The Journal of Education in Perioperative Medicine

Original Research

Development of a Multiple-Choice Test for Novice Anesthesia Residents to Evaluate Knowledge Related to Management of General Anesthesia for Urgent Cesarean Delivery

Allison J. Lee, MD
Stephanie R. Goodman, MD
Shawn E. Banks, MD
Meiko Lin, EdD
Ruth Landau, MD

Introduction

Cesarean delivery (CD) is the most commonly performed surgical procedure in American hospitals, representing over 30% of all births,[1] most of which are performed under spinal anesthesia (80% of elective cesarean deliveries in stratum III hospitals, which are centers that provide subspecialty care).[2] Even in tertiary care centers with high volumes of deliveries, the rates of GA have been reported to be as low as 0.5%.[3]

General anesthesia (GA) for CD is associated with persistently higher rates of anesthesia-related adverse events compared with regional anesthesia.[4,5] In New York State, despite a significant decrease in the proportion of CD performed under GA—from 7.5% in 2003 to 6% in 2012—the overall rate of anesthesia-related adverse events among women receiving GA for CD did not decrease. The declining utilization of GA for CD has concerned educators regarding insufficient training of anesthesiology trainees to manage this high risk clinical scenario.[3,6,7]

For an urgent CD, the learner needs to develop the skill of obtaining a quick, focused medical/surgical/obstetric history and airway exam. In devising the anesthesia plan, both fetal and maternal concerns must be taken into account. Finally, the learner must account for the physiological differences in the parturient, such as the effect of aortocaval compression, the increased risk of difficult intubation and pulmonary aspiration of gastric contents, the risk of uterine atony in response to inhaled volatile agents, the depressant effects of medication on the fetus, and the risk of maternal awareness.

In our institution, residents begin obstetric anesthesiology rotations as early as the third month of their first year of residency, and we must prepare them for the possibility of being involved in the management of a patient undergoing GA for emergent CD from day one of the rotation.

We identified the need for a valid and reliable assessment tool to evaluate trainee competency related to this critical scenario. We describe the multistage design process to create and validate a criterion-referenced knowledge test as an assessment tool for anesthesiology trainees with no clinical experience in obstetric anesthesiology.

Methods

The Columbia University (CU) and University of Miami (UM) Institutional Review Boards approved this study.

Instrument Development

Instrument development comprised four phases: (1) purpose and domain specification, (2) development of survey specifications, (3) content validation, and (4) empirical validation, based on Chatterji's process model (Figure 1).[8]

1. Purpose and domain specification

The target population was novice CA1 (first year) anesthesiology trainees, never exposed to obstetric anesthesia cases. The purpose of the instrument was to assess the degree of trainee proficiency for the domain, "GA for urgent CD." The design was a criterion-referenced test (CRT), the score of which is a measurement of performance against set criteria indicating mastery of the domain.[9]

2. Development of survey specifications

The essential areas of knowledge were based on a validated checklist.[10] A panel comprising three CU faculty content experts agreed upon the subdomains, (1) physiologic changes of pregnancy (PCP), (2) pharmacology (PHA), (3) anesthetic implications of pregnancy (AIP), and (4) crisis resource management principles (CRM). The competencies being tested for each subdomain were listed as keywords (Table 1). The content experts were asked to each submit ≥10 multiple choice questions (each stem with 1 correct answer and 3 distractors), which would be a representative sample covering the knowledge content of the indicators in the four subdomains at the level of a novice CA1 trainee. Equal numbers of questions for each subdomain was not expected. Thirty-six questions were submitted: (1) PCP (n=12), (2) PHA (n=4), (3) AIP (n=14), and (4) CRM (n=6). Three questions were discarded because they were too easy, resulting in 33 questions.

3. Content Validation

A Delphi process was conducted in three rounds for content validation of the questions.

Experts were recruited from the Society for Obstetric Anesthesia and Perinatology Research and Education Committees by email.

continued on next page

continued from previous page

Twenty experts initially agreed to participate. Panel members were asked to anonymously rate the 33 questions based on a 7-point Likert scale where 1 = "I feel this is not important at all" and 7 = "I feel this is extremely important." Feedback and suggestions for improving individual questions was encouraged.

4. Empirical Validation

Empirical validation was conducted after the three rounds of the Delphi process.

Participants for pilot testing

The knowledge test was administered to three different groups:

1. Uninstructed group (UG): The July 2016 CA1 (n = 26) class at UM was selected since they mimicked the CU CA1s with respect background characteristics. This group received no training regarding management of GA for CD.
2. Instructed group (IG): The 2017 CA1 (n = 26) class at CU received a 1-hour didactic lecture (delivered by A.L.) in the third week of July teaching about the domain, management of GA for CD, as part of the orientation month core lecture series. Using the case study of umbilical cord prolapse necessitating emergent CD under GA, the content taught covered the subdomains of physiologic and pharmacodynamic changes in pregnancy, the implications of the latter for anesthetic management in pregnancy, and the crisis management, teamwork and communication skills necessary to safely conduct GA for emergent CD.
3. Expert group (EG): Ten attending anesthesiologists (n = 10), volunteers from the UM CA2 (second-year residents) class (n = 10) and CU CA2 class (n = 7) took the same test (completed July 2016). This expert group was used for sensitivity analysis to further verify if the knowledge test is valid and reliable in assessing the GA for urgent CD knowledge domain.

Frequency polygons of the UG and IG were plotted to verify the consistency of the expert-selected cut-score. Internal consistency reliability was measured with Hoyt Analysis of Variance (ANOVA) method; a coefficient with a cutoff value of ≥ 0.70 was considered desirable. Item analysis with methods from Classical Test Theory was conducted.[11]

Based on the calculated item discrimination index (D), we used the following guidelines[12] to interpret CRT item analysis results:

- If D < 10%, the item should be removed.
- If 10% ≤ D < 20%, the item should be revised.
- If D ≥ 20%, the item is functioning well

Convergent validity was assessed through examining the intercorrelations among the four major subdomains on the survey (Table 3). The convergent validity coefficients were calculated with Spearman rank order correlation. All the aforementioned analyses were performed first with the UG and IG to establish evidence of validity and reliability for the criterion-referenced knowledge test. Then we performed a sensitivity analysis with the UG and EG to cross validate the UG/IG empirical validation results. Experts were consulted about setting a standard or cut-score for the bank of questions, the standard being the minimum competency expected of a CA1 resident after training in the clinical scenario. All analyses were performed using SPSS statistical software (version 20.0; IBM Corporation, Armonk, NY). A P value ≤ 0.05 was considered to be statistically significant.

Results

Content Validation Results

Fifteen experts participated in Round 1 (completed April 2016). The mean and median ranking for individual questions are in Table 2. The criteria for question elimination were established a priori as follows: Questions ranked ≥5 in importance by ≥70% of participants were retained. Questions eliminated were #6, #11, #15, #19, and #24, which received ratings of 5 to 7 in 66.7%, 53.3%, 66.7%, 60%, and 33.3% of participants, respectively. Several questions were revised, and 1 new question was added based on suggestions by experts.

In Round 2 (completed May 2016), participants rated in a manner similar to Round 1: the 28 questions remaining and the new question. Consensus was defined as (1) a change of ≤ 10% in the mean score for each item, and (2) after individuals were grouped into quartiles, a change of ≤ 5% in the average of the individual total scores all items by quartile. Fourteen responses were received. Consensus was reached in all except #17 and #28, the new question (#34), and 4 questions that had been revised based on feedback from Round 1 (#8, #12, #25, and #26).

For Round 3 (completed June 2016) experts rated in a manner similar to Round 1: 7 items for which consensus was not reached or for which significant revisions were made. Fourteen responses were received. All questions were found to have stabilized. The final number of questions within each category were (1) PCP (n = 8), (2) PHA (n = 3), (3) AIP (n = 11), (4) CRM (n = 7).

Empirical Validation Results

The overlapping frequency polygons of the UG and IG suggested an appropriate cut score should be between 20 and 21, where the two distributions first intersected (see Figure 2). A panel of three experts agreed that a high cut score of at least 25 was desirable to demonstrate mastery of the domain. The 29-item survey demonstrated acceptable internal consistency and reliability ($\rho = 0.67$). Table 3 shows the item analysis results for the UG and IG. Regarding the item discrimination index, only 3 items obtained the highest rating (D ≥ 20%) and suggested preservation. The convergent validity coefficients (Table 4) for the UG/IG suggested theoretical meaningfulness of the four subdomains: PCP correlated at 0.29 with PHA, 0.35 with CRM, and 0.25 with AIP. PHA correlated with CRM at 0.23, and AIP at 0.28. The CRM-AIP correlation was 0.29. The subdomains also demonstrated strong, positive correlations with total scores (correlations ranged from 0.54 to 0.74). Consistent with theoretical expectations, the positive intercorrelations suggested construct validity of the four measures in assessing knowledge pertinent to the conduct of GA for urgent CD.

To cross validate the UG/IG results, we performed a sensitivity analysis to compare the UG and EG (Tables 5 and 6); similar reliability and validity in terms of direction and magnitude were found. Six items were well-functioning and 6 more were borderline in terms of item discrimination index. The sensitivity analysis results further supported the evidence of validity and reliability

continued on next page

continued from previous page

in using this CRT to assess knowledge pertaining to GA for urgent CD in novice CA1s.

Discussion

This study describes the stages of development of a valid and reliable instrument to assess CA1 trainees' knowledge related to the conduct of GA for urgent CD. Reasonable internal consistency reliability and good convergent validty were demonstrated, but the instrument is currently lacking in internal structure evidence. Instrument validation is an iterative process (Figure 1). We believe that while the current test does have utility for measuring novice trainee knowledge, revisions are warranted to achieve greater robustness.

The discrimination indices betweeen the instructed and the uninstructed groups showed only 3 highly performing questions; however, the uninstructed and expert group comparison showed that 6 questions performed very well and 6 were borderline (D > 15), yielding 12 acceptable items (highlighted in Table 5). The lack of separation between the uninstructed and instructed groups may have been because the instructed group were still inexperienced novices, despite having received the lecture. The intent was not to test the effectiveness of the lecture. We acknowledge the limitation of applying a written test to verify competency in skills such as CRM.

Next steps will include consultation with experts to agree upon the disposition of the worst-performing items. If the underlying knowledge being tested for those items is considered important (as had been indicated by the Delphi process) those questions may need to be rewritten as opposed to being discarded, followed by additional rounds of pilot testing. To improve the ability to discriminate between experts and novices, we will consider weighting individual item scores by level of difficulty—easier items that are still considered to be critical knowledge would be assigned a lower score value.

With the shift towards competency-based milestones in graduate medical education, the development of reliable assessment tools to track training progress is invaluable.[13,14] We envision use of the finally validated instrument as a benchmark for trainees, which may allow faculty to identify and bridge knowledge gaps related to this infrequently encountered clinical scenario.

References

1. Hamilton BEPD, Martin JA, Osterman MMHS, Curtain SMA. Births: preliminary data for 2014. *Natl Vital Stat Rep.* 2015;64(6):1-19.
2. Bucklin BA, Hawkins JL, Anderson JR, Ullrich FA. Obstetric anesthesia workforce survey: twenty-year update. *Anesthesiology.* 2005;103(3):645-53.
3. Palanisamy A, Mitani AA, Tsen LC. General anesthesia for cesarean delivery at a tertiary care hospital from 2000 to 2005: a retrospective analysis and 10-year update. *Int J Obstet Anesth.* 2011;20(1):10-6.
4. Guglielminotti J, Wong CA, Landau R, Li G. Temporal trends in anesthesia-related adverse events in cesarean deliveries, New York State, 2003–2012. *Anesthesiology.* 2015;123(5):1013-23.
5. Hawkins JL, Chang J, Palmer SK, Gibbs CP, Callaghan WM. Anesthesia-related maternal mortality in the United States: 1979–2002. *Obstet Gynecol.* 2011;117(1):69-74.
6. Hawthorne L, Wilson R, Lyons G, Dresner M. Failed intubation revisited: 17-yr experience in a teaching maternity unit. *Br J Anaesth.* 1996;76(5):680-4.
7. Hawkins JL, Gibbs CP. General anesthesia for cesarean section: are we really prepared? *Int J Obstet Anesth.* 1998;7(3):145-6.
8. Chatterji M. *Designing and Using Tools for Educational Assessment,* Allyn & Bacon/Pearson, Boston, MA, 2003: 105- 110.
9. Chatterji M. *Designing and Using Tools for Educational Assessment,* Allyn & Bacon/Pearson, Boston, MA, 2003: 85.
10. Scavone BM, Sproviero MT, McCarthy RJ, et al. Development of an objective scoring system for measurement of resident performance on the human patient simulator. *Anesthesiology.* 2006;105(2):260-6.
11. Crocker LM, Algina J. *Introduction to Classical and Modern Test Theory*, Wadsworth Group/Thomas Learning, Mason, OH, 2006: 311-335.
12. Crocker LM, Algina J. *Introduction to Classical and Modern Test Theory*, Wadsworth Group/Thomas Learning, Mason, OH, 2006: 329.
13. Boulet JR, Murray D. Review article: assessment in anesthesiology education. *Can J Anaesth.* 2012;59(2):182-92.
14. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract.* 2014;19(2):233-50.

Allison J. Lee is an Assistant Professor of Anesthesiology and Stephanie R. Goodman is a Professor of Anesthesiology, both with Fellowships in Obstetric Anesthesia, Columbia University Medical Center, Columbia University, New York, NY; Shawn E. Banks is an Associate Professor of Clinical Anesthesiology, University of Miami, Department of Anesthesiology, Perioperative Medicine and Pain Management, Miami, FL; Meiko Lin is a Senior Research Assistant, Teachers College, Columbia University, New York, NY; Ruth Landau is a Professor of Anesthesiology at Columbia University Medical Center with a Fellowship in Obstetric Anesthesia, Columbia University Medical Center, Columbia University, New York, NY.

Corresponding Author: Allison J. Lee, Columbia University Department of Anesthesiology, 622 W 268th St, PH-5, New York, NY 10032, Telephone: 305-582-6077, Fax: 212-342-2742. Email address: al3196@cumc.columbia.edu

Funding: Funded by a Gertie Marx grant from the Society of Obstetric Anesthesia and Perinatology

Abstract

Background: Teaching trainees the knowledge and skills to perform general anesthesia (GA) for cesarean delivery (CD) requires innovative strategies, as they may never manage such cases in training. We used a multistage design process to create a criterion-referenced multiple-choice test as an assessment tool to evaluate CA1's knowledge related to this scenario.

Methods: Three faculty created 33 questions, categorized as: (1) physiologic changes of pregnancy (PCP), (2) pharmacology (PHA), (3) anesthetic implications of pregnancy (AIP), and (4) crisis resource management principles (CRM). A Delphi process (3 rounds) provided content validation. In round 1, experts (n = 15) ranked questions on a 7-point Likert scale. Questions ranked ≥ 5 in importance by ≥ 70% of experts were retained. Five questions were eliminated, several were revised, and 1 added. In round 2, consensus (N = 14) was reached in all except 7 questions. In round 3 (N = 14), all questions stabilized. A pilot test of the 29-question instrument evaluating internal consistency, reliability, convergent validity, and item analysis was conducted with the July CA1 classes at our institution after a lecture on GA for CD (n = 26, "instructed group") and another institution with no lecture (n = 26, "uninstructed group"), CA2s (N = 17), and attendings (N = 10).

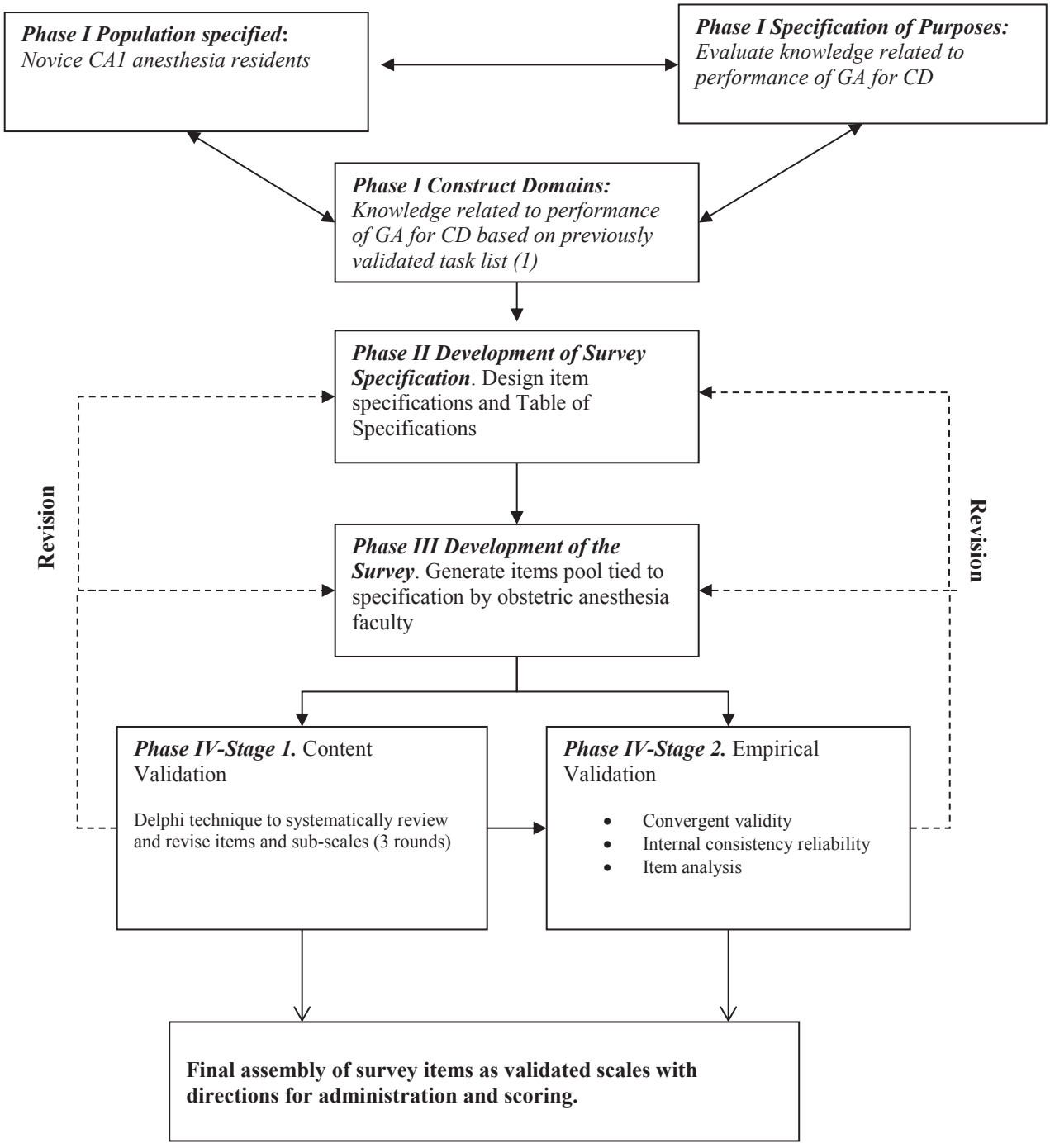
Results: Acceptable internal consistency and reliability was demonstrated ($\rho = 0.67$). Convergent validity coefficients between the CA1 uninstructed and instructed group suggested theoretical meaningfulness of the 4 sub-scales: PCP correlated at 0.29 with PHA, 0.35 with CRM, and 0.25 with AIP. PHA correlated with CRM and AIP at 0.23 and 0.28, respectively. The correlation between CRM and AIP was 0.29.

Conclusion: The test produces moderately reliable scores to assess CA1s' knowledge related to GA for urgent CD.

Key Words: Criterion-referenced test; content validation; empirical validation

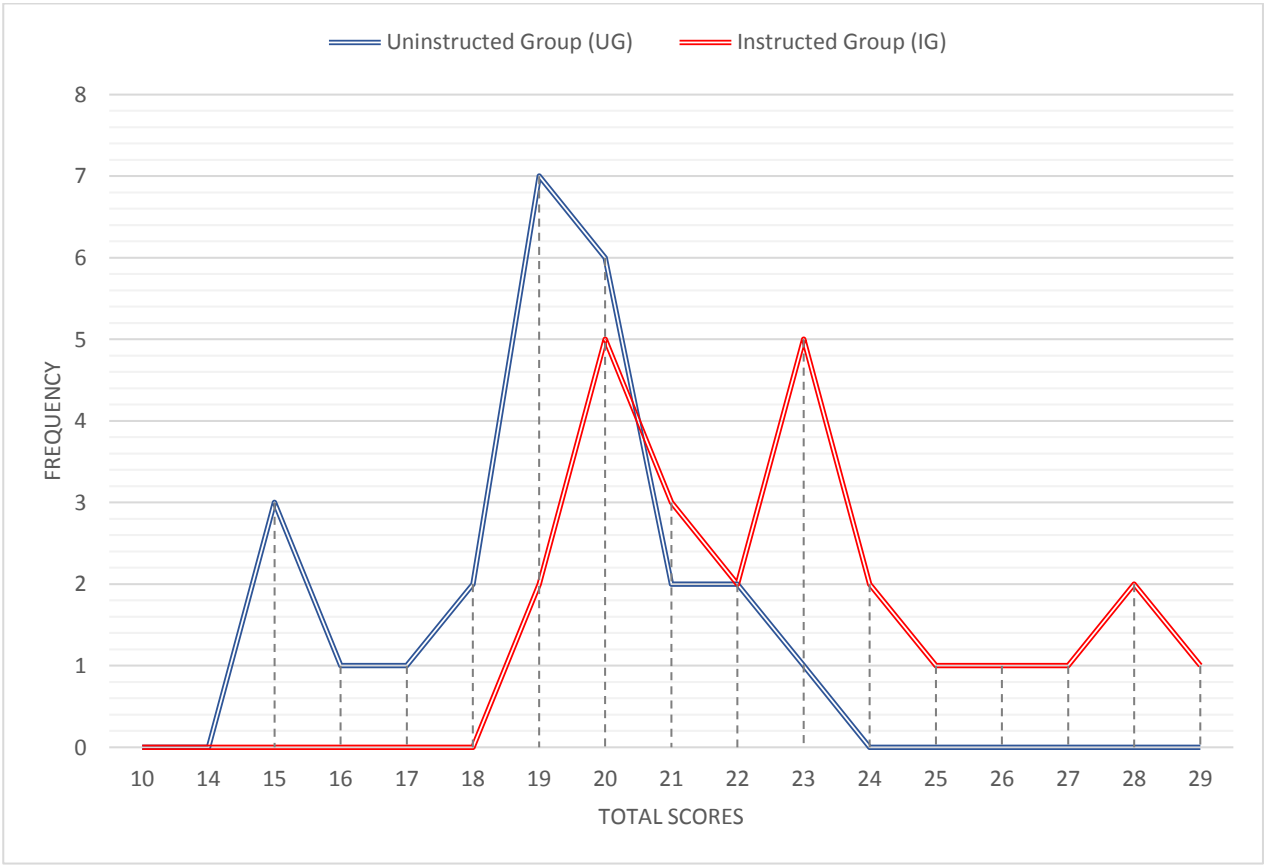
Figures

Figure 1. *Iterative process for designing and validating a knowledge test.*



Figures *continued*

Figure 2. *Standard-setting for the criterion-referenced knowledge test.*



Tables

Table 1

Subdomains	Keywords
Physiologic Changes of Pregnancy	Airway changes – increased edema, friability, risk of epistaxis
	Pulmonary changes – decreased FRC, decreased O_2 reserve, increased O_2 consumption
	Cardiovascular changes - increased CO, SV, HR; decreased SVR
	Gastrointestinal tract changes– pregnancy effect on gastric motility and acid secretion
	Central nervous system changes – decreased MAC for volatile agents
	Uteroplacental blood flow at term - mechanisms for stemming blood loss postpartum
	Normal fetal heart rate- recognize fetal bradycardia
	Hematologic changes - increased blood volume, increased red blood cell mass, hypercoagulability
	Renal changes - increased GFR, creatinine clearance
	Hepatalogic changes- decreased plasma proteins
Pharmacology	Effect of volatile agents on the uterus
	Succinylcholine metabolism – impact of lower pseudocholinesterase levels pregnancy
	Uterotonics -first vs. second line agents
	Sensitivity to catecholamines, vasopressors
Anesthetic implications of pregnancy	Gastrointestinal prophylaxis principles - use of nonparticulate antacids
	Adequate pre-oxygenation/denitrogenation - risk of rapid O_2 desaturation
	Principles of rapid sequence induction - preferred agents, avoidance of bag/mask ventilation, cricoid pressure
	Cricoid pressure - principles, how to apply
	Smaller endotracheal tube use in pregnanct
	Increased risk of failed/difficult intubation
	Factors underlying difficult intubation- engorged tissues, large breasts, obesity
	Failed airway rescue plan –glidescope, laryngeal mask airway
	Methods to confirm intubation - end tidal CO_2, auscultate for bilateral breath sounds
	Before delivery, volatile agent $\geq$ 1 MAC, Use of N_2O: O_2 50:50 ratio
	$FiO_2 \geq 0.5$ prior to delivery
	Risk of awareness
	Ventilation goals - end tidal CO_2
	Left uterine displacement -methods
	Oxytocin infusion - timing after delivery, dose
	After delivery, decrease volatile agent to $\leq$ 0.5 MAC
	After delivery, N_2O, hypnotics, opioids as needed
	Timing of abdominal prep/drape - different from non-pregnant GA

Tables *continued*

Table 2

Question	Round 1 Mean	Round 1 Median	Round 2 Mean	Round 2 Median	Round 3 Mean	Round 3 Median	Mean % Change Between Round 1 and 2	Mean % Change Between Round 2 and 3
1) AIRWAY: Which of the following physiologic changes MOST explains why pregnant patients are more difficult to intubate than non-pregnant?	6.07	6	5.64	5			7%	
2) PULMONARY CHANGES: A decrease in which of the following pulmonary parameters BEST explains the rapid oxygen desaturation after induction of general anesthesia in pregnancy?	6.20	6	6.29	6			1%	
3) RESPIRATORY SYSTEM: What is the normal arterial partial pressure of carbon dioxide (PaCO2) at term?	5.27	5	5.36	5			2%	
4) CARDIOVASCULAR SYSTEM: Which of the following cardiovascular changes normally occurs during pregnancy?	5.33	6	5.50	6			3%	
5) GASTROINTESTINAL SYSTEM: A decrease in which of the following factors MOST contributes to gastro-esophageal reflux during pregnancy?	5.47	5	5.43	6			1%	
6) CNS CHANGES: In pregnancy, the minimal alveolar concentration of volatile anesthetics changes in which direction?	4.80	5						
7) UTEROPLACENTAL FLOW: After delivery of the neonate, which of the following mechanisms is MOST responsible for decreasing maternal blood loss?	6.20	6	6.50	7			5%	
8) FETAL HEART RATE: What is the normal heart rate range in the term fetus?	6.33	7	6.29	7	5.86	6	1%	7%
9) HEMATOLOGIC CHANGES: Which of the following hematologic changes normally occurs during pregnancy?	5.47	6	5.36	6			2%	
10) VOLATILE AGENTS: Which physiologic effect is MOST commonly associated with administration of > 2 MAC of inhaled volatile agent during cesarean delivery?	6.07	6	6.21	6			2%	
11) SUCCINYLCHOLINE: An increase in which of the following pharmacokinetic parameters BEST explains the higher dose requirement of succinylcholine for rapid sequence induction in pregnancy?	4.40	5						
12) UTEROTONICS: Which of the following medications is the first-line uterotonic agent administered after delivery of the neonate at cesarean delivery?	6.07	6	5.93	6	6.07	6	2%	2%
13) EXOGENOUS CATECHOLAMINES: Which of the following BEST explains why a higher dose of vasopressors is needed in term parturients compared with non-pregnant women?	5.33	5	4.93	5			8%	
14) GI PROPHYLAXIS: Which of the following medications will raise the gastric pH fastest?	5.47	5	6.00	6			10%	
15) RISK OF ASPIRATION: Which of the following accounts for the higher risk of pulmonary aspiration of gastric contents during labor ?	4.80	5						
16) PREOXYGENATION: Which technique BEST describes adequate pre-oxygenation for emergent cesarean delivery?	6.47	7	6.07	6			6%	
17) RSI PRINCIPLES: The primary reason for performing a rapid sequence induction (RSI) of general anesthesia for cesarean delivery is to decrease the risk of which of the following outcomes?	5.80	6	6.43	6	6.21	6	11%	3%
18) Which technique BEST describes the correct application of cricoid pressure to prevent regurgitation of gastric fluid into the pharynx during rapid sequence induction?	5.27	5	5.36	5			2%	
19) ETT SIZE: What is the recommended endotracheal tube size for a term pregnant woman?	4.30	5						
20) DIFFICULT/FAILED INTUBATION RISK: What is the incidence of failed intubation in the obstetric setting?	5.53	6	5.93	6			7%	

Tables *continued*

Table 3

Table 3. Item analysis results for the uninstructed group (UG) and instructed group (IG)

Item	Item Difficulty Index for UG (n=25)	Item Difficulty Index for IG (n=25)	Item Discrimination Index (D)
Q1	14%	18%	4%
Q2	44%	46%	2%
Q3	12%	22%	10%
Q4	10%	16%	6%
Q5	18%	18%	0%
Q6	36%	44%	8%
Q7	38%	40%	2%
Q8	30%	36%	6%
Q9	12%	36%	**24%**
Q10	26%	26%	0%
Q11	26%	48%	**22%**
Q12	34%	48%	14%
Q13	32%	48%	16%
Q14	22%	32%	10%
Q15	32%	44%	12%
Q16	38%	40%	2%
Q17	14%	46%	**32%**
Q18	36%	48%	12%
Q19	26%	34%	8%
Q20	48%	50%	2%
Q21	42%	48%	6%
Q22	48%	50%	2%
Q23	24%	40%	16%
Q24	28%	46%	18%
Q25	34%	42%	8%
Q26	40%	46%	6%
Q27	48%	46%	-2%
Q28	42%	50%	8%
Q29	38%	34%	-4%

Reliability=0.67. Well-functioning discrimination index values (D≥20) are in bold text and highlighted. Borderline items (15<D<20) are highlighted.

Note: A case in UG (n=1) had missing data. To make both groups balanced in size, we randomly removed a case from the IG.

Tables *continued*

Table 4

Table 4. Convergent validity results for the uninstructed group (UG) (n=25) and instructed group (IG) (n=25)

	PCP	PHA	CRM	AIP	Total
PCP	1				
PHA	.293*	1			
CRM	.346*	.229	1		
AIP	.245	.275	.294*	1	
Total	.713**	.537**	.737**	.698**	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
Note: A case in UG (n=1) had missing data. To make UG and IG equal in numbers, one case was randomly removed from IG

Tables *continued*

Table 5

Table 5. Item analysis results for the uninstructed group (UG) and expert group (EG)

Item	Item Difficulty Index for UG (n=25)	Item Difficulty Index for EG (n=25)	Item Discrimination Index
Q1	4%	38%	**34%**
Q2	48%	50%	2%
Q3	16%	48%	**32%**
Q4	12%	36%	**24%**
Q5	26%	48%	**22%**
Q6	34%	46%	12%
Q7	44%	46%	2%
Q8	34%	50%	16%
Q9	18%	36%	18%
Q10	22%	38%	16%
Q11	26%	46%	**20%**
Q12	42%	50%	8%
Q13	44%	48%	4%
Q14	24%	46%	**22%**
Q15	38%	40%	2%
Q16	38%	50%	12%
Q17	28%	46%	18%
Q18	40%	48%	8%
Q19	22%	34%	12%
Q20	42%	50%	8%
Q21	44%	48%	4%
Q22	50%	50%	0%
Q23	16%	30%	14%
Q24	32%	48%	16%
Q25	30%	46%	16%
Q26	46%	46%	0%
Q27	46%	48%	2%
Q28	46%	48%	2%
Q29	40%	34%	-6%

Reliability=0.87. Well-functioning discrimination index values (D≥20) are in bold text and highlighted. Borderline items (15<D<20) are highlighted.

Note: A case in UG (n=1) had missing data. To make UG and EG equal in numbers, two cases were randomly removed from EG.

Tables *continued*

Table 6

Table 6. Convergent validity results for the uninstructed group (UG) (n=25) and expert group (EG) (n=25)

	PCP	PHA	CRM	AIP	Total
PCP	1				
PHA	.673**	1			
CRM	.598**	.648**	1		
AIP	.467**	.557**	.427**	1	
Total	.855**	.870**	.784**	.743**	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
Note: A case in UG (n=1) had missing data. To make UG and EG equal in numbers, two cases were randomly removed from EG.