

Data Comparability, Utility, and Collection Methods of the 2010 NRC Assessment of Doctorate Programs: Limitations and Suggestions for Improvement

Julie Carpenter-Hubin (The Ohio State University) and Lou McClelland (University of Colorado Boulder), representing the Association of American Universities Data Exchange (AAUDE) ¹

Presented at "Assessing Research-Doctorate Programs: A Convocation on Analytic Uses and Future Directions," Washington DC, March 4, 2011. Revised 3-21-11.

The Assessment of Research Doctorate Programs conducted by the National Research Council (NRC) provides data that allow comparisons to be made among similar doctoral programs around the United States, with the goal of informing efforts to improve current practices in doctoral education. . . . [It] offers accessible data about program characteristics that will be of interest to policymakers, researchers, university administrators, and faculty, as well as to students who are considering doctoral study. **Furthermore**, the assessment analyzes and combines these data to create ranges of rankings that allow the comparison of different doctoral programs within a field.

-- *A Guide to the Methodology of the National Research Council Assessment of Doctorate Programs* -
<http://www.nap.edu/catalog/12676.html> - Introduction (emphasis added)

This is an opportunity to look back at the vast endeavor of the NRC study, and forward to a more efficient, more useful, ongoing data collection, based on the goals stated in the Guide to Methodology.

Our perspective is that of the Association of American Universities Data Exchange (AAUDE). AAUDE (rhymes with "howdy") has been focusing on comparative data about research universities since 1974, with discipline or program-specific data exchanged since before 1980. AAUDE is loosely affiliated with the AAU, and all 63 members of the AAU participate in AAUDE. Most representatives are in institutional research offices; many were the Institutional Coordinators (IC's) for the NRC study, as were both authors. The 61 AAU US universities house over half the programs participating in the NRC study in every broad discipline, 63% overall; these programs in turn house over 60% of all enrollment reported by programs in the study.

¹ Carpenter-hubin.16@osu.edu; Lou.McClelland@colorado.edu. Both authors direct institutional research offices; they co-chair the AAUDE working group on graduate education data. Both are also on the (unpaid) institutional steering committee of Academic Analytics, LLC.

We believe that the “furthermore” portion of the goals – the “ranges of rankings,” especially the overall rankings -- became the tail wagging the dog, a hindrance to accessible, useful data rather than a help. Briefly, we believe this to be the case because

- The rankings were too complex, and the ranges too wide. Difficulties in use and interpretation stole attention from the measures on which they were based.
- The overall rankings attempted to combine facets of PhD programs that are incommensurate or at least more clearly considered separately: scholarly productivity, student process and outcomes, diversity of students and faculty, program size.
- The overall rankings purport to measure quality but did not take into account some of the most important components of quality, such as quality of publications. While it can be argued that citation counts address publication quality, excellent publications on specialized or niche topics may have fewer citations than very good but lesser quality publications on topics studied by a broader range of scholars. Quality of book publications was not measured.
- The ranking process took enormous effort and time on NRC’s part, which reduced the time available for other efforts and extended the lag between data collection and publication.
- The 21 measures used in the rankings gave a false sense that these were the 21 measures with all relevant information.
- The focus on the rankings meant a focus on quantitative measures available for almost all programs, with a supposed correlation to “quality.” This excluded publication of valuable information such as research specialties, affiliated fields, and numbers of post doctoral researchers.
- The ranking methodology required that program “similarity” be defined entirely by NRC field and that each program be in one and only one field.

For future iterations we suggest

- 1: Aiming for a sustainable, replicable, continuously improving process
- 2: Focusing on characteristics and comparable data, rather than on “assessing quality” with a combination of incommensurate facets of PhD programs. And measuring reputation, or reputation of faculty scholarly work, as one characteristic.
- 3: Aiming for easier public access to and use of results
- 4: Consider next steps now, and don’t do it again the same way, please.

We discuss each of these suggestions in turn, with a following section with selected examples.

Our suggestions recognize the multiple purposes of the NRC’s assessment: the overall assessment of quality; the compilation and publication of data about the doctoral student experience, including the diversity of the faculty and student bodies; and the comparison of faculty scholarly productivity across programs. We suggest that moving away from a single overall assessment toward development of a coordinated set of data collections and publications could provide faculty, administrators, current and prospective students, and the public with substantially improved information. In addition, we point to other collections and publications of higher education data that may serve as models for future doctoral program data compilations.

1: Aim for a sustainable, replicable, continuously improving process

The process must be easier on schools/programs and on NRC itself, and must produce results faster, to be sustainable and garner support from institutions.

- Streamline the process
 - Use IPEDS completions to define eligible programs. (IPEDS: Integrated Postsecondary Education Data System of the U.S. Department of Education)
 - IPEDS now distinguishes research doctorates from other types of doctoral degrees, making this possible. This would also allow expanding the universe of programs to include research doctorates in fields previously excluded, such as business and education.
 - Work with institutions to turn these into program lists, with new programs added, institutional names of programs attached, level of granularity aligned, and primary and secondary fields attached. See illustrations in the examples section.
 - Use CIP codes to facilitate mesh with IPEDS and other data and a coding system with continuous updates. CIP = US Dept of Education Classification of Instructional Programs.
 - Simplify the faculty list definitions and publish the lists
 - The faculty lists, designations, and allocations were
 - The most confusing and time-consuming portion of data collection from institutions. Even institutions that tried to follow the instructions in good faith had differences in their understanding, and implementing the definitions was an ordeal for many.
 - The most contentious portion after report publication
 - Critically important because they formed the basis for all research/scholarship measures.
 - Ask schools, programs to list all their tenured/tenure-track (TTT) faculty, and only TTT faculty.
 - Data users routinely assume that “the faculty” in a program means the tenured and tenure-track individuals in the department housing a program – all of them, and no one else. This is “the faculty” that users – both prospective students and institutions -- want to know about. Thus, the complicated definitions were not only onerous, but unnecessary and counterproductive.
 - Allow an individual to be listed as tenured/tenure-track in multiple programs.
 - Simplification would eliminate the core-new-associate distinction and the allocation concept. Individuals with multiple tenure loci would be counted as full members of each program.
 - Listing any and all TTT should markedly increase efficiency at the school level and markedly increase comparability of lists over schools and programs.
 - Invite programs to identify other faculty associated with their programs, but list these separately and do not use them for collection of scholarly productivity measures. Also invite program comment on faculty affiliations, specialties, changes, etc.
 - Publish all the lists, at the outset, with names and ranks; if possible, also with research specialties, important administrative duties, and other descriptive information.
 - This is done in the Philosophical Gourmet (Leiter), a long-standing set of information about graduate programs in philosophy.
 - The lists, and specialties, are data in and of themselves, as important in building a picture of each program as any numerical data. In addition, published lists mean peer scrutiny of lists and increase comparability.

- Make use of existing collections and collection methods of federal agencies, commercial entities, and organizations like CGS (Council of Graduate Schools) and AAUDE.

Examples:

- IPEDS completions, by CIP code, including research and practice doctorates, master's, and bachelor's in the discipline, for context. By gender, international status, and race/ethnicity. Plus numbers of completions by degree level at the university level, to provide context.
- National Science Foundation (NSF)-sponsored collections including the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS) and the Survey of Earned Doctorates (SED)
- Academic Analytics or similar commercial services which match faculty names to scholarly productivity measures
 - These have honed their methods for sweeping publications, citations, books, honors, and grants from public data, and matching to individual authors, PIs, recipients
 - Such a service might conceivably even provide faculty lists, at least for schools already using the service
 - Example: AAUDE completion rates and time to degree – based on, then refined from, CGS and NRC methods by repeated testing with a dozen schools. The AAUDE method, still in pilot, collects qualifiers on comparability and simplifies collection.
 - Last choice: Collection directly from schools and faculty with methods specific to this study. Even there, the goal should be incremental development of collections and measures to be taken over eventually by agencies, commercial entities, or organizations of institutions. A one-shot collection is worst of all.
- Collect frequently enough to allow routinization but also evolution of measures, instruments, process. Comparisons over time have value but it's vastly better to get it right than to stick with inappropriate measures.
- Challenges in following this approach
 - Staying simple
 - Collecting and conveying qualifying information, such as whether students enter a program after a master's or straight from the bachelor's
 - Business models for commercial services
 - IPEDS perturbation of small N's
 - Deciding what can't be done or isn't worth the effort
 - Dealing with programs with large variation in size
 - Adopting compromises – in using CIP codes, adding other fields, etc.
 - Moving to centralized collection from schools, rather than program by program, to increase efficiency, data comparability, and institutional capabilities
 - Dealing with interdisciplinary programs at all steps of the process
 - Balancing desires to compare over time with those to improve measures.

2: Focus on characteristics and comparable data, rather than on “assessing quality”

- Drop the overall combination and the rankings that went with it
 - The overall rankings attempted to combine facets of PhD programs that are incommensurate or at least more clearly considered separately: scholarly productivity, student process and outcomes, diversity of students and faculty, program size.

- If NRC focuses solely on collecting and providing data, other organizations – like PhDs.org – can use those data to create tools for potential graduate students and others. Institutions would be more likely to develop combinations that are meaningful to them, which is exactly what NRC stated as its goal of this assessment.
- Publish data of all types. A focus on publication of data, vs. listing of measures in rankings, would:
 - Require fewer data collection steps, reinforcing the streamlining already outlined above. Collection of faculty opinions on the relative importance of measures would not be needed, nor would the “rating questionnaire.”
 - Allow programs to be listed in multiple fields
 - Allow publication of data that can’t be readily included in a ranking, such as
 - statements about interdisciplinary work, research specialties, and the like
 - a list of the 3 (say) journals with the greatest N of publications by the program, or the agencies with over \$x in research funding
 - institutional comments on the measures. Because data don’t always speak for themselves, we recommend that NRC allow and encourage institutions to comment in public on the measures. Such a facility would allow institutions to provide context or explanation for numerical data and allow comment on which measures are more highly valued in their discipline/program.
 - per capita and total measures of publications, citations, books, awards, grant money. Both provide meaning, and many users are interested in totals.
 - Avoid the need to impute data where values are missing for some programs, thereby broadening the range of data available and eliminating publication of misleading imputed figures, such as average GRE’s for programs not requiring GRE’s.
 - Heighten the utility of measures not linearly related to quality (e.g., percent of students who are international; time to degree) and even measures with comparability issues, as long as those were explained and qualifiers listed. See illustrations in the examples section.
 - Provide richer data for comparison and program improvement
 - Avoid the controversy involved in ranking methodology.
- Treat reputation as a separate dimension, measured directly and separately when and if desired; this could be done every 5-10 years.
 - Collecting reputation ratings is something NRC may be better equipped to do than any federal agency, commercial entity, or private organization
 - Might be done off-cycle, at a different point than other collection/publication
 - Might focus explicitly on scholarly reputation of the program’s faculty
 - Explicitly informed by faculty lists to reduce the halo and historical effects inherent in reputation ratings
 - Philosophical Gourmet model (see more in the examples section)
 - Reputation as a “good place to be a student” is a different thing and does not necessarily exist, or exist widely, or for all programs.
 - Direct collection of reputation of quality of faculty scholarly work would produce ranks or ratings explicitly based on fellow faculty members’ assessments of the quality of past scholarly work and the potential for more, rather than ranks based primarily on the volume of scholarly work in some (distant) past period.
 - NRC 2010 did collect “quality ratings,” an attempt to get ratings on several facets combined. These were used as dependent variables in regression analyses with the 21

measures as predictors, and the resulting coefficients used in overall ratings and rankings. The 21 measures include volume of publications, books, and citations, but no measure of their quality. Study details:

- Ratings were collected from a sample of faculty for a sample of programs in each field.
- The collection was labeled “Survey of Program Quality.” What a rater saw:
 - “This survey asks for your judgment—and the judgment of other faculty members like you—about the quality of a sample of doctoral programs in your field.”
 - “Please reflect on what you consider important in a doctoral program as you decide on your ratings. To assist you, a link below each program’s name goes to an information page that lists several program and faculty characteristics, a list of the program’s faculty and a link to the program’s web site as well, should you want to seek additional information before finalizing your rating.”
 - A list of all programs in the field participating in the study
 - For each program to be rated
 - Names of core, new, and associated faculty members, labeled as such
 - N of PhDs granted, pct in academic positions, completion rate, median time to degree, and faculty pct female and pct non-white
 - Link to program website
 - Question: “On a scale from 1 to 6, where 1 equals not adequate for doctoral education and 6 equals a distinguished program, how would you rate this program?” Alternatives: Not Adequate For Doctoral Education, Marginal, Adequate, Good, Strong, Distinguished
- Build a body of knowledge about data comparability, working on
 - Documenting dead ends, measures that don’t work
 - Checking comparability in selected institutions and programs
 - Investigating outliers and relationships
 - Ways of using multiple years of data to increase N’s and stability in small programs, and of conveying the sometimes very small N’s on which measures are based
 - Ways of publishing data to facilitate combining programs for comparison
 - E.g., to create a single “foreign language program” at each institution
 - This requires publication of all numerators and denominators separately (e.g., for completion rates), to allow recombination.

3: Aim for easier public access to and use of results

- Publish material allowing users to select, group, and even combine programs in ways useful to them
 - Every program in the 2010 study is listed in one and only one NRC field. Virtually all search facilities – the NRC Excel, PhDs.org, etc. – are designed to select a single field first, then select and compare programs.
 - Whereas NRC published standardized values on individual measures for programs within a field, recalculating those values for a subset of programs in a field, or a combination of programs from two or more fields, is beyond most users.

- Ideally the user – prospective student, a faculty member, an administrator, the press – could start with any filter, see the effect of each filtering, get to a set of programs to compare, and then see raw and standardized values *within that set* on all measures.
 - E.g., has “biochemistry” in program name or field. Limit to programs in universities with at least 5 life-science programs and at least 2,000 total degrees per year. Limit to programs with at least 20 faculty and at least 10 new students per year. Compare.
 - E.g., select all foreign language programs at public universities with 10,000 or more total students. Combine the separate programs within each institution. Compare.
- Publish rich data, not just numbers
 - Publish the faculty lists
 - Publish research specialties -- Right now Peterson’s is the only comparison facility with words other than the name of the program. Peterson’s collects and publishes program research specialties, but the specialties do not feed into the program search facility – only program name does.
 - Publish program comments about the data and their importance. E.g., does this program tend to publish in journals, or in books?
 - List (some) programs in multiple fields; list multiple fields with each program. The association of a program with a field should serve as a guide to data users, not a determinant of its statistics.
- Explore models such as
 - Voluntary publication on institutional/program websites of specified measures in a suggested/standardized format
 - Like a “Common Data Set” for PhD program information
 - APLU, the Association of Public and Land-grant Universities, has a “task force on graduate study reporting” looking at a small set of measures
 - PhDs.org, which displays 2010 NRC results alongside data from IPEDS, GSS, SED
- Work with organizations such as CGS, APLU, and AAUDE, and outlets such as PhDs.org and even commercial services like Peterson’s or US News, to create routine outlets for comparable information – words, lists, numbers -- on research doctorate programs
- Learn from the experience of publishing the 2010 results – Papers from the March 2011 convocation are relevant; in addition
 - The *Chronicle of Higher Education* has a for-pay site -- what are people paying for and looking at?
 - Who’s downloading the big NRC Excel?
 - Who’s using PhDs.org, for what?
 - What’s on other websites?
 - Other?
- Document sources and calculations fully. E.g., publish examples of all calculations. As was done with the honors and awards, provide a listing of journals included in the publication and citation indices.

Examples illustrating limitations and opportunities

- Programs and fields
 - Programs listed in one field but with clear ties to multiple fields
 - U of Colorado Boulder, Chemistry and Biochemistry
 - North Carolina State, Marine, Earth, and Atmospheric Sciences
 - *Effect: Programs aren't listed where one might look, or are listed as "unranked"*
 - Psychology – a field with programs of different granularity and focus
 - 236 programs, the highest by a margin of 56 over Chemistry
 - Only 60% of the programs are straight comprehensive psychology departments as judged by program name and size and the pattern of other programs in the field from the same school
 - Others are sub-disciplines – e.g., Clinical Psychology (U of Florida), Psychometrics (Fordham), Social/Personality Psychology (SUNY Albany; this is one of five programs in the field at this institution)
 - And: 29 programs are named Clinical Psychology, which was not to be included in the study at all
 - Some programs are not even in psychology at all – Communication Sciences and Disorders (U of Texas-Austin); Human Development and Family Studies (Pennsylvania State)
 - *All these were known from 2007. All affect the meaning of rankings and standardized scores for programs in the field*
 - Two fields with overlapping programs: Applied mathematics, and mathematics
 - All programs named "Applied Mathematics" are in the field "Applied Mathematics"
 - But so too are five programs named "Mathematics"; these are often the only program at the school in either field
 - *Is this a problem? CU-Boulder's math department thinks so.*
- Issues with measures – see also "Summary of issues with measures released by NRC" following the text and posted as <http://www.colorado.edu/pba/nrc/NRCMeasuresIssuesSummary.pdf>
 - Enrollment: A deceptively simple measure with comparability issues
 - (Question C8): How many doctoral students, whether or not they were yet admitted to candidacy, were enrolled in your program during fall of 2005?
 - This measure is easy only if the institution/program can unambiguously discriminate students intending doctorates from those intending master's. Many cannot do so.
 - The measure is not comparable if programs differ
 - Colorado counted every student who applied for PhD work at CU-Boulder and was admitted to a PhD program
 - In one program, Ohio State didn't count any of the graduate students in their first two years. Nearly all of those students think of themselves as doctoral students – that's what they applied for and what they intend, but the department admits them to a master's program and informs them at the end of two years as to whether they can continue for a PhD [and they're counted as doctoral students only after that]. That's not really any different from a program that starts students in a doctorate-only program, determines that they should not continue at the end of two years and gives them a consolation MA.

- *Issues with defining the “starting point” for students in a comparable way affect measures of enrollment, first-year enrollment, time to degree (time from what?), and graduation or completion rates*
- Continuous enrollment policies at some schools and not others pose another threat to comparability
- Super non-comparability: Full support
 - (Question E5): How many full-time first-year doctoral students 2005-06 had "full financial support"?
 - “Full support” was not defined. AAUDE work in 2007 or 2008 determined that institutional and program answers – even among AAU schools – differed radically in the definition of “full,” from \$1 over tuition and required fees, to an unknown higher amount
 - *But AAUDE did not publish these findings, so the non-comparability is generally not known.*
- Mismatch between what was rated by faculty as “important” and what was measured: Percentage of Students in Academic Positions
 - This measure was included in ratings because faculty rated it as important. But the faculty questionnaire asked for the importance of "Placement of students after graduation" -- not "academic" placement. Very different!
 - In addition, the denominator of the percentage, taken from student self-report to the SED, is the total N of PhDs reported in the program, not the N who replied to the employment questions on the SED. This differentially affected schools with lower SED response rates. We understand that this calculation will be revised for the 2011 re-release of the data.
 - And, some institutions have found a poor match between the placement reported on the SED, and later placement information collected from faculty and other sources.

Median GRE scores of entering students

- This measure was included in ratings because faculty rated it as important as well. The faculty questionnaire did not distinguish between verbal and quantitative scores; however, a decision was made to include verbal scores only for all humanities disciplines, and quantitative scores only for all others. Again, very different!
- Measures with a curvilinear relationship to “quality” – the most desirable value is not the maximum
 - Time to degree – Too short a time says the program was too easy or wasn’t measuring the starting point in a comparable way
 - Graduation or completion rates – If the rate is 100%, something is amiss in the program, or the starting point has been defined differently, or the program works differently from most others.
 - Percent of students who are international - 100% means that the program cannot attract U.S. students
 - Gender diversity – is 100% women maximally better than zero percent?
 - *The measures themselves may be fine; it’s the use in linear combinations of measures to create overall rankings, where higher values necessarily mean better, that’s problematic*
- Measures and information never released

- Faculty lists, research specialty lists
- Numbers of post-docs, whether students leave before or after candidacy, patterns of dissertation committee supervision, criteria for candidacy, placement, . . .
- *Why? Were the data hopelessly incomplete, wacky, or what?*
- *At least, document why not released, for next time*
- *At best, release these data outside the rankings context*
- AAUDE completion rates – Example of an ongoing collection
 - Refined and simplified from CGS and NRC methods
 - Based on trial and error, intensive review at the operational level, revisions. Still in pilot with a dozen schools.
 - For a cohort, at a time point after entry, report
 - N entering
 - N who have received the PhD, not received but are enrolled, or neither. The sum = the N entering.
 - N who have achieved candidacy – a subset of N entering, larger than or equal to the N who have received the PhD.
 - Plus categorization of a program's cohorts as
 - a) All applied for PhD work and were classified as PhD students upon entry, generally with no graduate work intervening between college graduation and entry to the program.
 - b) Program is entered from a master's program or from an umbrella or home PhD program; students are classified as PhD students in the program and counted in a cohort only after prior graduate work, usually a master's degree
 - c) Neither clearly a) nor b).
 - These alternatives are paraphrased and still in development. Programs not clearly a) will also have space to explain their situation.
 - Designed to simplify, facilitate comparison with similar programs, and shed light on differences
- *Philosophical Gourmet* reputational rankings
 - Focus is on the "quality of philosophical work and talent represented by the faculty." Note the emphasis on the quality (not just volume) of scholarly work, and on potential ("talent"), not just work in the past
 - Raters get faculty lists by institution, but without the name of the institution. One rater remarked that it made it "surprisingly tough to say what I think, without the institutional halo effect front loaded."
 - Both ratings and rankings are published, along with extensive "how to use" information which discusses student placement, differences in atmosphere of different departments, , paying attention to ratings rather than rankings, etc.
 - The entire endeavor appears transparent. Faculty lists and rater lists are both published on the website. Raters are nominated by a large advisory board to represent research-active faculty. Ratings are collected every 2-3 years.
 - Rankings cover philosophy departments in English-speaking universities. However, not all PhD-granting programs are rated. Rather, the focus is on "the top programs in each region . . . plus a few additional programs . . . to 'test the waters.'" Ninety-nine programs were rated in 2009 (across five countries); this compares to 110 PhD programs in philosophy in the US alone at the time, and 90 programs in NRC 2010.

- A set of ratings of departments in specialties within philosophy is also published; this gives a rich picture of a department's interests and strengths.
- The website (at least) is sponsored by Blackwell Publishing

4: Consider next steps now, and

Don't do it again the same way, please

- Issues in wrapping up "this time"
 - Issuing error corrections, recalculations, etc.
 - The unpublished data
 - Sponsoring concerted looks at the data
 - NCES faculty and doctoral student fellowships as model?
 - Learning from publishers – PhDs.org, *Chronicle*, etc.
 - Learning from collection, preparation, and publication problems
 - Making the most of convocation presentations and discussion
- Issues in "next time"
 - Is it best to treat doctoral student experience, faculty scholarly productivity, and faculty scholarly reputation as three separate projects?
 - Is there any need for an "overall" combination, or is this like flattening the facets of a diamond into something that no longer glitters?
 - Who are audiences for each facet?
 - What entities are most appropriately responsible for each collection? For coordination? For publication?
 - How often should data be updated?
 - Where should the data reside?
 - Can institutions take responsibility for publishing comparable data about themselves? Would it matter if some did not?
 - Does a sustainable business or funding model exist? How can commercial publishers, even "world ranking" efforts, fit in?
 - Are comparisons covering US institutions only still relevant?
 - What should and could NRC's role be?

References

- Main site for NRC Assessment of Research Doctorate Programs
<http://sites.nationalacademies.org/pga/resdoc/index.htm>
- AAUDE
 - Introduction <http://aaude.org/intro-to-aaude>
 - Public site on the NRC study; includes letter to NRC Jan. 2011, summary of issues with measures, AAU representation in the NRC study, and more
<http://aaude.org/nrc>
- Organizations, data collections, and data outlets mentioned in the text
 - Academic Analytics: <http://www.academicanalytics.com/>
 - APLU: <https://www.aplu.org/netcommunity/>
 - CGS: <http://www.cgsnet.org/>
 - Chronicle: <http://chronicle.com/>
 - <http://chronicle.com/article/Tables-Doctoral-Programs-by/124789/>
 - IPEDS: <http://nces.ed.gov/ipeds/>

- Peterson's: <http://www.petersons.com/graduate-schools.aspx>
- PhDs.org: <http://www.phds.org/>
- Philosophical Gourmet: <http://www.philosophicalgourmet.com/>
- NSF GSS <http://www.nsf.gov/statistics/srvygradpostdoc/>
- NSF SED: <http://www.nsf.gov/statistics/srvydoctorates/>
- Colorado IR NRC site <http://www.colorado.edu/pba/nrc/>
- This paper: <http://aaude.org/system/files/documents/public/nrc/convocation-mcclelland-hubin.pdf>

Summary of issues with measures released by NRC – see following pages

Also posted as <http://www.colorado.edu/pba/nrc/NRCMeasuresIssuesSummary.pdf>

Published measures about programs, with issues - NRC 2010

Lou McClelland, U of Colorado Boulder, 10/7/10, 2/8/11, 2/28, 3/20/11

Measure	NRC #	In rankings	Source	Issues w measure ("dated" not listed; could refer to any/all)	Issues in linear combos/rankings including weights from the "importance" section of faculty questionnaire	Source detail
Publications per Allocated Faculty	V 1	Overall; Rsrch	NRC	Matching faculty names w pubs/books, journal lists & coverage, approp of denominator, totals meaningful too, effect of young faculty w/o same chance, no demo of validity, appears books mixed into some non-humanities programs	Books are key for many NON-humanities depts but not used for them; mix of pub & books differs by program w/i field. ALSO: Most ranking systems use total pubs, not per-faculty; NRC rating questionnaire did say per faculty	ISI (2000-2006) Humanities: N pub'd books and articles, from CVs, 1986 to 2006; books weighted 5:1
Cites per Publication	V 2	Overall; Rsrch	NRC	See pubs. Also: 2002 inadvertently omitted from pubs and/or cites in 2010 release, to be fixed in 2011 release.	See pubs	Average number of citations in 2000-2006 to papers published in 1981-2006, per Allocated Faculty in 2006. Not used in humanities, computer science
Pct Faculty with Grants	V 3	Overall; Rsrch	Faculty Q	Not \$\$, self-report, based on survey respondents only, totals meaningful too	Fac Q said "Receipt of extramural grants for research" -- poor match to measure used; totals meaningful too.	(E1): Is your work currently supported by an extramural grant or contract? (yes or no). Based only on respondents to faculty questionnaire.
Pct Faculty Interdisciplinary	V 4	Overall	Faculty list	No credit for own core/new faculty who are listed in other programs. A program can have tons of interdisciplinary activity w/o listing any associate. Levels differ over institutions w/o apparent real bases, and depend on institutional org. Second+ fields listed by programs never published. Penalizes the designed-to-be-interdisciplinary pgm w its own faculty.	Fac Q said "Involvement in interdisciplinary work" -- not at all the same. AND relationship to quality could be curvilinear.	N of faculty categorized as "associate" / total N listed in program. Other measures from program Q not used.
Pct Non-Asian Minority Faculty	V 5	Overall; Div	Program Q	Small N of core/new in some interdisciplinary programs makes goofy. See also measure 11 for students.	Relationship to desirable diversity might be curvilinear.	(B7); [See q. for students, measure 11] - Pct of core and new
Pct Female Faculty	V 6	Overall; Div	Program Q	Ditto	Some fields (e.g. nursing) are majority female; what does an "important" weight mean in those? AND, might be curvilinear	Pct of core and new

Measure	NRC #	In rankings	Source	Issues w measure ("dated" not listed; could refer to any/all)	Issues in linear combos/rankings including weights from the "importance" section of faculty questionnaire	Source detail
Awards per Allocated Faculty	V 7	Overall; Rsrch	NRC	Matching faculty names to awards, coverage, no demo of validity		NRC from a review of 1,393 awards and honors from various scholarly organizations for names on faculty lists. Honors and awards were differentiated between "highly prestigious" awards, which received a weight of 5, and other awards, which received a weight of 1.
Average GRE	V 8	Overall	Program Q	Many programs say they use both V and Q. Set to field average for programs not using or reporting GRE's.	Fac Q said "GRE" period; a classic "input" measure not necessarily reflecting what a program does	Verbal and quantitative collected for 3 entering classes. Verbal alone used for humanities, quantitative alone for all others.
Pct 1st yr. Students with Full Support	V 9	Overall; Stu support	Program Q	Not comparable, no definition of "full" (we know this ranged widely), some figures from central records, some from statements by pgms	Fac Q did not restrict to first-year. AND, if some students elect to fund themselves, should that reflect on a program?	(E5): How many full-time first-year doctoral students 2005-06 had "full financial support"? [full support not defined]
Pct 1st yr. Students with External Funding	V10	Overall	Program Q	See pct w full support	Fac Q said "portable fellowships"	(E8) N first yr students w full support from externally-funded fellowships either alone or in combo w internal funding.
Pct Non-Asian Minority Students	V11	Overall; Div	Program Q	Excluding unknown from denominator threatens comparability.	Relationship to desirable diversity might be curvilinear.	(C9c): Of the doctoral students who were U.S. citizens or permanent residents, how many were... White, Non-Hispanic; Black, Non-Hispanic; Hispanic; Asian or Pacific Islander; American Indian or Alaska Native; Race/ethnicity Unknown. Unknown apparently excluded from denominator.
Pct Female Students	V12	Overall; Div	Program Q		Relationship to desirable diversity might be curvilinear.	
Pct International Students	V13	Overall; Div	Program Q		Relationship to desirable diversity definitely curvilinear -- too high = inability to attract US students	

Measure	NRC #	In rankings	Source	Issues w measure ("dated" not listed; could refer to any/all)	Issues in linear combos/rankings including weights from the "importance" section of faculty questionnaire	Source detail
Average PhDs 2002 to 2006	V14	Overall	Program Q	Averaging only non-zero years is weird. [And - available from IPEDS if use CIPs to characterize programs]	Relationship to quality might be curvilinear - if want not too big and not too small, no way to indicate on Fac Q.	(C1): ****Average over non-zero years, so series 2-0-0-0-3 = 5/2 = 2.5, not 5/5 = 1.
Pct Completing within 6 Years (8 years Humanities)	V15	Overall; Stu support	Program Q	Not comparable if some students entering from master's or from other programs	Relationship to quality could be curvilinear. Too high = too easy.	(C16 and C17) average completion rate = average over 3 cohorts for the humanities and over 5 cohorts for other fields. Cohorts = humanities 96-97 thru 98-99, others 96-97 thru 00-01
Time to Degree Full and Part Time	V16	Overall; Stu support	Program Q	See completion rate	Relationship to quality could be curvilinear. Too short = too easy.	(C2): Of the doctoral graduates who received doctoral degrees in the period 2003-2004 through 2005-2006, what was the median time to degree?
Pct Students in Academic Positions	V17	Overall; Stu support	NRC	Self-report before graduation (on SED); denominator in 2010 release is all PhDs not just respondents (may change in 2011 release); NRC matching to programs; inaccurate when a school has >1 pgm in an SED field	Fac Q said "Placement of students after graduation" -- <u>not</u> "academic" placement. Very different!	Doctoral records file from Survey of Earned Doctorates (SED), 2001 thru 2005. Academic includes post-docs. Proxied at field average if not available. Used but not displayed if N is 4 or less.
Student Work Space	V18	Overall	Program Q	99% and 0% both --> no	Is 100% the only desirable situation?	(D12): Approximately what % of doctoral students have a workspace for their exclusive use? 100% = yes, all other = no.
Health Insurance	V19	Overall	Institutional Q			(A1): Is university-supported health care insurance part of the financial support provided to enrolled doctoral students? (yes/no). Same for all programs in an institution.
Number of Student Activities Offered	V20	Overall	Program Q	Uncertain comparability	Is more necessarily better?	(D8): N offered by program or campus, of 18 listed. E.g., orientation, grant-writing instruction

Measure	NRC #	In rankings	Source	Issues w measure ("dated" not listed; could refer to any/all)	Issues in linear combos/rankings including weights from the "importance" section of faculty questionnaire	Source detail
Program Collects Outcomes Data	V21	Stu support	Program Q	Self-report yes/no	Is "collecting" really the point?	(D10): Does your program collect data about employment outcomes for all of your doctoral graduates? (yes/no)
Total Faculty		-	Faculty list	Instructions confusing and relied on records not readily available in many schools.		Faculty list, new + core + associated. Fall 2006
Number of Allocated Faculty		-	Faculty list	Confusing instructions for overrides exacerbated non-comparability of lists.		Faculty list plus allocation algorithms and institutional overrides
Assistant Professors as Pct of Total Faculty		-	Faculty list	More meaningful as pct of tenured/tenure-track.		Faculty list
Tenured Faculty as a Pct of Total Faculty		-	Faculty list	Ditto		Faculty list
Number of Core and New Faculty		-	Faculty list	Use of "core" and "new" and "associate" may have differed across programs		Faculty list
Number of Students Enrolled		-	Program Q	Easy only if institution/program can unambiguously discriminate students intended doctorates from those intending master's. Many cannot do so.		(C8): How many doctoral students, whether or not they were yet admitted to candidacy, were enrolled in your program during fall of 2005?
Average Annual First Year Enrollment		-	Program Q	Same issues as N enrolled.		(C3) For 2002-2006
Pct of Students with Research Assistantships		-	Program Q	Calc'd only from % with RA's <u>only</u> ; those with combinations including RA's excluded. Affects programs differentially. May change in 2011 release.		(E7) 5-year pattern, or (probably) E8, types of support; derivation not clear. Fall 2005
Pct of Students with Teaching Assistantships		-	Program Q	See pct with RA's - same issue here.		(E7) 5-year pattern, or (probably) E8, types of support; derivation not clear. Fall 2005