To SEE or not to SEE
Intended Purposes vs. Actual Uses of the Self-Evaluation Examination

Timothy Muckle, Nancy Moriber and Gwendolyn Randall
December 2012

NBCRNA
National Board of Certification & Recertification for Nurse Anesthetists

Promoting patient safety by enhancing provider quality.

TABLE OF CONTENTS

1.   Introduction

For over 20 years, the Self-Evaluation Examination (SEE) has been administered by the National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA) to provide formative assessment feedback to nurse anesthesia students and program faculty. While the SEE certainly has value in providing information on students' progress in mastering core concepts of anesthesia, there have been several instances of uses of SEE scores which go beyond their intended purposes. The very first of the joint standards for testing - published by the American Educational Research Association (AERA), National Council on Measurement in Education (NCME), and the American Psychological Association (APA) - deal with validity and making clear the intended uses of test scores: "The test developer should set forth clearly how test scores are intended to be interpreted and used...If the interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be cautioned about making unsupported interpretations" (AERA, 1999). The purpose of this article, in part, is to satisfy NBCRNA's responsibility to clearly delineate the appropriate uses of SEE scores and to establish evidence for the SEE's intended purposes.

First, a brief chronology of the development of the SEE is provided.   Then, we provide the clearly stated purposes of the SEE and explore some uses of SEE scores which exceed the intended purposes, and the reasons why they are problematic. Next, a brief quantitative study uses survey results to understand the predictive validity of the SEE for the National Certification Examination (NCE). Finally, the article presents the results of a recent focus study of students and faculty viewpoints about the SEE, leading to discussion of how the SEE may be modified to meet the needs of future stakeholders.

2. Chronology of the SEE

In April 1987, Sandra Maree, a program director at North Carolina Baptist Hospital, sent a letter to the Council of Certification requesting consideration of an in-training examination. Ms. Maree stressed that "the future of nurse anesthetists was dependent upon the marriage between education and practice."

On May 21, 1987, a memorandum was sent to program directors from the Chairman of the Council on Certification of Nurse Anesthetists (CCNA) announcing it was investigating the feasibility of offering an interim examination for students who desire to practice taking a national examination and to identify areas that may require their further study. Assessment Systems, Incorporated (ASI) was contacted for a proposal and feasibility study. In October 1987, the CCNA agreed to develop a proposal for an in-training exam for either 12/88 or 6/89.

In March 1988, the Council charged the Examination Committee to explore the development of an in-training examination with the testing agency (ASI) staff at their September meeting. In October 1988, the Council agreed to proceed with an in-training exam and rename it the CCNA Self-Evaluation Examination (SEE). At the October 1989 Council meeting, the implementation of the SEE was postponed until June 1991.

In March 1990, test parameters for the Self-Evaluation Examination were approved. The SEE Committee was added to the Council bylaws in September 1990. The Council agreed that CRNA faculty would not be allowed to take the SEE. On June 21, 1991, the first SEE was administered via the American College Testing (ACT) organization (Certification History Notes, 1937-2006).

Between 1991 and 1996, the SEE was a 250-item, paper-and-pencil examination. The examination consisted of 250 multiple choice questions to be answered over a four-hour period. It tested six content areas and two cognitive levels: Level 1 – Recall, and Level 2 - Application, Synthesis and Evaluation. In April 1996, the SEE was converted to a 160-question, computer-adaptive examination (Fagerlund &McShane, 2005).

In October 1994, computer adaptive testing parameters were set for the SEE: (a) a test length of 140 questions with 20 pretest questions per examination; (b) a maximum length of three hours, including tutorial; (c) no option for review; and (d) administration twice a year in September and March within a 2 week window.

In 1996, the SEE parameters were changed (a) 160 multiple choice questions; (b) a maximum length of three hours; (c) no option for review; (d) administration twice a year in May and September, within a four-week window; and (e) a fee of $110.00. On April 15, 1996, the first CAT test was offered at the Sylvan Technology Centers. At the October 1996 meeting, the Council agreed to make the SEE available year-round, effective April 1, 1997. The SEE has been offered on-demand in computerized adaptive test (CAT) format since this time.

3.   Stated Purposes and Current Problematic Uses of the SEE

The Self-Evaluation Examination (SEE), administered by the NBCRNA, has three objectives: 1) to provide information to students about their progress in the nurse anesthesia program; 2) to provide information to program directors on how well their programs are preparing students with the knowledge they need for anesthesia practice; and 3) to prepare students for the National Certification Examination (NCE) experience. Given the current stated purpose of the SEE, there are a number of questionable or problematic uses and interpretations of SEE scores which go beyond the stated, valid uses.

For instance, many nurse anesthesia program faculty desire to use not only the total SEE domain score as a reliable indicator of student knowledge and ability, but also the sub-domain scores as indicators of ability in the smaller content areas (Professional Legal Aspects, Pharmacology, etc.). The sub-domain scores are then used to inform student remediation, with the student emphasizing future study in the content areas in which they scored poorly.

While the desire for and potential value of such information is understandable, the SEE is not constructed in a way to yield this type of information in a reliable manner. In testing, *reliability* refers to the precision of the measurement instrument and the degree to which scores arising from a test can be trusted as indicators of an examinee's true ability or level of knowledge. A number of reliability coefficients are used to quantify the degree of internal consistency that exists in test scores. Reliability is largely dependent on the number of questions we ask examinees, and the appropriateness of the questions given the examinee's ability level. The more questions we ask, the more opportunities we have to observe and understand where an individual examinee lies on a continuum of knowledge attainment.

Generally speaking, a reliability coefficient of 0.80 is considered a rule-of-thumb minimum level for good reliability for test scores. In previous years, annual analysis of the total SEE scores has exhibited good reliability, with coefficients over 0.85, indicating a high level of reliability. This favorable result is partially due to the number of scored questions on the SEE – 100. A large number of scoring opportunities will usually yield a reliable score. However, the sub-domain scores do not benefit from the same level of reliability. The largest content area on the SEE is Anatomy, Physiology, and Pathophysiology, comprising 29% of the exam, or 29 scored questions. The associated reliability with these sub-area scores is, at maximum, 0.72, which is not sufficient for make decisions about student remediation. The problem is, of course, compounded for the other content areas, which comprise even less of the test: Pharmacology (25 items, reliability of 0.68), Basic Principles (14 items, reliability of 0.50), Advanced Principles (19 items, reliability of 0.55), Professional and Legal Aspects (13 items, reliability of 0.43). While these are positive reliabilities, their smaller magnitude does not support the use of the scores for informing future student study. More questions would need to be administered in each content area in order to generate scores with higher reliability. Therefore, interpretation of the SEE scores should be limited to use of overall score as the most reliable indicator of the student's knowledge, and use of the SEE Interpretive Guide as normative data to evaluate the student's standing compared to peers.

Another widespread practice is the use of minimum scores on the SEE to determine a student's continuation in the program. In this practice, the SEE is used somewhat as a final exam. The student must achieve a cutoff score, predetermined by the nurse anesthesia program, in order to "weed out" weaker students from progressing from Year 1 to Year 2, or even in order to graduate from the program and go on to be eligible to take the NCE. Due to the reliability issues associated with the sub-scores, the NBCRNA does not recommend use of the scores in this manner.

The SEE is not intended to be an end-of-year or end-of-program comprehensive exam for students in nurse anesthesia programs. This misuse of the SEE is supported neither in the stated purpose of the SEE in public documents such as the SEE Student Handbook (NBCRNA, 2012), nor in other public documents related to the SEE, including this article. If nurse anesthesia programs wish to administer a comprehensive assessment to assess cumulative student learning, the program must develop and administer themselves an examination appropriate to their curriculum.

Another problem with this practice is that the passing standards (or "cut scores") used by nurse anesthesia programs are determined arbitrarily. When these arbitrary cut-off points are adopted by individual programs, the process for establishing them rarely follows recommended procedures for establishing passing standards on tests (AERA/NCME standards 4.19 and 4.20), which require that a passing standard be a specific criterion closely linked to a meaningful

3

8725 W. Higgins Rd.
Suite 525
Chicago, IL 60631

NBCRNA

855-285-4658
www.NBCRNA.com
708-669-7636 :: fax

performance level on the score scale, and that the rational and procedures for setting the standard be documented. For example, a common practice by anesthesia programs is to adopt a passing standard on the SEE of 450 because this is the passing standard on the NCE. However, the score scales of the SEE and NCE are not empirically or causally linked in any way, and a 450 on the SEE has no equivalence to a 450 on the NCE.

Unfortunately, in the vast majority of cases, the established program-dependent cutoff point for the SEE is not linked to a meaningful performance level, but rather to an arbitrary score. As a result, these programs may be making poorly supported decisions about a student's academic standing.

4.   Quantitative Study

Up until the present time, it has not been clear (a) to what extent examinees prepare for the SEE, (b) what consequences, if any, educational programs are imposing as a result of a student's performance on the SEE, or (c) whether it is advisable to use SEE scores to predict correlation with NCE. Many students and faculty want to use SEE scores as a predictor of students' performance on the National Certification Examination. Predictive use of SEE scores may be problematic because it has not been clear to what extent examinees' SEE scores correlate to NCE scores. If a strong statistical correlation (i.e. a Pearson correlation greater than +.50) could be identified between SEE and NCE scores, then we may say with some degree of confidence that a student with a high score (or low) on the SEE also will score high (or low) on the NCE.

However, it has been difficult to establish such a predictive relationship due to the variation in both student preparation for the SEE and programmatic use of SEE scores. Many students are required to attain a minimum score on the SEE to continue in their program or to advance from Year 1 to Year 2. As a result these students devote a great deal of time to preparation for the SEE so that they may pass their specific program requirements. Many others, however, are not subject to the same requirements, and thus may take the SEE with very little or no preparation, in the hopes of receiving a more critical or honest assessment of their current mastery of anesthesia knowledge. In these instances, a student who performs poorly on the SEE may very possibly go on to "ace" their NCE. This nationwide inconsistency in preparation for the SEE tends to "muddy the waters" of a correlation between scores for the two examinations.

In order to bring the picture of the SEE landscape into sharper focus, a brief post-examination survey has been administered to each SEE test taker (N=4,085) since August 2009. The examinees were asked to respond to the following questions:

Does your nurse anesthesia program require that you attain a minimum score on the SEE in order to continue in the program or to graduate?

○ A.   Yes

○ B.   No

How many hours did you study for the SEE that you took today?

○ A.   I did not study for the SEE exam.

○ B.   Less than 10 hours

○ C.   10-20 hours

○ D.   21-30 hours

○ E.   31-40 hours

○ F.   41-50 hours

○ G.   51-60 hours

○ H.   61-70 hours

○ I.   Over 70 hours

Just under one-third (32.9%) of examinees responded that their program required a minimum score on the SEE to continue in the program or to graduate. Figure 1 depicts the pattern of responses for amount of time studied, overall and broken down by whether the examinee reported a minimum score program requirement. Overall, less than half (43.1%) of the SEE examinees spent more than 20 hours in preparation for the SEE. More than one-quarter (27.8%) reported that they did not

prepare for the SEE at all. Not surprisingly, preparation for the SEE clearly differs depending on whether a minimum score is required for program continuance. Examinees with program-imposed minimum score requirements spend more time in preparation for taking the SEE.
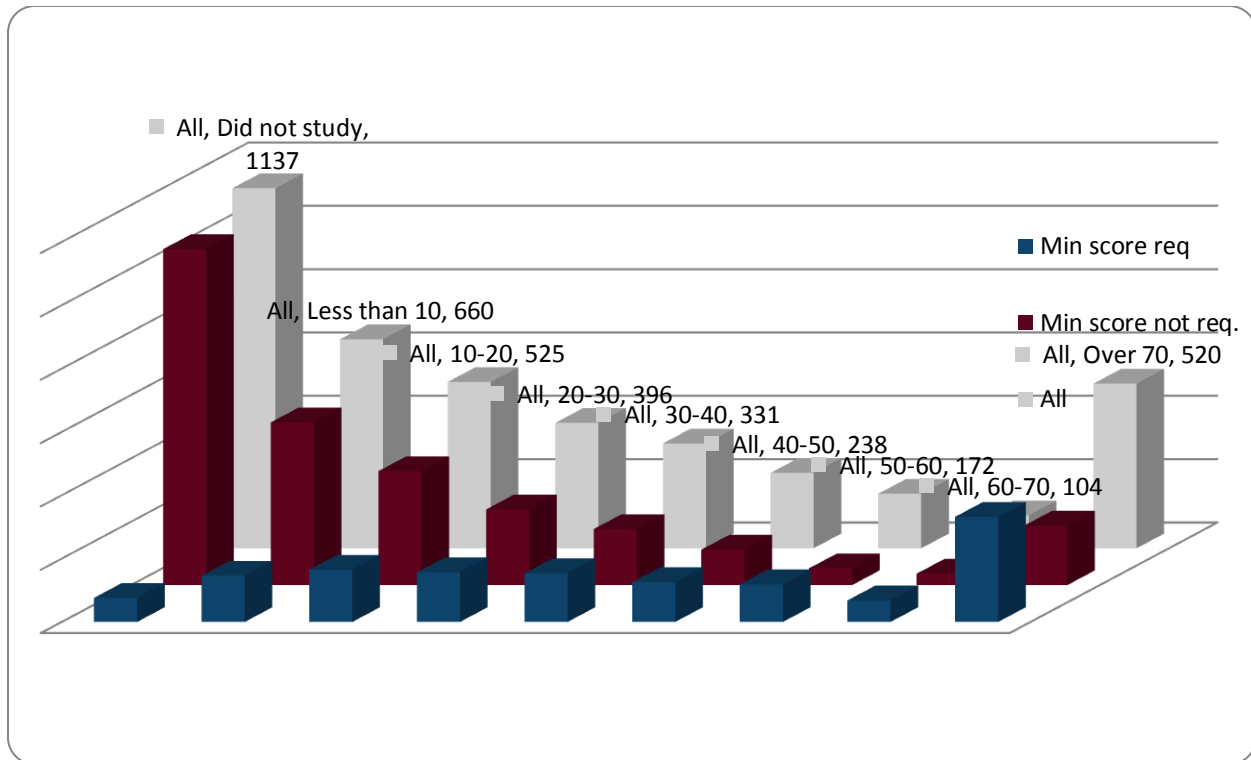

Figure 1. Amount of preparation time (hours) reported by SEE examinees

By pairing the survey responses with SEE score data and year in program, a correlation study was conducted in 2011. A sample of Year 2 nurse anesthesia students who were administered the SEE in the years 2009-2011 was identified. This sample was further refined to include only those students who indicated that they had studied a substantial amount of time (at least 50 hours total) for the SEE. The total SEE scores of this sample were correlated to the same students' total score resulting from their first NCE administration. The observed Pearson correlation was ρ = +0.57, p < 0.01 (statistically significant, i.e., the correlation did not arise by chance). This suggests a moderately strong predictive relationship between SEE and NCE scores for those students who studied at least 50 hours. Figure 2 depicts an X-Y plot of SEE examinees' scores (horizontal axis) with their eventual first-time NCE score.
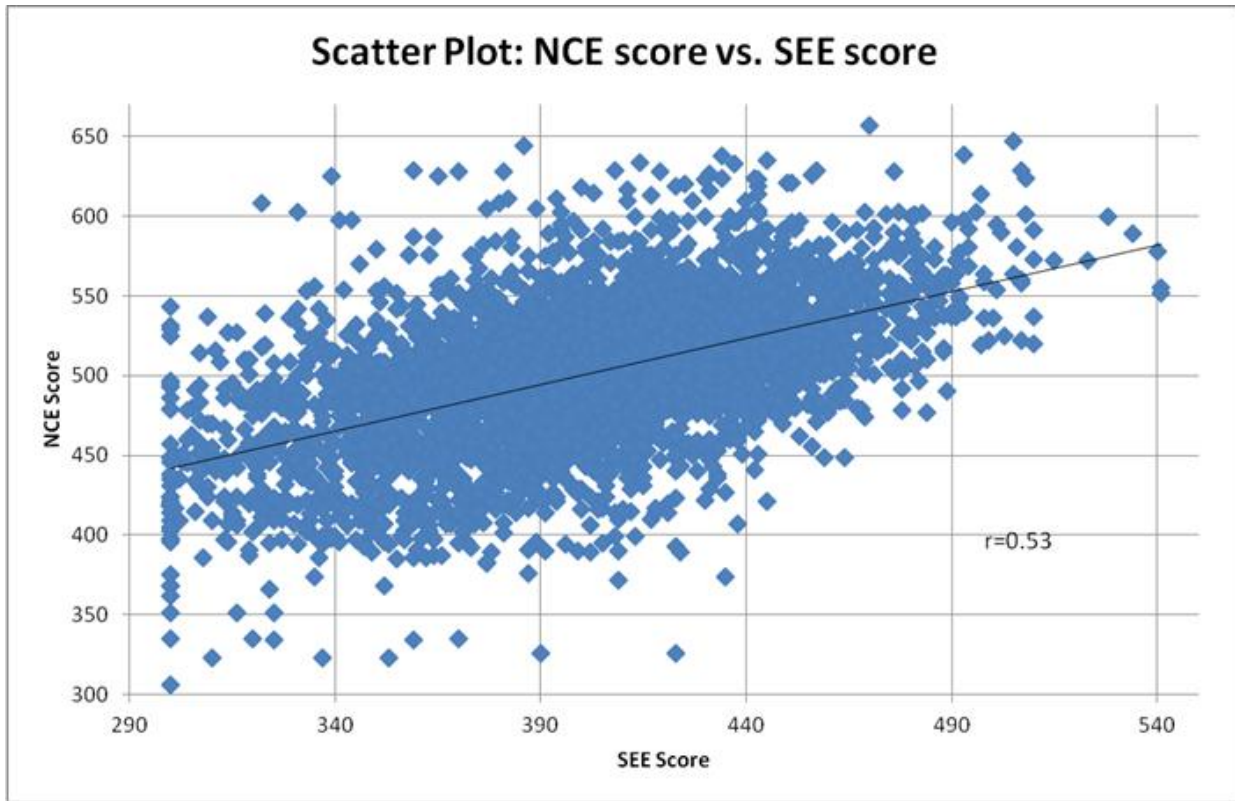
Figure 2. X-Y plot of SEE scores vs. eventual first-time NCE score

5.   Current Views of Faculty and Students

A focus group was held in September 2011 at the New England Assembly of School Faculty (NEASF) Development Workshop to determine the current views of program administrators, didactic instructors and clinical preceptors regarding the current and future uses for the SEE. Seventy-five CRNAs involved in various aspects of nurse anesthesia education in programs predominantly in the northeast were asked three questions about the SEE: (1) How is the SEE currently being used in your program? (2) What do you feel are acceptable uses for the SEE in its current format? and (3) What do believe the "ideal" SEE would look like and what would it do? Prior to this workshop, many attendees had minimal knowledge regarding the structure and psychometric properties of the SEE.

Overwhelmingly, workshop attendees stated that currently the SEE is being utilized by programs as a "self-test" of student performance and as a guide for student remediation in preparation for the NCE. Some programs require students to take the exam in both the first and second years, expecting to see improvement in overall student performance. Others require students to take the exam only once prior to graduation. Most program faculty felt that the SEE served as a "reality check" for students with respect to their existing knowledge base, and therefore probably motivated them to begin studying for the certification exam. However, there did not seem to be any consensus regarding the degree of preparation students were encouraged to complete prior to taking the examination. While all programs required students to take the SEE, none represented at this focus session stated that they utilized it as a tool for weeding students out of programs. Faculty did, however, state that they knew of programs elsewhere in the country that were utilizing the SEE as a benchmark, pass-fail exam.

Program administrators and faculty felt that the "ideal SEE" would help to prepare students for the certification examination by providing sub-scores that reflected student knowledge in each content domain and composite scores that were predictive of future performance on the NCE. They felt that the content outlines of the SEE and the NCE should be aligned so that students would be tested on information that was truly germane to clinical practice.   They also hoped that the students could use this information to guide their own study plans. Finally, faculty felt that the SEE in its current format did not meet their current needs for evaluation of program effectiveness and student performance because programs are structured differently, with some front-loaded and others integrated. Therefore, they would like to be provided with clear recommendations regarding the best time for students to take the SEE and with an interpretation of results based upon program type.

6.   Future Conceptions for the SEE

It is clear that the current use of the SEE has strayed somewhat from the intended purposes set forth by the NBCRNA. Many factors have influenced this shift, including changes to Council on Accreditation (COA) standards for nurse anesthesia education, increasing rigor of state boards of nursing for graduate registered nurse anesthetist (GRNA) practice, pressure placed upon programs to meet mandatory federal benchmarks for first-time NCE pass rates, and fears of punitive action from regulatory agencies if these standards are not met. Programs and students alike are looking for ways of ensuring that students pass the NCE on the first attempt. While we all understand the forces that have led us to this point, the real challenges will come as we strive to move forward and develop the SEE into something that meets the needs of all interested parties: the NBCRNA, program faculty members, and the students who will eventually become the practitioners and educators of tomorrow. Therefore, based upon feedback from the communities of interest, several possibilities for the future of the SEE have been proposed. No one conceptualization is seen as the best solution, and by no means is this list all-inclusive. It does, however, provide us with a starting point for future discussion and idea generation.

One of the most frequently heard wishes of program faculty and students is to be able to utilize the SEE as a future predictor of NCE performance. It appears from the analysis in this paper that there is a moderately strong, positive correlation between SEE and NCE scores for second-year students who spend a substantial amount of time in SEE preparation. If the predictive validity of the SEE were to become a reality for all students, it would first be necessary to develop an examination that is truly capable of predicting future NCE success. This would likely involve first aligning the SEE content outline with that of NCE.

The content outlines of the SEE and NCE are not equivalent. The current SEE content outline is designed to assess students' progression through their respective nurse anesthesia programs, and it reflects curriculum requirements set forth by the COA. The NCE is designed to measure the knowledge, skills and abilities necessary for entry-level nurse anesthesia practice. This shift in focus results in a greater emphasis on basic and advanced anesthesia principles on the SEE versus basic sciences on the NCE. In addition, content related to research principles is omitted, as it is not considered essential knowledge for entry into practice. Because the two examinations do not test the same construct in their current format they should not be used to predict NCE performance. Once the content outline of the NCE is revised using information obtained from the current professional practice analysis, it might be worthwhile for the NBCRNA to adopt the same content outline for the SEE so that the two exams are evaluating equivalent content domains.

Program administrators also currently utilize the SEE to provide guidance and meaningful feedback to students regarding weakness in key knowledge domains in preparation for the NCE. While many physician specialties have developed reliable "in-training" examinations, the SEE in its current format is not capable of providing this type of feedback to examinees. There are not enough scored questions in the SEE's sub-domains to provide reliable information about a student's knowledge base in any one particular content area. If the SEE is to truly be utilized as an "in-training" examination, it would be necessary to increase the number of scored questions in each sub-category, paying particular attention to the content domains with the fewest questions and, therefore, the least predictive value. This would, however, require additional time and resources, including a larger item bank, statistical validation, and longer seat time at testing centers, which could potentially increase the cost of administering the SEE. It is possible that the overall benefits of this type of initiative to students, programs, and the NBCRNA could offset any increased expense. However, a cost-benefit analysis would need to be carried out before such an initiative could be undertaken.

While there are many possible ways in which the SEE can evolve, if it is to provide the information desired by students and program faculty, several key logistical issues will need to be addressed as well. Nurse anesthesia programs come in all shapes and sizes, some are front-loaded, some integrated; some are educating at the master's level, others at the doctoral level, and some are housed within schools of nursing, while others within schools of arts and sciences or allied health programs. Therefore, issues of when to test and how to address different curricular requirements become real ones. When students complete nurse anesthesia educational programs, it is understood that they should all possess an equivalent knowledge base because all programs are required to meet the same standards for accreditation. However, students actively involved in the training process may have vastly different knowledge at the same point in their programs. Therefore, students would need to take the SEE not after a set time period, but when equivalent curriculum content has

been covered. In addition, programs will need to set similar benchmarks for all students scheduled to take the SEE because the degree of preparation and consequences for poor performance would dramatically impact both actual SEE scores and the generalizability of those scores to success on the NCE.

References

American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

National Board of Certification and Recertification for Nurse Anesthetists (2012). *Self-Evaluation Examination Student Handbook*. NBCRNA, Park Ridge, IL.