

Using an Innovative Curriculum Evaluation Tool to Inform Program Improvement: The Clinical Skills Fair

ADITEE P. NARAYAN, MD, MPH
SHARI A. WHICKER, EdD, MEd
BETTY B. STAPLES, MD
JACK BOOKMAN, MAT, PhD
KATHLEEN W. BARTLETT, MD
KATHLEEN A. MCGANN, MD

Abstract

Background Program evaluation is important for assessing the effectiveness of the residency curriculum. Limited resources are available, however, and curriculum evaluation processes must be sustainable and well integrated into program improvement efforts.

Intervention We describe the pediatric Clinical Skills Fair, an innovative method for evaluating the effectiveness of residency curriculum through assessment of trainees in 2 domains: medical knowledge/patient care and procedure. Each year from 2008 to 2011, interns completed the Clinical Skills Fair as rising interns in postgraduate year (PGY)-1 (R1s) and again at the end of the year, as rising residents in PGY-2 (R2s). Trainees completed the Clinical Skills Fair at the beginning and end of the intern year for each cohort to assess how well the curriculum prepared them to meet the intern goals and objectives.

Results Participants were 48 R1s and 47 R2s. In the medical knowledge/patient care domain, intern scores improved from 48% to 65% correct ($P < .001$). Significant improvement was demonstrated in the following subdomains: jaundice (41% to 65% correct; $P < .001$), fever (67% to 94% correct; $P < .001$), and asthma (43% to 62% correct; $P = .002$). No significant change was noted within the arrhythmia subdomain. There was significant improvement in the procedure domain for all interns ($\chi^2 = 32.82$, $P < .001$).

Conclusions The Clinical Skills Fair is a readily implemented and sustainable method for our residency program curriculum assessment. Its feasibility may allow other programs to assess their curriculum and track the impact of programmatic changes; it may be particularly useful for program evaluation committees.

Editor's Note: The online version of this article contains sample intern medical knowledge/patient care questions used in this study and scoring instructions examples for sample questions.

All authors are at Duke University. **Aditee P. Narayan, MD, MPH**, is Associate Professor, Department of Pediatrics, and Associate Program Director, Pediatrics Residency Program; **Shari A. Whicker, EdD, MEd**, is Director of Education Development, Office of Pediatric Education; **Betty B. Staples, MD**, is Associate Professor, Department of Pediatrics, and Program Director, Pediatrics Residency Program; **Jack Bookman, MAT, PhD**, is Professor of the Practice Emeritus, Department of Mathematics; **Kathleen W. Bartlett, MD**, is Associate Professor, Department of Pediatrics, and Associate Program Director, Pediatrics Residency Program; and **Kathleen A. McGann, MD**, is Professor, Department of Pediatrics, and Vice Chair of Education, Office of Pediatric Education.

Funding: This study was funded by the Duke Graduate Medical Education Innovations Fund.

Portions of these results were presented at the Annual Spring Meeting of the Association of Pediatric Program Directors, San Antonio, TX, March 28–31, 2012.

The authors would like to thank the curriculum committee members, Duke pediatrics and combined internal medicine-pediatrics residents, and J. Marc Majure, MD.

Corresponding author: Aditee P. Narayan, MD, MPH, DUMC Box 3127, Durham, NC 27710, 919.684.2356, aditee.narayan@duke.edu

Received May 21, 2013; revision received September 12, 2013; accepted September 17, 2013.

DOI: <http://dx.doi.org/10.4300/JGME-D-13-00190.1>

Introduction

The Accreditation Council for Graduate Medical Education (ACGME) requires that training programs “must document formal, systematic evaluation of the curriculum at least annually” through program evaluation committees (PECs).¹ Curriculum evaluation is consistent with sound educational principles^{2,3} and is increasingly important in the ACGME’s Next Accreditation System.⁴

Despite the availability of program evaluation models, there is a need for consistent curriculum evaluation processes that are sustainable and well-integrated into program evaluation.⁵ One such program evaluation model, Musik’s model,⁶ includes needs assessment, focused methodology, and presentation and documentation of results. Other evaluation models have incorporated similar principles.^{2,7}

We present a Clinical Skills Fair as a rigorous curriculum evaluation method that assesses clinical skills based on level-specific learning objectives, thereby providing meaningful educational outcomes. This method helps programs identify curricular strengths and deficiencies, effectively informing program improvement.

Methods

This study was a quasi-experimental, pre-post design using a purposeful sample of Duke University pediatrics and combined internal medicine-pediatrics residents (interns) in postgraduate year (PGY)-1 from 2008 to 2011.

In 2006, the Duke Pediatric Residency Program formed a curriculum committee (CC) comprising program leadership and staff who represented a range of expertise, including primary care, education, hospital medicine, critical care, cardiology, infectious diseases, child abuse, and newborn medicine. The CC was created to supervise the development, implementation, and evaluation of the curriculum and identified a need for more rigorous evaluation. This led the CC to create the Clinical Skills Fair to assess how well the curriculum prepared trainees to meet level-specific goals and objectives. Using a modified Delphi technique,⁸ the CC reviewed the curriculum goals and objectives to identify the highest-yield areas, which included asthma, fever, jaundice, arrhythmias, and lumbar puncture. Corresponding goals and objectives were organized into 2 domains: medical knowledge/patient care (MK/PC) and procedure.

For MK/PC, the CC developed a written test based on intern-level goals and objectives, using the blueprinting process.⁹ Fourteen questions were organized into 4 subdomains: asthma, fever, jaundice, and arrhythmias. All questions were case-based and either multiple choice or short answer; 21 points were possible. Revisions were made based on pilot-testing with nonparticipating senior residents. Questions were entered into an electronic survey system (provided as online supplemental material). Criterion-referenced standards were not identified for this study; the intent was to demonstrate improvement in scores over time. The CC developed written scoring instructions based on consensus (provided as online supplemental material).

For the procedure domain, interns performed a lumbar puncture using a low-fidelity simulator. They were scored by a trained faculty member using scoring instructions based on procedural skills training modules and expert opinion.¹⁰ The test was criterion-referenced, and 10 of 13 points was set as the minimum standard, as determined by CC consensus to meet acceptable clinical performance. Copies of the intern curriculum and all test and scoring instructions are available from the authors.

We examined the construct validity of the Clinical Skills Fair using multiple sources of evidence. Content evidence was demonstrated through systematic test development as previously described. Two CC members scored subsets of the tests at separate points using written scoring instructions. Scores were compared between the 2 CC members and for each CC member separately over time.

What was known

Program evaluation committees must document formal, systematic evaluation of the curriculum at least annually.

What is new

A Clinical Skills Fair assesses program effectiveness through resident assessment in medical knowledge/patient care and procedure domains.

Limitations

The single-institution intervention reduces the ability to generalize; the intervention requires modest investment of faculty/staff time and equipment.

Bottom line

The Clinical Skills Fair is an effective and sustainable approach to program evaluation in a continuous improvement process.

We calculated the interrater and intrarater reliability using Cohen’s alpha. Once it was deemed reliable, the tool was used and scored by 1 faculty member for subsequent years.

The outcome measures included changes in the MK/PC and procedure test scores from beginning to end of the intern year. We also measured feasibility outcomes, including resident/faculty/staff time, equipment costs, information technology resources, staff training, and space.

A paired *t* test was performed annually on the aggregate data for each cohort. Clinical Skills Fair scores from the rising interns in PGY-1 (R1s) were compared to scores from the rising residents in PGY-2 (R2s). For example, scores from the 2008 R1s were compared with scores from the 2009 R2s, thereby comparing the same class of interns. We also compared all R1s to all R2s.

Data analyses were conducted in aggregate because the intent of the Clinical Skills Fair was to assess the curriculum, not the individual. Anonymity codes were used for all participants and were maintained by the research assistant. To enable participation in the intervention, interns were excused from clinical duties, which were covered by senior residents. They were not provided with individual results from the Clinical Skills Fair as part of the study design.

This study was deemed exempt by Duke University Medical Center’s Institutional Review Board.

Results

A total of 48 R1s and 47 R2s participated from 2008 to 2011. TABLE 1 provides participation rates per year. Using Cohen’s kappa, the interrater and intrarater reliability for the written test ranged between 0.89 and 0.95. The overall comparison between R1s and R2s increased from 48% to 65% correct (TABLE 1). There was also significant

TABLE 1 CLINICAL SKILLS FAIR MEDICAL KNOWLEDGE/PATIENT CARE DOMAIN RESULTS

	2008–2009		2009–2010		2010–2011		Overall	
Subdomains	R1	R2	R1	R2	R1	R2	R1	R2
n	14 ^a	17 ^a	18 ^b	17 ^b	16 ^c	13 ^c	48 ^d	47 ^d
Jaundice	R1	R2	R1	R2	R1	R2	R1	R2
Mean (SD)	0.464 (0.292)	0.662 (0.215)	0.389 (0.230)	0.735 (0.257)	0.391 (0.302)	0.538 (0.320)	0.411 (0.270)	0.654 (0.269)
Gain	0.198		0.346		0.148		0.243	
P value	.04		< .001		.21		< .001	
95% CI	(0.011, 0.384)		(0.179, 0.514)		(0.090, 0.386)		(0.133, 0.353)	
Effect size (Cohen's d)	0.782		1.421		0.476		0.901	
Fever	R1	R2	R1	R2	R1	R2	R1	R2
Mean (SD)	0.636 (0.268)	0.953 (0.118)	0.728 (0.214)	0.929 (0.145)	0.644 (0.333)	0.938 (0.100)	0.673 (0.271)	0.940 (0.121)
Gain	0.317		0.201		0.295		0.268	
P value	.001		.003		.003		< .001	
95% CI	(0.155, 0.480)		(0.075, 0.328)		(0.111, 0.478)		(0.182, 0.353)	
Effect size (Cohen's d)	1.59		1.099		1.51		1.27	
Arrhythmias	R1	R2	R1	R2	R1	R2	R1	R2
Mean (SD)	0.429 (0.164)	0.585 (0.150)	0.508 (0.206)	0.568 (0.187)	0.453 (0.182)	0.446 (0.218)	0.467 (0.186)	0.54 (0.190)
Gain	0.159		0.06		−0.0007		0.074	
P value	.01		.38		.93		.06	
95% CI	(0.041, 0.272)		(−0.076, 0.195)		(−0.159, 0.146)		(−0.003, 0.150)	
Effect size (Cohen's d)	1.003		0.301		0.035		0.393	
Asthma	R1	R2	R1	R2	R1	R2	R1	R2
Mean (SD)	0.339 (0.287)	0.614 (0.242)	0.538 (0.268)	0.629 (0.296)	0.395 (0.302)	0.615 (0.308)	0.432 (0.292)	0.620 (0.275)
Gain	0.28		0.09		0.22		0.19	
P value	.007		.35		.06		.002	
95% CI	(0.080, 0.469)		(−1.03, 0.284)		(−0.0125, 0.454)		(0.072, 0.303)	
Effect size (Cohen's d)	1.044		0.321		0.725		0.661	
Overall	R1	R2	R1	R2	R1	R2	R1	R2
Mean (SD)	0.458 (0.110)	0.670 (0.111)	0.501 (0.143)	0.684 (0.132)	0.454 (0.113)	0.576 (0.149)	0.476 (0.125)	0.649 (0.135)
Gain	0.212		0.183		0.122		0.173	
P value	< .001		.001		.02		< .001	
95% CI	(0.13, 0.29)		(0.080, 0.269)		(0.022, 0.222)		(0.120, 0.226)	
Effect size (Cohen's d)	1.920		1.269		0.935		1.332	

Abbreviations: CI, confidence interval; R1, rising intern in postgraduate year 1; R2, rising resident in postgraduate year 2.

^a 21 possible.

^b 22 possible.

^c 22 possible.

^d 65 possible.

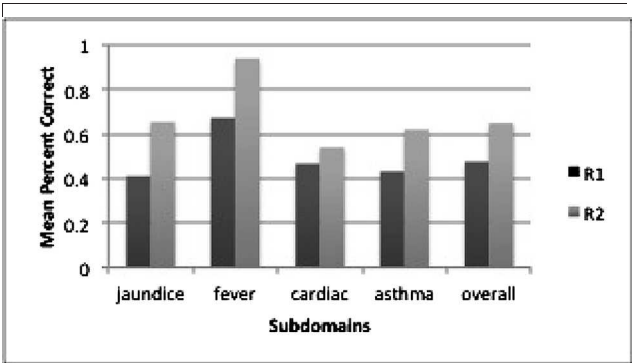


FIGURE **COMPARISON OF RISING PEDIATRICS INTERN (R1) AND RISING SECOND-YEAR RESIDENT (R2) CHANGES IN MEANS ACROSS THE SUBDOMAINS**

improvement in Clinical Skills Fair MK/PC scores for each cohort per year.

Subset analyses were performed for each subdomain (TABLE 1). We demonstrated significant improvements in jaundice, fever, and asthma. There was no significant improvement in the arrhythmia subdomain. Some variability in significance was noted per year. The FIGURE compares the changes in scores between R1s and R2s across the subdomains. There was significant improvement in overall lumbar puncture scores for 2 of 3 years (TABLE 2).

We also measured feasibility outcomes. Moderate faculty/staff investment was required but was part of their existing responsibilities. The Clinical Skills Fair also required moderate time commitments and collaboration within the department. TABLE 3 enumerates the resources needed.

Discussion

The Clinical Skills Fair allowed for effective program evaluation and may serve as a model for PECs. The MK/PC

tests consistently identified strengths in asthma, fever, and jaundice, whereas the procedure test demonstrated mastery of lumbar puncture skills. Therefore, the CC believed the curriculum was effective in these subdomains and no significant interventions were necessary. In contrast, the CC focused on deficiencies in arrhythmias, regardless of the fact that all interns had completed Pediatric Advanced Life Support training before their intern year. These results were shared with program leadership and faculty who identified recent system changes that likely impacted learning. Changes included a new cardiac intensive care unit that did not incorporate residents. Previously, cardiac patients were seen by residents in the general and neonatal intensive care units. Before implementing the Clinical Skills Fair, these unintended consequences were unrecognized. We collaborated with cardiology and nursery faculty to incorporate cardiac teaching into the newborn rotation. We are planning for additional electrocardiogram and arrhythmia teaching in cardiology electives and noon conferences. Performance in subsequent Clinical Skills Fair assessments should determine whether these changes were effective.

The feasibility outcomes suggested that the Clinical Skills Fair required moderate investments, particularly in faculty/staff time. Leadership commitment was required to support the program. Faculty development models can support faculty in such roles without financial compensation.¹¹ Programs may need to consider using existing committees, such as PECs, to develop a Clinical Skills Fair similar to the approach described here.

The Clinical Skills Fair required a moderate initial cost investment in equipment. The same equipment has been used for multiple years and represented a reasonable investment for the returns. Programs may need to identify funds to implement this type of program evaluation or leverage existing resources.

Our study demonstrated new findings. Other studies have shown how objective structured examinations can

TABLE 2	CLINICAL SKILLS FAIR PROCEDURE (LUMBAR PUNCTURE) DOMAIN RESULTS							
	2008–2009		2009–2010		2010–2011		Overall	
	R1	R2	R1	R2	R1	R2	R1	R2
n	14 ^a	17 ^a	18 ^b	17 ^b	16 ^c	13 ^c	48 ^d	47 ^d
Mean (SD)	6.7 (2.3)	10.09 (1.44)	9.8 (1.52)	11.9 (0.33)	9.5 (1.63)	11.9 (0.33)	8.79 (2.23)	11.28 (1.21)
χ ² test	χ ² = 28.31, P < .001		χ ² = 5.26, P = .07		χ ² = 8.2, P = .02		χ ² = 32.82, P < .001	

Abbreviations: R1, rising intern in postgraduate year 1; R2, rising resident in postgraduate year 2.

^a 21 possible.

^b 22 possible.

^c 22 possible.

^d 65 possible.

TABLE 3 CLINICAL SKILLS FAIR RESOURCES USED FOR EACH DOMAIN

Domain	Activities	Resources Needed
Medical Knowledge/ Patient Care	• Creation of written test (Web-based) • Use of existing computer lab • Residency program staff assistant to monitor the computer lab • Faculty training on scoring process • Trained faculty member scores the answer sheets	• CC semimonthly hourly meetings for 1 year • No additional cost • 1.5 hours/year • 1 hour • 2 hours/year
Procedures	• Low-fidelity simulator • Use of expired equipment (needles, LP kits that were expired/surplus and donated by the hospital) • Faculty training on scoring process • Trained faculty member to observe the procedures and score in real time	• < \$500 • No additional cost • 1 hour • 1.5 hours/year
Overall	• Resident time to complete the Clinical Skills Fair • Staff training on Clinical Skills Fair process • Use of existing conference rooms (1 for each domain) • Staff assistant to facilitate residents moving through the 2 stations • Staff assistant or chief resident time to schedule clinical coverage while residents complete Clinical Skills Fair • CC review of Clinical Skills Fair results	• 1.5 hours per resident/year • 2 hours • No additional cost • 1.5 hours/year • 2 hours/year • 2 hours/year

Abbreviations: CC, competency committee; LP, lumbar puncture.

provide useful assessment of clinical competence.¹² Our results supported these findings. Moreover, we demonstrated that systematic test development can be sustainable and long-term components of program evaluation. The Clinical Skills Fair was replicated annually for 4 years. It correlated well with other measures of performance supported by the literature to assess knowledge, such as in-training examination (ITE) scores.¹³ The ITE scores from our same cohort of residents similarly identified the need for improvement in cardiology knowledge. However, our study provided additional information beyond the ITE. The Clinical Skills Fair measured growth through the pre-post design to track improvements with observable outcomes tailored to program needs.

This study has limitations. Data were analyzed in aggregate for curriculum evaluation rather than to assess individual resident competency. A Clinical Skills Fair can be used for individual resident tracking as part of the overall assessment process but not as the sole tool for high-stakes assessments. Further validity testing should be done before such use.

Although low Clinical Skills Fair scores may be related to the curriculum, other factors must be considered, including issues related to the test, such as phrasing, scoring, delivery, and environment. As with other assessment tools, the data must be considered in context and alternative explanations explored. We attempted to address this limitation through systematic test development.⁹

Next steps in the development of the Clinical Skills Fair approach may include adaptation for individual resident assessment. We analyzed data in aggregate using codes to preserve resident anonymity while the Clinical Skills Fair was

tested and to avoid potentially adverse impressions of the resident based on pilot-test performance. We showed that the Clinical Skills Fair is a reasonable approach; therefore, it may be used as a component of individual assessment.

Competence as measured by the Pediatrics Milestones is ideally based on multiple assessment methods.¹⁴ Development of a single valid test to assess overall competence is difficult given the variability of performance across settings.^{9,14} Using multiple assessment methods provides a broader sampling per resident and a more meaningful holistic assessment.⁹ The Clinical Skills Fair offers 1 measure of how well residents meet goals and objectives using a standardized approach.

Conclusion

The Clinical Skills Fair provides focused curriculum and program evaluation in a structured and replicable manner, is implemented relatively easily, and is sustainable. It can serve as a model for PECs to refine curricula and track outcomes as continuous process improvement.

References

- 1 Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Pediatrics. http://www.acgme.org/acgmeweb/Portals/0/PFAssets/2013-PR-FAQ-PIF/320_pediatrics_07012013.pdf. Accessed December 9, 2013.
- 2 Kern DE, Thomas PA, Hughes MT. *Curriculum Development for Medical Education: A Six-Step Approach*. 2nd ed. Baltimore, MD: Johns Hopkins University Press; 2009.
- 3 Morrison GR, Ross SM, Kemp JE, Kalman H. *Designing Effective Instruction*. 6th ed. Hoboken, NJ: John Wiley & Sons Inc; 2010.
- 4 Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056.
- 5 Reed D. Nimble approaches to curriculum evaluation in graduate medical education. *J Grad Med Educ*. 2011;3(2):264–266.
- 6 Musick DW. A conceptual model for program evaluation in graduate medical education. *Acad Med*. 2006;81(8):759–765.

- 7 Branson RK, Rayner GT, Cox JL, Furman JP, King FJ, Hannum WH. *Interservice Procedures for Instructional Systems Development*. 5 vols. Ft Monroe, VA: US Army Training and Doctrine Command; 1975. TRADOC Pam 350-30 NAVEDTRA 106A. NTIS No. ADA 019 486 through ADA 019 490.
- 8 Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. *Manage Sci*. 1963;9(3):458–467.
- 9 Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945–949.
- 10 Ellenby MS, Tegtmeyer K, Lai S, Braner DA. Videos in clinical medicine. Lumbar puncture. *N Engl J Med*. 2006;355(13):e12.
- 11 Narayan AP, Whicker SA, McGann KA. An innovative process for faculty development in residency training. *Teach Learn Med*. 2012;24(3):248–256.
- 12 Carraccio C, Englander R. The objective structured clinical examination: a step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med*. 2000;154(7):736–741.
- 13 Althouse LA, McGuinness GA. The in-training examination: an analysis of its predictive value on performance on the general pediatrics certification examination. *J Pediatr*. 2008;153(3):425–428.
- 14 Hicks PJ, Englander R, Schumacher DJ, Burke A, Benson BJ, Guralnick S, et al. Pediatrics milestone project: next steps toward meaningful outcomes assessment. *J Grad Med Educ*. 2010;2(4):577–584.