

Research in Educational Evaluation and Examination

A TREND REPORT

PRITAM SINGH

VED PRAKASH

Ever since the formalization of the institution of examinations they have played a key role in influencing the educational process as well as in classifying students for various purposes. Because of this they have attracted the attention of researchers. Research in examinations during the last four decades calls for special attention as it was in these years that it ceased to be just a haphazard pursuit and became a consciously directed effort to study the subject in a concrete frame of reference. Surveys and reviews of studies in this field began to acquire significance. Long and Mehta (1966) published the first mental measurement yearbook about researches done in examination and evaluation. This was followed by the *Directory of Behavioural Science Research in India* (Pareek and Kumar, 1966). Dave (1968) reviewed researches in the area of examination and evaluation. Subsequently, research in examination and evaluation was reviewed by Buch (1972), Buch and Passi (1974), Passi and Padma (1974) and Passi and Hooda (1986). These surveys followed a system of classifying examination research and identified the problems and suggested the priority areas of research in this field. The previous surveys were based on researches done in India up to 1983. The present report takes into view the researches done till March 1988. There have been a total of 225 studies done up to date. Out of these, 88 are reported to have been submitted as theses for the Ph.D in Education, 11 in Psychology and two in Physical Education. Besides these, 124 are project reports. All these investigations have been classified in six areas and their periodical distribution is shown in Table 18.1.

Research in Educational Evaluation and Examinations seems to have been taken seriously by researchers

Table 18.1

AREA-WISE AND PERIOD-WISE DISTRIBUTION OF STUDIES

Year	Up to 1949	1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	1980-84	1985-87	Total
Area										
Achievement	-	-	2	10	16	17	17	10	1	73
Diagnostic Test	-	-	-	-	5	5	2	2	-	14
Examinations	-	-	2	12	17	16	17	13	5	82
Factors Affecting Achievement	1	2	5	2	4	3	2	2	-	21
Prediction-Admission Promotion Studies	-	2	2	6	4	3	3	1	1	22
Failures	-	-	2	4	2	-	3	1	1	13
Total	1	4	13	34	48	44	44	29	8	225

during 1960-79 when a number of Ph.D. studies and research projects in this area were undertaken. Though the problems undertaken were related to various allied fields also, the main emphasis was on Achievement and Examination. The maximum number of studies, i.e. 48 are reported to have been completed during 1965-69

A slump period in the history of research in Educational Evaluation and Examinations again set in during eighties. Figures reveal that only 29 studies were undertaken during 1980-84 as compared to 44 during 1975-79. The situation got still worse in 1985-87 when only eight studies were conducted and the area of diagnostic testing remained completely untouched. Two out of six areas were mostly favoured by the researchers. The maximum number of studies (82) are reported to have been done on the Examinations side, followed by Achievement (73). The work done in areas like Failures (13) and Diagnostic Testing (14) is very meagre. Studies on Prediction-Admission Promotion (22) were also not adequate in number.

METHODOLOGY

(a) Sample

Investigators, while standardizing tests of various kinds, have used samples of different sizes and nature. A large number of studies were carried out on a sample of less than 3000 subjects. It was only in a very few studies that the sample used was as large as 7000 students. The sample for all these studies was drawn from various parts of the country. For studies like construction and standardization of an entrance test, the sample was drawn from a wide range of population. The techniques employed for drawing the samples differed very widely and included stratified sampling, random stratified, clustered multistage randomisation and simple randomisation. This was done on the basis of the nature of the related population and the purpose of the study. Besides, in some studies of special kinds, the sample was drawn from amongst teachers, guardians and students.

(b) Tool Construction

By and large, the researchers either used examination marks or developed achievement tests for completing their studies. The usual standard steps were employed for constructing achievement tests and question papers. Some researchers rigorously followed procedures such as content analysis, identification of concepts and objectives, development of design, preparation of blueprint, constructing test items, editing, trying out and item analysis. Besides estimating the coefficient of reliability and validity using different methods, norms for

certain tests were also established.

(c) Item Analysis

It appears from the review that only two item parameters, namely, item difficulty and item discrimination, were widely used. Further, it seems that in quite a few of the investigations conditions required for carrying out item analysis were not seriously taken into consideration. It may be appropriate to mention here that if test performance is dependent upon various demographic and other variables, separate item analysis ought to be carried out for these subgroups. By and large, this was not taken into account.

(d) Reliability

Various methods have been employed for computing reliability such as analysis of variance, split-half, parallel form, KR-20, KR-21, and test-retest. Though many researchers used the test-retest method, the time gap between two administrations had been fixed casually. None of the studies introduced multiple time gaps. Multiple time gaps with test-retest or parallel forms should be studied to establish the stability coefficients. It is also observed that, in quite a few of the studies related to essay-type tests, content reliability had not been computed. The marks reliability for essay-type tests can be obtained as the product of content reliability and examiner reliability. In essay-type tests, the marks reliability is the content reliability attenuated by examiner unreliability. Besides, standard error of measurement which adds to the interpretation of the reliability of a test seemed to have not been reported in many studies. It is suggested that large-scale multi-facet studies of reliability, aiming to find out the error components due to various identifiable factors such as due to items, examiners, examinees, instructions, occasions, etc. be carried out.

(e) Statistical Design

Of 225 studies included in the report, as many as 180 are of a descriptive and correlation type, 23 are factor analytical, 15 regression analytical, one regression and factor analytical and six experimental. Experimental studies have used simple pre-test and post-test design with one treatment and one control group. The factor analytical studies have aimed at either evaluating the factorial validity of different instruments or classifying different school subjects into new families of subjects or

examining the nature of factors involved in the test batteries. The regression studies have aimed at establishing multiple regression equations and multiple correlation between predictors like SSC examination marks and criteria like college grades.

ACHIEVEMENT TESTS

A review of studies on achievement tests indicates that investigators have constructed and standardized achievement tests in various areas such as general scholastic achievement tests, achievement tests in language, social sciences, mathematics and sciences. Besides these, efforts have also been made to develop some tests of miscellaneous types. A brief review of all these investigations following the said order is given below.

(a) General Scholastic Achievement Tests

Tests of general scholastic achievement are tests of general educational development. They are of special interest because they measure complex learning outcomes that cut across subject-matter lines and are common to the major content areas of school. They emphasize understanding, interpretative skills and the ability to apply knowledge and skills to new situations. Since the learning outcomes measured by these tests are closely related to the ultimate objectives of education, such tests are especially likely to have a desirable influence on curriculum and teaching methods. These tests can be made use of to measure the general rate of student learning and to group students for various educational tasks. Keeping in view the usefulness of such tests for teachers as also for guidance workers, concerted efforts are needed for their development. Some efforts have been made by Parikh (1946 b), Liddle (1965), Jha (1974), Sharma (1975), Sharma (1976), Patel (1977), De (1979), and Shah (1982). Lele and Parikh (1965) constructed a Scholastic Aptitude Test for admission to preparatory science courses which comprised three sub-tests in English, Numerical Ability and Abstract Reasoning. Liddle (1965) standardized an Academic Aptitude Test for high school students of Uttar Pradesh. It consisted of tests of vocabulary, numerical computation, sentence completion and mathematical reasoning. The coefficients of reliability for each subtest as well as for the total test ranged from 0.83 to 0.89. The concurrent validity coefficients against scholastic achievement in terms of total scores ranged from 0.46 to 0.76. It ap-

pears from the review that the general scholastic achievement tests have been developed covering the subjects, English, Hindi, general science, mathematics, history and geography. These tests were constructed for the students of grades V to XI, in the states of UP, Gujarat, Punjab, Rajasthan and West Bengal. However, they have not been developed for the subjects of social studies, physics, chemistry and regional languages. Such tests for affective and psychomotor domains have not been developed at all. Some of the tests developed were content-oriented rather than oriented to basic skills and general educational development, which restrict their usefulness in these areas. Efforts should be made to overcome this limitation.

(b) Achievement Tests in Language

It is revealed from the review of research studies that language-skill-based achievement tests have been developed in English, Hindi, Gujarati, Oriya, Marathi and Kannada. In the English language, as many as 11 tests for Grades VI to XI have been constructed. A good deal of work has been done in this area by Aram, Rangaswamy & Feroze (1957), Buch, Patel and Kotwal (1960), Misra (1970), Deshpande (1972), Sinha (1967), Chatterji *et al.* (1970), Patel (1971), and Skariah (1981). As regards the Hindi language, a few achievement tests have been constructed by Shukla and Tutoo (1959), the CIE (1962), the Gujarat Research Society (1963), Jha *et al.* (1964), Sharma (1967), Deshpande (1972), Gaur (1973), Giri (1976), Verma (1977) and Joshi (1980). These studies provided tests for grades V to SSC level.

Amongst regional languages, Gujarati got the maximum attention from researchers in the area of achievement test construction. Efforts in this direction have been made by Buch, Patel and Kotwal (1960), Bhagatwala (1960), Maniar (1961), the Gujarat Research Society (1963), Bhatt (1971), Krishnamurti (1971), Maniar (1973), Pandya (1973), Parekh (1973), Desai (1974), Gohil (1974), Modi (1975), Bisnagari (1976), Patel (1978) and Upadhyaya (1979). Achievement tests in Gujarati are available for all grades from V to pre-university and for children in the age group three to five years. Besides, there are three other regional languages, namely, Kannada, Oriya, and Marathi, for which tests have been constructed. Dash (1967) standardized an achievement test in Oriya for Grade VII students, whereas Deshpande (1972) developed an objective assessment tool in Marathi for students appearing for the secondary education examination in

Maharashtra. Shivananda (1981) standardized Reading Tests in Kannada for pupils of standards V to VII, separately.

It appears from the review that except in four regional languages namely, Gujarati, Marathi, Oriya and Kannada, no systematic and sustained efforts have been made to standardize tests in regional languages. Therefore, it is not only proper but highly desirable to gear up work on the construction of achievement tests in languages at the national level as well as at the regional levels. It is noteworthy to mention that the studies in the areas of reading speed, reading comprehension and listening comprehension are restricted to the Gujarati language. Efforts should be made to coordinate research studies carried out at the M. Ed. level also to have a better perspective. Attempts should also be made to develop achievement tests in each regional language for all grade levels.

(c) Achievement Tests in Social Sciences

The present inadequacy of tests in the social sciences urgently demands the designing of tests in this important area. Some efforts for the development of achievement tests in the social sciences were made by Aram *et al.* (1957), Shukla and Tutoo (1959), Buch *et al.* (1960), the Gujarat Research Society (1963), Saraf (1964), the SIE, Kerala (1965), Dash (1967), Muzaffar (1967), Srivastava (1967), Misra (1968), Vanajakshi (1970), Misra (1970), Deshpande (1972), and Sharma (1981). They developed tests in social studies, history, geography and civics only. The social studies tests were constructed for students of grades IV to VIII, in history for grades V to XI, in geography for grades V to VIII, X and XI and in civics for grades IX to XI.

It is observed that, by and large, achievement tests in social studies, history and geography are available for grades V to XI devised on the basis of studies of samples from various states. However, the general observation is that achievement tests in the social sciences have not been developed in all the subjects, even within a state, and for subjects like economics, sociology, etc., they have not been standardized at all. An interesting investigation was carried out by Tiwari (1982) in which he tried to make a comparative study of trends of achievement measurement in civics in higher secondary examinations of various Boards of Secondary Education. His findings reveal lot of inconsistency with regard to difficulty level, objectives tested, etc.

(d) Achievement Tests in Mathematics

Compared to the position in other disciplines, more tests have been constructed in mathematics, including arithmetic, algebra, geometry and trigonometry. Some investigators like Aram *et al.* (1957), Maniar (1961), the SIE, Kerala (1965), Dash (1967), Kulkarni *et al.* (1970), Misra (1970), Vanajakshi (1970) and Bhatt (1971) constructed achievement tests in mathematics either for a doctorate degree or for institutional projects. As regards test construction in arithmetic, Chickermane (1943), Dave (1958), Buch *et al.* (1960), Pendharkar (1965) and Basu (1969) made some efforts and provided tests for grades III to X for the states of Maharashtra, Gujarat, Mysore and West Bengal. Besides these, Gokhale (1954) developed an achievement test for geometry, Jha (1974) for arithmetical and geometrical concepts, Gupta (1974) for algebra, trigonometry and geometry, and Tewari (1975) for arithmetic, algebra and geometry. Sharma (1978) constructed a battery of sequential achievement tests for classes V to X. Besides, Ketkar (1982) standardized unit achievement tests for standard VIII for pupils studying in Maharashtra. With reference to the categories of Guilford's structure of Intellect Model, Chauhan (1982) constructed and standardized an achievement test in algebra for class IX.

It is observed that, by and large, validation of achievement tests in mathematics has been done using achievement scores as the external criterion. It is, therefore, desirable to establish other types of validity coefficients. Another important thing which needs mention here is that, with change of curriculum in mathematics, the available achievement tests would become obsolete. Therefore, it is not only proper but highly desirable to develop corresponding measuring tools.

(e) Achievement Tests in Science

A fairly large number of studies have been undertaken in this area. Of the various studies, there happen to be many of test construction in general science, physics, chemistry, botany, zoology and home science. Studies carried out in the area of general science were by Aram *et al.* (1957), Buch *et al.* (1960), Saxena (1960), Gupta (1962), the Gujarat Research Society (1963), the SIE, Kerala (1965), Dash (1967), Sheth (1967), Rup Prakash (1968), Vanajakshi (1970), the SCERT, Hyderabad (1971), Bhatta (1971) and Hira Devi (1973). By and large, the tests are available for grades V to VIII of the

states of Tamil Nadu, Punjab, Haryana, Maharashtra, Andhra Pradesh, Gujarat, Kerala and Orissa. For subjects like physics and chemistry, a few investigators like Bountra (1970), Gupta (1974), Tewari (1975), Sali (1977), Chhaya (1978) and Khandelwale (1981) standardized tests for high school and college students. They have drawn their samples from Uttar Pradesh, Haryana and Maharashtra, Kapoor (1968) and Garg (1969) standardized achievement tests in home science for secondary and higher secondary students of Uttar Pradesh respectively.

It appears from the review that achievement tests in other science subjects have not been constructed so far. Thus, it is desirable to develop more and more tools to keep pace with changing syllabi in science.

(f) Miscellaneous Tests

Besides development of tests in specific subjects, some have been developed for the purposes of admission, selection of subjects, and classification and selection of candidates on the basis of their skills. Such tests were constructed by Lele and Bhagatwala (1954) for university entrance, Mascarenhas (1964) for medical fitness, Satyamurthy (1965) for subject selection, Shukla (1957) for physical education, Pillai *et al.* (1967) for handwriting, De (1965) for educational psychology, the GCPI, Allahabad (1976) for evaluation of questions, Datar (1984) for B.Ed. in educational psychology, and Sangral (1986) for physical education.

The review of these investigations reveals that there is no coordination amongst various organizations and agencies which are busy developing tests of these types. There should, therefore, be some centrally organized agencies giving thought to future requirements of such tests in different areas as also carrying out the job of monitoring and evaluation of these tests.

DIAGNOSTIC TESTS

It appears from the review that very little work has been done in the area of diagnostic testing. Only 14 studies have been conducted so far. Of these, seven are Ph.D. studies, and the remaining seven sponsored research projects. Among the universities and organizations where these investigations have been carried out are the universities of Bombay, Kerala, Agra, Punjab, Kurukshetra and Lucknow, the Bureau of Psychology, Allahabad, and the SIE, Gujarat. It was way back in

1966 that the first attempt to investigate this field was made by Mehta (1966), who undertook doctoral research work in diagnostic tests. Later, some more studies were undertaken. These covered the subject areas of Hindi, English, Gujarati, arithmetic, algebra, physics and chemistry. Diagnostic tests have not been prepared in areas like social studies, drawing, and life sciences, from the methodological point of view, the survey method and the experimental method have been used. The items included in these tests are completion, matching, alternate response, short-answer and multiple choice. As regards the coefficients of validity, only content validity and concurrent validity have been established. The coefficients of reliability have been established by split-half, test-retest, parallel forms and the K-R formula.

(a) Languages

There appears to be only five studies regarding the construction of diagnostic tests in language. Sinha (1971) constructed a diagnostic test in Hindi for Bangla-speaking students of grade VI of Haryana. A diagnostic reading test in Hindi was standardized by the I.T. College, Lucknow (1971). This test was made for the grade III students of Uttar Pradesh. Gomathy Ammal (1972) conducted a study to locate the specific difficulties of pupils learning Hindi in Kerala. The SIE, Gujarat (1969) constructed a diagnostic test on Gujarati spellings and the use of the *anuswara* for students of grades IV and V. A diagnostic test in English preposition use for students of the pre-university class was constructed by Sharma (1982).

The review of the studies reveals that diagnostic tests for language skills, such as listening, speaking, reading and writing have not been developed adequately. Not even a single diagnostic test has been prepared for the English language for any school stage. The construction of diagnostic tests in languages at the national and regional levels will need more and more attention over the years to come.

(b) Mathematics

A diagnostic test related to the use of four fundamental rules of arithmetic was constructed by Mehta (1966). The SIE, Gujarat (1969) also carried out similar work. Patel (1976) developed a battery of diagnostic tests in arithmetic for Gujarati-medium students studying in grades V, VI and VII in Greater Bombay. Thakore

(1980) constructed diagnostic tests on fractions and decimal fractions for students of grade V. Sharma (1969) constructed a diagnostic test in algebra for students of grade VIII of Uttar Pradesh. Ashar (1972) standardized a diagnostic test in basic algebraic skills for Gujarati-medium students of grades VIII, IX and X from Greater Bombay. A diagnostic test of skill in using geometrical instruments for grades VI and VII was developed by the SIE, Gujarat (1969). Hussien (1967), while comparing the achievement of pupils in mathematics of 12 countries, reported that the achievement of pupils in this area in India is much lower than that of the pupils of other countries. This underlines the need for paying more attention to diagnostic testing in mathematics and developing remedial programmes.

EXAMINATIONS

While constructing any measuring tool, we not only strive to make it reliable and valid but also to ensure that it has a beneficial rather than a harmful backwash effect on curriculum and teaching. Besides, we also strive to ensure that, in a regional or national examination, grading standards should be comparable.

Of the 225 studies in the area of Educational Evaluation and Examinations, as many as 82 have been classified in the area of examinations. However, no work was reported in this area till 1954. The subject gained some importance only in the sixties. The studies of this period are related to various aspects of research on examinations, such as achievement in annual examination, inter-examiner and intra-examiner reliability, question papers and their nature, pass percentage, external and internal assessments, correlation between theory and practical marks, mass copying and innovations in examinations. Besides the Ph.D. studies by Rao (1968), Misra (1970), Deshpande (1972), Shah (1972), Tluanga (1974), Sharma (1977), Mascarenhas (1977), Sinha (1977), Singh (1978), Nath (1980), Verma (1981), Singha (1983), Kushwaha (1985), Singh (1986), and Rashid (1987), there are institutional projects conducted at Kharagpur, Gauhati, Baroda, Allahabad, New Delhi, Anand, Bombay, Annamalai, Nagpur, Saugar, Karnal, Calcutta, Pune, Madras, Hyderabad, Kerala, Jammu, Shimla, Rajasthan, Karnataka and Kurukshetra. The institutional projects were carried out by Bokil (1956-63), Shukla (1959), Lele *et al.* (1962-63), Taylor (1963), Taylor and Tluanga (1963), Taylor (1964 a, 1964 b), Raina (1964), Venkubai (1965), Dave

and Patel (1966), Taylor and Tluanga (1966), Sharma (1966), Taylor, Tluanga and Misra (1966), Bose *et al.* (1967), Misra (1968 a, 1968 b), Kamat (1968), Jhaveri and Patel (1968), Patel (1968), Rao (1968), Gayen *et al.* (1961-70), Misra (1969), Misra (1970 a, 1970 b), Misra (1971), Harper (1962-70), Nath (1972-74), Chatterji and Mukherjee (1973), Deo (1974, 1980), Tluanga (1974), Tewari (1975), Bhatt *et al.* (1978), Sali (1978), Mandal (1978), Sali and Umathe (1979), Kaul (1979), Gunasekaran and Jayanthi (1979-81), Rasool *et al.* (1981-83), Sen (1981), Rajasekharan (1982), Nagaraju and Usha (1983) and Gupta and Verma (1985).

(a) Achievement in Examination

With a view to evaluating the achievement in different subjects, Gayen *et al.* (1961-70), have taken the results of the school final examinations and that of higher secondary examinations conducted by the West Bengal Board of Secondary Education and applied various statistical techniques to them. They observed that the reliability and validity of most aspects of the present examination system were very low and this was so because the performance of students was assessed in a single, final examination, by a large number of examiners on a set of questions mostly of the essay type. The *J* effect was observed, having a cluster of marks at the pass mark and other critical zones, followed by big gaps just below these.

(b) Examiner Reliability

Researchers have observed time and again that essay-type examinations are not as valid and reliable as objective-type examinations. It has been established by several investigators that chance plays a very significant role in determining the achievement level of a pupil. Inter-examiner reliability was reported to be very low. The low inter-scorer reliability was attributed to the essay type of examinations, types of questions, and inadequate instructions and training of examiners. Different examiners showed a large variation in the mean and standard deviation of their evaluation. It was because of this that researchers have suggested calibration of marks to make them more comparable. In some investigations it has been proved that about 25 per cent of the candidates benefited as a result of re-valuation. Besides this, the level of intra-examiner reliability, too, has been reported to be a matter of concern in the conventional type of examination.

(c) Question Papers

Question papers are the real backbone of the system. One might very well conceive a situation in which candidates encounter defective question papers. Sometimes candidates fail not because they do not know the subject but because of some external factors like defective question papers, carelessness, etc. In most of such cases, the questions in the question paper were found to be not properly balanced as far as the difficulty value of the question was concerned. Question papers in certain cases seemed to fail to discriminate between higher ability and lower ability groups as they mostly had questions which required mental operation at a lower level. Open book examination did not greatly help students when the questions were set in a reformed way.

(d) Question Bank

Some efforts have been made to develop question banks using Bloom's taxonomy as a tool for prevalidation. Singh (1978), while validating Bloom's taxonomy of educational objectives by analysing the product and process-oriented approaches, found that objective synthesis in the hierarchy of taxonomy seemed to be misplaced while the objective analysis appeared to be misleading. Gnanapragasam (1975) developed a question bank using Bloom's taxonomy of educational objectives. Natarajan (1975-80) at the Association of Indian Universities has developed a series of question-bank books for various university courses, both at undergraduate and postgraduate levels. Hooda (1976) also made an attempt to develop a question bank in education. The NCERT and the Examination Reform Unit of Madras University have also made attempts to develop question banks.

(e) Pass percentage

Some very important decisions concerning the system of examination are being taken quite arbitrarily, such as deciding the cut-off point, awarding of grace marks and fixing the pass-fail percentage. In certain cases, a pre-determined cut-off-point is being used, without any rationale. Dave and Patel (1966) found that there was a substantial variation among the pass percentages in a school examination in a given year. Not only this, they also found a substantial increase in variation in the pass percentages over a five-year period, from 1960 to 1964. Sharma (1966) while analysing the UP Board results,

found that, from year to year, the average performance continuously increased in Hindi, mathematics and science; but in English a reverse trend was observed from 1961 to 1965. The SCERT, Haryana (1970), also observed a substantial increase in variation in pass percentages over a five-year period, 1964-68, of middle-standard examination results. Rajasekharan (1982), while comparing the results of Kerala, Calicut, Madras, Mysore and Tirupati Universities, found significant variations in the percentage of passes across the years. In another investigation, Shukla (1959), while analysing Delhi University results over 1950-57, found that the level of marks obtained by examinees of different subjects at the M.A. examination had different meanings. Nath (1974) revealed that the subjects of economics, English and geography contributed significantly to label a college as below average in attainment.

(f) Internal-External Assessment

Raina (1964), Kamat (1968), Deshpande (1972), Singh (1972), Gunasekharan and Jayanthi (1980), Nath (1980) and Rasool *et al.* (1981-83) studied the relationship between external and internal assessments. Some of them reported a significant correlation between them while some observed that it was less in arts than in science subjects. However, Rao (1968) found that correlation coefficients between external and internal assessments were greater in language subjects than in other subjects. Gunasekaran and Jayanthi (1980) found a significant correlation between the marks of continuous internal assessment and the university examination. Nath (1980) found that internal assessment marks possessed some predictive value for external assessment marks but there was a tendency towards overmarking in internal assessment. Venkubai (1965) found internal assessment marks responsible for higher percentage at the public examination. Bennur (1971) reported mixed results. Reddy (1979) in his study revealed that teachers of the college where internal assessment had not yet started did not differ in their attitude towards the internal assessment system from those of the college where the system was in practice. Most of the investigations carried out in this area signify that there has always been liberal marking in internal assessment.

(g) Theory and Practical Examinations

Very few investigations have been carried out to study the relationship between marks obtained in theory and

in practical examinations. Sali and Umathe (1979) studied the relationship between these two at the SSC Examination of Maharashtra. Hooda (1976) studied the relationship of attainments in theory papers with student teaching in secondary training colleges. These studies revealed that marking in theory and practical aspects of examinations had a wide disparity. Efforts are required to settle the issue of marking pattern in theory and practical examinations.

(h) Mass Copying

With a view to studying this growing menace in public examinations, Mandal (1978) carried out an investigation to find out the factors associated with mass copying behaviour among students in the final secondary examinations, with particular reference to their home and school conditions. The findings of the study reveal that 'mass copying students' had relatively poor socio-economic status, low parental education and inadequate facilities at home.

(i) Innovations

Various suggestions have been put forward by several experts to improve upon traditional examinations. Taylor (1963) recommended the scaling of marks. Bose (1965) suggested the screening of examiners. Harper suggested the appointment of independent examiners for each question. Taylor and Tluanga (1963) suggested the use of standard error of marking instead of the arbitrary principle of grace marks and others suggested the vertical system of promotion. Kaul (1979) studied the acceptability and feasibility of thirteen innovations in the examination system at the university stage. He observed that innovations like the semester system, grading system, supplementing easy questions with objective-type questions, internal assessment, etc. were acceptable both to teachers and students. These recommendations, though quite valid and useful, need careful examination and preparation before being implemented.

PREDICTION—ADMISSION—PROMOTION STUDIES

For ages now tests have been used to make predictions. These predictions are, however, verified only with regard to those who have been admitted on the basis of

some predictors and not on the basis of the success-failure ratio of those who have not been admitted. This is one of the snags which we should always keep in mind while interpreting these results.

Mathur (1971) found that Jalota's General Mental Ability Test was a good predictor for general science. Satyamurthy (1965) used a battery of psychological tests for selection purposes and found that success ratios in respect of diversified courses ranged from 0.93 to 0.80. Chakrabarti (1974) developed a battery for predicting probable success of students in the humanities, science, commerce and technical streams. Singh (1975) developed a battery for predicting success in achievement in science at the intermediate level. Ramachandran *et al.* (1971) found that students could be classified with considerable precision into science and humanities streams on the basis of their interest pattern. Mitra *et al.* (1978) found that interest in science could be employed as a differentiating factor in the identification of more capable science students. Bokil (1958b), Buch (1963) and Sengupta (1963) used the final examination marks for predicting achievement in the next final examination. Dutt (1954) observed that subject choice at the degree level by those who secured less than 45 per cent marks should be made with caution. Dharm Bir and Srivastava (1965) found that subjects like Sanskrit, mathematics, arithmetic and household accounts had higher prediction coefficients than subjects like geography and English. Bhatt and Indiresan (1981) found that high school scores were not a good predictor of success in polytechnics. Buch (1963) found that SSC examination composite scores of physics, chemistry and mathematics were better predictors than the individual set of scores while predicting success in the next higher examination. Palsane (1965) found that the language group marks had the highest predictive potentiality in the faculties of arts and engineering. Hiriyaniah (1963) found that the predictive value of the pre-university examination in science was better than that in the arts and commerce groups. Nath (1973) found that unscaled marks turned out to be as good predictors as scaled marks. Lele and Bhagatwala (1954) developed a standardized entrance test. Narayana (1979) found that in the selection of candidates by the Andhra Pradesh Public Service Commission, those with first division predominated over those with second division and so on. Singh (1986), while studying the effect of teaching strategy for competitive examinations, found that lecture-cum-diagrammatic presentation followed by question-answer was the most effective strategy for achieving better success.

FAILURES

As many as 13 studies have been carried out in this area. Of these, two are doctoral studies and 11 are institutional studies related to the area of failures in examinations. The studies were conducted by Bokil (1956c, 1958a, 1963), Gadgil (1978, 1979b), the five institutional studies of the DEPSE (1964), the NCERT (1964, 65), the GCPI, Allahabad (1964), the Directorate of Higher Education, Hyderabad (1966), Kumar (1982) and Deka (1985).

The chance factor plays a pivotal role in our examinations. Bokil (1956c) found that size of school influenced the number of candidates scoring less than 20 per cent of the total marks. The DEPSE (1964) found that there were various extrinsic factors such as teachers' qualifications, teaching methods, location of school, equipment, etc. responsible for the pass-failure percentage of schools. The NCERT (1965) reported in their studies that compartmental examinations conducted by various boards did not seem to make much difference to the large-scale failures. The Directorate of Higher Education, Hyderabad (1966), findings were in tune with the findings of the DEPSE (1964). However, Gadgil (1978, 1979a, 1979b) in his studies concluded that a large number of failures were mainly due to teachers' incompetence. Deka (1985) while studying school failures, suggested that less academically oriented students should be directed to vocational courses rather than academic courses.

This is such an important area that more and more studies should be undertaken to fill in the gaps which have not been uncovered so far.

GAPS AND PRIORITIES

Gaps

This review of the studies carried out in the area of examinations and evaluation indicates that not enough work has been done on these subjects. The few studies we have are neither integrated nor goal-directed. In achievement test construction, even though some dent has been made at middle, secondary and higher secondary levels, test construction at the primary stage and at higher education level is almost negligible. There appears to be lot of duplication in studies in examinations and evaluation. Quite surprisingly, no organization has undertaken well-knit studies over a longitudinal period

covering a large number of school subjects. In quite a few studies, construction of achievement tests and evaluation procedures have been treated as an end itself.

Most of the studies were conducted following the traditional steps of tool construction within the framework of norm-referenced testing. New areas like criterion-referenced testing and mastery-learning need to be incorporated in tool construction. The sample drawn in most of the studies was neither adequate nor properly drawn. Group-dependent statistics have been used to carry out item analysis. Efforts have not been made to use sample-free item analysis. In only a few studies was content-validity established by matching course content with course objectives. Concurrent validity has been established by using valid criterion. Coefficient or reliability has been established by using various methods. In certain cases, the Kuder-Richardson formula has been used without fulfilling the conditions for its application.

There is an urgent need to develop more and more tools to match the needs of the changing syllabi in school subjects. It appears from the review that tools to measure general scholastic achievement have not been developed for all the subjects at all levels. By and large, the achievement tests developed have not been standardized. Very little work has been done in the area of diagnostic testing. Besides improving upon their quality, efforts must be made to develop diagnostic tests in different subjects at different levels. Concerted efforts to develop these should be made at national level as well as at state level.

As many as 82 studies have been carried out in the area of examinations, covering various aspects such as inter-examiner variability, internal assessment, mass copying, grading, scaling, etc. Studies must be undertaken to fix the proportion of objective-type, short-answer type and essay-type questions. There is an urgent need to undertake more studies on the problems of scaling in order to avoid disparity. More studies should be undertaken in areas like question-wise analysis, question banks and open-book examination. Studies related to factors affecting achievement at primary level and at higher education level are almost negligible in number.

Prediction-admission-promotion studies reveal that achievement test marks are invariably used for making predictions which may be practical but definitely non-academic. Therefore, admission tests, selection tests, entrance tests and promotion tests should be developed independently. The problem of failure is quite

a serious one as one-half of the students fail in their final examinations. By and large no scientific method is being used to determine the cut-off points for different subjects. Questions used in the test papers are such that they do not require mental operations at a higher level. Efforts should, therefore, be made to include such questions in the test papers as require mental operations at a higher level. Questions seeking less factual information should be avoided. Evaluation of affective and psychomotor domains, which have been ignored for all these years, needs to be done as systematically as that of the cognitive domain. Organizations responsible for testing should make use of the latest electronic aids for development of a computerized question bank and scoring of answer scripts by an optical mark reader with a view to introducing more objectivity and accuracy and also to make the system more economically viable. Above all, premier organizations like the NCERT, UPSC, and AIU, should make concerted efforts to provide guidelines to new researchers to undertake studies aimed at making the system more scientific and rational.

Priorities

(a) Survey of research in this area has revealed that the measurement aspect of evaluation has been over-emphasized. Almost all the studies relate to written examinations and testing. Studies relating to oral testing, performance testing and non-testing tools and techniques, like observation, inquiry and analysis seem to have altogether escaped the attention of researchers and advocates of research. Even the new or emerging fields of testing like diagnostic and criterion-referenced testing have not attracted the needed attention of the researchers.

(b) An even more discouraging feature is that the societal context and ecology of examination and evaluation are almost totally neglected or relegated to the background. This is in spite of the fact that the social consequences of examinations vis-a-vis evaluation systems are of great significance and concern for students, parents, teachers, employees and, in fact, the society at large. Therefore, besides the measurement value of examinations which hitherto has been quite appreciably taken care of, the fact remains that the academic or pedagogical value of reform in examination and evaluation has yet to be appreciated even by our evaluators.

Another area related to examinations that has still to

be recognized as a cogent field for research is the management of examinations and evaluation. Since the quality of measurement instruments depends on the quality of people who write or edit them, their selection and their training is as important as the strategy and logistics of paper-setting and conduct of examinations and the mechanics of processing and declaration of results.

(c) Hundreds of functionaries are involved at various levels at different times to perform different tasks to ensure the efficient conduct of an examination as per a stipulated schedule. The efficacy of the various steps taken to ensure the intended results, the need for proper planning, coordination, monitoring, evaluation and decision-making cannot be overemphasized. For this, suitable systems of management and control have to be worked out and operated within a given time-frame. This aspect has been overlooked and considered of so little relevance that not more than two researches on this aspect of examination were reported during the period under review. Thus we find that, out of the three major aspects of reform, viz. the measurement aspect, the pedagogical aspect and the management aspect, it is only the first which has received its due of attention, perhaps more than its due, while the other two have been almost neglected.

(d) If we compare likewise, studies in the cognitive, affective and psychomotor domains, again we have an almost nil report to present in the affective as well as the psychomotor domains. We have on record only three or four studies. Thus, almost all the studies are undertaken in the cognitive area. This shows that the other two areas are either relatively more difficult to probe or that, perhaps the absence of well-established credo in these two domains, like the one we have in Bloom's Taxonomy of Educational Objectives in the cognitive domain, discourages researchers. The most probable reason is that researchers tend to follow an easy and well-known track rather than enter an unexplored or less-explored field, requiring more rigour in applying less sophisticated cannons of proof to arrive at conclusions and pass judgements. The count of researches involving the use of data at local, national and international level indicates that convenience rather than usability of data for reform was preferred. There is not a single study reported that uses data at the international level while there are only five studies which demand national level data and samples. All the other studies are local, having only a few findings of conse-

quence that might influence the system even at local level. More comparative studies at the national and international levels are necessary for self-evaluation and knowing the regional imbalance, besides their relevance for national identity and international understanding.

(e) About 7 per cent of the studies were undertaken at the primary stage, about 25 per cent of those at the middle stage, about 45 per cent at the secondary stage and about 25 per cent at the university stage. An interesting feature that emerges is the preference for researches at the secondary stage (about 50 per cent of the studies) in comparison with elementary and university stages of education. Whatever the reasons, it is quite obvious that, willy nilly, the research student has to choose this stage which provides the broadest base and the firmest anchor to safeguard researchers' interests. In spite of the priority allotted to improving elementary education and universalization of primary education, as indicated in the Seventh Plan, it has not attracted researchers to select topics related to this stage. Whatever the trend may be, it seems quite reasonable to expect researchers and their guides to take cognizance of the priority areas as reflected in the plans, policies and official documents of the government. When more than 50 per cent of the investment in education is wasted due to failure and stagnation at the elementary stage, why should it not become a priority area for our researchers? It is a social need and a compelling reason for preferring this stage to the secondary and university stages of education.

(f) If the researchers are classified as fundamental and functional, no particular trend is identifiable; there is hardly any study that could be regarded as fundamental, the findings of which could be used as a framework or theoretical construct applicable to a specific population or have far reaching implications for changing, adapting and improving existing practices or models. Likewise, among the functional researches, which are all applied in nature, too few can stand the test of rigour of sampling, construction of tools and techniques and canons of evidence, besides, of course, the subject areas of demand.

(g) Regarding extent of coverage, if the studies are clas-

sified into pupil evaluation, teacher evaluation and institutional evaluation, we find that there is only one research undertaken at Ph.D. level related to evaluation of teachers and only one study on evaluation of an institution. This is indeed a dismal picture because in the National Policy on Education (1986) there is a clear emphasis on accountability of teachers and therefore, construction of measuring instrument for evaluating teachers is the need of the hour. Similarly, institutional evaluation as reflected in the Programme of Action (1986) demands development of evaluation models to work on and identify indicators of success of an institution, which become the basis for development of various tools. This area of evaluation of learning milieu within the social setting of an institution requires attention of researchers.

(h) For obvious reasons the time factor and other inputs required—few longitudinal studies are undertaken. Likewise, case studies of individual institutions or agencies is yet another area which requires more attention. Pace-setting institutions are visualized and the offer of autonomy to selected institutions is on the anvil, as recommended in the NEP. In this context, both case studies and longitudinal studies are to be preferred to the traditional researches. Feedback of results of such studies, to the institutions would provide the needed data and evidence for better planning and management of such institutions.

(i) From the foregoing discussion, it would appear that there are certain areas which have not been touched at all by researchers. In other cases, within a given area, the emphasis or preference of the researcher has been to undertake research in that sub-area which has already been traversed well and a well-beaten track of methodology, data collection and judgement making, as reflected in the earlier studies and findings, is available. For example, in testing, it is limited mostly to achievement testing, not criterion-referenced testing. In achievement testing, it is mostly in science, mathematics and languages, not the vocational subjects. By and large, we have tended to conform to traditional postures rather than to look forward to an innovative and pragmatic outlook. Table 18.2 indicates the gaps and the coverage of various aspects and areas of investigation.

Table 18.2

AREA-WISE CLASSIFICATION OF RESEARCH STUDIES
(Incorporated in 1st, 2nd, 3rd, & 4th Surveys)

Area	No. of Studies
Pupil Evaluation	221
Teacher Evaluation	-
Institutional Evaluation	1
Basic Research	178
Functional Research	43
Action Research	01
Academic Subjects	216
Vocational & Professional Subjects	-
Psychology, Sociology	2
Cognitive	219
Affective	00
Psychomotor	03
State Level	216
National Level	6
International Level	-
Measurement Aspect	219
Academic Aspect	1
Management Aspect	2
Elementary Stage	66
Secondary Stage	105
University Stage	51
Achievement Testing	208
Diagnostic Testing	14
Criterion Referenced Testing	-
Written Examination	220
Oral Examination	-
Practical Exam.	2
LESS REPRESENTED AREAS	
Internal Assessment	8
Continuous Comprehensive Evaluation	-
Question Banking	1
Causal Comparative	1
Case Studies	-
Survey	-

LESS REPRESENTED AREAS	No. of Studies
Physical Education	2
S.U.P.W.	2
Arts	-
Need Assessment Studies	-
Implementation Studies	-
Performance, Monitoring & Evaluation	-
Scaling	1
Grading	1
Moderation of Results	-
Selection tests	-
Moderation of Question Papers	-
Marking Schemes	-

TO CONCLUDE

It might offend researchers, and those who sponsor or coordinate researches, if we say that most researches are undertaken in those areas wherein the researchers or their guides already have expertise, but unfortunately, this is actually so. This leads to perpetuation of the conservative posture in research. The needs, the gaps, the priorities, the social demand, and policy documents did not influence the researchers in the field of examinations and evaluation. It was the convenience, the availability of time and expertise and the beaten track of methodology that were the deciding factors for choice of topics of research. This is why measurement aspects of examination vis-a-vis evaluation is over-emphasized, while the ecological aspect or societal context of examinations is almost totally neglected. Similarly, the management aspect did not find favour with researchers to the desired extent.

New challenges are universalization of education, care of the weaker sections, identifying and nurturing of rural talent, accelerated development of rural and backward areas to remove regional imbalances, removing disparities between boys and girls and general and scheduled caste/scheduled tribe populations. These are some of the serious challenges that the Indian educational system faces today. Alternative strategies of instruction and evaluation are needed to cope with the massive upsurge of numbers. The social dynamics of external examinations, teacher-based assessment in the school system of evaluation, evaluation as a teaching and feedback device, increasing dependence on teach-

ers, better managerial services for conduct of examinations, teachers' accountability and institutional evaluation as a whole rather than pupil evaluation in isolation of the instructional setting and learning milieu and the focus on use of examinations and evaluation as learning experience and a device for improvement of students' learning and development are the focal points for researchers to consider for identifying their topics of research.

The need of the hour is to appreciate the shift from the measurement aspect to the ecological and pedagogi-

cal aspects of examinations, from an experimental to an illuminative approach to evaluation, from a judgemental to a diagnostic role of examinations, from an authoritarian and bureaucratic approach to a democratic and mastery-learning approach, from excellence and elitist focus to equity and the commoner, from the exploitative to mass welfare and from tools as master of men to men as master of tools. More faith in men in general and teachers in particular is the immediate need. Researchers in the field should now clearly show a big tilt in these directions.

ABSTRACTS : 1004—1033

- *1004. DABIR, P.M., *A Critical Analysis of the Marks at the B.Ed. Examination to Study the Trends and Reliabilities of the Assessment*, Ph.D. Edu., Nag. U., 1984

The study was advanced on the basis of the following hypotheses: (1) There is a low correlation between marks in the theory examination (external assessment) and marks in the practical examination (internal assessment). (2) Performance of pupil-teachers in practical subjects is better than that in theoretical subjects. (3) The internal assessment of pupil-teachers in government colleges is less than that of pupil-teachers in private colleges. (4) Private colleges are more liberal in giving internal marks than other colleges. The purpose of the study was to critically appraise the B.Ed. Course of Nagpur University.

The marks in the B. Ed. examination for five consecutive years were collected from the records of Nagpur University. An interview schedule was used as a tool to collect information from principals, lecturers and pupil-teachers about the practical aspect of this examination.

The following were some of the major findings: 1. The general percentage of passes in theory was more than 90 for all the years. 2. There was an increasing tendency of passing and a diminishing tendency of failing. 3. The percentage of successful candidates in theory was remarkably higher in government colleges for all the five years. 4. The ranking order was government colleges as first, private colleges as second, and university colleges as third. 5. The increasing trend of admissions in university colleges was one of the causes of deterioration in the standard of this professional course. 6. There was wide disparity between the marks in the theory examination and the practical examination. 7. There were instances of students in private colleges scoring as high as 90 per cent to 99 per cent marks in the internal assessment.

1005. DATAR, V.V., *Development of an Achievement Test in Educational Psychology for B.Ed. (Marathi Medium) Students and Preparation of a Question Bank*, Ph.D. Edu., SNDTU, 1984

The objectives of the study were (i) to construct objec-

tive type items to test the objectives and specifications of teaching educational psychology, (ii) to test the validity of the items in terms of their difficulty value and discriminating power, (iii) to prepare a question bank consisting of objective based and objective type item having varied difficulty values and positive discriminating powers, and (iv) to develop a reliable and valid achievement test which would consistently test the attainment of B.Ed. (Marathi medium) students in educational psychology.

The content in educational psychology was analysed and objectives and specifications were determined. Two hundred and thirteen objective type test items were constructed. The test was administered to a sample of 166 B.Ed. students. Item analysis was done and valid and discriminating items selected.

The characteristics of the test were: 1. The difficulty value of the items varied from 0.10 to 0.85. 2. The discriminating power ranged from 0.01 to 0.75. 3. The distribution of test scores was approximately normal. 4. The achievement test had a reliability of 0.94. The achievement test scores significantly and positively correlated with the marks obtained by student teachers in educational psychology of the university examination.

1006. GUPTA, S.M., and VERMA, L.K., *Significant Correlates of J and K High Schools Showing Consistently Above and Below Average Results at the Board's Examinations for the Last Five Years*, Dept. of Education, Kur. U., 1985 (NCERT financed)

The objectives of the study were (i) to demarcate and identify the schools showing consistent result—above average pass percentage and below average pass percentage—from 1980 to 1984 in the matriculation examination conducted by the Board of School Examination of J & K State, (ii) to study the views of heads of selected schools showing consistently above average and below average results, (iii) to study the organizational pattern of two types of institutions, (iv) to scale the factors responsible for showing consistently above average and below average results as perceived by heads of both the categories, (v) to compare the significance of differences in the proportion of heads of schools showing consistently above and below average results, (iv) to study the differences in dimensions of adjustment of teachers with respect to sex, locality, economic status,

competency of teachers, and consistent type of results, i.e. above and below average, and (vii) to study the first and second order interaction between sex and type of results, sex and competency of teachers, locality and economic status, locality and type of results, economic status and type of results, sex and competency of teachers and type of results, etc.

A sample of 16 high schools showing above average and nine high schools showing below average results was randomly selected from the various parts of Jammu province of J & K State. From each school four teachers teaching English, science, mathematics and Hindi to students of Class X were selected. Thus, a total sample of 100 teachers was selected from 25 schools of Jammu province. The teachers were administered the Pandey Teacher Adjustment Inventory, and Baroda General Teaching Competency Scale, a checklist, a schedule and a questionnaire specially prepared for the study. The checklist included items like number of teachers in the school, average workload of teachers, school, building, classrooms, teaching aids, etc. The schedule included items concerning discipline, truancy, cocurricular activities, etc. The questionnaire consisted of various factors influencing the results. The data collected with the help of these tools were analysed through percentage, biserial correlation, critical ratio, and analysis of variance.

The findings of the study were: 1. The average number of teachers, both trained and untrained, was more in schools showing consistently above average results as compared to institutions showing below average results. 2. The workload of teachers working in schools showing consistently above average results was less in comparison to schools showing below average results. 3. Experience of heads was more in case of schools showing above average results. 4. Most of the schools showing above average results were situated in urban areas; the students of these schools had not to cover long distances to reach the school. 5. The institutions showing above average results had pucca and planned buildings. These institutions had facilities like a dispensary, library, laboratories, science room, staff room, auditorium, study hall, craft room, canteen, etc. There was a separate office for the head in these institutions. 6. Most of the schools showing above average results had the facility of electricity, heating by charcoal, and fans. 7. The classrooms in the institutions showing above average results were ventilated. 8. The institutions showing above average results had notice-boards for the students; the number of desks in the classrooms and number of chairs in

staff rooms was more in these institutions. 9. More institutions showing above average results used models, charts and maps during the teaching process than schools showing below average results. 10. In the institutions showing above average results, debates, quiz, music competitions, dances, poetic symposia, painting competitions, science fairs, etc. were organized more frequently. 11. Moral education was regularly provided in most of the schools showing above average results. 12. The economic condition of schools showing above average results was better than those showing below average results. 13. The heads in schools showing consistently above average percentage of results ranked in five positions the following factors affecting the matriculation results: ability of students, teachers' style of dealing with the child, teachers' expression, teachers' general ability, and teachers' fund of knowledge. The heads of institutions showing below average results ranked in five positions the following factors: institutional environment, ability of the students, teachers' expression, teachers' style of dealing with the child, and seriousness among students. 14. There was a low correlation between competence of teachers and their adjustment scores.

The study has its implications for school managements, headmasters and the government. (1) The schools should be provided with facilities like library, laboratories, science room, staff room, auditorium, study hall, canteen, playground, common room, etc. (2) Trained teachers should be employed in the schools and care should be taken that each teacher has a reasonable teaching workload. (3) More emphasis should be given to cocurricular activities of the schools. (4) Moral education should be a regular feature of the school education programme.

1007. JESUDASON, D., *A Study of the Perceptions of the School Community in the City of Madras about Reforming of the Present Examination System of the Higher Secondary School Stage in Tamil Nadu*, Ph.D. Edu., SPU, 1986

The objectives of the study were (i) to design and to validate an Examination Reform Description Questionnaire (ERDQ) that would measure the perceptions of the teaching community of higher secondary schools in the city of Madras, (ii) to examine critically the perceptions of the teaching community of higher secondary schools about the academic and administrative compo-

nents and issues related to the introduction of the internal assessment scheme and question bank as reforms in the examination system, (iii) to determine the extent to which biographical and institutional variables such as sex, age, experience, subject of specialization and type of management of the sampled teachers of higher secondary schools had a relationship with their perceptions about the introduction of an internal assessment scheme and a question bank at the higher secondary school stage.

The tool for measuring perceptions, known as the Examination Reform Description Questionnaire, was standardized. Initially 150 statements were coined and were scaled on a five-point scale. The statements were selected on the basis of the Q-value. For the pilot study 135 statements were selected, and again the questionnaire was administered to teachers. On the basis of the correlation matrix, 115 items were included in the final form. The reliability of the questionnaire was determined by the split-half method and was found to be 0.97. The study was carried out on a sample of 500 teachers. The t-test was used for verifying the hypotheses.

The major findings of the study were: 1. Eleven factors were identified and named as (a) validity factor, (b) difficulty factor, (c) reliability factor, (d) utility factor, (e) human ability factor, (f) facility factor, (g) student ability factor, (h) teacher-student cordiality factor, (i) essentiality factor, (j) variety factor, and (k) teacher ability factor. 2. Out of six independent variables, three variables showed significant difference. They were type of management, age (above 40 and below 40 years), and experience (above 15 and below 15 years). The mean score of the perceptions of the teachers in the non-government schools was significantly higher than the mean score of the perceptions of the teachers in government schools. The mean score of the perceptions of the teachers above 40 years of age was significantly higher than the mean score of teachers below 40 years of age. The mean score of perceptions of teachers having more than 15 years of experience was significantly higher than the mean score of teachers having less than 15 years of experience. 3. There was no significant mean difference between the perceptions of men and women teachers. 4. There was no significant mean difference between the perceptions of science and humanities teachers. 5. There was no significant mean difference between perceptions of the postgraduate and inducted teachers. 6. In the validity factor, the mean score of the teachers of non-government schools was higher than

that of the teachers of government schools. In the utility factor, facility factor and student ability factor, the mean scores of the teachers of government schools were higher than those of the teachers of non-government schools. 7. With regard to the difficulty factor, women teachers had a higher mean score than men teachers, while on the teacher-student cordiality factor, men teachers had a higher mean perception score than women teachers. 8. In the validity factor, teachers above 40 years of age had a significantly higher mean score than teachers below 40 years of age. Teachers having above 15 years experience had a significantly higher mean score on the validity factor than teachers having less than 15 years of experience. 9. In the facility factor, the teachers of humanities had a significantly higher mean score than the teachers of sciences.

1008. KHANAPURKAR, U.H., *Construction and Standardization of Silent Reading Comprehension Test in Marathi, for Pupils studying in Standard VII, in the Schools in Osmanabad District of Maharashtra State*, Ph.D. Edu., SNDTU, 1984

The major objectives of the study were (i) to select reading material suitable for pupils of standard VII, (ii) to construct test item based on the selected material, (iii) to determine reliability and validity of the test, and (iv) to establish norms.

Passages from different subjects were selected for preparing the test. The researcher prepared a number of objective based multiple choice items for each sub-test. A preliminary try-out of the test was done on a sample of 740 pupils. Items for the final test were selected on the basis of difficulty value and discriminating power. The final form of the test consisted of 90 items. The final test was administered to a representative sample of 2,990 pupils. The obtained data were analysed with the help of various statistics such as measures of central tendency, variability, skewness, kurtosis. Reliability of the test was determined with the help of test-retest, split-half, and K-R formula 20. Validity of the test was determined by correlating the performance of the pupils on the test with scores on different tests. The performance of 1,000 pupils was taken into account for computing the inter-correlations of 13 sub-tests. Factor analysis was done by Thurstone's Centroid Method. Norms were calculated in the form of standard score norms, percentile norms, etc.

The characteristics of the test were: 1. Performance

of the tested sample was normally distributed. 2. The reliability of the test varied from 0.83 to 0.90. 3. The validity coefficients ranged from 0.57 to 0.77. 4. The test consisted of two factors which were ability of general reading comprehension and ability of deciding technical symbols. 5. There was no significant difference between performances of pupils from urban and rural areas and between boys and girls. 6. The performance of the pupils on the test was not very satisfactory in respect of the different aspects of reading comprehension.

1009. KUSHWAHA, A.S., *System of Examinations in Kanpur University: A Critical Study*, Ph.D. Edu., Kan. U., 1985

The investigation was designed to make a critical study of the system of examinations in Kanpur University with a view to suggesting improvements.

The data were collected through study of the university records and views of people were collected with the help of a questionnaire.

The main findings of the study were: 1. The majority of the respondents felt that the system of internal assessment should be introduced. 2. Marks obtained in the internal assessment and external assessment should be added up. 3. Central evaluation would be helpful in preparing the results quickly. 4. The practice of re-evaluation should be abolished. 5. The cases of unfair means should be decided at the time of the examination. 6. The roll numbers on the answer books should be changed into code numbers. 7. The members of the checking squads to check unfair means should not belong to the same college. 8. The use of unfair means should be made a cognizable offence. 9. The publication of guess papers and guide books should be banned.

On the basis of his findings the investigator has suggested that internal assessment should be done by the subject teachers concerned. The university should have its own press for printing of the question papers and other relevant papers. There should be correspondence courses for private candidates.

1010. MISRA, V.S., *A Study of the Effect of Randomization and Scaling on the Errors in Examination Marks*, Examination Research Unit, Gau. U., 1969

The main aim of the study was to find out the effect of

'randomization' and 'scaling' on the errors in examination marks.

The students of one college who appeared for English in the T.D.C. Part I examination of Gauhati University in the year 1965, old system, (N=85) and 1967, new system, (N=163) were selected. 'Randomization' and 'scaling' were employed in 1967, and not in 1965. Each of the selected students had two assessments—one by the college and the other by the university. The difference between the marks given to the students in the two assessments was taken as the error.

The major conclusions of the study were: 1. The randomization of scripts and scaling of marks as practised by Gauhati University reduced the errors in examination marks by 50 per cent approximately. 2. In the new system the chances of a fail-deserving student being passed, or a pass-deserving student being failed, were much less than the chances in the old system. 3. In the new system the agreement in the college and university assessments had increased by approximately 2.5 times. 4. There seemed no reason to believe that what was true for English was not true for other subjects.

1011. MISRA, V.S., *An Investigation into Admission Criteria*, Examination Research Unit, Gau. U., 1968

The main aims of the study were (i) to find out whether the different papers of a subject measured the same ability, (ii) to find out whether different subjects of an examination measured the same ability and in case there were more than one ability measured by an examination, what was the contribution of each ability to the variance of examination scores, and (iii) to determine the admission criteria for university courses.

The Matriculation Examination of Gauhati University held in 1963 was chosen and a sample of 1441 candidates was selected. In all, papers related to four subjects (English papers I, II and III, history, geography and mathematics) were selected. Every selected test was of the essay-type (three hours duration) and carried 100 marks. The coefficients of correlation between the scores of all the possible pairs of papers were calculated by the product-moment method. The values of 'r' were then converted into 'z's. Next, the factor analysis was carried out. In the absence of any exact criterion to sort out significant factors, the value of Tucker's Phi was calculated. Humphrey's test was also applied. The number of minimum variables required to determine the factors

was also found out. The obtained factors were then rotated by the orthogonal method.

The major findings and recommendations of the study were :- 1. The estimates of the population parameters of true non-overlapping variances among the three English papers ranged from 0.14 to 0.43. Each paper of English measured to some extent an ability which was not measured by the remaining two papers. 2. On the extraction of factors, three significant factors showed up, which were named 'verbal', 'problem solving', and 'memorization'. No 'quantitative ability' showed up in mathematics. Factors seemed to overlap among different subjects. Each subject appeared to have a different combination of factors. Assuming that the factorial structure remained, more or less, the same for each subject at higher stages also, it was recommended that for the graduate and postgraduate courses the marks obtained in the preceding examination in the subject which the candidate proposed to offer should be given the major weightage. 3. The study seemed to support the practice of admitting the students having the honours course in the subject. 4. For the selection to the P.U. Course it was suggested that a factor analysis study of H.S.L.C. and P.U. examination marks be made to discover the common factors. For the present it was suggested that for selection to the science stream, weightage should be given to the marks in mathematics and geography, and for selection to the arts stream, weightage should be given to the marks in English and history. 5. As regards the objectives of teaching, it seemed that emphasis was laid on the 'acquisition of knowledge' rather than on the 'application of knowledge to a new situation'

1012. MISRA, V.S., *Difficulty of Essay Questions*, Examination Research Unit, Gau. U., 1970

The major objective of the study was to investigate into the item indices of essay items.

A random sample consisting of 191 students of English I paper of the P.U. Examination of Gauhati University (Tezpur Zone, 1966) was drawn. Two methods were employed to find out the difficulty indices of questions. The first method was suggested by Harper and the second method was adopted by Gayen and others in their studies. The difficulty of each item was first estimated in terms of proportion and then it was transformed to Davis's linear scale with the help of Harper's Item Analysis Chart. Ideal and actual means for two sets of ques-

tions were estimated to study the provision of choice.

The study revealed that the two methods gave quite different results. The second method gave smaller values. Even the rank order of the questions in terms of difficulty was not the same for the two methods. Due to certain advantages the second method was recommended for the estimate of essay item indices. Supposed parallel items were not equal in difficulty, and the provision of choice sometimes penalized bright students and rewarded weak ones.

The significant educational implication is that there is a need for developing a pool of suitable essay items. The method used in the present study seems worth trying for the estimate of essay item indices. The evidence of the study is against the provision of choice.

1013. MISRA, V.S., *Discrimination Values of Essay-Type Questions*, Examination Unit, Gau. U., 1970

The main aim of the study was to investigate the discrimination indices of the English First Paper (revised course) of the Pre-University examination (1966) of Gauhati University.

The sample consisted of 157 students assumed to be a representative sample of the entire examinees of the Tezpur Zone. All these scripts were examined by one examiner. Every sub-question was considered a question. All the unanswered questions were treated as not reached. To make discrimination indices of essay questions comparable with those of objective questions, tetrachoric 'r's between the marks on the question and the total marks on the composite of the remaining questions were calculated. The values of tetrachoric 'r's were transformed to a linear scale with the help of the Item Analysis Chart prepared by Harper and others.

The study mainly revealed the following: 1. The discrimination indices of most questions of the English First Paper were satisfactory. 3. There was a need for introducing short answer-type questions in the examination. 3. The study did not support the practice of providing alternatives, as the alternatives were not equal in discrimination.

1014. MISRA, V.S., *Effect of Examiner Variability on Difficulty Indices of Essay Items*, Examination Research Unit, Gau. U., 1971

The major objective of the study was to investigate

whether for traditional essay type tests there could be some estimate of item difficulty independent of examiners.

In all 501 scripts of the English I paper (revised course) of the P.U. Examination of 1966 of Gauhati University (Tezpur Zone) examined by three examiners were analysed. The mark on the item that would contribute to the mean aggregate of the composite of the remaining items was considered the ideal mean. The proportion of students (p) who recorded more than the ideal mean was considered the difficulty index of the item. Another method used for the estimate of difficulty of an item was to compare the ideal and actual means for the item.

The study revealed that while the values of ' p ' changed significantly from examiner to examiner, an idea of the difficulty of an item could be had by comparing the ideal and actual means. The study did not support the practice of giving choices in the question paper.

- 1015.** NAGARAJU, C.S., M.D. USHA DEVI, *Multiple-point Entry Scheme—A Study of Private Candidates Appearing for SSLC Examination of Karnataka*, ISEC, 1983

The study mainly focused on the profile of private candidates, to know their study habits and problems, their motivations and aspirations. Private candidates are those who appear in the SSLC examination for the first time without being full-time students of any recognized high school during the academic year in which the examination is being conducted.

The study covered 5 per cent of private candidates appearing in the SSLC examination of Karnataka Secondary Education Examination Board during March – April 1982. Through a mailed questionnaire, about 550 private candidates (based on a 5 per cent sample selected randomly from the list of private candidates) were approached out of which 270 candidates responded.

The major findings of the study were: 1. An analysis of the socio-economic background of candidates revealed that parents of most of the candidates had at least school education. 2. Occupational backgrounds of the parents/guardians revealed that they were agriculturists or unskilled workers. 3. Male candidates outnumbered female candidates among the private candidates. 4. Candidates from urban and rural areas were evenly distributed. But rural representation was a little

less. 5. Most of the candidates belonged to the single-year age of 18 years. 6. The majority of them were high-school dropouts. 7. Slightly higher representation of the SC & ST category was observed in the study.

- 1016.** NEHRA, S.S., *Standardisation of Athletic Norms in Field Events for Boys in Haryana Secondary Schools (Age Group from 12 to 16)*, Ph.D., Phy. Edu., Kur. U., 1984

The objectives of the study were (i) to measure the existing level of performance in athletic field events of boys belonging to the age group 12 to 16 years, (ii) to establish norms in the field events (athletics) for boys of the age group 12 to 16 years, (iii) to find out the difference between the achievements of urban boys and rural boys in athletic field events, (iv) to determine what can reasonably be expected of the untrained boys in these events, (v) to discover individual and group deficiencies in certain events, and (vi) to compare the achievements of students in different events.

The sample of the study consisted of 2400 students (1200 from rural schools and 1200 from urban schools) with age ranging from 12 to 16 years. During the selection process 12 urban and 12 rural schools were selected from various districts of Haryana. Out of each school 100 students were selected. The selected subjects were divided into two groups, that is, subjects under the age of 15 formed one group (lower age group) and those 15 and above 15 years formed another group (upper age group). The performance of all the students of the lower age group was recorded on four field events, viz. shot put, discus throw, long jump, and high jump. The performance of the upper age group students was recorded for six field events, viz. shot put, discus throw, javelin throw, long jump, high jump, and hammer throw. Three trials were given to each student for each item and out of these the best effort was taken into account as a score. The data so collected were analysed with the help of percentile ranks and t-test.

The findings of the study were: 1. The mean performance of rural and urban boys for different events did not differ significantly in the case of the lower age group. 2. The mean performance of rural and urban boys did not differ significantly for different events in the case of the upper age group. 3. The distribution of scores on different events in the case of boys of the upper age group and lower age group was normal. 4. The performance on different events was better in upper age group boys than in

lower age group boys. 5. The lower limit of scores in shot put was 1.20 metres and the highest limit was 14.36 metres. 6. In discus throw the lower limit was 3.85 metres and the upper limit was 33.07 metres. 7. The minimum height crossed in high jump was 0.61 metres and the maximum height was 1.75 metres. 8. In long jump the shortest and longest jumps were 1.50 metres and 6.58 metres respectively. 9. In javelin throw the minimum and maximum distances covered were 5.20 metres and 39.36 metres respectively. 10. In case of hammer throw the minimum distance covered was 2.50 metres and the maximum distance covered was 60.25 metres.

- 1017.** PATEL, B.V. and VORA, I.A., *Construction and Standardization of Silent Reading Comprehension Tests in Gujarati for Pupils of Classes V, VI and VII to Study the Effect of Reading Improvement Programme*, M.B. Patel College of Education, SPU, 1985 (NCERT financed)

The objectives of the study were (i) to prepare reliable and valid tools for measuring reading comprehension in Gujarati for pupils of classes V, VI and VII separately, (ii) to establish norms for speed and reading comprehension in Gujarati for pupils of classes V, VI and VII, (iii) to prepare and provide programmes for developing reading speed and comprehension, (iv) to study sex differences with regard to reading speed and comprehension, and (v) to study whether there were any differences in reading speed and comprehension scores of pupils coming from different socio-economic status.

The investigators constructed and standardized reading comprehension tests in Gujarati which measured different components of reading comprehension—significant details of what is read, meaning of words and phrases, sequence of events or ideas, captions of the paragraphs read, and drawing generalizations. The tests were standardized over a sample of 500 students of each class, selected at random. The reliability of the tests was established by the K.R. formula, split-half method and analysis of variance approach. The reliability of the test for class V ranged between 0.75 and 0.83, for class VI between 0.82 and 0.86, and for class VII between 0.80 and 0.89. The concept validity, factorial validity, and concurrent validity were established. The reading improvement, and concurrent validity were established. The reading improvement programme (RIP) for each class was prepared for improving word recognition, word perception, vocabulary, word and phrase mean-

ing, sentence comprehension, column reading, and paragraph comprehension. The equivalent group design was used. The experiment was carried out on pupils of rural schools only. The schools were selected at random with 100 students in the experimental group and 100 in the control group for each class. The analysis of covariance was used for drawing the conclusions.

The major findings of the study were: 1. There were no sex differences with regard to reading comprehension of students of classes V, VI and VII. 2. There were no sex differences with regard to reading speed of students of classes V, VI and VII. 3. The experimental group students of classes V, VI and VII did better on reading comprehension tests after taking the reading improvement programme (RIP) than the control group students. The students of the experimental group of classes V, VI and VII showed better improvement in speed of reading after taking the RIP than those of control group. 5. SES was found to be an effective variable influencing the reading comprehension of pupils of class V. The students coming from a high SES level benefited more by the RIP than their counterparts from low SES families. 6. RIP treatment given to pupils of classes VI and VII proved more effective with students coming from a high SES level than with students coming from a low SES level. 7. The reading improvement programme treatment given to pupils of classes V, VI and VII proved more effective in increasing the rate of reading per minute with students coming from a high SES as compared to students coming from a low SES level.

The three reliable and valid tests to measure reading comprehension and rate of reading of students of classes V, VI and VII would be useful to the teachers working with classes V, VI and VII and would help them to know the actual level of their students in reading comprehension and rate of reading. The teachers who are concerned about improving the levels of achievement in school subjects by improving reading comprehension could use the reading improvement programme. If the teachers are rightly made aware of such programmes and available material, they can help their students in acquiring basic skills of reading comprehension which are very useful in self-study or auto-learning.

- 1018.** PATEL, M.J., *A Comparative Study of Rate of Reading and Comprehension of Students of Standard VIII in Different School Subjects*, Ph.D. Edu., SPU, 1986

The objectives of the study were (i) to determine the

rate of reading of students of standard VIII in various school subjects, (ii) to determine the level of comprehension of the students in different subjects, (iii) to compare the rate of reading of students in different school subjects, (iv) to compare the level of comprehension of students in different school subjects, (v) to find out the relationship between the rate of reading and the level of comprehension of the students in different school subjects, (vi) to find out the relationship between rate of reading in different school subjects and I.Q., (vii) to find out the relationship between rate of reading in different school subjects and SES, (viii) to find out the relationship between comprehension of different school subjects and I.Q., (ix) to find out the relationship between comprehension of different school subjects and SES, (x) to study the effect of sex differences on the rate of reading, (xi) to study the effect of sex difference on the level of comprehension of students and (xii) to study the relationship between rate of reading and comprehension after partialling out the effects of sex, I.Q. and SES.

The Intelligence Test standardized by Desai and Bhatt, and the SES scale constructed by Pathak and Patel were used as tools for collecting the data. The rate of reading and comprehension were measured through paragraphs selected from mathematics, science, Gujarati, English and History. The data were collected from 400 students consisting of 200 boys and 200 girls selected from the schools of Cambay city. The 2x2x2 factorial design was used for analysing the data. At the same time t-test and correlational techniques were also used to study the significance of difference between means and relationship between the scores of different variables respectively.

The major findings of the study were: 1. Girls were superior to boys in their rate of reading in mathematics. 2. Sex and I.Q. both affected the rate of reading in mathematics. 3. The interaction effect of SES and I.Q. was not significant. 4. The girls with high SES and high I.Q. had a higher mean rate of reading in mathematics than other groups, except girls with low SES and high I.Q. 5. The girls with low SES and high I.Q. had a higher mean rate of reading in mathematics than the groups of boys with high SES and high I.Q., boys with high SES and low I.Q. and boys with low SES and low I.Q. 6. Boys and girls did not differ significantly in reading comprehension in mathematics. 7. The interaction effects of sex and I.Q., and sex and SES were not significant. 8. The girls with high SES and high I.Q. were found superior in reading comprehension in mathematics to all other groups. 9. The coefficients of correlation of rate of read-

ing and comprehension in mathematics for boys and girls were 0.239 and 0.154 respectively. 10. Girls were found to be superior in rate of reading in science, Gujarati, English and history, to boys. 11. Boys were superior in reading comprehension in science to girls. 12. Sex, SES and I.Q. affected significantly the rate of reading in science. 13. Comprehension in science was dependent on SES and I.Q. and not on sex. The correlation between the rate of reading and comprehension in science was 0.308. 14. Sex and I.Q. both affected the rate of reading in history. 15. The interaction effect between sex, SES and I.Q. was not significant. 16. The girls with high SES and high I.Q. had a higher mean rate of reading in history than other groups, except girls with low SES and high I.Q. and girls with high SES and low I.Q. 17. Boys and girls were found equal in comprehension in history. 18. The correlation between rate of reading and comprehension in history was found to be 0.289. 19. The correlations between rate of reading in history and SES and I.Q. were significant. 20. The correlation between comprehension in history and SES was not significant. The interaction effect of SES, sex and I.Q. was not significant in the case of rate of reading in Gujarati. 21. The girls with high SES and high I.Q. had a higher mean rate of reading in Gujarati than other groups. Boys and girls did not differ significantly in reading comprehension in Gujarati. 22. The interaction effect of all the three variables was not significant. 23. The correlation between the rate of reading and comprehension in Gujarati was found to be 0.128. The correlation of rate of reading in Gujarati with I.Q. was significant at 0.01 level. The correlation between comprehension in Gujarati and I.Q. was not significant. 24. The interaction effect of sex, SES and I.Q. was not significant in the case of rate of reading in English. 25. Boys and girls were equal in case of reading comprehension in English. 26. The I.Q. and SES correlated significantly with comprehension in English. 27. The interaction effect of sex, SES and I.Q. was significant in the case of comprehension in English. 28. The correlation between rate of reading and comprehension in English was significant. The correlations of rate of reading in English with SES and I.Q. were also significant. The correlations between comprehension in English and SES and I.Q. were significant.

1019. POPLI, L.R., *Causes of Mass Failure in Matric/Hr. Sec. Part-I, II Examination-1973*, SIE, Haryana, 1974

The main objective of the study was to find out the cause

es of low performance of the schools in the Board (of School Education) Examination for the year 1973.

The study was confined to only higher secondary schools. A questionnaire was sent to all seventy higher secondary schools which had shown a pass percentage less than the state average. Out of these, forty schools responded. In addition to this, five schools near Gurgaon were also visited for personal discussion with the members of the staff about the causes of their low performance. A questionnaire for obtaining the responses about the causes of low performance was developed. The responses were tabulated and analysed.

The main causes of mass failures were as follows: 1. Change in the system of promotion from Class VIII to Class IX which made it obligatory to pass in English, mathematics and Hindi, was a major cause. 2. Lenient Promotion rules for Class IX was the second cause. 3. Some of the questions set in the Board Examination were beyond the comprehension of average students. 4. A few schools reported that they had shortage of permanent staff and in some schools there even were no permanent heads. 5. In some schools non-availability in time of textbooks was also a factor. 6. Abolition of internal assessment also contributed to large scale failure. 7. Lack of proper guidance and counselling at the higher secondary school level, the government teachers' strike from November 1972 to March 1973, emergency measures adopted for getting the answer books checked, and the introduction of a centralized system of marking were other causes.

1020. SANGRAL, M.S., *Construction and Standardization of Skill Test For Hockey*, Ph.D. Phy. Edu., Punjabi U., 1986

The objectives of the study were (i) to identify major fundamental skills and skill combinations in hockey, (ii) to construct and standardize a battery of skill tests purely on objective lines, (iii) to prepare norms in each skill for players at different levels—intercollege, inter-university, combined university, state and national levels, and (iv) to present a standardized procedure for grading hockey players.

The sample for test construction consisted of 63 male experienced hockey players. The sample included 17 intercollege players, 15 interuniversity players, nine combined university players, 19 state-level players, and three national level players. However, for working out norms, the sample consisted of 500 male hockey players

who participated at different levels in the year 1984-85. The skills and skill combinations selected for playing hockey were 25 yard goal shooting, dribble and push, scoop for accuracy, dribble and hit, and self hitting and stopping. The test was constructed on the basis of items linked with these five skills. The reliability of the test was established against each item with the test-retest method. The reliability coefficient varied from 0.78 to 0.91. The validity for each of the test items was established against the rating by judges about playing ability. The validity coefficient varied from 0.40 for dribble to 0.77 for self-hit and stop. On factor analysis of test scores, self hit and stop, scoop for accuracy, and dribble and push came out as independent factors. The multiple regression equation showed that the factors—self hit and stop, scoop for accuracy, and dribble and push, measured total hockey playing ability sufficiently. So the test battery was formed of three tests, namely dribble and push, scoop for accuracy, self hit, and stop. After factor analysis the test battery had test-retest reliability coefficient 0.91, and validity against judgement by judges was 0.81. The norms were established for intercollege, interuniversity, combined university, state, and national level players separately.

The study showed the following: 1. The mean difference in performance in dribble and push amongst three groups—intercollege, interuniversity and state level players, was not significant but the mean differences were significant in case of intercollege, interuniversity and national level players. 2. In case of scoop for accuracy, the national level group was significantly better in performance in comparison to intercollege, interuniversity, combined university and state level groups. 3. The national level group was found to be much superior on the self hit and stop test than any other group.

1021. SINGH, P.M., *Physical Fitness Norms of Punjab State High School Boys*, Ph.D. Phy. Edu., Pan. U., 1986

The objectives of the study were (i) to establish norms for the various components of physical fitness of Punjab State high school boys (age group 12 to 15 years), (ii) to compare the physical fitness of school boys according to age groups on various components of fitness, (iii) to compare the physical fitness of school boys belonging to rural and urban schools, and (iv) to compare the physical fitness of school boys belonging to different socio-economic groups on various components of fitness.

A sample of 5000 students was randomly selected from high school boys of Punjab State. The study employed three-way factorial design (3x2x3) with unequal replications. The three factors were age having three levels—young (12–13), middle (13–14), and old (14–15), area having two levels—rural and urban, and socio-economic status (SES) having three levels—low, middle and high. The tools employed in the study were: (i) Standing Broad Jump for explosive muscle power, (ii) Illinois Agility Run for change of direction and agility, (iii) Sit and Reach for flexibility, (iv) Sit Ups Bent Knees for abdominal muscular endurance, (v) 50 Metre Dash for speed, (vi) Push Ups Chair for muscular endurance of arms, (vii) Cricket Ball Throw for strength and coordination, (viii) 600 Metre Run-Walk for cardio-vascular endurance. Percentile norms were established for all fitness components, viz. agility, flexibility, speed, muscular power, muscular strength, coordination, muscular endurance and cardio-vascular endurance.

The findings of the study were: 1. Percentile norms for physical fitness components were suitable to assess the physical fitness level of Punjab State high school boys in the age range 12–15 years. 2. There was a significant relationship between age and performance of subjects in physical fitness tests. 3. The performance of older boys was better than that of younger boys. 4. Urban subjects performed significantly better than rural subjects in standing broad jump, agility run, 50 metre dash and cricket ball throw. 5. Rural subjects performed significantly better than urban subjects in sit and reach and 600 metre run-walk tests. 6. There was no significant difference between the three SES groups in respect of standing broad jump, sit and reach test, 50 metre dash, chair push ups and cricket ball throw tests. 7. There was a significant interaction effect of age and area for physical fitness test items except the sit ups bent knees test. 8. The interaction effect of age and SES was not significant for physical fitness test items except in the case of the sit ups bent knees test. 9. The interaction effects of age × area × SES were not significant for physical fitness test items except for the agility run test.

1022. SINGHA, H.S., *The Pathology of Public Examinations in India*, Ph.D. Edu., JNU, 1983

The objectives of the enquiry were (i) to study the effect of the present system of public examinations on schooling, education and society at large, with a special focus

on the pathology or deviations from pedagogical, economic, political and/or social propriety, and (ii) to examine the relevance of public examinations with special reference to India in order to bring out the contradictions, paradoxes and dilemmas produced by them.

The study was planned at the macro level, adopting the qualitative method in the form of the historico-philosophical approach. Thus the methodology depended heavily on the logic of extreme cases, on the citing of key examples and on the opinions of outstanding authorities.

The major findings of the study were: 1. An overview of literature showed that the externalities, i.e. social, political and economic institutions, had been neglected by thinkers and researchers. 2. The indigenous examinations were informal, unidimensional and individualized. However, the British as colonizers imposed examinations which were formal, rigid and impersonal for serving their purpose, disregarding the socio-economic conditions in India. 3. These examinations adversely affected education by curbing the teacher's initiative, by stereotyping the curriculum, by promoting mechanical and lifeless methods of teaching, by discouraging all spirit of experimentation, and by encouraging a concern with trivialities in education. 4. Examinations operated as a selecting process for jobs in various sectors. 5. Examinations became the tools of dehumanizing, deindividualizing and depersonalizing man. 6. On sociological grounds, they have necessitated the adoption of the cult of testocracy, thereby creating a triangular contest between three competing principles namely testocracy, democracy and hereditary stratification. 7. In spite of their irrelevance, examination would continue to flourish, although in a substantially changed condition, and would possibly be supplemented by alternative strategies by 2001 A.D.

1023. SINHA, S.K., *A Study of Attitudes towards the Present System of Examination*, Ph.D. Psy., Ran. U., 1977

The main aims of the inquiry were to study (a) the attitudes of students, their teachers, and their guardians towards the academic, evaluative and administrative aspects of the existing system of university examination, (b) their attitudes towards different measures of improving the examination system, (c) the relevance of some important factors like sex, faculty (subject of study), degree of education, achievement level (previ-

ous examination results) and ethnicity which were supposed to influence the attitudes of students towards the present examination system.

In all, 560 Ranchi University students were selected by adopting proportionate stratified random sampling design, and their teachers (N=90) and guardians (N=30) were selected randomly. The student sample was classified into five sub-samples on the basis of faculty, degree of education, sex, achievement level, and ethnicity (matched groups), respectively. A Likert type questionnaire was constructed and used. Kelley's technique was used for item selection. Theme-wise and item-wise analysis was done. Analysis based upon percentage of agreement with each item on suggestion relating to improvement in academic, evaluation and administrative fields, respectively, was done. Percentage median, chi-square, etc. were used.

Some of the major conclusions of the study were: 1. The existing examination system had both merits and demerits. 2. According to students, teachers, and guardians it had more demerits than merits. 3. A very high percentage of them had agreed with various suggestions for improvement in academic, administrative and evaluative aspects of the existing examination system. 4. Most students, teachers and guardians were dissatisfied with the existing examination system and wanted immediate and adequate improvement in it.

*1024. SRIVASTAVA, A., *Construction of an Achievement Test for Class VIII Students*, Ph.D. Psy., RSU, 1981

The purpose of this study was to construct a general achievement test in Hindi based on the syllabi for class VIII prescribed by the M.P. Madhyamik Shiksha Mandal.

The items for the test were taken from all the areas of study, i.e. general science, mathematics, social studies and languages. In all, 652 items were framed from the course contents of which half were of the true-false type and half were of the multiple choice type. These items were circulated to five teachers, who had a minimum of five years of teaching experience to teach class VIII, to determine their content validity. This list of items was then administered to 500 students of class VIII of Chhattisgarh region, taking care to ensure representation of sex, type of management of the school, location, etc. The item analysis was carried out taking 27 per cent of the students who were top scorers and 27 per cent who were low scorers. This was done primarily to esti-

mate the difficulty level and the discrimination power of the items. Of the 652 items, 120 items met the joint criteria of difficulty level 0.40 to 0.60 and discrimination power of above 0.40. These 120 items were thus retained for the final test and were presented to 300 students for the fixation of time limits. The test was administered to a final sample of 5575 students of grade VIII of Chhattisgarh region of Madhya Pradesh. The sample was selected giving full representation to sex, location and management of schools. The reliability of the test was determined by using Test-retest Split-half and Rational Equivalence methods, which was found to be 0.961, 0.94, and 0.96 respectively. The validity of the test was worked out by correlating the test scores with examination scores (Concurrent validity). This was found to be 0.88. This was also done with the help of the teacher's judgement. The teacher's judgement of the pupils obtained a coefficient of 0.78 with the test scores. The construct validity of the test was ascertained with the help of inter-correlation of items, which yielded a significant coefficient of correlation. The objectivity of the test was maintained by keeping objectivity in administration and scoring.

1025. TAYLOR, H.J. and TLUANGA, L.N., *Scaling Tables for Examination Marks*, Examination Research Unit, Gau. U., 1963

The main aim of the study was to construct scaling tables for adjusting the marks of the various examiners to a common standard.

The scaling tables were constructed by adopting a scientific procedure in such a way that in any transformation the marks 100 and zero remained unchanged, and marks between these limits and the median changed proportionately. Fractions were rounded off to the nearest whole number, and exact halves to the next higher number.

The study revealed that the scaling tables provided a sufficient and satisfactory answer to the problem of mark adjustment. These tables were useful chiefly for scaling sets of marks relating to the same question paper for which variations in the standard deviation were usually small.

1026. TAYLOR, H.J., and TLUANGA, L.N., *The Persistence Effect in Marking*, Examination Research Unit, Gau. U., 1965

The main purpose of the study was to present evidence

showing that the persistence effect was not only detectable but often quite large, and that it affected the marking of a surprisingly high proportion of examiners.

The marksheets of 112 examiners were selected covering English, mathematics, history, geography and Assamese. The marksheets used were a sample drawn from the Gauhati Matriculation Examination of 1963. In each examiner's set of scripts the order of merit of the candidates was known to be random. A preliminary test was first carried out. Then the Mean Difference Test and the Serial Correlation Test were applied.

The major findings of the study were: 1. All the tests clearly indicated the presence of persistence. 2. The persistence effect appeared definitely and strongly in about one-third of all the marksheets tested, and was probably present in at least half of them. 3. No significant difference was found between one subject and another. The effect was almost always positive. 4. High serial correlations appeared in at least one-third of the marksheets. These correlations were interpreted in terms of the persistence effect. The correlations might also be due, in part, to fluctuations in the examiner's standard of marking. For one examiner in three, the average error due to these causes was probably about ± 4 per cent and occasional errors might well reach ± 10 per cent.

1027. TAYLOR, H.J., TLUANGA, L.N. and MISRA, V.S., *A Study in Multiple Marking*, Examination Research Unit, Gau. U., 1966

The chief objective of the study was to find out how reliable essay-type examinations actually were.

The scripts used were a sample of actual university scripts submitted by 6635 candidates at the B.A. Part I examination of Gauhati University in 1964, English Paper I was chosen. The distribution of the original marks was first examined and a sample of 100 scripts was finally selected by adopting the stratified random sampling technique. These scripts were duplicated and marked independently by 19 highly qualified and independent examiners. The average marks of a script was taken as the best approximation to its 'true marks', and a study was made of the errors with respect to the true marks. Total range, mean marks, total variance, standard deviation, pass percentage, mean error, error variance, standard error of estimation, non-error variance, standard spread, coefficient of reliability, average correlation coefficient and correlation coefficient with true marks were computed from the marksheets. Ranking by

examiners, correlation between examiners, relation between different measures of performance, comparison with the original results, the question of unanimity, unanimity of ranking and order of merit in relation to several examiners were studied.

The study mainly revealed that the mean standard error of marking was 4.5, and the average error was almost independent of the merit of the script. The examiners showed a high degree of concordance, nearly half of the inter-correlation coefficients were 0.80 or higher. The degree of unanimity amongst the examiners in regard to passing a script, or ranking two scripts, was discussed as a function of the true marks, and criteria for passing or ranking with 95 per cent confidence were developed.

The significant educational implication is that essay-type examining, when properly conducted and analysed, can give more reliable results than is commonly supposed.

1028. TAYLOR, H.J., TLUANGA, L.N., and MISRA, V.S., *Examinations as Predictors*, Examination Research Unit, Gau. U., 1966

The main aim of the study was to find out to what extent a candidate's performance in one examination indicated his probable performance in the next.

The examinations selected for the study were the 1963 Matriculation Examination of Gauhati University and the subsequent 1964 Pre-university Examination. In both examinations the marks of examiners were scaled to common norms before tabulation. The study included a random sample of 347 P.U. candidates (233 arts and 114 science). The Matriculation results of a random sample of 1441 candidates were also analysed for obtaining some statistics of the Matriculation marks distribution. The regression line was determined and the correlation coefficient between the marks of candidates in the two successive examinations was computed.

The major findings of the study were: 1. The true correlation coefficient between the marks of candidates in two successive examinations (Matriculation 1963 and Pre-university 1964) was 0.72. 2. Of 16,542 Matriculation failures, some 1600 would actually have passed the P.U. if they had been allowed to appear. Of these 1600, 39 would have gained 2nd division in the P.U. Examination. 3. Of the candidates who gained 2nd division marks (45 per cent) in the Matriculation Examination, nearly one-fifth failed in the P.U. Examination.

A significant educational implication of the study is that the universities can maintain cumulative records not only for two consecutive examinations but also for longer periods so that the academic history of a group of candidates could be ascertained without laborious special investigation and follow-through studies could be made to highlight the efficiency and validity of the examinations.

1029. TAYLOR, H.J., TLUANGA, L.N., and MISRA, V.S., *The Inconstant Examiner*, Examination Research Unit, Gau. U., 1966

The main aim of the study was to study the extent to which an examiner's marking standard varied as he worked through the scripts.

The analysis was based on a body of material extending over 28,000 examination scripts, 112 examiners, eight major papers and five different subjects. Diurnal and long range fluctuations in marking standard were studied. Homogeneity of means and variances were tested. Fluctuations of standard and accuracy were found out. The influence of 'persistence effect' on the examiners was also studied.

The major findings of the study were: 1. Nearly half the examiners were subject to large fluctuations of standard in the course of their marking. 2. One-third of the examiners were subject to large fluctuations of accuracy. 3. One-third were subject to the 'persistence effect', by which the impression formed on one script was carried over to the next. 4. These three tendencies occurred independently, and three examiners out of four displayed one or more of them. 5. There were indications of a diurnal fluctuation in marking, and clear evidence of occasional periods when an examiner's judgement effectively ceased to operate.

The significant educational implication of the study is that the large errors which are inherent in conventional marking can be reduced to some extent by using scaling procedures before combining marks, and such procedures should be used wherever applicable. But certain types of errors are not removed by scaling, including those which arise from fluctuations in the standard or accuracy of marking. Marking of a single examiner is not an absolute measure which provides a reliable estimate of a candidate's attainment.

1030. TEWARI, L.M., *Evaluation and its Problems in Upper Primary Schools*, SIERT, Rajasthan, 1975

The objectives of the study were (i) to find out the strengths and weaknesses of the examination papers set by teachers of upper primary schools, and (ii) to find out problems of teachers in the process of evaluation at the upper primary level.

For analysis, the question papers set for upper primary schools of Udaipur and Dungarpur districts were selected. Seventy-six question papers from 30 schools of Udaipur and Dungarpur districts were analysed. The basis of the study was 210 sets of test papers of Hindi, mathematics, social studies and general science and 46 question papers of half yearly and annual examinations. Eighteen teachers were interviewed to find out the difficulties regarding evaluation. The questions were analysed on the basis of the objectives in setting the examination paper. The tools used were an observation schedule, an interview schedule and a questionnaire.

The study revealed the following: 1. More questions were based on the knowledge aspect. The percentage of questions on understanding and skill was very low. 2. In classwork, essay type questions were mainly used. In tests, half yearly and annual examinations the position was the reverse. 3. In multiple choice questions, most of the distractors did not function well but the stems were alright on the whole. 4. The teachers did not know anything about the preparation of the blue print. The headmaster did not give them general instructions. The teachers did not have mastery over their subjects. 5. About 55.5 per cent of paper setters did not find it difficult to frame multiple choice type questions and 19 per cent felt it was not easy to frame essay type questions. 6. Teachers were ignorant about the objectives of unit-wise teaching. 7. Teachers were not trained in evaluation techniques. 8. About 43.7 per cent teachers were of the opinion that there was no need to set a word-limit for the answers.

1031. TIWARI, C., *A Comparative Study of Trends of Achievement Measurement in Civics in Higher Secondary in different Boards of Secondary Education*, Ph. D., Edu., Raj. U., 1982

The objectives of the study were (i) to compare the objectives of teaching civics with respect to evaluation, in

different boards of secondary education, (ii) to compare the curriculum of civics, (iii) to compare the differential capacity and difficulty levels of question papers in civics, and (iv) to compare the reliability and validity of question papers in civics of different boards of secondary education.

In all five boards of secondary education were taken. These were the Rajasthan Board of Secondary Education, Central Board of Secondary Education, Delhi Board of Secondary Education, Madhya Pradesh Board of Secondary Education and Haryana Board of Secondary Education. The syllabi and question papers set in the previous five years (1973–1977) in these boards in civics for part A and part B of the class XI curriculum formed the sample of the study. Apart from this, 300 students from each board who attained first, second and third division in the examinations conducted by these boards were also taken as the sample of the study. In this way the answer books of 1500 students in all, who appeared in the civics examination conducted by these boards were taken for analysis. These were analysed through checklists for cognitive domain with respect to knowledge, comprehension and application objectives.

The findings of the study were: 1. The analysis of the question papers of five years revealed that these contained more questions pertaining to knowledge level than comprehension level. In the Rajasthan and Delhi Boards, knowledge questions were more than the comprehension questions in the first year but in the later years the number of comprehension questions went on increasing. But in the M.P. Board, the comprehension questions were more than the knowledge questions. In the Haryana Board there had been a balance between knowledge and comprehension questions. The same was the case with the Central Board of Secondary Education. But in all these five boards, questions regarding application level were non-existent. 2. Regarding the representation of various chapters of the syllabus in the question papers, all units prescribed in the syllabi of the five boards had not been given full representation in the question papers. In the Rajasthan Board the representation was 52 per cent, in the Delhi Board, it was 64 per cent, and in the other boards it was less than 50 per cent. The same was the case with question paper II of the syllabus. The representation ranged from 38 per cent to 71 per cent in the various boards. 3. Almost all question papers set by the boards of secondary education had a lower level of differential capacity. These question papers were not able to differentiate

between the specific ability and higher ability levels. All the question papers could differentiate the success level of students up to 33 per cent. 4. From the internal analysis of question papers of the five boards it was found that there was a significant difference in the difficulty level of questions of different boards. There were only a few questions in the question papers of different boards which had 50 per cent difficulty level. 5. The ideal score percentage of all these boards was also different. It was 82.31 per cent for the Rajasthan Board, 96 per cent for the Central Board, 95 per cent for the Delhi Board, 87 per cent for the M.P. Board and 100 per cent for the Haryana Board of Secondary Education. 6. The reliability coefficient of question papers of all these boards was also different. It was 0.75, 0.88, 0.75, 0.46 and 0.52 respectively for the Rajasthan, Central, Delhi, M.P. and Haryana Boards of Secondary Education. 7. The error variance analysis of the question papers showed that there was a significant difference in categorization of the different boards. On this account the students of different boards placed in different categories could not be taken as equal in achievement. 8. The distinction level of these boards was also different.

1032. TLUANGA, L.N. and TAYLOR, H.J., *The Influence of Scaling on Examination Results*, Examination Research Unit, Gau. U., 1964

The main aim of the study was to find out the influence of scaling on examination results.

A sample survey was carried out. The Gauhati University Matriculation Examination of 1963 was carried through with the help of new techniques including (1) randomization of scripts, (2) marksheet scaling, (3) adjustment of standard between subjects, and (4) borderline adjustment. The total number of candidates involved (excluding supplementary and other odd cases) was 33,571. A little over 4 per cent (N=1441) candidates were randomly selected. Mark distribution in single subjects, and the effect of scaling in single subjects as well as on the final examination result were studied. A detailed analysis was carried out and the influence of the new technique on the examination results was examined.

The study mainly revealed the following: 1. The errors which had arisen from variations in the standard of marking were by no means negligible and their removal made a striking difference

the examination. 2. The scaling increased the pass percentage and produced significant changes in the classification of the candidates and the order of merit.

The significant educational implication of the study is that scaling techniques should be introduced wherever circumstances permit. The new techniques have been shown to be practicable in a major examination covering more than 33,000 candidates and the whole work was completed several weeks earlier than in previous years. There is no technical difficulty in using the same methods for other examinations. Scaled marks, whatever its intrinsic uncertainty, is clearly a more just estimate of a candidate's performance than raw marks, and results which are based on the scaled marks have a higher degree of validity than those based on the raw marks.

1033. TRIVEDI, J.R., *Construction and Standardization of Reading Achievement Test in Hindi for Pupils of Class VIII of Secondary Schools of Gujarat State*, Ph. D. Edu., SPU, 1984

The objectives of the study were (i) to provide schools with a valid and reliable tool for measuring reading achievement in Hindi of pupils studying in Std. VIII, (ii) to establish norms of reading achievement in Hindi for pupils of Std. VIII, (iii) to study whether there were any sex differences with regard to reading achievement in Hindi, and (iv) to study reading achievement in Hindi of pupils coming from rural and urban areas.

The usual method of constructing and standardizing a test was followed by the investigator. The test purported to measure the following components of reading comprehension: (i) ability to note the significant details of what is read, (ii) ability to give the meaning of words, phrases, etc., (iii) ability to give the sequence of events or ideas, (iv) ability to draw generalizations and to give captions to the paragraphs, and (v) ability to find out the relationship between ideas. These components were measured through nine different sub-tests. The first test aimed at measuring the speed of reading. The usual method of item analysis was adopted for the purpose of selecting the items for the inclusion in the final form of the test. For the final form, 90 items out of 159 items were selected on the basis of discriminative index and difficulty value. The total time required to administer the test was 65 minutes. The final form was administered to 1800 students selected at random from 70 high schools of Gujarat.

boys and 598 were girls. The percentile norms and standard score norms were established along with letter grade norms. The reliability and validity of the tests were established by different methods.

Some of the findings and some characteristics of the test were as follows: 1. There was no significant difference between the mean performance of boys and girls from urban areas. Similarly the difference between the mean performance of boys and girls from rural areas was not significant. 2. There was no significant difference between the mean performance of urban and rural students. 3. The average difficulty index of the test was found to be 51.6. 4. The reliability of the test, as established by the test-retest method, split-half method, K.R. formula and Rulon method, was found to lie between 0.92 and 0.98. 5. The concept validity, concurrent validity and predictive validity of the test were established. The factorial validity revealed that there was only one factor, namely, reading comprehension, consisting of ability to grasp the significant details, ability to give word meaning, and ability to give sequence of events and draw generalizations.

ALSO SEE

811. ANSARI, A.M., *Construction and Standardization of Achievement Tests in General Science for Standards V, VI and VII for Children studying through Hindi as the Medium of Instruction in Greater Bombay*, Ph. D. Edu., Bom. U., 1984
924. DEKA, U., *School Failure: A Causal Comparative Study of High and Low Achievers*, Ph. D. Edu., Gau. U., 1985
798. MONDKAR, S.M., *Factor Analysis of some Tests in Number Systems in the Mathematics Syllabus prescribed from June 1972 for Standard VIII in the State of Maharashtra*, Ph. D. Edu., Bom. U., 1984
1282. SEN, A.J., PAL, M.N., and GHOSH, S.K., *Calcutta University Examination System — A Management Study*, Indian Institute of Management, Calcutta, 1981
1197. SINGH, H., *Effective Teaching Strategies used for preparing for the Examination as perceived by the Students*, Ph. D. Edu., Mee. U., 1986
1436. SRIVASTAVA, S., *Cognitively Oriented Pre-School Programme for Children : Formative and Summative Evaluation*, Dept. of Child Development, J.B.A.S. Women's College, Madras, 1987.

LDD, NCERT



F22794

1202 were

National Institute of Education
Division of Library, Documentation
& Information (N.C.E.R.T.)
Acc. No. [unclear]
Date [unclear]