

Using an Accountability Program to Improve Psychiatry Resident Scores on In-Service Examinations

Brandon T. Ferrell, PhD
William E. Tankersley, MD
Clayton D. Morris, MD

ABSTRACT

Background The Psychiatry Resident-In-Training Examination (PRITE) is a standardized examination that measures residents' educational progress during residency training. It also serves as a moderate-to-strong predictor of later performance on the board certification examination.

Objective This study evaluated the effectiveness of an accountability program used by a public psychiatric hospital to increase its residents' PRITE scores.

Methods A series of consequences and incentives were developed based on levels of PRITE performance. Poor performance resulted in consequences, including additional academic assignments. Higher performance led to residents earning external moonlighting privileges. Standardized PRITE scores for all residents ($N = 67$) over a 10-year period were collected and analyzed. The PRITE examination consists of 2 subscales—psychiatry and neurology. Change in the overall level of PRITE scores following the implementation of the accountability program was estimated using a discontinuous growth curve model for each subscale.

Results Standardized scores on the psychiatry subscale were 51.09 points, approximately 0.50 SD change, which was higher after the accountability program was implemented. Standardized scores on the neurology subscale did not change.

Conclusions An accountability program that assigns consequences based on examination performance may be moderately successful in improving scores on the psychiatry subscale scores of the PRITE. This likely has longer-term benefits for residents due to the relationship between PRITE and board certification examination performance.

Introduction

In-training examinations are used by most residency programs in the United States to measure residents' educational progress. These examinations are moderately predicted by trainees' performance on the United States Medical Licensing Examination (USMLE).¹⁻³ More important, performance on the in-training examination is a moderate-to-strong predictor of performance and pass rates on the American Board of Medical Specialties member board examinations,⁴⁻⁹ giving residency program directors motivation to improve their residents' in-training examination (ITE) scores.

The majority of research on interventions to improve ITE performance has focused on residency programs' academic elements. The most frequently used strategy is adding a course to prepare trainees for the in-training or board-certifying examination. These courses have been shown to improve ITE scores,¹⁰⁻¹⁵ although not universally, particularly

with shorter courses.^{16,17} Peer-led courses or study groups have been shown to work in neurology,¹⁸ but have had mixed results in a psychiatry^{19,20} program. Academic half-days improved ITE scores in family medicine²¹ and internal medicine^{22,23} programs, but not in psychiatry.²⁴ Duty hour restrictions have had no effect in either emergency medicine²⁵ or psychiatry²⁴ programs.

We could find little research in nonsurgical specialties regarding the effect of a consequences-based accountability program to improve ITE performance. In surgery, where ITE scores are used in promotion and fellowship decisions,²⁶ 3 studies have assessed the impact of a mandatory remediation program,²⁷⁻²⁹ with 1 showing statistically significant improvement.²⁹ One study in family medicine involved mandatory ITE remediation but showed no statistically significant score improvement.¹⁷

The objective of this study was to assess the effectiveness of a tiered, consequences-based accountability program in improving psychiatry residents' ITE scores.

DOI: <http://dx.doi.org/10.4300/JGME-D-14-00722.1>

Methods

Setting and Participants

Griffin Memorial Hospital is a public psychiatric hospital in central Oklahoma, with sister facilities for child and adult outpatient services and child inpatient services located on the same extended campus. We examined all psychiatry residents from Griffin Memorial Hospital's residency program between 2004 and 2013.

Educational Intervention

Between 2009 and 2011, residency program faculty began restructuring the program's didactics by (1) recruiting more board-certified faculty, (2) limiting rotation sites to those with academic faculty, (3) revising all course syllabi to include more structured teaching, (4) adding required scholarly projects, and (5) increasing residency responsibilities for teaching medical students. The restructuring produced no improvement in performance on the Psychiatry Resident-in-Training Examination (PRITE) scores, an annual 300-item test, with items divided unevenly between 2 subscales—psychiatry and neurology.³⁰ In response to this, the first iteration of the accountability program was put into place for the 2012 PRITE. Residents with scores at or below the 25th percentile were assigned 1 hour of study hall Friday morning before their usual duties, and were required to retake the examination.

For the 2013 PRITE, the second iteration of the accountability program was put into place. This version involved tiered consequences. All residents scoring below the 10th percentile were required to retake the examination and to have regular mandatory meetings with a mentor who assigned practice problems for a structured study hall. Residents scoring below the 30th percentile were required to retake the examination. To earn external moonlighting privileges, residents had to score above the 50th percentile or retake the examination if they did not. These consequences were applied cumulatively, such that a resident scoring at the fifth percentile, for example, would be subject to all consequences.

This study was approved by the Oklahoma Department of Mental Health and Substances Abuse Services Institutional Review Board.

Outcome Measures

The effectiveness of the accountability program was determined by residents' scores on the PRITE. Every resident received 4 scores for each subscale: (1) a raw

What was known and gap

Programs are interested in improving residents' in-training examination performance.

What is new

A multitier accountability program consisting of required retaking of the test, academic assignments, and rewarding high performers with external moonlighting privileges.

Limitations

Single site, single specialty study reduces generalizability; lack of a comparison group; inability to rule out alternative reasons for score improvement.

Bottom line

The accountability program was associated with improved resident performance on the in-training examination.

score, reflecting the number of items answered correctly; (2) a score standardized on a distribution with mean and SD of 500 and 100, respectively; (3) percentile ranks for comparison to all other residents in the same training year; and (4) percentile ranks for comparison to all other residents. Using data from 2008 examinations, the internal consistency coefficients for the subscales were 0.90 for psychiatry and 0.61 for neurology.³⁰

Analyses

Primary analysis of the PRITE data was conducted using a discontinuous growth curve model,³¹ in which score levels were allowed to vary before and after the accountability program was implemented. This model allowed for the separate estimation of the effect resulting from within-resident change over time and the effect from the accountability program. In a secondary model, 2 parameters were added to the model to test for differences between US and international medical graduates at baseline and after the accountability program was implemented. Parameters were estimated using full maximum likelihood estimation in the "nlme" package³² in R version 3.0.3 (The R Foundation).

Secondary analyses included calculating 95% CIs for annual mean PRITE scores between 2011 and 2013. To assess different levels of resident aptitude, USMLE Step 1 or Comprehensive Osteopathic Medical Licensing Examination (COMLEX-USA) Level 1 scores were used to compare residents entering the program before or during 2010 and after 2010 using an independent *t* test. Each resident's score was standardized against the minimum passing score for that year's examination.

TABLE 1
Growth Curve Analysis: Psychiatry Subscale Scores

Parameter	Estimate	P value
Intercept	350.22	< .001
Annual rate of change	52.23	< .001
Accountability program	51.09	.004

Results

The study encompasses data from 67 residents with a mean age of 38 years old. Of these, 32 residents (48%) were women, and 49 (73%) were international medical graduates. The USMLE Step 1 or COMLEX-USA Level 1 scores showed that the scores of residents admitted to the program during or before 2010 were not statistically significantly different from those of residents admitted after 2010 ($t[56] = 0.92$; $P = .36$; $d = 0.26$).

The number of PRITE scores per resident varied, ranging from 1 to 4 scores per resident, with a mean and median of 2.8 and 3 scores per resident, respectively. Fifteen residents (22%) had at least 1 PRITE score before and after the implementation of the accountability program.

Results from the primary analysis of the standardized psychiatry subscale scores are found in TABLE 1. The mean standardized psychiatry subscale score for first-year residents was 350.2 points, approximately 0.75 SDs below the national first-year average, and increased at a mean rate of 52.3 standardized points per year, approximately 0.50 SD annually. The accountability program was associated with an increase in psychiatry subscale scores by 51.1 standardized points. In terms of effect size, this is approximately a 0.50 SD change, a Cohen's d of 0.51, and the impact of an additional year of residency training. US medical graduates showed no difference in scores before and after the accountability program, compared to international medical graduates ($P = .06$).

We also compared the 95% CIs for the standardized psychiatry subscale scores for the years between 2011 and 2013. These years represent the period before any accountability program (2011), the first iteration of the accountability program (2012), and the second iteration of the program (2013). As shown in TABLE 2, the CIs for 2011 (mean = 436.5) and 2012 (mean = 436.2) overlap entirely, and the mean CIs for 2011 and 2013 (mean = 537.5) are almost entirely different. The mean CIs for 2012 and 2013 are entirely distinct.

TABLE 2
Means and CIs for Standardized Psychiatry Subscale Scores

Year	Mean	95% CI LB	95% CI UB
2011	436.53	370.73	502.32
2012	436.24	391.62	480.86
2013	537.53	502.23	572.82

Abbreviations: LB, lower bound; UB, upper bound.

To determine if specific subgroups of residents differed in their improvement on the psychiatry subscale, we examined the 15 residents who had taken the PRITE at least once before and after the accountability program was implemented. There were no differences seen in these residents based on their first-year PRITE percentile scores or their postgraduate year status when the accountability program was implemented.

The mean standardized neurology subscale score for first-year residents was 459.4 points, approximately 0.10 SD below the mean for first-year residents, and increased at a mean rate of 26.9 standardized points annually. The effect of the accountability program was an increase in scores of 6.47 ($P = .80$), suggesting that the accountability program was not associated with an increase in performance on the neurology subscale.

Discussion

Our results suggest that an accountability program improves ITE performance. The program may improve PRITE psychiatry subscale scores at a level comparable to that of an additional year of residency training. No effect was seen for the neurology subscale, potentially because of that subscale's lower reliability,³³ the residents' initially higher scores on that subscale, or another factor.

The accountability program was associated with improved PRITE scores, similar to the results for other interventions^{15,19} in psychiatry programs. However, our study more closely maps onto a mandatory ITE remediation program in surgery.²⁹ Both studies involved more independent work on the part of residents, and had improvements in ITE performance. A potential advantage of the accountability program was that it permitted residents to study for the PRITE how, and for as long as, they chose. As such, different residents reported using different strategies shown to be successful in the literature, including reviewing questions from old examinations,¹⁵ studying with peers,¹⁹ or focusing on practice questions.²⁹ A variable strategy approach might perform better than a more singular and

homogeneous intervention; however, future research is needed to explicitly test this.

Our study has limitations, including that it lacked a control group, which makes it difficult to rule out alternative causes for the observed effect. In addition, while the selection committee might have attempted to recruit better students, leading to a group of residents with higher PRITE scores in the absence of an accountability program, comparisons between older and newer residents found no differences between the groups in terms of USMLE Step 1 or COMLEX-USA Level 1 scores. Our study also involved only 1 residency program at 1 location, and the results may not necessarily be generalizable to other programs.

We will continue to use the accountability as it (1) has been associated with an increase in psychiatry ITE scores, (2) has been well received by faculty and most residents, and (3) has required almost no institutional resources. Resident scores have improved to the point that no residents have required mentoring. Future research should assess if the improvements in PRITE scores translate into improved clinical performance and improved performance on the board certification examination. This should take into consideration the relative effectiveness of different preparation methods among the residents.

Conclusion

After institution of an accountability program, scores on the PRITE's psychiatry subscale were 0.50 SD higher than before. The improvement on the psychiatry subscale associated with the accountability program was the equivalent of an extra year of residency training. Though low scorers showed more improvement in scores, an effect was also seen in low and high average scorers as well.

References

1. Gaiser R. Subtest scores from the in-training examinations: an evaluation tool for an obstetric-anesthesia rotation. *J Grad Med Educ*. 2010;2(2):246–249.
2. Mirkes C, Myers JD, Song J, Cable C, McNeal TM, Colbert CY. Examining the relationship between internal medicine resident moonlighting and IM-ITE performance. *Am J Med*. 2014;127(2):163–167.
3. Miller BJ, Sexson S, Shevitz S, Peebles D, Van Sant S, McCall WV. US medical licensing exam scores and performance on the psychiatry resident-in-training examination. *Acad Psychiatry*. 2014;38(5):627–631.
4. Levy D, Dvorkin R, Schwartz A, Zimmerman S, Li F. Correlation of the emergency medicine resident in-service examination with the American Osteopathic Board of Emergency Medicine part I. *West J Emerg Med*. 2014;15(1):45–50.
5. Leigh TM, Johnson TP, Pisacano NJ. Predictive validity of the American Board of Family Practice In-Training Examination. *Acad Med*. 1990;65(7):454–457.
6. Kay C, Jackson JL, Frank M. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE Step 1 examination. *Acad Med*. 2015;90(1):100–104.
7. Juul D, Flynn FG, Gutmann L, Pascuzzi RM, Webb L, Massey JM, et al. Association between performance on neurology in-training and certification examinations. *Neurology*. 2013;80(2):206–209.
8. 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.
9. Juul D, Schneidman BS, Sexson SB, Fernandez F, Beresin EV, Ebert MH, et al. Relationship between resident-in-training examination in psychiatry and subsequent certification examination performances. *Acad Psychiatry*. 2009;33(5):404–406.
10. Chang TP, Pham PK, Sobolewski B, Doughty CB, Jamal N, Kwan KY, et al. Pediatric emergency medicine asynchronous e-learning: a multicenter randomized controlled Solomon four-group study. *Acad Emerg Med*. 2014;21(8):912–919.
11. Millstein LS, Charnaya O, Hart J, Habicht R, Giudice E, Custer J, et al. Implementation of a monitored educational curriculum and impact on pediatrics resident in-training examination scores. *J Grad Med Educ*. 2014;6(2):377–378.
12. Sharma R, Sperling JD, Greenwald PW, Carter WA. A novel comprehensive in-training examination course can improve residency-wide scores. *J Grad Med Educ*. 2012;4(3):378–380.
13. Gillen JP. Structured emergency medicine board review and resident in-service examination scores. *Acad Emerg Med*. 1997;4(7):715–717.
14. Mathis BR, Warm EJ, Schauer DP, Holmboe E, Rouan GW. A multiple choice testing program coupled with a year-long elective experience is associated with improved performance on the internal medicine in-training examination. *J Gen Intern Med*. 2011;26(11):1253–1257.
15. Hettinger A, Spurgeon J, El-Mallakh R, Fitzgerald B. Using Audience Response System technology and PRITE questions to improve psychiatric residents' medical knowledge. *Acad Psychiatry*. 2014;38(2):205–208.

16. Cheng D. Board review course effect on resident in-training examination. *Int J Emerg Med*. 2008;1(4):327–329.
17. Shokar GS. The effects of an educational intervention for the “at-risk” residents to improve their scores on the in-training exam. *Fam Med*. 2003;35(6):414–417.
18. Schuh L, Burdette DE, Schultz L, Silver B. Two prospective educational interventions in a neurology residency: effect on RITE performance. *Neurologist*. 2007;13(2):79–82.
19. Mariano MT, Mathew N, Del Regno P, Pristach CA. Improving residents’ performance on the PRITE: is there a role for peer-assisted learning? *Acad Psychiatry*. 2013;37(5):342–344.
20. Vautrot VJ, Festin FE, Bauer MS. The feasibility and effectiveness of a pilot resident-organized and -led knowledge base review. *Acad Psychiatry*. 2010;34(4):258–262.
21. Steinweg KK, Cummings DM, Kelly SK. Are some subjects better taught in block rotation? A geriatric experience. *Fam Med*. 2001;33(10):756–761.
22. Ha D, Faulx M, Isada C, Kattan M, Yu C, Olender J, et al. Transitioning from a noon conference to an academic half-day curriculum model: effect on medical knowledge acquisition and learning satisfaction. *J Grad Med Educ*. 2014;6(1):93–99.
23. Batalden MK, Warm EJ, Logio LS. Beyond a curricular design of convenience: replacing the noon conference with an academic half day in three internal medicine residency programs. *Acad Med*. 2013;88(5):644–651.
24. Cooke BK, Garvan C, Hobbs JA. Trends in performance on the psychiatry resident-in-training examination (PRITE®): 10 years of data from a single institution. *Acad Psychiatry*. 2013;37(4):261–264.
25. Pepper DJ, Schweinfurth M, Herrin VE. The effect of new duty hours on resident academic performance and adult resuscitation outcomes. *Am J Med*. 2014;127(4):337–342.
26. Kim RH, Tan TW. Interventions that affect resident performance on the American Board of Surgery In-Training Examination: a systematic review. *J Surg Educ*. 2015;7(3):418–429.
27. Kosir MA, Fuller L, Tyburski J, Berant L, Yu M. The Kolb learning cycle in American Board of Surgery In-Training Exam remediation: the Accelerated Clinical Education in Surgery course. *Am J Surg*. 2008;196(5):657–662.
28. Borman KR. Does academic intervention impact ABS qualifying examination results? *Curr Surg*. 2006;63(6):367–372.
29. Harthun NL, Schirmer BD, Sanfey H. Remediation of low ABSITE scores. *Curr Surg*. 2005;62(5):539–542.
30. Prometric. *Summary Report on the October 2008 Administration of the American College of Psychiatrists Psychiatry Resident-In-Training Examination* [examination information insert]. Baltimore, MD: Prometric; 2009.
31. Singer JD, Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford, UK: Oxford University Press; 2003.
32. Pinheiro J, Bates D, DebRoy S, DebRoy S, Sarkar D, EISPACK authors, et al. *Package ‘nlme’: Linear and Nonlinear Mixed Effects Models*. 2014 ed. R package version 3.1-113. <http://cran.r-project.org/web/packages/nlme/nlme.pdf>. Accessed May 11, 2015.
33. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company; 2002.



All authors are in the Residency Department, Griffin Memorial Hospital, Oklahoma Department of Mental Health and Substance Abuse Services. **Brandon T. Ferrell, PhD**, is Research Coordinator, Griffin Memorial Hospital Residency Program; **William E. Tankersley, MD**, is Medical Director, Children’s Recovery Center, and Professor of Psychiatry, Griffin Memorial Hospital Residency Program; and **Clayton D. Morris, MD**, is Program Director and Professor of Psychiatry, Griffin Memorial Hospital Residency Program.

Funding: The authors report no external funding source for this study.

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

Corresponding author: William E. Tankersley, MD, Children’s Recovery Center, 320 12th Avenue NE, Norman, OK 73071, 405.684.7262, wtankersley@odmhsas.org

Received November 11, 2014; revisions received February 9, 2015, and March 12, 2015; accepted March 16, 2015.