

Top 10 Responses to the Commentaries on Dixon, Reed, Smith et al. (2015)

Mark R. Dixon¹ · Derek D. Reed² · Tristram Smith³ · Jordan Belisle¹ · Rachel E. Jackson⁴

Published online: 5 October 2015

© Association for Behavior Analysis International 2015

Abstract In a previous article (Dixon et al. *Behavior Analysis in Practice*, 8(1), 7–15, 2015), we put forward data suggesting that most behavior analytic faculty do not publish in major behavior analytic journals, and in only about 50 % of behavior analysis programs have faculty combined to produce ten or more empirical articles. Several commentaries followed the release of our article, with content that ranged from supporting our endeavors and confirming the dangerous position our field may be in to highlighting the need for further refinement in procedures used to rank the quality of behavior analysis graduate training programs. Presented in the present article are our “top 10” responses to these commentaries.

Keywords Research productivity · Behavior analysis · Behavior analyst certification board

In our now seemingly controversial article (Dixon et al. 2015, we put forward data indicating that only about 50 % of behavior analysis programs have faculty who have produced ten or more empirical articles in major behavior analytic journals and that most behavior analytic faculty have not published a single article. Further, we provided “top 10” lists of programs and their faculty in terms of research productivity, as a metric for evaluating the degree to which faculty who train aspiring

behavior analysts are actively creating knowledge that provides the foundation of our field. We presented what we see as a discouraging truth about our field, and members of the field responded. Twelve expert commentaries, appearing in this issue of *Behavior Analysis in Practice*, provide interesting perspectives on the current state and future of applied behavior analysis (ABA) practice. Below are our top 10 responses to the commentaries.

1. Mission accomplished.

Our goal was never to offer a definitive exploration of research productivity or clinical competency. Rather, we hoped to alert others to the possibility that, despite the essential role that empirical work plays in ABA, faculty in most of the field’s graduate programs do not publish much (we assumed that this means they do not conduct much research; more on this assumption below). Our findings are the first of their kind to appear in print, are striking in their clarity, and, if we may say so without being immodest, are therefore worthy of the attention they received in the commentaries. The commentaries, both complimentary and critical, can exist because we made a concrete matter about which previously the interested observer could only speculate.

We realize that quantifying productivity is a tricky business. We expected that some readers would disagree with our decisions about how to accomplish this, and we anticipated that our procedures could be improved upon. Several critical commentaries later, these expectations have been realized, but clearly our most basic goal was accomplished. In ways that were not true before the publication of our article, members of the field are now debating the functional significance of, and means of quantifying, the scholarly work of ABA faculty. The sleeping bear has been poked.

✉ Jordan Belisle
jbelisle@siu.edu

¹ Behavior Analysis and Therapy Program, Southern Illinois University, Carbondale, IL 62901, USA

² University of Kansas, Lawrence, KS, USA

³ University of Rochester Medical Center, Rochester, NY, USA

⁴ University of Houston–Clear Lake, Houston, TX, USA

Only upon accepting this discouraging (to us, at least) truth about the research productivity of those most responsible for disseminating our field's knowledge to its next generation, we can move on to the corollary issue—does the research productivity of ABA faculty really matter in practitioner training. Several commentators (e.g., Detrich 2015; Maguire and Allen 2015) suggested that low faculty research productivity does not necessarily signal that a training program is inferior. This is a point worthy of debate, and it can only be good for our field to evaluate how specific training experiences impact professional competency. Even if we were to concede that good practitioners might arise without research training; however, our data carry a disturbing implication that was not addressed in our original article or the commentaries. If ABA training program faculty define the pool of potential contributors to the expansion of knowledge in our field, our data suggest that actual contributors are few. A field with too few “knowledge generators” is at risk of stagnating.

2. Methods and data are intertwined.

The procedure used in our study was relatively simple—We report the number of articles identified for individual authors in specific journals by Google Scholar, which we chose because it is publically accessible and easy to use, allowing any interested reader to replicate our analyses. Upon the paper's release, we received communications from a number of individuals asserting that their independent searches revealed “errors” in our faculty productivity rankings. A slightly different interpretation is that data are only as good as their source, and Google Scholar has its limitations. A recent paper in the *Journal of the American Medical Association* (Kulkarni et al. 2009) compared citation metrics of articles published in general medical journals using Web of Science, Scopus, and Google Scholar. Results differed significantly across databases, so it would not be surprising if different “investigators” using different databases replicated specifics of our findings to varying degrees. A possible further consideration is the stability of the Google Scholar database itself. During our own data collection, we observed that the article count generated by a specific search could vary somewhat from day to day.¹ Consequently, we are not surprised that several commentators proposed alternative metrics for evaluating scholarly productivity (e.g., Maguire and Allen 2015; Wilder et al. 2015), and

¹ In at least one case, the day-to-day variation was not trivial. Our initial count omitted more than 20 publications of one faculty member that, for unknown reasons, appeared in the Google Scholar database only months later after we collected our data. We did not publish an erratum because our article accurately described the data we obtained using the procedures we described. Still, this deviation underscores the difficulties of quantifying scholarly productivity.

we concur with them that details of faculty productivity rankings will vary as a function of the specific metrics employed.² We are nearly as certain, however, that any credible metric will lead to replication of our main findings, that most ABA faculty do not publish much.

We reject three too-simple objections to this conclusion. First, it is possible that some ABA faculty conduct research but, due to competing demands (e.g., oppressive teaching duties) are not able to create the written reports that peer-reviewed journals demand. This is an untestable hypothesis because no obvious means exist to verify the existence of the “file drawer” studies that would document research productivity. Second, some faculty may not have an independent research program but instead expertly guide students through the process of conducting high quality research. This is a partially untestable hypothesis. Although it may be possible to determine how many of a faculty member's students completed theses or dissertations, unless that research undergoes peer review, there is no easy measure of whether it meets conventional quality standards. If student research does pass muster in peer review, of course, the faculty member, as co-author, would receive credit in analyses like our own. Third, perhaps our procedures underestimated faculty productivity by focusing only on prominent behavior analytic journals (i.e., ABA faculty might be publishing elsewhere). This is a testable hypothesis, and we welcome discussion about which other journals ought to count. In the meantime, we have difficulty imagining circumstances that would lead numerous productive ABA researchers to never publish in major behavior analysis journals. For the time being, we believe that no credible alternative has been offered to the conclusion that most ABA faculty do not publish.

3. Beware the fundamental attribution error (or do not shoot the messenger).

² Another factor that influences data is how variables are operationally defined. Some who contacted us about our article suggested that our rankings omitted certain individuals who work primarily in clinical research settings, are affiliated in some fashion with a training program, and therefore contribute to the mentoring of graduate students. We had a similar concern while gathering our data but chose to uniformly apply the objective search criteria that were described in our Methods section. One difficulty that we experienced is that web sites and other public descriptions of graduate programs do not always accurately identify program faculty or specify the role that affiliated faculty play in a program. In the latter case, for instance, a faculty member might teach courses in a BCBA-approved course sequence, be listed as program faculty but teach only in other areas like behavioral neuroscience, or supervise students' clinical work without teaching didactic courses at all. Similarly, a program-affiliated researcher may or may not routinely involve program students in his or her research program. If the goal is to hold programs accountable for how practitioners are trained, an obvious initial step is to standardize what is meant by “program faculty.” We emphasize, however, that the present lack of standardization is a characteristic of the field, not a weakness that is peculiar to our specific data collection methods.

Among the suggestions made by correspondents who contacted us about our article was that our findings were somehow an artifact of our personal biases regarding the importance of research in practitioner training, regarding which journals really matter, and so forth. Like all authors, we embrace certain professional values, but data, it is important to stress, have no values (other than the numerical kind). We took the usual precautions to promote objectivity in our research (e.g., we defined our measures, employed a standardized data collection process, and verified observations using interobserver agreement) and were transparent about our methods. It is fine to suggest better ways of collecting data, but the resulting data should be discussed on their own merits.

4. The “best” should not be enemy of the “good.”

Because there can be many different approaches to quantify scholarly productivity, it may prove difficult to develop an uncontroversial way of ranking faculty and training programs. Our position is that the need for better quality control over practitioner training is sufficiently acute that our field cannot afford to wait around for a “perfect” method. Burgeoning demand for applied behavior analysts has spurred massive growth in the graduate training “industry,” and consumers deserve a means of distinguishing between better and worse programs and the practitioners trained by them. For now, an “imperfect” set of rankings is better than none, and as the current discussion may illustrate, it is easier to refine an existing ranking system than to wait for a “perfect” one to emerge.

5. When you name names, people pay attention.

Our original article suggested that program and faculty rankings serve a discriminative function, and responses to our article help to validate this point. Within hours of the article’s release, we received several queries about our results, most commonly from faculty asserting that they (or their programs) belonged in the “Top 10” or asking how close they (or their programs) came to inclusion in the top 10. A recurring theme in the published commentaries concerned how better (and, by extension, worse) programs are to be identified. We believe that our results would have received much less scrutiny had we presented only actuarial data (e.g., the number of faculty or programs without any publications).

As we suggested in our original article, the purpose of rankings is to harness social and professional contingencies. At the least, responses to our article illustrate how “naming names” gets attention, and once people are paying attention, the potential exists for change. In this regard, a critic might argue that our article drew attention to the wrong outcomes, because our top 10 (and ensuing discussion about them) lists emphasized the most productive individuals and programs when, according to

our own logic, the field’s real worry is over the least productive ones. Imagine, however, the reaction that might have ensued had we chosen to publish “Bottom 10” lists. This is, in effect, what comprehensive rankings do, and we make no apologies for the potential of rankings to create an aversive situation for low-ranking faculty and programs. To serve as an agent of quality control, rankings must exert exactly this kind of pressure.

6. Assuming non-research programs produce better practitioners is wrong.

At the time of original article was written, we knew that there was no objective evidence that research training makes for better practitioners. Our position on the value of research training was a logical one, derived from the observation that ABA has always been research-informed. We would have been (and should have been) pilloried if we asserted that research training must promote practitioner competence because there is no definitive evidence to the contrary. In the empirical world, no position ever is validated simply by the absence of disconfirming evidence.

Yet, this is essentially the type of logic advanced by some of our critics in claiming that training programs without a strong research emphasis must produce better practitioners than those in which faculty are distracted from students’ clinical pursuits by the demands of their own research programs. There exist no more data to support this proposition than the one we advanced in our article, so in point of fact, the practitioners who graduate from a “research-light” program, could be terrific or terrible. One thing seems clear, however. In the absence of informative data, promoting research light programs runs counter to the original conception of ABA as a profession in which science and practice are essentially the same activity (Bailey and Burch 2002).

Quite obviously, training standards are best guided by data on the relationship between various aspects of graduate training and the subsequent competence of practitioners (e.g., see Critchfield, this issue). To provide a very preliminary illustration, we evaluated the relationship between program publication counts reported in our article and the percentage of program graduates who passed the 2013 certification examination (data reported by the Behavior Analyst Certification Board (BACB)). A Pearson correlation revealed a positive correlation between faculty research productivity and student pass rate ($r=0.584$, $p<0.005$). Within the limits of the available data (e.g., there is currently no objective basis for assuming that passing the certification exam predicts field competence), this finding is contrary to the assumption that that programs whose faculty publish frequently produce weaker practitioners.

7. The research-practice divide is real.

A strong link between science and practice is thought to be a defining feature of ABA (Bailey and Burch 2002), but in all clinical fields, there are concerns about whether this link is maintained rigorously enough. In ABA, worries about a “research-practice divide,” in which practice is inadequately informed by science, trace back at least 40 years, but only rarely have been informed by objective data (e.g., S.C. Hayes 1978). Our data provide a useful contemporary snapshot of one possible manifestation of the “divide.” As suggested by L. J. Hayes (2015), engaging in research is central in promoting research values and creating a community that shapes research behavior (broadly defined, see below) in its students. This research culture appears to be nearly absent in about 50 % of ABA programs. If graduate training does not guide practitioners in developing research-informed clinical work, what will? It may be too much to expect practitioners to conduct research (e.g., Critchfield 2015a, b), but it is imminently reasonable to expect them to align clinical practice with advances in research. Practitioners therefore must be able to consume research. Research training may not be the only means of learning how to do this, but it is a time-tested means that our data suggest is being ignored in too many current training programs.

8. Those who do, do not necessarily teach.

Our article indicated that many who teach in ABA graduate programs infrequently engage in research, but do those who engage in research necessarily teach? It is not a stretch to conceive that teaching and research assignments are not evenly distributed across faculty in graduate programs. Different faculty are likely to have different strengths and faculty assignments may reflect this. For example, some programs may support scholarship by releasing productive researchers from teaching duties.³ This approach has its benefits but raises questions about the amount of contact that future practitioners will have with the most productive researchers on their faculty. Both common sense and a more formal conceptual analysis (e.g., L.J. Hayes 2015) suggest that students must actually interact with a faculty member to benefit from his or her expertise.

Our analysis did not take into account the possibility that that faculty who publish most frequently might teach few courses and therefore have only limited interactions with students. In other words, our analyses might have overestimated the research climate in some programs

with productive faculty. An alternative approach to ours, as some commentators suggested, would be to rank ABA programs strictly on the basis of the productivity of faculty who play a direct role in ABA training by teaching courses in a BACB-approved sequence or by supervising certification-relevant field work. In this way, no program would receive undue credit for the research productivity of colleagues who are only tangentially affiliated with it. As we noted previously, however, because of inconsistencies in program promotional materials (e.g., web sites), it is currently impossible to accurately define what contributions each “program faculty member” makes toward training future practitioners.

9. We need a blue ribbon panel.

A useful conversation about practitioner training has begun, and it is apparent that intelligent people disagree about what constitutes exemplary training. At the present stage of our field’s development, diversity of opinion probably is unavoidable. Among other influences, people who are actively engaged in research (like the authors of our original article) probably are biased to see an important role for research training whereas people who are primarily engaged in clinical work will see more value in other things. Following ABA’s origins as a synthesis of science and practice (Bailey and Burch 2002), there is no reason to assume mutual exclusivity in the development of research and practice skills. The debate, presumably, is over how these skills should be developed and balanced within the time constraints of graduate training.

Our current lack of a clear definition of what constitutes good training therefore serves as a major impediment to the evaluation of training program quality, and we remain steadfast in our assertion that our field has a pressing need for the consumer education and protection that program rankings can provide. As a hedge against inertia, we propose that a panel of experts, representing all reasonable perspectives, be brought together to determine the best way to evaluate program quality. With a bit of luck, discussion may erode the subjective biases that all of us risk bringing to bear on the idea of competitive rankings and sow the seeds of a workable—not perfect—means of assessing the quality of ABA training programs. Whether the resulting system reflects the procedures we described in our original article is far less important than the fact that it would reflect the input of many types of stakeholders.

Once a workable—not perfect—evaluation system exists, it can yield at least three kinds of empirically informed benefits. First, programs that are favorably ranked could use this to market themselves to potential students. They could also levy critical support from university administration for things that influence rankings, such as hiring good faculty or maintaining costly research operations. No university seeks to lose national ranking of its academic

³ It is important here not to endorse stereotypes uncritically. Many of the people listed in our “Top 10” also have distinguished reputations for teaching frequently and effectively. Research productivity does not necessarily preclude teaching, and heavy teaching loads do not necessarily preclude research productivity.

programs. In our experience, however, when no ranking is available, administrative whims often prevail. Second, unfavorably ranked programs could know precisely what is lacking and what is needed to achieve a better ranking. Universities typically dislike poor program rankings, so justification may exist for such things as increasing faculty hires, reducing teaching loads, improving research space, and decreasing class sizes.⁴ Third, objective program ranking identify clear models for universities that seek to create new behavior analytic training programs. Some guidance on how to design a program already exists in the form of BACB-approved course sequences and ABAI program accreditation guidelines; a ranking system would expand upon and complement these mechanisms.

10. Critiques are easier to gather than data.

At the risk of sounding unappreciative of the contributions of those who commented on our original article, we close by reiterating a theme that runs through the present essay: Discussion is necessary, but it must be supported by data. Although we are a data-driven field, as a community of scholars and professionals, we sometimes seem quite comfortable discussing our field's foundations—including the steps that will be taken to train our scholarly and professional successors—on the basis of reason and opinion. To us, our original article, the collected commentaries, and the accompanying essay by Critchfield (this issue) all point to a need for data to drive our graduate-training practices. All that is left is actually gathering the data. It should not be enough for programs to be designed solely around the hunches of program faculty and the BACB's minimal standards for certification, as we believe is too often the case. We are proud of any role we have played in shifting the conversation toward relevant data and their merits, because as behavior analysts data is what we all do. From here on,

let the conversation about graduate training continue, but let it be data-driven, and let those who care about this important issue join with us in generating new data that inform our conception of what constitutes good graduate training and who is (and is not) providing this to future practitioners.

References

- Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks: Sage.
- Critchfield, T.S. (2015). In dreams begin responsibility: why and how to measure the quality of graduate training in applied behavior analysis. *Behavior Analysis in Practice*.
- Critchfield, T. S. (2015b). What counts as high-quality practitioner training in applied behavior analysis? *Behavior Analysis in Practice*, 8(1), 3–6.
- Detrich, R. (2015). Are we looking for love in all the wrong places? Comment on Dixon et al. *Behavior Analysis in Practice*, 1-3.
- Dixon, M. R., Reed, D. D., Smith, T., Belisle, J., & Jackson, R. E. (2015). Research rankings of behavior analytic graduate training programs and their faculty. *Behavior Analysis in Practice*, 8(1), 7–15.
- Hayes, L. J. (2015). There's a man goin' round taking names. *Behavior Analysis in Practice*, 1-2.
- Hayes, S. C. (1978). Theory and technology in behavior analysis. *Behavior Analyst*, 1, 35–41.
- Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Journal of the American Medical Association*, 302, 1092–1096.
- Maguire, R. W., Allen, R. F. (2015). Another perspective on research as a measure of high-quality practitioner training: a response to Dixon, Reed, Smith, Belisle, and Jackson. *Behavior Analysis in Practice*, 1-2.
- Wilder, D. A., Lipschultz, J. L., Kelley III, D. P., Rey, C., Enderli, A. (2015). An alternative measure of research productivity among behavior analytic graduate training programs: a response to Dixon et al.(2015). *Behavior Analysis in Practice*, 1-3.

⁴ A ranking system also carries risks. One plausible administrative response to a low program ranking is to discontinue the program. That is not necessarily a bad thing.