



The Structured Oral Examination: A Method to Improve Formative Assessment of Fellows in Pediatric Endocrinology

Ranjit V. Shenoy, MD; Dorothee Newbern, MD; David W. Cooke, MD; Dennis J. Chia, MD; Leonidas Panagiotakopoulos, MBBS; Sara DiVall, MD; Lournaris Torres-Santiago, MD; Sitaram Vangala, MS; Nidhi Gupta, MD

From the Division of Pediatric Endocrinology, Department of Pediatrics, UCLA Mattel Children's Hospital (RV Shenoy), Los Angeles, Calif; Division of Pediatric Endocrinology & Diabetes, Phoenix Children's Hospital (D Newbern), Phoenix, Ariz; Division of Pediatric Endocrinology, Department of Pediatrics, John Hopkins University School of Medicine (DW Cooke), Baltimore, Md; Division of Pediatric Endocrinology, Department of Pediatrics, UCLA Mattel Children's Hospital (DJ Chia), Los Angeles, Calif; Division of Pediatric Endocrinology, Department of Pediatrics, Emory University (L Panagiotakopoulos), Atlanta, Ga; Division of Endocrinology, Department of Pediatrics, Seattle Children's Hospital/University of Washington (S DiVall), Seattle, Wash; Division of Endocrinology, Diabetes & Metabolism, Nemours Children's Health (L Torres-Santiago), Jacksonville, Fla; UCLA Department of Medicine Statistics Core (S Vangala), Los Angeles, Calif; and Division of Pediatric Endocrinology and Diabetes, Department of Pediatrics, Vanderbilt University Medical Center (N Gupta), Nashville, Tenn. Dr Shenoy is now with Division of Endocrinology, Diabetes & Metabolism, Nemours Children's Health, Jacksonville, Fla. The authors have no conflict of interest to disclose.

Address correspondence to Ranjit V. Shenoy, MD, Division of Endocrinology, Diabetes & Metabolism, Nemours Children's Health, 807 Children's Way, Jacksonville, FL 32207 (e-mail: ranjit.shenoy@nemours.org).

Received for publication July 19, 2021; accepted December 13, 2021.

ABSTRACT

OBJECTIVE: A structured oral exam (SOE) can be utilized as a formative assessment to provide high-quality formative feedback to trainees, but has not been adequately studied in graduate medical education. We obtained fellow and faculty perspectives on: 1) educational effectiveness, 2) feasibility/acceptability, and 3) time/cost of a SOE for formative feedback.

METHODS: Four pediatric endocrinology cases were developed and peer-reviewed to generate a SOE. The exam was administered by faculty to pediatric endocrinology fellows individually, with feedback after each case. Fellow/faculty perspectives of the SOE were obtained through a questionnaire. Qualitative thematic analysis was utilized to analyze written comments generated by faculty and fellows.

RESULTS: Seven of 10 pediatric endocrinology fellowship programs and all 18 fellows within those programs agreed to participate. Thematic analysis of fellow and faculty comments resulted in 5 perceived advantages of the SOE: 1) improved identification of clinically relevant knowledge deficits, 2) improved assessment of clinical reasoning, 3) immediate

feedback/teaching, 4) assurance of adequate teaching/assessment of uncommon cases, and 5) more clinically relevant assessment. Mean time to administer one case was 15.8 minutes (2.0) and was mentioned as a potential barrier to implementation. Almost all fellows (17/18, 94%) and faculty (6/7, 86%) would recommend or would most likely recommend implementation of the SOE into their curriculum.

CONCLUSIONS: The SOE utilized for formative feedback was perceived by fellows and faculty to have several educational advantages over current assessments and high acceptability. Objective educational advantages should be assessed on future studies of the SOE.

KEYWORDS: fellowship assessments; fellowship evaluations; graduate medical education assessments; medical education assessment; medical education evaluation; medical trainee assessment; medical trainee evaluation

ACADEMIC PEDIATRICS 2022;22:1091–1096

THE SHIFT TO competency-based medical education resulted in innovative methods to assess trainee competence in graduate medical education (GME), such as the Accreditation Council for Graduate Medical Education (ACGME) milestones, and more recently, Entrustable Professional Activities. Despite these advances in how we assess competence, the inadequacy of formative feedback provided to trainees continues to be a major concern of physician-educators and trainees alike.^{1–3} This may be

the case as the primary assessment tools utilized in most GME programs to provide formative feedback have inherent limitations.

For example, the In-Training Exam (ITE) is an annual multiple-choice assessment used to assess trainees' medical knowledge and readiness for their board certification exam. However, because the exam is designed to simulate board examinations, trainees' may receive limited feedback to protect exam content,

and testing of basic science and rare clinical scenarios may predominate in some specialties, rather than common clinical scenarios.⁴ Furthermore, the multiple-choice format can cue trainees to recognize correct answers, thus questioning its effectiveness in measuring higher-order thinking relevant for clinical practice.^{4–8} Indeed, the ITE has not been shown to correlate well with assessments of clinical performance.^{9–11}

Direct observation, another popular assessment utilized in GME, is usually reported on a global assessment form (“milestones” and “Entrustable Professional Activities”).⁴ It can be subject to rater bias, inaccurate recall, and depends on the availability of skilled, trained, and motivated faculty.^{4,12} Limited faculty time to observe and provide immediate feedback is a common and critical barrier.¹³ As a result, trainees frequently cite feedback from direct observation as inadequate and nonspecific.¹³

Thus, given significant limitations of these widely utilized assessments, trainees may not receive specific, objective, and clinically relevant formative feedback. However, the oral examination has potential to provide this high-quality feedback. Indeed, the ACGME has cited the oral examination as the most desirable tool for assessing medical knowledge and practice-based learning and improvement.¹⁴ Strengths of the oral exam include its ability to assess higher levels of Miller’s pyramid such as critical reasoning, problem solving, and judgment (“shows how”), as well as ability to express ideas, synthesize material, and think on one’s feet (“does”).¹²

Traditional oral exams lack standardization as the number, difficulty, and content of questions can vary significantly. Its use in GME in the United States has fallen out of favor in many specialties for summative examination (ie, board certification) for concerns of poor reliability, potential for bias, high cost/time, and threatening nature to trainees.¹⁴ In contrast, *structured* oral exams (SOE) follow a standardized rubric of questions and answers. Studies of the SOE utilized for *formative* feedback in undergraduate medical education suggest this format is more acceptable compared to traditional oral examinations as the standardized rubric inherently increases reliability, decreases bias and cost, allows for more efficient administration, and its formative purpose was not found to be threatening to trainees.^{15,16}

In GME settings, the SOE has favorable reliability, feasibility, acceptability, and low cost.^{17,18} However, detailed trainee and faculty perspectives of a SOE implemented for formative feedback are lacking, particularly with regards to its educational effectiveness. In addition, the SOE has not been studied in fellowship programs previously. In this study, our primary objectives were to provide detailed perspectives of fellow and faculty physicians on the: 1) educational effectiveness, 2) feasibility and acceptability, and 3) time/cost of a SOE for formative feedback.

METHODS

SETTING AND PARTICIPANTS

Ten pediatric endocrinology fellowship programs at tertiary care children’s hospitals across the United States were invited to participate. Eligible participants included one pediatric endocrinology faculty per program and pediatric endocrinology fellows at any level of training. The study took place in April and May 2019 and received certified exempt status by each participating program’s institutional review board. Verbal consent was obtained from each participating fellow.

EXAMINATION DEVELOPMENT

Messick’s validity framework was utilized.¹² The “target attribute” was fellows’ clinical knowledge and reasoning skills required to evaluate and manage common cases in pediatric endocrinology. Therefore, the lead author (R.S.) developed 4 common pediatric endocrinology cases (Congenital hypothyroidism is available in Supplement 1. Materials for precocious puberty, ambiguous genitalia, and late endocrine effects in childhood cancer survivors are available from the author). Objectives were developed based on this target attribute to assess patient care and medical knowledge competency. For each objective, at least one question was developed, along with an evidence-based scoring rubric. Each case consisted of 8 to 14 objectives. To support “content evidence,” all assessment tasks, phrasing/formatting of items, and scoring rubrics were peer-reviewed and edited by all 7 pediatric endocrinology faculty responsible for examination administration so that all were suitably aligned with the target attribute. To support “response process evidence,” the SOE was pilot tested by four fellow/faculty dyads to determine whether examinees’ cognition and behaviors when responding to assessment tasks were consistent with the intended interpretation and uses of scores and was modified accordingly. Fellows who participated in the pilot study were excluded from the full-scale study.

Passing scores for each case were determined by following “Eight Steps to Standard Setting” as described by Yudkowsky et al.¹² Four judges (participating faculty) utilized the Angoff Method to individually rate each item. After consensus, passing scores were determined: congenital hypothyroidism 20/27 (74.1%); precocious puberty 32/40 (80.0%); ambiguous genitalia 21/27 (77.8%); late endocrine effects 29/41 (70.7%); total score 102/135 (75.6%).

Prior to the study, there were no formal formative assessments utilized by participating fellowship programs other than the ACGME milestones and ITE.

QUESTIONNAIRE DEVELOPMENT

A combination of Likert scale and free response questions were utilized to gain insight into faculty and fellows’ perspectives on the educational effectiveness, time/cost, and feasibility and acceptability of a SOE for formative feedback (Supplement 1). This information helped determine “consequences evidence” as part of Messick’s

Table 1. Assessment Scores by Fellowship Year

Assessment	First Year Fellows (n = 7)	Second Year Fellows (n = 5)	Third Year Fellows (n = 6)	Total (n = 18)
ITE score*	119.6 (22.0)	155.0 (35.4)	158.6 (20.9)	144.4 (30.9)
PC3 milestone [†]	4.3 (1.3)	6.0 (1.3)	6.7 (1.0)	5.7 (1.5)
Congenital hypothyroidism [‡]	20.6 (3.2) (76.3%)	24.2 (2.4) (89.6%)	22.3 (3.0) (82.6%)	22.2 (3.1) (82.2%)
Precocious puberty [§]	28.6 (4.4) (71.5%)	31.0 (6.7) (77.5%)	31.5 (4.7) (78.8%)	30.2 (5.1) (75.5%)
Ambiguous genitalia	19.4 (3.9) (71.9%)	21.8 (4.7) (80.7%)	20.3 (3.8) (75.2%)	20.4 (4.0) (75.5%)
Late endocrine effects in CCS	26.3 (4.8) (64%)	30.4 (6.1) (74.2%)	32.5 (2.3) (79.3%)	29.5 (5.1) (71.9%)
Total SOE score [#]	94.9 (8.6) (70.3%)	107.4 (19.0) (80.0%)	106.7 (10.3) (79.0%)	102.3 (13.4) (75.8%)
Number of fellows who passed SOE**	1/7 (14.3%)	4/5 (80.0%)	3/6 (50.0%)	8/18 (44.4%)

CCS indicates childhood cancer survivors; SOE, structured oral exam.

Data presented as mean (SD) and (% correct) where applicable.

*Based on 2019 Scaled Subspecialty In-Training Examination Score. National mean scores (SD) in 2019 by fellowship year: first year 138 (35); second year 164 (36); third through fifth year 170 (33).

†Based on the most recent ACGME PC3 (patient care) milestone: "Develop and carry out management plans"; 1 to 9 scale with score of 5 = level 3 ("competent"), 7 = level 4 ("proficient"), 9 = level 5 ("expert").

‡27 total points.

§40 total points.

||27 total points.

¶41 total points.

#135 total points.

**Based on passing score of 102 out of 135 total points (75.6%).

validity framework.¹² The questionnaire was also pilot tested and modified based on feedback. Each fellow answered the questionnaire at completion of their SOE while faculty did so after all fellows in their program completed the SOE.

EXAMINATION ADMINISTRATION

At each represented institution, one faculty administered the SOE to a fellow, individually, in the faculty's office. Faculty read questions aloud, one at a time, allowing the fellow 60 seconds to respond. Fellows could take notes on scratch paper. Growth charts and laboratory results were provided when necessary. Start/stop times for each case were recorded. Fellows were asked not to share information with co-fellows postexamination.

FEEDBACK

At the end of each case, faculty reviewed answers with the fellow. At the end of the SOE, faculty provided scoring sheets to the fellow, which contained performance on all objectives (Supplement 1). For each objective, articles and/or book chapters were listed that provide answers/explanations. In addition, faculty provided written feedback and recommendations.

DATA ANALYSIS

Data were managed using REDCap (Vanderbilt University, Nashville, Tenn) and descriptive statistics were computed.¹⁹ Survey responses were stratified by rank (fellow

vs faculty) and fellow year. Associations were evaluated using Pearson correlations. Analyses were performed using R v 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria) (<http://www.r-project.org/>). We used qualitative, thematic analysis to analyze written comments generated by faculty and fellows. Two of the study authors (R.S. and N.G.) individually reviewed responses to free response questions, generating an inductive list of codes to create categories of responses. These were refined and combined into a mutually accepted coding structure.

RESULTS

All 10 faculty at 10 pediatric endocrinology fellowship programs initially agreed to participate. However, 3 faculty withdrew just prior to the start of the study for various reasons. All 18 fellows within the 7 participating programs agreed to participate in the study (Table 1).

EDUCATIONAL EFFECTIVENESS

All faculty (7/7) and 94% of fellows (17/18) perceived the formative SOE to be at least moderately effective in preparing a fellow for unsupervised clinical practice, with the majority perceiving it as very effective or extremely effective (average 3.92; 1 = not effective, 2 = slightly effective, 3 = moderately effective, 4 = very effective, 5 = extremely effective) (Table 2). Five primary advantages regarding perceived educational effectiveness emerged from free response questions (Table 3). Based on

Table 2. Perceived Effectiveness of Structured Oral Exam in Preparation for Unsupervised Practice

	Not Effective	Slightly Effective	Moderately Effective	Very Effective	Extremely Effective
Fellows (n = 18)	0	1 (5.6%)	5 (27.8%)	8 (44.4%)	4 (22.2%)
Faculty (n = 7)	0	0	2 (28.6%)	2 (28.6%)	3 (42.9%)

Table 3. Fellow and Faculty Perspectives of Advantages and Disadvantages of the Structured Oral Exam for Formative Feedback

Advantage	Representative Fellow Quotations	Representative Faculty Quotations
1. Improved identification of clinically relevant knowledge deficits (12 fellows, 6 faculty)	<p><i>"The exam was helpful to identify knowledge gaps that may not have been easily identified in an academic environment where other providers are readily available to assist and fill in knowledge gaps"</i></p> <p><i>"Bunch of scenarios at once helps identify areas you may be weak in and can focus on those prior to practicing."</i></p>	<p><i>"This exam gives the opportunity to directly examine the fellow's knowledge in core topics in real time and give immediate feedback after deficiencies are identified."</i></p> <p><i>"It is the only way of walking through a specific medical condition from start to finish to ensure understanding of the rationale for decision making and specifics of management."</i></p>
2. Direct observation of clinical reasoning (4 fellows, 3 faculty)	<p><i>"I felt that having to talk through my decisions and work through the cases without any feedback/opinions from other fellows or faculty was helpful. It made me more thoughtful in my consideration of the cases."</i></p> <p><i>"I realized that I have to be more organized in my thought process."</i></p>	<p><i>"Evaluation of clinical thinking, reasoning, and skills, in addition to knowledge."</i></p> <p><i>"Direct observation of thinking process."</i></p>
3. Immediate feedback and teaching (4 fellows, 2 faculty)	<p><i>"Structured one-on-one format with the attending (the immediate feedback) is useful for discussing cases; this format of teaching often missing from the busy clinic day or group didactics."</i></p> <p><i>"Discussing cases right away allows for learning and further questioning on cases or relevant cases you have on same topic."</i></p>	<p><i>"Immediate and in-person/face to face feedback"</i></p>
4. Ensuring adequate teaching and assessment of less commonly encountered cases (2 fellows, 2 faculty)	<p><i>"Covering less commonly seen cases like cancer survivor—are helpful in preparing for clinical practice."</i></p> <p><i>"Could focus more on cases we may not see as frequently in hospital/clinic so would have the opportunity to discuss."</i></p>	<p><i>"Often faculty will assume that trainees have covered some materials, yet it has not been explicitly taught. Therefore, it's useful to go through these scenarios to truly assess knowledge base."</i></p>
5. More clinically relevant assessment compared to assessments currently utilized (11 fellows, 3 faculty)	<p><i>"Better correlation to clinical practice than SITE exam. You can defend your answers better and not have to choose "the next best step". This approach is more realistic to real practice than tests."</i></p> <p><i>"More naturally follows challenges/thought process of actual clinical work."</i></p> <p><i>"This style of exam teaches and tests more relevant clinical information, gives immediate feedback, focuses on core fundamental clinical knowledge"</i></p>	<p><i>"ACGME milestones are rather non-specific and do not provide practical clinical benefit/assessment to the fellows. SITE exams are a great way to assess fellows' theoretical knowledge and get them to read and memorize. But it is not an indicator of fellow's clinical capabilities. This assessment gives a better idea about the real preparation a fellow has to practice endocrinology. Something a multiple choice exam would never provide."</i></p> <p><i>"More practical. Can assess fundamental concepts better."</i></p>
Disadvantages Various	<p>Representative Fellow Quotations</p> <p><i>"There are other aspects of a successful clinical practice that cannot be specifically tested for (e.g. patient interactions)."</i></p> <p><i>"Seeing patients in a different contexts is important for clinical practice. These cases by design are textbook. That said, the set of 32 cases would be more effective than board exams for clinical practice."</i></p>	<p>Representative Faculty Quotations</p> <p><i>"32 cases may take too long to administer even over the course of 3 years."</i></p> <p><i>"Agree could identify gaps in knowledge; but there is no guarantee they would fill this gap without retesting them."</i></p> <p><i>"The time constraints and oral nature of the test put added pressure on the fellows that sometimes is not present in clinical practice."</i></p>

the examination results, 89% of fellows (16/18) indicated that they would or would most likely utilize their scoring sheets from the SOE to read on areas of knowledge deficit and most faculty (5/7, 71%) would most likely re-administer failed cases, after a period of study (Table 4). Most fellows agreed with the statement: "This case identified clinically relevant knowledge deficits I was previously unaware of" for all topics (congenital hypothyroidism 15/18, 83%; precocious puberty 15/18, 83%; ambiguous

genitalia 17/18, 94%; late endocrine effects 18/18, 100%). Similarly, most faculty agreed with the statement: "This case identified clinically relevant knowledge deficits in the fellow that I was previously unaware of" for all topics (congenital hypothyroidism 12/18, 71%; precocious puberty 14/18, 78%; ambiguous genitalia 13/18, 72%; late endocrine effects 15/18, 83%). The SOE was significantly correlated with ITE score (0.78, $P < .001$) but not with ACGME PC3 milestone (0.51, $P = .053$).

Table 4. Questionnaire Results

		No	Unlikely	Unsure	Most Likely	Yes
Recommend implementation of exam into curriculum?	Fellows (n= 18)	0	0	1 (5.6%)	11 (61.1%)	6 (33.3%)
	Faculty (n= 7)	0	0	1 (14.3%)	3 (42.9%)	3 (42.9%)
Re-administer failed cases to fellow after a period of study?	Faculty (n= 7)	0	2 (28.6%)	0	4 (57.1%)	1 (14.3%)
Utilize scoring sheets from SOE to read and take notes on areas of knowledge deficits?	Fellows (n= 18)	0	1 (5.6%)	1 (5.6%)	7 (38.9%)	9 (50.0%)

SOE indicates structured oral exam.

TIME/COST

Time to develop one case was approximately 3 to 4 hours. The mean time to administer one case was 15.8 minutes (2.0) and by topic: congenital hypothyroidism 14.6 (2.8); precocious puberty 19.1 (3.6); ambiguous genitalia 14.4 (2.5); late effects 16.3 (3.9). Time for feedback was not included but was estimated to be 5 to 7 minutes per case on average. Cost to administer exam including printer paper, ink/toner, etc., was minimal.

FEASIBILITY/ACCEPTABILITY

All but one fellow (17/18, 94%) and one faculty (6/7; 86%) would recommend or would most likely recommend implementation of the exam into their curriculum (Table 4). A minority of fellows and faculty agreed with the following statements: “The exam introduced excessive subjectivity into the scoring process” (4/18 fellows, 22.2%; 0/7 faculty, 0%), “Faculty need additional specialized training in order to properly administer our structured oral exam” (3/18 fellows, 17%; 1/7 faculty, 14%), “The exam seemed threatening to me (the fellow)” (1/18 fellows, 6%; 2/7 faculty, 29%). There were no significant concerns regarding clarity of questions. Most fellows (12/18, 67%) and faculty (6/7, 86%) prefer faculty to administer the exam rather than co-fellow administered, or self-administered (Table 5).

DISCUSSION

Fellows and faculty described five distinct educational advantages of the SOE over traditional assessments such as the ACGME milestones and ITE (Table 3), while also finding it feasible and acceptable as almost all fellows and faculty would likely or most likely recommend implementation into their curriculum (Table 4).

The most cited educational advantage by fellows and faculty was the SOE’s ability to identify clinically relevant knowledge deficits (Table 3). There were unexpected knowledge deficits in many fellows. For example, 3/6 (50%) of third year fellows, who were in their final months of fellowship training, were unable to achieve a

passing total score of at least 75.6% on the SOE. Yet 2 of these 3 fellows achieved a “proficient” PC3 milestone rating (7 out of 9). Three other fellows had similar discrepancies: SOE scores ranging from 57.0% to 76.3% with PC3 milestone ratings of 6 to 8 (“competent” = 5; “expert” = 9). It is unclear whether stress from high expectations, presence of a time limit for responses, an emphasis on research over clinical practice during the third year of training, or other factors could have contributed to these discrepancies. Our small sample size limits our ability to make strong conclusions. However, the existing literature along with our performance data with fellow/faculty perspectives raises concerns that currently utilized formative assessments such as ACGME milestones and ITE may not be adequate in revealing clinically relevant knowledge deficits and supports the use of an additional assessment to allow knowledge deficits to be more readily apparent and remedied prior to graduation.

As described above, Messick’s validity framework was applied in the development of the SOE (*content, response process*).¹² Data obtained (*relations to other variables, consequences*) provide further support for validity. For example, the SOE was shown to have a strong relationship with ITE score (0.78, $P < .001$). Additionally, our postexam questionnaire provides support that the SOE may have positive consequences. For example, the majority of fellows (16/18) reported they will or will most likely read and take notes on areas of knowledge deficit (Table 4) and most faculty (5/7) report they would or would most likely retest fellows who failed a case after a period of study (Table 4).

The SOE has previously been shown to have favorable reliability, feasibility, acceptability, and time/cost in GME settings.^{17,18} We did not measure reliability, but we report favorable feasibility as described above. Acceptability was high as nearly all fellows and faculty would recommend or would likely recommend implementation into their curriculum.

Regarding time/cost, each case took, on average, 15.8 minutes to administer, not including feedback and review of answers. The *comprehensive* SOE would likely consist of approximately 30 cases. We anticipate that 3 to 4 hours per fellow, per year, would be required to administer and provide feedback for this examination over the course of a 3-year fellowship. While most faculty found this to be acceptable, some faculty noted this to be a potential barrier. However, several methods exist to decrease faculty burden such as recruiting multiple faculty members for exam administration, senior fellow administration, or self-administration (written exam), although most fellows and

Table 5. Exam Administration Preference

	Self-Administered	Fellow to Co-Fellow	Faculty to Fellow
Fellows (n = 18)	2 (11.1%)	3 (16.7%)	12 (66.7%)
Faculty (n = 7)	0	1 (14.3%)	6 (85.7%)

faculty would prefer faculty administration (Table 5). Faculty could utilize already scheduled weekly educational sessions for exam administration. Feedback and review could be completed with all fellows simultaneously. Additionally, examination questions could be limited to those that require critical thinking, rather than simple recall, and/or focusing on knowledge gaps that could have detrimental effects on patient care. However, these methods to decrease faculty time could alter effectiveness of the SOE. For example, a self-administered (written) version would not provide insight of the fellow's clinical reasoning or immediate in-person feedback and teaching, which were highlighted as distinct advantages of the SOE (Table 3).

One limitation is that faculty who agreed to participate were likely more accepting of this type of assessment, and therefore were more likely to view the SOE positively. Also, no follow-up was completed to determine the objective educational effect of the SOE on fellows' learning, knowledge retention, or clinical performance during or after fellowship; only fellow and faculty perspectives were analyzed. Finally, although this study was multi-institutional, this SOE was administered in a single pediatric subspecialty with relatively small numbers of participants. Thus, larger studies of the SOE should be pursued in a variety of specialties with follow-up assessments to determine objective educational effectiveness, in addition to feasibility, acceptability, and time/cost.

CONCLUSIONS

Our next steps are to work with an organization, such as the Pediatric Endocrine Society, to develop a comprehensive SOE consisting of approximately 30 cases. By utilizing a large organization, the SOE could be further enhanced in 3 ways; 1) improved access to national and international experts along with robust peer review to ensure the most relevant knowledge and skills required to practice pediatric endocrinology safely and effectively are being assessed; 2) to facilitate widespread exam dissemination and administration; and 3) to allow for larger studies to be completed.

Both fellows and faculty noted that the SOE administered for formative feedback had several educational advantages over currently utilized assessments. Future studies should investigate objective educational advantages of the SOE in a variety of GME settings to gain insight on the exam's ability to enhance preparation for unsupervised clinical practice.

ACKNOWLEDGMENTS

We thank the following pediatric endocrinologists for their review of the SOE: Paul Kaplowitz, MD, Alaina Vidmar, MD, Sara Ali Akhtar, MD, Kristin Favela, MD, and Christine Cherella, MD.

Financial statement: The authors report no external funding source for this study. We thank the following pediatric endocrinologists for their review of the SOE: Paul Kaplowitz, MD, Alaina Vidmar, MD, Sara Ali Akhtar, MD, Kristin Favela, MD, and Christine Cherella, MD.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at <https://doi.org/10.1016/j.acap.2021.12.032>.

REFERENCES

- Dujin C, Welink LS, Mandoki M, et al. Am I ready for it? Students' perception of meaningful feedback on entrustable professional activities. *Perspect Med Educ*. 2017;6:256–264.
- Anderson PA. Giving feedback on clinical skills: are we starving our young? *J Grad Med Educ*. 2012;4:154–158.
- Ende J. Feedback in clinical medical education. *JAMA*. 1983;250:777–781.
- Schwartz A, ed. *Assessment in Graduate Medical Education: A Primer for Pediatric Program Directors*. Chapel Hill, NC: American Board of Pediatrics; 2011.
- Newble DI, Baxter A, Elmslie RG. A comparison of multiple choice and free response tests in examinations of clinical competence. *Med Educ*. 1979;13:263–268.
- Case SM, Swanson DB. Extended-matching items: a practical alternative to free-response questions. *Teach Learn Med*. 1993;5:107–115.
- Veloski JJ, Rabinowitz HK, Robeson MR. A solution to the cueing effects of multiple-choice questions. *Med Educ*. 1993;27:371–375.
- Schuwirth LWT, van der Vleuten CPM, Donkers HJLM. A closer look at cueing effects in multiple-choice questions. *Med Educ*. 1996;30:44–49.
- Minhaj MM, Klafta JM, Tung A. Good doctor? Nice person? Correlations between ACGME competencies in an anesthesia training program. *American Society of Anesthesiologists Annual Meeting*. 2007.
- Hawkins RE, Sumption KF, Gaglione MM, et al. The in-training examination in internal medicine: resident perceptions and lack of correlation between resident scores and faculty predictions of resident performance. *Am J Med*. 1999;106:206–210.
- Clarke SL, Eydghi A. The relationship between In-training examinations and simulation assessments in an emergency medicine residency program. *SMRJ*. 2016;1: 4941. <https://doi.org/10.51894/001c.4941>.
- Yudkowsky R, Park YS, Downing SM, eds. *Assessment in Health Professions Education*. 2nd ed. New York, NY: Routledge; 2019.
- Cheung WJ, Patey AM, Frank JR, et al. Barriers and enablers to direct observation of trainees' clinical performance: a qualitative study using the theoretical domains framework. *Acad Med*. 2019;94:101–114.
- Accreditation Council for Graduate Medical Education, ACGME Outcomes Project. Toolbox of assessment methods. 2000. Available at: <http://dconnect.acgme.org/Outcome/assess/Toolbox.pdf>. Accessed May 12, 2019.
- Shenwai MR, Krishnakant BP. Introduction of structured oral examination as a novel assessment tool to first year medical students in physiology. *J Clin Diagn Res*. 2013;7:2544–2547.
- Patidar A, Chaturvedi JN. Introduction of a structured oral examination in formative assessment of pharmacology for 2nd professional MBBS students. *Int J Basic Clin Pharmacol [Online]*. 2019;8:253–258.
- Mount CA, Short PA, Mount GR, et al. An end-of-year oral examination for internal medicine residents: an assessment tool for the clinical competency committee. *J Grad Med Educ*. 2014;6:551–554.
- Jefferies A, Simmons B, Ng E, et al. Assessment of multiple physician competencies in postgraduate training: utility of the structured oral examination. *Adv Health Sci Educ*. 2011;16:569–577.
- Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–381.