Impact of Residency Training Redesign on Residents’ Clinical Knowledge

Elaine Waller, BA; M. Patrice Eiff, MD; Eve Dexter, MS; Jason C.B. Rinaldo, PhD; Miguel Marino, PhD; Roger Garvin, MD; Alan B. Douglass, MD; Robert Phillips, MD, MSPH; Larry A. Green, MD; Patricia A. Carney, PhD

BACKGROUND AND OBJECTIVES: The In-training Examination (ITE) is a frequently used method to evaluate family medicine residents’ clinical knowledge. We compared family medicine ITE scores among residents who trained in the 14 programs that participated in the Preparing the Personal Physician for Practice (P4) Project to national averages over time, and according to educational innovations.

METHODS: The ITE scores of 802 consenting P4 residents who trained in 2007 through 2011 were obtained from the American Board of Family Medicine. The primary analysis involved comparing scores within each academic year (2007 through 2011), according to program year (PGY) for P4 residents to all residents nationally. A secondary analysis compared ITE scores among residents in programs that experimented with length of training and compared scores among residents in programs that offered individualized education options with those that did not.

RESULTS: Release of ITE scores was consented to by 95.5% of residents for this study. Scores of P4 residents were higher compared to national scores in each year. For example, in 2011, the mean P4 score for PGY1 was 401.2, compared to the national average of 386. For PGY2, the mean P4 score was 443.1, compared to the national average of 427, and for PGY3, the mean P4 score was 477.0, compared to the national PGY3 score of 456. Scores of residents in programs that experimented with length of training were similar to those in programs that did not. Scores were also similar between residents in programs with and without individualized education options.

CONCLUSIONS: Family medicine residency programs undergoing substantial educational changes, including experiments in length of training and individualized education, did not appear to experience a negative effect on resident’s clinical knowledge, as measured by ITE scores. Further research is needed to study the effect of a wide range of residency training innovations on ITE scores over time.

(Fam Med. 2017;49(9):693-8.)
Navy family medicine residents showed that interruptions in training, to complete General Medical Officer tours of duty, did not significantly impact residents’ ITE scores. McDonald et al found that regular curricular conference attendance and self-directed learning activities were both associated with improved ITE scores among internal medicine residents. To our knowledge, no other studies have reported on the impact of residency innovation and redesign on residents’ ITE scores.

The P4 Project was designed to test a significant redesign of family medicine residency training, including changes in the length of training and experiments with individualized education options. One specific area of concern within the programs was that the experimentation would adversely affect residents’ clinical knowledge scores and subsequent success on the ABFM certification examination. In this paper, we explore the extent to which innovations in residency redesign appear to influence residents’ ITE scores. The primary hypothesis we explored was that residents exposed to P4 innovations in training would have clinical knowledge test scores similar to those of residents who were not part of P4. Secondary hypotheses we explored included that neither extended length of training nor individualized training would affect ITE scores.

Methods

Study Setting

The P4 project was a 5-year (2007 through 2012) national demonstration project that included 14 residency training programs, redesigning to improve the graduate medical education of family physicians so they were better prepared to work in the emerging new model of care, the Patient Centered Medical Home (PCMH). Details of the P4 project and characteristics of the programs, including program type, size, and their specific innovations are described elsewhere. Briefly, the participating programs conducted experiments that included changes in the length, structure, content, and location of training, as well as expanded measurements of competency. Each P4 site and the central evaluation team at Oregon Health & Science University (OHSU) underwent IRB review and were granted exemptions, waivers, or approvals by their sponsoring institutions for their educational innovations. Additionally, all IRBs reviewed and approved the consent forms used to obtain residents’ permission to use their ITE scores for analyses of the innovations being tested. Participating residents provided active informed consent for the study activities reported here.

Data Sources and Data Management

The ITE scores for all consenting residents in the 14 P4 programs from 2007 through 2011 were obtained from the ABFM via a secure, password-protected file. Eight hundred and forty residents underwent training during this time period. Thirty eight residents (4.5%), from five of the 14 programs did not provide informed consent and were excluded, leaving data on 802 (95.5%) for analyses. Linkages were created between the ITE scores and residents’ characteristics using unique study identifiers. Residents’ demographic information included age, sex, race, ethnicity, marital status, and parental status. After analyses were completed, the file containing ITE scores was destroyed and analytic files were archived. National data was obtained from the ABFM In-Training Examination Score Results Handbooks (years 2007 through 2011). Residents from P4 programs are included in the national data and represent approximately 3 to 4% of all residents nationally.

To test the association between ITE scores and exposure to individualized education (INDV), we again used the resident cohort who started residency in 2009, and residents were categorized based on whether or not they trained in a program with this innovation. A program was categorized as having individualized education if they had defined curricular tracks or a customized curriculum based on individual resident learning plans (seven P4 programs).

Statistical Analyses

Resident characteristics were compared using an independent samples t-test for continuous variables, and a Pearson chi-squared test for categorical variables. For categorical variables with low cell counts, Fisher’s Exact test was used.

We computed yearly mean ITE scores between 2007 and 2011 according to Program Year (1, 2, and 3) to describe the differences in scores of P4 program residents and scores of residents nationally to examine our primary hypothesis.

Because ITE scoring methods changed between 2007 and 2010, we were unable to perform a pre-P4 vs post-P4 analysis or a trend analysis including all years of study data. Before 2008, standard scores were used to describe resident performance within and across training cohort. Starting in 2008, as a result of a 2004 predictive validity study conducted by the ABFM, scoring was designed to approximate how a resident would do on the certification exam. The scoring changed again in
2009, when ITE performance was placed on the same scale as the certification exam.

For secondary analyses of LOT and INDV innovations, we used linear mixed modeling adjusted for age, gender, race/ethnicity, marital and parental status, as well as if they attended a US medical school, to determine if the adjusted mean scores differed between the innovation subgroups. We used two-tailed tests of significance, and an alpha level of <0.05 was considered statistically significant. Analyses were conducted using R version 3.2.1 and STATA version 12.

### Results

Characteristics of the 802 consenting and the 38 nonconsenting residents are shown in Table 1. The mean age of the consenting residents was 31 years, and 59% were female. Seventy-five percent of consenting residents attended medical school in the United States. Nonconsenting residents were more likely to be older (35.1 vs 30.4; \(P<0.001\)), from a minority racial group (62.5% non-white vs 35.6% white; \(P<0.001\)), to have children (51.4% vs 30%; \(P=0.01\)), and to have attended medical school outside the US (58.6% vs 24.7%; \(P<0.001\)) compared to consenting residents.

Table 2 illustrates our primary analysis of P4 residents’ ITE scores compared to all residents nationally between 2007 and 2011, according to program year. National data were only available in aggregate, thus we are only able to present these results using descriptive statistics. Scores for all residents increased for each year of training. The ITE scores of P4 residents were between 15.2 and 21 points higher than the national scores across years. For example, in 2011, the mean P4 score compared to the national average for PGY1 was 401.2 vs 386; for PGY2, it was 443.1 vs 427; and for PGY3, it was 477 vs 456. Between 2009 and 2011, when

### Table 1: Resident Characteristics According to Consenting Status

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Consenting P4 Residents (n=802)</th>
<th>Non-Consenting P4 Residents (n=38)</th>
<th>(P) value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Age in Years (SD(b))</strong></td>
<td>30.4 (5.2)</td>
<td>35.1 (6.5)</td>
<td>&lt;0.001(c)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>323 (40.8%)</td>
<td>10 (29.4%)</td>
<td>0.252</td>
</tr>
<tr>
<td>Female</td>
<td>469 (59.2%)</td>
<td>24 (70.6%)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td>&lt;0.001(e)</td>
</tr>
<tr>
<td>White</td>
<td>502 (64.4%)</td>
<td>12 (37.5%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>58 (7.4%)</td>
<td>7 (21.9%)</td>
<td></td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>133 (17.1%)</td>
<td>12 (37.5%)</td>
<td></td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>5 (0.6%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Other(d)</td>
<td>81 (10.4%)</td>
<td>1 (3.1%)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td>0.749(e)</td>
</tr>
<tr>
<td>Hispanic Origin</td>
<td>63 (8.0%)</td>
<td>3 (8.8%)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic Origin</td>
<td>723 (92.0%)</td>
<td>31 (91.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td></td>
<td></td>
<td>0.275(e)</td>
</tr>
<tr>
<td>Single (never married)</td>
<td>289 (36.5%)</td>
<td>8 (23.5%)</td>
<td></td>
</tr>
<tr>
<td>Married/Partnered</td>
<td>485 (61.3%)</td>
<td>26 (76.5%)</td>
<td></td>
</tr>
<tr>
<td>Divorced/Separated</td>
<td>17 (2.1%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td><strong>Have Children</strong></td>
<td></td>
<td></td>
<td>0.012</td>
</tr>
<tr>
<td>Yes</td>
<td>237 (30.0%)</td>
<td>18 (51.4%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>554 (70.0%)</td>
<td>17 (48.6%)</td>
<td></td>
</tr>
<tr>
<td><strong>Attended Medical School in the United States</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>569 (75.3%)</td>
<td>12 (41.4%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>187 (24.7%)</td>
<td>17 (58.6%)</td>
<td></td>
</tr>
</tbody>
</table>

\(P\) values from chi-square  
\(SD\) Standard Deviation  
\(t\) value from two-sample t-test  
\(e\) Includes multi-racial residents  
\(f\) P values from fisher's exact test due to low cell counts
ITE scoring was stable, we found that scores decreased over time within program year for PGY1 and PGY2 among the P4 residents, though they did remain consistently higher than national scores. A similar decrease in national scores occurred between 2009 and 2011.

Table 3 illustrates ITE covariate adjusted mean scores by program year between 2009 and 2011, according to LOT innovation. We found no significant differences among residents trained in programs with this innovation compared to those who were not. Table 4 shows ITE covariate adjusted mean scores by program year between 2009 and 2011, according to INDV innovations. Our analysis revealed no significant differences between the groups.

Discussion
To our knowledge, this is the first study on such a large scale to examine the effects of implementing substantial innovations in residency training on performance on clinical knowledge tests. The family medicine residency programs in this project undergoing substantial educational changes did not experience a negative effect on residents’ clinical knowledge, as measured by ITE scores. Because of the predictive validity of the ABFM’s ITE, the results from our study suggest that outcomes of the certification examination would similarly not be adversely impacted. This is encouraging news for residencies redesigning for the future to produce more highly skilled family physicians, indicating that disruptive training innovations need not disrupt the level of clinical knowledge of their residents. Further research is needed to study the effect of a wide range of residency training innovations on ITE scores over time.

Our finding that ITE scores of P4 residents were higher for each resident group compared to national norms in each year could be
attributed to the substantial selection bias in the P4 Project. Selected programs were in good standing with the ACGME and had a higher percentage of US medical school graduates compared to national averages (75% vs < 50%) during the same years which may have affected the programs’ overall scores. Though ITE scores of P4 residents dropped between 2009 and 2010, this mirrors a similar drop in the national data during this time period. Rather than residency factors, it is likely that alterations in the scoring method that occurred three times between 2007 and 2010, influenced this drop in scores.

Although there was some variation in ITE scores in the LOT innovation analysis, none of the differences were statistically significant. These findings are important, as the length of training in family medicine has been undergoing recent scrutiny and a pilot study is currently underway. The Length of Training Pilot case control study design is much more robust than the case series design used in P4, and thus results yielded by studies of that design will be more meaningful in answering this important question.

A potential risk of individualized education is that it could lead to a narrower educational experience, and therefore affect broad clinical knowledge. Our finding that residents exposed to it as part of the P4 innovations scored similarly to their peers is useful to programs undertaking education approaches based on more individualized learning plans. Previously reported P4 findings revealed that individualized education innovations improved residency program match results, suggesting that such innovations attract strong candidates.

Strengths of our study include a group of residency programs diverse in size, setting, and geographic region. Another strength is the size of our resident sample and the high percentage of P4 residents who consented to share their scores for the purpose of this study. Although non-consenting residents were not incorporated into the analyses, they represent only a small proportion of the population.

Limitations include the likely bias associated with this select group of residencies and the residents choosing such programs. It is possible that residents who chose P4 programs because of their innovations are different than other residents in some way, and those unknown variables could contribute to the higher ITE scores compared to the national scores. For example, 75% of all consenting P4 residents were US medical school graduates, compared to...
less than 50% nationally during the same years. Additionally, the P4 project represented a case series design that was not powered to fully test any specific hypothesis; thus our approach was necessarily exploratory in nature. Because there were only three study years (2009 through 2011) in which the scoring of the ITE was the same, we were limited in our ability to do a longitudinal analysis on all cohorts. Finally, the variability of length of training models and the varied approaches to individualized education in the P4 residencies, in addition to the fact that those innovations were not necessarily mandatory for all residents, may have affected our analysis of the influence of these specific innovations on ITE scores.

In conclusion, substantial changes in family medicine residency training during the P4 study, including experiments in length of training and individualized education, did not appear to adversely affect clinical knowledge as measured by ITE scores. These findings should encourage program directors to continue to experiment with new training innovations while monitoring ITE scores over time.

ACKNOWLEDGEMENTS: The authors thank Steele Valenzuela for additional statistical assistance.

Financial Support: The Preparing the Personal Physician for Practice (P4) Project, is jointly sponsored by the American Board of Family Medicine Foundation and the Association of Family Medicine Residency Directors.

CORRESPONDING AUTHOR: Address correspondence to Elaine Waller, research associate, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, FM, Portland, OR 97239, 503-494-3397. Fax: 503-494-4496. waller@ohsu.edu.

References