THIEME
OPEN
ACCESS

Electronic Fetal Monitoring Credentialing Examination: The First 4000

Mark W. Tomlinson, MD, MBA[1]   Sara A. Brumbaugh, MS[2]   Marin O'Keeffe, RN[1]

Richard L. Berkowitz, MD[3]   Mary D'Alton, MD[3]   Michael Nageotte, MD[4] on behalf of the Perinatal Quality Foundation

[1] Perinatal Quality Foundation, Northwest Perinatal Center/Women's Healthcare Associates, Providence Health and Services, Oregon, Women and Children's Program, Portland, Oregon

[2] Perinatal Quality Foundation Consulting Statistician, Ceres Analytics, LLC, Oklahoma City, Oklahoma

[3] Perinatal Quality Foundation, Department of Obstetrics and Gynecology, Columbia University, New York, New York

[4] Perinatal Quality Foundation, Department of Obstetrics and Gynecology, Miller Children's and Women's Hospital, University of California, Irvine, California

Address for correspondence Mark W. Tomlinson, MD, MBA, Perinatal Quality Foundation, Northwest Perinatal Center/Women's Healthcare Associates, Providence Health and Services, Oregon, Women and Children's Program, 9701 SW Barnes Road, Street 299, Portland, OR 97225 (e-mail: mark.tomlinson@providence.org).

Abstract

Objective   Recognized variability in fetal heart rate interpretation led the Perinatal Quality Foundation (PQF) to develop a credentialing exam. We report an evaluation of the 1st 4000 plus PQF Fetal Monitoring Credentialing (FMC) exams.

Study Design   The PQF FMC exam is an online assessment for obstetric providers and nurses. The exam contains two question types: traditional multiple-choice evaluating knowledge and Script Concordance Theory (SCT) evaluating judgment. Reliability was measured through McDonald's Total Omega and Cronbach's Alpha. Pearson's correlations between knowledge and judgment were measured.

Results   From February 2014 through September 2018, 4,330 different individuals took the exam. A total of 4,057 records were suitable for reliability analysis: 2,105 (52%) physicians, 1,756 (43%) nurses, and 196 (5%) certified nurse midwives (CNMs). As a measure of test reliability, total Omega was 0.80 for obstetric providers and 0.77 for nurses. There was only moderate correlation between the knowledge scores and judgment scores for obstetric providers (0.38) and for nurses (0.43).

Conclusion   The PQF FMC exam is a reliable, valid assessment of both Electronic Fetal Monitoring (EFM) knowledge and judgment. It evaluates essential EFM skills for the establishment of practical credentialing. It also reports modest correlation between knowledge and judgment scores, suggesting that knowledge alone does not assure clinical competency.

Keywords
► electronic fetal monitoring
► Fetal Monitoring Credentialing
► interpretation of fetal monitoring

In essentially every labor and delivery (L&D) unit in the United States, Electronic Fetal Monitoring (EFM) is considered a basic competency for all nurses and providers of obstetrical care.[1] Despite ongoing debates about its accuracy in predicting neonatal outcomes, EFM is utilized in the overwhelming majority of deliveries in the United States.[2,3] Standard EFM definitions and guidelines were established in a 2008 conference sponsored by the Eunice K. Shriver

National Institute of Child Health and Human Development (NICHD), the American College of Obstetricians and Gynecologists (ACOG), and the Society for Maternal-Fetal Medicine (SMFM).[4]

A subsequent ACOG Practice Bulletin presented these definitions and guidelines. The purpose of the bulletin was not only to improve communication among clinicians who care for patients in labor but also to improve outcomes of perinatal morbidity and mortality.[5] The Joint Commission, which accredits and certifies over 20,000 health care organizations and programs in the United States, had previously identified inadequate interpretation of fetal monitoring as a root cause of 34% of sentinel events in L&D.[6] Specifically implicated in these sentinel events were communication issues (72%) and EFM orientation and training processes (40%).

Training in EFM interpretation is a foundational component of obstetrical resident education. Similarly, EFM training is routinely built into the basic obstetric orientation of nurses. Many textbook chapters, research articles, and continuing medical education (CME) courses are devoted to various aspects of EFM. However, a reliable, standardized approach to validate EFM competency was needed. To address this need, the Perinatal Quality Foundation (PQF) launched a Fetal Monitoring Credentialing (FMC) exam in 2014. The PQF's aim is to provide meaningful credentialing for EFM interpretation to those providing obstetrical care.[7]

Examinations developed by other organizations are available and focus primarily on EFM knowledge or they target specific caregiver groups. In contrast, the PQF developed separate but similar exams for such caregiver groups: nurses (registered nurses [RNs]) and obstetric providers (physicians and certified nurse midwives [CNMs]). While the questions may be very similar between the two exams, they differ by reflecting each group's different roles and responsibilities in EFM interpretation and management decisions. The FMC exam is novel in that it assesses competency not only through traditional knowledge-based questions but also unique judgment-based questions.

The judgment-based questions employ Script Concordance Theory (SCT), a method of testing that measures the consistency of a clinician's organization of knowledge for clinical action with that of experts.[8] This approach is particularly suited to evaluation of clinically uncertain scenarios that lack a clearly defined course of action.[9] The unique format of SCT questions is unfamiliar to most obstetric practitioners, although it has been validated and used in other disciplines.[10–13] A typical question presents a clinical situation, along with a previously determined course of action. Then new information is introduced, and the examinee is asked to rate his or her agreement with the previously determined action. Agreement is measured through a Likert-type scale. The questions are structured to mimic the dynamic clinical situations frequently encountered on L&D.[10]

The SCT is useful in situations where there is no clear single best approach and individual clinicians presented with the same clinical scenario may act somewhat differently. This is often the situation found on L&D when delivery decisions

involving a fetal heart rate (FHR) tracing entail multiple patient-specific factors such as parity, estimated time until delivery, and other clinical morbidities. One clinician may decide that the risk of an adverse outcome balanced with the anticipated time to achieve a vaginal delivery warrants proceeding with a cesarean section. In contrast, another may believe it reasonable to allow labor to proceed while continuing to monitor progress and the FHR tracing.

We report an evaluation of the first 4000 plus PQF FMC exams.

Methods

Data Collection

The PQF FMC exam is administered online at https://fmc.perinatalquality.org/. Nurse and obstetric provider exams each contain ~70 scored questions, of which 40 are classified as knowledge and 30 are classified as judgment (SCT) questions. Physicians and CNMs take the same provider exam, as their decision-making responsibilities on L&D are similar. An examinee that fails the initial exam is offered an opportunity to retake an exam similar in both length and content. Over time, some questions have been modified to improve validity based on feedback from examinees and ongoing statistical analysis. For example, wording was changed to make questions more clear. Statistical analysis included application of Item Response Theory, to estimate each question's difficulty, along with its ability to differentiate among examinees.[14]

To support continuing evolution of the exam, three to five unscored new questions are evaluated for future use. Each question is associated with a higher-level objective associated with important aspects of EFM interpretation and management. There are five such objectives:

- Interpretation of tracing per National Institute of Child Health and Human Development (NICHD) Guidelines
- Characterization of Category II
- Evaluation of Category II management
- Assessment of Category III risk
- Evaluation of Category III management

Each objective is addressed with a mix of both knowledge and judgment questions with the exception of the first objective, for which only knowledge questions are utilized. Relatively more questions are allocated to the objectives evaluating Category II tracings because Category II patterns account for the majority of tracings encountered in clinical practice.

At the beginning of each exam, demographic information is collected and includes clinician type (medical doctor [MD], RN, and CNM), years of clinical experience, degree for RNs (associate, bachelor, master, PhD), level of training (such as resident, fellow, attending), medical specialty (such as OB/Gyn generalist, OB hospitalist, MFM, family medicine), board certification, and faculty status.

Data filters were applied to identify valid first-time examinees. When necessary for reliability calculations, incomplete examinations were excluded. Incomplete exams often arise when the examinee does not reach the last question(s) because of the exam's time constraint. Over

time, monitoring of individual question performance led to dropping one RN and two MD exam questions due to poor statistical performance. For reliability analysis, these changes in exam content (e.g., removed questions) were retrospectively applied to all records.

Exam Scoring

Knowledge questions are comprised of two types, either a single best or multiple response answer (e.g., "Select all that apply"). For the multiple response questions, partial credit is granted when some, but not all responses are correct. Similarly, partial credit is deducted when some, but not all responses are incorrect. The final score is based on a composite of the knowledge and judgment components.

SCT questions are designed to reflect the inherent variability routinely experienced on L&D. Daily decision-making requires the obstetric provider to choose between allowing labor to continue and expediting delivery for fetal indications. As such, experts are not expected to always agree on the "best" course of action. Consistent with this expectation, there were no questions included in the exam where the experts were in complete agreement. Consequently, judgment (SCT) questions are subject to the aggregate scoring method.[15] In aggregate scoring, the answer chosen by most experts is awarded full credit. Remaining answers receive partial credit commensurate with the percentage of experts who chose the answer. ►Fig. 1 shows an example SCT question along with credit applied and an explanation of the expert responses from the exams study guide.

Thirty-four experts (nineteen MD/CNM, fifteen RN) were chosen based on recognized excellence in the area of EFM through publication or established superior clinical performance and leadership. Selection of experts followed guidelines previously established in the literature. For example, sample panels of 15 to 20-members were previously shown to correlate highly with a larger population of experts.[16] Selected experts came from both academic and community backgrounds demonstrating a breadth of experience in EFM, consistent with prior recommendation for such breadth over narrow niches of expertise.[9] Breadth of experience was

1. Case Content:

A 40-year-old G1P0 patient is admitted with a cervical exam of 7/0/80%/vertex. You review her FHR tracing shown in Panel A.

Your management plan is...

Expectant management while initiating conservative measures.

...and then you learn this additional information:

One hour later, a repeat pelvic examination reveals the cervix to be complete/+2/100%/vertex, and you review her FHR tracing shown in Panel B.

How does this additional information affect your thinking about the management plan?

- **Strongly invalidates**: full credit (the majority of the experts chose this option. Although the patient has progressed in labor, variability has decreased to minimal to absent and the decelerations have increased in depth and duration consistent with a Category III tracing. Preparations should be made either for an operative vaginal delivery or a cesarean section.)
- **Could invalidate**: partial credit (Some of the experts chose this option. The patient has progressed in labor fairly rapidly, however the variability has decreased to minimal to absent and the decelerations have increased in depth and duration consistent with a Category III tracing. Labor progress can be observed while preparations are be made either for an operative vaginal delivery or a cesarean section if she doesn't imminently deliver spontaneously vaginally.)
- **No impact**: no credit (none of the experts chose this option
- **Could support**: no credit (none of the experts chose this option)
- **Strongly supports**: no credit (none of the experts chose this option)

Fig. 1 Example of Script Concordance Theory Question from the Fetal Monitoring Credentialing Study Guide.

confirmed by statistical evaluation of potential experts' impacts on reliability indices.

Participation was confidential and voluntary.

Statistical Analysis

Reliability of the exam was calculated through coefficients known as Alpha and Omega. (Cronbach's) Alpha measures the reliability of a test with respect to one underlying concept. Alpha is considered to be biased low when an exam is known to measure multiple concepts.[17] Although not an optimal measure for this multiobjective exam, Alpha is reported here because of its frequent and familiar use. (McDonald's total) Omega measures the reliability of a test with respect to both a general, common factor and several supporting, specific factors.

Negative impacts on reliability coefficients are expected from several skewed items.[18] These items test important basic concepts and are usually answered correctly (knowledge questions) or answered in agreement with the experts (SCT questions). Further reductions of Alpha and Omega are expected in exchange for the exam's construct validity.[19] For example, while reliability coefficients could be increased simply by adding questions of redundant content, doing so would compromise the breadth of skills that the exam is intended to address within a reasonable time frame. Hence, we aim to satisfy, but not to exceed commonly accepted standards of reliability: 0.7 as minimal, and 0.8 as good.[20]

For purposes of comparison, knowledge and judgment scores were normalized to mean zero and a standard deviation of one. Pearson's correlations between knowledge and judgment scores were measured.

Results

The exam was introduced in early 2014. From introduction through September 30, 2018, 4,330 different individuals took the exam. Of these, 4,196 were determined to be valid first attempts. After exclusion of an additional 139 exams with some unanswered questions, 4,057 were suitable for reliability analysis (►Fig. 2). Those suitable attempts include 2,105 (52%) physicians, 1,756 (43%) nurses, and 196 (5%) CNMs. The majority of examinees registered through their organizations representing a broad range of practice settings, from midwife only to small community hospitals to large academic centers. Approximately a third registered as individuals through the PQF website. Exam completion over time is shown in ►Fig. 3. The pass rates for physicians/CNMs and nurses were 96.4 and 91.0%, respectively. The examinees' characteristics are detailed in ►Table 1.
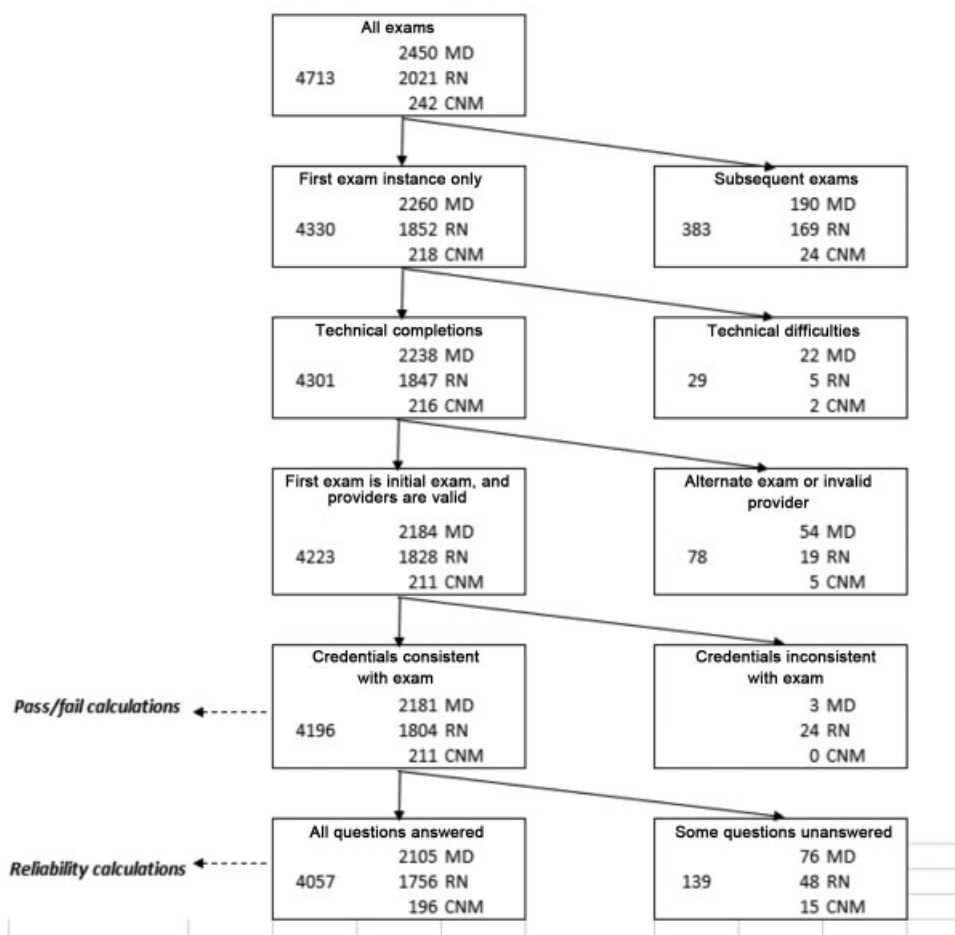

Fig. 2 Data filters. CNM, certified nurse midwife; MD, medical doctor; RN, registered nurse.
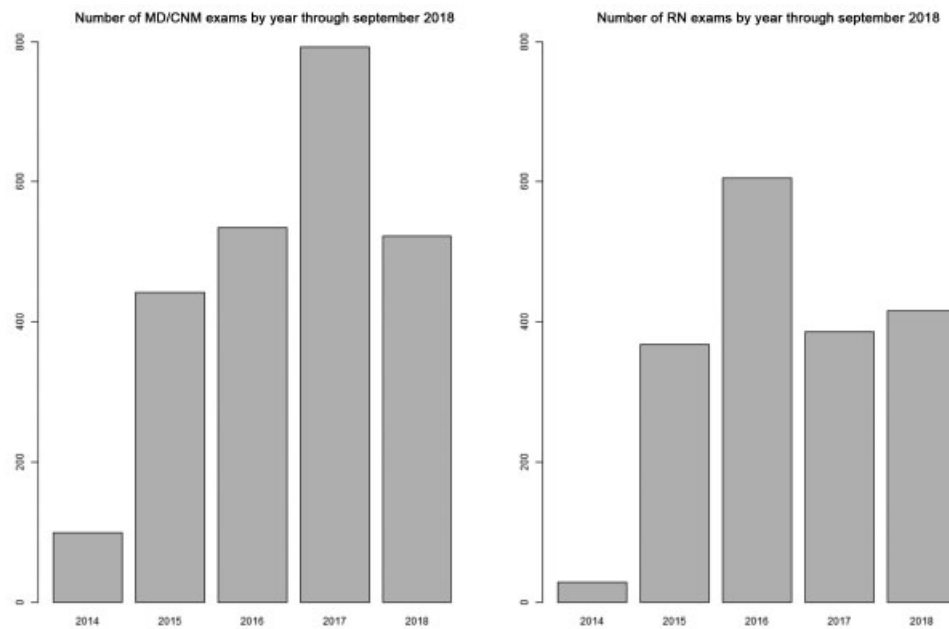
Fig. 3 Exam completion for MD/CNMs and nurses since introduction. CNM, certified nurse midwife; MD, medical doctor; RN, registered nurse.

Reliability

Alpha and total Omega were measured as follows, with reported values comprising the 4000+ exams administered to date:

- The MD/CNM exam Omega is 0.0.80; the RN exam Omega is 0.77.
- The MD/CNM exam Alpha is 0.77; the RN exam Alpha is 0.75.

Knowledge, Judgment, and Exam Objectives

Knowledge and judgment scores vary considerably with only moderate correlation between the two: 0.38 for physicians/CNMs and 0.43 for nurses. Scatter plots of the normalized knowledge and judgment scores are shown in ►**Fig. 4**.

Discussion

Analysis of initial 4000 plus PQF FMC exams utilizing knowledge and judgment components demonstrates its reliability. The overall pass rates were high; however, wide variability was noted in both knowledge and judgment scores. In addition, only moderate correlation was noted between knowledge and judgment scores.

The exam combines traditional questions that test basic knowledge concepts related to fetal monitoring with unique questions designed to assess obstetrical clinical judgment. The pass rates show the majority of individuals practicing on L&D have the necessary basic knowledge and clinical judgment skills to adequately interpret FHR tracings, although a wide range of both knowledge and judgment scores among both physicians and nurses were noted. We believe this range suggests a likely benefit from ongoing training and assessment. Analysis revealed only moderate correlations between individual knowledge and judgment scores. The lack of a strong correlation implies that an individual's basic knowledge of FHR monitoring principles does not automatically translate into practical clinical skills and judgment. This lack of translation leads to questions about the utility of traditional assessments that would focus only on knowledge.

A working definition of professional competence is "the habitual and judicious use of communication, knowledge, technical skill, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served."[21] An important component of competence includes the integration of basic knowledge with clinical judgment, an aspect not reliably assessed in traditional multiple-choice questions in a knowledge-only format. That format also cannot reliably incorporate the inherent ambiguity that often accompanies actual clinical practice.

Training and assessment of FHR monitoring interpretation have been almost accomplished primarily through didactics and observation.[22] After residency completion for physicians, or L&D orientation for nurses, there is no national standard to assure ongoing competency. To address this lack of standard, Berkowitz et al[7] articulated the case for a credentialing exam pointing out that individual obstetric providers and nurses who had been interpreting FHR tracings for years believed they were competent without any objective supporting evidence. At the same time chart reviews showed differences among these same providers' FHR interpretation. In 2005, the Medical Center Insurance Company, insuring several East Coast academic medical centers, required all caregivers to pass an EFM exam to work on L&D. Pettker and Grobman[23] in an expert review of quality and safety obstetrics also suggest that the standard of assuming competency based on completion of training or orientation program is inadequate.

Table 1 Examinee characteristics

n (% total)	MD	RN	CNM
	2,181 (52%)	1,804 (43%)	211 (5%)
Experience (% total)[a]			
Years experience (mean ± SD)	15.2(11.1)	10.7(9.8)	17(11.2)
n for years experience (% total)	1,483(35.3%)	1,625(38.7%)	198(4.7%)
Not reported (all exams)	890(21.2%)		
Education (% exam)			
Associate's degree		449 (24.9%)	4 (1.9%)
Bachelor's degree		1,131 (62.7%)	3 (1.4%)
Master's degree		154 (8.5%)	172 (81.5%)
None		3 (0.2%)	1 (0.5%)
Other		65 (3.6%)	26 (12.3%)
PhD		2 (0.1%)	5 (2.4%)
MD	1,979 (90.7%)		
DO	202 (9.3%)		
Clinical role (% exam)			
Ob/Gyn generalist	1,026 (47%)		
Maternal fetal medicine Subspecialist	202 (9.3%)		
Ob/Gyn hospitalist	102 (4.7%)		
Family medicine doctor	83 (3.8%)		
Fellow	65 (3%)		
Resident in training	673 (30.9%)		
Nurse midwife (CNM)			211 (100%)
Labor and delivery nurse		1,746 (96.8%)	
Nurse practitioner		29 (1.6%)	
Other1	30 (1.4%)	29 (1.6%)	
Board certified (% exam)[a]			
Yes	1,247(57.2%)		
Resident year (% exam)[a]			
PGY1	310 (14.2%)		
PGY2	155 (7.1%)		
PGY3	107 (4.9%)		
PGY4	87 (4%)		
Faculty status (% exam)[a]			
Yes	774 (35.5%)	372 (20.6%)	85 (40.3%)
Other	870 (39.9%)	175 (9.7%)	145 (68.7%)

Abbreviations: CNM, certified nurse midwife; MD, medical doctor; PGY1, postgraduate year 1; RN, registered nurse; SD, standard deviation.
[a]Not available for all examinees, category percentages shown may sum to less than 100%.

A strength of the PQF credentialing exam is that it aims to address such inadequacies by using SCT questions to measure integration of core EFM knowledge with clinical judgment. This type of question simulates real life by incorporating ambiguity routinely encountered in daily practice. Assessment using SCT has evaluated decision-making during gynecologic surgery[12] as well as surgical specialties and radiology.[10,11,13] In addition, the large number of individuals taking the exam allows robust statistical measurement of its reliability.

Despite the theoretic benefits of the PQF credentialing examination, one of the limitations is that it cannot overcome the well-known inability of FHR monitoring to predict accurately fetal compromise and long-term neurologic outcome.[24] However, EFM can identify fetuses at risk of acidemia and it currently remains a foundational component of current labor management.[13,25,26] As such, the exam can strengthen knowledge and judgment surrounding management of FHR tracings. While there is not yet direct evidence
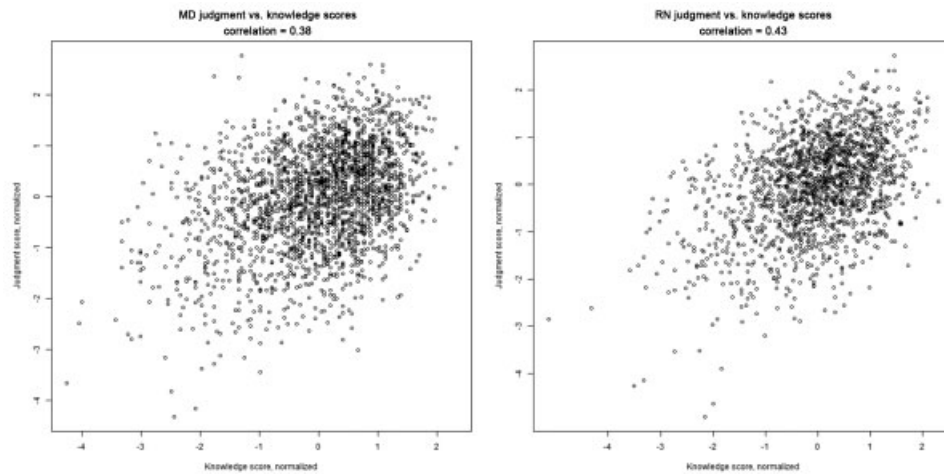
Fig. 4 Variation and distribution of knowledge scores compared with judgment scores. MD, medical doctor; RN, registered nurse.

that the exam is associated with improvement in clinical outcomes, comprehensive safety programs that have included training in FHR monitoring interpretation have demonstrated a decrease in adverse events and medicolegal costs.[27,28] Another limitation of the exam is that it only assesses individual performance and not participation in teamwork, which is an important skill necessary for a highly functioning L&D unit.

Opportunities for future work include analysis of knowledge and judgment scores by provider attributes, such as experience, training, and other characteristics collected in the FMC registration process. Given collection of over 4,000 exam results, opportunities for enhanced reporting include percentile rankings to suggest opportunities for focused training based on individual results. To assure the success of such training in the context of SCT, reliable methods to augment clinical judgment will need to be developed.

Conclusion

Analysis of more than 4,000 PQF FMC exams demonstrated a statistically reliable exam. For both physicians and nurses, knowledge and judgment scores are only moderately correlated, with such scores demonstrating clinically important variation. We believe that the modest correlation, together with the variation in scores, identifies opportunities for continued training and monitoring of competency with respect to both knowledge and judgment. Consequently, we can envision opportunities for training that address each a through a different mechanism. Lectures could improve knowledge and provide a foundation for judgment, while clinical simulation could advance development of judgment. Through improvement in both knowledge and judgment, FHR performance in L&D could be optimized.

Disclosure
There was no outside financial support or services.

References

1 Association of Women's Health. Obstetric and neonatal nurses. Fetal heart rate monitoring. J Obstet Gynecol Neonatal Nurs 2015; 44:683–686
2 Curtin SC, Park MM. Trends in the attendant, place, and timing of births, and in the use of obstetric interventions: United States, 1989-97. Natl Vital Stat Rep 1999;47(27):1–12
3 Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, Munson ML. Births: final data for 2002. Natl Vital Stat Rep 2003;52(10):1–113
4 Macones GA, Hankins GD, Spong CY, Hauth J, Moore T. The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines. Obstet Gynecol 2008;112(03):661–666
5 American College of Obstetricians and Gynecologists. Practice bulletin no. 116: Management of intrapartum fetal heart rate tracings. Obstet Gynecol 2010;116(05):1232–1240
6 The Joint Commission. Preventing infant death and injury during delivery. Sentinel Event Alert 2004;(30):1–3
7 Berkowitz RL, D'Alton ME, Goldberg JD, et al. The case for an electronic fetal heart rate monitoring credentialing examination. Am J Obstet Gynecol 2014;210(03):204–207
8 Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. Teach Learn Med 2000;12(04):189–195
9 Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. BMC Med Inform Decis Mak 2008; 8:18
10 Charlin B, Brailovsky CA, Brazeau-Lamon-tagne L, Samson L, Van der Vleuten CP. Script questionnaires: their use for assessment of diagnostic knowledge in radiology. Med Teach 1998;20:567–571
11 Meterissian S, Zabolotny B, Gagnon R, Charlin B. Is the script concordance test a valid instrument for assessment of intra-operative decision-making skills? Am J Surg 2007;193(02): 248–251
12 Park AJ, Barber MD, Bent AE, et al. Assessment of intraoperative judgment during gynecologic surgery using the Script Concordance Test. Am J Obstet Gynecol 2010;203(03):240.e1–240.e6
13 Sibert L, Darmoni SJ, Dahamna B, Hellot MF, Weber J, Charlin B. On line clinical reasoning assessment with Script Concordance test in urology: results of a French pilot study. BMC Med Educ 2006;6:45
14 Hableton RK, Swaminathan H. Item Response Theory: Principles and Applications. Boston, MA: Kluwer-Nijhoff Publishing; 1985
15 Groves M, Ed. The Diagnostic Process in Medical Practice: The Role of Clinical Reasoning. New York: Nova Science Publishers; 2007

16 Gagnon R, Charlin B, Coletti M, Sauvé E, van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? Med Educ 2005;39(03):284–291

17 Novick MR, Lewis C. Coefficient alpha and the reliability of composite measurements. Psychometrika 1967;32(01):1–13

18 Sheng Y, Sheng Z. Is coefficient alpha robust to non-normal data? Front Psychol 2012;3:34

19 Bandalos DL, Enders CK. The effects of nonnormality and number of response categories on reliability. Appl Meas Educ 1996;9(02): 151

20 Nunnally JC. Psychometric Theory. 2nd edition. New York: McGraw-Hill; 1978

21 Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002;287(02):226–235

22 Murphy AA, Halamek LP, Lyell DJ, Druzin ML. Training and competency assessment in electronic fetal monitoring: a national survey. Obstet Gynecol 2003;101(06):1243–1248

23 Pettker CM, Grobman WA. Obstetric safety and quality. Obstet Gynecol 2015;126(01):196–206

24 Grimes DA, Peipert JF. Electronic fetal monitoring as a public health screening program: the arithmetic of failure. Obstet Gynecol 2010;116(06):1397–1400

25 Williams KP, Galerneau F. Intrapartum fetal heart rate patterns in the prediction of neonatal acidemia. Am J Obstet Gynecol 2003; 188(03):820–823

26 Cahill AG, Tuuli MG, Stout MJ, López JD, Macones GA. A prospective cohort study of fetal heart rate monitoring: deceleration area is predictive of fetal acidemia. Am J Obstet Gynecol 2018;218(05): 523.e1–523.e12

27 Clark SL, Meyers JA, Frye DK, Perlin JA. Patient safety in obstetrics– the Hospital Corporation of America experience. Am J Obstet Gynecol 2011;204(04):283–287

28 Pettker CM, Thung SF, Lipkind HS, et al. A comprehensive obstetric patient safety program reduces liability claims and payments. Am J Obstet Gynecol 2014;211(04):319–325