INNOVATIONS in pharmacy

Volume 6 | Number 3 Article 206

2015

Comparing Vertical and Horizontal Screening Methods for Pharmacy Resident Candidates Before Interviews

Michael J. Peeters
michael.peeters@utoledo.edu

Todd E. Gundram

Julie A. Murphy

Follow this and additional works at: http://pubs.lib.umn.edu/innovations

Recommended Citation

Peeters MJ, Gundram TE, Murphy JA. Comparing Vertical and Horizontal Screening Methods for Pharmacy Resident Candidates Before Interviews. *Inov Pharm.* 2015;6(3): Article 206. http://pubs.lib.umn.edu/innovations/vol6/iss3/2

INNOVATIONS in pharmacy is published by the University of Minnesota Libraries Publishing.

Comparing Vertical and Horizontal Screening Methods for Pharmacy Resident Candidates Before Interviews

Michael J. Peeters, PharmD, MEd[1]; Todd E. Gundrum, PharmD[2]; Julie A. Murphy, PharmD[1]
[1]University of Toledo College of Pharmacy & Pharmaceutical Sciences, Toledo OH; [2]University of Toledo Medical Center, Toledo OH

The authors have no conflicts of interest to disclose.

Abstract

***Purpose**: Prior studies have examined autobiographical screening methods among medical student applicants, and demonstrated halo bias with single-rater scoring; though others have questioned its practical significance. Comparing with traditional vertical screening method, we evaluated a horizontal method for initial screening of Post-Graduate Year-1 (PGY-1) pharmacy practice resident candidate applications prior to interviews.*

***Methods**: Our screening rubric for PGY-1 pharmacy residency candidates consisted of eight criteria, each scored using a 5-point Likert scale. During the 2014 residency recruitment season, two single-evaluators (A&B) scored all eight criteria and their scores were summed into total application scores (vertical method). Meanwhile two other evaluators (C&D) each evaluated only two criteria for all applications. The four combined-evaluators (A-D) scores, on two criteria each, were summed together into total application scores (horizontal method). For statistical comparison of single-evaluator and combined-evaluators, inter-component reliabilities were analyzed for each evaluator, while inter-rater consistency was also examined. For practical significance, actual selection differences were reviewed.*

***Results**: Forty-six applications were evaluated to determine 24 invitations for on-site interviews. Inter-component reliability differed among evaluatorA, evaluatorB, combined-evaluators A-D (Cronbach's alpha of 0.74, 0.73, 0.58, respectively; lower better). Among raters, inter-rater consistency was excellent (0.86 by intraclass correlation, p<0.001). In practice, single-evaluators and combined-evaluators agreed on 21 interview invitations (88% of 24 invitations), while single-evaluators did not agree on any others (0%). One single-evaluator agreed with combined-evaluators on one further invitation (4%), while all differed on the remaining two (8%) invitations. Combined-raters reported faster, more confident scoring of applications, specific to those criteria evaluated.*

***Conclusion**: Halo bias was seen with the single-evaluators (vertical method); two interview invitations were negatively impacted. For pharmacy resident screening, a horizontal screening method appears to be rigorous in promoting fairness for applicants. As pharmacy residency applications continue to grow, a fair and time-efficient method of screening seems imperative.*

Background

There was a 50% increase in the number of Post-Graduate Year-1 (PGY-1) applicants participating in the American Society of Health-System Pharmacists (ASHP) Resident Matching Program from 2010 to 2015.[1] As the number of pharmacy residency applicants continues to increase, programs will spend more time screening candidates to determine the individuals they choose to invite for an on-site interview. Locally, we also saw dramatic growth in applications (30 in 2012, 41 in 2013, 46 in 2014, and 61 in 2015; a100% increase over four interview cycles). With that increase, we sought a method to divide the work of screening many applications among a small group of evaluators, while maintaining the rigor of our screening process. This initial screening of candidates typically involves review of their written application materials (i.e., curriculum vitae, letter of intent, letters of recommendation, and academic transcripts). The initial screening procedure should be as reliable and fair to candidates as the actual interview process itself.

Traditionally, applicant packets have been screened on a vertical basis where a single evaluator will screen an applicant's entire information packet. Evaluators intuitively want to divide complete applicant packets into separate groups for separate individual evaluator screening. However, this approach introduces rater biases. Because each evaluator may not review all application packets from all applicants, there is issue with inter-rater reliability with one notable rater bias being the halo effect. Halo bias comes from an evaluator that inflates (subconsciously or consciously) the score for an applicant that they perceive as noteworthy;[2-4] an evaluator may take one criteria that is important to them, and overemphasize it by incorporating it into other criteria

Corresponding Author: Michael J. Peeters, PharmD, MEd, FCCP, BCPS; Clinical Senior Lecturer
University of Toledo College of Pharmacy & Pharmaceutical Sciences; 3000 Arlington Ave, MS1013, Toledo, OH 43614
Phone: 419-383-1946; Email: michael.peeters@utoledo.edu

scores as well, or leniently give their benefit-of-the-doubt on other scoring criteria. Halo bias is not systematically applied to all candidates and may unfairly advantage some applicants. To try and overcome this bias, multiple evaluators can be used—though these further evaluators are often added to provide more reviews of entire application packets using this same vertical screening method.

An alternative approach is to have multiple evaluators use a horizontal screening method.[3-5] In this approach, each evaluator reviews one or two screening criteria across all applicants. This method can overcome inter-rater reliability concerns (i.e., stringency, leniency, halo bias) by each evaluator scoring all applicants on the same criteria. If the evaluator is stricter than others, it affects all candidates' scoring similarly, and if the evaluator is lenient, all applicants are treated equally. As well, it would be very unlikely that each evaluator has enough information to produce any halo bias; applications could even be blinded to minimize this potential further. Figure 1 illustrates these vertical and horizontal screening methods.

Three prior studies described use of these traditional vertical screening method and the novel horizontal methods for medical school admissions. Dore et al. found the horizontal screening method to have promising findings with a lower inter-component reliability, higher inter-rater reliability and improved predictive capacity.[3] Hanson et al. confirmed this and explained that using the horizontal method decreased the halo effect compared to the vertical approach when screening medical school candidates.[4] While confirming the theory of this approach, Allalouf et al. suggested there was not a practical difference with doing this, and it was less convenient in some evaluators' opinions.[5]

In an effort to avoid halo effect influences, and make our screening process more consistent, rigorous, and fair, we wanted to compare this novel horizontal method to our traditional vertical method for screening pharmacy residency applicants. We also wanted to see if any practical difference could be ascertained with use in pharmacy residency applications. Would application to pharmacy residency applicant screening show the positive effects already demonstrated in medical school admissions, and how much practical difference does it show? The purpose of this study was to evaluate a horizontal method as compared to a vertical method for initial screening of PGY-1 pharmacy practice resident candidate applications prior to on-site interviews.

Methods

The University of Toledo's Institutional Review Board approved this investigation as exempt. During the 2014 residency recruitment season, we evaluated two methods for initial screening of PGY-1 pharmacy residency applicants, to determine those we would invite for on-site interviews. The two screening methods used were vertical and horizontal scoring (see Figure 1).

For both methods, our evaluators used the same scoring rubric. Our screening rubric consisted of a 5-point rating scale for eight criteria. The criteria by which each of the residency applicants were evaluated were: quality of advanced pharmacy practice experiences, professional organization involvement, professional scholarship, writing skills, career goals consistent with the program strengths, pharmacy work experience, recommendation letters, and grades earned within their Doctor of Pharmacy program.

For application reviews, the traditional *vertical* method used one evaluator to score all criteria for each applicant; we had two evaluators review all criteria for all applications, and summed their criteria scores into an applicant's total score (see vertical screening with evaluator A, evaluator B in Figure 1). The novel *horizontal* method has each evaluator score *all* applicants on only two criteria; multiple evaluators' criteria scores are then summed into each applicant's total score (see horizontal screening evaluators A, B, C and D in Figure 1). Using this method, we had two additional evaluators (evaluator C and evaluator D) score all applicants on only two criteria; their two scores were combined with two criteria scores taken from each of evaluator A and evaluator B. Thus with the horizontal method, the multiple evaluators combined to score all eight criteria together (combined-evaluators A-D) with each evaluator scoring the *same* criteria for *all* applicants.

Vertical and horizontal screening processes for screening for residency candidates were compared statistically using two indices. First, *inter-component reliability* (a coefficient for the association among the different screening criteria) was analyzed using Cronbach's alpha. While Cronbach's alpha is commonly used elsewhere for internal consistency, its application was slightly different herein; because it is calculated from inter-item correlations, it would be higher for items that are asking a similar concept. However, screening criteria should be different characteristics from one another. Evaluators should not rate similar candidate characteristics as

this would not help distinguish strong candidates from weaker ones.[1] Additionally for horizontal screening, a

Generalizability theory model[6] was used to better describe inter-component reliability among the four *independent* evaluators, using *G String IV* (McMaster University, Hamilton, ON). Using only a Cronbach's alpha to determine inter-component reliability would assume that only one evaluator scored all criteria, instead of four independent raters each scoring two criteria; using Generalizability theory will add uncertainty (e.g., a lower reliability coefficient) to better describe four independent evaluators as different from a single evaluator. Second, as a comparison of each criterion score for all applicants among the raters, *inter-rater consistency* was compared for single-evaluator and multiple-evaluator data, using a 3,1-type of intraclass correlation (ICC). Both Dore[3] and Hanson[4] used these indices in their prior investigations of medical school applicants. The *a priori* significance was set as 0.05. *SPSS version 19 for Mac* (IBM, Armonk, NY) was used for inter-component reliability and inter-rater consistency analyses. To determine practical significance[7] in our circumstance, the ranked candidate lists from both vertical evaluators were compared with the rank list from horizontal scoring results. Absolute inter-rater agreements of high (all 3 evaluators), medium (2 of the 3 evaluators), and low (no evaluators agree) were used.

Results

Forty-six applications were evaluated to determine 24 invitations for on-site interviews. Using Cronbach's alpha, inter-component reliabilities differed. Evaluator A alone was 0.74 (vertical screening), evaluator B alone was 0.73 (vertical screening), while combined-evaluators A-D together were 0.59 (horizontal screening). Furthermore, using Generalizability theory to correct combined-evaluators A-D because they were independent from one another, inter-component reliability was slightly lower at 0.58.

Meanwhile, at 0.86 the overall inter-rater consistency was excellent among all evaluators (evaluator A, evaluator B, and combined-evaluators A-D; $p<0.001$). On pairwise comparisons, evaluator A and evaluator B were 0.83 ($p<0.001$), evaluator A and combined-evaluators A-D were 0.90 ($p<0.001$), while evaluator B and combined-evaluators A-D were 0.84 ($p<0.001$). The ICC was lowest between single-evaluators, and so least agreement was between the single-evaluators (vertical screening).

Practically speaking, high-agreement was 88%, where evaluator A, evaluator B, and combined-evaluators A-D all agreed on most (21 of 24) interview invitations. Evaluator A and evaluator B did not agree with one another on any other invitations (0%). Medium agreement was 4%, where one single-evaluator agreed with combined-evaluators on one further invitation. Low agreement was 8%, where all evaluators differed on the remaining two invitations; these remaining two invitations required supplementary discussion of candidates by all evaluators.

Discussion

This novel approach to pharmacy residency applicant screening is a pharmacy-specific investigation confirming prior evidence from medical school admissions[3,4] as well as a best-practice in educational grading.[8,9] Herein, using the horizontal scoring method for initial screening of PGY-1 pharmacy practice resident candidate applications prior to interviews appears to provide excellent inter-rater consistency among evaluators. As mentioned previously, a further study in medical school admissions had questioned the practical significance of this horizontal approach over a conventional approach;[5] investigating pharmacy residency admissions, we found a 8% practical difference in candidates receiving interview invitations. Considering the potential career impact to graduates with obtaining and completing a residency to further their development towards becoming a clinical pharmacy specialist, this percentage should not be understood as inconsequential. This screening should be seen as a high-stakes evaluation needing thorough evaluation. And so inter-rater consistency is critical, and foundational for process rigor.

Building on its sound consistency herein, the horizontal method's inter-component reliability was lower than using the vertical method. Screening criteria items correlated less with one another among combined-evaluators using the horizontal method than among single-evaluators using the vertical screening method. The higher inter-component reliability with our vertical screening suggests a halo bias, which agrees with prior studies in medical school admissions.[3-5]

[1]For later interpretation: The higher this number *inter-component reliability* is, the more similarity among screening criteria. At some point (>0.9-0.95 possibly), redundancy is suggested; time is being wasted on scoring redundant screening criteria that are not helping statistical discrimination among candidates. Instead, if we approach the screening criteria mentioned above as essential and all as different, than the inter-component reliability should be *lower*. Having a high inter-component reliability can suggest 'halo bias' as raters may be taking one criteria that is important to them and conflate it with scoring of other screening criteria based on their perspective. Halo bias is not systematically applied to all candidates, but unfairly advantages only some.

An analogy of this method of evaluation would be grading student responses on long answer questions from examinations. In this didactic academic setting, educators would do best to grade each question for all students (i.e., horizontal grading), and *not* grade all questions for each student before moving on to the next student's responses (i.e., vertical grading);[8,9] this horizontal grading should minimize intra-rater (and also lower inter-rater) variation.[8] Additionally, this method should be more time-efficient for that instructor.[9] In our current study of resident screening, we observed these advantages as well.

In *post-hoc* discussions both evaluators C and D reported fast and confident scoring of screening criteria using the horizontal method. However, they also noted that confidence was specific to those criteria evaluated. While evaluators' confidence was specific to the scoring of those criteria evaluated, their confidence in understanding an *entire* applicant for later interviews was not as helpful with horizontal scoring as with vertical scoring. Similar to past findings by Allalouf,[5] we also found reviewers mixed on whether they preferred horizontal or vertical screening; evaluators C & D differed in their preference. Scoring with this horizontal method appears to avoid rater biases and be more time-efficient immediately, but appears to come at a cost of needing to re-review interviewees in greater depth prior to interviews. This study was also limited by the single residency program and single interview cycle analyzed, though this study, using Boyer's scholarship of integration,[10] was a pharmacy-specific application of a wider practice with evidence in medical education, as well as being analogous to a best-practice in classroom grading. It is important to reiterate that if high-stakes decisions are poorly reliable, that is a very shaky foundation for any later inferences.[6]

Improving the selection process for pharmacy residents is crucial. Fairness to applicants is paramount, while consistency is vital when attributing any consequences to screening decisions. Interview processes (including screening) are notoriously problematic and poorly reliable.[11,12] The results herein using a horizontal method, complement highly reliable results from an interview format that also used a multiple independent sampling (MIS) method.[13] Using MIS, multiple reviewers *independently* score one smaller portion of an entire set of performance stations (i.e., objective structured clinical exam), interview stations (i.e., multiple mini-interviews) or application criteria (i.e., horizontal scoring), and the multiple independent evaluators' scores summed into one integrated total score. In this way, the decision does not rest on any one person, for a halo effect to emerge. Screening and interview processes have high-stakes consequences for candidates.

Some or no agreement among evaluators was concerning. In this era where about one-half of pharmacy residency applicants do not match with a residency program,[1] screening and interview decisions can be career-changing for those unmatched applicants. Our screening and interview decisions need to be as rigorous as possible—in ethical fairness to applicants and using best-practice in measurement to confidently evaluate downstream consequences of rigorous decisions. In an investigation of validity evidence for pharmacy application scores, Nisly reported criteria that best predict performance in that specific residency program's interviews.[14] Future research could extend validity evidence through correlation with successful residency completion, or other success markers such as obtaining board certification or publishing a manuscript from their residency.

Addendum. In 2015, a four faculty-member team screened over 60 applications; a 25% increase in applications from 2014. Once again, this horizontal screening method was used. The eight criteria were divided among those four faculty-members to evaluate on all applications. The program preceptors who screened applications reported quick scoring for the relatively large number of applications. Their scores were collated and summed by the residency program director into screening total scores. Applicants were subsequently ranked by these total scores. The top 36 applicants were invited for on-site interviews. Additionally, the residency program director had reviewed each application in entirety to become acquainted with applicants before their interviews; this was helpful in his communication with them during and following screening within our resident selection process. From reviewer feedback afterwards, three of the four reviewers preferred this horizontal screening method, while one reviewer would rather have used the conventional vertical method of screening.

Conclusion

As pharmacy residency applications continue to grow, a consistent, fair and time-efficient method of screening seems imperative. Halo bias was seen with the single-evaluator vertical screening method; two interview invitations were negatively impacted. For pharmacy resident screening, the multiple-evaluator horizontal screening method appeared to be rigorous in promoting fairness for applicants.

Acknowledgements

This manuscript was presented at the 2014 American College of Clinical Pharmacy Annual Meeting in Austin, TX.

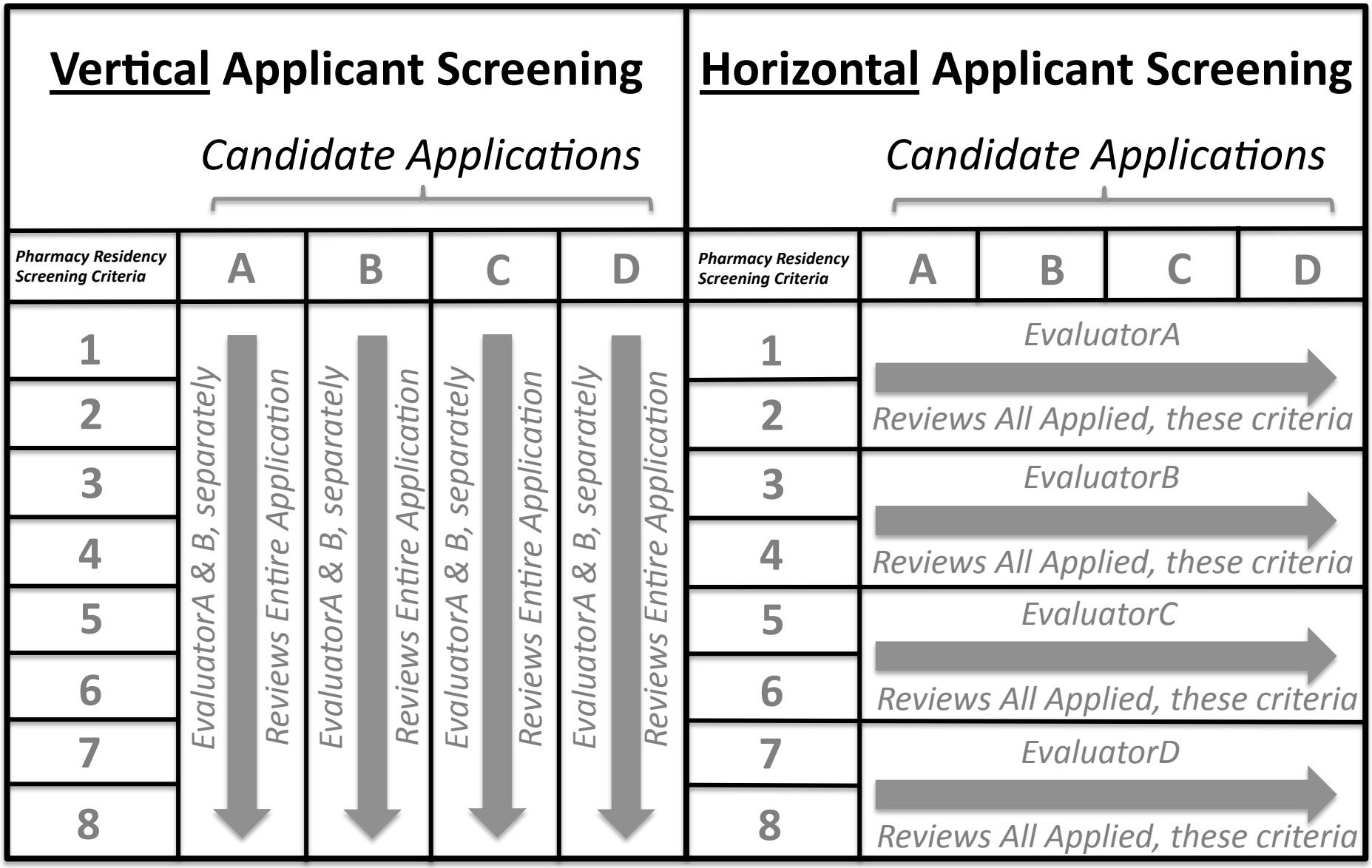
Contributor roles: MJP and JAM conceived this multiple independent sampling application for pharmacy residency screening. TEG & JAM designed the framework. MJP analyzed the data. MJP & JAM drafted the manuscript, while all

authors revised it critically. Final approval was granted from all authors; everyone accepts responsibility for the manuscript contents.

References

1. ASHP Resident Matching Program. Match Statistics. https://www.natmatch.com/ashprmp/aboutstats.html. Accessed September 12, 2015.
2. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* 4th ed. New York, NY: Oxford University Press; 2008:55,121-122.
3. Dore KL, Hanson M, Reiter HI, et al. Medical school admissions: enhancing the reliability and validity of an autobiographical screening tool. *Acad Med.* 2006; 81(suppl 10): S70-73.
4. Hanson MD, Kulasegaram KM, Coombs DL, et al. Admission file review: applying the multiple independent sampling (MIS) methodology. *Acad Med.* 2012; 87(10): 1335-1340.
5. Allalouf A, Klapfer G, Fronton M. Comparing vertical and horizontal scoring of open-ended questionnaires. *Practical Assessment, Research and Evaluation..* 2008; 13(8). http://pareonline.net/getvn.asp?v=13&n=8. Accessed on September 12, 2015.
6. Peeters MJ**,** Beltyukova SA, Martin BA. Educational testing and validity of conclusions in the scholarship of teaching and learning. *Am J Pharm Educ*. 2013; 77(9):article 186.
7. Peeters MJ. Practical significance: moving beyond statistical significance. *Curr Pharm Teach Learn.* 2016; 8(1). In press.
8. Oray JC, Ryan KE. Tips for Improving Testing and Grading. Newbury Park: SAGE Publications; 1993.
9. Davis BG. Tools for Teaching. San Francisco: Jossey-Bass; 2009.
10. Dauphinee D, Martin JB. Breaking down the walls: thoughts on the scholarship of integration. *Acad Med.* 2000; 75(9):881-886.
11. Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Health Sci Educ.* 2001; 6(2):159-71.
12. Peeters MJ, Schmude KA, Steinmiller CL. Inter-Rater Reliability and false confidence in precision: using standard error of measurement within PharmD admission essay rubric development. *Curr Pharm Teach Learn.* 2013; 6(2): 298-303.
13. Peeters MJ, Serres ML, Gundrum TE. Improving reliability of a residency interview process. *Am J Pharm Educ.* 2013; 77(8):article 168.
14. Nisly SA, Howard ML, Isaacs AN, Trujillo T. Association between application scores and onsite interviews of pharmacy resident applicants. *Am J Health Syst Pharm.* 2014; 71(24):2110-2113.

Figure 1. Horizontal and Vertical methods for screening applications



Modified from ***Dore et al. Acad Med. 2006; 81(10):S71***