

# SUPREME COURT OF INDIA

Sanjay Singh

Vs.

U.P. Public Service Commission, Allahabad

W.P.(Civil)No.172,409,466 and 467 of 2005.

(Y.K.Sabharwal and C.K.Thakker and R.V.Raveendran,JJ.)

09.01.2007

## JUDGMENT

### **Raveendran,J.**

1. These petitions under Article 32 of the Constitution of India have been filed by the unsuccessful candidates who appeared in the examinations conducted by the Uttar Pradesh Public Service Commission ('Commission for short) for recruitment to the posts of Civil Judge (Junior Division).

2. On the request of the Allahabad High Court, to conduct the examination for filling 347 posts of Civil Judge (Junior Division), the Commission issued an advertisement in the Employment News dated 28.11.2003. As many as 51524 candidates appeared for the "U.P. Judicial Service Civil Judge, (Junior Division) Preliminary Examination, 2003" conducted by the Commission on 21.3.2004. The preliminary examination was of 'objective' type consisting of two papers General Knowledge and Law. The result was declared on 30.6.2004 and 6046 candidates were declared qualified to appear for the "U.P. Civil Judge (Junior Division) Examination (Main), 2003" which was of 'descriptive' (conventional) type. The Main examination consisted of five papers (each carrying 200 marks) - General Knowledge, Language, Law I, II and III - and was held between 5th and 7th October, 2004. The number of candidates who took the said examination was 5748.

3. The answer scripts relating to each subject were distributed to several examiners for valuation, as it was not possible to get the large number evaluated by a single examiner. The number of examiners, to whom the answer-scripts were distributed for valuation, were as follows: General Knowledge 18, Language 14, Law-I 11, Law-II 10, and Law-III 14. The marks assigned by the examiners were subjected to 'statistical scaling' and the results of written examination based on such scaled marks, were declared on 7.3.2005. Thereafter, 1290 candidates were interviewed between 14.4.2005 and 26.4.2005. After such interview, the Commission declared the final results of the examination on 1.5.2005 based on the aggregate of 'scaled marks' in the written (Main) examination and the marks awarded in the

interview. On the recommendations made by Commission, appointments were made to 347 posts of Civil Judge, Junior Division.

4. The petitioners, who were unsuccessful, are aggrieved. They contend that the statistical scaling system adopted by the Commission is illegal as it is contrary to the Uttar Pradesh Judicial Service Rules, 2001. They also contend that conversion of their raw marks into scaled marks, is illegal as it was done by applying an arbitrary, irrational and inappropriate scaling formula. It is submitted that the Commission's exercise of subjecting the marks secured by the candidates to scaling, has resulted in meritorious students being ignored, and less meritorious students being awarded higher marks and selected, thereby violating the fundamental rights of the candidates.

4.1) W.P. [C] No.165/2005 was filed on 5.4.2005 even before the final results were declared, praying (i) for a direction to the Commission not to adopt the system of scaling and to declare the results of the Main Examination on the basis of actual marks obtained by the candidates; and (ii) for a direction that the petition be heard by a Bench of three or more Judges as the decision of a Bench of two Judges of this Court in *U.P. Public Service Commission v. Subhash Chandra Dixit*<sup>1</sup> upholding the system of scaling adopted by the Commission does not lay down the correct law.

4.2) The other petitions were filed after declaration of the final results, in effect, for the following reliefs : (a) for quashing the results of the U.P. Civil Judge (Junior Division) Main Examination-2003 declared on 7.3.2005 and the final results declared on 1.5.2005 on the basis of scaled marks and direct the Commission to declare the results on the basis of actual marks secured by the candidates; (b) to direct an inquiry by an independent agency into the irregularities committed by the Commission in the said examination; (c) for a declaration that the use of 'statistical scaling' in regard to the examinations for the subordinate judiciary is unconstitutional; and (d) to reconsider the law laid down in *Subhash Chandra Dixit* (supra).

5. The respondents raised the threshold bar of maintainability. It is submitted that this Court in *S. C. Dixit* (supra), has rejected identical grounds of attack and upheld the statistical scaling method adopted by the Commission in the examination conducted in 2000. It is contended that the prayers in these petitions under Article 32, in effect, seek setting aside or review of the decision in *S. C. Dixit*, and that is impermissible. Reliance is placed on the Constitution Bench decision of this Court in *Rupa Ashok Hurra v. Ashok Hurra*<sup>2</sup> to contend that a writ petition under Article 32 would not lie to challenge any judgment of this Court or that of a High Court, as superior courts are not 'State' within the meaning of Article 12 and their judgments cannot be termed as violative of fundamental rights. It is also pointed out that Review Petition (Civil) No. 162/2004 and Curative Petition No.43/2004 filed in respect of *S. C. Dixit* (supra) were rejected on 04.2.2004 and 6.10.2004 respectively.

6. In regard to merits, the Commission contended that the 'statistical scaling' method adopted in regard to Civil Judge (Junior Division) Examination is legal, scientific and sound and its policy to apply statistical scaling to marks of written examination, was based on experts'

opinion as also the experience gained in conducting several examinations. It is submitted that under the proviso to Rule 50 of the U.P. Public Service Commission (Procedure and Conduct of Business) Rules, 1976, it is entitled to adopt any formula or method or device to eliminate variation in marks; that it found variation in the marks awarded by different examiners on account of a phenomenon known as 'examiner variability' and to eliminate it, statistical scaling was introduced. It is further submitted that matters relating to the conduct of Examination, evaluation of answer-scripts, application of methods to bring in uniformity in evaluation are matters of policy involving technical and scientific decisions based on expert opinion; that courts are not equipped to pronounce upon such matters and, therefore, should not interfere in the absence of manifest arbitrariness or mala fides; and that, at all events, in the absence of an opinion by a body of experts in the field of statistics certifying that the system of scaling adopted by the Commission is unsound and irrational, there should be no interference. Lastly, it is submitted that if the court, for any reason, should hold that the existing scaling system should be substituted, that should be done prospectively.

7. On the contentions urged, the following questions arise for our consideration :

(i) Whether the writ petitions are not maintainable ? (ii) Whether 'scaling' of marks is contrary to or prohibited by the relevant rules ? (iii) Whether the 'scaling system' adopted by the Commission is arbitrary and irrational, and whether the decision in S. C. Dixit (supra) approving the 'scaling system' requires reconsideration ? (iv) If the statistical scaling system is found to be illegal or irrational or unsound, whether the selections already made, which are the subject-matter of these petitions, should be interfered with? Re : Question (i) :

8. It is true that a judgment of this Court cannot be challenged in a petition under Article 32. It can, however, be reviewed under Article 137 or in exceptional circumstances reconsidered in exercise of inherent power, on a curative petition (See Rupa Ashok Hurra). It is equally true that a final judgment of a High Court can be challenged only by an appeal under Articles 132 to 134 or by obtaining 'special leave' under Article 136 and not by a petition under Article 32. But that is not the issue here.

9. In regard to decisions of civil courts in suits governed by Civil Procedure Code or appeals therefrom, the term 'judgment' refers to the grounds of a decree or order, 'decree' refers to the formal expression of an adjudication in a suit and 'order' refers to formal expression of any decision of a civil court which is not a decree. In regard to the decisions of High Court and Supreme Court in writ jurisdiction, the term 'judgment' is normally used to refer to the 'judgment and order', that is the grounds for the decision and the formal expression of the decision. The petitioners do not seek to upset the 'order' part of the judgment in S. C. Dixit (supra) which decided the validity of UP Civil Judge (Junior Division), Examination, 2000, held under the UP Nyayik Sewa Niyamawali 1951. The grievance of the petitioners is in regard to the UP Civil Judge (Junior Division) Examination, 2003, held under the UP Judicial Service Rules 2001. They, however, contend that the ratio decidendi of the decision in S.C. Dixit upholding the Commission's system of scaling of marks in written examination, requires reconsideration. Therefore, these petitions are neither for 'review' nor for 'setting

aside' or 'questioning' the decision in S.C. Dixit. Therefore, the bar, referred to in Rupa Ashok Hurra, will not apply.

10. The contention of Commission also overlooks the fundamental difference between challenge to the final order forming part of the judgment and challenge to the ratio decidendi of the judgment. Broadly speaking, every judgment of superior courts has three segments, namely, (i) the facts and the point at issue; (ii) the reasons for the decision; and (iii) the final order containing the decision. The reasons for the decision or the ratio decidendi is not the final order containing the decision. In fact, in a judgment of this Court, though the ratio decidendi may point to a particular result, the decision (final order relating to relief) may be different and not a natural consequence of the ratio decidendi of the judgment. This may happen either on account of any subsequent event or the need to mould the relief to do complete justice in the matter. It is the ratio decidendi of a judgment and not the final order in the judgment, which forms a precedent. The term 'judgment' and 'decision' are used, rather loosely, to refer to the entire judgment or the final order or the ratio decidendi of a judgment. Rupa Ashok Hurra (supra) is of course, an authority for the proposition that a petition under Article 32 would not be maintainable to challenge or set aside or quash the final order contained in a judgment of this Court. It does not lay down a proposition that the ratio decidendi of any earlier decision cannot be examined or differed in another case. Where violation of a fundamental right of a citizen is alleged in a petition under Article 32, it cannot be dismissed, as not maintainable, merely because it seeks to distinguish or challenge the ratio decidendi of an earlier judgment, except where it is between the same parties and in respect of the same cause of action. Where a legal issue raised in a petition under Article 32 is covered by a decision of this Court, the Court may dismiss the petition following the ratio decidendi of the earlier decision. Such dismissal is not on the ground of 'maintainability' but on the ground that the issue raised is not tenable, in view of the law laid down in the earlier decision. But if the court is satisfied that the issue raised in the later petition requires consideration and in that context the earlier decision requires re-examination, the court can certainly proceed to examine the matter (or refer the matter to a larger Bench, if the earlier decision is not of a smaller Bench). When the issue is re-examined and a view is taken different from the one taken earlier, a new ratio is laid down. When the ratio decidendi of the earlier decision undergoes such change, the final order of the earlier decision as applicable to the parties to the earlier decision, is in no way altered or disturbed. Therefore, the contention that a writ petition under Article 32 is barred or not maintainable with reference to an issue which is the subject-matter of an earlier decision, is rejected. Re : Question (ii) :

11. Article 234 of the Constitution requires appointments to the Judicial Service of a State (other than District Judges) to be made by the Governor of the State in accordance with the Rules made by him in that behalf, after consultation with the State Public Service Commission and with the High Court exercising jurisdiction in relation to such State. The UP Judicial Service Rules, 2001 (for short 'Judicial Service Rules') were made by the Governor of Uttar Pradesh in exercise of powers conferred by Article 234 and Article 309 of the Constitution, in consultation with the Commission and the Allahabad High Court, to regulate the recruitment and appointment to Uttar Pradesh Judicial Service. The Judicial Service Rules replaced the 'Uttar Pradesh Nyayik Sewa Niyamawali, 1951' which was in

force earlier. The Judicial Service Rules were amended by the Uttar Pradesh Judicial Service (Amendment) Rules, 2003.

11.1) Rule 7 of the Judicial Service Rules provides that recruitment to the post of Civil Judge (Junior Division) shall be by direct recruitment on the basis of a competitive examination conducted by Commission. Part V of the said rules lays down the procedure for recruitment to Judicial Service. Rule 16 provides for competitive examination and Rule 19 deals with the syllabus. The said rules are extracted below :

"16. Competitive Examination The examination may be conducted at such time and on such dates as may be notified by the Commission and shall consist of (a) a written examination in such legal and allied subject including procedure, as may be included in the Syllabus prescribed under rule 19, unless the same is otherwise modified by the Governor in consultation with the court and the Commission;

(b) an examination to test the knowledge of the candidates in Hindi, English and Urdu;

(c) an interview for assessing merit of the candidate giving due regard to his ability, character, personality, physique and general suitability for appointment to the service."

19. Syllabus The syllabus and the rules relating to the competitive examination shall be such as given in the Appendix II, provided that the syllabus and rules may be amended by the Governor in consultation with the Commission and Court." Appendix II to the Rules contains the syllabus for the competitive examination. It enumerates the details of the five subjects for the written examination and the number of marks carried by each subject (200 each). It also provides for a Personality Test (interview) to find out the suitability of the candidates (carrying 100 marks). Note (i) to Appendix-II provides that "the marks obtained in the interview will be added to the marks obtained in the written papers and the candidate's place will depend on the aggregate of both".

12. Sub-Rule (1) of Rule 20 of the Judicial Service Rules requires the Commission to prepare the result of the written examination and thereafter, invite such number of candidates, who in the opinion of the commission have secured minimum marks as may be fixed. Sub-Rule (2) provides for participation of a sitting Judge in the interview of candidates. Sub-rule (3) provides that the Commission shall prepare a final list of selected candidates in order of their proficiency as disclosed by aggregates of marks finally awarded to each candidate in the written examination and the interview. The proviso thereto provides that if two or more candidates obtain equal marks in the aggregate, the name of the candidate who is elder in age shall be placed higher and where two or more candidates of equal age obtain equal marks in the aggregate, the name of the candidate who has obtained higher marks in the written examination shall be placed higher. Rule 21 provides that the Governor shall on receipt of the list of candidates submitted by the Commission under Rule 20(3) make appointment on the posts of Civil Judge (Junior Division) in the order in which their names are given in the

list provided. Thus the Judicial Service Rules constitute a complete code in itself in regard to recruitment to Judicial Service. It is also evident that the marks finally awarded to each candidate in the written examination and interview are crucial both for appointment as also for purposes of inter se seniority.

13. The petitioners point out that the Judicial Service Rules do not provide for substituting the actual marks obtained by a candidate by scaled marks. It is contended that the words "marks obtained in the written papers" in Note (i) of Appendix II clearly indicate that the actual marks obtained in the written examination alone should be taken into account and not any moderated or scaled marks; that in the absence of any provision for scaling in the Judicial Service Rules, the Commission had no authority to substitute the actual marks by 'scaled marks'; and that the places/ranks of the candidates should be determined strictly on the basis of the aggregate of the actual marks obtained in the main written examination plus the marks obtained in interview.

14. The Commission contends that the manner of conducting examination by the Commission, even in regard to recruitment to Judicial Service, is governed by the Uttar Pradesh Public Service Commission (Procedure and Conduct of Business) Rules, 1976 (for short 'PSC Procedure Rules') made by the Commission in exercise of the power conferred by the UP State Public Service Commission (Regulation of Procedure and Conduct of Business) Act, 1974. Rule 26 provides for preparation of a panel of Examiners or constitution of a Committee for the purpose of holding examination in each subject. Rule 28 provides that the question papers set by the examiners shall be placed before the Commission to ensure conformity with the required standard of examination and the Commission may moderate the question papers or constitute a Committee to perform the work of moderation. Rule 30 provides for advertisement of vacancies for which selections are to be made and scrutiny of applications received. Rule 33 provides for the determination of place, dates and time of examination and the centres for examination. Rule 34 provides for the list of persons suitable to be appointed as invigilators and appointment of invigilators. Rule 37 provides for fictitious roll numbers (code numbers) to be allotted to each candidate before the answer books are dispatched to the examiners for assessment. Rule 38 provides that the number of answer books to be sent to each examiner shall be fixed by the Commission. Rule 44 requires the Secretary of the Commission to take steps for tabulation of marks obtained by each candidate as soon as the answer-scripts are received after valuation, after scrutiny of scripts, removal of discrepancies and corrections. Rule 45 provides for random checking of the tabulation to ensure correctness and accuracy of tabulation. Rule 47 provides that the original roll numbers of candidates shall thereafter be restored to the answer-scripts and for issue of interview letters. Rule 49 authorizes the Commission to decide the number of candidates to be called for interview to appear before a Board on any day. Rule 50 provides that the interview marks awarded shall be kept in safe custody. Rule 51 provides that mark-sheets shall be opened on the last day of interview and immediately thereafter the marks of interview/personality test shall be added to the marks obtained by the candidates in the written examination, and thereafter on the basis of the total so obtained, the merit list shall be prepared and placed before the Commission for final declaration of the result. The proviso to Rule 51 provides that the Commission with a view to eliminate variations in the ranks awarded to candidates at

any time at any examination or interview, adopt any method, device or formula which they consider proper for the purpose. The Commission contends that having regard to the proviso to Rule 51 which specifically enables them to adopt any method, device or formula to eliminate variations in the marks awarded to any at any examination, they are entitled to adopt the scaling system to eliminate variations in marks.

15. The petitioners point out that the PSC Procedure Rules were not made in consultation with the High Court. On the other hand, the Judicial Service Rules, 2001 which came into effect from 1.7.2000, were made in consultation with both Commission and the High Court. It is, therefore, submitted that the Judicial Service Rules alone will regulate and govern the recruitment of Civil Judges (Junior Division) including examinations and interviews and the proviso to Rule 51 of PSC Procedure Rules will not apply to recruitment of Civil Judges. Reliance is placed on the decisions of this Court in *State of Bihar v. Bal Mukund Sah*<sup>3</sup> *Union of India v. Hansoli Devi*<sup>4</sup> and *Union of India v. Deoki Nandan Aggarwal*<sup>5</sup> in regard to interpretation of the Rules.

16. This question was considered briefly by this Court in *S. C. Dixit* wherein it was held that the PSC Procedure Rules made in exercise of power under the U.P. State Public Service Commission (Regulation of Procedure and Conduct of Business) Act, 1974 give the guidelines for any examination to be held by the Commission and therefore, all the provisions of the said Rules will be applicable to an examination for recruitment to judicial service also.

17. It is no doubt true that Judicial Service Rules govern the recruitment to Judicial Service, having been made in exercise of power under Article 234, in consultation with both the commission and the High Court. It also provides what examinations should be conducted and the maximum marks for each subject in the examination. But the Judicial Service Rules entrust the function of conducting examinations to the Commission. The Judicial Service Rules do not prescribe the manner and procedure for holding the examination and valuation of answer-scripts and award of the final marks and declaration of the results. Therefore, it is for the Commission to regulate the manner in which it will conduct the examination and value the answer scripts, subject, however, to the provisions of the Judicial Service Rules. If the Commission has made Rules to regulate the procedure and conduct of the examination, they will naturally apply to any examination conducted by it for recruitment to any service, including the judicial service. But where the Judicial Service Rules make a specific provision in regard to any aspect of examination, such provision will prevail, and the provision of PSC Procedure Rules, to the extent it is inconsistent with the Judicial Service Rules, will be inapplicable. Further, if both the Rules have made provision in regard to a particular matter, the PSC Procedure Rules will yield to the Judicial Service Rules.

18. The manner in which the list of candidates as per merit should be prepared is provided both in the Judicial Service Rules and the PSC Procedure Rules. Relevant portion of Rule 20(3) and Note (i) of Appendix-II of the Judicial Service Rules and Rule 51 of the PSC Procedure Rules providing for the aggregation of marks and preparation of the merit list, are extracted below :- Judicial Service Rules PSC Procedure Rules. Rule 20(3). The Commission

then shall prepare a final list of selected candidates in order of their proficiency as disclosed by aggregate of marks finally awarded to each candidate in the written examination and the interview. Note (i) of Appendix-II. - Marks obtained in the interview will be added to the marks obtained in the written papers and the candidates' place will depend on the aggregate of the both.

“Rule 51. The marks-sheets so obtained shall be opened on the last day of interview and immediately there after the marks of interview/ personality test shall be added to the marks obtained by the candidates in the written examination. Thereafter, on the basis of the totals so obtained the merit list shall be prepared and place before the Commission for final declaration of the result. Provided that the Commission may, with a view to eliminating variation in the marks awarded to candidates at any examination or interview, adopt any method, device or formula which they consider proper for the purpose. (different emphasis supplied) As the field is occupied by Rule 20(3) and Note (i) of Appendix-II of Judicial Service Rules, they will prevail over the general provision in Rule 51 of PSC Procedure Rules.”

19. Rule 20(3) provides that the final list of selected candidates in order of their proficiency as disclosed by the aggregate of 'marks finally awarded to each candidate in the written examination and the interview". Note (i) to Appendix II of the Judicial Service Rules provides that the "marks obtained in the interview" will be added to "the marks obtained in the written papers" and that the candidate's place will depend on the aggregate of both. Though Judicial Service Rules refers to 'marks finally awarded', the said Rules do not contain a provision similar to the proviso to Rule 51 of PSC Procedure Rules, enabling the Commission to adopt any method, device or formula to eliminate variation in the marks. It is not possible to read the proviso to Rule 51 or words to that effect into Rule 20(3) or Note (i) of Appendix-II of Judicial Service Rules. It is well settled that courts will not add words to a statute or read into the statute words not in it. Even if the courts come to the conclusion that there is any omission in the words used, it cannot make up the deficiency, where the wording as it exists is clear and unambiguous. While the courts can adopt a construction which will carry out the obvious intention of the legislative or rule making authority, it cannot set at naught the legislative intent clearly expressed in a statute or the rules. Therefore, Rule 20(3) and Note (i) of Appendix-II has to be read as they are without the addition of the proviso to Rule 51 of PSC Procedure Rules. If so, what can be taken into account for preparing final list of selected candidates, are 'marks finally awarded to a candidate' in the written examination and the interview. The marks assigned by the examiner are not necessarily the marks finally awarded to a candidate. If there is any error in the marks awarded by the examiner it can always be corrected by the Commission and the corrected marks will be 'the final marks awarded to the candidate'. Where the Commission is of the view that there is 'examiner variability' in the marks (due to strict or liberal assessment of answer scripts) or improper assessment on account of erratic or careless marking by an examiner, they can be corrected appropriately by moderation. The moderation is either by adding (in the case of strict examiners) or deducting (in the case of liberal examiners) a particular number of marks which has been decided with reference to principles of moderation applied. If there is erratic or careless marking, then moderation is by fresh valuation by another examiner.

Therefore, the marks assigned by the examiner as moderated will be the marks finally awarded to the candidates or marks obtained by the candidates. Moderation, it has to be held, is inherent in the evaluation of answer scripts in any large scale examination, where there are more than one examiner.

20. We cannot accept the contention of the petitioner that the words "marks awarded" or "marks obtained in the written papers" refers only to the actual marks awarded by the examiner. 'Valuation' is a process which does not end on marks being awarded by an Examiner. Award of marks by the Examiner is only one stage of the process of valuation. Moderation when employed by the examining authority, becomes part of the process of valuation and the marks awarded on moderation become the final marks of the candidate. In fact Rule 20(3) specifically refers to the 'marks finally awarded to each candidate in the written examination', thereby implying that the marks awarded by the examiner can be altered by moderation.

21. But the question is whether the raw marks which are converted into scaled scores on an artificial scale which assumed variables (assumed mean marks and assumed standard deviation) can be considered as 'marks finally awarded' or 'marks obtained'. Scaled scores are not marks awarded to a candidate in a written examination, but a figure arrived at for the purpose of being placed on a common scale. It can vary with reference to two arbitrarily fixed variables, namely 'Assumed Mean' and 'Assumed Standard Mean'. We have dealt with this aspect in greater detail while dealing with question (iii). For the reasons given while considering question (iii), we hold that 'scaled scores' or 'scaled marks' cannot be considered to be 'marks awarded to a candidate in the written examination'. Therefore, scaling violates Rule 20(3) and Note (i) of Appendix-II of Judicial Service Rules.

22. Rule 20 of Judicial Service Rules requires the Commission to call for interview such number of candidates, who in its opinion have secured the minimum marks fixed by it. Because of application of scaling system by the Commission, it has not been possible for the Commission to fix such minimum marks either for individual subjects or for the aggregate. In the absence of minimum marks, several candidates who secured less than 30% in a subject have been selected. We note below by way of illustration, the particulars of some candidates who have been selected in spite of securing less than 20% in a subject :

| S.No. | Roll No. | Subject Actual marks(%) | Scaled Marks | Rank in Selection |
|-------|----------|-------------------------|--------------|-------------------|
| 1     | 012610   | 8%                      | 79           | 225               |
| 2     | 032373   | 8%                      | 79           | 290               |
| 3     | 002454   | 11%                     | 79           | 196               |
| 4     | 008097   | 13%                     | 89           | 85                |
| 5     | 017808   | 13%                     | 76           | 317               |
| 6     | 010139   | 14%                     | 85           | 333               |
| 7     | 012721   | 15%                     | 100          | 172               |
| 8     | 002831   | 16%                     | 89           | 263               |

|   |        |     |    |     |
|---|--------|-----|----|-----|
| 9 | 004998 | 17% | 91 | 161 |
|---|--------|-----|----|-----|

Thus scaling system adopted by the Commission, contravenes Rule 20(1) also.

Re : Question (iii) :

23. When a large number of candidates appear for an examination, it is necessary to have uniformity and consistency in valuation of the answer- scripts. Where the number of candidates taking the examination are limited and only one examiner (preferably the paper-setter himself) evaluates the answer-scripts, it is to be assumed that there will be uniformity in the valuation. But where a large number of candidates take the examination, it will not be possible to get all the answer-scripts evaluated by the same examiner. It, therefore, becomes necessary to distribute the answer-scripts among several examiners for valuation with the paper-setter (or other senior person) acting as the Head Examiner. When more than one examiner evaluate the answer-scripts relating to a subject, the subjectivity of the respective examiner will creep into the marks awarded by him to the answer- scripts allotted to him for valuation. Each examiner will apply his own yardstick to assess the answer-scripts. Inevitably therefore, even when experienced examiners receive equal batches of answer scripts, there is difference in average marks and the range of marks awarded, thereby affecting the merit of individual candidates. This apart, there is 'Hawk- Dove' effect. Some examiners are liberal in valuation and tend to award more marks. Some examiners are strict and tend to give less marks. Some may be moderate and balanced in awarding marks. Even among those who are liberal or those who are strict, there may be variance in the degree of strictness or liberality. This means that if the same answer-script is given to different examiners, there is all likelihood of different marks being assigned. If a very well written answer-script goes to a strict examiner and a mediocre answer-script goes to a liberal examiner, the mediocre answer-script may be awarded more marks than the excellent answer-script. In other words, there is 'reduced valuation' by a strict examiner and 'enhanced valuation' by a liberal examiner. This is known as 'examiner variability' or 'Hawk-Dove effect'. Therefore, there is a need to evolve a procedure to ensure uniformity inter se the Examiners so that the effect of 'examiner subjectivity' or 'examiner variability' is minimised. The procedure adopted to reduce examiner subjectivity or variability is known as moderation. The classic method of moderation is as follows :

“(i) The paper-setter of the subject normally acts as the Head Examiner for the subject. He is selected from amongst senior academicians/scholars/senior civil servants/Judges. Where the case of a large number of candidates, more than one examiner is appointed and each of them is allotted around 300 answer-scripts for valuation.

(ii) To achieve uniformity in valuation, where more than one examiner is involved, a meeting of the Head Examiner with all the examiners is held soon after the examination. They discuss thoroughly the question paper, the possible answers and the weightage to be given to various aspects of the answers. They also carry out a sample valuation in the light of their discussions. The sample valuation of scripts by

each of them is reviewed by the Head Examiner and variations in assigning marks are further discussed. After such discussions, a consensus is arrived at in regard to the norms of valuation to be adopted. On that basis, the examiners are required to complete the valuation of answer scripts. But this by itself, does not bring about uniformity of assessment inter se the examiners. In spite of the norms agreed, many examiners tend to deviate from the expected or agreed norms, as their caution is overtaken by their propensity for strictness or liberality or erraticism or carelessness during the course of valuation. Therefore, certain further corrective steps become necessary.

(iii) After the valuation is completed by the examiners, the Head Examiner conducts a random sample survey of the corrected answer scripts to verify whether the norms evolved in the meetings of examiner have actually been followed by the examiners. The process of random sampling usually consists of scrutiny of some top level answer scripts and some answer books selected at random from the batches of answer scripts valued by each examiner. The top level answer books of each examiner are revalued by the Head Examiner who carries out such corrections or alterations in the award of marks as he, in his judgment, considers best, to achieve uniformity. (For this purpose, if necessary certain statistics like distribution of candidates in various marks ranges, the average percentage of marks, the highest and lowest award of marks etc. may also be prepared in respect of the valuation of each examiner.) (iv) After ascertaining or assessing the standards adopted by each examiner, the Head Examiner may confirm the award of marks without any change if the examiner has followed the agreed norms, or suggest upward or downward moderation, the quantum of moderation varying according to the degree of liberality or strictness in marking. In regard to the top level answer books revalued by the Head Examiner, his award of marks is accepted as final. As regards the other answer books below the top level, to achieve maximum measure of uniformity inter se the examiners, the awards are moderated as per the recommendations made by the Head Examiner.

(v) If in the opinion of the Head Examiner there has been erratic or careless marking by any examiner, for which it is not feasible to have any standard moderation, the answer scripts valued by such examiner are revalued either by the Head Examiner or any other Examiner who is found to have followed the agreed norms.

(vi) Where the number of candidates is very large and the examiners are numerous, it may be difficult for one Head Examiner to assess the work of all the Examiners. In such a situation, one more level of Examiners is introduced. For every ten or twenty examiners, there will be a Head Examiner who checks the random samples as above. The work of the Head Examiners, in turn, is checked by a Chief Examiner to ensure proper results. The above procedure of 'moderation' would bring in considerable uniformity and consistency. It should be noted that absolute uniformity or consistency in valuation is impossible to achieve where there are several examiners and the effort is only to achieve maximum uniformity.”

24. In the Judicial Service Examination, the candidates were required to take the examination in respect of the all five subjects and the candidates did not have any option in regard to the subjects. In such a situation, moderation appears to be an ideal solution. But there are examinations which have a competitive situation where candidates have the option of selecting one or few among a variety of heterogenous subjects and the number of students taking different options also vary and it becomes necessary to prepare a common merit list in respect of such candidates. Let us assume that some candidates take Mathematics as an optional subject and some take English as the optional subject. It is well-recognised that a mark of 70 out of 100 in mathematics does not mean the same thing as 70 out of 100 in English. In English 70 out of 100 may indicate to an outstanding student whereas in Mathematics, 70 out of 100 may merely indicate an average student. Some optional subjects may be very easy, when compared to others, resulting in wide disparity in the marks secured by equally capable students. In such a situation, candidates who have opted for the easier subjects may steal an advantage over those who opted for difficult subjects. There is another possibility. The paper setters in regard to some optional subjects may set questions which are comparatively easier to answer when compared some paper setters in other subjects who set tougher questions difficult to answer. This may happens when for example, in a Civil Service examination, where Physics and Chemistry are optional papers, examiner 'A' sets a paper in Physics appropriate to a degree level and examiner 'B' sets a paper in Chemistry appropriate for matriculate level. In view of these peculiarities, there is a need to bring the assessment or valuation to a common scale so that the inter se merit of candidates who have opted for different subjects, can be ascertained. The moderation procedure referred to in the earlier para will solve only the problem of examiner variability, where the examiners are many, but valuation of answer scripts is in respect of a single subject. Moderation is no answer where the problem is to find inter se merit across several subjects, that is, where candidates take examination in different subjects. To solve the problem of inter se merit across different subjects, statistical experts have evolved a method known as scaling, that is creation of scaled score. Scaling places the scores from different tests or test forms on to a common scale. There are different methods of statistical scoring. Standard score method, linear standard score method, normalized equi- percentile method are some of the recognized methods for scaling.

25. A. *Edwin Harper Jr. & V Vidya Sagar Misra*<sup>6</sup> in their publication "Research on Examinations in India" have tried to explain and define scaling. We may usefully borrow the same. A degree 'Fahrenheit' is different from a degree 'Centigrade'. Though both express temperature in degrees, the 'degree' is different for the two scales. What is 40 Degrees in Centigrade scale is 104 Degrees in Fahrenheit scale. Similarly, when marks are assigned to answer-scripts in different papers, say by Examiner 'A' in Geometry and Examiner 'B' in History, the meaning or value of the 'mark' is different. Scaling is the process which brings the mark awarded by Examiner 'A' in regard to Geometry scale and the mark awarded by Examiner 'B' in regard to History scale, to a common scale. Scaling is the exercise of putting the marks which are the results of different scales adopted in different subjects by different examiners into a common scale so as to permit comparison of inter se merit. By this exercise, the raw marks awarded by the examiner in different subjects is converted to a 'score' on a common scale by applying a statistical formula. The 'raw marks' when converted to a

common scale are known as the 'scaled marks'. Scaling process, whereby raw marks in different subjects are adjusted to a common scale, is a recognized method of ensuring uniformity inter se among the candidates who have taken examinations in different subjects, as, for example, the Civil Services Examination.

26. The Union Public Service Commission ('UPSC' for short) conducts the largest number of examinations providing choice of subjects. When assessing inter se merit, it takes recourse to scaling only in civil service preliminary examination where candidates have the choice to opt for any one paper out of 23 optional papers and where the question papers are of objective type and the answer scripts are evaluated by computerized/ scanners. In regard to compulsory papers which are of descriptive (conventional) type, valuation is done manually and scaling is not resorted to. Like UPSC, most examining authorities appear to take the view that moderation is the appropriate method to bring about uniformity in valuation where several examiners manually evaluate answer-scripts of descriptive/ conventional type question papers in regard to same subject; and that scaling should be resorted only where a common merit list has to be prepared in regard to candidates who have taken examination of different subjects, in pursuance of an option given to them.

27. But some Examining Authorities, like the Commission are of the view that scaling can be used, not only where there is a need to find a common base across different subjects (that is bringing the performance in different subjects to a common scale), but also as an alternative to moderation, to reduce examiner variability (that is where different examiners evaluate answer scripts relating to the same subject).

28. Let us now examine the reasons as to why the Commission adopted 'scaling' instead of moderation. The Committee states that the anomalies caused on account of 'examiner variability' was engaging its attention. It found that a candidate's score may depend upon the "chance" factor of whether his answers script is assessed by a lenient or a strict examiner; and that in an extreme case, while a candidate of a given merit may get a First Class/Division, another student of equal merit may be declared to have failed. Therefore, the Commission constituted a Committee to carry out an indepth study into the matter and suggest appropriate means to ensure that the evaluation was on more equitable basis. The Committee by its Report dated 2.9.1996 suggested statistical scaling system as the remedy and recommended the linear standard score method which operates on the following formula :

“ $Z = \frac{X - M}{S.D.}$  SD Z= is the Scaled Score.

X = is the Raw mark.

M = is the mean of Raw Marks of the group/subject.

S.D. is the Standard Deviation of Raw Marks of the group/subject.

The Committee suggested the following 'assumptions' or 'parameters' for applying the formula :

(i) Assumed Mean will be taken as Half of the maximum marks of the group/subject.

(ii) Assumed S.D. will be taken as one-fifth of the assumed mean.

(iii) If scaled score is less than zero after scaling, then candidates will be allotted zero marks in the said group/subject.

(iv) If scaled score after scaling is more than maximum marks, then candidate will be allotted maximum marks in the said group/subject.”

29. Eversince then, the Commission has been following the statistical scaling. According to the Commission, the scaling method is rational, scientific and reasonable and would lead to assessment of inter se merit of the candidates in a just and proper manner. The use of the said method was reviewed by an Expert Committee on 31.7.2000 and it was reiterated that the formula and method presently used for scaling can be continued to be used in future also and there was no need to change the same. Thus the scaling is continued.

30. We may at this stage refer to the condition to be fulfilled, for scaling to be effective. For this purpose, we are referring to passages from the Authors/Experts relied on by the Commission itself.

30.1) A. Edwin Harper & Vidya Sagar Misra (in 'Research on Examinations in India) make it clear that scaling will be useful and effective only if the distribution of marks in the batch of answer scripts sent to each examiner is approximately the same as the distribution of marks in the batch of answer scripts sent to every other examiner.

30.2) A similar view is expressed by J.P. Guilford & Benjamin Fruchter (in their treatise 'Fundamental Statistics in Psychology and Education' page 476-477). They say that two conditions are to be satisfied to apply scaling :

(i) The population of students from which the distributions of scores arose must be assumed to have equal means and dispersions in all the abilities measured by the different tests; and

(ii) the form of distribution, in terms of skewness and kurtosis, must be very similar from one ability to another. He proceeds to refer to the disadvantages of scaling thus :

"Unfortunately, we have no ideal scales common to all these tests, with measurements which would tell us about these population parameters. Certain selective features might have brought about a higher mean, a narrower dispersion, and a negatively skewed distribution on the actual continuum of ability measured by one test, and a lower mean, a wider dispersion, and a symmetrical distribution on the continuum of another ability represented by another test. Since we can never know definitely about these features for any given population, in common scaling we often have to proceed on the assumption that actual means, standard deviations, and form of distribution are uniform for all abilities measured. In spite of these limitations, it is almost certain that derived scales provide more nearly comparable scales than do raw scores." 30.3) V.

Natarajan & K. Gunasekaran in their treatise 'Scaling Techniques what, why and how', have warned :

"If one studies the literature in this field, he can find that there are a number of methods available ranging from simple to complex. Each has its own merits and demerits and can be adopted only under certain conditions or making certain assumptions." The Authors describe the Linear Standard Score method (which is used by the Commission) thus :

"Unlike Z-score (Standard score) which has a mean of 'zero' and standard deviation 'one', the linear standard score has some pre-determined mean and standard deviations. ..the choice of the mean and standard deviations is purely arbitrary. Each has its own advantages and disadvantages and useful for specific purpose only. It may be emphasized here that both the standard scores and linear standard scores retain the shape of the original distribution of raw marks. Therefore, if the original distribution is 'normally' distributed, then any type of Linear Standard Scores will also be 'normally' distributed. Taking the Normal Curve as the model, various points in other scales are plotted. It should be, however, noted that the kind of relationship shown in Figure -2 between normal curve vis-à-vis the other scores are valid only if the raw score distribution can be assumed to approximately normally distributed. (emphasis supplied) 30.4) The Kothari Report, 1976 ('Policy & Selection Methods' published by UPSC) while referring to scaling in regard to papers in different subjects, by using appropriate statistical techniques as a recognized procedure for improving the reliability of examination as a tool for selection, however cautions that the method should be under continuous review and evaluation, that continuing improvement in the light of experience and new developments, taking into account advancement of knowledge, is essential.

31. The entire basis for applying scaling in regard to marks awarded by different examiners in the same subject is the assumption that all answer scripts have been thoroughly mixed, and that equal number of answer scripts drawn at random and sent to each examiner for valuation will contain answer scripts of candidates with equal distribution of abilities. When the distribution of abilities in each batch is approximately equal, the mean marks and standard deviation of the scaled marks of each batch will be identical. To put it differently, if each examiner is sent 300 answer scripts and each batch of 300 candidates have almost equal number of good, average and poor standard students, they can all be brought to a common scale for comparing their merit inter se. But we find that there is no such broad equal distribution in the examination with which we are concerned. We find from the Tables furnished that the range of marks awarded and the range of deviation have varied enormously from examiner to examiner in the same subject. We extract below these ranges, which demonstrate the wide diversity, in turn indicating that scaling method was inappropriate for bringing uniformity in valuation : Subject No. of Examiner No. of Scripts Examined (range) Mean Marks of the examiner (range) Standard Deviation of marks allotted (range) Minimum Marks (awarded by the Examiner) Maximum Marks (awarded by the Examiner)

- “1. General Knowledge
2. Language
3. Law-I

4. Law-II

5. Law-III 18 14 11 10 14 50 to 800 231 to 800 300 to 900 200 to 1402 150 to 1000  
47.4 to 83.91 37.51 to 82.43 30.83 to 56.90 70.57 to 94.40 63.14 to 86.74 12.24 to  
20.49 14.16 to 31.75 12.45 to 17.85 11.48 to 20.05 13.16 to 19.54 10 to 43 0 to 30 0  
to 10 0 to 40 0 to 31 84 to 126 105 to 145 83 to 113 to 132 99 to 134”

32. The formula heavily relies upon the standard deviation among the candidates in a given pool or batch. The standard deviation is a measure of the range and distribution of marks awarded by an examiner. It depends on the set of students in any given pool. If an examiner has a set of extremely good or poor standard candidates and another examiner has a more even set of average candidates, the standard deviation would be high for the first examiner and low for the second examiner, having regard to the range of distribution of marks. Consequently the scaled marks of a candidate calculated on a formula heavily relying on standard deviation, would be based on the cumulative standard deviation of all the candidates in his pool rather than the strictness or liberality of the examiner. Therefore, standard deviation has only a bearing on ascertaining the range of capabilities of the candidates in a given examination and in no way eliminates the anomalies arising out of the strictness or liberality of the examiner. We may demonstrate the fact that the scaled marks vary with reference to the extent of standard deviation (and has nothing to do with the issue of strictness or liberality of the examiner), from the following examples : Actual Marks Average (Mean) Marks Strict Examiner No. I Strict Examiner No. II Standard Deviation Scaled Marks Standard Deviation Scaled Marks 0 5 20 50 50 50 15 15 15 33 40 60 25 25 25 60 64 76 Actual Marks Average (Mean) Marks Liberal Examiner No. I Liberal Examiner No. II Standard Deviation Scaled Marks Standard Deviation Scaled Marks 50 120 150 90 90 90 15 15 15 47 140 180 25 25 25 68 124 148 The reason given for introducing scaling is to cure the disparity on account of strictness or liberality of the examiners. But the effect of the scaling formula adopted by Commission is to average the marks of a batch of candidates and convert the raw marks of each candidate in the batch into scaled marks with reference to the average marks of the batch and the standard deviation. The scaling formula therefore, does not address or rectify the effect of strictness or liberality of the examiner. The scaling formula is more suited and appropriate to find a common base and inter se merit, where candidates take examinations in different subjects. As the scaling formula has no nexus or relevance to give a solution to the problem of eliminating the variation or deviation in the standard of valuation of answer scripts by different examiners either on account of strictness or liberality, it has to be concluded that scaling is based on irrelevant considerations and ignores relevant considerations.

33. We will next refer to apparent anomalies which show scaling of marks is arbitrary. The Commission has furnished five Tables relating to the five subjects showing the following particulars : (i) The number of examiners, (ii) Number of answer scripts allotted to each examiner; (iii) Mean marks of each examiner; (iv) Standard deviation of the marks allotted by each examiner; (v) Minimum raw marks secured by a candidate in the batch of answer-scripts corrected by each examiner; (vi) Maximum raw marks secured by a candidate in the batch of answer-scripts corrected by each examiner. The Commission has also furnished the

tabulation of scaled and actual marks of all the candidates. An examination of the particulars furnished discloses several glaring anomalies.

I. Award of high scaled marks to those who secured zero marks :

“We find from Table-II (furnished by the Commission) that the answer scripts relating to Language Paper were distributed among 14 examiners. Several candidates whose papers were evaluated by examiners 2, 3, 4, 5, 6, 8, 13, & 14 have secured zero marks. Evidently only those who did not attempt any answer or had absolutely no knowledge of either Hindi or English would have got zero marks. But such candidates who actually secured zero marks have strangely been assigned scaled marks ranging from 36 to 67, depending upon the examiner, in whose pool, they fell. We give below scaled marks obtained by different candidates who secured zero marks with reference to the examiners. Subject : Language Examiner No. Raw Marks of the candidate Scaled Marks 2 0  $(100)+(0-66.58 \times 20) = 44$  23.73 3 0  $100+(0-55.29 \times 20) = 47$  20.91 4 0  $100+(0-74.88 \times 20) = 0$  (-5 to be taken as zero) 14.20 5 0  $100+(0-44.48 \times 20) = 58$  20.06 6 0  $100+(0-61.52 \times 20) = 50$  24.8 8 0  $100+(0-52.86 \times 20) = 67$  31.75 13 0  $100+(0-43.11 \times 20) = 66$  25.50 14 0  $100+(0-54.77 \times 20) = 36$  17.02 But unfortunately in the same subject, candidates who secured 32 to 30 marks, assessed by Examiner No.10, got their marks reduced to 31 to 28 on scaling. (Mean being 80.93 and SD being 14.16). The devastating effect of awarding such high scaled marks, that too ranging from 36 to 67, to those who have secured '0' need not be stressed. In fact UPSC has clarified that whenever they follow scaling procedure, no scaling is applied to '0' marks. But the Commission had not applied its mind to this aspect when applying 'scaling'

II. Equalization of marks of persons who secured very high marks. The scaling has equalized the different high end marks of candidates, where the mean marks is low. To give a hypothetical example if the mean marks is 70 and the standard deviation is 15, all candidates securing raw marks 145 to 200 will be assigned the equal scaled marks of 200. If the mean marks are 60 and the standard deviation is 15, all candidates securing 135 to 200 will be awarded the scaled marks of 200. Similarly, if the mean marks are 80 and the standard deviation is 20, all candidates securing raw marks between 180 to 200 will be awarded equal scaled marks of 200. In addition to the above hypothetical examples, we may give a concrete example. In regard to Examiner No. 14 in Language Paper, Table-II shows that the highest marks secured is 145. In regard to that examiner, the mean marks is 54.77 and standard deviation is 17.02. By applying the scaling formula, the marks of 145 secured by that candidate becomes 206 which is taken as 200 as per the formula. All candidates who were awarded raw marks of 140 to 145 by Examiner No. 14 in Language paper will be assigned the equal scaled marks of 200. This leads to unequals being treated as equals. In case of candidates securing marks in higher ranges on scaling, there is likelihood of their marks being equalised with those who secured lesser marks thereby losing the benefit of their higher marks and inter se merit.

III. Equalization of marks of persons who secured low marks. The scaling has also equalized the different low end marks of candidates, where the mean marks is high. To give a hypothetical example, if the mean marks is 95 and the standard deviation is 11, then all

candidates securing 40 and below will be awarded only '0'. To give a concrete example, in regard to Examiner No. 7 in Law Paper-II, one candidate has secured 32. In respect of that examiner, the mean marks is 94.4 and standard deviation is 11.48. By applying the scaling formula, the scaled marks of the said candidate who secured 32 becomes '0'. Not only that. Scaled marks of all candidates who were given raw marks of 37 and less by that examiner, becomes '0'. This leads to unequals being treated as equals and candidates who secured marks in the lower ranges (from that examiner) losing out to candidates who performed much worse but were in the pool of other examiners.

#### IV. Inadequate mixing of answer scripts and improper distribution of answer scripts :

The basic requirement for scaling is that all answer scripts will be mixed thoroughly and that approximately equal number of answer scripts drawn at random will be allotted to each examiner so as to infer equal distribution of ability of candidates in each batch of answer scripts. But that was apparently not done by the Commission. We give below the details of distribution of answer scripts which demonstrate that they were nowhere equal :

General Knowledge Paper (18 Examiners) The distribution of answer scripts is : 50 papers (2 examiners), 100 (3 examiners), 150 (1 examiner), 200 (2 examiners), 250 (2 examiners), 300 (1 examiner), 350 (1 examiner), 400 (1 examiner), 500 (2 examiners), 648 (1 examiners) and 800 (2 examiners).

Language Paper (14 Examiners) The distribution of answer scripts is : 231 papers (1 examiner), 300 (5 examiners), 350 (1 examiner), 400 (2 examiners), 450 (3 examiners), 700 (1 examiner), 800 (1 examiner).

Law Paper-I (11 Examiners) - The distribution of answer scripts is : 100 papers (1 examiner), 300 (2 examiners), 400 (2 examiners), 450 (1 examiner), 600 (1 examiner), 700 (1 examiner), 775 (1 examiner), 800 (1 examiner), 900 (1 examiner).

Law paper-II (10 examiners) - The distribution of answer scripts is : 200 papers (1 examiner), 300 (1 examiner), 350 (1 examiner), 450 (1 examiner), 500 (2 examiners), 650 (2 examiners), 700 (1 examiner), 1402 (1 examiner).

Law paper-III (14 examiners) The distribution of answer scripts is : 150 papers (3 examiners), 200 (1 examiner), 250 (1 examiner), 300 (1 examiner), 350 (2 examiners), 400 (1 examiner), 444 (1 examiner), 500 (1 examiner), 550 (1 examiner), 900 (1 examiner), 1000 (1 examiner).

Very large variation in the number of answer scripts allotted to each examiner has a bearing on the mean marks and the standard deviation. The fact that there was no proper randomization and distribution is also evident from the fact that though approximately equal number appeared in each segment of 10000 from among the roll nos. 1 to 51524, selection is inexplicably high in the first segment of roll nos. 1 to 10000. The particulars of roll number segments and the number of persons who appeared for the main examination from each segment are as follows :

“Roll Numbers No. of Persons

1. 1-10000 1072

2. 10001 to 20000 1115
3. 20001 to 30000 1124
4. 30001 to 40000 1031
5. 40001 to 50000 1112
6. 50001 to 51524 170

If there was proper randomization and distribution leading to equal distribution of the candidate capacity, it would have been expected that the number of selected candidates also would have been proportionate to each segment. But we find that out of 347 candidates selected, as many as 139 candidates fall in first segment alone (within Roll nos. 1 to 10000) and 208 fall in the next five segments put together. Significantly out of the top 150 selected candidates, as many as 68 candidates also fall within Roll nos. 1 to 10000. Be that as it may.

V. Low raw marks were further lowered (or made into '0') and higher raw marks were further increased due to scaling Example : Law Paper-II. Examiner No. 5 : 33 became 9; and 120 became 146 Examiner No. 6 : All marks between 9 and 1 became 0; and 119 became 139 Examiner No. 7 : All marks between 37 and 1 became 0; and 132 became 165 Examiner No. 9 : 4 became 0; and 122 became 156 In contrast, in some cases all raw marks whether low or high, became higher.

Example : Law Paper-I.

Examiner No. 4 : 1 became 56; and 102 became 177.

Examiner No. 6 : 9 became 66; and 85 became 184.

Examiner No. 9 : 1 became 60; and 107 became 184.

Examiner No. 10 : 9 became 49; and 83 became 156.

The petitioners have referred to certain other absurdities arising from the application of scaling, with reference to the results of 2000 examination which was the subject matter of S.C. Dixit. (For example, it was demonstrated that in some cases, the low marks awarded by liberal examiners had increased and high marks awarded by strict examiners had reduced, thereby achieving the opposite of the goal sought to be achieved -- that marks given by liberal examiners should be reduced and marks given by strict examiners should be increased). We however consider it appropriate to rely only on the anomalies/absurdities demonstrable with reference to the 2003 examination which is the subject matter of these petitions, and do not propose to rely on the anomalies noticed in regard to the 2000 examination.”

34. When selections are made on the basis of the marks awarded, and the inter se ranking depends on the marks awarded, treating unequals equally, or giving huge marks to candidates who have secured zero marks in some subjects make the process wholly irrational, virtually bordering on arbitrariness. It is no doubt true that such irrationality may adversely affect only those cases which are at either end of the spectrum, and if they are excluded, by and large the scaling system may be functional. But if the extreme cases are even 20 out of 5000 for each of the subjects, it becomes 100 for 5 subjects, which means that the results of as many as 100 are likely to be affected. It may be more also. In that process, at least 5% to 10% of the vacancies are likely to be filled up by less meritorious candidates. This will lead to

considerable heart-burn and dissatisfaction. When the object of the selection process is to try to select the best, and even one mark may make the difference between selection or non-selection, the system of scaling which has the effect of either reducing or increasing the marks in an arbitrary manner will lead to unjust results. This is in addition to the main disadvantage that scaling does not remedy the ill-effects of examiner variability arising out of strictness or liberality in valuation.

35. The illustrations given above with reference to the 2003 examinations clearly demonstrate the arbitrariness and irrationality of scaling, particularly in cases falling at the two ends of the spectrum. We, therefore, hold that scaling system as adopted by the Commission is unsuited for the Civil Judge (Junior Division) Examination.

36. We may now summarize the position regarding scaling thus :

(i) Only certain situations warrant adoption of scaling techniques.

(ii) There are number of methods of statistical scaling, some simple and some complex. Each method or system has its merits and demerits and can be adopted only under certain conditions or making certain assumptions.

(iii) Scaling will be useful and effective only if the distribution of marks in the batch of answer scripts sent to each examiner is approximately the same as the distribution of marks in the batch of answer scripts sent to every other examiner.

(iv) In the Linear Standard Method, there is no guarantee that the range of scores at various levels will yield candidates of comparative ability.

(v) Any scaling method should be under continuous review and evaluation and improvement, if it is to be a reliable tool in the selection process.

(vi) Scaling may, to a limited extent, be successful in eliminating the general variation which exists from examiner to examiner, but not a solution to solve examiner variability arising from the 'hawk-dove' effect (strict/liberal valuation). The material placed does not disclose that the Commission or its expert committee have kept these factors in view in determining the system of scaling. We have already demonstrated the anomalies/absurdities arising from the scaling system used. The Commission will have to identify a suitable system of evaluation, if necessary by appointing another Committee of Experts. Till such new system is in place, the Commission may follow the moderation system set out in Para 23 above with appropriate modifications.

37. We may now refer to the decision of this Court in S. C. Dixit. The validity of scaling was considered in paras 31 to 33 of the judgment extracted below :

"31. There is a vast percentage difference in awarding of marks between each set of examiners and this was sought to be minimized by applying the scaling formula. If

scaling method had not been used, only those candidates whose answer-sheets were examined by liberal examiners alone would get selected and the candidates whose answer-sheets were examined by strict examiners would be completely excluded, though the standard of their answers may be to some extent similar. The scaling system was adopted with a view to eliminate the inconsistency in the marking standards of the examiners. The counsel for the respondents could not demonstrate that the adoption of scaling system has in any way caused injustice to any meritorious candidate. If any candidate had secured higher marks in the written examination, even by applying scaling formula, he would still be benefited.

32. The Division Bench of the High Court observed that the process of scaling was done examiner-wise only and the scaling formula did not take into consideration the average of mean of all the candidates in one particular paper but took the mean of only that group of candidates which has been examined by one single examiner. The counsel for U.P. PSC submitted that the observation made by the High Court is incorrect. The scaling formula was adopted to remove the disparity in the evaluation of 14 examiners who participated in the evaluation of answer-sheets and the details have also been furnished as to how the scaling formula was adopted and applied. Therefore, we do not think that the observation of the Division Bench that the Commission did not take care of varying standards which may have been applied by different examiners but has sought to reduce the variation of the marks awarded by the same examiner to different candidates whose answer-sheets had been examined, is correct. The Division Bench was of the view that as a result of scaling, the marks of the candidates who had secured zero marks were enhanced to 18 and this was illegal and thus affected the selection process. The finding is to be understood to mean as to how the scaling system was applied. 18 marks were given notionally to a candidate who secured zero marks so as to indicate the variation in marks secured by the candidates and to fix the mean marks.

33. In that view of the matter, we do not think that the application of scaling formula to the examinations in question was either arbitrary or illegal. The selection of the candidates was done in a better way. Moreover, this formula was adopted by U.P. PSC after an expert study and in such matters, the court cannot sit in judgment and interfere with the same unless it is proved that it was an arbitrary and unreasonable exercise of power and the selection itself was done contrary to the Rules. Ultimately, the agency conducting the examination has to consider as to which method should be preferred and adopted having regard to the myriad situations that may arise before them." S. C. Dixit, therefore, upheld scaling on two conclusions, namely (i) that the scaling formula was adopted by the Commission after an expert study and in such matters, court will not interfere unless it is proved to be arbitrary and unreasonable; and (ii) the scaling system adopted by the Commission eliminated the inconsistency arising on account of examiner variability (differences due to evaluation by strict examiners and liberal examiners). As scaling was a recognized method to bring raw marks in different subjects to a common scale and as the Commission submitted that they introduced scaling after a scientific study by experts, this Court apparently did not want to interfere. This Court was also being conscious that any new method, when introduced, required corrections and adjustments from time to time and should not be rejected at the

threshold as unworkable. But we have found after an examination of the manner in which scaling system has been introduced and the effect thereof on the present examination, that the system is not suitable. We have also concluded that there was no proper or adequate study before introduction of scaling and the scaling system which is primarily intended for preparing a common merit list in regard to candidates who take examinations in different optional subjects, has been inappropriately and mechanically applied to a situation where the need is to eliminate examiner variability on account of strict/liberal valuation. We have found that the scaling system adopted by the Commission leads to irrational results, and does not offer a solution for examiner variability arising from strict/liberal examiners. Therefore, it can be said that neither of the two assumptions made in S.C. Dixit can validly continue to apply to the type of examination with which we are concerned. We are therefore of the view that the approval of the scaling system in S.C. Dixit is no longer valid.

38. Learned counsel for the Commission contended that scaling has been accepted as a standard method of evaluation in the following decisions and therefore it should be approved

:- (i) *Kamlesh Haribhai Goradia vs. Union of India*<sup>6</sup> upheld by this Court by order dated

11.3.1987 in SLP (C) No. 14000/1986

(ii) *Muhesh Kumar Khandelwal vs. State of Rajasthan*<sup>7</sup> upheld by this Court by order dated 22.1.1996 in SLP(c) No. 15682-15684 of 1994.

(iii) *K. Channegowda vs. Karnataka Public Service Commission*<sup>8</sup>

All the three cases related to moderation and not scaling. There are, however, passing references to scaling as one of the methods to achieve common standard of assessment. The fact that scaling is a standard method of assessment, when a common base has to be found for comparative assessment of candidates taking examinations in different optional subjects, is not in dispute. In fact the Commission may continue to adopt the said system of scaling, where a comparative assessment is to be made of candidates having option to take different subjects. The question is whether scaling, in particular, linear standard scaling system as adopted by the Commission, is a suitable process to eliminate 'examiner variability' when different examiners assess the answer scripts relating to the same subject. None of the three decisions is of any assistance to approve the use of method of 'scaling' used by the Commission.

39. Learned counsel for the Commission also referred to several decisions in support of its contention that courts will be slow to interfere with matters affecting policy requiring technical expertise and leave them for decision of experts. (*State of U.P. v. Renusagar Power Co. Ltd.*<sup>8</sup> - *Tata Iron & Steel Co. Ltd. v. Union of India*<sup>9</sup> *Federation of Railway Officers Association v. Union of India*<sup>10</sup>) There can be no doubt about the said principle. But manifest arbitrariness and irrationality is an exception to the said principle. Therefore, the said decisions are of no avail.

40. We should, however, record the fair submission on behalf of the Commission that it is not irrevocably committed to any particular system and will adopt a different or better system if the present system is found to be defective.

Re : Point No. (iv).

41. The petitioners have requested that their petitions should be treated as being in public interest and the entire selection process in regard to Civil Judge (Junior Division) Examination, 2003 should be set aside. We are unable to accept the said contention. What has been made out is certain inherent defects of a particular scaling system when applied to the selection process of the Civil Judges (Junior Division) where the problem is one of examiner variability (strict/liberal examiners). Neither mala fides nor any other irregularities in the process of selection is made out. The Commission has acted bona fide in proceeding with the selection and neither the High Court nor the State Government had any grievance in regard to selections. In fact, the scaling system applied had the seal of approval of this Court in regard to the previous selection in S.C. Dixit (supra). The selected candidates have also been appointed and functioning as Judicial Officers. Further as noticed above, the scaling system adopted by the Commission has led to irrational and arbitrary results only in cases falling at the ends of the spectrum, and by and large did not affect the major portion of the selection. We, therefore, direct that our decision holding that the scaling system adopted by the Commission is unsuited in regard to Civil Judge (Junior Division) Examination and directing moderation, will be prospective in its application and will not affect the selections and appointments already made in pursuance of the 2003 Examination.

42. However, in so far as the petitioners are concerned, we deem it proper to issue the following directions to do complete justice on the facts of the case :

“a) If the aggregate of raw marks in the written examination and the marks in the interview of any petitioner is less than that of the last selected candidate in the respective category, he will not be entitled to any relief (for example, the petitioners in WP(C) No. 165/2005 belonging to the Category 'BC' have secured raw marks of 361 and 377 respectively in the written examinations, whereas the last five of the selected candidates in that category have secured raw marks of 390, 391, 397, 438 and 428 respectively. Even after adding the interview marks, the marks of the petitioners in W.P. [C] No.165/2005 is less than the marks of the selected candidates).

b) Where the aggregate of raw marks in the written examination and the interview marks of any petitioner, is more than the aggregate of the raw marks in the written examination and interview marks of the last selected candidate in his category, he shall be considered for appointment in the respective category by counting his appointment against future vacancies. (For example, we find that petitioner Archana Rani, one of the petitioners in WP (C) No. 467/2005 has secured 384 raw marks which is more than the raw marks secured by the last five selected candidates [347, 337, 336, 383 and 335] under the SC category and even after adding the interview marks, her marks are more than the five selected candidates. Hence, she should be

considered for appointment). This relief will be available only to such of the petitioners who have approached this Court and the High Court before 31st August, 2005.

43. The petitions are allowed in part accordingly.

Judgment Referred.

<sup>1</sup>(2003) 12 SCC 0701

<sup>2</sup>(2002) 4 SCC 0388

<sup>3</sup>(2000) 4 SCC 0640

<sup>4</sup>(2002) 7 SCC 0273

<sup>5</sup>(1992) Supp. 1 SCC 0323

<sup>6</sup>(1987) 1 Guj.LR 0157

<sup>7</sup>(1994) 1 Raj.LR 0533

<sup>8</sup>(1988) 4) SCC 0059

<sup>9</sup>(1996) 9 SCC 0709

<sup>10</sup>(2003) 4 SCC 0289