

A Pilot Study of the Creation and Implementation of a Teaching Development Assessment Tool

Jennifer K. O'Toole, MD, MEd
 Melissa D. Klein, MD, MEd
 Daniel McLinden, EdD
 Heidi Sucharew, PhD
 Thomas G. DeWitt, MD

ABSTRACT

Background The importance of effective clinical teaching skills is well established in the literature. However, reliable tools with validity evidence that are able to measure the development of these skills and can effectively be used by nonphysician raters do not exist.

Objective Our initiative had 2 aims: (1) to develop a teaching development assessment tool (TDAT) that allows skill assessment along a continuum, and (2) to determine if trained nonphysicians can assess clinical teachers with this tool.

Methods We describe the development of the TDAT, including identification of 6 global teaching domains and observable teaching behaviors along a 3-level continuum (novice/beginner, competent/proficient, expert) and an iterative revision process involving local and national content experts. The TDAT was studied with attending physicians during inpatient rounds with trained physician and nonphysician observers over 6 months.

Results The TDAT showed emerging evidence of content, construct, and viable validity (the degree to which an assessment tool is practical, affordable, suitable, evaluable, and helpful in the real world) for the evaluation of attending physicians on inpatient rounds. Moderate to near perfect interrater reliability was seen between physician and nonphysician raters for the domains of promotion of clinical reasoning, control of the learning environment, ability to teach to multiple levels of learners, and provision of feedback.

Conclusions The TDAT holds potential as a valid and reliable assessment tool for clinical teachers to track the development of each individual's teaching skills along the continuum from early development to mastery.

Introduction

The development of an individual's skills as a clinical teacher occurs over time and, similar to other skills in medicine, can be tracked through the achievement of discrete observable milestones. While there are well-accepted skills for effective clinical teachers,^{1,2} previous studies of tools to assess these skills have not measured the longitudinal development of these skills.³⁻⁶ Most of these tools have been studied in simulated teaching sessions that utilize physicians or physicians-in-training as raters.⁷⁻¹¹ Only a few were studied during live teaching encounters,^{3,5,7} and none to date have used nonphysician raters, a valuable quality in an era of increasing demands on physicians' time. Additionally, many of these tools do not have established evidence of validity.

In 2011, a General Pediatric Master Educator Fellowship was established at Cincinnati Children's Hospital Medical Center (CCHMC) to formally train pediatricians in medical education.¹² Our fellows, who are graduates of accredited pediatrics residencies, receive training in curriculum development, evaluation methodologies, and educational scholarship. They are also encouraged to develop their teaching skills in a variety of clinical settings for which we sought an assessment tool to document this development over time. However, we were unable to find an assessment tool with evidence of validity that tracked skill development longitudinally.

To assess the evolution of one's teaching skills on inpatient rounds we sought to construct a novel teaching development assessment tool (TDAT). This educational innovation had 2 aims: (1) to develop a tool based on rigorous descriptions of observable teaching behaviors (OTBs) along a continuum from beginner to mastery for 6 key teaching domains; and (2) to determine if this tool could reliably be used by trained nonphysician observers. We hypothesized that the TDAT would show evidence of validity and would

DOI: <http://dx.doi.org/10.4300/JGME-D-14-00439.1>

Editor's Note: The online version of this article contains the Cincinnati Children's Hospital General Pediatrics Master Educator Fellowship Teaching Development Assessment Tool-INPATIENT.

prove to be a viable alternative to scarce faculty resources through reliable measurement by a trained nonphysician.

Methods

We describe the construction of the TDAT, and the assessment of its validity and interrater reliability for physician and trained nonphysician observers.

The Institutional Review Board at CCHMC approved this study as exempt.

Development of the TDAT

Construction of the TDAT began with a review of the literature on best practices in teaching and evaluation tools to identify the core teaching domains for an inpatient attending physician.^{1,2,4-6,13} This review yielded 12 potential teaching domains. Through an iterative process of in-person and phone meetings, using medical education experts locally and nationally, these domains were consolidated into 6 key teaching domains: (1) promoting clinical reasoning; (2) taking control of the inpatient learning environment; (3) using learner-centered teaching techniques; (4) having the ability to teach to multiple levels of learners; (5) instructing learners in physical examination techniques; and (6) providing feedback to learners.

Each of these key teaching domains was then broken down into discrete OTBs. Each OTB contained 3 distinct levels that demonstrate the pathway from novice to expert. The relationship between teaching domains, OTBs, and the continuum of levels of achievement are depicted in the online supplemental material.

During the development of the TDAT a trained physician spent 5 hours training a nonphysician research assistant (RA) on the components of effective teaching skills and the use of the TDAT. Together they piloted the tool during 6 inpatient rounding sessions, after which the RA completed another 4 sessions on her own, for a total of 16 hours of dedicated training. The RA then provided feedback on the OTBs to ensure they were easily identifiable to a nonphysician observer. The tool also underwent meticulous review by members of the CCHMC Medical Education Scholarship Team, the General Pediatric Master Educator Fellowship Oversight Committee, and mentors and scholars in the Academic Pediatric Association's Educational Scholars Program.¹⁴ Edits and revisions were made following each pilot use and review session.

What was known and gap

Teaching skills are important, yet there are no reliable tools for assessing them that can be used by nonphysician raters.

What is new

A teaching development assessment tool showed acceptable validity and high interrater reliability.

Limitations

Single institution, small sample study limits generalizability; validity evidence does not include comparisons with accepted measures of teaching proficiency.

Bottom line

The tool can be used by trained nonphysician observers, and shows promise for tracking the development of teaching skills.

Each of the inpatient rounding sessions included a team comprised of an attending physician, 1 to 2 senior residents, 2 to 3 interns, and 3 to 4 medical students for a time period of 1 to 3 hours (depending on the team census). The attending physicians were responsible for supervising and educating the various levels of learners, as well as overseeing the team's medical decision making. The attending physicians consisted of faculty or fellows from the Division of Hospital Medicine in the Department of Pediatrics at CCHMC. Only 1 attending physician was observed during each rounding session. At the end of an observed inpatient rounding session, consisting of 5 to 12 patient encounters on average, each observer selected 1 level for each OTB they observed the attending physician frequently executing. In scoring the tool, if a particular OTB was not observed, the observer was instructed to leave that OTB blank. Lack of rating did not imply performance at the novice level; rather it indicated that the attending physician did not demonstrate the particular OTB during a given rounding session. The tier where the majority of OTBs clustered determined the global rating assigned (ie, novice/beginner, competent/proficient, or expert) for that domain. During this pilot period, if there was a discrepancy between the 2 observers, they discussed their assessments and reconciled differences.

Study of TDAT During Inpatient Rounding Encounters

Two physicians (the fellowship director or associate director) with advanced training in educational theory and assessment and the previously trained

TABLE
Teaching Development Assessment Tool Interrater Reliability

	No. Times Rated by Both Observers	% Perfect Agreement	Weighted κ (95% CI)
Promotion of clinical reasoning (global rating)	11	91	0.88 (0.65, 1)
OTB1–Method of teaching clinical reasoning	8	100	
OTB2–Encourages use of problem representation	9	89	
OTB3–Encourages learners to think beyond the clinical diagnosis	9	78	
Control of the inpatient learning environment (global rating)	10	80	0.71 (0.31, 1)
OTB4–Ability to incorporate teaching at the bedside	9	56	
OTB5–Ability to handle high or complex patient census	7	43	
OTB6–Appearance of being stressed or hurried	7	100	
OTB7–Displays confidence in one’s teaching skills	7	86	
Use of learner-centered teaching techniques (global rating)	11	82	0.39 (–0.30, 1)
OTB8–Exhibits learner-centered teaching	9	56	
OTB9–Adapts teaching to the needs of various learning styles	8	75	
OTB10–Completes needs assessments with their learners	3	100	
Ability to teach to multiple levels of learners (global rating)	10	80	0.74 (0.42, 1)
OTB11–Ability to teach to multiple levels of learners	7	86	
OTB12–Ability to engage multiple levels of learners during rounds	9	89	
OTB13–Adapts conversation to include all levels of learners	7	71	
Provision of feedback to learners (global rating)	7	57	0.51 (0.09, 0.93)
OTB14–Provides formative feedback to learners during rounds	7	26	
OTB15–Encourages learners to self-assess their performance	4	75	
Instruction in physical examination techniques (global rating)	2	50	0
OTB16–Finds opportunities to have learners display their physical examination skills	2	50	
OTB17–Models their own physical examination skills for learners	2	50	

Abbreviation: OTB, observable teaching behavior.

nonphysician RA then spent 21½ hours jointly observing 11 inpatient rounding sessions over a 6-month period. During these observations, the 2 observers (1 physician and the RA) were instructed not to discuss their observations or scoring. All attending physicians being observed were informed of the study and provided verbal consent prior to the observation. The selection of the attending physicians was a convenience sample based on the observers’ schedules.

Data Management and Statistical Analysis

Study data were managed with REDCap electronic data capture tools at CCHMC.¹⁵ Percent perfect agreement and interrater reliability using weighted κ value were calculated for each global teaching

domain rating, as well as the percent perfect agreement for each corresponding OTB.

Results
The Creation of an Assessment Tool With Emerging Evidence of Viable Validity

The TDAT (provided as online supplemental material) was reviewed and its content approved by 15 local and 10 national content experts over a 6-month period, conferring evidence of both content and construct validity.^{16,17} For the 17 observations completed during the pilot and study phases, all had scores for at least 4 of the 6 skill domains on the TDAT. Variability in the scores was noted, indicating that not everyone received the same score. The ability of a trained nonphysician to easily observe and score attending physicians on 17 inpatient rounding sessions supports emerging evi-

dence of viable validity. Chen¹⁸ describes viable validity (1 of the 3 components of integrative validity) as the degree to which intervention or assessment tool is practical, affordable, suitable, evaluable, and helpful in the real world. Our experiences in implementing the TDAT have shown that it meets most of those qualities. One issue in the broader implementation and affordability of the tool is that we relied on a paid RA for our study. However, if a nonphysician observer can be used in lieu of a physician, it enhances the feasibility and ability to implement the tool.

TDAT Rater Agreement

The percent perfect agreement and weighted κ for each of the 6 global teaching domain ratings and the 3 levels across the developmental continuum (novice/beginner, competent/proficient, or expert) are displayed in the TABLE. Moderate to near-perfect interrater reliability was seen for the domains of promotion of clinical reasoning, control of the learning environment, ability to teach to multiple levels of learners, and provision of feedback.

Discussion

Our TDAT is promising as a reliable assessment tool with evidence of validity. It holds potential for faculty developers who train clinical teachers and aim to track skill development over time. When compared to similar teaching assessment tools, our tool is novel for the following reasons. First, it provides a discrete framework consisting of OTBs that are easily identifiable and allows tracking of skill development over time. Second, it is an in vivo assessment tool constructed to provide feedback to teachers in a “live” educational setting. Third, a trained nonphysician can feasibly use the tool to observe teaching performance, allowing one to not solely rely on busy faculty members.

Our intervention has several limitations. Our pilot was conducted with a limited number of observations on a small sample of physicians, thus limiting generalizability. We also did not compare our results to other measures of teaching proficiency, such as learner evaluations. Finally, we did not investigate the tool’s ability to detect progression of teaching skills over time. We are currently investigating the TDAT with cohorts of fellows and faculty to determine its ability to track teaching skills over time. While initial data are promising, further observations are needed

to fully identify the reliability between the 2 observers for all teaching domains and OTBs.

Conclusion

This teaching development assessment tool has potential as a valid and reliable assessment tool to assess the development of an individual’s teaching skills along the continuum from early development to mastery, and it is suitable for use by nonphysician observers.

References

1. Bing-You RG, Lee R, Trowbridge RL, Varaklis K, Hafler JP. Commentary: principle-based teaching competencies. *J Grad Med Educ*. 2009;1(1):100–103.
2. Khandelwal S, Bernard AW, Wald DA, Manthey DE, Fisher J, Ankel F, et al. Developing and assessing initiatives designed to improve clinical teaching performance. *Acad Emerg Med*. 2012;19(12):1350–1353.
3. Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. *Med Teach*. 2003;25(2):131–135.
4. Iblher P, Zupanic M, Härtel C, Heinze H, Schmucker P, Fischer MR. The Questionnaire “SFDP26-German”: a reliable tool for evaluation of clinical teaching? *GMS Z Med Ausbild*. 2011;28(2):Doc30.
5. Mookherjee S, Monash B, Wentworth KL, Sharpe BA. Faculty development for hospitalists: structured peer observation of teaching. *J Hosp Med*. 2014;9(4):244–250.
6. Williams BC, Litzelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty’s clinical teaching performance. *Acad Med*. 2002;77(2):177–180.
7. Conigliaro RL, Stratton TD. Assessing the quality of clinical teaching: a preliminary study. *Med Educ*. 2010;44(4):379–386.
8. Julian K, Appelle N, O’Sullivan P, Morrison EH, Wamsley M. The impact of an objective structured teaching evaluation on faculty teaching skills. *Teach Learn Med*. 2012;24(1):3–7.
9. McAndrew M, Eidtson WH, Pierre GC, Gillespie CC. Creating an objective structured teaching examination to evaluate a dental faculty development program. *J Dent Educ*. 2012;76(4):461–471.
10. Wamsley MA, Julian KA, Vener MH, Morrison EH. Using an objective structured teaching evaluation for faculty development. *Med Educ*. 2005;39(11):1160–1161.
11. Ottolini M, Wohlberg R, Lewis K, Greenberg L. Using observed structured teaching exercises (OSTE) to

- enhance hospitalist teaching during family centered rounds. *J Hosp Med.* 2011;6(7):423–427.
12. Klein M, O'Toole JK, McLinden D, DeWitt TG. Training tomorrow's medical education leaders: creating a general pediatric master educator fellowship. *J Pediatr.* 2013;162(3):440–441.e1.
 13. Stanford School of Medicine. Stanford faculty development center for medical teachers. <http://sfdc.stanford.edu>. Accessed June 18, 2015.
 14. Academic Pediatric Association. Academic Pediatric Association's Educational Scholars Program. http://www.academicpeds.org/education/education_scholars_program.cfm. Accessed June 18, 2015.
 15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377–381.
 16. Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ.* 2011;3(2):119–120.
 17. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006;119(2):166.e7–e116.
 18. Chen HT. The bottom-up approach to integrative validity: a new perspective for program evaluation. *Eval Prog Plan.* 2010;33(3):205–214.



All authors are at the University of Cincinnati College of Medicine and Cincinnati Children's Hospital Medical Center. **Jennifer K. O'Toole, MD, MEd**, is Associate Director, General Pediatric Master Educator Fellowship; **Melissa D. Klein, MD, MEd**, is Director, General Pediatric Master Educator Fellowship; **Daniel McLinden, EdD**, is Associate Director, General Pediatric Master Educator Fellowship; **Heidi Sucharew, PhD**, is Assistant Professor, Biostatistics and Epidemiology; and **Thomas G. DeWitt, MD**, is Associate Chair for Education and a Designated Institution Official.

Funding: The development of this project was supported by a Title 7 faculty development grant from the Health Resources and Services Administration.

Conflict of interest: The authors declare they have no competing interests.

The authors would like to thank the members of the Medical Education Scholarship Team, faculty members, fellows, and chief residents at Cincinnati Children's Hospital who participated in the development of this teaching development assessment tool (TDAT). The authors would also like to thank the scholars and faculty of cohort 6 of the Academic Pediatric Association's Educational Scholars Program who provided feedback on the TDAT during the development and revision process.

Corresponding author: Jennifer K. O'Toole, MD, MEd, Cincinnati Children's Hospital Medical Center, MLC 5018, 3333 Burnet Avenue, Cincinnati, OH 45229, 513.349.1488, jennifer.otoole@cchmc.org

Received August 1, 2014; revisions received January 5, 2015, and May 13, 2015; accepted May 31, 2015.