UNITED STATES
NUCLEAR REGULATORY COMMISSION
REGION IV
611 RYAN PLAZA DRIVE, SUITE 400
ARLINGTON, TEXAS 76011-8064

January 2, 2002

EA-01-298

David L. Wilson, Vice President of
Nuclear Energy
Nebraska Public Power District
P.O. Box 98
Brownville, Nebraska 68321

SUBJECT: NRC INSPECTION REPORT 50-298/0112; PRELIMINARY WHITE FINDING (COOPER NUCLEAR STATION)

Dear Mr. Wilson:

On December 3, 2001, the NRC completed an inspection of your Cooper Nuclear Station. The enclosed report documents the inspection findings, which were discussed onsite on October 18, 2001, and in a telephone conversation on December 3, 2001, with members of your staff.

The inspection examined activities conducted under your license as they relate to safety and compliance with the Commission's rules and regulations and with the conditions of your license. The inspectors reviewed selected procedures and records, observed activities, and interviewed personnel. The inspection consisted of a review of the licensed operator requalification program. Findings identified during the inspection are discussed in detail in the enclosed inspection report.

This report discusses a finding that appears to have low to moderate safety significance. As described in Section 1.R11 of this report, your process for examination development and validation combined with the similarity of each examination resulted in an apparent compromise of your 2000 biennial requalification written examinations. This finding was assessed using the operator requalification human performance significance determination process as a potentially safety significant finding that was preliminarily determined to be White; i.e., a finding with some increased importance to safety, which may require additional NRC inspection. The issue has a low to moderate safety significance because it represented a failure to recognize an apparent examination compromise and the resultant failure of operators on the written requalification examination, a number of whom performed licensed duties without remediation.

The finding also appears to be an apparent violation of NRC requirements and is being considered for escalated enforcement action (EA-01-298) in accordance with the "General Statement of Policy and Procedure for NRC Enforcement Actions" (Enforcement Policy), NUREG-1600.

Before the NRC makes a final decision on this matter, we are providing you an opportunity to request a Regulatory Conference where you would be able to provide your perspectives on the significance of the finding, the bases for your position, and whether you agree with the apparent violation. If you choose to request a Regulatory Conference, we encourage you to submit your evaluation and any differences with the NRC evaluation at least one week prior to the conference in an effort to make the conference more efficient and effective. If a Regulatory Conference is held, it will be open for public observation. The NRC will also issue a press release to announce the conference.

Please contact Mr. Anthony Gody at (817) 860-8159 within 10 days of the date of this letter to notify the NRC of your intentions. If we have not heard from you within 10 days, we will continue with our significance determination and enforcement decision and you will be advised by separate correspondence of the results of our deliberations on this matter.

Since the NRC has not made a final determination in this matter, no Notice of Violation is being issued for this inspection finding at this time. In addition, please be advised that the number and characterization of the apparent violation described in the enclosed inspection report may change as a result of further NRC review.

In accordance with 10 CFR 2.790 of the NRC's "Rules of Practice," a copy of this letter and enclosure will be available electronically for public inspection in the NRC Public Document Room or from the Publicly Available Records (PARS) component of NRC's document system (ADAMS). ADAMS is accessible from the NRC Web site at http://www.nrc.gov/reading-rm/adams.html (the Public Electronic Reading Room).

Sincerely,

/RA/

Arthur T. Howell III, Director
Division of Reactor Safety

Docket: 50-298
License: DPR-46

Enclosure:
NRC Inspection Report 50-298/01-12

cc w/enclosure:
G. R. Horn, Senior Vice President
of Energy Supply
Nebraska Public Power District
1414 15th Street
Columbus, Nebraska 68601

John R. McPhail, General Counsel
Nebraska Public Power District
P.O. Box 499
Columbus, Nebraska 68602-0499

D. F. Kunsemiller, Risk and
Regulatory Affairs Manager
Nebraska Public Power District
P.O. Box 98
Brownville, Nebraska 68321

Dr. William D. Leech
Manager - Nuclear
MidAmerican Energy
907 Walnut Street
P.O. Box 657
Des Moines, Iowa 50303-0657

Ron Stoddard
Lincoln Electric System
1040 O Street
P.O. Box 80869
Lincoln, Nebraska 68501-0869

Michael J. Linder, Director
Nebraska Department of Environmental
Quality
P.O. Box 98922
Lincoln, Nebraska 68509-8922

Chairman
Nemaha County Board of Commissioners
Nemaha County Courthouse
1824 N Street
Auburn, Nebraska 68305

Sue Semerena, Section Administrator
Nebraska Health and Human Services System
Division of Public Health Assurance
Consumer Services Section
301 Centennial Mall, South
P.O. Box 95007
Lincoln, Nebraska 68509-5007

Ronald A. Kucera, Deputy Director
for Public Policy
Department of Natural Resources
205 Jefferson Street
Jefferson City, Missouri 65101

Jerry Uhlmann, Director
State Emergency Management Agency
P.O. Box 116
Jefferson City, Missouri 65101

Vick L. Cooper, Chief
Radiation Control Program, RCP
Kansas Department of Health
and Environment
Bureau of Air and Radiation
Forbes Field Building 283
Topeka, Kansas 66620

Electronic distribution from ADAMS by RIV:
Regional Administrator **(EWM)**
DRP Director **(KEB)**
DRS Director **(ATH)**
Senior Resident Inspector **(JAC)**
Branch Chief, DRP/C **(KMK)**
Senior Project Engineer, DRP/C **(WCS)**
Staff Chief, DRP/TSS **(PHH)**
RITS Coordinator **(NBH)**
RidsNrrDipmLipb
Scott Morris **(SAM1)**
NRR Event Tracking System **(IPAS)**
CNS Site Secretary **(SLN)**

R:_cns\cn0112rp-pcg.wpd

OE:OB	SOE:OB	C:OB	C:PBC	ACES	C:OB	D:DRS
RELantz	PCGage	ATGody	KKennedy	GFSanborn	ATGody	ATHowell
1/02/02	1/02/02	1/02/02	1/02/02	1/02/02	1/02/02	1/02/02

OFFICIAL RECORD COPY T=Telephone E=E-mail F=Fax

ENCLOSURE

U.S. NUCLEAR REGULATORY COMMISSION
REGION IV

Docket:	50-298
License:	DPR 46
Report No.:	50-298/01-12
Licensee:	Nebraska Public Power District
Facility:	Cooper Nuclear Station
Location:	P.O. Box 98 Brownville, Nebraska
Dates:	October 15 - December 3, 2001
Inspectors:	Paul C. Gage, Senior Operations Engineer Ryan E. Lantz, Operations Engineer
Approved By:	Anthony Gody, Chief, Operations Branch Division of Reactor Safety

SUMMARY OF FINDINGS

IR 05000298/2001-12; on 10/15-12/03/2001; Nebraska Public Power District; Cooper Nuclear Station, licensed operator requalification report.

The inspection was conducted by two regional operations engineers. The inspection identified one Green finding, and one finding whose significance was preliminarily evaluated as White. The significance of each issue was indicated by their color (green, white, yellow, red) and was determined by the Operator Requalification Human Performance Significance Determination Process in Inspection Manual Chapter 0609.

Cornerstone: Mitigating Systems

- TBD. The licensee had compromised their 2000 Biennial Requalification Written Examinations. This constitutes an apparent violation of 10 CFR Part 55.49 for engaging in activities, which compromised the integrity of an examination.

  The finding was preliminarily evaluated as having low to moderate safety significance because after identification of the compromise, the corrective action process (compensatory actions) failed to adequately evaluate the requalification examinations for the effects of the compromise. Had the licensee performed a detailed question analysis and regraded the requalification examinations by removing those questions where compromise was indicated, at least two licensed operators would have failed instead of receiving their original passing grade. Subsequently, at least two operators were returned to licensed duties without completion of the required retraining and testing for having failed the examination (Section 1R11).

- Green. During the 2000 Annual Operator Requalification Operating Test, two out of seven total crews failed the dynamic simulator portion of their operating test.

  The safety significance of this finding was very low because the overall crew failure rate was less than 34 percent, the crews were not performing licensed duties, and the failed crews were appropriately retrained and retested prior to being returned to licensed duties (Section 1R11).

Report Details

1. REACTOR SAFETY

R11 Licensed Operator Requalification

a. Inspection Scope

This inspection evaluated licensed operator performance in mitigating the consequences of events, since poor licensed operator performance results in increased risk through increased operator recovery rates and licensed personnel-induced common-cause error rates assumed in the licensee's individual plant examinations (IPEs). This inspection effort of the licensed operator requalification program included the following major areas: (1) facility operating history, (2) requalification written examinations and operating tests, (3) licensee training feedback system, (4) licensee remedial training program, and (5) conformance with operator license conditions.

Operator performance since the last requalification program evaluation was assessed to determine if performance deficiencies have been addressed through the requalification training program.

Examination security measures and procedures were evaluated for compliance with 10 CFR 55.49. The licensee’s sample plan for the written examinations was evaluated for compliance with 10 CFR 55.59 and NUREG-1021 as referenced in the facility requalification program procedures. In addition, the inspectors: (1) reviewed the number of applicants and pass/fail results of the written examinations, individual operating tests, and simulator operating tests; (2) interviewed personnel regarding the policies and practices for administering examinations; (3) observed the administration of three dynamic simulator scenarios to one requalification crew by facility evaluators, including an operations department manager assistant and plant manager, who participated in the crew and individual evaluations; and (4) observed two facility evaluators administer two job performance measures in the control room simulator in a dynamic mode.

The inspectors reviewed the licensee's process for revising and maintaining an up-to-date licensed operator continuing training program, including the use of feedback from plant events and industry experience information.

The inspectors verified the adequacy and effectiveness of the remedial training conducted since the last requalification examinations and the training planned for the current examination cycle to ensure that identified licensed operator or crew performance weaknesses during training and plant operations were addressed. Remedial training and examinations for examination failures were reviewed for compliance with facility procedures and responsiveness to address areas failed. The inspectors also reviewed the remediation documented for three individuals, one of which involved two written examination failures and one a simulator examination failure.

Maintenance of license conditions was evaluated for compliance with 10 CFR 55.53 by review of facility records, procedures, and tracking systems for licensed operator training, qualification, and watchstanding.

b. Findings

Written Examination Administration

The inspectors found that the written examinations developed for the 2000 requalification cycle were adequate; however, the examinations were compromised during the validation process. This finding was preliminarily determined to have low-to-moderate safety significance (White). Final significance determination of the apparent violation is still under NRC review. An apparent violation of 10 CFR 55.49, “Integrity of Examinations and Tests,” was also associated with this finding.

The licensee’s 2000 requalification written examination was developed from the requalification training program sample plan. From this sample plan, 35 learning objectives were selected on the percentage of time that the various objectives were taught. Those same 35 learning objectives were used for each of the seven examinations that were developed and administered.

The five operating crews were given their respective examinations over the first five weeks of training cycle 00-20, one crew the first week, the next crew the next week, and so on. Inactive and other staff active licensed operators were given their examinations in the sixth and seventh weeks of the cycle. The first-week crew performed validations of the second-week crew's examination, and the second-week crew performed validations for the third-week crew's examination, and so on. The fifth-week crew also validated the examination for the staff crew of the sixth week. The crews were administered their respective examination the day after completion of validation of the next week’s examination.

The validation was conducted by the entire crew, with half of the crew taking one half of the examination to be validated, and the other half of the crew taking the other half of the examination. The entire crew then met together with members of the training staff to review comments on each question, if there were any, and gather validation times. This validation methodology did not continue after the fifth week.

To ascertain if the examination was administered in an equitable and consistent manner, the inspectors compared each question on each examination to the corresponding question validated by that crew. The inspectors noted that the examinations had numerous similarities week to week. First, the order in which the learning objectives were tested was not changed from week-to-week. The first question on the examination administered during the first week tested the same objective as the first question on each of the subsequent week’s examinations. Second, more than half of the questions on the first week’s examination for both the senior and reactor operators were very similar and tested the same knowledge as their counterpart questions on the second week’s examination. The inspectors' analysis showed that four

of the five senior operator examinations and one of the five reactor operator examinations contained more than 50 percent of questions that were similar in knowledge required to answer the question as those that the respective crews validated.

The inspectors' independent analysis of the similarity between requalification examinations was conducted as follows: Each set of questions was reviewed to determine what skill or knowledge was being tested, and if that skill or knowledge was similar to that required to answer the question that was validated by a crew. The inspectors then compared crew performance week-to-week to determine if: a) any crew that had validated a similar question performed substantially better than crews that had not validated a similar question, or b) any crew that had not validated a similar question performed substantially worse than any crew that had validated a similar question. This was done for each crew and question set on the examination. There were 35 questions on each examination, 7 of which were unique to the senior operator examination, 7 unique to the reactor operator examination, and 28 questions common to both examinations. Therefore, a total of 42 question sets were reviewed.

The inspectors concluded that the validation process combined with the similarity of the examination questions resulted in a significant potential for inequitable and inconsistent examination administration. To assess the impact of this finding, the inspectors evaluated each operating crew's performance on the written examination from question to question.

Of the 42 question sets, the inspectors' analysis determined that at least 8 questions demonstrated clearly that the examination was not administered in an equitable or consistent manner. Four of these questions were common to both the senior reactor operator and reactor operator examinations, 3 were unique to the senior reactor operator examination, and 1 was unique to the reactor operator examination. Crew performance on Questions 10, 15, S31, S32, and R35 supported a conclusion that the examination was not administered in an equitable and consistent manner.

For example, Question S31 on the senior reactor operator examinations administered during the first three weeks of the examination cycle were determined to be similar by the inspectors. Each question required the examinee to determine what sodium pentaborate solution concentration was required for system operability given either standby liquid control tank temperature or level. This question was answered by interpreting one of two graphs in Technical Specification 3.1.7. Question S31 on each senior reactor operator examination administered during the fourth through seventh weeks of the examination cycle was also similar to one another, but were very different from the first three week's questions. These required the examinee to evaluate rod scram time testing data and evaluate what technical specification action may be required. Each of these last four questions required reference to Technical Specification 3.1.4. The learning objective being tested by Question Set S31 was knowledge of entry level conditions for technical specifications. All crews performed very well on this question, with the exception of Crew 3, which was the only crew that did not validate a similar question, since they validated the week four question. The inspectors concluded that the crew performance on this question supported a conclusion that the examination was not administered in an equitable and consistent

manner. No evidence supporting that Crew 3 should have done more poorly on this question than the other crews for any reason other than the validation process was identified.

During the administration of the 2000 requalification examinations the licensee observed that the content and structure of the examinations coupled with the validation methodology, created a potential for pre-conditioning the crews. Problem Identification Report (PIR) 4-10812 entered this observation into the licensee's corrective action process in August 2000. The inspectors found the resolution of PIR 4-10812 narrowly focused on adherence to station procedures and did not effectively investigate the potential compromise. In August 2001, a licensee internal notification again questioned the potential compromise issue. The root cause report that followed this notification again failed to address the 2000 requalification examination compromise issue, and made it a long-term corrective action to investigate the potential compromise and its effects on the 2000 requalification examination. The licensee's final analysis of the compromise issue, which was entered into the corrective action process in November 2001, concluded that the results of the question analysis did not support a finding of compromise, and that no further actions were warranted. The facility analysis results are provided in Attachment 2.

The inspectors reviewed the licensee's final examination question analysis that was provided to the NRC on November 15, 2001. The inspectors noted that for 18 of the 42 question sets, the licensee's analysis of examination question similarity between crews agreed with the inspectors' analysis. For 9 additional question sets, the analysis differed by only 1 question of the set, and for 4 others the analysis differed by 2 questions of the set. Therefore, the facility's and inspector's similarity analyses were in substantial agreement for 31 of the 42 question sets.

While the licensee's analysis was very similar and consistent in many cases with the inspectors' analysis, the licensee's conclusion was nonconservative, in that, it did not address whether the examination was administered in an equitable and consistent manner. This nonconservative conclusion appears to have been based on an inappropriate interpretation of the data. For example, the licensee's analysis stated that there were 6 questions for which an apparent compromise could not be ruled out based on the results. The analysis also concluded that for 6 questions, where an apparent advantage should have been realized, the results showed that crews who reviewed a similar question performed worse than a comparative crew that reviewed a different question. The licensee analysis stated that this inconsistency supported a conclusion that no compromise occurred on the examination. The inspectors did not agree with this conclusion because the validation method itself would produce inconsistent results, in that, the degree to which any individual in a crew gained benefit could vary significantly, and that compromise could not be ruled out for at least 6 questions. The inspectors concluded that if the examinations were regraded with the removal of questions that showed evidence of compromise, at least two individuals would have failed. These individuals were returned to shift duties without the retraining and re-evaluation that would have been required for an examination failure.

10 CFR 55.49 states, “Applicants, licensees, and facility licensees shall not engage in any activity that compromises the integrity of any application, test, or examination required by this part. The integrity of a test or examination is considered compromised if any activity, regardless of intent, affected, or, but for detection, would have affected, the equitable and consistent administration of the test or examination. This includes activities related to the preparation and certification of license applications and all activities related to the preparation, administration, and grading of the tests and examinations required by this part.” 10 CFR 55.59 (a)(2) requires each licensee to pass a comprehensive requalification written examination which covers a 24-month requalification program.

Contrary to the above, the licensee developed and administered their 2000 Biennial Licensed Operator Requalification Examinations in a manner which compromised the integrity of the examinations for all five operating crews that took the examinations. The development of similar examinations and subsequent use of the operating crews to validate the next week’s examination prior to taking their own examination resulted in a compromise of the integrity of the requalification examinations. The two staff crew examinations were not compromised during the validation process. This was an apparent violation of 10 CFR Part 55.49 (AV 50-298/0112-01).

The inspectors determined that the examination compromise was more than minor. Failure to administer an equitable and consistent requalification examination as part of a continuing training program had a credible impact on safety because operators were returned to shift duties without demonstrating an adequate level of knowledge. Failure of operators to properly conduct licensed duties could credibly affect the operability and availability of mitigating systems.

This operator requalification written examination security compromise was preliminarily determined to have low to moderate safety significance (White) using Appendix I to NRC Manual Chapter 0609, “Operator Requalification Human Performance Significance Determination Process.” The inspectors determined that the validation methodology used by the licensee combined with the similarity of the questions from examination to examination resulted in the compromise in the security of the examinations. The inspectors further determined that once the issue was discovered by the licensee in August 2000, immediate compensatory actions were not taken.

Operating Test Administration

During the 2000 Annual Operator Requalification Operating test, two out of seven total crews failed their dynamic simulator test. The finding was determined to be more than minor because inadequate crew performance could adversely alter expected initiating event frequencies, or given an accident sequence, demonstrate an inability to carry out expected actions to mitigate the consequences of the accident. Both conditions would increase the overall plant risk and have a credible impact on safety (FIN 50-298/0112-02).

Using the Operator Requalification Human Performance Significance Determination Process in NRC Inspection Manual Chapter 0609, Appendix I, the inspector determined the finding concerning crew failures to be of very low safety significance (Green). This determination was based on the fact that two of seven operating crew failures in August 2000 were Green, as determined by the matrix, the prior year operating test had no findings of Green or higher, and that the failed crews were appropriately remediated before being returned to shift work.

No additional findings were identified.

4OA6 Meetings, including Exit

The team discussed preliminary findings and other minor observations with Mr. Wilson, Vice President, Nuclear, and other members of the licensee's staff, on October 18, 2001. A final telephonic exit meeting was conducted on December 3, 2001, when the team discussed these findings with members of the training and operations staffs. The licensee provided no further comment on other findings or issues.

Licensee management did not identify as proprietary any materials examined during the inspection.

ATTACHMENT 1

KEY POINTS OF CONTACT

Licensee

M. Boyer, Senior Manager, Technical Services
T. Chard, Manager, Radiation
J. Christensen, Assistant Training Manager
D. Cook, Manager, Training
M. Coyle, Assistant Vice President
P. Flemming, Manager, Licensing
J. Fox, Assistant Operations Manager
R. Gardner, Senior Manager, Quality Assurance
M. Gillan, Manager, Work Control
B. Houston, Manager, Quality Assurance Operations
J. Hutton, Plant Manager
D. Kimball, Assistant Manager, Radiation
D. Kunsemiller, Manager, Risk and Regulatory Affairs
D. Linnen, Senior manager, Training
W. Macecevik, Manager, Operations
D. Pease, Operations Supervisor
M. Schaible, Supervisor, Operations Training
D. VanderKamp, Licensing
D. Werner, Lead instructor, Licensed Operator Requalification
N. Wetherell, Assistant Plant Manager
D. Wilson, Vice President, Nuclear

NRC

M. Hay Resident Inspector

ITEMS OPENED AND CLOSED

Opened

50-298/0112-01	AV	Licensed operator requalification written examination compromise involving an apparent violation of 10 CFR 55.49 (EA-01-298)

Opened and Closed

50-298/0112-02	FIN	Two of seven simulator operating test crew failures occurred during the 2000 annual requalification operating test.

DOCUMENTS REVIEWED

The following documents were selected and reviewed by the inspectors to accomplish the objectives and scope of the inspection and to support any findings:

Procedures Reviewed:

Operation Desk Guide (ODG) 206, Development of Non-Faulted and Faulted JPMs

Operation Desk Guide (ODG) 210, Operations Department Examination Security

Operation Desk Guide (ODG) 213, Annual Licensed Operator Requal Examination Develpment

PIR OTHER NAIT 4-10812, 8/10/2000

Nuclear Training Procedure (NTP) 4.1, Training Material Development and Revision

Nuclear Training Procedure (NTP) 4.2, Examination Development

Nuclear Training Procedure (NTP) 5.2, Examination

Surveillance Procedure 6.SLC.601, “SLC Tank Sampling,” Revision 2, Nov 26, 1998

RCR 2001-0715, “Potential Examination Compromise (Near Miss) During the Licensed Operator Requalification Biannual Written Examination - 2000.” Aug 6-16, 2001

NPPD: Notification: 10096568, 7/10/2001

Results Summary for 00-20 Requal Written Examinations

Weeks 1-7 Licensed Operator Requalification Written Examinations, Cycle 00-20

ATTACHMENT 2

Facility Analysis Results

ACTION ITEM TRACKING STATUS FEEDBACK FORM

CAP ID NUMBER: RCR 2001-0715

CAP ACTION NUMBER: #7

SAP CAP ORDER NUMBER: 4208368

SAP ORDER TYPE: CAP (CAP Order, Tech Support Order)

ASSIGNED DEPARTMENT: TRAINING

ACTION REQUESTED:

[X] Closure [ ] Extension [ ] Action Owner Transfer

[ ] Resp. Transfer [ ] Status Update [ ] Reopen PMActType: _____

(Present Over Due Date: / /) [ ] (New Over Due Date: / /)

RESPONSE/EXTENSION/TRANSFER JUSTIFICATION:

Detailed analysis is attached. Also attached is the revision to ODG 213 "LO Requal Exam Development" which contains specific guidance on exam validation.

SAFETY IMPACT/GENERIC IMPLICATIONS:

No safety or generic implications.

SUBMITTED BY: Mark Schaffh DATE: 11/14/01

RESPONSIBLE MGR: Tom Christensen DATE: 11/14/01

OTHER REVIEW #1: _____ DATE: _____

OTHER REVIEW #2: _____ DATE: _____

OTHER REVIEW #3: _____ DATE: _____

Order 4208368

RCR 2001-0715 #7 - REOPEN ORIGINAL ANALYSIS. A SECOND, MUCH MORE DETAILED ANALYSIS IS BEING COMPLETED (THAT SUPPORTS THE ORIGINAL ANALYSIS) AND THIS SHOULD BE CAPTURED IN THE CAP SYSTEM.

November 8, 2001

2000 Licensed Operator Requal Analysis

The following report was completed the week of November 5, 2001 following a conference call held between CNS and NRC Region IV personnel. During this discussion a request was made by the Region for additional detailed information regarding the 2000 licensed operator exam validation issue. More information is provided in this report concerning the specific process used to validate the written examinations, and the possible effects of that process on the exam results. Different results are provided based upon the different assumptions made during the analysis. These assumptions are critical in determining the actual effect on the examination results and will be discussed in detail later in the report. The results of this additional analysis support the results from the original analysis.

Determination of Similar Questions

The first step used in conducting the exam analysis was to determine which questions were potentially similar in the 42 question sets. This analysis in itself is very subjective. To date, there have been five independent individuals who have reviewed the question sets and each one came up with a different list of questions that are similar. One of the five individuals used to evaluate this issue was an independent contractor with extensive NRC exam development experience. Most recently, a group of four individuals reviewed the question sets to determine which questions are potentially similar, and then detailed discussions took place between the four individuals to reach a consensus on the similarity. It should be pointed out that this review was extremely conservative, and any possible similarities, no matter how unlikely, were labeled as potentially similar. These extreme assumptions were made to ensure that the worst-case analysis would bound the impact of the problem. In general, if the methodology used to answer two questions was the same, it did not always mean the questions were similar. If the methodology was the same and the question covered the same specific knowledge, it was considered a potentially similar question. Another criteria used to determine similarity was if the validation question contained any possible information that answered the question seen on the individual's exam. The detailed analysis below will only discuss those questions deemed to be potentially similar in successive weeks (with the exclusion of question sets six and seven, which were given the same week). Questions that are similar on non-consecutive weeks have no bearing on the analysis.

Questions sets 1, 2, 4, 6, 7, 8, 11, 12, 14, 17, 19, 20, 21, 23, 26, 27, S33, S34, S35, R29, R32, R33, and R34 have no similar questions on consecutive examinations.

Question sets 3, 5, 9, 10, 13, 15, 16, 18, 22, 24, 25, 28, S29, S30, S31, S32, R30, R31, and R35 have potentially similar questions on consecutive exams. Below is a detailed analysis of these questions.

Assumptions and Analysis Methodology

As is stated above, depending on the assumptions made, there is a range of impact on the results of the examination. To ensure reasonable and accurate assumptions are being used, significant discussions with the instructors involved in the validation process have taken place, and all of the operators were also interviewed. Based on the results of the interviews, the validation process was conducted as listed below:

1. The exam was divided into halves to distribute to the crew to review. In most cases this was done using odd/even questions, but in some cases it was done using first half/second half, or other similar ways. While it cannot be totally ruled out, it is not believed that the complete exam was ever handed out to a group of individuals. All of the individuals who took the exam were interviewed, and three individuals indicated they thought they remembered reviewing the entire exam, while everyone else remembered taking only a portion of the exam. In addition, the instructors conducting the validation remember only giving half of the exam to each individual.
2. The individuals took their portion of the exam and then handed it in to the instructor.
3. The instructor then graded each exam and when an individual missed the question, the instructor noted the correct answer on the answer sheet/exam. The instructors also reviewed the overall results from all exams to determine if there were any potential 'problem' questions. These questions were noted as potential issues if more than half of the individuals missed the question. The instructors stated that these averaged 5-6 questions per week. No additional focus was necessarily put on these during the review, they were just flagged as potentials.
4. The exams were returned to the individuals and the instructor asked "Are there any questions/concerns on number 1, 2, 3, etc.?" It should also be stated that the individuals taking the exam were not exposed to the answer key that gave explanations why a correct answer was correct and why incorrect answers were incorrect. It was noted by all of the instructors that if an individual missed a question because of a knowledge deficiency, the question was not discussed in any detail. All of the instructors noted (independently) that there were relatively few discussions during the review process. The only discussions that took place were if there was a challenge to the technical correctness of the question, or if there was a concern with wording, layout, etc. All instructors unanimously agreed that, on average, there were only three or four questions discussed each week. They also unanimously agreed that of these three or four discussions, half were on technical accuracy, and half were on wording of the stem, wording of distractors, etc.

Understanding the above validation process is critical to developing the correct bases used during the analysis. It would be a correct statement that in a vast majority of cases, an individual was only exposed to fifty-percent of the next week's written exam. It would be incorrect to assume that all individuals were exposed to all questions. The following questions were asked when conducting this analysis:

- Which questions on consecutive exams are potentially similar?

- Did the individuals validating a similar question score higher than the average on this question?
- How did the individuals taking completely dissimilar exams score?
- Are there other items involved that would affect the average score? (e.g. two correct answers for the question)
- Is the sample size statistically significant? (e.g. if only two individuals answered a question correctly, can a conclusion of advantage be statistically shown?)
- What is the impact on the results if an individual question is thrown out?
- What is the overall cumulative impact of worst-case for each individual question?
- What do the overall statistics show? As an example, if ten similar questions are found and individuals score higher than average on five and lower than average on five, what is the conclusion? Would it be correct to assume the worst for the five where the crew scored higher than average and ignore the other data?

Below is a detailed discussion of the questions that were identified as potentially similar.

Question set 3

Question set three contains questions that relate to time-to-boil calculations and from that sense the questions are all potentially similar. All questions contain different initial conditions. Questions 1, 5, and 6 all require the student to use a different graph to get the correct answer. Questions 2, 3, and 4 all use the same graph. All candidates answered this question correctly.

Analysis: While all of the questions are potentially similar, since the assumption has been made that approximately half of the individuals had no prior knowledge of this question, it is difficult to reach any statistical conclusion. It should also be noted that this question is relatively easy and it would not be unexpected to have all candidates get this question correct.

Worst Case: If just question three is thrown out for weeks one through five, three licensed individuals who received 80% will have their scores drop to 79.4%.

Question set 5

Question set five, exams two and three, and four and five contain potentially similar questions. Exams two and three contain questions on determining the SRV tailpipe temperature for a constant enthalpy process. Exam two, however, also requires the determination of the 'state' of the steam (i.e. superheated, saturated). The individuals taking exam two averaged 61.5%. Questions four and five both require the determination of specific enthalpy of steam for a constant enthalpy process. In question four it must also be deduced what the steam quality is based upon reactor water level. The individuals that took exam four averaged 100%. The remaining individuals averaged 73.5%.

Analysis: All of the questions in set five require the ability to use the steam tables for a constant enthalpy process. Each question requires the individual to analyze a set of initial conditions, remember to convert psig to psia, and accurately plot and retrieve data from

the steam tables. Because the scores on this question are so divergent (range from 57.1% to 100%), it is difficult to draw any statistical conclusions. In addition, the individuals taking exams two and four seem to have gained roughly the same advantage, yet one group averaged 61.5% and one averaged 100%.

Worst Case: If just question five is thrown out for week four, there is no impact on the overall results for these individuals.

Question set 9

Question set nine contains questions on conditions that would require the declaration of a General Emergency. From a methodology standpoint, all of these questions are potentially similar. Exams five and six are potentially similar because both questions contain a failure of two out of three fission product barriers (fuel failure and a steam leak). The individuals that took exam five averaged 100%. The overall average for this question 83.3%.

Analysis: The individuals taking exam one and three also averaged 100%. It could also be argued that questions four and five are similar and yet the individuals taking exam four averaged 37.5%. Since the assumption has been made that approximately half of the individuals had no prior knowledge of this question, it is difficult to reach any statistical conclusion.

Worst Case: If just question nine is thrown out for week five, one licensed individual will drop from 80% to 79.4%.

Question set 10

Question set ten, examinations two and three, and four and five are potentially similar. Exams two and three deal with an ATWS condition at high flux and the required actions to reset the reactor scram. The individuals that took exam two averaged 84.62%. Exams four and five deal with the expected response to a variable leg failure of a level instrument. The group that took exam four averaged 87.5%. The overall average for question set 10 was 77%.

Analysis: As can be seen by the overall percentage on this question, the sets of individuals taking exams two and four did score higher than the overall average. In addition, these two sets of individuals scored higher than any other groups on this question. While these two groups did not average 100% and the difference is not statistically significant, the numbers cannot rule out a possible advantage for these two groups.

Worst Case: If just question 10 is thrown out for groups two and four, no impact on any individual exam result would occur.

Question set 13

Question set thirteen contains questions on how reactor water level instrumentation will change in response to various temperature changes when given initial level readings. All of these questions require the use of different level instruments and EOP graphs except for exams three and four. Exams three and four, however, require different knowledge to answer correctly. Exam three deals with shutdown range instruments and a drywell temperature increase. Exam four deals with shutdown range instruments and minimum indicated level. After much discussion, it was determined that these questions are not similar.

Analysis: While all of these questions deal with level instrument ranges, there is enough difference in the required knowledge between consecutive week exams to make these questions dissimilar. That fact coupled with the fact that that approximately half of the individuals had no prior knowledge of this question rules out any statistical measurable advantage.

Worst Case: No impact.

Question set 15

Question set fifteen, exams four and five have potentially similar questions. While exam four requires knowledge that the bypass valves must start to open within 0.1 seconds. Exam five requires knowledge that the bypass valves must be 80% open within 0.3 seconds. The individuals taking exam four averaged 100% for this question. The overall average for this question was 83.5%.

Analysis: While exams four and five contain potentially similar questions, each question requires a distinctly different knowledge to answer correctly. It was noted that knowing the correct answer to five may help eliminate a distractor in exam four. We believe this is an extremely subtle point for any individual to catch, and because only half of the individuals would have been exposed to this question, we believe no advantage was gained. In addition, two other groups that did not have any advantage also averaged 100% on this question.

Worst Case: If just question fifteen is thrown out for group four no impact on any individual exam results would occur.

Question set 16

Question set sixteen contains potentially similar questions on exams one, two, and three. These questions all concern turbine startup and hold times required based upon metal temperatures. The individuals taking exam one averaged 100% on this question. Individuals taking exam 2 averaged 84.6%. The overall average on this question was 95%.

Analysis: While individuals taking exam one and two validated similar questions, the results for these two groups would lead to different conclusions. In one case the group

scored higher than average and in the other case the group scored less than the average. In addition, those individuals taking exam three who validated a dissimilar question also averaged 100%. No statistical evidence of an advantage can be shown.

Worst Case: If just question sixteen for exam one is thrown out, there is no impact on any individual exam results.

Question set 18

Question set eighteen exams three and four contain potentially similar questions. Both questions require knowledge of LPCI injection during shutdown cooling. The exam four stem actually contains the answer to the exam three question. The individuals taking exam three averaged 62.5%. The overall average for this question is 80.7%.

Analysis: If an advantage was gained on this question, it would be expected that the individuals that took exam three would statistically show this advantage. That is not the case. In fact, the individuals taking exam three did significantly worse than the remaining individuals.

Worst Case: Not applicable.

Question set 22

Question set twenty-two exams one and two contain potentially similar questions. Both questions require knowledge regarding the ability to restart reactor recirculation pumps with power above or below 1%. The individuals taking exam one averaged 62.5%. The overall average for this question set was 85.8%.

Analysis: If an advantage was gained on this question, it would be expected that the individuals taking exam one would statistically show this advantage. That is not the case. In fact, the individuals taking exam one did significantly worse than the remaining individuals.

Worst Case: Not applicable.

Question set 24

Question set twenty-four exams two and three, and exams five and six contain potentially similar questions. Exams two and three require knowledge of requirements for tripping the turbine if temperatures reach the required setpoint. The individuals taking exam two averaged 84.6%. Exams five and six require knowledge of actions to take during startup with high turbine exhaust hood temperatures. Exam five question 24 was changed to accept two correct answers, and the average for these individuals was 100%. The overall average for this question set was 86.7%.

Analysis: As questions two and three are similar, the individuals taking exam two may have gained an advantage. The individuals taking exam two, however, did not show any

statistical advantage. In fact, they averaged less than the overall average for this question, so no advantage can be shown. The individuals taking exam five answered a question with two correct answers, which negates the ability to determine if any statistical advantage occurred. No advantage can be demonstrated in either of these cases.

Worst Case: If just question twenty-four for exam five is thrown out, one licensed individual will drop from 80% to 79.4%.

Question set 25

Question set twenty-five exams two and three contain potentially similar questions. Both questions require knowledge of what actions will isolate the reactor water sample valves. Exam 2 question 25 was changed to accept two correct answers. The individuals taking exam 2 averaged 100%. Overall, the average for this question set was 94%.

Analysis: The individuals taking exam two answered a question with two correct answers, which negates the ability to determine if any statistical advantage occurred. In addition, a total of four weeks of individuals averaged 100% on this question. No advantage can be demonstrated in this case.

Worst Case: If just question twenty-five for exam two is thrown out, one licensed individual will drop from 80% to 79.4%.

Question set 28

Question set twenty-eight contains several potentially similar questions. All questions in the question set deal with operation of the TIP system in response to isolation signals. In particular, exams one and two, and four and five are potentially similar. The individuals taking exam one averaged 87.5%. The individuals taking exam two averaged 84.6%. Exam four was changed to accept two correct answers. The individuals taking exam four averaged 100%. The overall average for this question is 86.3%.

Analysis: The individuals taking exam one scored slightly higher than the overall average which neither supports or rules out any advantage. While the individuals taking exam four averaged 100%, two correct answers were accepted which statistically rules out any possibility of advantage. It should be pointed out that the individuals that took exam five validated a completely different question and averaged 100% on this question. Overall, the results do not statistically support any advantage gained.

Worst Case: If just question twenty-eight for exam four is thrown out, there is no resulting impact on anyone's exam results.

Question set SRO 29

Question set twenty-nine contains potentially similar questions covering TRM chemistry requirements. Individuals taking exams three and four averaged 100%. The overall range of scores on this was from 40% to 100%. The overall average on this question set was 75.3%.

Analysis: Information received during the first analysis of the exam indicated that two crews had received extensive TRM training after taking their validation exams. It cannot be determined which weeks this occurred, but an assumption was made that the groups that received 100% were the groups that received this training.

Worst Case: If just question twenty-nine is thrown out for exams three and four, there is no overall impact on anyone's exam results.

Question set SRO 30

Question set thirty exams four and five have potentially similar questions. Both of these questions require knowledge of technical specification operating modes, and specifically involve MODE 2. The individuals taking exam four averaged 100%. The overall average for this question set is 93.7%.

Analysis: The individuals taking exam four averaged 100% on this question. Four other groups of individuals averaged 100% on this question as well. While these two questions are potentially similar, since the assumption has been made that approximately half of the individuals had no prior knowledge of this question, it is difficult to reach any statistical conclusion. It should also be noted that this question is relatively easy and it would not be unexpected to have all candidates get this question correct. No advantage can be shown or ruled out for this question.

Worst Case: If just question SRO thirty is thrown out for exam four, there is no resulting impact on anyone's exam results.

Question set SRO 31

Question set SRO thirty-one exams one, two, and three, and exams four, five, and six contain potentially similar questions. Exams one, two, and three require the ability to take data points for SLC and plot them correctly on a technical specification graph. The individuals taking exams one and two averaged 100% on this question. The individuals taking examination three averaged 20% on this question. Exams four through seven require knowledge on scram time testing, slow, and inoperable control rods. The individuals taking exam four averaged 100% and the individuals taking exam five averaged 66.7%. The overall average for this question was 79%.

Analysis: While exams one, two, and three contain similar knowledge, each question requires the application of different data to a T.S. graph. In addition, exam two contained two correct answers, which makes it statistically impossible to show any advantage. The

analysis also showed that of the ten individuals who took exam two, seven originally answered the question incorrectly. It was only after a second correct answer was found that all ten individuals passed. While the overall exam results might indicate some advantage for these two groups, only half of these individuals would have been exposed to these questions. It is more of a surprise that one group did so poorly on this question than the fact that two groups did well on it. This question is quite easy and we would expect the average operator to correctly answer the question. While question 31 on exams four, five, and six initially seemed similar, it was determined that the specific knowledge for each subsequent question was discreetly different, and no advantage could be gained. Each question requires the individual to go to a different technical specification section, note, or procedure to obtain all of the information to successfully answer the question. Exam four requires application of T.S. Table 3.1.4-1 note 2. Exam five requires knowledge not only of T.S. LCO 3.1.4, but requires knowledge of procedure 10.9 and T.S. bases. Exam six requires knowledge of what a 'slow' control rod is. The results from exam five would also confirm that no advantage had been gained. Overall, while an advantage cannot be ruled out for exam groups one, two, and four, the data does not statistically support any advantage gained.

Worst Case: If just question SRO thirty-one is thrown out for groups one, two and four no impact on any individual exam results would occur.

Question set SRO 32

Question set SRO thirty-two has potentially similar questions on exams one and two, and five and six. Exams one and two require knowledge of the MCPR safety limit (power > 25% with pressure less than 785 psig). While the initiating cues are different, these questions are potentially similar. The individuals taking exam one averaged 83.3%. Exams five and six require knowledge of the RPV pressure safety limit. The individuals taking exam five averaged 100% on this question. The overall average on this question was 81%.

Analysis: While the individuals taking exam one and five may have acquired an advantage, only half of these individuals would have been exposed to these questions. In addition, the individuals taking exam six averaged higher than the individuals who took exam one, so statistical data cannot confirm nor rule out any advantage.

Worst Case: If just question SRO thirty-two is thrown out for exams one and five, one licensed individual will drop from 80% to 79.4%.

Question set RO 30

Question set RO thirty exams three and four are potentially similar. Both of these questions require the ability to calculate reactor period using IRM count rate increases. The individuals taking exam three averaged 100%. The overall average for this question was 83.3%.

Analysis: While the exam results for the individuals taking exam three may indicate that they gained an advantage, the sample size is extremely small (three individuals), and

overall ten of twelve individuals answered the question correctly. It should again be noted that only half of the individuals would have been exposed to a similar question. While the statistical results cannot rule out an advantage, they also cannot show any advantage.

Worst Case: If just question RO thirty is thrown out for exam three, one licensed individual will drop from 80% to 79.4%.

Question set RO 31

Question set RO thirty-one examinations four and five contain potentially similar questions. Both of these questions require knowledge of the MCPR safety limit. The group of individuals taking exam four averaged 100%. The overall average for this question was 83.3%.

Analysis: While the exam results for the individuals taking exam four may indicate that they gained an advantage, the sample size is extremely small (two individuals), and overall ten of twelve individuals answered the question correctly. It should again be noted that only half of the individuals would have been exposed to a similar question. While the statistical results cannot rule out an advantage, they also cannot show any advantage.

Worst Case: If just question RO thirty-one is thrown out for exam four, there is no impact on the results of any exam.

Question set RO 35

Question set RO thirty-five exams one, two, and three, and exams four and five contain potentially similar questions. Exams one, two, and three require knowledge of 24 VDC charger operations and setpoints. Exams four and five require knowledge of 125 VDC charger operations and setpoints. The individuals taking exam one, two, and four averaged 100% on this question. The overall average on this question was 83.3%.

Analysis: While the exam results for the individuals taking exams one, two and four may indicate that they gained an advantage, the sample size is extremely small (two, three, and two individuals respectively), and overall ten of twelve individuals answered the question correctly. It should again be noted that only half of the individuals would have been exposed to a similar question. While the statistical results cannot rule out an advantage, they also cannot show any advantage.

Worst Case: If just question RO thirty-five is thrown out for exams one, two and four, there is no impact on the results of any exam.

Cumulative Analysis

The first analysis was conducted using worst-case assumptions. First, it was assumed that all operators had knowledge of all of the questions on the validation exam. Second, it was assumed

that discussions took place on every question during validation. Finally, it was assumed that any score greater than the average on similar questions was solely due to an advantage and should be thrown out. It should be noted that none of the above three assumptions can be corroborated. An analysis using these assumptions was conducted to determine the cumulative impact on the results and to bound the overall possible impact of the validation process. This worst-case analysis resulted in the following question sets being removed from the exam:

- question set 3 exams 1-5
- question set 5 exam 4
- question set 9 exam 5
- question set 10 exams 2 and 4
- question set 15 exam 4
- question set 16 exam 1
- question set 24 exam 5
- question set 25 exam 2
- question set 28 exam 4
- question SRO 29 exams 3 and 4
- question set SRO 30 exam 4
- question set SRO 31 exams 1, 2 and 4
- question set SRO 32 exams 1 and 5
- question set RO 30 exam 3
- question set RO 31 exam 4
- question set RO 35 exams 1, 2, and 4

The overall cumulative result of removing the above questions from the exam is that two licensed individuals who originally passed would now receive failing grades. These individuals both originally received a score of 80% on their exams and this scenario would drop them to 78.8% and 77.4% respectively.

The cumulative impact of removing the above questions is actually less limiting than that of removing question set three for exams one through five. In this situation three individuals would drop from 80% to 79.4%.

An individual case-by-case analysis was conducted using worst-case data, and the worst-case bounded result is that three licensed individuals would drop to less than 80%. It should be noted that the worst-case analysis did not throw out questions that individuals missed. The worst-case analysis only threw out questions individuals got correct. This will be most limiting in all situations. It was also determined that the requirements of NUREG-1021 were still met under worst-case conditions.

Of the three individuals affected by the most limiting situation, none are currently standing licensed duties in the control room. A letter has been sent to the NRC requesting the removal of one of these individuals from licensed duties. The second individual is in the process of having his license removed. The action to remove these individuals from licensed operator duties is unrelated to the exam validation issue. The third individual is standing Work Control Center

SRO duties. This third individual has successfully passed all five written exams given since the 2000 annual written exam.

A second analysis was completed using the most accurate information available based on interviews of all individuals involved in the validation process. First, it was based upon the fact that only 50% of the operators had knowledge of any one question on the validation exam. Second, it was based on the fact that during validation technical discussions only took place on two or three questions per week. Finally, it was determined that while individual question results may show a higher score by individuals that validated a similar question, these must be analyzed cumulatively with the questions where individuals scored lower than average. A conclusion must be drawn using all of the data and not just isolated individual questions. Using these bases, no statistical advantage can conclusively be shown as a result of the validation process. To support this conclusion, it can be shown by the analysis that while some individuals scored higher than average on questions that are similar to ones they validated, there are also examples where individuals scored worse than average. Examples of criteria that were determined to rule out any conclusive evidence of advantage are: 1) Questions where a group scored better than average or averaged 100%, but several other groups also scored better than average or 100%. 2) Groups containing only two individuals make it virtually impossible to determine if a statistical advantage occurred. 3) Questions with two correct answers will skew the statistical results. These questions cannot be used to determine any possible advantage because the odds of getting the correct answer are already doubled.

Eliminating questions that fall in any of the above three categories leaves six exam questions where the possibility of an advantage cannot be ruled out. These six questions are:

- question set 10 exam 2
- question set 10 exam 4
- question set SRO 29 exam 3
- question set SRO 29 exam 4
- question set SRO 31 exam 1
- question set SRO 32 exam 5

However, there are also six questions where a group validated similar questions and their group scored below the average for the question set. In many of these cases the score was well below the overall average. These questions are:

- question set 5 exam 2
- question set 16 exam 2
- question set 18 exam 3
- question set 22 exam 1
- question set 24 exam 2
- question set SRO 31 exam 5

The fact that there is the same number of questions that show a possible advantage and that show no advantage would indicate that no overall advantage occurred. Once again, the key factor is how many individuals were actually exposed to a similar question. The fact remains that the

exposure was limited to approximately 50% of the individuals taking an exam. This fact in turn statistically reduces the possibility of advantage by roughly 50%. This also changes the logic for those questions where groups scored 100% on the exam. For instances where there is a possible advantage, since it most probably only occurred for 50% of the group, the fact that everyone got the question correct would support the conclusion that no one could be shown to have gained an advantage.

Conclusion and Recommendations

Overall, the statistics do not support any consistent advantage gained by the validation process. This process limited the exposure to approximately one half of the candidates, which significantly reduces the ability to determine if any advantage was gained at all. While it is clear that individuals were exposed to some questions similar to ones they found on their exams, it is not clear that any specific advantages were gained.

The 2000 annual operator written examination results were also compared with the results of the previous six cycle written exams. This review was conducted to determine if the annual exam yielded the expected results based upon previous operator performance. In general, operators performed consistent with previous performance, the number of failures was consistent with previous exams, and the average score for the annual exam was lower than the previous six exams by approximately 2%.

Of the three individuals that would be effected by the worst-case analysis, two are in the process of having their licenses removed and the third has an inactive license and is working in the WCC. The third individual has received 100%, 86.7%, 99.2%, 94.6%, and 86.7% on the five written exams since the 2000 annual written exam. It is therefore recommended that no further actions outside of correcting the validation process be taken.