Main and Regional Campus Assessments of Applicants to a Rural Physician Leadership Program: A Generalizability Analysis

Terry Stratton, Clarence Kreiter, PhD, Carol Elam, EdD

DOI: https://doi.org/10.24926/jrmc.v2i1.1981


z.umn.edu/JRMC

All work in JRMC is licensed under CC BY-NC
Main and Regional Campus Assessments of Applicants to a Rural Physician Leadership Program: A Generalizability Analysis

Terry Stratton, Clarence Kreiter, PhD, Carol Elam, EdD

Introduction
At its most basic level, the selection of qualified candidates remains a defining characteristic of any profession. In medicine, emphasis has shifted in terms of the desired qualities of future physicians and the means to assess those qualities. For example, holistic approaches advocate consideration of multiple factors other than academic performance, including an individual’s “fit” with a medical program’s social mission. Yet, underlying most selection criteria or methods are individual ratings or judgments of some personal quality, aptitude, or behavior is information limited by the reliability of the ratings and the representativeness of the encounter.

For programs with specific foci, such as those accepting applicants directly into rural paths, tracks, or concentrations, the admission process may accommodate supplemental values, interests, or stakeholder perspectives. As such, training targeted to certain practice locales (e.g. rural or underserved), medical specialties (e.g. primary care), or career interests (e.g. physician scientists) may demand an expanded approach to selecting qualified applicants for a specific programmatic fit. A changing dimension of the admission process, and the academic qualifications, personal qualities, and/or demographic considerations it entails, is the increasing presence of regional medical campuses (RMCs). RMCs are defined as “campuses of medical schools at which a portion of pre-clinical or clinical education of medical students occurs”, and play a significant role in calls for increased enrollment. RMCs are classified into 4 models (basic science, clinical, longitudinal, and combined) based on curricular years taught and/or type of training provided and can target a specific mission, demographic, specialty focus, and/or delivery model (e.g. community-based care). While RMCs may provide all aspects of medical training, they are usually considered extensions of the main or “parent” program with selection decisions made by a single, overarching admission committee.

Located in the southeastern United States (US), the Rural Physician Leadership Program (RPLP) was created in 2008 at the University of Kentucky College of Medicine (UKCOM) to attract and train applicants interested in practicing rural medicine, ideally in the state of Kentucky. Located in a city of approximately 320,000, students’ pre-clerkship training (years M1-M2 in our program) occurs at the main urban campus, while their clinical instruction and leadership training (years M3-M4) is completed at a smaller, rural (population ~7,500) medical campus about one hour away. Ten students are admitted annually, with preference given to applicants with rural backgrounds, interests, or experiences. Like their main campus counterparts, RPLP students are free to pursue any medical specialty.

RPLP Admissions Process
While myriad factors underlie the selection of applicants to medical school, the challenge for more focused programs like the RPLP is twofold: To gauge preparedness for and fit within the medical profession, and to discern interest in rural medical practice. With final admission decisions made by a single committee, regional input into these assessments was deemed essential. To compliment applicants’ written responses to items contained on our secondary form, semi-structured, face-to-face interviews are conducted with applicants meeting certain academic standards. Interviews typically last between 30-45 minutes and are conducted by a wide range of individuals including active and retired faculty, administrative staff, community members, and current medical students and residents. Interviewer assignment is not systematic, though specific individuals may be paired with applicants with similar personal (e.g. geographic area) or professional (e.g. medical specialty or research area) backgrounds. Standardized training includes a review of program missions, the admission process (including instrumentation), and interview protocol. Over 2 consecutive days, RPLP applicants complete interviews at both main and regional medical campuses. At each site, 2 interviewers with access to standardized applicant data (e.g. prior academic performance, standardized test scores, demographic characteristics, residency status, relevant activities/experiences, and letters of evaluation) independently offer subjective, narrative assessments of applicants’ backgrounds and qualifications as well as a global (overall) assessment. This assessment is assigned a 1-7 numeric rating ranging from “unacceptable” to “outstanding, clearly superior”. Using a scale from 0 (“no chance”) to 100 (“absolute certainty”), interviewers at both sites are also
Method

The study protocol received institutional review board approval to use preexisting data from 232 RPLP applicants who, from 2009-2017, completed admission interviews. From the 22 re-applicants who interviewed in multiple years, initial rating data were used. Interviews were granted via an internal screening process which included a holistic review of each applicant but tended to emphasize measures of cognitive ability and geographical background of applicants. Of those RPLP applicants interviewed during the 9-year study period, 90 were accepted for admission and subsequently enrolled in the program.

Results

Descriptive Statistics

Complete sets of ratings (2 x 2 = 4) for overall acceptability and likelihood of rural in-state practice were available for 211 and 174 RPLP applicants, respectively, due primarily to main campus interviewers’ failure to consistently rate the latter. Applicants with incomplete rating data did not differ significantly by race, gender, or geographic origin. The
proportion of cases excluded due to missing data varied by year, but followed no discernible pattern. Since these represented key variables of interest, and because sample sizes remained sufficient, cases with missing data were excluded rather than substituted with imputed values. Mean ratings of applicants’ overall acceptability were 5.27 (median = 5.50, SD = 0.98) and 5.35 (median = 5.50, SD = 1.20) for main and regional campus interviewers, respectively. For the likelihood of practicing in rural Kentucky, mean ratings were 77.7% (median = 87.5%, SD = 20.4%) and 78.5% (median = 85.0%, SD = 19.8%) for main and regional campus interviewers, respectively. Ratings of applicant acceptability and likelihood of rural in-state practice were not significantly correlated for main (r = 0.04, p = .61) or regional (r = 0.09, p = .25) campus interviewers.

**Univariate G Study Results**

Table 1 displays univariate G study results [r : (p x c)] of overall applicant rating and likelihood of rural in-state practice. As shown, the percentages of “true score” variance associated with the object of measurement (p) were 36% and 51%, respectively. Compared to likelihood of rural practice, the rating in variances attributable to the person-by-rater (pr) and person-by-campus (pc) interaction was notably larger for overall applicant rating (53% versus 44% and 11% versus 5%, respectively).

### Table 1. Univariate Mixed Model G Study Results [r : (p x c)]

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>Mean Square</th>
<th>Variance Component</th>
<th>Standard Error</th>
<th>% Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons (p)</td>
<td>210</td>
<td>3.99063</td>
<td>0.65440</td>
<td>0.08177</td>
<td>36</td>
</tr>
<tr>
<td>Campus (c)</td>
<td>1</td>
<td>0.16669</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0</td>
</tr>
<tr>
<td>Interviewers: pc</td>
<td>422</td>
<td>0.97305</td>
<td>0.97304</td>
<td>0.06834</td>
<td>53</td>
</tr>
<tr>
<td>Overall (p x c)</td>
<td>210</td>
<td>1.36086</td>
<td>0.19391</td>
<td>0.07406</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>843</td>
<td>1.82135</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Multivariate G Study Results [r : (p x c)]

Multivariate G Study Results

Table 2 displays the multivariate G study results in a matrix format, with VCs for applicant acceptability ratings and likelihood of rural practice reported by campus on the diagonals of each of the 2 matrices (p, r). As shown in the left matrix, the proportion of variance attributable to systematic differences in applicants (p) is considerably greater for regional campus interviewers (53% of the total variance) than their main campus counterparts (30% of the total variance). This implies that the reliability of a single interview is 0.30 for the main campus and 0.53 for the regional campus. The average reliability of one interview across the 2 campuses is 0.42. The relationship between campuses, reflected in the observed covariance (0.56), is reported in the lower left cell of the p matrix. However, more readily interpretable is the universe score correlation (0.83) shown in the upper right cell – which, again, is synonymous with “true score” correlation in classical test theory. This universe score correlation indicates a strong positive relationship between interviewer ratings of overall applicant acceptability at main and regional campuses, suggesting that raters on the 2 campuses were assessing similar but not identical applicant characteristics.

Results from a corresponding D study for the [r : (p x c)] mixed model with campus fixed are shown in Figure 1. For overall applicant acceptability, one admissions interview conducted on each of the 2 campuses resulted in an observed reliability of 0.57 which increased to 0.73 when doubled to 2 interviews per campus. Ratings of RPLP applicants’ likelihood of practicing medicine in rural Kentucky tended to be more reliable: The reliability of 2 interviews, one on the main campus and one the regional medical campus, was 0.70 and rose to 0.82 when increased to 2 interviews per campus (4 in total). For both measures, the effects on reliability of increasing the number of interviews beyond the present configuration of 2 were modest.

### Figure 1. Univariate Mixed Model D Study Results: [r : (p x c)]
In the right matrix, similar results are presented for RPLP applicants’ likelihood of rural in-state practice. Compared to overall applicant acceptability, the proportion of “true score” variance attributable to applicants (p) is somewhat greater and less disparate for both main and regional campus interviewers (49% and 61% of the total variance, respectively). The universe score correlation (0.91), listed in the upper right cell of the top matrix, indicates that raters on the 2 campuses were assessing very similar aspects in arriving at their judgments regarding applicants’ likelihood of practicing medicine in rural Kentucky. Since raters were nested within persons, it was not possible to disentangle the specific error attributable to the rater-person interaction and systematic rater stringency. Hence, in this study, the \( r : p \) \( VC \) reflected the sum of the interaction and systematic VCs.

### Multivariate D Study Results

Figures 2 and 3 present multivariate D study results, by campus, for each of the measures collected in the RPLP admissions process: Overall applicant acceptability and likelihood of rural in-state practice, respectively. In addition, composite estimates are presented which combine information on interviewer ratings from both main and regional campuses. Estimated reliability is projected for up to 4 interviews per campus and, correspondingly, 8 total interviews per applicant.

In Figure 2, the G coefficients corresponding to our current protocol of 2 independent raters (one per interview) from each campus are 0.46 (main) and 0.69 (regional) which, respectively, would increase incrementally to 0.56/0.77 and 0.63/0.82 with an additional 2-3 interviews per campus. With main and regional campus ratings weighted equally, the composite reliability averaged across 4 interviews (the current configuration) is 0.73 and would increase to 0.84 if doubled to 8 total interviews (4 per campus).

**Figure 2. Multivariate D Study Results: Overall RPLP Applicant Acceptability**

![Image](https://example.com/image.png)

**Figure 3 plots comparable generalizability estimates for applicants’ likelihood of rural in-state practice. Here, ratings are both more reliable and more comparable, with G coefficients for 2, 3, and 4 interviews being 0.66, 0.74, and 0.79 for main campus interviewers and 0.76, 0.82, and 0.86 for their regional counterparts. Composite reliabilities, again weighted equally across campuses, are 0.82 (4 total interviews), 0.87 (6 total interviews), and 0.90 (8 total interviews) – good to excellent for most purposes.**

**Figure 3. Multivariate D Study Results: Likelihood of Rural In-State Practice**

![Image](https://example.com/image.png)

### Discussion

As the growth of regional campuses continues, admission committees may seek to expand or better formalize the roles played by these partner programs. Especially where such programs have uniquely-targeted (e.g. rural) missions, the need may exist to access specific expertise or incorporate local stakeholders in selecting qualified candidates.

Coordinating this process in a logistically and psychometrically optimal fashion requires a robust plan for establishing and monitoring this process.

This study examined admission interview ratings from a regional campus which, via a Rural Physician Leadership Program, provides clerkship training to 10 UK COM students interested in rural medical practice, preferably in rural Kentucky.21 Established in 2008, this “clinical model” RMC is an extension of the “parent” academic medical center located...
about one hour away, which retains ultimate authority for all
enrollment under a single, centralized admissions committee.
However, primary feedback from RPLP stakeholders is
routinely provided through a parallel process in which each
applicant, over a 2-day period, meets independently with 2
interviewers at each campus.
This structure allowed reliability of interviewer assessments
to be examined using a generalizability framework. Weighted
equally and averaged across campuses, combined reliabilities
of both overall RPLP applicant rating (0.73) and likelihood of
rural practice (0.82) were adequate for the purposes at hand.
However, main campus assessments tended to contain more
error variance for both measures, particularly overall
applicant acceptability. While effects of alternative weighting
schemes on overall reliability were not formally examined,
slight improvements appear possible by assigning a greater
contribution to regional campus ratings.
Several possible explanations exist for the observed variation
in main and regional campus reliabilities. First, as previously
mentioned, the regional campus interviewer pool is smaller
and more homogeneous in term of program knowledge and
focus; that is, most interviewers are intimately familiar with
the RPLP history, curriculum, and objectives. In contrast, main
campus interviewers reflect a much broader array of
backgrounds, expertise, and interests reflective of an
academic medical center. Second, while the association
between applicants’ overall acceptability and likelihood of
rural in-state practice varied somewhat by campus, it tended
to be weak: Applicants’ academic qualifications were largely
unrelated to raters’ judgments about their propensity to
eventually practice medicine in rural Kentucky.
Due to confounding factors, the data collection design was
unable to accommodate occasion and rater as separate
random facets. However, the observed results are fairly
consistent with other reports on admission interview
reliability.13,32,33 Further, these findings support the utility of
having RPLP candidates interview at both the regional
campus and main campus. In this study, the observed
interview reliability effectively comprised 4 interviews per
applicant, the likely reason for the reliable mean interview
scores. In addition, results show little benefit to expanding
the RPLP admission process beyond 2 interviews per campus.
While the medical school personal interview (MSPI) remains
part of the admissions process, it is only one source of data
considered in committee assessments of applicant
qualifications and professional/program fit.34 With rare
exceptions,35,36 how this and other information is used in
committee deliberations or weighted in decisions to accept,
reject, or hold applicants has not been widely examined.13
Holistically, MSPI ratings are considered part of applicants’
overall “dossier”, but no algorithm or guidelines standardize
their role or degree of influence. As such, it is unknown
whether their use consistently constitutes a “high stakes”
application, which has obvious implications for the level of
rigor required in their collection and measurement. In a
holistic review of applicants, a low rating of overall
acceptability, based in part on face-to-face interviews, could
prove detrimental to admission.
The study objective was not to establish the superiority of a
single approach to assessing applicant fit, but rather to
empirically explore the apparent tension between the
diversity of stakeholder input and the reproducibility of
resulting scores. Indeed, recent research suggests a hybrid
model containing selected elements from various approaches
might be optimal.33,37 Predicting future events, be it academic
performance, specialty choice, or eventual practice locale,
remains an inexact science37 involving both tangible38 and
intangible39-40 considerations. From prior research on rural
medical practice, considerable attention has been paid to
applicants’ related backgrounds, interests, and experiences.41-
43 Indeed, of the 107 (38.8%) RPLP applicants not invited for
interviews during the study time frame, most lacked
meaningful rural experience and/or sufficient academic
performance.23
Since the overriding goal of the RPLP is to recruit and train
physicians who will practice medicine in rural Kentucky, a
shared understanding of program goals is essential. In the
case of the RPLP, this was explicit – incorporating the major
program outcome (practicing in rural Kentucky) into the
actual interview process. The purpose was not to develop a
precise measure, but rather to help direct focus on the task at
hand. Whether this was effective or caused RPLP interviewers
to cognitively approach the process differently is a question
for future research. The meager correlation of this measure
with overall acceptability suggests interviewers were able to
discern between them.
These findings are limited by several factors. First, this study
is based on a singular rural track training program at one US
institution. As a result, how widely these findings may
generalize beyond this context is unknown. Second, although
all interviewers follow the same semi-structured format,
there is some flexibility in the specific questions that can be
asked. Moreover, regional campus interviewers used a
differently interview narrative form. Lastly, interviewers on
both campuses were guided by “rural qualities” gleaned
inductively via a nominal group process – not a standardized,
demographic definition of “rural”. While by design, this more
qualitative operationalization was likely implicated in
interviewers’ assessments of likelihood of rural practice.
Another issue worth mentioning is the potential disconnect
between the composite reliability estimates averaged across
campus and, in actual practice, the disaggregated use of
interviewer ratings by the admissions committee. That is, the
informal assessment of agreement made by committee
members in their review and comparison of individual
interviewer ratings, some of which, not being anonymous,
may be afforded more credence than others.

Conclusion
Dedicated rural medical tracks or programs have been shown
to be effective strategies in producing primary care physicians

DOI: https://doi.org/10.24926/jrmc.6666

Original Reports

Journal of Regional Medical Campuses, Vol. 1, Issue 6
for practice in rural, often underserved areas, especially when provided in settings (like RMCs, for example) that offer meaningful learning experiences outside the larger, urban environment.44 Key to the success of these efforts is the selection of candidates most qualified to meet programmatic goals. In Kentucky, the RPLP was designed to meet this need by admitting applicants who prefer rural practice and training them in settings with appropriate physician and community role models.

Study results found composite (combined) reliabilities of RPLP applicants’ overall acceptability and likelihood of rural in-state practice to be encouraging. On both measures, however, ratings from regional campus interviewers tended to have less error variation than their main campus counterparts. It is possible that better training and calibration, perhaps combining interviewers from both campuses, might narrow the observed differences in reliability. Mean ratings, it should be noted, did not differ between campuses. Various weighting schemes could, in concert with the number of interviews, be more closely examined as a means of maximizing overall reliability. This study highlights a methodology for developing and monitoring the inclusion of additional stakeholders to the admission process, and may prove useful for programs seeking to strategically tap a wider range of perspectives – especially as they relate to a specific, targeted mission. In the present context, eliciting input from interviewers with complimentary backgrounds resulted in more reliable composite ratings of applicants’ acceptability and likelihood of practicing in rural Kentucky. In addition to estimating the reliabilities of these combined scores, the multivariate approach allowed estimation of the relationship between the 2 groups of interviewers.

While broadening the universe of generalization is typically associated with reduced reliability, in this application, the increase in reliability from additional interviews was found to outweigh these effects. Put another way, the addition of regional stakeholders to the pool of potential interviewers may not necessarily result in a less reliable composite measure. As regional campuses proliferate, the use of multivariate generalizability approaches to examine assessments of applicants or students at multiple “fixed” locales may hold promise.

Future studies should include a focus on the validity of interviewer ratings. That is, whether or not assessments validly reflect applicants’ academic success or eventual likelihood of practicing in rural Kentucky, for example. Although early data appear encouraging, a continued follow-up of RPLP graduates will help determine the accuracy of these long-term projections made during admission into the profession. On a broader scale, studies may wish to explore underlying differences and similarities in determining the psychometric impact of expanding stakeholder input.

References

12. Pau A, Chen YS, Lee VKM, Sow CF, De Alwis R. What does the multiple mini-interview have to offer over


38. Daniels ZM, VanLeit BJ, Skipper BJ, Sanders ML, Rhyne RL. Factors in recruiting and retaining


