widely recognized and established mathematics assessment, such as a national math proficiency exam, which serves as an external criterion measuring the same mathematical skills. However Criterion validity essentially evaluates how well a test aligns with an external criterion that reflects the same underlying construct. Below

5.4.1 Concurrent Validity

Concurrent validity assesses the degree of correlation between a newly developed test or scale and a pre-established, validated measure. This method is particularly valuable in establishing the validity of a newly created tool. However, it is essential to note that concurrent validity can only be applied when there is an already standardized and validated test available for comparison. Concurrent validity specifically examines how effectively a test aligns with measurements or scores from another test that is known to be valid. For example, if a newly developed intelligence test exhibits high concurrent validity, it should display a robust positive correlation when administered alongside an existing intelligence test that is recognized for its validity.

5.4.2 Predictive Validity

Predictive validity is a form of validity that examines how well a test or measurement can forecast future outcomes or behaviours associated with the construct. In other words, it evaluates whether the scores obtained from a test can accurately forecast future events or criteria. This form of validity is occasionally known as "statistical validity" or "empirical validity" due to its primary reliance on empirical and statistical methodologies, enabling the empirical testing of validity. It allows for empirical testing of the tool's ability to predict future events. For instance, a university uses an entrance exam for B.Ed. applicants to predict their future academic success. They track students' performance throughout the program, examining if higher entrance exam scores correlate with better grades, on-time graduation, and strong student teaching performance. Strong positive correlations indicate effective predictive validity for the entrance exam. Predictive validity gains particular relevance when the external criterion is measured at a subsequent point in time.

5.5 Factorial validity

The term "factorial validity" describes the connection between diverse factors within a test and the test as a whole. Factorial validity examines the correlation between distinct factors or components within a test and the test as a whole. Achieved through factor analysis, a statistical technique, this process involves scrutinizing inter-item correlations to unearth latent factors representing the constructs the test intends to measure. Factor extraction identifies these factors, and factor loadings quantify the strength of associations between individual items and these factors. Factorial validity evaluation entails determining how well these identified factors align with the intended theoretical framework or construct of the test. Calculating correlations between the test and each factor elucidates the contribution of each factor to the overall test performance. This method provides crucial insights into the test's underlying structure,

helping refine and validate its content, making it pivotal in test development and assessment. It offers the most transparent understanding of what a test measures and should unquestionably be prioritized over all other types of validity.

6.0 Conclusion

Reliability and Validity are fundamental principles in educational measurement. The pursuit of reliable and valid measurement in education is essential for informed decision-making, fair evaluations, and credible research. Understanding the characteristics and types of reliability and validity equips educators, researchers, and policymakers with the tools to create assessments that accurately reflect the knowledge and abilities of individuals. These principles serve as a cornerstone for building a robust and equitable educational system based on precision, integrity, and transparency. By employing reliable and valid assessment tools and evaluation methods, educators and policymakers can make better-informed decisions, promote fairness, and enhance the overall quality of education and evaluation processes.

References

Anastasi, A. (1976). Psychological Testing (4th ed.). *Macmillan Publishing* Co. Inc.

Gronlund, N. E. (1981). *Measurement and Evaluation in Teaching* (4th ed.). Macmillan Publishing Co. Inc.

IGNOU. (2010). Criteria of a Good Tool (Unit 6; Block 2*). In Educational Evaluation* (ES-333, B.Ed.). New Delhi: IGNOU.

Singh, A. K. (2002). *Tests, Measurements and Research Methods in Behavioral* Sciences (3rd ed.). Bharti Bhawan.

Sansanwal, D. N. (2020). Research Methodology and Applied Statistics. *Shipra Publications*

Index

A

Abilities of individuals, 239
Academic achievement, 23, 63, 175
Academic competition, 168
Academic inquiry, 211
Academic journey, 173
Academic knowledge, 178
Academic progress, 63
Academic pursuits, 136
Actual assessment, 132
Adaptive learning, 174
Additional coaching, 169
Administrative utility, 44
Aligning assessment, 108
American Psychological Association, 192
Appreciate learning, 168
Artificial intelligence, 41
Assessing incremental, 226
Assessing learners, 164
Assessing treatment outcomes, 193
Assessment approaches, 96
Assessment instrument, 235
Assessment manager, 32
Assessment method, 136, 176
Assessment methodologies, 72
Assessment of teaching, 51
Assessment system, 31
Assessment technique, 52
Asynchronous learning, 39
Avoid plagiarism, 134

B

Behavioral Aspect, 185
Behavioural sciences, 80
Bloom's Taxonomy, 4, 182

C

Classification of learning, 3
Classroom activities, 48
Classroom group, 167
Classroom instruction, 78
Classroom management, 32
Classroom participation, 175, 178
Classroom teachers, 150
Clever classroom, 62
Clinical psychology, 226
Cognitive psychologists, 5
Cognitive system, 7, 8
Communication medium, 150
Communication skills, 38, 40, 109, 175
Communication technolog, 36
Competitive assessment, 98, 236
Comprehensive Evaluation, 165, 172, 173, 176
Concept of evaluation, 16
Concurrent information, 237

Construct validity, 220, 237
Contextual influences, 204
Conventional examination procedures, 122
Conventional methods, 109
Covid 19 pandemic, 123
Credit System, 173
Critical principles, 232
Critical thinking, 125, 135, 175
Cultural sensitivity, 70
Curriculum Evaluation, 79

D

Decision-making, 226
Demographic characteristics, 74
Demographic groups, 229
Demonstrate mastery, 119
Determination methods, 236
Diagnostic assessments, 57
Digital classroom, 66
Digital resources, 36, 132
Discriminant validity, 220, 222
Distance learning, 123
Diverse evaluation, 174
Dynamic landscape, 119

E

Ecological validity, 225
Economic development, 35
Education aims, 156
Education system, 10, 19
Educational assessment, 25, 70, 92. 96
Educational evaluation, 23, 24
Educational growth, 76
Educational improvement, 26
Educational institution, 30, 31, 173
Educational journey, 137
Educational measurement, 26, 44, 103
Educational policies, 70
Educational programme, 20
Educational studies, 4
Educational system, 239
Educational technology, 35, 41
Effective communication, 179
E-Learning platforms, 40
Empirical evidence, 224
Empirical studies, 123
Empirical validity, 238
Ensuring consistency, 117
Environmental impact, 204
Essential component, 25, 132
Ethical implications, 228
Evaluates knowledge, 78
Evaluating educators, 51
Evaluating information, 118
Evaluation happens, 50
Evaluation involves, 22
Evaluation judges, 17
Evaluation processes, 239
Evaluation system, 161, 186
Expert curriculum judgements, 92
Expert judgment, 228
External assessment, 27
External criterion, 219
External environment, 30
External validity, 216

F

Financial benefits, 41
Financial statements, 148
Formal classrooms, 136

Formative assessments, 174
Formative Evaluation, 18, 156
Framing strategies, 10
Fundamental advantage, 118
Fundamental components, 46
Fundamental knowledge, 140
Fundamental need, 230
Fundamental principles, 239

G

Good evaluation, 48
Good indicator, 210

H

Health questionnaires, 195
Holistic development, 175
Holistic evaluation, 176
Holistic learnin, 174
Holistic scoring, 153
Human being, 18
Human development, 4
Human judgment, 195, 235
Human life, 15
Human or systematic, 202
Hypothetical constructs, 229

I

Important method, 199
Incremental validity, 226
Independent variables, 212
Innovative methods, 126
Internal consistency, 197
Interpersonal abilities, 109

J

Judgments align, 195

K

King abilities, 140
Knowledge and skills, 201
Knowledge application, 124

L

Learning achievements, 7
Learning activities, 52
Learning assessment, 153
Learning environment, 64, 175
Learning experience, 97, 173
Learning experiences, 23, 57, 117
Learning journey, 135
Learning objectives, 55, 71, 114
Learning outcomes, 62, 165
Learning paths, 47
Learning process, 35, 130
Learning system, 21
Lifelong learners, 136
Lifelong learning, 116, 118
Lifelong passion, 136
Logical approach, 85
Logistically challenging, 118
Longitudinal research, 194

M

Making judgments, 28
Marking Scheme, 84
Mathematical ideas, 9
Mathematical skills, 237
Meaningful conclusions, 215
Meaningful judgments, 228
Measurable quantity, 207
Measurement and assessmen, 232
Measurement instrument, 197, 200, 218, 219, 220, 222, 229

Measurement quality, 230
Measurement techniques, 44
Medical diagnosis, 227
Mental measurement, 18
Methodology promises, 174
Microlearning, 41
Modern day teaching-learning, 184
Modern economy, 35
Modern taxonomies, 10
Modern workplaces, 118
Moment Correlation, 233
Money travelling, 39

N

National Association of Colleges, 179
National Curriculum, 31
National Education Policy (2020), 122
Numerical values, 26

O

Online environment, 35
Open Book Examination, 110, 113
Ormative evaluation, 153

P

Pedagogical philosophies, 108
Personnel selection, 227
Physical growth, 24
Physical location, 40
Physical measurement, 18
Portfolio assessment, 149
Portfolio construction, 150
Potential disparities, 118
Professional contexts, 114
Professional growth, 136
Psychological assessments, 198, 216, 228
Psychological measurement, 18
Psychological tests, 193

Q

Quality of education, 44
Question Bank, 181
Quick information, 131
Quickly determine, 100

R

Random sampling techniques, 74
Rational equivalence method, 234
Relative performance, 167
Relevant information, 108, 131, 133
Relevant statistical, 201
Reliability analysis, 74
Reliability of assessment, 84
Research skills, 136
Residential classes, 123

S

School-based assessment, 164
Scientific concept, 16
Self-evaluation, 144
Self-system, 7
Skills and abilities, 51
Smart classroom, 62, 67
Social behaviors, 197
Social learning, 38
Socio-economic background, 9, 70
Specific knowledge, 236
Split confidence, 211
Split-Half dependability, 208
Split-half method, 234
Split-Half Reliability, 201
Statistical analysis, 200
Summative assessment, 56

Summative evaluations, 157
Surprise lamp, 162
Synchronous learning, 39
Systematic organization, 182

T

Teaching experience, 186
Teaching-learning process, 31, 46, 49, 156, 183
Technological advancements, 35
Technological learning, 40
Technological tools, 35
Technology integration, 26
Theoretical development, 221
Theoretical implications, 220
Thinking skills, 184
Traditional classroom, 63, 65, 135
Traditional Examination, 112
Traditional systems, 177
Traditional teaching methods, 184

U

Ultra sonography, 15
Unconventional assessment, 109
Unique academic, 108
Unique characteristics, 92, 211
Unlike traditional, 108

V

Value judgment, 47
Virtual classrooms, 39
Virtual reality, 38

W

Writing skills, 182