

81 FR 6361-1

2/5/2016

(21)

PUBLIC SUBMISSION

As of: 5/9/16 8:49 AM
 Received: May 06, 2016
 Status: Pending_Post
 Tracking No. 1k0-8ph4-mfji
 Comments Due: May 06, 2016
 Submission Type: Web

Docket: NRC-2016-0006

Operator Licensing Examination Standards for Power Reactors, NUREG-1021, Revision 11

Comment On: NRC-2016-0006-0013

Operator Licensing Examination Standards for Power Reactors; Reopening of Comment Period

Document: NRC-2016-0006-DRAFT-0031

Comment on FR Doc # 2016-07907

Submitter Information**Name:** Michael Petersen**Address:**

1660 Wakonade Drive

Welch, MN, 55089

Email: michael.petersen@xenuclear.com

RECEIVED

2016 MAY -9 AM 8:55

RULES AND DIRECTIVES
FRANCH
1000**General Comment**

Comments are attached.

Attachments

NUREG 1021 Draft Rev 11 Comments Attachment Xcel

SUNSI Review Complete**Template = ADM - 013****E-RIDS= ADM-03****Add=** M. Scheetz (mcs7)

T. Kolb (TCK)

Michael Petersen
Xcel Energy
1660 Wakonade Drive
Welch, MN 55089
Michael.Petersen@xenuclear.com
651-267-7253

The following comments are organized by topical area and in most cases do not list the specific NUREG 1021 changes by section as there are several proposed changes that are addressed under each topical area. It should be clear by title and description which specific changes are included.

Standards for Educational and Psychological Testing

The NRC is deficient in failing to update its examination practices, as delineated in NUREG 1021, by not using the most recent version of the overriding testing standard. From NUREG 1021 Rev. 11, Draft, Appendix A;

"The testing industry endorsed this approach to the development of content valid licensing examinations in the 1985 revision of the "Standards for Educational and Psychological Testing" published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education."

This standard, though artfully crafted 31 years ago, was not intended to be beyond reproach. As a result, it has been updated and improved many times, the last and most recent version being published in 2014. Notably this revision has significant changes from the 1985 version in the areas of content validity, determination of cut scores and workplace testing and credentialing (the specific area that these type tests are governed).

Recommendation: the NRC should perform a detailed review of "Standards for Educational and Psychological Testing (2014)" and make appropriate changes to NUREG 1021 to bring it in line with the most recent standard practices. This should include significant emphasis on the importance of content validity and revisions to the process to ensure that examiner practices, like inappropriate removal of valid questions on "low level of difficulty" are guarded against. Additional significant emphasis should be placed on refocusing the process for determining the written exam cut score on the importance of using currently licensed operators as a benchmark for competence. The current method which relies heavily on examiner judgement is fundamentally flawed; since the examiners have typically never been licensed on the facility where the licenses are being sought.

Proposed changes to scenario overlap requirements

The changes to ES-301 and Appendix D that are associated with scenario overlap limitations are inappropriate and fail to reasonably account for the finite number of appropriate scenario malfunctions and failures. Specifically, the changes in requirements for scenario event overlap would require that a significantly broader set of malfunctions and major events be utilized in subsequent operating tests. Because of the finite number of major events (a handful) and limited number of malfunctions that are appropriate for use on an NRC examination, once these are exhausted (which occurs on exams now) events and malfunctions which have low PRA or are not appropriate for evaluation will have to be

utilized. Using these low probability events or events that do not lend themselves to good evaluations further lowers the discriminatory ability of the exam and will result on inappropriate license decisions.

Recommendation: the NRC should leave the current overlap requirements in place.

Proposed changes to ES-303

There are multiple proposed changes to NUREG 1021 related to the specific grading practices of the Simulator Operating Test that are in direct opposition to Section 107 of the Atomic Energy Act of 1954:

Sec. 107. Operators' Licenses

The Commission shall—

- a. prescribe uniform conditions for licensing individuals as operators of any of the various classes of production and utilization facilities licensed in this Act;

Additionally, NUREG 1021 highlights the importance of and basis for compliance with the *uniform conditions* requirements. From NUREG 1021, Rev 11 Draft, Appendix A, Section B (emphasis added):

The internal attributes of the examination, such as its level of knowledge, level of difficulty, and use of item banks, also impact the operational and discriminatory validity of the examination, which, in turn, can affect its consistency and reliability. **If the internal and external attributes of examinations are allowed to vary significantly, the uniform conditions that are required by Section 107 of the Atomic Energy Act of 1954, as amended, and the basis upon which the NRC's licensing decisions rest are challenged. The NRC must reasonably control and structure the examination processes to ensure the integrity of the licenses it issues.**

Acceptable levels of examination consistency, uniformity, and fairness would be impossible to achieve without quantitative and qualitative acceptance criteria. The examination standards identify many of the quantitative criteria necessary for a well-balanced and consistent examination. Although the NRC's Knowledge and Abilities Catalogs for pressurized and boiling-water reactors (NUREG-1122, 1123, 2103 and 2104) have brought a degree of consistency to the qualitative issue of safety-significance, there is no comparable mechanism to aid in determining an examination's level of knowledge or difficulty before it is administered. **In the end, the validity and consistency of the NRC's examinations depend largely on the individual and collective judgments of the people who write and review the examinations. The discussions herein clarify the intent of the NRC's examination criteria, thereby decreasing the likelihood that inconsistencies among examinations, particularly with regard to the level of knowledge and difficulty, will jeopardize the validity of the NRC's licensing decisions.**

It is clear from the above statements that any significant variance in the attributes of the exams are unacceptable. Considering that the NRC is currently struggling in many areas with consistency not only among examiners in each region but significant differences between regions, any additional allowance for examiner judgment in the process only serves to exacerbate the already unacceptable subjectivity. This increase in subjectivity, as stated above, "...will jeopardize the validity of the NRC's licensing decisions."

The following changes can be addressed together as their impact on the process and reasons for being unacceptable are the same;

- Changes to ES-303 which expand the grading scale from 1-3 to 0-3.
- Changes to ES-303 which remove the "points back" methodology. The following is to be removed from NUREG 1021 Rev 10, ES-303; 'If an applicant makes two errors related to a rating factor, circle an "RF Score" of "1" for that rating factor unless the applicant correctly performed another activity (or activities) related to the same rating factor, in which case the "RF Score" shall remain at "2."'
- Changes to ES-303 which expand the scope of grading for SRO applicants to include every missed TS entry as an individual performance deficiency.
- Changes to ES-303 the clarify there is no limit to the number of rating factors that can be assigned to a single performance deficiency.

The above changes all serve to provide more latitude to the NRC examiners to subjectively grade applicants in a manner that will ONLY result in additional failures. This has been confirmed by the various studies performed by the NRC which have stated that if the new (NUREG 1021 Rev 11, Draft) methodology was used to grade previous successful applicants it would result in additional denials of licenses as a result of failures on the Simulator Operating Test.

The above changes are inappropriate for the following reasons;

- Allowing more latitude in the subjectivity of the grades provided by the NRC examiners is in direct opposition to NUREG 1021 Appendix A and thus Section 107 of the Atomic Energy Act of 1954. Specifically, as quoted, "If the internal and external attributes of examinations are allowed to vary significantly, the uniform conditions that are required by Section 107 of the Atomic Energy Act of 1954, as amended, and the basis upon which the NRC's licensing decisions rest are challenged." In this case, the external attributes of the examinations, the grading criteria, would be allowed to vary significantly depending on NRC examiner judgement. This added subjectively, inappropriately and excessively challenges the ability of the NRC to make appropriate licensing decisions.
- The NRC has acknowledged that using the new grading methodology for the Simulator Operator Test would have resulted in additional license denials for currently licensed individuals. The NRC has not provided any justification why the previous methodology was unacceptable and in need of change. There has been no change to any regulations that would predicate increasing the difficulty of the performance standard. It is not appropriate to make changes to the process for the sole purpose of making it "easier" for examiners to obtain the results they would like.
- The NRC has not provided any data that shows that initial licensed operator performance has been deficient in any way or any individuals previously licensed by the NRC were inappropriately determined to be competent. In fact, it appears by most metrics that operator performance throughout the industry has improved. With this in mind, there has been no valid supporting reason for raising the performance standards for license applicants on the Simulator Operating Test. It is not appropriate to make changes to the process for the sole purpose of making it "easier" for examiners to obtain the results they would like.
- Standard 2.7 of Standards for Educational and Psychological Testing (2014) states; "When subjective judgement enters into test scoring, evidence should be provided on both interrater consistency in scoring and within-examinee consistency over repeated measurements...task-to-task variations in the quality of an examinee's performance and rater-to-rater inconsistencies in

scoring represent independent sources of measurement error." Clearly subjective judgement is a source of measurement error and increasing the subjectivity and thus the measurement error is not only a poor practice but in direct opposition to the uniform conditions requirement and results in less reliable license decisions.

Because of the above reasoning the NRC should make the following changes to NUREG 1021 and the associated processes:

Recommendation: the NRC should discard the above noted changes related to changing the Simulator Operating Test grading and maintain the current (Rev. 10) methodology.

Recommendation: the NRC should provide data and justification for why the current testing methodology is unacceptable for determination of initial license competency and requires changing. At present the only justification has been to make it "easier" for examiners to fail applicants that they "feel" should not pass. Though the emotional state of the examiners is an important input into the final grading, it is already over emphasized and increasing the allowance for examiner feelings only serves to diminish the reliability of the licensing process.

Recommendation: the NRC should provide data supporting the licensed operator performance has either degraded over time or shown a sharp decline which would predicate a change to the examination practices.

Proposed changes relating to Critical Tasks

The changes to ES-301 and Appendix D relating to critical tasks are fundamentally flawed. The following statements from NUREG 1021 Rev. 11, Draft Appendix D are examples;

Note: An unintentional RPS or ESF actuation does not equate to a failure to perform a CT. A post-scenario CT may be created if that unintentional actuation results in a significant plant degradation or significantly alters a mitigation strategy. The active of initiating the actuation is not a CT.

There may be times that either an individual or the crew "create" a new critical task. As will be discussed in Section D, the critical aspect of an unintentional or unnecessary RPS or ESF actuation is to ensure that they do not occur. Therefore, if an unintended protection action does occur during a scenario, or would have occurred, but different applicant prevented it (as described in Appendix E), CT criteria has been met. For example, if an unintended reactor trip or ESF actuation occurs, the CT to "take action during an event to prevent a reactor trip or ESF actuation" was not met.

These changes equate to giving the examiner more opportunity to subjectively create opportunities to downgrade candidates on events that in their opinion should have been performed differently. These decisions will be made AFTER the scenario is complete and thus allow the examiner to create failure criteria that would otherwise have little or no basis. It is clear from the statement above, "the [act] of initiating the actuation is not a CT", that the intent is to NOT give credit for actions performed only to provide additional avenues to fail applicants. A fair use of this practice would be to make every RPS or ESF manual actuation a critical task, however, this does not meet the definition of a critical task, nor does the new proposed methodology.

Recommendation: the NRC should remove the above noted changes related to critical tasks and return to the previous methodology.

Errors in NUREG 1021 Assumptions

There are two significant errors in the assumptions of NUREG 1021 that have existed for several revisions. Though historically these errors have been somewhat compensated for, current examiner practices and the proposed changes to NUREG 1021 result in the unacceptable deficiencies in the examination process and thus lead to inappropriate licensing decisions (failures of applicants that should be determined competent).

The first incorrect assumption of NUREG 1021 is based around the concept of content validity. NUREG 1021 uses the K/A catalogs to determine and limit the content for the initial examinations. The incorrect belief is that;

NUREG 1021 assumes that the initial license candidates will be trained using training programs that include the content of the K/A catalogs and little more. It also assumes that those training programs have not changed or expanded to include training on topics significantly beyond that which is listed in the K/A catalogs. This is based on an outdated belief that utilities would find it acceptable to have failures on the licensing examination and thus would not implement changes to the training programs to improve performance.

The entirety of this assumption is incorrect. Though many years ago the training programs aligned closely with the content of the K/A catalogs, those programs have increased in length, depth and breadth significantly as a result of the systematic approach to training methodology. Specifically, because of the high cost of preparing license candidates to take the NRC exam, it is no longer acceptable (and has not been for many years) to have more than isolated failures on the initial licensing exams. A high failure rate results in millions of dollars in stranded investment and nearly that in additional costs for root cause evaluations and corrective actions. Because of this the training feedback process has increased course content to a level such that courses that several years ago could be completed in 8-10 months are now taking 18-24 months. This has occurred with no corresponding increase in applicant performance; doubling the course length has not doubled the license applicant scores. In reality, applicant performance on the NRC written exams has been relatively stable. This is a result of the inappropriate escalation of the difficulty of the license exams to ensure candidate performance is stable. This is driven by statements in NUREG 1021 Appendix B;

Authors should develop examinations that are estimated to center around the 80-percent cut score level, with individual item difficulty estimated to fall in the 70- to 90-percent difficulty range. (These parameters should not be viewed as precise benchmarks, but rather as approximate end points.) Examination authors should consider the results of past examinations when preparing a new one. Past performance on individual test questions may provide a basis for generating new questions and for estimating the level of difficulty of the examination. For example, questions that everyone answered incorrectly may indicate that the topic did not receive sufficient emphasis in training or that the item was poorly worded. Conversely, questions that everyone answered correctly may indicate that the item was written at too low a level or the distractors were not very plausible.

Training program length, breadth and depth have increased over time which has resulted in applicant knowledge increasing over time which resulted in an associated increase in exam scores. To compensate for this, question difficulty was increased to bring test scores down. The K/A catalog content has been static for more than a decade and thus any increase in exam difficulty to ensure lower scores, when training content has nearly doubled, is inappropriate. Most currently administered NRC initial exams contain flaws that require applicants to possess skills that have no relation to their ability to competently perform the required job skills. This results in an inappropriate bias on applicant skills which results in inappropriate licensing decisions (individuals that should be given a license are denied).

Recommendation: the NRC should change the NUREG 1021 methodology for determining exam difficulty to use methodologies recommended in Standards for Educational and Psychological Testing (2014) which rely more heavily on incumbent performance on examinations and less on developer and examiner judgement. This would include removing recommendations on performance ranges for individual item difficulty, "...with individual item difficulty estimated to fall in the 70- to 90-percent difficulty range..." The new methodology should acknowledge that current training programs are designed to ensure applicant success and that candidate performance greater than 90% on written exams is not only acceptable, but expected.

The second incorrect assumption of NUREG 1021 is related to the benchmark of competence for currently licensed operators;

NUREG 1021 incorrectly assumes that currently licensed operators perform at below the level of competence and thus their performance on an initial license exam should be at or below the cut score of 80%.

This wrong assumption is based on the flawed idea that currently licensed operators are trained to the content of the K/A catalog and little more and that somehow, unlike every other profession, licensed operators become "dumber" once they are initially qualified and spend time performing actual job duties. In reality, currently licensed operators are initially trained well beyond the level of competence and the continuing training program extends that knowledge. Their performance on initial exams should be at or near the perfection mark (100%) and any shortfalls in their performance are indicative of shortfalls in the exam construction. This can be validated by noting that there have been few if any events in the industry for years that rise to the level of a shortfall in competence by currently licensed operators; no actual operator caused accidents that resulted in increased dose to the public, major damage to safety related equipment or a significant degradation in the ability of the plant to respond to an accident.

Recommendation: the NRC should change the NUREG 1021 methodology for determining exam difficulty to use methodologies recommended in Standards for Educational and Psychological Testing (2014) which rely more heavily on incumbent performance on examinations and less on developer and examiner judgement. Cautionary statements should be included in NUREG 1021 that highlight that low validation performance on initial exams by incumbent operators that are currently qualified and proficient are indicative of significant flaws in the exam content.

Generic Fundamentals Examinations (GFE)

The most recent change to the number of offerings of the Generic Fundamentals Examination creates unreasonable hardship on the utilities to align the licensed operator initial training programs with the limited number of administrations.

Recommendation: the NRC should change the Generic Fundamentals Examination process to more closely align with the current NRC initial examination process where-by the utilities develop and administer the GFE on an as needed basis, under the oversight, guidance and approval of the NRC.