



State of Louisiana

OFFICE OF

STATE INSPECTOR GENERAL

**MUNICIPAL FIRE AND POLICE CIVIL SERVICE
OFFICE OF STATE EXAMINER**

Report by

Inspector General Bill Lynch

Prepared for

Governor M.J. "Mike" Foster, Jr.

March 15, 1999

File No. 1-99-0038



State of Louisiana

OFFICE OF
STATE INSPECTOR GENERAL

Municipal Fire and Police Civil Service
Office of State Examiner

February 23, 1999

Report by

A handwritten signature in cursive script, appearing to read "BL", written over a horizontal line.

Inspector General Bill Lynch

Approved by

A handwritten signature in cursive script, appearing to read "MJC", written over a horizontal line.

Governor M.J. "Mike" Foster, Jr.

File No. 1-99-0038

Municipal Fire and Police Civil Service Office of State Examiner

Examinations for firemen and policemen, administered by the Office of State Examiner, Municipal Fire and Police Civil Service, have been routinely altered in the grading process providing an advantage to some candidates who have failed and to the detriment of those who passed. The process for deleting questions and allowing dual credits for others gives the appearance of allowing the Examiner's office to change any candidate's failing score to a passing score virtually at will.

Paul Daly, the state examiner, denied that the practice was aimed at helping any particular individuals and no direct evidence was found to confirm that it was. However, the potential for manipulation is compounded by the fact that reviewers had the names and original scores of the test takers as they were deciding which questions to delete.

Mr. Daly acknowledged that there may be the perception of an opportunity for bias when examined by an outside party, but emphasized that the only purpose for the process is to insure that the agency, prior to the grades being finalized, has not inadvertently disqualified someone who deserved to have a passing grade. He insisted that the program has been conducted with integrity and professionalism. He also stated that the agency is taking steps to eliminate problems found in the review.

Background

The Office of State Examiner of Municipal Fire and Police Civil Service was constitutionally created to provide a general personnel system for firefighters and police officers in all municipalities in the state having populations of 7,000 to 400,000 and in all parish fire departments and fire protection districts, based upon merit, efficiency, fitness, and length of service. As of March, 1998, there were 92 jurisdictions, each with a local civil service board, representing 7,300 employees covered under the system.

The agency is within the Department of Civil Service, but is essentially autonomous since the State Civil Service Commission is prohibited by law from exercising administrative control over the state examiner and the deputy state examiner, two classified civil servant

who are responsible for managing the agency. The state examiner has a unique position in that he is not supervised by anybody or any board.

Test Development and Administration

One of the primary duties of the agency is to administer tests for original entrance and promotion to applicants for positions in the respective classified service of the municipalities. This includes developing the tests, scoring the tests, and submitting the results to the local board for which the tests are given.

The agency administers approximately 500 tests and develops about 250 new tests annually. The agency administers tests on either a competitive or promotional basis with the law providing that 75% shall constitute a passing score for all tests administered.

The appointing authority of the local jurisdiction for which the test was administered may select anyone from a test score, regardless of ranking, list to fill a competitive vacancy. A promotional vacancy must be filled by the person having the greatest seniority and passing the test. Each successive vacancy must be filled with the next highest seniority person.

1. Development

The agency constructs two basic types of tests, standard exams and non-standard exams. Standard exams are developed to use in multiple jurisdictions for the same job class. These tests are typically used for entry-level positions such as firefighter or police officer. Non-standard exams are custom tailored by job class for a specific jurisdiction. For example, a different fire chief exam would be developed for use in Alexandria than would be used in Houma. Most non entry-level classes are tested using non-standard exams.

All tests use multiple-choice questions, but some also include a practical component in which the candidate (test taker) may be required to give an oral presentation or prepare a written exercise. The scope of this report is limited to discussion of the multiple-choice component.

According to Mr. Daly, the test development process is done in accordance with federal and industry guidelines as found in the Uniform Guidelines on Employee Selection Procedures and Principles for the Validation and Use of Personnel Selection Procedures.

The process begins with his agency conducting a job analysis, which includes determining what tasks are performed in the job by sending questionnaires to incumbents and verifying this information with first-line supervisors and the chief of the department. The agency then designs an exam plan from the information which identifies the different areas of skill, e.g., administration, traffic control, criminal investigation, etc., that will be tested. Using the plan, the agency develops a test by selecting multiple-choice questions from a large data base it maintains, which target the knowledge, skills, and abilities needed for that position in that jurisdiction. The test development is assigned to one of the agency's human resource program consultants.

Then the proposed test goes through an extensive review process which includes peer review by another program consultant, review and approval by Dennis Bartlett, the agency's human resources program consultant supervisor, and final review and approval by Melinda Livingston, the deputy state examiner.

Mr. Daly further explained that a similar but more extensive validation process is used for the standard exams since they are used in multiple jurisdictions. He said it can take over a year to develop these exams and the agency contracts industrial psychologists to assist.

2. Administering

The agency assigns one of its human resources program consultants to administer the test at the jurisdiction site. The candidates mark their multiple-choice answers on a Scantron form. The program consultant secures the test and the Scantron forms and returns them to the agency for grading by clerical staff.

3. Grading and Grade Reviewing

Clerical staff use computerized equipment to grade the forms.

Agency policy provides that a grade review will be done on all tests except 1) competitive standard exams, 2) tests in which all candidates pass and 3) tests in which all the failures are below 64% unless all candidates fail.

On those tests meeting the grade review criteria, the clerical staff generates a Frequency Report showing each question and the percentage of candidates for that jurisdiction who answered the question correctly. The Frequency Report also shows each of the answer choices and the number of candidates choosing each one.

The policy further provides that the clerical staff identify those questions with a passing rate of 50% or below, considered low performing items by the agency, and post the question number along with passing statistics to a Review Form. Clerical also posts statistics showing the percentage of candidates previously answering the question correctly statewide when the question was used in other jurisdictions for the same job class.

The Review Form and the Frequency Report are subsequently given to Fred Dressel, a human resources program consultant assigned responsibility for reviewing the low performing items. According to Mr. Dressel, he analyzes the questions and statistics, checks the reference source to make sure the information is correct, checks to make sure the item was keyed correctly when graded, and basically looks to see if there is any flaw in the question.

Mr. Daly said in addition to these reasons there is a multitude of other reasons for taking action on a question. Some of these include the wrong question being included on the test, departmental policy varying from the authoritative source, and a question being identified as needing revision after the test booklets are printed.

After Mr. Dressel has reviewed the question he will recommend either no action be taken or the question be deleted from the test or credit be given for more than one answer (dual credit). His recommendations and justifications are then posted to the Review Form.

Mr. Bartlett reviews Mr. Dressel's recommendations and then Mr. Daly gives final approval. Sometimes Ms. Livingston is involved in the review and approval process also. If questions are deleted or dual credit is given, clerical re-grades all candidates' papers with some candidates going from failing to passing scores. Ms. Livingston said that if the deletion process results in a candidate's raw score going from passing to failing, the agency does not penalize the candidate and assigns a passing score of 75.

4. Submission

After scores are finalized, they are reported to the local civil service board for its approval. The scores do not become official until the board approves them.

Opportunity for Manipulation

A review of 18 targeted test files for the period January, 1996, to December, 1998, revealed that the agency's system for reviewing questions and altering test grades exposes the system to potential manipulation for the benefit of targeted individuals. Of the 18 targeted test files reviewed, there were 9 tests in which candidates benefited after the agency applied its grade review process. In these 9 tests, 11 candidates, whose preliminary scores were failing, were subsequently assigned passing scores after questions were deleted or dual credit was allowed. Of these 11 candidates, 4 were subsequently promoted to the job class for which the test was administered.

Mr. Daly estimated that tests are re-graded 30% of the time and Mr. Bartlett estimated that 50% of the tests, excluding entry-level exams, are affected by the process. Mr. Daly later said that an analysis by his office revealed that for the period July 1, 1998, to Dec. 17, 1998, that 19 out of 199 or 9.5% of the total exams administered had revisions during the grading process. His statistics did not exclude entry-level exams, on which answers are rarely changed because the tests are standardized.

When questioned about the prevalent re-grading considering the extensive test validation process, Mr. Daly stated that the deletions are mostly on the non-standard tests and the agency tries to do the best it can given the available resources.

There is an obvious disadvantage for those persons who passed the test as originally presented when additional persons are added to the passing category. In the case of competitive exams, the initial passers face additional competition. In the case of a promotional exam, the initial passers may be displaced by someone with more seniority.

The primary problem with the agency's policy is that the individuals with review and approval authority over subsequent test grade changes are given unnecessary information regarding the identity of the individuals and their respective test performance. While conducting the grade review, the four agency personnel involved were provided information that could have given each of them the opportunity to manipulate the process to insure that certain candidates pass the tests.

Agency policy requires that a preliminary grade sheet be generated which list all candidates' names, social security numbers, sex, race, and scores on the multiple-choice. Policy also provides that clerical generate an audit report on each candidate making a grade of 72

through 76. The audit report shows the candidate's name, his answer to each multiple-choice question and whether his answer was correct or wrong. The reviewers confirmed that the preliminary grade sheet and audit reports are provided to them when conducting the grade review.

Mr. Dressel acknowledged that he sometimes obtained seniority information on candidates when performing the grade review. Mr. Dressel said he used the information as a tool by seeing if a person with a lot of seniority scored low on the test. When asked on which tests he chose to review seniority information, he said he had no specific criteria for selecting which seniority information he would review. However, in our opinion, using seniority information could provide an opportunity to manipulate the system for positions filled on a seniority basis. Additionally, seniority indicates length of time worked and not necessarily knowledge obtained.

According to an industrial psychologist, currently under contract to assist the agency in an unrelated area, the process should be done blindly where the reviewers do not see candidate's names and do not have seniority information.

In a letter dated Jan. 15, 1999, to IG auditors, Mr. Daly agreed that there might be the appearance of an opportunity for bias when an outside party examines the process. The letter advised that the agency had changed several policies and procedures to avoid any future perception that there is an opportunity for bias. The revised policies and procedures provide that the preliminary grade sheets and audit reports will no longer include names. However, he said the agency will continue to provide the reviewers with the social security number, race, and sex. The letter also advised that the agency had already eliminated the practice of reviewing seniority information during the grade review process several months prior to the IG audit.

We question the wisdom of continuing to use social security number, race, and sex information in determining the validity of questions for specific examinations. This information can still potentially identify the candidate to the reviewers. We also question the need for the reviewers to even be provided the preliminary grade sheet and the audit reports. The Review Form and the Frequency Report alone provide the necessary information for the reviewers to identify low-performing questions requiring further analysis.

We found no direct evidence that changes were intended to help favored individuals. However, in some instances the circumstances created the appearance that changes were made in order to give selected individuals passing grades. Regardless of whether changes were influenced by intentional favoritism, the appearance of favoritism impairs the integrity of the testing process.

Examples of grade changes on the nine examinations with grade alterations where perception may have raised questions about the agency's actions are detailed as follows:

Fire Chief Test

In one extreme instance involving a fire chief test administered to 8 candidates in 1997, the agency altered 26 out of 148 questions after the examination, eliminating 16 and accepting more than one of the multiple-choice answers on the other 10. The result was that three persons who had failed the test originally were added to the two who had passed.

This particular example reflects a number of problems created in the agency's review process detailed as follows:

1. This examination was a competitive test. The two persons who passed initially were disadvantaged by having three more persons with whom to compete.
2. The alteration of so many questions gave the appearance that this was done to insure passage by a particular person, even though that person eventually did not get the job. This person benefited from 23 of the 26 questions altered. In this case the three persons who were moved from failing to a passing grade of at least 75, went from a 71 to a 78, a 70 to a 79 and a 66 to a 77.

The person with a 66 was the acting fire chief at the time. Ironically, one of the three other persons who failed the test despite the re-grading, initially was two points above the acting fire chief who was moved to passing. A second person had a 66 also initially and a third had 61, but none was passed on the re-grading.

3. The reasons for altering a question are stated on the Review Form, some of them citing only statistics on performance. Ms. Livingston said that the explanation given may not have been as comprehensive as it might have been under ideal circumstances, but that the Review Form was always intended as an internal document. She said that statistics alone were not used as the basis for making decisions on questions. She added that statistics are certainly tools which are useful in making informed review decisions, but one must consider the substance of the questions in any review process.
4. Normally questions answered correctly by more than 50 per cent of the candidates are not reviewed. However, there may be exceptions, according to Ms. Livingston. In this test six of the questions above the 50 percent mark were among the 26 altered. Ms. Livingston said this was because the test was considered too difficult.

Promotional Exams

Of the nine examinations in which there were alterations of questions, eight were promotional tests in which persons were promoted on the basis of seniority if they passed the test.

All of the promotional exams had far fewer changes in the questions than did the competitive test's 26 alterations. One of the promotional tests had seven questions altered and the rest a lesser number.

The following is an analysis of the eight promotional exams:

1. District Fire Chief Test

In this example involving a promotional district fire chief test administered to 15 candidates in 1997, the agency altered 2 questions out of 136 resulting in 1 candidate having his failing score of 74 changed to a passing score of 75. This disadvantaged 12 candidates who initially passed the test because the 1 candidate benefiting from the alterations had more seniority and was subsequently promoted to the position.

2. Police Lieutenant Test

In this example involving a promotional police lieutenant test administered to 13 candidates in 1997, the agency altered 2 questions out of 128 resulting in 1 candidate having his failing score of 74 changed to a passing score of 75. This disadvantaged 9 candidates who initially passed the test because the 1 candidate had more seniority and was the first to be promoted to the position. Of the 9, 3 candidates were subsequently promoted to the job class based on their seniority.

3. Assistant Chief of Police Test

In this example involving a promotional test administered in 1996, the agency altered 7 questions out of 140 which boosted the only candidate taking the test from a failing score of 71 to a passing score of 76. The candidate benefited from all 7 affected questions. The candidate was subsequently promoted to the position.

Even though there was only one candidate, it could be perceived that the seven questions were altered so that this one candidate passed.

4. Police Corporal Test

In this example involving a promotional test administered to 27 candidates in 1998, the agency altered 1 question out of 112 resulting in 1 candidate having his failing score of 74 changed to a passing score of 75. Of the 27 candidates, 17 initially passed the test. As of the completion of the audit work, no one had been promoted to the position.

5. Police Lieutenant Test

In this example involving a promotional test administered to 11 candidates in 1998, the agency altered 1 question out of 128 resulting in 1 candidate having his failing score of 74 changed to a passing score of 75. Of the 11 candidates, 6 initially passed the exam. As of the completion of the audit work, no one had been promoted to the position.

6. Assistant Fire Chief Test

In this example involving a promotional test administered to 2 candidates in 1996, the agency altered 1 question out of 140 resulting in 1 candidate having his failing score of 74 changed to a passing score of 75. The other candidate initially passed the test. Agency records show no one was promoted to the position and the promotional list expired.

7. Deputy Fire Chief Test

In this example involving a promotional test administered to 6 candidates in 1996, the agency altered 4 questions out of 142, which boosted the only failing candidate from a score of 73 to a passing score of 75. This candidate had initially missed all 4 of the deleted questions. Of the 5 candidates initially passing, 1 failed to be promoted because he was displaced by the person whose score was raised from failing to passing.

8. Police Captain Test

In this example involving a promotional test administered to 16 candidates in 1996, the agency altered 1 question out of 136 resulting in 1 candidate having

his failing score of 74 changed to a passing score of 75. Of the 16 candidates, 14 initially passed the test. The one candidate who benefited from the alteration was never promoted to the position. Of the 14 who initially passed, 12 were subsequently promoted to positions in this job class.

Conclusions:

1. The Office of State Examiner's policy and procedures for reviewing and changing test scores provide an opportunity for manipulation of test results to favor certain persons. By obtaining identifying information of the testing candidates along with the specific answers to test questions provided by these candidates, the agency's review team has the opportunity to change the test grades of any favored individual. However, no direct evidence was found to confirm that this was the case.
2. The Review Form does not always reflect the actual substantive reason for deleting a question.

Recommendations:

1. The Office of State Examiner should implement policies and procedures which will allow questions to be evaluated on their own merit without any information about the individuals taking the test.
2. The Review Form should reflect the actual substantive reason for deleting a question.

Management Response:

See attached.



State of Louisiana

OFFICE OF

STATE EXAMINER
MUNICIPAL FIRE AND POLICE CIVIL SERVICE

8401 United Plaza Blvd., Suite 280
P.O. Box 44366
Baton Rouge, Louisiana 70804-4366
(225) 925-4400
FAX (225) 925-4567

PAUL K. DALY
State Examiner

February 19, 1999

Mr. Bill Lynch
State Inspector General
P. O. Box 94095
Baton Rouge, LA 70804-9095

Re: File No. 1-99-0038

Dear Mr. Lynch:

The Office of State Examiner conducts reviews of test grades to accomplish one, well-defined mission: To insure that we do not inadvertently disqualify any candidate who possesses the minimum knowledge, skills, and abilities needed in order to begin a working test period. References are made in the report to "disadvantaging" candidates as a result of the grade review process. If we were to fail to review examinations and correct any errors that may be present as a result of inappropriate questions, this office would seriously fail to fulfill our obligation under the law. The Office of State Examiner has never manipulated test scores to benefit any individual or group of individuals. We have operated, and will continue to operate, the Office of State Examiner for Municipal Fire and Police Civil Service with the highest degree of honesty, integrity, and professionalism. Despite our limited resources and the high volume of tests we are required to develop and administer, we have often undertaken more extensive research and gone to greater lengths than other testing entities to insure the fairness of our testing program. One of these efforts has been our grade review process.

For a three month period of time, we cooperated with investigators from the Office of the Inspector General by providing unlimited access to our confidential files in order that all aspects of our testing program could be closely examined. In addition we instructed our staff to fully cooperate in answering any and all questions the investigators might have. During the course of the investigation, the investigators identified that there may be a perception that manipulation of scores could possibly occur, and their report contained the following two recommendations:

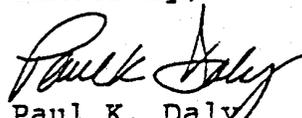
Mr. Bill Lynch, State Inspector General
February 19, 1999
Page 2

1. The Office of State Examiner should implement policies and procedures which will allow questions to be evaluated on their own merit without any information about the individuals taking the test.
2. The Review Form should reflect the actual substantive reason for deleting a question.

We agree that such changes should occur and have instituted changes in our policies in order that our review process is completely blind and our record for revisions is fully detailed (see enclosed policies). While we had previously removed the name from the draft score sheet, under our new policies the persons reviewing the examination results will also no longer be provided social security numbers, nor an identification by race, gender, or any other identifying information. In making this change, we are eliminating that which we feel is a beneficial tool: the use of past exam performance and race/sex information. This capability must, unfortunately, be a trade off to insure the elimination of any perceived bias. In addition, I have also instituted a policy of recording complete and detailed information about actions taken on the grading of specific questions as a result of an exam review in order that the information contained on the review document will create full understanding when reviewed by an outside observer.

While we feel that the methods and procedures employed in our grading of examinations have always been conducted with integrity and objectivity, it has been a beneficial experience for us to have an outside entity, such as your office, to critically evaluate our methods. In this regard, we appreciate the input and thoroughness of the investigators representing the Office of the Inspector General. We respectfully request that this official response to your report be included with and become a part of the report.

Sincerely,


Paul K. Daly
State Examiner

Enclosures

GUIDELINES FOR EXPLAINING THE ADJUSTMENTS TO AN EXAMINATION AS A
RESULT OF THE EXAM REVIEW PROCESS

1. DELETING AN ITEM FROM SCORING:

A. If, for any reason, an item has been removed from the item bank subsequent to approval of the final draft of the examination, and before the actual administration of the examination, the item will be deleted from scoring for that particular examination.

B. If an item is determined to be seriously flawed and/or has an extremely poor performance statewide, or has been repeatedly modified and revised and continues to draw an extremely poor performance, the item will be deleted from scoring for the examination.

C. If an item is found to be restricted for use at a higher or lower level than the examination being reviewed, or is restricted for use on specialized classes or in specific jurisdictions other than the one under review, or for any other legitimate criterion that would render the item inappropriate for the particular jurisdiction, the item will be deleted from scoring for that particular examination.

D. If an item has been significantly revised subsequent to approval of the final draft of the examination, and before the actual administration of the examination, the item will be deleted from scoring for that particular examination.

E. If an item cannot be adequately documented by a professionally recognized reference source, or if the information in the item has changed, the item will be deleted from scoring for that particular examination.

F. If an item is deemed to be unfair or inappropriate for an examination due to processing or administrative errors, the item will be deleted from scoring for that particular examination. Examples of such errors would be: (1) incorrect unique number entered on exam key (2) fire service item included on police service exam or vice versa (3) words omitted from premise or distractors (4) incorrect instructions/materials included with or omitted from examination (5) major grammatical or punctuation errors (6) confusing or awkward language or sentence construction

2. GIVING CREDIT FOR MORE THAN ONE ANSWER

If more than one answer can be adequately defended as a correct response based on any of the following: (1) citing from a recognized and reputable reference source (2) local departmental policy conflicts with standard (3) possible dual interpretation of wording of premise (4) change of information or procedure (5) change in applicable laws (6) logical reasoning based on premise, credit will be given for both answers in scoring the examination.

3. SUBSTANTIVE EXPLANATION AND DOCUMENTATION FOR ANY ADJUSTMENT TO EXAM

For each and every case where items have been deleted from scoring for an examination, or where credit for more than one answer has been recommended, a complete and comprehensive explanation will be offered for each adjustment. The substantive reasoning and documentation shall be sufficient to enable an outside observer to comprehend the rationale for the recommended change.

4. REVIEW ADJUSTMENT CAUSES A DRAFT GRADE TO GO FROM 75 OR GREATER TO LESS THAN 75

If an examinee had a draft score of 75 or greater, and subsequent review adjustments would cause that score to fall below 75, then the score would be reported as a 75.

GRADE REVIEW PROCESS

NEW POLICY

EFFECTIVE 02/15/99

In order to minimize the possibility and/or perception of bias with regard to the procedures used by this agency in the grades review process, the following guidelines have been implemented effective February 15, 1999:

1. The preliminary draft grade report submitted to the HRPC III for the initial review will consist ONLY OF THE PERCENTAGE SCORES ON THE EXAMINATION, beginning with the highest score and descending in numerical order to the lowest score. Administrative Support will also submit a "Review Form" of those items on the examination on which candidates as a group scored 50% or less. EFFECTIVE JANUARY 7, 1999, THE EXAMINEE'S NAME WAS REMOVED FROM THE PRELIMINARY DRAFT GRADE REPORT. EFFECTIVE FEBRUARY 15, 1999, IN ADDITION TO THE REMOVAL OF THE EXAMINEE'S NAME, THE SOCIAL SECURITY NUMBER, RACE, AND GENDER WILL ALSO BE REMOVED FROM THE PRELIMINARY DRAFT GRADE REPORT. THE PRELIMINARY DRAFT GRADE REPORT WILL CONSIST ONLY OF THE PERCENTAGE (%) SCORES.

2. During the review process, any recommendations by the HRPC III for the removal or modification of an item from that particular examination shall be accompanied with complete and substantive reasoning sufficient for an outside observer to comprehend the rationale for the recommended change.