

JEFFREY ALLEN, et al.,

Appellants,

v.

Administrator, DIVISION OF
MERIT RECRUITMENT AND SELECTION

Respondent.

Case No. 89-0124-PC

FINAL
DECISION
AND
ORDER

••ISSUED UNDER SEAL••

A proposed decision and order was issued in this matter on February 20, 1990. The respondent filed objections and requested oral arguments. A copy of the proposed decision and order is attached hereto. After hearing the parties' arguments, reviewing the written arguments filed by the parties and consulting with the hearing examiner, the Commission rejects certain portions of the proposed decision and order as noted below. The remaining portions of the proposed decision and order are adopted.

Finding of fact #14 is revised to read:

14. The scoring levels developed by the job experts did not explicitly take into account differences in size and scope of responsibility of the "organizations" employing the various applicants. However, the scoring levels were logical and were all clearly related to the five evaluation criteria.

Conclusions of law #3 and #4 are revised to read:

3. The appellants have failed to sustain that burden.
4. The examination was conducted in accordance with the civil service requirements.

That portion of the Discussion section of the proposed decision and order commencing with the second full paragraph on page 13 and ending before the second full paragraph on page 16 is replaced with the following language:

The appellants also argue that the application of the scoring system was inconsistent with the goal of accurately assessing the experience of the various applicants because the system failed to contemplate organizations of vastly different sizes and overemphasized mere status as a supervisor. In finding of fact #14, the Commission has noted that the scoring levels "did not explicitly take into account differences in size and scope of responsibility of the 'organizations' employing the various applicants." It may have been preferable if the scoring levels had made specific reference to the size of the employing entities. However, the Commission is unwilling, on this record, to second guess the judgment of the job experts who adopted scoring levels which were logical and were all clearly related to the five evaluation criteria. Those job experts did not testify and therefore could not explain why the scoring levels did not specifically refer to employer size. In addition, the job experts who scored the resumes did not testify and, therefore, could not explain why they assigned scores in a particular manner. The Commission cannot say, on this record, that the scoring system applied by the raters was invalid because of the emphasis it placed on supervisory experience, where the goal of the examination process was to develop a list of eligibles for filling vacancies in a supervisory classification.

The proposed decision hinged on a comparison of the scores assigned by the raters to one applicant whose resume (AE#12) reflected supervisory experience, and to certain other applicants whose resumes (AE#34 and 54) did not show any supervisory experience. One scorer appeared to have simply made a mistake by giving 0 points to applicant AE#34 for criterion E (Cash Management Reporting Functions--Invoice Billings and/or Letter of Credit Process) where the applicant's resume specifically mentioned letter of credit responsibilities. However, the other comparisons discussed in the proposed decision don't necessarily lead to the conclusion that the scores assigned by the raters were inconsistent with the actual relative experience reflected on the resumes of the applicants. The record in this matter is not clear enough for the Commission to conclude that the scores were inconsistent with relative levels of experience and that, for example, applicant AE#54 should have received more points on criterion C than applicant AE#12. The fact that applicant #54, in DHSS's central accounting office, may have reviewed accounting

work coming out of the correctional institution where applicant AE#12 was in charge of the accounting system, doesn't necessarily mean that applicant AE#54 should be awarded more points for directing reconciliations or audit performance functions where applicant AE#12 supervises other employees while applicant AE#54 is a leadworker.

The first full paragraph on page 17 of the Discussion section of the proposed decision and order is replaced with the following language:

Here, the scoring system was established by job experts and it was applied by persons who were also designated as job experts. None of these persons testified during the course of the proceeding. Under the York analysis, the Commission must determine whether the scoring system was clearly ridiculous or offended common sense. Without the testimony of the job experts who developed the system and those who applied it, it is very difficult for the Commission to say that the results were either clearly ridiculous or offended common sense. The respondent opted for a scoring system which placed a high degree of importance on the supervisory experience which had the effect of dropping out those applicants who lacked such experience. The fact that the system, as applied, resulted in lower scores for persons in central office accounting positions with no supervisory responsibilities than persons with supervisory responsibilities in a satellite facility is not clearly ridiculous given that the examination process was for a supervisory position.

The first sentence under the heading of "Private sector applicants" on page 17 of the Discussion section of the proposed decision and order is modified to read:

The record does not support the appellant's arguments regarding the level of experience in governmental accounting for the private sector applicants who were ranked as the top four after the resume screen.

The Summary paragraph found on the bottom of page 21 and the top of page 22 of the proposed decision and order is replaced with the following:

The appellants' case may be summarized as consisting of 1) their own testimony as to their training and experience and the training and experience of certain of the applicants who made it past the resume screen and into the oral examination stage of the Fiscal Supervisor 1 examination process, 2) Mr. Bell's testimony and 3) analyses of the results of the screening process based on comparisons among groups of applicants. The appellants failed to offer any expert testimony which placed into question the conclusions of validity testified to by Mr. Bell and failed to call the exam raters in an effort to establish some rating impropriety. The appellants' case was sufficient to raise questions about the resume screen process involved here. Those questions arise because of the apparent discrepancies between the 1988 and 1989 Fiscal Supervisor 1 results, the markings on the exams made by the raters and by the testimony relating to the relative qualifications of certain applicants. Because of the absence of contrary expert testimony or other evidence which directly undermined the criteria, benchmarks or scoring, the Commission must sustain the resume screen process used in the Fiscal Supervisor 1 examination.

The Order is revised to read:

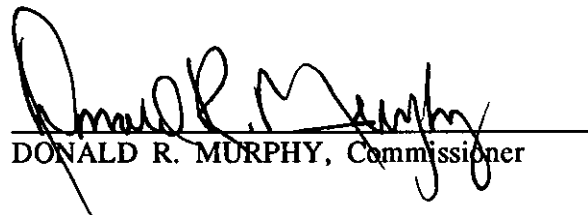
The respondent's use of the resume screen process for the Fiscal Supervisor 1 examination process is affirmed and this appeal is dismissed. Because this decision sets forth examination information contained in exhibits which were initially submitted to the Commission and made available to the appellants pursuant to a protective order, this decision is issued under seal and

the parties are directed not to disclose the examination materials contained in this decision to the public or outside the confines of this proceeding.

Dated: May 17, 1990 STATE PERSONNEL COMMISSION


LAURIE R. MCCALLUM, Chairperson

KMS:kms


DONALD R. MURPHY, Commissioner


GERALD F. HODDINOTT, Commissioner

Parties:

Jeffrey Allen
310 South Yellowstone #1
Madison, WI 53705

Michael A. Doyle
Rt. 2, Box 319A
Mineral Point, WI 53565

Roger Birkett
5321 Mosinee Lane
Madison, WI 53704

Yuh-Wen Chi
5 South Hillside Terrace
Madison, WI 53705

Daniel J. Steeger
2 Green Ridge Court
Madison, WI 53704

Hugh Henderson
Acting Administrator, DMRS
Box 7855
Madison, WI 53707

Harlan Olson
1421 Drewry Lane
Madison, WI 53704

JEFFREY ALLEN, et al.,

Appellants,

v.

Administrator, DIVISION OF
MERIT RECRUITMENT AND SELECTION

Respondent.

Case No. 89-0124-PC

PROPOSED
DECISION
AND
ORDER
••ISSUED UNDER SEAL••

This matter arises from an examination conducted by the respondent. The parties agreed to the following issue for hearing:

Did the decision of the respondent to use a "resume screen" for the Fiscal Supervisor 1 vacancy announced on August 27 and September 1, 1989, violate §230.16, Stats?

The parties filed posthearing briefs.

FINDINGS OF FACT

1. At all times relevant to this proceeding, the appellants have been employed by the Department of Health and Social Services (DHSS) in the Bureau of Fiscal Services.
2. Some time prior to September of 1989, John Packard, a Fiscal Supervisor 2 in the Department of Industry, Labor and Human Relations, contacted Alan Bell, a personnel specialist with respondent Division of Merit Recruitment and Selection (DMRS), and informed Mr. Bell of the need to fill a Fiscal Supervisor 1 vacancy.
3. Since 1981, Mr. Bell has been employed by DMRS (or its predecessor) and has had responsibilities which include examination development. Mr. Bell has a masters of education degree in psychology and guidance, teaches statistics at a technical college, has a "substantial background" in statistics, measurement and test development and has completed the necessary coursework for a doctorate in personnel management. By statute, DMRS is responsible for

examining applicants and, based on the results, preparing a list of eligibles to be considered by an appointing authority for a specific vacancy.

4. Since approximately January of 1988, Mr. Bell has used a resume screen procedure as part of more than 75 examinations. The resume screen procedure requires each applicant to submit a resume which is reviewed by raters based on previously established standards. Those persons whose resumes receive the highest scores then proceed to the second stage in the examination process, usually an oral examination.

5. The resume screen procedure is one of a variety of examination options available to the respondent. Other options include achievement history questionnaires, oral examinations and written examinations. Written examinations may take many forms, including short answer questions, multiple choice questions, narrative questions and hypothetical questions.

6. Mr. Bell and other DMRS staff had received one or more directives to try to speed up the examination/certification process. One of the advantages of the resume screen procedure is that it requires less time for the raters to complete their ratings than for other methods of examination.

7. Mr. Bell and Mr. Packard discussed the various examination options available for the Fiscal Supervisor 1 position and concluded that the resume screen process followed by an oral examination was appropriate.

8. Later, Gary West, a Fiscal Supervisor 3 at DHSS, became involved in the Fiscal Supervisor 1 exam development process because he also anticipated a vacancy in his work unit for a Fiscal Supervisor 1.

9. Both Mr. Packard and Mr. West had previously served as job experts for DMRS for one or more accounting-related job titles.

10. Before Mr. Bell decided to use Mr. Packard and Mr. West as the job experts for the development of the Fiscal Supervisor 1 exam process, he checked on their current classification titles and on how long they had served in their positions.

11. Messrs. West, Packard and Bell reviewed the position descriptions for the two known Fiscal Supervisor 1 vacancies and determined the most important job tasks. The position summaries for the two positions read as follows:

DILHR (Unit Supervisor, Accounts Receivable and Reconciliation,
General Accounting Section, Bureau of Finance and Management
Analysis)

Supervises a staff of professional and para-professional employees in the reconciliation processes of the Department's highly complex computerized accounting system. Responsible for directing the cash management processes of federal letter of credits and preparation of invoice billings of federal and non-federal grants and contracts. Directs and controls submission of the Department's financial transactions to the DOA-Bureau of Finance. Directs the internal and external reporting requirements for generation of accurate financial reports. Directs and reviews the ancillary financial operations for IRS reporting requirements, travel advance requirements, audit determinations and additions or modifications to administrative practices and procedures manual.

DHSS (Deputy Section Chief, General Accounting Section, Bureau of Fiscal Services, Division of Management Services)

The Deputy Section Chief of the General Accounting Section in the Bureau of Fiscal Services assists the Section Chief in all phases of the Section's operation. Under administrative delegation, the Deputy Section Chief supervises the day-to-day operations of the General Accounting Section.

12. Based on their analysis, the evaluators developed a list of five criteria for use in evaluating applicant resumes. The five criteria were: 1) "supervisory experience," 2) "direct or coordinate accounting system or business management function," 3) "directed reconciliations or audit performance functions," 4) "knowledge of governmental accounting, auditing and financial reporting principles" and 5) "cash management reporting functions--invoice billings and/or letter of credit process." For each criteria, an applicant could be awarded between 0 and 4 points.

13. The most points (4) were to be awarded to applicants who performed those responsibilities identified in the criteria at the "organizational" level. If the responsibilities were supervised at the "component" level, the applicant was to receive 3 points. Two points were to be awarded for performing the responsibilities at the "functional" level. Applicants received 1 point for "minimal involvement" with the responsibility and no points if there was no such involvement.

14. The scoring levels failed to adequately account for differences in size and scope of responsibility of the "organizations" employing the applicants. Applicant A who had supervisory responsibilities could be awarded

more points for a particular criterion than Applicant B who was not a supervisor, even though Applicant B reviewed A's work.

15. Both Mr. Packard and Mr. West completed a "job expert affidavit" in which they 1) specified they were qualified to participate in the development of the Fiscal Supervisor 1 examination, and 2) certified the examination was "representative, necessary, and at an appropriate level of complexity" for the position.

16. Late in August of 1989, the respondent issued a job announcement for the Fiscal Supervisor 1 vacancy at DILHR and noted that the register created could be used to fill other Fiscal Supervisor positions in state service. The announcement was directed to the general public and read, in part:

KNOWLEDGE AND SKILLS REQUIRED: State accounting and auditing policies and procedures; state and federal reconciliation requirements; federal grant system requirements and reports; demonstrated supervisory and leadership skills with both professional and paraprofessional employees. **Apply with a letter of application and resume which focuses your qualifications on the job duties, and knowledge and skills cited above to Merit Recruitment and Selection** **NOTE:** Application materials will be screened and those individuals who appear best qualified will be invited to participate in an oral examination. [Emphasis in original]

17. Fifty-four persons submitted application materials.

18. The objective established prior to the grading of resumes was to develop a list of candidates who could be orally examined on one day as the second step in the examination process. This meant that the goal was for between 10 and 16 candidates to proceed to the second step of the process.

19. At Mr. Bell's request for the names of possible raters for the resume screening process, Mr. Packard suggested two persons: Ray Wyss, who had been a Fiscal Supervisor 1 for a number of years, and James Hoelzel, who had been a legislative auditor and audit supervisor for a number of years before being promoted to a bureau director position.

20. Before the screening began, Mr. Bell and Mr. Packard met with Mr. Wyss and Mr. Hoelzel, explained the examination process, explained various rater tendencies, emphasized the rater's ability to withdraw in the event they felt they were biased as to any applicants and discussed the fact that the raters would probably be more familiar with certain employing organizations than

others. The group discussed how various hypothetical resumes should be scored and Mr. Bell explained the importance that the raters develop their ratings independently of each other and consistently for all the applicants. Mr. Bell divided the applications (resumes plus cover letters) into four approximately equal groups, had the raters score 5 or 6 and then had them look at how they had done before they scored the remaining resumes.

21. At least one of the raters circled various items of information on the resumes during the screening process. Thirty-seven of the 54 applications have one or more words or phrases circled. For the most part, the circled information relates to supervisory or leadwork experience indicated on the resume. Mr. Bell was unaware that the information was circled until after the screening process was completed but he was aware of the marks before the register for the position was created.

22. For those 37 applications which had one or more words or phrases circled, the two raters disagreed on 74 of the 185 scores, or 40%. For those 17 applications which had no rater markings on them, the two raters disagreed on only 21 of 85 possible scores, or 25%.

23. After the two raters had completed their scoring, Mr. Bell added together the rater's scores for each applicant and also compared the results of the two raters by running a Pearson Product Moment Correlation (PPMC). The PPMC is designed to compare the degree of agreement between the raters. If two raters invariably scored the candidates in an identical way, the correlation would be +1.0. If two raters always came up with different ratings of the candidates, the correlation would be -1.0. A correlation of 0.7 was considered to be the lowest acceptable value. Any correlation in the 0.9's was considered to be quite high. The PPMC value for the Fiscal Supervisor 1 resume screen scores was 0.937.

24. PPMC is a commonly used statistic and is neither an overly liberal nor an overly conservative estimator.

25. Because of the distribution of the results, the respondent chose to expand the number of candidates who would proceed to the second step in the examination process and selected 20 candidates from the original applicant pool as a consequence of the resume screen. This meant that the oral examination had to be conducted over a period of 1 1/2 days rather than 1 day. The ranking from the resume screen was such that there was no natural break in

the scores assigned to the candidates between rankings 10 and 16. However, by including those candidates whose score would have put them among the top 16 candidates for either rater, respondent came up with the list of 20 candidates.

26. As a consequence of the resume screening procedure, 34 of the applicants were dropped from further participation in the Fiscal Supervisor 1 examination process. The remaining 20 applicants were invited to an oral examination and based on the results of that examination, a certification list was prepared.

27. None of the appellants were invited to the oral examination.

28. The Department of Industry, Labor and Human Relations filled the vacant Fiscal Supervisor 1 position from the certification list.

29. The appointing authority for the DILHR vacancy and members of the FS 1 oral board informed Mr. Bell that the persons they had encountered during their portions of the examination/selection process were, for the most part, qualified to perform FS 1 responsibilities.

30. Both Mr. Wyss and Mr. Hoelzel completed a "job expert affidavit" in which they specified they were qualified to participate in rating the Fiscal Supervisor 1 examination applicants.

31. For a number of years preceding 1989, the respondent had used a written examination for Fiscal Supervisor 1 vacancies. The written exam was largely unchanged during this period.

32. In 1988, respondent had issued a service-wide promotional announcement for the Fiscal Supervisor 1 examination. The announcement generated between 40 and 60 applicants. Fourteen people took both the 1988 and the 1989 Fiscal Supervisor 1 examinations. Their scores on the 1988 written exam and on the resume screen component (only) of the 1989 exam are listed below:

<u>Candidate</u>	<u>Ranking on 1988 Exam</u>	<u>1989 Resume Screen</u> <u>Score</u>	<u>Inc./Excluded</u>
A	2	17	Excl.
B	3	13	Excl.
C	6	30	Included
D	8	20	Excl.

E	9	8	Excl.
F	11	26	Included
G	12	14	Excl.
H	14	16	Excl.
I	16	7	Excl.
J	17	29	Included
K	18	18	Excl.
L	23	19	Excl.
M	27	26	Included
N	Did not pass test	26	Included

CONCLUSIONS OF LAW

1. This matter is properly before the Commission pursuant to §230.44(1)(a), Stats.
2. The appellants have the burden of establishing that the Fiscal Supervisor 1 examination violated §230.16, Stats., or the related administrative rules.
3. The appellants have sustained that burden.
4. The examination was not conducted in accordance with the civil service requirements.

OPINION

This appeal, filed by six persons who were prevented from proceeding further in the Fiscal Supervisor 1 examination process by the results of a resume screen procedure, alleges violation of certain provisions in §230.16, Stats., and ER-Pers 6.05, Wis. Adm. Code. The relevant statutory provisions read, as follows:

(4) All examinations, including minimum training and experience requirements, for positions in the classified service shall be job-related in compliance with appropriate validation standards and shall be subject to the approval of the administrator. All relevant experience, whether paid or unpaid, shall satisfy experience requirements.

(5) In the interest of sound personnel management, consideration of applicants and service to agencies, the administrator may set a standard for proceeding to subsequent steps in an examination, provided that all applicants are fairly treated and due notice has been given. The standard may be at or above the passing point set by the administrator for any portion of the examination. 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. [Emphasis added]

Specifically, the appellants contend that the respondent has failed to properly apply the language which is underlined above. The appellants also contend the following provisions of §ER-Pers 6.05, Wis. Adm. Code, were violated:

(3) All examinations shall be:

(a) Based on information from job analysis, position analysis or other equivalent information documenting actual job tasks to be performed or skills and knowledges required to perform job tasks, or both;

(b) Developed in such a manner as to establish the relationship between skills and knowledges required for successful performance on the test and skills and knowledges required for successful performance on the job;

(c) Supported by data documenting that the skills and knowledges required for successful performance on the test are related to skills and knowledges which differentiate among levels of job performance if the examination results are to be used as a basis for ranking candidates;

(d) Sufficiently reliable to comply with appropriate standards for test validation;

Exam notice

Appellants contend they were not given due notice of certain "circumstances" which amounted to Mr. Bell's expectation that applicants should be able to sell themselves via their resumes and that a good resume would combine functional and chronological parts. However, the Fiscal Supervisor 1 announcement very specifically advised the applicants that they would be screened based upon their application materials, which were identified as a letter of application and resume. See finding of fact 16. The Commission rejects the appellants' contention that this announcement was somehow insufficient. The examination announcement clearly informed the applicants that they should put their best foot forward when filing their resume and cover letter.

The appellants have raised a series of arguments against the use of the resume screen procedure, both as part of any exam and as part of the Fiscal Supervisor 1 examination.

Inherent reliability of the resume screen

Appellants contend that the resume screen procedure is inherently unreliable because applicants may give false information on their resume as to their education, experience or level of skill. The appellants argue that the information found on resumes is less reliable than information supplied by applicants as part of other examination techniques. Mr. Bell testified that he was unaware that the resume screen process would generate any more falsification than the various other examination alternatives. Clearly, there is a greater potential for falsification whenever an applicant is asked to describe his or her experience in a certain area instead of being asked a technical question designed to test the existence of that knowledge. In a resume screen, there is no way in which to ask such a technical question. By definition, the resume screen process is only seeking an applicant's assertion that she or he has certain knowledge or experience. The problem here is that the mere potential for falsification does not make the resume screen process invalid unless there is some evidence that applicants are making use of the potential and actually submitting falsified resumes so as to cause invalidity.¹ The appellants didn't offer any evidence to show that the resume screen process is viewed as unreliable by persons in the field of test development. In contrast, Mr. Bell testified that the respondent has used the resume screen process on a regular basis for the past several years and the submission of resumes is common practice in private sector hiring. Appellants also argue that somehow the information in the resumes should have been verified, but they have failed to point to any language in the rules or the statutes or in the principles incorporated therein which would require such verification. Therefore, based on the evidence presented in this proceeding, the Commission rejects the appellants' contention that the resume screen process is inherently unreliable or requires verification of the resume information.

Resume screen versus other examination options

¹Even though there is clearly a potential for applicants to inflate their resumes, the consequences of being found out after having falsified a resume may act to deter such conduct.

The appellants also argue that the other examination options which could have been used for the Fiscal Supervisor 1 examination were so far superior to the resume screen process that the respondent's decision to use the resume screen as the first stage of the Fiscal Supervisor 1 examination must be overturned. Clearly there were many other examination alternatives that were available, including an essay exam, a multiple choice exam, an achievement history questionnaire and an oral exam. Mr. Bell testified that there were several factors which were considered when deciding to use the resume screen process as the initial aspect of the Fiscal Supervisor 1 examination. He stated that: 1) because the position is a fairly high level job, one could expect that the applicants would have a substantial amount of experience, 2) the job was open to the public and the process of submitting resumes generally is a familiar one to the public and 3) the exam for the Fiscal Supervisor 1 position had been relatively unchanged over the previous several years and persons who had taken the exam several times during this period could have an advantage should it be repeated in 1989. In addition, the time necessary to complete the examination process could be shortened by including a resume screen in the process.

The respondent presumably could have used another examination process which did not include a resume screen. All of the different examination techniques have their trade-offs. Here, the respondent offered various reasons in support of the use of the resume screen. Even assuming, for the sake of argument, that the overriding motivation behind the decision was to get a certification list as soon as possible, there is insufficient evidence to reject the decision to utilize a resume screen as the first stage.

Examination validity

As noted above, §230.16, Stats., requires that the resume screen, as part of the examination process, be "job-related in compliance with appropriate validation standards."

The various methods of establishing test validation are described in 1 C Sullivan, M. Zimmer, R. Richards, Employment Discrimination §4.5.5 (2d ed.)

Three different techniques or strategies are available to validate a test -- content validation, criterion-related validation, and construct validation. The easiest to understand is content validation, which is appropriate where the test can approximate a sample of the critical elements of the job. The classic example is a typing

test for a typing job. The second is criterion-related validation. Unlike content validation, criterion validation tests are for more abstract skills, which are determined by job evaluation to be involved in the job in question. In the typing example the test designer might conclude that finger dexterity is critical to success on the job and employ a dexterity test to show which worker would be likely to learn to be a good typist. Because the dexterity test may not accurately predict job performance, however, the test designer will conduct an empirical study to determine whether success on the test predicts success on the job. Criterion-related validation is the strategy preferred by testing experts because it involves undertaking a statistical comparison of test scores and job performance scores. . . . The third strategy, construct validation, amounts to the same kind of study as criterion-related validation; its use is contemplated for psychological diagnosis rather than employment testing.

Mr. Bell testified that the examination system employed here relied on the opinions of the various job experts involved in the exam to confirm that the entire examination, including the process followed and the criteria used, had "content validity." Mr. Bell used the term "content validity" to mean that the content of the applicant evaluation process was based on the content of the duties to be performed in the vacant position.

The Commission has already addressed the appellants' contention that all resume screens are inherently unreliable due to the potential for falsification. The appellants have not offered any argument or evidence to the effect that the resume screen process is inherently not job-related.

At one point in their brief, respondents appear to acknowledge the job-relatedness of the job criteria set forth in finding of fact 12: "We believe that the five job criteria . . . were generally appropriate, if applied correctly, to the Fiscal Supervisor 1 position." However, later in their brief, the appellants contend that the "failure to evaluate educational level, or certification (as a public accountant), resulted from a professional bias of the job experts." Mr. Bell testified that the three persons who developed the criteria had specifically considered whether to include a criterion based upon educational level but had decided instead to focus on work experience because they expected the applicants to have had a substantial amount of work experience after receiving their educational training. Mr. Bell also testified that CPA certification could be considered within criteria 4 and that the possibility of doing so was part of the preliminary discussions with the two raters. The appellants did not provide their own expert witness to testify as to the validity of the criteria. The

persons identified as the "job experts" by the respondent did not testify. To the extent the appellants contend that the 5 job criteria somehow invalidate the Fiscal Supervisor 1 process, they have failed to sustain their burden of proof.

Scoring system

While the evidence supports the appropriateness of respondent's action of employing a resume screen as part of the Fiscal Supervisor 1 examination process and also supports the appropriateness of the 5 job criteria established by the job experts, the appellants have raised numerous arguments relating to the application of the 5 criteria to various groups within the 54 applicants. Those arguments may be summarized as follows:

1. The rating is invalid because none of the 8 applicants employed within DHSS's Bureau of Fiscal Services were invited to the oral examination.
2. The rating is invalid because all 6 of the applicants employed within DHSS but outside of the Bureau of Fiscal Services were invited to the oral examination.
3. The rating is invalid because none of the applicants employed as audit supervisors within the Legislative Audit Bureau were invited to the oral examination.
4. The rating is invalid because those applicants ranked in the top 4 were all employed in private business and their resumes showed little indication of any knowledge of governmental accounting.
5. The rating is invalid because an insufficient number of auditors were invited to the oral examination.
6. The rating is invalid because the results were inconsistent with the results of the 1988 Fiscal Supervisor 1 essay examination.

Within these 6 arguments, there are a few recurring themes raised by the appellants in terms of the rating methods. The appellants contend there was an over-emphasis on the existence of supervisory responsibilities and that an applicant's status as a supervisor generated higher scores for each of the 5 criteria rather than just for the "supervisory experience" criterion (A). Closely related to this argument is the appellants' contention that after the raters awarded an applicant points for criterion A, the raters were likely to award the same number of points for criteria B through E.

There can be little question that supervisory experience is an appropriate factor for screening resumes for a Fiscal Supervisory 1 vacancy. Here, re-

spondent chose to set up its "resume evaluation criteria" to look at five different areas of experience. The first related to supervisory experience. The most points (4) for the supervision criterion would be awarded to an applicant who had "overall management responsibilities of Financial Operations at organizational level". Fewer points would be awarded to applicants whose supervisory experience was of "professional and paraprofessional employees at component level", of a "small staff of employees at a functional level" or consisted only of "minimal supervision of paraprofessional employees." Applicants without a reference on their resume to supervisory experience would receive no points. This scoring system took into account not only the preliminary question of whether the applicant had supervisory experience but also attempted to consider whether the supervision was of professional employees or paraprofessional employees and whether the scope of the operation supervised was at the functional, component or organizational level.

The remaining four criteria used in the resume screen addressed separate areas of fiscal experience, but did so in a way which not only looked at whether the applicants had such experience but went on to evaluate the level at which that experience was gained. The four areas of fiscal experience which were screened for were business management, reconciliation/audit, governmental financial reporting and cash management. (See finding of fact 12). As to each of these four areas of experience, the resume screen was designed to differentiate between applicants whose experience was gained at the organizational level, the component level or the functional level. For example, someone with business management responsibilities at an organizational level received 4 points, while someone who had exercised business management responsibilities at the functional level received only 2 points. Persons with "minimal business activity involvement" received 1 point while those with no business activity involvement at all didn't receive any points. Obviously, with this type of scoring structure, an applicant who had gained experience in all four areas while serving in a position with organization-wide responsibilities, should get the highest overall ranking.

The application of this scoring system has been shown by the appellants to have been inconsistent with the goal of accurately assessing the experience of the various applicants because the system failed to contemplate or-

ganizations of vastly different sizes and overemphasized mere status as a supervisor.

For example, the resume and cover letter (Appellants' Exhibit 34) for someone employed as an Accountant 4 (Leadworker) in pay range 14 in DHSS includes the following information:

In my current position I am responsible for the federal reporting of the Medical Assistance and Aid to Families with Dependent Children (AFDC) programs. These two programs total over \$1 billion, per fiscal year, in Federal aid alone. Part of my duties include reconciling the Department of Health and Social Services' accounting system to the Department of Administration's records for projects assigned to me. This also includes reconciling expenditures to revenue drawn on the Single Letter of Credit (SLC) System. If an expenditure does not draw on the SLC, necessary adjusting entries are prepared to effect the draw.

* * *

- Reconciliation, maintenance and monitoring of assigned documents on the Single Letter of Credit system (SLC). Includes analyzing and managing the department accounting system for drawing federal funds for various programs, reconciling expenditure reports to State records and the SLC, reconciling expenditure reports to State records and the SLC, reconciling monthly Department of Administration (DOA) Project Balance Reports to projects being monitored.
- Consultation/problem resolution with Federal officials, auditors, DOA Single Letter of Credit personnel, and program employees regarding assigned federal programs and projects.
- Designated lead worker in section - attend meetings in absence of Section Chief, attend audit entrance and exit conferences.

Despite these references to letter of credit responsibilities for programs totalling over \$1 billion per fiscal year, applicant AE #34 received received zero and 2 points, respectively, from the two raters for criterion E (Cash Management Reporting Functions--Invoice Billings and/or Letter of Credit Process).

In contrast, another employe of DHSS, by virtue of his title as an Account Specialist 3 - Supervisor (pay range 12) and status as Assistant Business Administrator, received 2 and 3 points, respectively, from the two raters for the same criterion. The resume and cover letter for that person, applicant AE#12, read in part:

As Assistant Business Administrator at the Oshkosh Correctional Institution, I am very familiar with state accounting and auditing

policies and procedures and reconciliation requirements. I am in charge of the entire accounting system for our institution, directing the cash management processes, accounts receivable, accounts payable and accurate and timely financial reporting. I am also involved and very knowledgeable in the purchasing process, assuring that all purchasing laws, policies and procedures are complied with.

The same individual (AE#12) received 3 points from each rater for criterion C (Directed Reconciliations or Audit Performance Functions). In addition to the information set out above, his resume also referred to his audit responsibilities as follows:

Conduct audits of accounting documents, cash receipts, bank statements, inmate payroll and checkbooks. . . . Reconcile the Institution store accounts, physical inventory and capital equipment on an annual basis.

For purposes of comparison, another applicant (AE#54) employed by DHSS as a Lead Accountant was awarded 1 and 2 points by the two raters for criterion C despite the following language in the resume and cover order:

For the last four years I have been the Lead Accountant for all of the institutions in the Department. In my present lead worker position, I am directing the work of paraprofessional staff, overseeing the reconciliation including cost allocations for more than sixty appropriations, and providing primary financial accounting services to the institutions.

The consistently high marks awarded by the raters to applicant AE#12 generated an overall score of 29 which tied the applicant for 6th highest among all 54 of the applicants. These scores resulted directly from that applicant's civil service title as a supervisor rather than from an assessment of the actual knowledge and responsibility he carried out in the areas of cash management/letter of credit and reconciliation/audit. In contrast, applicants who were lead workers rather than supervisors, but otherwise were in higher level classifications and had far more extensive responsibilities in these areas, were awarded fewer points. As "lead Accountant for all of the institutions in the Department", applicant AE#34 provided "primary financial accounting services" and had review responsibilities over the accounting work coming out of the Oshkosh Correctional Institution, where applicant AE#12 was employed. Yet, according to the scoring system applied by the respondent, appli-

cant AE#12 was rated as having superior experience relative to applicant AE#54 as to all five criteria for assessing the resumes. Neither applicant AE#34 nor #54 came close to receiving enough points to reach the oral examination.¹

These results strongly support the appellants' contention that the scoring system adopted by the respondent to implement the 5 criteria overemphasized an applicant's status as a supervisor and failed to take into account the size of the organization in which the applicant was employed or potential for review of the applicant's work by others.

The respondent cites York v. DP, 78-42-PC, 7/18/80, as establishing the standard to be used when reviewing benchmarks:

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. . . .

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.

¹Applicant AE#34 received 14 points, which tied him at the rank of 38th among the 54 applicants while applicant AE#54 received a total of 18 points, tying him for the rank of 32nd.

In York, the appellant was contesting the appropriateness of certain benchmark answers for essay questions which were part of a written examination for sergeant positions in the state patrol. It was the appellant in York who contended that some of the benchmark answers were "clearly ridiculous" or offended common sense, while the respondent's exam experts defended the benchmarks as being consistent with state patrol policy. A logical reading of York is that the Commission would only have sustained the appellant's attack on the policy which served as the basis for the benchmark, if it could have been shown that the policy itself was clearly ridiculous.

In contrast to the situation in York, the arguments of the appellants in the instant case relate to scoring standards which are not premised on some underlying performance standard. The appellants, therefore, need not show that the scoring system employed here was clearly ridiculous or offends common sense. The appellants have met their burden by showing that the application of the scoring system to certain applicants resulted in scores which are inconsistent with the screening criteria.

Private sector applicants

While the record supports the appellants' arguments regarding the failure of the scoring system to adequately consider organizational size and its overemphasis on supervisory status, the record does not support their arguments regarding the level of experience in governmental accounting for the private sector applicants who were ranked as the top four after the resume screen. "Knowledge of governmental accounting, auditing and financial reporting principles" was one of five criteria utilized in evaluating the applicant's resumes. The rating system for this evaluation criterion was as follows:

- 4 - Directed financial reporting at organizational level.
- 3 - Supervised financial reporting at component level.
- 2 - Performed financial reporting responsibilities at functional level.
- 1 - Involved with financial reporting responsibilities.
- 0 - No knowledge of governmental accounting reporting requirements.

The cover letter to the resume of the top ranked applicant (#34) included the following statement:

In my position as Controller at [XYZ] Corporation, a private non-profit organization which had to follow state accounting and auditing procedures per the Job Training Partnership Act, I worked with state and federal reconciliation guidelines and grant system requirements. I have experience analyzing account receivables and reconciling that with the state.

Applicant #34 was awarded 1 point for criterion D by each of the two reviewers. The cover letter to the resume of the second ranked applicant (#38) included the following statements:

I have performed audits of various State of Wisconsin agencies and departments during my eight years in public accounting. I am familiar with the compliance aspect of accumulating and reporting fiscal transactions along with the related empirical data in report formats that enable cross utilization for state, federal and local requirements.

Applicant #38 was not awarded any points by either reviewer for criterion D. There were no specific references to governmental accounting, auditing or financial reporting principles in the cover letter or the resume of the third ranked applicant (#50). Applicant #50 received 1 point from one reviewer and no points from the other reviewer. The resume of the private sector applicant who was tied at the ranking of fourth (#20) included the following experience:

Executive Director [ABC], Inc.

A non-profit, 501(c)(3), corporation, created in cooperation with [a] State University. Report to Board of Trustees. Direct all personnel and activities of the organization, with a \$300,000 budget. Develop programs to stimulate new business development and creation of small business incubators for technology based businesses. . . . Administer grants, keep records and generate reports as required by grant sources.

Applicant #20 received 2 points from one reviewer and no points from the other reviewer for criterion D.

In their brief, the appellants contend that the resumes of these top four applicants "show no indication of knowledge of governmental accounting." Despite the appellants' contention, the Commission finds that the application materials of 3 of these 4 applicants can be interpreted as describing certain experience in governmental accounting. That experience has been described in the preceding paragraphs. The two resume reviewers did not testify so

there is no direct evidence as to how they interpreted the excerpted passages set out above. However, the scores awarded to these four applicants for criterion D do appear to generally correspond to the level of experience described in their respective application materials.

Comparison with results of 1988 examination

The appellants also argue that because the results of the 1989 Fiscal Supervisor 1 resume screen are radically different from the scores from the 1988 Fiscal Supervisor 1 exam scores for those persons who took both exams, the 1989 resume screen process is invalid.

It is difficult to draw this conclusion as to the validity of the 1989 exam process without some additional supporting evidence. A mere discrepancy could otherwise just as logically point to the invalidity of the 1988 process as to the 1989 resume screen procedure. In addition, there are factual distinctions between the two exam processes which tend to undermine a comparison of their results. Fourteen of the 54 applicants for the 1989 exam also took the 1988 exam. That means that there is no way to know how the other 40 applicants would have fared had they taken the 1988 exam. The 1988 process also was a promotional opportunity for persons already within the classified service while the 1989 process was opened up to anyone, irrespective of their existing civil service status. This presumably had the effect of including more applicants in the process who had experience from outside of state service. The 1989 resume screen also appeared to place a much higher emphasis on supervisory experience than did the 1988 exam. Having pointed out these factual distinctions, a review of the scores for those 14 persons who participated in both processes shows that many of these persons had widely divergent results from the two processes. Eleven of the fourteen experienced shifts of more than 10 places between 1988 and 1989:

Candidate A dropped from #2 to #34.
Candidate B dropped from #3 to #41.
Candidate D dropped from #8 to #25.
Candidate E dropped from #9 to #42.
Candidate G dropped from #12 to #38.
Candidate H dropped from #14 to #35.
Candidate I dropped from #16 to #49.
Candidate J moved up from #17 to #6.
Candidate K dropped from #18 to #32.
Candidate M moved up from #27 to #15.
Candidate N moved up from "did not pass" to #15.

The results for 3 of the 14 remained relatively constant between the two exams:

Candidate C moved up from #6 to #4.
Candidate F dropped from #11 to #15.
Candidate L dropped from #23 to #28.

It is certainly possible that someone who did very well in the 1988 oral exam could have prepared an incomplete resume in 1989 which failed to address the salient responsibilities of the Fiscal Supervisor 1 position. This scenario could account for at least some of the discrepancy between the 1988 and 1989 results without drawing into question the validity of 1989 resume screen process. Another possible explanation is that the validity of the 1988 exam was compromised because it had been used for many years in a similar format so that applicants who had taken it before had an advantage. However, these scenarios are simply conceivable explanations for the various discrepancies in the results. The Commission concludes that the discrepancies noted above do provide support for the conclusion that the 1989 resume screen process was invalid.

Rater reliability

The Pearson Product Moment Correlation (PPMC) measures reliability of a particular aspect of the exam process rather than validity of the entire exam. Appellants argue that Mr. Bell should have used some other measure of reliability because he testified that the PPMC was probably not the "best possible" correlation to use. While alternative statistics were available to Mr. Bell, the only testimony on this point is that it was unnecessary to go through other analyses because the PPMC showed high reliability between the two raters. Therefore, the Commission cannot conclude that the respondent had to do something more than it did in order to establish rater reliability.

Rater markings on the resumes

Even though the Commission accepts the conclusion that the high PPMC score meant it was unnecessary for Mr. Bell to carry out other statistical tests, those same PPMC results provide support for a contention that the markings made by at least one of the raters on the resumes undermined the independence of the two raters and resulted in an invalid examination process. Mr. Bell admitted that he would question an exam which resulted in a PPMC value of 1.0. Here, the PPMC value was .937 and there were markings on 37 of the 54 resumes. Neither rater was called as a witness, but the Commission in-

fers from the evidence in the record that at least one of the raters circled information on the resumes and that up to 37 of the resumes reviewed by the second rater included markings made by the first rater. There is no evidence of collusion between the two raters. One logical explanation for the markings is that the rater who made them only was thinking they would help in making his own ratings. However, the marks were on the resumes when they first reached the second rater. The second rater had to have been aware of the markings although there is very little evidence as to what effect the markings had on the second rater. The evidence does show that the inter-rater reliability was extremely high which suggests that the markings did have an impact on the scoring of the second rater. On the other hand, an analysis of the scores awarded by the two raters indicates that the markings actually did not undermine the independence of the second rater: the two raters had a much higher percentage of disagreement between their scores for those 37 applications with markings on them (40% disagreement) than for the 17 applications which had no markings (25% disagreement). The Commission cannot know which of the 37 resumes were marked by the time they reached the second rater. Without this information, the reliability of the preceding analysis is undercut somewhat. However, absent any other evidence tending to support a conclusion that the second rater's independence was compromised by the markings made by the first rater, the Commission will not conclude that the respondent failed to use "appropriate scientific techniques and procedures in administering the selection process [and] in rating the results of examinations." §230.16(5), Stats.

Summary

This case could have been substantially easier to decide had the appellants offered expert testimony relating to the conclusions of validity drawn by Mr. Bell. Instead, the appellants relied on 1) their own testimony as to their training and experience and the training and experience of certain of the applicants who made it past the resume screen and into the oral examination stage of the Fiscal Supervisor 1 examination process, 2) Mr. Bell's own testimony and 3) analyses of the results of the screening process based on comparison among groups of applicants. The appellants' case was sufficient to indicate inconsistencies between the resume scoring system and the identified criteria in that the scoring system failed to adequately take into account dis-

crepancies in organizational size and overemphasized an applicant's status as a supervisor. In addition, the appellants pointed out discrepancies between the results of the 1988 and 1989 Fiscal Supervisor 1 examinations. Those discrepancies support the conclusion that the 1989 resume screen, as scored, did not accurately measure an applicant's experience with respect to the 5 job criteria. The respondent's case consisted merely of the general conclusion of content validity proffered by Mr. Bell rather than on a more specific analysis supporting the scoring system that was utilized. Under these circumstances and when the evidence is viewed as a whole, the appellants have shown that the scoring system used in the Fiscal Supervisor 1 resume screen process was incompatible with the identified screening criteria and, therefore, the resume screen was not job-related in compliance with appropriate validation standards as required by §230.16, Stats.

ORDER

The respondent's use of the resume screen process for the Fiscal Supervisor 1 examination process is rejected and the respondent is directed to discontinue the use of any list of eligibles generated as a consequence of the use of that process.

Dated: _____, 1990 STATE PERSONNEL COMMISSION

LAURIE R. MCCALLUM, Chairperson

KMS:kms

DONALD R. MURPHY, Commissioner

GERALD F. HODDINOTT, Commissioner

Parties:

Jeffrey Allen
310 South Yellowstone #1
Madison, WI 53705

Michael A. Doyle
Rt. 2, Box 319A
Mineral Point, WI 53565

Roger Birkett
5321 Mosinee Lane
Madison, WI 53704

Yuh-Wen Chi
5 South Hillside Terrace
Madison, WI 53705

Daniel J. Steeger
2 Green Ridge Court
Madison, WI 53704

Dan Wallock
Administrator, DMRS
Box 7855
Madison, WI 53707

Harlan Olson
1421 Drewry Lane
Madison, WI 53704