Problematic Methods in the Assessment of Scholarly Productivity in Clinical PhD Programs

Barbara L. Andersen, Steven J. Beck, Robert A. Bornstein, Charles F. Emery, Mary A. Fristad, Janice K. Kiecolt-Glaser, Daniel R. Strunk, Julian F. Thayer, Michael W. Vasey, and Keith O. Yeates, Ohio State University

We review the methods in the ranking of clinical psychology doctoral programs provided by Stewart, Roberts, and Roy (2007). Using our own program as an example, we identify several areas of concern (e.g., authorship credits, criteria applied, faculty attrition). The inaccuracies identified for our program ranking, in combination with methodological concerns highlighted by previous commentaries, suggest that the validity of the rankings can be called into question.

Key words: **academic reputations, program evaluation, rankings, scholarly productivity.** ***[Clin Psychol Sci Prac 15: 102–104, 2008]***

We read with interest the article by Stewart, Roberts, and Roy (2007), reporting faculty publication rates over a five-year period (2000–2004) in 166 clinical psychology PhD programs approved by the American Psychological Association. A primary focus of their report was the number of authorships per program and the number of authorships per faculty member in each program. In addition to providing "normative" data on authorship rates in these programs, Stewart et al. ranked programs based on the quantity of faculty authorships. We also read the accompanying commentary by Heesacker and Elliott (2007) and found ourselves in agreement with their thesis that this kind of ranking does not provide information regarding the quality of clinical psychology training programs. Moreover, we were surprised by the data presented for our institution (Ohio State University [OSU]). Upon closer examination, we discovered areas of concern in the calculation of the data for our program, as well as in data for other programs.

According to Stewart et al., the training program at OSU obtained a rank of 146 for mean number of publications produced per faculty member. This rank differs markedly from the *US News and World Report* ranking of 38. On the basis of such discrepant rankings, Stewart et al. identified OSU as having a reputation exceeding its research productivity. In fact, OSU was at the top of the list of such programs, strongly suggesting that the OSU training program reputation is overrated. Our ranking of 146 is indicated in the original article and our position at the top of the list of the apparently overrated programs is indicated in a table ("Top 10 programs whose reputations are beyond their total research productivity rank") available from the authors upon request. Following our review of their report, we identified the following areas of concern:

(1) Stewart et al. reported a total authorship count of 18 for OSU. Our own PsycINFO search identified 36 authorships for the five faculty members Stewart et al. included in our program. The most significant

Address correspondence to Daniel R. Strunk, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210. E-mail: strunk.20@osu.edu.

error for our program was traced to the misspelling of a name of one of our faculty members, but the total authorship count was also incorrect for at least one other faculty member whose name had been spelled correctly. Dr. Roberts acknowledged that the spelling correction would increase the mean number of authorships per faculty member from 3.6 to 7.2 (M. E. Roberts, personal communication, July 2, 2007), elevating the mean rank for OSU to 110. At our request, Dr. Roberts sent us the data for our program and the other nine most overrated programs included in the supplementary table (based on mean number of faculty authorships). In our review of the data for the other nine programs, we noted additional errors. A total of 27 faculty members in these other programs received credit for zero authorships. Our own search found that among this group of faculty, four produced publications during the 2000–2004 time frame. As in the case of OSU, some of these errors apparently resulted from Stewart et al. conducting searches with misspelled faculty names. Thus, the considerable increase in the OSU ranking (and the likely changes to the ranking of other programs) underscores the tenuous nature of the rankings due to clerical errors.

(2) Our program at OSU was listed with five faculty members. Identification of faculty members for each program was determined by asking Directors of Clinical Training to respond to an e-mail request or by consulting public information about each program. Dr. Roberts provided us with a copy of the e-mail request that was sent. The e-mail asked for a list of "core" faculty and asked that this list not include appointments of the following types: "adjunct courtesy, clinical professor or supervisor, instructor, or professor emeritus." Whether other types of appointments were to be included was not indicated. The e-mail did not identify the purpose of the request. Examination of the tables in the Stewart et al. article reveals surprisingly large variability across training programs in the number of program faculty identified. It appears that inclusive criteria were used for a number of programs, whereas a restricted approach was used for others. For example, in 2004, our training program included six faculty members with primary appointments in other departments but with a joint appointment (the term used at OSU) in psychology. These faculty members regularly admit and advise graduate students in the clinical program, and also teach graduate courses. We regard such faculty as central to our training program, but they are not identified as "core" because they have primary appointments outside of our department. If our six joint faculty members had been included, the mean number of authorships per faculty member would have been approximately 11.5, and the mean rank for our program would increase to approximately 50. Given the markedly increased ranking of OSU when all training faculty are included, the validity of OSU's ranking is undermined. In addition, of the 27 faculty with zero publications referred to in item 1 above, 13 have appointments that appear inconsistent with the Stewart et al. criteria (e.g., Clinical Assistant Professor, Research Assistant Professor, Assistant Professor in a VA Medical Center). Because differing standards for the inclusion of faculty appear to have been used across programs, the validity of the rankings is again undermined.

(3) Only faculty members present at the end of 2004 were included in the publication count. Thus, for example, if an assistant professor was hired and/or a senior faculty member left/retired in 2003, the publication count would not reflect accurately publications emanating from the training program during the entire period (2000–2004). In our case, two highly productive faculty members had left the program shortly before 2004 and had not yet been replaced. Movement of faculty may not significantly alter publication counts for larger programs, but it dramatically affects the data for smaller programs and programs in transition.

We have highlighted only a sampling of the more obvious sources of error and variability in the data used to determine the rankings. As noted by Heesacker and Elliott (2007), there were other potential methodological shortcomings, such as counting authorships rather than publications, with no correction for publications by multiple faculty members at a single institution. Some training programs, by design, include faculty who form a collaborative publishing group, whereas other programs do not form a collaborative group for publishing. In our program, for

example, faculty represent three tracks of research and training (adult, child, and health), and joint publications have been uncommon. Also germane to this issue is the increasingly inter- and cross-disciplinary nature of psychology. Many journals in which interdisciplinary researchers publish are not indexed by PsycINFO and thus publication counts would be underestimated for such programs. However, an accurate counting of publications still would not index quality or impact of the work.

Stewart et al. undertook the difficult task of providing comparative data for a diverse group of training programs in clinical psychology. As reflected in the example of our program, the validity of the rankings can be questioned. Developing reliable and valid measures of the productivity of faculty at training programs and program success is a highly complex endeavor. The problems revealed in Stewart and colleagues' approach underscore the need for more work in the development of uniformly acceptable ratings of clinical psychology training programs. We hope that future work in this area will address the identified methodological and conceptual needs.

REFERENCES

Heesacker, M., & Elliott, T. R. (2007). My dog's better than your dog: Publication counts and quality of clinical psychology PhD training. *Clinical Psychology: Science and Practice*, *14*, 175–178.

Stewart, P. K., Roberts, M. C., & Roy, K. M. (2007). Scholarly productivity in clinical psychology PhD programs: A normative assessment of publication rates. *Clinical Psychology: Science and Practice*, *14*, 157–171.

Received August 24, 2007; revised August 24, 2007; accepted August 27, 2007.