

\* \* \* \* \*

CLARENCE YORK,  
 Appellant,

v.

Administrator, DIVISION OF  
 PERSONNEL,  
 Respondent.

Case No. 78-42-PC

\* \* \* \* \*

OFFICIAL

DECISION  
AND  
ORDER

NATURE OF THE CASE

This is an appeal pursuant to §230.44(1)(a), Stats., of an examination.

FINDINGS OF FACT

1. During the week ending November 5, 1977, Jack Lawton, a Personnel Specialist 5 with the Division of Personnel, was notified by Robert Barnes of the Department of Transportation personnel office that there were three sergeant vacancies in the state patrol and that the department wanted the positions filled as soon as possible.

2. Mr. Lawton has been employed by the respondent and its predecessor agency, the Bureau of Personnel, and involved in examination development and analysis since July, 1974. He has a Bachelor's Degree in Psychology, a semester of work on a Masters Degree in Public Administration, and has attended conferences and workshops on content validity and police selection.

3. On November 4, 1977, Mr. Lawton met with Jerome Blied, George Ryan, and Gerald Johnson who had been selected as job experts

with respect to the sergeants' exam.

4. All of the aforesaid job experts were employees of the State Patrol, Division of Enforcement and Inspection, Department of Transportation. By way of background:

- A. Mr. Blied was a field staff captain with about 21 years of experience in the state patrol including about four years as sergeant.
- B. Mr. Ryan was a planning analyst with about 21 years of experience in the State Patrol including about 11 years as a sergeant.
- C. Mr. Johnson was a sergeant with about 19 years of experience in the State Patrol including about 9 years as a sergeant. At the time of serving as job expert, his assignment included coordination between the Division of Enforcement and Inspection and DOT personnel.

5. At this meeting the job experts under the direction of Mr. Lawton, verified the accuracy of the pre-existing position description and identified what they believed to be the high importance job content, or those knowledges, skills, and abilities which they felt were important for immediate performance by newly-hired sergeants.

6. Also at this meeting the job experts reviewed for job-relatedness a number of multiple choice questions presented by Mr. Lawton, which included questions that had been used on the previous sergeant's exam in 1976 and approximately 60 additional questions.

7. The job experts rejected the additional questions and then evaluated the 81 old questions on item rating forms (Respondent's

Exhibit 4A-C), rating each question on a scale of 1-5. A one rating indicated that "the specific knowledge, skill, or ability measured by this item is CONTINUALLY NECESSARY for successful work performance" while at the opposite extreme a five rating indicated that the item "is either unrelated to any of the listed knowledges, skills or abilities is obsolete or outdated - or is poorly worded."

8. Shortly before the meeting Mr. Lawton had reviewed the 81 items and had determined that 13 of the questions should be excluded because they did not meet either the respondent's internal or federal testing guidelines. These items related to spelling, grammar, and verbal reasoning. Mr. Lawton felt that the items were insufficiently job-related and were too easily memorized and that there might be some retention from the previous exam.

9. The job experts were not informed prior to doing the item review that Mr. Lawton had decided to reject certain of the items.

10. As a result of the evaluation by the job experts of these items, four additional items were selected for exclusion, including two items which had already been selected for exclusion by Mr. Lawton, around December 1, 1977.

11. The respondent's policy with respect to item review of multiple choice items with a small number of raters such as was the case here is that normally a 5 rating by any of the raters would result in exclusion of the item, while an average rating of 4 indicates that there is a serious problem with the item.

12. Item 72 was rated 5 by all three raters but was neither excluded from the exam nor rejected from the scoring following analysis

of the exam results. Item 80 received a "2" from two raters and a "4" from the third rater but also remained on the exam and counted as part of the final score.

13. As of about December 1, 1977, it was determined that there were insufficient multiple choice items for a complete exam, and the respondent informed the Department of Transportation that it could develop additional multiple choice items or essay-type questions.

14. During the first week in December 1977, DOT decided it would proceed to develop essay questions.

15. Sometime in late December 1977 or early January 1978, Captain Blied, Sergeant Johnson, and Mr. Ryan met several times, along with, on some occasions, Major Lacke, to develop the essay questions and bench mark answers.

16. Major Lacke, director of the Bureau of Support Services, which includes the personnel function, has been employed by the State Patrol since approximately 1957, and was a sergeant for approximately 3 years.

17. Major Lacke participated in the development of the essay questions and in the development of two of the three bench mark answers.

18. The aforesaid job experts had with them the identified and weighted compilation of knowledges, skills, and abilities that had been developed at the November 4, 1977, meeting.

19. The three essay questions which ultimately were included on the examination were developed by these job experts in these sessions and as they referred to the aforesaid compilation of knowledges,

skills and abilities.

20. The job experts then developed the bench mark answers which represented, in their opinion, the best to the worst responses that could be made to each question.

21. Mr. Lawton picked up the essay questions from DOT on December 16, 1977. He contacted the divisional paraprofessional staff on that day or the next day and was told that at least partly because of the press of business associated with the January exam date, there would not be sufficient time to retype the 1976 exam with the excluded items deleted. He then made the decision to have the 1976 exam booklet reprinted with all the items included, to avoid delaying the administration of the exam.

22. The answer sheet for the multiple choice items was a standard printed form used for all multiple choice examinations which was designed to be machine-gradable. To have deleted some items manually or by instructions would have created the possibility of some confusion inasmuch as the answer sheet had consecutively numbered spaces for the answers.

23. The documentation that was submitted by Mr. Lawton to provide instructions to the printer contained a form which indicated which of the multiple choice items had been determined should not be scored.

24. Shortly after receiving a copy of the essay questions and prior to the printing of the exam, Mr. Lawton received a verbal summary of the bench mark answers. He also received a copy of the written bench mark answers around February 1, 1978, before the exam

was scored.

25. Mr. Lawton reviewed the essay questions and bench marks, and made a determination that the bench marks were useable and that a reliable score could be drawn from the bench marks. He did not question the ranks assigned to the bench marks or the content of the bench marks. This review was conducted prior to the grading of the exam.

26. The examinees were not told prior to the examination what the relative weights would be of the multiple-choice and the essay portions of the exam.

27. All 81 of the multiple-choice items were included on the examination to be completed by the examinees, and they were not informed prior to the examination that some of the items would not be counted in the final score or should not be completed by the examinees.

28. The examination was administered on January 7, 1978, to approximately 85 applicants, including appellant. The appellant had an adequate amount of time to take the examination and completed it before the end of the allotted time.

29. After the administration of the examination the answer sheets were returned to the State Division of Personnel for scoring.

30. Sergeants Nash and Johnson graded the essay questions on February 9 and 10, 1978.

31. Sergeant Nash has been employed by the State Patrol a little over 16 years and has been a sergeant since July 1973.

32. With respect to the answers to the essay questions, numbers

were placed on the answer booklets in place of names and the scorers did not know the identity of the examinees.

33. When Nash and Johnson graded the essay exams they had no knowledge of how the examinees had done on the multiple-choice portion of the exam.

34. In grading the essays, the answers were evaluated on a scale of 9-0 according to the bench marks as much as possible. Where the answer did not fit the bench marks at all, the graders evaluated the question on the basis of their evaluation of the answer based on their knowledge and experience.

35. With respect to the first essay question, one of the bench marks identified three subparts (a, b, and c) for a desirable answer (9 points). The bench mark worth 7 points was associated with an answer giving only subpart a while the bench mark with 6 points was associated with an answer giving either subpart a or b or c.

36. With respect to the first essay question on the exam, Sergeant Nash felt it was difficult but not impossible to grade a three-part question with only one bench mark answer.

37. On the third essay question, Sergeant Johnson's interpretation of requiring that complaints be put in writing as a poorer answer than rationalizing away the officer's action was that the former response would be equivalent to a brush-off, e.g., for the sergeant to say that the complaint could not be looked into or responded to at all unless it were filed in writing.

38. Sergeant Johnson defended ranking of bench marks for essay question two on the basis of his opinion of patrol policy.

39. The appellant received a grade of 3 on the third essay question from one grader.<sup>1</sup> The second grader failed to record a numerical score but wrote the comment "Did not understand reason for investigating complaints."

40. The paraprofessional recording the scores recorded, by interpolation, a 3 as the appellant's score from the second grader on the third answer.

41. The third essay question had a relatively high inter-rater reliability as determined by mathematical analysis, and the interpolation was valid.

42. The reliability figure for the first two essay questions indicated that this part of the test was a reliable measuring device.

43. The multiple-choice portion of the exam was scored by computer.

44. The respondent's policy is that it is permissible, under appropriate circumstances, to delete questions from scoring at any time in the examination process, including after the examination has been administered.

45. Two of the multiple-choice items were deleted by Mr. Lawton after the examination following a computer analysis which lead to a determination that they did not contribute to the overall reliability of the examination.

46. The appellant's rank on the multiple-choice portion of the exam would have been the same regardless of whether all the questions were scored or only the 62 questions that were not excluded.

47. Sometime in the end of February or the beginning of March,

---

<sup>1</sup> The appellant received scores of a 9 from both graders on the first and second essay questions.



1978, and prior to March 7, 1978, the decision was made by Mr. Lawton with the knowledge of DOT to weight the multiple choice part of the written exam the same as the essay portion.

48. The decision to weight equally the oral and written parts of the exam was made in the first instance by DOT and approved by the respondent. This decision was made at the outset of the development of the examination.

49. The decision as to where to establish the passing point on an exam usually is based by the Division of Personnel on a number of factors, including the needs of the appointing authority and the distribution of the scores, e.g., if there is a natural break in the scores, this would be a factor that would favor setting the passing point there.

50. In this case there was a natural break in the scores at the average raw score (86) and this was recommended to DOT. This would have resulted in 48 examinees passing with 17 of those above a higher cutoff point and eligible to take the oral exam.

51. When Major Lacke heard that it had been determined that approximately 17 people would be invited to take the oral exam he was very concerned that this number was insufficient, based on the number of vacancies, the anticipation that there would be additional vacancies, and the anticipation, based in part on prior problems, that persons on the register would be reluctant and would refuse to accept promotions to sergeant because of their concerns about relocation and inadequate financial remuneration. He presented these concerns to the DOT personnel management who also were concerned that the examinees

would be upset at not being allowed to proceed with the oral after having passed the written portion of the exam.

52. DOT requested of the respondent that more people be allowed to proceed to the oral exam and that the cutoff point for the oral be made the same as the passing point so that all examinees who received a passing grade on the written portion of the exam would be permitted to proceed to the oral.

53. The respondent then determined to set the passing point and the cutoff point at a raw score of 91. This score produced the maximum number (34) of candidates for the oral exam that DOT felt that it could examine efficaciously.

54. The appellant received a passing grade on the written portion of the exam and was allowed to take the oral portion of the exam.

55. The appellant's was the last name on the register that was not certified.

56. The appellant's civil service score was 0.42 less than the person on the register ahead of him who was in fact certified.

57. At the time of the written examination there were four vacancies for sergeants in the State Patrol.

58. Ultimately there were five vacancies and five persons were appointed from among those certified.

59. With respect to protection against the manipulation of the exam by way of the elimination of a question or questions after the exam to help or hurt a candidate or candidates, safeguards include the number of people involved in the examination administration and scoring

process and the concern by someone who might consider such manipulation that someone else in the process might discover the manipulation; the anonymity of examinees resulting from the fact that the item analysis conducted with respect to the multiple-choice items after the examination utilizes statistics based on the total group of scores, so that an individual's test scores could be influenced only if the personnel specialist involved in the item analysis could ascertain in what areas an applicant did or would have been likely to score well or poorly on; and the criminal penalties provided by statute for illegal manipulation of the test process.

60. The examination was job-related in compliance with appropriate validation standards and was administered, including rating the results and determining the relative ratings of the competitors, in accordance with appropriate scientific techniques and procedures.

#### CONCLUSIONS OF LAW

1. This matter is properly before the Commission pursuant to §230.44(1)(a), Wis. Stats. (1977).
2. The burden of proof is on the appellant to establish that the examination was not conducted in accordance with §230.16 Stats. (1977)
3. The appellant has not sustained that burden.
4. The examination was conducted in accordance with §230.16, Stats. (1977).

OPINION

The Bill of Particulars submitted by the appellant (Commission's Exhibit 3), raises the following grounds of error with respect to the examination in question:

"1. Appellant objects to and complains as to any and all changes made in the content and or format of the January 7, 1978 State Patrol Sergeant CP-Transportation exam given by the Bureau of Personnel, Department of Administration which were made after the date said exam was validated or after the date said exam was administered.

2. Appellant objects to the method and procedures used in designing and grading the essay portion of the exam referred to in Paragraph No. 1."

Section 230.16, Stats. (1977), provides, in part:

(4) All examinations ... shall be job-related in compliance with appropriate validation standards ....

(5) The administrator shall utilize appropriate scientific techniques and procedures in administering the selection process, in rating the results of examinations and in determining the relative ratings of the competitors."

One of the appellant's primary focuses is on the actions that were taken by the respondent with respect to the exclusion of certain of the multiple-choice items from the scoring of the exam, both before and after it was administered.

The items selected for exclusion before the exam were rated negatively by the job experts or the exam expert. The job experts had an impressive background of relevant experience. The reasons advanced by the exam expert were relatively convincing. The items that were eliminated on the basis of item analysis after the exam were determined by statistical analysis to have been unreliable. In the opinion of the Commission, the exclusion of all these items contributed to the validity of the examination. It cannot

be said that their exclusion was erroneous.

The appellant went to some lengths to attempt to argue that the respondent could and should have deleted the questions from the exam booklet before the test was administered. The respondent's explanation for why this was not done was at least reasonable. On the other hand, there is nothing in the record to indicate that there was not adequate time provided to take the exam or that the appellant did not have plenty of time to finish the entire exam, including all the multiple-choice items. The failure to eliminate these items prior to the administration of the exam was not erroneous.

The appellant's argument that the examinee should have had notice before the written exam of the relative weight of its parts also is unconvincing in light of the adequate time provided.

It is clear that at least one and possibly two multiple-choice items were not eliminated that should have been, based on the evaluation by the job experts. Since the Bill of Particulars objected only to changes that were made in the exam, this point, in and of itself, is outside the scope of the issues. However, while not cognizable as a separate issue, it is probative as to the contentions concerning the degree of precision in the exam processes.

The appellant also argues that the changes in the exam content and the flexibility and timing of the decisions on the relative weights of the various parts of the exam created the potential for possible manipulation of the scores to help or hurt particular candidates. However, it was not argued that such manipulation in fact occurred.

After the exam had been given, deletions of multiple-choice items were made by a testing expert from the State Division of Personnel, which of course is completely separate from DOT. Manipulation would require that that person ascertain the raw score of the condidate or candidates to be affected and then select items for elimination that would have the desired effect. Such actions would have to be taken in the knowledge that such action is subject to criminal sanctions and that the decision to eliminate test items has to be supported on the basis of mathematical analysis to be sustainable. The decisions on the weight to be accorded various parts of the exam also had to be made or concurred in by the test expert, subject to similar restrictions. Undoubtedly, more extensive safeguards could have been utilized, but the Commission cannot conclude that the process used here did not comply with the requirements of §230.16, Stats. (1977).

The appellant objects to the content of certain of the bench mark answers to the essay questions, which are set forth in Respondent's Exhibit 2. It is argued that some of these answers are clearly ridiculous or offend common sense.

These bench mark answers were developed by a panel of well-qualified job experts. The Commission agrees that if the content of particular bench marks were clearly ridiculous or clearly offensive to common sense that they would not be valid. There are many other ways that the validity of such bench marks could be attacked, such as, for example, demonstrating that they conflict with actual State Patrol policy.

The Commission cannot conclude that any of the bench marks are

of the "clearly ridiculous" variety. While the Commission might well disagree as a matter of the Commission's own ideas of program management with some aspects of the bench marks, such disagreement can not constitute the basis for a conclusion of invalidity. In addition, it is noted that essay exams by their nature are more flexible than multiple-choice. The bench marks cannot possibly cover all answers precisely and the graders have to use some individual judgment. The graders interpreted some of the bench marks in a manner that avoided to some extent some of the appellant's criticisms. See, for example, finding 37.

The Commission is of the opinion that part of the bench marks for the first question which rated an answer containing part (a) of the best answer at a 7 and an answer containing part (a), (b), or (c) as 6, was not an appropriate measuring device. A candidate could give exactly the same answer and get either a 6 or a 7. However, there is no indication that this aspect of the bench mark resulted in a low reliability figure for this question or affected adversely the overall validity of the exam.

One of the raters failed to assign a numerical score to the appellant's answer to the third essay question. The paraprofessional who recorded the scores interpolated a "3" based on the other rater's score of "3." At this time, the Commission cannot conclude that this was error in light of the facts that the mathematical analysis showed a high degree of correlation between the two raters on this question and that the reater's recorded comment on the answer indicated a low score was called for.

Although the record reflects some deficiencies in the administration of this examination, particularly with respect to the benchmarks for the first essay question, in the opinion of the Commission the process must be reviewed as a whole to determine if the administrator has complied with statutory requirements. For example, if one multiple-choice item out of 62 were found to be defective and the multiple-choice items were weighted to constitute 25% of the entire exam, it may well be that that item would not affect the validity of the exam as a whole.

In this case the testing expert employed by respondent who administered the exam testified as to how the exam was developed and administered and testified that in his opinion the Division operated in accordance with professional testing standards. Well-qualified job experts were consulted to develop the high importance job content and the content of all questions and answers. The elimination of items and weighting of the parts of the exam either were done or approved by the exam expert. A mathematical analysis was accomplished to identify and eliminate unreliable multiple-choice items. Another mathematical analysis was accomplished to ascertain the reliability of the essay questions. Under all of these circumstances and considering the entire record the Commission cannot conclude on the basis of the errors that were established that the respondent violated §230.16, Stats., by changes made in the content and format of the exam, or in the design and grading of the essay portion.



ORDER

The actions of the administrator are affirmed and this appeal is dismissed.

Dated: July 18, 1980.

STATE PERSONNEL COMMISSION

Charlotte M. Higbee  
Charlotte M. Higbee  
Commissioner

Donald R. Murphy  
Donald R. Murphy  
Commissioner

Gordon H. Brehm  
Gordon H. Brehm  
Commissioner

AJT:jmg