Use of Resources and Method of Proctoring During the NBCRNA Continued Professional Certification Assessment: Analysis of Outcomes

Dennis Spence, PhD, CRNA; Robyn Ward, PhD, CRNA; Steve Wooden, DNP, CRNA, NSPM-C; Mary Browne, MA; Hao Song, PhD; Robert Hawkins, PhD, CRNA, DNP; Mary Wojnakowski, PhD, CRNA

Objective: The National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA) conducted a study to determine whether use of resources and method of test administration for the Continued Professional Certification Assessment (CPCA) impacts performance, testing experience, and fidelity of the examination. **Methods:** A total of 1,500 certified registered nurse anesthetists (CRNAs) were randomized into one of six groups based on method of proctoring (in person or live online remote) and use of resources (closed book, open book with NBCRNA-provided resources [e-books], or open book with use of a choice of two of nine approved hard copy resources). Participants completed a baseline survey, the CPCA, and a post-CPCA survey. **Results:** A total of 1,217 participants were included in the analyses. Scores on the CPCA were 5 points higher in the resources groups than in the closed-book groups ($p = .0005$), but examination times were twice as long ($p = .0005$). Performance standard rates were similar ($p = .26$). In the resources groups, 84.2% of participants used resources to confirm an answer. In the e-books groups, approximately 40% of participants' comments indicated problems with the e-books. Participants in remote proctoring groups had a moderate number of problems with remote proctoring; thus, scores were lower than expected. However, 55% to 73% of participants indicated remote proctoring would reduce their anxiety on future examinations. Half of all participants strongly agreed or agreed that the CPCA accurately reflected core knowledge all CRNAs should know. **Conclusion:** It is recommended that the CPCA be delivered as a closed-book examination with choice of in-person or remote proctoring that does not require live online proctors.

Keywords: Continued professional certification, electronic books, nurse anesthetists, open book, recertification, remote proctoring, testing

Over the past 20 years, healthcare credentialing organizations have been challenged with the incorporation of lifelong learning and evaluation of performance standards in the processes of licensure and certification for healthcare professionals (Institute of Medicine, 2001). State regulators, hospitals, and the public expect advanced practice nurses to remain competent throughout their career and have the knowledge and skills beyond those needed for entry into practice (Wooden, Krogh, Waters, & Plaus, 2017; Swankin, LeBuhn, & Morrison, 2006).

Medical and nursing credentialing organizations are charged with incorporating more rigorous methods for granting continued certification to include periodic examination for the assessment of knowledge (Brennan et al., 2004; Swankin et al., 2006). In a systematic review that evaluated testing as a method to assess continued competency in nurse anesthesia practice, the authors concluded that if testing was used as a component of recertification, it should be utilized for the purpose of assessing knowledge as one component of competence (Riddle, Baker, & Sapp, 2016). An additional recommendation was to include a means of reflection for the examinee to promote lifelong learning.

As part of its mission to promote patient safety by enhancing provider quality, the National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA) develops and implements credentialing programs that support lifelong learning among certified registered nurse anesthetists (CRNAs). In 2016, the NBCRNA implemented the Continued Professional Certification (CPC) program, which consists of continuing education in the form of Class A and Class B credits, the use of core modules, a 2-year check-in, and an assessment of knowledge (Continued Professional Certification Assessment [CPCA]).

The CPCA is a 150-item computer-based examination designed to assess knowledge in the four core domains of nurse anesthesia practice. The examination used conventional single-best multiple-choice items and multiple-select items that require selection of more than one answer to score (i.e., select two). Participants had up to 4 hours to complete the assessment. The domains and the corresponding weighted

percentages for the CPCA are specified according to the examination blueprint, which is established based on an extensive survey of CRNAs in practice. The domains included (a) Airway Management (34%), (b) Applied Clinical Pharmacology (24%), (c) Human Physiology and Pathophysiology (24%) and (d) Anesthesia Equipment and Technology (18%). CPCA items are considered "walking around" knowledge of experienced practicing CRNAs rather than entry-level knowledge, which is assessed on the National Certification Examination. CPCA questions underwent a rigorous development and review process to ensure each item tests knowledge important in clinical practice, differentiates adequate and inadequate performance, has a verified reference taken from an NBCRNA-approved list of anesthesia textbooks, is relevant to the CRNA population, and has an appropriate difficulty level.

In consideration of feedback from the CRNA population on the connection of the assessment to the practice of nurse anesthesia, respect for CPCA examinees' busy professional lives, and the recommendation that examinees be allowed access to resources (Riddle et al., 2016), the NBCRNA Board of Directors approved a study to evaluate the implications of allowing examinees to have reference resources available during the CPCA and the opportunity to test at different locations while being observed by real-time remote online proctoring through available technology in addition to traditional testing centers.

Emerging research examining aspects of open-book testing, utilization of resources during open-book testing, and the use of remote proctoring was carefully evaluated by the NBCRNA. A comprehensive review of studies comparing open-book to closed-book examinations (Durning et al., 2016) reported three key findings: (a) there was no difference in the amount of examinees' preparation time for either examination format; (b) there was inconclusive evidence overall that open-book examinations lead to lower test anxiety, and (c) there was no statistically significant difference in most studies in terms of examination performance. However, of studies reporting statistical significance, performance was better in the closed-book condition, which the authors attributed to better examination preparation (Durning et al., 2016).

In a randomized controlled trial (Lipner, Brossman, Samonte, & Durning, 2017) examining the effect of closed- and open-book experimental conditions, the investigators concluded that no clear differences were found in test-taking strategies between closed- and open-book conditions using an electronic medical resource (UpToDate). In addition, a survey of participants confirmed that allowing reference searches aligned with the way resources are used in practice and helped to reduce test anxiety, but it did not necessarily reduce test preparation time. A survey of CRNAs conducted by the NBCRNA (Ferris & Muckle, 2018) revealed that CRNAs utilize a variety of professional resources in their daily practice. These resources were primarily web-based, but there was no single pivotal resource used by all CRNAs. These findings gave the NBCRNA options regarding how to best operationalize the open-book test method.

Remote testing has been used in education in various forms but has only recently been applied to the licensure and certification processes (Lipner et al., 2017). Remote test administration may be proctored in real time (live online proctoring), recorded and then reviewed at a later time, or not proctored or reviewed at all (most longitudinal assessments). Literature on remote testing is focused largely on education or is conducted by the vendors who own remote proctoring technologies. Few medical boards are using remote proctoring. The American Board of Internal Medicine (ABIM) launched a record and review remote proctoring platform to securely administer their 2-year Knowledge Check-In (ABIM, n.d.). In contrast, the American Board of Anesthesiology's Maintenance of Certification in Anesthesiology (MOCA) Minute is a longitudinal examination program consisting of 120 unproctored questions annually, with no more than 30 questions answered per quarter, that can be accessed via weekly email reminders, a portal account, or the mobile app (American Board of Anesthesiology, n.d.).

Objectives

The NBCRNA launched the CPCA Beta Research Study in 2018 to compare differences in examinees' performance under different testing modalities (closed vs. open book [e-books or personal hard copy books] and in-person vs. remote live online proctoring). The primary objective of the study was to determine whether the use of resources and type of administration of the CPCA impacted examinee performance (test scores, duration, meeting performance standard). A second objective was to compare self-reported differences in perceived CPCA difficulty, test performance, fidelity of the assessment, and testing experience.

Methods

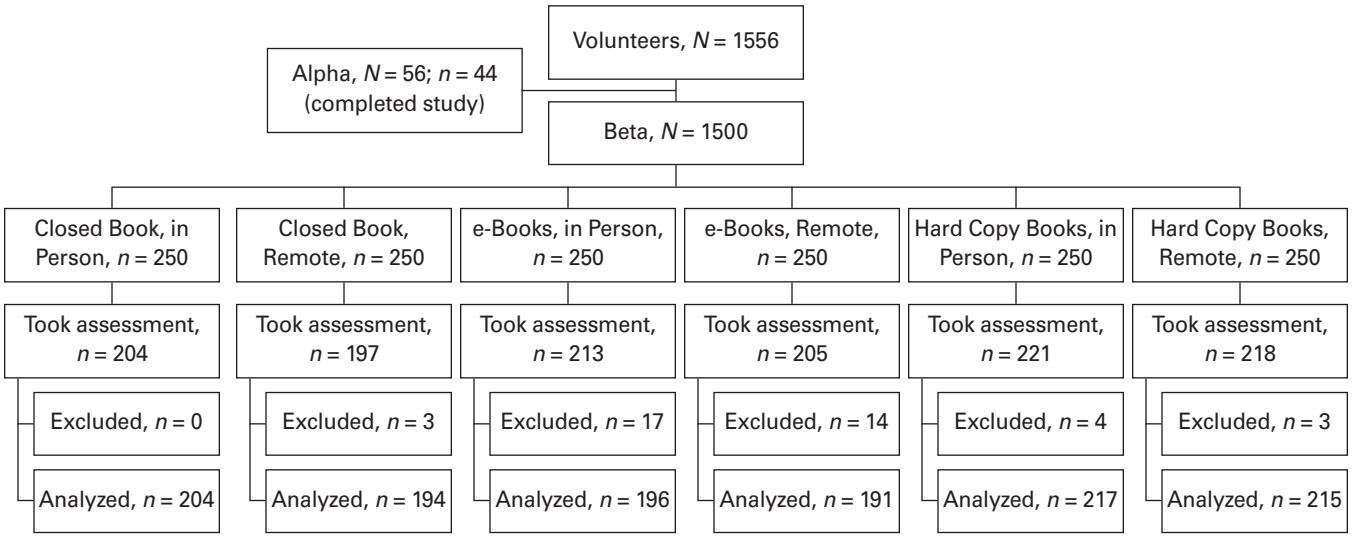
This was a prospective, stratified (gender, age group, and years of certification), randomized (1:1), parallel-group (2 x 3 experimental design, Figure 1), controlled study consisting of three phases: (a) a baseline survey, (b) the CPCA, and (c) a post-CPCA survey. The research protocol was approved as exempt by the Midwestern University Institutional Review Board. Inclusion criteria were CRNAs with full certification with an unrestricted registered nurse license, currently participating in the CPC program, eligible to take the CPCA, willingness to be randomized to different testing modalities, willingness to sign a participation agreement, ability to travel to test center, and having a computer with the required technical specifications for remote proctoring, such as reliable internet connection, webcam, and microphone. NBCRNA Board members, staff, and CPCA item writers were excluded.

Call and Selection of Volunteers

The population for this study included all CRNA examinees who met inclusion criteria. Investigators sent an email request for volunteers for the study on February 12, 2018, to 22,389 CRNAs and again on February 26, 2018, to 22,010 CRNAs via Constant Contact (no overlap occurred in email requests sent to CRNAs). More than 3,000 responses were received by March 14, 2018, when the call for volunteers closed. Volunteers were selected through stratified random sampling on the main demographic characteristics of gender, age group, and years of

FIGURE 1

Consort Flow Diagram



Exclusion reasons:

Closed book, remote: delay completing assessment, *n* = 2; technical issues—Unable to complete second section, *n* = 1.

e-Books, in person: e-book technical issues, *n* = 9; technical issues—Unable to complete second section *n* = 7; delay completing assessment, *n* = 1.

e-Books, remote: e-book technical issues, *n* = 9; technical issues—Unable to complete second section, *n* = 3; purposely did poorly on assessment, *n* = 1; backed out mid-assessment, *n* = 1.

Hard copy books, in person: technical issues—Unable to complete second section, *n* = 3; extended time, *n* = 1.

Hard copy books, remote: delay completing assessment, *n* = 2; technical issues—Unable to complete second section, *n* = 1.

Note. e-Books included pdf copies of *Miller's Anesthesia,* 8th edition, and *Nurse Anesthesia,* 6th edition. Thirty days before the assessment, participants in the e-books groups were provided access to a maximum of 35% of the textbook chapters so they could familiarize themselves with navigating the e-books. Participants in the hard copy resources groups could bring any two of the following (any edition): *Miller's Anesthesia; Nurse Anesthesia; Clinical Anesthesia; Stoelting's Anesthesia and Co-Existing Disease; A Practical Approach to Anesthesia Equipment; Clinical Anesthesia Procedures of the Massachusetts General Hospital; Handbook of Clinical Anesthesia; Handbook of Nurse Anesthesia; Clinical Anesthesiology*.

certification. These criteria were used for sampling to ensure selection of a representative sample of the CRNA population in terms of demographics, and years of certification was used as a surrogate marker for years of experience. These characteristics were chosen because in our experience, they are more associated with certification examination performance than other characteristics. Six representative sample groups of the general CRNA population were formed and randomly assigned with a 1:1 allocation ratio to six parallel testing conditions (Figure 1). Participants were informed of their group assignment after randomization and then scheduled the CPCA accordingly based on their testing window. Participants in the remote groups took the CPCA with live online proctoring, and those in the in-person groups took the CPCA in traditional NBCRNA-approved testing centers. Of the 3,000 volunteers, 56 were selected for the Alpha phase and 1,500 for the Beta research study. The Alpha phase was a pilot study to determine whether changes to the protocol were necessary as a result of logistical or technical issues prior to launch of the Beta research study. Participants were not compensated for participation in the study.

Baseline Survey and Post-CPCA Survey

After enrollment and electronic consent, but prior to participation in the CPCA, participants completed a baseline survey via Survey Monkey. Survey questions were developed by the NBCRNA Evaluation and Research Advisory Committee and were based on expert opinion, American Association of Nurse Anesthetists' professional practice surveys, and review of the literature (Durning et al., 2016; Lipner et al., 2017). Age, date of birth, and core module completion rates were obtained from NBCRNA databases. Completion of a core module was defined as taking at least one module between the time the examinee started his or her CPC cycle to the date they took the CPCA. The baseline survey data collection included demographics, CPC program data (Class A and B credits), level of familiarity with online textbooks (0–10 visual analog scale), and anticipated comfort level with use of resources during the CPCA (Likert scales). The Westside Test Anxiety Scale was used to measure baseline test anxiety in all study participants (Driscoll, 2006). The scale has high face validity: in college students, the correlation (*r*) between anxiety reduction on the Westside scale and improvement in test scores was .49 (Driscoll, 2006). After participants took the CPCA and prior to receiving their results, they were emailed a link via Survey Monkey to complete a post-CPCA survey to collect information on their preparation time, perception of level of CPCA difficulty, fidelity of assessment content, their own performance, and testing experience.

Open Book Resources

Participants in the e-books groups were provided access to PDF copies of *Miller's Anesthesia,* 8th edition, and *Nurse Anesthesia,* 6th edition. Thirty days before the assessment, participants in the e-books groups were provided access to a maximum of 35% of the textbook chapters so they could familiarize themselves with navigating the e-books. Participants in the hard copy books groups could bring to a testing center, or use during remote proctoring, any two of the following: *Miller's Anesthesia; Nurse Anesthesia; Clinical Anesthesia; Stoelting's Anesthesia and Co-Existing Disease; A Practical Approach to Anesthesia Equipment; Clinical Anesthesia Procedures of the Massachusetts General Hospital; Handbook of Clinical Anesthesia; Handbook of Nurse Anesthesia; Clinical Anesthesiology*. Participants in both of these groups (ie, the resources groups) were instructed that they may refer to these books at any time. These resources were chosen based on previous research published by the NBCRNA, which indicated they were the most commonly used and cost-effective textbooks/handbooks (Ferris & Muckle, 2018).

During the CPCA, access to websites and smart phone applications commonly used by CRNAs in clinical practice (Riddle et al., 2016) were not permitted due to (a) cheating potential; (b) risk of theft of intellectual property; (c) lack of access to websites outside the secure testing platform; (d) lack of a single, all-inclusive anesthesia resource website (Ferris & Muckle, 2018); and (e) the examination questions were referenced to textbook resources (Ferris & Muckle, 2018; NBCRNA, n.d.).

Remote Proctoring

Participants in the remote proctoring groups used live online proctoring when taking the CPCA. During the check-in process, participants were required to present valid photo identification and had to use their webcam to scan their testing location. If participants took a break after the first 75 questions, they were required to repeat the same validation sequence. The remote proctoring ratio was one proctor for every three participants. The time to check in or to take a break was not included in the overall assessment time. All testing sessions were recorded, and participants' computers were locked down to prevent activities such as copying test material or searching the Internet.

CPCA Delivery

The CPCA is a 150-item, computer-based examination (multiple-choice and multiple-select items) designed to assess knowledge in the four core domains of nurse anesthesia practice. Participants had up to 4 hours to complete the assessment. The NBCRNA offered CPCA practice examinations online to all CRNAs; however, participants were not specifically directed to use these preparation materials. Participants completed the examination during one of three testing windows based on their group assignment (Figure 1). Based on feedback from participants in the Alpha phase pilot study, the CPCA administered during the Beta study was divided into two 75-item sections, with the option for a 10-minute break in between to allow participants to use the restroom. Participants were required to log back in and confirm their identity after the break.

CPCA Performance Standard

In November 2018, the NBCRNA convened a Standard Setting Panel to establish the performance standard for the CPCA. The Modified Angoff and Hofstee Methods were used to establish the performance standard. The NBCRNA Board of Directors reviewed the recommendations and approved a standard. Participants who meet the performance standard will not be required to complete the CPCA during their initial CPC period. Those participants who did not meet the performance standard were provided with a score report describing their areas of weakness in the four core domains and will be required to retake the CPCA during their initial CPC period.

Sample Size Calculation

This proposal used a 2 × 3 experimental design where there are two conditions for the testing environment (i.e., in-person at a test center and remotely proctored at the examinee's own location) and three conditions for access to resources (i.e., closed book, open-book provided resources, and open-book candidate-provided resources), resulting in six experimental groups.

A number of factors were considered in determining the sample size for this study.

First, a power analysis was conducted to determine the sample size that would be needed to achieve a high probability of detecting group differences in test scores. With an assumption of power being 0.8, a two-tailed statistical test (assuming no directional hypothesis), a significance level of 0.05, and a small to moderate Cohen's *d* effect size (0.35), it was estimated 130 participants would be required in each group.

Second, consideration was given to sample size needed for reliable computations for the psychometric models being used for the analysis of the testing data. For the purposes of this study, we were interested in representing the population of examinees, but more important was to have sample sizes large enough for estimating test-level psychometric characteristics for the examination. To obtain these types of estimates in classic test theory, minimum sample sizes of 100 are recommended (Jones, Smith, & Talley, 2006). However, because we anticipated using item response theory in addition to classic test theory, larger sample sizes would likely yield more stable empirical estimates for the selected item response theory model (Jones et al., 2006).

Third, the desire to obtain large samples is tempered by logistical and operational costs (e.g., testing fees, monetary incentives) involved with managing a study of this size.

Taking the considerations of power analysis, psychometric models, and logistical/operational constraints, a sample size of 200 per condition (total 1,200) was adopted for this study. A sample size of 200 to 250 is ideal for item response theory analysis; therefore, we invited 1,500 to participate in the Beta study, with 250 per group to account for attrition.

Statistical Analysis

Descriptive and inferential statistics were used to analyze the results. Chi square or Fisher's exact test, where appropriate, were used to test for

TABLE 1

Demographics of CPCA Participants in the Beta Research Study

Variable	Overall (*N* = 1,217)	Closed Book, in Person (*n* = 204)	Closed Book, Remote (*n* = 194)	e-Books, in Person (*n* = 196)	e-Books, Remote (*n* = 191)	Hard Copy Books, in Person (*n* = 217)	Hard Copy Books, Remote (*n* = 215)
Mean Age, y	46.2 ± 9.5	46.8 ± 9.3	45.9 ± 9.7	46.3 ± 9.4	45.4 ± 9.8	46.0 ± 9.8	46.9 ± 9.5
Gender							
Male	458 (37.6)	73 (35.8)	70 (36.1)	84 (42.9)	75 (39.3)	78 (35.9)	78 (36.3)
Female	759 (62.4)	131 (64.2)	124 (63.9)	112 (57.1)	116 (60.7)	139(64.1)	137 (63.7)
AANA Region[a]							
1	118 (9.7)	18 (18.9)	18 (9.3)	29 (14.8)	17 (8.9)	18 (8.3)	18 (8.4)
2	244 (20.1)	34 (16.7)	48 (24.7)	44 (22.4)	32 (16.8)	44 (20.3)	42 (19.6)
3	138 (11.4)	24 (11.8)	25 (12.9)	18 (9.2)	24 (12.6)	25 (11.5)	22 (10.3)
4	129 (10.6)	21 (10.3	17 (8.8)	19 (9.7)	25 (13.1)	26 (12.0)	21 (9.8)
5	139 (11.4)	29 (14.3)	13 (6.7)	23 (11.7)	21 (11.0)	30 (13.8)	23 (10.7)
6	190 (15.6)	33 (16.3)	27 (13.9)	26 (13.3)	33 (17.3)	38 (17.5)	33 (15.4)
7	257 (21.2)	44 (21.7)	46 (23.7)	37 (18.9)	39 (20.4)	36 (16.6)	55 (25.7)
Years in Practice							
≤ 5 years	255 (21)	35 (16.7)	49 (25.3)	39 (19.9)	44 (23.0)	48 (22.2)	41 (19.1)
6–10 years	331 (27.2)	55 (27.0)	43 (22.2)	66 (33.7)	56 (29.6)	56 (25.8)	55 (25.6)
11–20 years	385 (31.6)	74 (36.3)	57 (29.4)	61 (31.1)	53 (27.7)	71 (32.7)	69 (32.1)
21–30 years	165 (13.6)	29 (14.2)	31 (16.0)	20 (10.2)	20 (10.5)	27 (12.4)	38 (17.7)
31–>40 years	81 (6.6)	12 (5.9)	14 (7.2)	10 (5.1)	18 (9.4)	15 (6.9)	12 (5.6)
Highest Degree							
BS or BSN	51 (4.2)	6 (2.9)	13 (6.7)	6 (3.1)	6 (3.1)	6 (2.8)	14 (6.5)
Master's degree	986 (81.0)	175 (85.8)	147 (75.8)	159 (82.1)	155 (81.2)	178 (82.0)	172 (80.0)
Practice doctorate	131 (10.8)	14 (6.9)	23 (11.9)	24 (12.2)	22 (11.5)	24 (11.1)	24 (11.2)
PhD	20 (1.6)	5 (2.5)	4 (2.1)	3 (1.5)	4 (2.1)	2 (0.9)	2 (0.9)
EdD/other	29 (2.4)	4 (2.0)	7 (3.6)	4 (2.0)	4 (2.1)	7 (3.2)	3 (1.4)
Primary Position							
Practice	1142 (93.8)	194 (95.1)	183 (94.3)	176 (89.8)	173 (90.6)	207 (95.4)	209 (97.2)
Education	46 (3.0)	4 (2.0)	3 (1.5)	10 (5.1)	10 (5.2)	5 (2.3)	4 (1.9)
Department	31 (2.5)	6 (2.9)	3 (1.5)	8 (4.1)	7 (3.7)	5 (2.3)	2 (0.9)
Other	8 (0.7)	0	5 (2.6)	2 (1.0)	1 (0.5)	0	0
Patient Population (may select >1)							
Neonates/infant, <12 m	293 (24.1)	43 (21.1)	43 (22.3)	47 (24)	63 (33)	49 (22.6)	48 (22.4)
ASA III neonates	117 (9.6)	24 (11.8)	9 (4.6)	25 (12.8)	18 (9.4)	26 (12)	15 (7.0)
Pediatrics, >12 m to 12 y	720 (59.2)	119 (58.3)	114 (58.8)	117 (59.7)	116 (60.7)	131 (60.4)	123 (57.2)
ASA III pediatrics	216 (17.7)	37 (18.1)	20 (10.3)	46 (23.5)	42 (22)	38 (17.5)	33 (15.3)
Adolescent, 13-17 y	882 (72.5)	149 (73.0)	144 (73.7)	146 (74.5)	141 (73.8)	149 (68.7)	154 (71.6)
ASA III adolescents	426 (35)	72 (35.3)	63 (32.5)	75 (38.3)	75 (39.3)	74 (34.1)	67 (31.2)
Adults, 18-64 yrs	1157 (95.1)	191 (93.6)	186 (95.9)	186 (94.9)	186 (97.4)	205 (94.5)	203 (94.4)
ASA III adults	1130 (92.9)	190 (93.1)	181 (93.3)	182 (92.9)	179 (93.7)	201 (92.6)	197 (91.6)
Geriatrics, ≥65 yrs	1123 (92.3)	184 (90.2)	180 (92.8)	177 (90.3)	182 (95.3)	203 (93.5)	197 (91.6)
ASA III geriatrics	1097 (90.1)	181 (88.7)	176 (90.7)	174 (88.8)	176 (92.1)	192 (88.5)	198 (92.1)
Obstetrics	632 (51.9)	111 (54.4)	101 (52.1)	101 (51.5)	96 (50.3)	112 (51.6)	111 (51.6)
Gastroenterology	1066 (87.6)	179 (87.7)	166 (85.6)	176 (89.9)	168 (88)	187 (86.2)	190 (88.4)
Cardiac	284 (23.3)	38 (18.6)	46 (523.7)	50 (25.5)	49 (25.7)	51 (23.5)	50 (23.3)
Thoracic	518 (42.6)	89 (43.6)	74 (38.1)	87 (44.4)	88 (46.1)	90 (41.5)	90 (41.9)
Neuro-anesthesia	632 (51.9)	101 (49.5)	92 (47.4)	108 (55.1)	108 (56.5)	115 (53)	108 (50.2)
Transplant	169 (13.9)	22 (10.8)	29 (14.9)	24 (12.2)	38 (19.9)	32 (14.7)	24 (11.2)
Chronic Pain	295 (24.2)	37 (18.1)	52 (26.8)	39 (19.9)	58 (30.4)	50 (23.0)	59 (27.4)
Other	64 (5.3)	11 (5.4)	9 (4.6)	7 (3.6)	13 (6.8)	11 (5.1)	13 (6.0)

Demographics of CPCA Participants in the Beta Research Study *(continued)*

Variable	Overall	Closed Book, in Person	Closed Book, Remote	e-Books, in Person	e-Books, Remote	Hard Copy Books, in Person	Hard Copy Books, Remote
	(*N* = 1,217)	(*n* = 204)	(*n* = 194)	(*n* = 196)	(*n* = 191)	(*n* = 217)	(*n* = 215)
Mean Test Anxiety Score[b] (range, 1–5)	2.54 ± 0.8	2.61 ± 0.8	2.55 ± 0.83	2.49 ± 0.75	2.55 ± 0.76	2.57 ± 0.81	2.54 ± 0.8

Note. AANA = American Association of Nurse Anesthetists; ASA = American Society of Anesthesiologists; CPCA = Continued Professional Certification Assessment. Results presented as N (%) or mean ± SD.

[a] AANA Regions represent multiple states; this variable represents where examinees live. See https://www.aana.com/states/state-associations

[b] The Westside Test Anxiety Scale is comprised of 10 Likert-scale questions (1 = not at all true to 5 = extremely or always true). The sum of the 10 questions is divided by 10 to calculate the test anxiety score. Totals may not add up due to missing responses.

associations between group assignment and survey question responses and determine whether a participant met the performance standard. Survey responses with Likert scales (strongly agree to strongly disagree) were analyzed using Chi square or Fisher's exact test because results were easier to interpret and easier for readers to understand. Analysis of variance was used to test for group differences in interval-level demographic and survey variables, overall CPCA raw scores, and test duration (in minutes). If group differences were detected (in the *F* statistic), appropriate post-hoc tests (e.g., Tukey) were conducted to determine which groups demonstrated significant differences. To account for differences introduced by use of resources and administration conditions, the linear equating method was used to link the scores from the remote administrations and with resource conditions to the closed book testing center administration conditions to adjust the performance standard rate across groups. During analysis, some survey variable responses were collapsed to reduce the number of categories.

Only results from participants who completed the Beta study were analyzed. A per protocol analysis was used; participants who experienced technical difficulties, withdrew after starting the CPCA, or were unable to complete the CPCA according to the testing parameters were excluded from analysis. Exemplar quotes from participants' comments were used to highlight some findings. A $p < .05$ was considered significant. Statistical Package for the Social Sciences (SPSS; IBM) was used to analyze the results. Study personnel conducting the statistical analyses were blinded to participant group assignment.

Results

Results in this report are for participants in the Beta study. The 1,500 participants were randomized to one of six assessment conditions (Figure 1). Of these, 1,258 participated in the CPCA. There were 41 participants who were excluded from the analysis, with a higher percentage excluded from the e-books groups when compared to the other groups (75.6% vs. 24.4%, $p = .0005$; Figure 1). We examined results among participants who completed the CPCA according to their group assignment and testing parameters. A higher frequency of participants in the e-books groups experienced technical difficulties with the e-books or were unable to complete the CPCA due to technical difficulties. There were 1,217 participants who took all 150 items of the CPCA and were included in the final analysis.

The groups had similar baseline demographics (Table 1). The average age of participants was 46 years with a majority being female (62.4%); 48.2% had 10 years or less, 31.6% had 11–20 years, 13.6% had 21–30 years, and 6.6% had more than 30 years in practice. Most participants had a master's degree (81%), and 12.4% had a doctoral degree. Most participant's primary position was in clinical practice (91.8%), and most took care of adults (95.1%), ASA III (American Society of Anesthesiologists Physical Status level of severe systemic disease) adults (92.9%), and adolescents (72.5%). There were 24.1% who reported taking care of neonates/infants aged less than 12 months, and only 9.6% reporting caring for ASA III neonates/infants. A majority reported caring for obstetric patients (51.9%). Approximately a quarter of the sample reported taking care of chronic pain patients (24.2%); however, we are unable to determine whether the participants performed chronic pain procedures or provided anesthesia for these procedures. Average test anxiety scores were 2.5, indicating participants had normal to high-normal test anxiety.

For the entire sample, the level of familiarity (0 = lowest, 10 = highest) with online textbooks was 4.95 ± 3.28, indicating participants had a moderate amount of experience with this resource type; most (66.8%) strongly agreed or agreed they would feel comfortable using online anesthesia textbooks (e-books) to find information related to the test questions during the CPCA. Most (85.5%) strongly agreed or agreed that they would feel comfortable using personal resources (i.e., anesthesia textbooks/handbooks) to find information related to the test questions during the CPCA. A majority of participants (17.1%) studied 0–4 hours per week, 17.1% studied 5–8 hours per week, and 12.1% studied more than 8 hours per week prior to taking the assessment. The post-CPCA responses indicated that participants in the closed book groups were more likely to take the NBCRNA CPCA practice examinations when compared to those in the resources groups (57%–63.5% vs. 48.2%–50.7%, $p = .01$).

CPCA Results

Average raw scores by group ranged from 113 to 118 (Table 2). Scores ranged from a low of 62 to a high of 143 (maximum score was 150). Participants in the closed book conditions scored significantly lower

TABLE 2

Participant Outcomes on the CPCA in the Beta Research Study

Variable	Overall (*N* = 1,217)	Closed Book, in Person (*n* = 204)	Closed Book, Remote (*n* = 194)	e-Books, in Person (*n* = 196)	e-Books, Remote (*n* = 191)	Hard Copy Books, in Person (*n* = 217)	Hard Copy Books, Remote (*n* = 215)	*p*
Scores								
Mean ± SD	116 ± 12.6	113 ± 11.5	113 ± 13	118 ± 13	117 ± 13	118 ± 12	117 ± 12	0.0005
Median (IQR)	118 (16)	112 (15)	115 (19)	121 (17)	119 (16)	120 (14)	119 (16)	
Range	62–143	82–138	76–139	71–140	64–143	65–141	62–140	
Length (min)								
Mean ± SD	150.5 ± 61.1	95 ± 37.9	86 ± 32	181 ± 46	174 ± 51	183 ± 48	179 ± 50	0.0005
Median (IQR)	151.9 (111)	86 (43)	81 (43)	187 (75)	177 (83)	192 (68)	191 (83)	
Range	31–240	32–240	31–192	55–240	51–240	48–240	56–240	
Met Performance Standard								
Yes	95%	95.10%	91.80%	94.90%	94.80%	96.30%	96.70%	0.26

Note. CPCA = Continued Professional Certification Assessment; IQR, interquartile range. Analysis of variance with Tukey's post-hoc test was used to analyze mean differences between groups for test scores and test length with a $p < .001$ considered significant. Chi square was used to analyze association between group assignment and meeting performance standard with $p < .05$ considered significant.

on the CPCA when compared to the resources groups ($p = .0005$). Examination scores in the resources groups were similar ($p > .05$). The time to complete the examination ranged from 31 to 240 minutes (maximum time allowed was 240 minutes). Participants in the resources groups took significantly more time to complete the examination ($p = .0005$). Group assignment explained 3.5% (small effect size; partial eta squared = .035) of the variance in assessment scores and 45% of the variance in examination length (large effect size; partial eta squared = .47). Approximately 95% of participants met the CPCA performance standard. No significant differences were found in the frequency of participants meeting the performance standard across the six assessment conditions ($p = .26$; Table 2).

CPCA Duration

Overall, 90.3% of participants strongly agreed or agreed that the time allotted for the CPCA was adequate. However, participants in the resources groups (e-books and hard copy books) were less likely to strongly agree or agree than the closed book groups that the time allotted was adequate ($p = .0005$; Table 3).

CPCA Difficulty

Participants who took the CPCA rated the examination difficulty as a mean of 7 ± 1.47 on a scale from 0 to 10, which suggests they perceived the examination to be moderately to highly difficult (Table 3). Participants in the closed book, in-person group rated the CPCA as being significantly less difficult (6.7 ± 1.5) when compared to the remote hard copy books group (7.2 ± 1.4, mean difference = -0.5; $p = .01$). All other groups were similar.

CPCA Performance

A significant difference was noted in self-rated CPCA performance between the groups ($p = .003$; Table 3); however, Tukey post hoc tests revealed no group-by-group differences. On average, participants in both closed book groups rated their performance slightly lower than those in the e-books or hard copy books groups.

CPCA Fidelity

No association was found between group assignment and participant's response to the question, "The CPCA accurately reflected CORE KNOWLEDGE ALL CRNAs SHOULD KNOW" ($p = .21$). Overall, 51.9% strongly agreed or agreed, 20.3% were neutral, and 27.8% disagreed or strongly disagreed with this statement. No association was found between group assignment and response to the question, "The CPCA accurately reflected CORE KNOWLEDGE I NEED TO KNOW IN MY CURRENT PRACTICE" ($p = .66$). Overall, 49.4% strongly agreed or agreed, 20.6% were neutral, and 30% disagreed or strongly disagreed with this statement (Table 3).

Use of Resources

Participants in the resource groups (e-books or hard copy books) were asked to rate their frequency of resource use on a scale of 0 (did not refer to a resource for any question) to 10 (referred to a resource for every question). The overall mean resource use was 5.6 ± 2.5, which suggests a moderate use of resources during the assessment. Participants used resources most commonly to confirm an answer (84.2%), followed by looking up an answer (55.3%) or both (47.6%). There was no association between group assignment and reason for resource use ($p > .05$). Participants in the e-books groups reported using resources less often and reported them as being less helpful when compared to hard copy books conditions ($p < .05$). Approximately 40% of comments by participants in the e-books groups reported problems with the resources. There were interactive effects found between accessed book types and proctoring conditions.

TABLE 3

Self-Rated Post-CPCA Survey Results in the CPCA Beta Research Study

Variable	Overall (*N* = 1,147)	Closed Book, in Person (*n* = 200)	Closed Book, Remote (*n* = 189)	e-Books, in person (*n* = 191)	eBooks, Remote (*n* = 182)	Hard Copy Books, in Person (*n* = 184)	Hard Copy Books, Remote (*n* = 201)	*p*
The time allotted for the CPC exam was adequate.								
Strongly agree	521 (45.4)	125 (63)	118 (62.4)	72 (37.7)	70 (38.5)	66 (35.9)	70 (34.8)	<.0005
Agree	515 (44.9)	68 (34.0)	66 (34.9)	95 (49.7)	90 (49.5)	94 (51.1)	102 (50.7)	
Neutral	53 (4.6)	4 (2.0)	4 (2.1)	7 (3,7)	13 (7.1)	14 (7.6)	11 (5.5)	
Disagree	51 (4.4)	2 (1.0)	1 (0.5)	15 (7.9)	8 (4.4)	9 (4.9)	16 (8.0)	
Strongly disagree	7 (0.6)	1 (0.5)	0	2 (1.0)	1 (0.5)	1 (0.5)	2 (1.0)	
CPC exam difficulty (0 = lowest to 10 = highest)								
	7 ± 1.5	6.7 ± 1.5	7.1 ± 1.5	6.9 ± 1.5	7 ± 1.5	7.1 ± 1.4	7.2 ± 1.4	0.01
CPC exam performance (0 = extremely poorly to 10 = extremely well)								
	6 ± 2	5.6 ± 1.9	5.7 ± 1.8	6.2 ± 1.9	6.1 ± 2.1	6.3 ± 2	6.2 ± 2.1	0.003
The CPC exam accurately reflected CORE KNOWLEDGE ALL CRNAs SHOULD KNOW.								
Strongly Agree	166 (14.5)	36 (18)	29 (15.3)	33 (17.3)	28 (10.9)	20 (10.9)	20 (10)	0.028
Agree	431 (37.6)	77 (38.5)	70 (37)	72 (37.7)	67 (38.8)	70 (38)	75 (37.3)	
Neutral	229 (20)	44 (22)	39 (20.6)	35 (18.3)	36 (19.8)	42 (22.8)	33 (16.4)	
Disagree	246 (21.4)	37 (18.5)	41 (21.7)	42 (22)	39 (21.4)	42 (22.8)	45 (22.4)	
Strongly disagree	75 (6.5)	6 (3)	10 (5.3)	9 (4.7)	12 (6.6)	10 (5.4)	28 (13.9)	
The CPC exam accurately reflected CORE KNOWLEDGE I NEED TO KNOW IN MY CURRENT PRACTICE.								
Strongly agree	161 (14)	31 (15.5)	26 (13.8)	33 (17.3)	29 (15.9)	20 (10.9)	22 (10.9)	0.069
Agree	407 (35.5)	72 (36)	69 (36.5)	63 (33)	66 (36.3)	66 (35.9)	71 (35.3)	
Neutral	233 (20.3)	39 (19.5)	45 (23.8)	38 (19.9)	35 (19.2)	41 (22.3)	35 (17.4)	
Disagree	262 (22.8)	51 (25.5)	29 (20.6)	47 (24.6)	37 (20.3)	44 (23.9)	44 (21.9)	
Strongly disagree	84 (7.3)	7 (3.5)	10 (5.3)	10 (5.2)	15 (8.2)	13 (7.1)	29 (14.4)	

Note. CPCA = Continued Professional Certification Assessment; CPC = Continued Professional Certification; CRNA = certified registered nurse anesthetist. Results are presented as *N* (%) or mean ± SD. Total may not add up due to missing responses. Likert scale results were analyzed using chi square or Fisher's exact test. A $p < .05$ was significant. Mean differences in CPCA difficulty and CPCA performance were analyzed with an analysis of variance with Tukey post-hoc test used to analyze mean differences between groups for test scores and test length with $p < .001$ considered significant.

Several participants commented that they had difficulty using the resources during the CPCA. Exemplar quotes included, "The e-books had really small print and I had to enlarge every reference, and this took time." Another participant commented, "The navigation of the e-books was so difficult I had to stop using them…. Eventually, I gave up using the resources." Another in the in-person hard copy books group commented, "Testing sites are not geared for textbooks; very difficult to find space for books.... Time management was a big issue." A participant in the remote hard copy books group wrote, "Finding items in books was very time consuming and not always attainable. Online books with search feature would be nice!"

Remote Proctoring Experience

Participants were asked to rate their experience with remote proctoring, with 0 representing many problems and 10 representing no problems. The overall rating was 5.64 ± 3.37, indicating participants experienced a moderate number of problems with remote proctoring. No significant differences were found in the proctoring experience across the remote groups ($p = .051$). In response to the question, "Did you have any difficulty using your computer's camera or audio equipment as you took the test?" 16.4% of participants in the remote proctoring conditions answered *yes* and 83.3% answered *no.* The rate of computer or camera audio problems was similar across the remote proctored groups ($p = .99$).

In response to the statement, "I was able to get a response from the remote proctoring staff quickly when I needed assistance," most participants (70.6%) in the remote proctoring groups answered *yes*, 16.9% answered *no*, and 12.2% selected *N/A* (not applicable). Responses were similar across the groups ($p = .19$). Assessment times were 23 minutes longer among those who had delays in getting proctor assistance when compared to those who did not experience a delay ($p = .001$). Despite these findings, 64% of participants strongly agreed or agreed that being able to take the CPCA with remote proctoring would reduce their anxiety; however, participants in remote groups were more likely to strongly agree or agree (27%–47%) with this statement when compared to those in the in-person groups (25%–29%, $p < .0005$).

Comments by several participants in the remote proctoring condition reported they experienced technical issues and difficulty understanding proctors due to their strong accent. The language barriers

were most likely due to the fact that the remote proctors used in this study were located in Southeast Asia and India. One participant wrote, "I struggled to get to the test, and I couldn't understand my proctor at times. Once I was on it went well. I think the older anesthetists may struggle who did not grow up with technology." Another wrote,

> *It took almost 2 hours to start the CPCA from the time I initially logged in to the system. There were issues with passwords linking up etc. mainly on the proctor end. There also was a language barrier on the proctor's end.... During the break in between the two assessments the system logged out and it took almost an additional 30 minutes to start the second assessment. These situations need to be remedied! My level of frustration and anxiety was way too high to start an assessment, let alone take it and remain focused....*

Discussion

At the core of the evaluation of various testing conditions and groups is the desire to determine the relative equivalence of both the testing conditions and the results of the examination by group and testing condition. Additionally, and perhaps more importantly, a practical and pragmatic approach to understanding these results and their implications to overall knowledge assessment is paramount. For example, although there is a statistically significant difference in raw test scores on the CPCA based on testing conditions, when examining the difference in performance, the mean raw score range of performance is 5 points (113 to 118), and the performance standard rates across the six testing conditions were similar after equating of scores. The difference in raw scores across conditions may be explained by some test items being less difficult with access to reference books. This difference in score, while statistically significant, is arguably not practically important in terms of the impact that knowledge level difference may have on overall clinical practice.

In addition to evaluating performance, other key variables that should be considered include self-assessment of performance, testing circumstances (e.g., time to complete the CPCA), and other qualitative variables such as the perceived difficulty of the CPCA based on testing condition, perceived value of the CPCA, and perceived application of the knowledge assessed to actual clinical practice.

It is interesting to note that the perceived difficulty of the CPCA was dependent on the method of delivery with those taking the open book CPCA reporting a higher level of difficulty. This is perhaps partially explained by the greater amount of time each examinee in the open books group took to complete the CPCA. Another finding was that the use of hard copy books or e-books during the CPCA produced some dissatisfaction with the amount of time that was consumed looking for answers and in the e-books group with the difficulty in utilizing the resource. One concern with open book testing is what is being tested (i.e., testing knowledge or testing one's ability to look up an answer) (Durning et al., 2016; Johanns, Dinkens, & Moore, 2017). When considering the difficulties in terms of time to take the CPCA and utilize the resources, it is difficult to justify this method of assessment delivery to the larger population of more than 50,000 CRNAs.

Additionally, some comments from participants in the remote proctored groups indicated they had difficulty understanding remote proctors' instructions. Despite these issues, there was a strong preference for remote proctoring on future CPCAs, and participants' responses suggest use of remote proctoring would decrease their test anxiety. However, some participants in remote proctoring groups expressed a preference for taking the CPCA in a testing center. This could be mitigated if a form of remote proctoring is utilized that does not involve actual interaction with the remote proctor. One option is to record and review the remotely proctored examination—an option that records the test check-in and administration for later review by a proctor. Direct proctor communication is not involved unless requested for technical assistance, and in those instances, it is managed via chat, not voice.

This study is one of the few randomized controlled trials examining the effect of administration of a recertification assessment remotely or in a testing center with closed or open book testing (Lipner et al., 2017). Post-CPCA results indicated that participants in the closed book groups were more likely to report using the NBCRNA practice examinations than participants in the resource groups. This finding suggests participants who knew they would not have access to resources prepared differently for the assessment. A systematic review on open- vs. closed-book testing found performance was better in the closed-book condition, which the authors attributed to better examination preparation (Durning et al., 2016). Future studies should examine which CPCA preparation methods (i.e., core modules, NBCRNA practice assessments, hours studied) are predictive of performance. Also, these results only apply to the use of anesthesia textbooks/handbooks as resources. Previous research by the NBCRNA (Ferris & Muckle, 2018) found most CRNAs use web-based resources in daily practice. A future study with web-based resources would be more reflective of practice.

Implications for Nursing Regulations

The Citizen Advocacy Group (Swankin et al., 2006) recommended that state licensing boards require healthcare providers to participate in continuing professional development programs and that these programs include an assessment of knowledge required for clinical practice that is legally defensible, psychometrically sound, and evidence based. The CPC program and the CPCA meet these requirements, and our results demonstrate that use of resources has minimal impact on a examinees ability to meet the performance standard on the CPCA, confirming that they possess the mastery of the knowledge determined necessary for practice in nurse anesthesia. Moving forward, the NBCRNA should ensure those examinees who do not meet the performance standard be required to complete focused continuing education in domains in which they show weakness. State regulators should consider these results and encourage other advanced practice registered nursing certifying organizations to develop continuing professional development programs that include an objective, rigorously developed

assessment of knowledge needed for continued clinical practice rather than one that tests entry-level knowledge or relies soley on completion of continuing education hours.

Limitations

There were some differences in participants' practice environments and patient populations (Table 1); however, these differences were not substantial enough to break the randomization design and group comparability. Second, there was a statistically significant higher rate of exclusions of participants in the e-books groups because of technical difficulties. Participants in the in-person e-books group who were excluded were approximately 8 years older when compared to those who were not excluded (53.97 ± 10.4 vs. 46.29 ± 9.37, $p = .002$); however, no differences were seen in age in the e-books remote group between excluded and nonexcluded participants (46.53 ± 9.52 vs. 45.34 ± 9.34, $p = .65$). This suggests that older CRNAs may have more difficulty with accessing and using e-book references. Future investigations should explore this finding.

Conclusion

Overall, the results of this study indicated that use of resources had no effect on whether participants met the performance standard. On average, use of resources increased test scores by 4 to 5 points; however, the use of resources doubled the amount of time needed to complete the CPCA. Additionally, several participants reported difficulties using the e-books. Despite technical difficulties, participants' responses suggest a strong preference for remote proctoring. Therefore, it is recommended that future CPCAs be delivered as a closed book assessment with a choice between in-person testing and record and review remote proctoring.

References

American Board of Anesthesiology. (n.d.). Part 3: MOCA Minute®. Retrieved from http://www.theaba.org/MOCA/MOCA-Minute

American Board of Internal Medicine. (n.d.). Knowledge Check-In. Retrieved from https://www.abim.org/maintenance-of-certification/assessment-information/assessment-options/knowledge-check-in.aspx

Brennan, T. A., Horwitz, R. I., Duffy, F. D., Cassel, C. K., Goode, L. D., & Lipner, R. S. (2004). The role of physician specialty board certification status in the quality movement. *JAMA, 292*(9), 1038–1043. https://doi.org/10.1001/jama.292.9.1038

Driscoll, R. (2006). Westside Test Anxiety Scale validation. American Test Anxiety Association. Retrieved from https://files.eric.ed.gov/fulltext/ED495968.pdf

Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino, A. R., Jr., Boulet, J. R., & Eva, K. (2016). Comparing open-book and closed-book examinationinations: A systematic review. *Academic Medicine, 91*(4), 583–599. https://doi.org/10.1097/ACM.0000000000000977

Ferris, M., & Muckle, T. J. (2018). Reference resource usage among nurse anesthetists: A survey. *AANA Journal, 86*(4), 299–309.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10027

Johanns, B., Dinkens, A., & Moore, J. (2017). A systematic review comparing open-book and closed-book examinations: Evaluating effects on development of critical thinking skills. *Nurse Education in Practice, 27*, 89–94. https://doi.org/10.1016/j.nepr.2017.08.018

Jones, P., Smith, R., & Talley, D. (2006). Developing test forms for small-scale achievement testing systems. In S. M. Downing & T. M. Haladyna (Eds.). *Handbook of test development* (pp. 487–526). Mahwah, NJ: Lawrence Erlbaum.

Lipner, R. S., Brossman, B. G., Samonte, K. M., & Durning, S. J. (2017). Effect of access to an electronic medical resource on performance characteristics of a certification examination: A randomized controlled trial. *Annals of Internal Medicine, 167*(5), 302–310. https://doi.org/10.7326/M16-2843

National Board of Certification & Recertification for Nurse Anesthetists. (n.d.). CPC assessment bibliography. Retrieved from https://www.nbcrna.com/examinations/cpce/CPC-Assessment-Bibliography

Riddle, D., Baker, K., & Sapp, A. (2016). Evaluation of testing as a method to assess continued competency in nurse anesthesia practice: A systematic review. *AANA Journal, 84*(4), 239–245.

Swankin, D., LeBuhn, R. A., & Morrison, R. (2006). *Implementing continuing competency requirements for health care practitioners.* Washington, DC: American Association of Retired Persons. Retrieved from https://www.cacenter.org/files/ImplementingContinuingCompetencyRequirements.pdf

Wooden, S., Krogh, M. A., Waters, E., & Plaus, K. (2017). Developing the Continued Professional Certification program for nurse anesthetists. *Journal of Nursing Regulation, 8*(1), 31–37. https://doi.org/10.1016/S2155-8256(17)30072-8

Dennis Spence, PhD, CRNA, is a Director, National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA), and board liaison to the Evaluation and Research Advisory Committee, NBCRNA, Chicago, Illinois. **Robyn Ward, PhD, CRNA,** is a member of the Evaluation and Research Advisory Committee, NBCRNA, and an Associate Professor and Director, School of Nurse Anesthesia, Texas Christian University, Fort Worth, Texas. **Steve Wooden, DNP, CRNA, NSPM-C,** is the Chairman, Evaluation and Research Advisory Committee, NBCRNA, and a Staff Nurse Anesthetist, Boone County Health Center, Albion, Nebraska. **Mary Browne, MA,** is the Senior Director of Testing, NBCRNA. **Hao Song, PhD,** is the Director of Psychometrics, NBCRNA. **Robert Hawkins, PhD, CRNA, DNP,** is a Consultant, Evaluation and Research Advisory Committee, NBCRNA. **Mary Wojnakowski, PhD, CRNA,** is President, NBCRNA, and a Professor, Department of Surgery and Anesthesia, Midwestern University, Glendale, Arizona.

This study could not have been completed without the assistance of Tim Muckle, Matthew Ferris, Cindy Davenport, and Elizabeth Johnson. We would also like to thank the CRNAs who volunteered to take part in this study.