Journal of Management Inquiry
2015, Vol. 24(3) 340–344

DOI: 10.1177/1056492615574230
jmi.sagepub.com
SAGE

Assessing Candidate Quality: Lessons From Ethnography (and Accountants)

Michael G. Pratt[1]

Abstract

This essay examines how we as academics assess the quality of our colleagues through the lens of how we hire them. I begin by reviewing some critiques that have arisen regarding the use of largely quantitative critieria for assessment, and suggest that we might learn something about assessing quality by viewing it as a cultural issue. Building on insights from ethnograpy, I then offer some suggestions for how to build upon, and hopefully improve, our current practices.

Keywords

quality, hiring, ethnography

If you have never felt that good prospective doctoral students or qualified faculty members failed to get a position they deserved, then you can put this article away now. If you are happy with how you hire colleagues, then spend your time buying a lottery ticket instead of reading this article, for you are one lucky person. I realize that no system is perfect, that bad things happen to good people, and that at times, life is not fair. However, there is increasing evidence that there are some systemic issues with how we, as a field of management and organizational professionals, evaluate the quality of our prospective colleagues. Such criteria may be incomplete at best, or more seriously, deeply flawed. While I realize that we assess the quality of our colleagues at various points in their careers, one of the most critical is when we are determining whether a person, relative to others, is worthy to hire. If an academic doesn't get a position, later assessments (e.g., tenure) are moot.

There have been numerous repudiations of what is referred to as an "audit culture" in our profession by our peers (Walsh, 2011, p. 217; see also Baum, 2012; Macdonald & Kam, 2011), and I confess that my initial motivation for this "Provocations and Provocateurs" entry was to spend the entire article on the numerous ways we get it wrong when it comes to making critical judgments about assessing the quality of scholars and scholarship at critical junctures (e.g., hiring, annual reviews, promotion, tenure). However, a few things quickly became apparent. First, we already know a lot about what is wrong with such assessments—and have known them for some time. I will add a bit to what has been said, but on the whole, I do not think I have much new to add. Second, the topic is enormous, so I picked hiring as an example of the issues we face writ large. Third, there is much more written about problems than solutions and I was raised to not complain unless I had some idea about how to make things better. So I decided to ask myself what, if anything, I could add about addressing shortcomings regarding how we assess people—especially during hiring. The net result of these ponderings, dear reader, is in your hands (or on your screen) right now.

In preparing to write this article, it dawned upon me that despite numerous critiques about the over-reliance on largely (faulty) quantitative measures, few (if any) have stepped back and have asked what kind of assessments about candidates are we trying to make, and what skills can we draw upon to make them? In doing so, I realized that we may not think enough about how assessing data on a job candidate is like assessing data in a qualitative research project. While not a perfect analogy, the quality of a job candidate does involve the gathering and analysis of archival data (e.g., publications) and cultural artifacts—especially in terms of what is considered "countable" in making an assessment. I use the term *cultural artifacts* because what we decide to count (e.g., journal articles) and not count (or to count less, such as books) is specific to our professional culture. Other professions (e.g., physicians) may count different things (e.g., proportion of surgeries that patients survive). Moreover, we collect a lot of data from conversations and informal interviews (e.g., with candidate colleagues and mentors)—even before we decide whether or not to invite someone out for formal interview. In addition, like an inductive study, we often have large amounts of different types of data in front of us that need to be combined and interpreted in some fashion.

[1]Boston College, Chestnut Hill, MA, USA

Corresponding Author:
Michael G. Pratt, Carroll School of Management, Boston College, 424C Fulton Hall, Chestnut Hill, MA 02467-3808, USA.
Email: michael.pratt@bc.edu

While hiring can be approached as a more deductive (H1: People from ____ institution are good) or inductive process, I hope you can agree there are at least some parallels between assessing candidates and the assessment of qualitative data. If so, then I am going to take the next logical leap and suggest that lessons from qualitative methods may help enrich how we assess quality in our profession. Before suggesting such lessons, however, I think it makes sense to provide some evidence that such lessons are needed.

"Houston, We He Have a Problem"

Initial assessments of prospective faculty member quality, at least in the research institutions with which I am familiar, rely on *at least* the following "Big 4" criteria:

1. How many top publications a candidate has (relative to career stage);
2. How many citations a candidate has accumulated (again, relative to career stage);
3. The quality of the letters written on behalf of the candidate;
4. The quality of writing samples (e.g., "job market paper").

The last two criteria are the most qualitative of the Big 4. However, even these often involve quantitative assessments. The quality assessment of a paper, for example, is often at least partially based on the quality of the journal where it appears and the citations it garners. Criteria for assessing the quality of letters include, in part, the ranking of the letter writers' institutions. Certainly there are other criteria, such as the formal job interview, which is important downstream in the assessment process.[1] I focus on these four criteria, however, not only due to their prevalence but also because they are used in other assessments of quality in our profession (e.g., promotion and tenure).

These criteria are common enough that each has been criticized to varying degrees:

- The issues with the number of top publications have been twofold. First, there can be disagreement with what counts as a good journal from school to school. Some suggest that this issue can be addressed by looking at the journal's impact factor. But this leads to a second issue: Impact factor scores are problematic because they can be "gamed" such that journals can artificially increase their scores (Davis, 2014).
- The use of citation counts of a candidate's articles have also been found wanting. For example, (a) existing measures are not always accurate; (b) they are an incomplete measure of impact as they only take into account one set of stakeholders (other academics); and (c) like journal impact scores, can be artificially inflated (e.g., Aguinis, Shapiro, Antonacopoulou, & Cummings, 2014; Baum, 2012; Davis, 2014). In addition, citation counts—especially for co-authored pieces—may be largely dependent on the fame and network strength of a scholar's co-authors, and may not say anything about the candidate's unique contribution to the manuscript. Counts may also be influenced by the size of the "built in" audience for a given paper. All things being equal, a highly impactful piece in an established field would likely have a higher potential for garnering citations than a piece that is carving out a new field of research -at least in the short horizons we often use to assess others. The impact of the latter may eventually be even bigger than the former, but it may take many more years to assess.
- The issues of letters are twofold: What they say and where they are from.[2] First, whether motivated by potential litigation or not wanting to disadvantage a candidate vis a vis others, letter writers may be overly positive about a given candidate's potential or impact on the field, leaving the reader in the uncomfortable position of parsing out degrees of positivity. Second, letters are also assessed by the quality of the institutions of the letter writers.[3] However, just like each of the other criteria noted above, these assessments can be "gamed" (see Tushman, 2009). Even prestigious schools may even falsify data to look better in their rankings (Markus, 2013).
- Finally, some have discussed the assessment of the quality of writing samples, arguing that the quality of the publication outlet is often used as a proxy for the quality of the journal article itself. As noted above, this can be problematic given issues with journal rankings. Moreover, the quality of the journal does not necessarily reflect the quality of any single submission; that is, even top journals can publish less-than-stellar papers (Baum, 2012). Depending on the methods used, one might also assess the quality of a paper based on the significance-level of its findings. But one need only to look at one of our "feeder" disciplines, social psychology, to observe dangers of over-relying on *p* values for significance (Nuzzo, 2014).

Quality Criteria: "The Good, the Bad, and the Ambivalent"

At best, the use of these criteria can be an attempt at fairness by providing common metrics on which all candidates can be assessed because such metrics (e.g., number of publications) may help minimize the over-politicization of the assessment process. If your department is looking for a "best athlete," for example, then it is easier to have a standard metric than to simply go with personal preferences (though even here some

may argue that quantity of output may vary depending on a candidate's chosen methodology).

At worst, if the critics are indeed correct, these criteria can each be deeply flawed and may even promote unethical behaviors on the part of candidates (e.g., such as manipulating data—see Callaway, 2011; Ferguson, 2012), and as noted above, institutional actors. Moreover, just as stringing together several weak studies does not result in a superior one, it seems unlikely that a string of questionable criteria will somehow lead to a superior assessment. This is especially true because the criteria are not independent. For example, evaluation of business schools in the United States and abroad is based, at least in part, on the research productivity (e.g., top-level publications) of their faculty. In the United Kingdom, the Research Excellence Framework (http://www.ref.ac.uk/), which assesses the quality of research institutions, is largely based on faculty publications, which has led, at least anecdotally, to a surge in the hiring of faculty with top-tier publications prior to the assessment. Thus, journal impact, scholarly impact, and institutional quality seem inherently bound together. So what to do?

Responses to the current way we assess the quality of potential colleagues seem to reflect known responses to ambivalence. For many, the problem seems too big, so the response is paralysis—do nothing to create change. Others call for an overhaul of our current "system" (see Aguinis & colleagues', 2014, thoughtful revision of impact scores), while others seem to call for scrapping the entire system altogether and starting over. The latter seems premature, however, because we don't have good alternatives to replace what we currently have. Another response, and one that I am advocating, is to recognize the criteria for what they are (cultural artifacts), and—at least in the short run—to supplement what data we have with other types of (albeit, still culturally influenced) data. Toward that end, I would like to add to the conversation about assessing quality by incorporating some lessons learned from inductive qualitative research, and in particular, ethnography.

Lessons From Ethnography

I focus on ethnography because ethnography, broadly defined, is the study of culture and the problems we face as a field appear to be at grounded, at least in part, in our culture: a culture expressed in how we reward others, how we socialize others, and as noted, how we count. In addition, one of the key reasons for doing an ethnography is that the researcher does not have all the answers and expects that wisdom will be found in consulting other people . . . and for me in this article, "people" even includes an accountant! Accompanying these lessons are suggested action items, some of which are big and some smaller. Overall, my main intention is to suggest additional data for "triangulating" our assessments of worth and quality.

Lesson 1: Understand Your Own Assumptions

One of the lessons I am constantly faced with when doing inductive qualitative research in a particular setting is that what I think is important is often not what the people in that setting feel is important. The same can be true of our assessments of what makes for a quality academic. I am imprinted with a particular University-of-Michigan-organizational-psychology view of what a quality candidate looks like: someone who asks "big, messy, and important questions" and has the potential to produce ground-breaking research. Thus, from a quantitative standpoint, I am often fine with a lower number of top publications if the research is creative—in that it is both novel and useful (subjective assessments to be sure). Although my views have changed over the years, I was also taught that while teaching and professional service are important, they are often weighted less than research. As a reader, you might say, "that's what I think" or "wow, that is really wrong." Both responses are understandable. Different people and different schools have different views of what "quality" is. My sense is that people tend to view quality from their own cultural viewpoints and that what is important to each of us may be different—even if we use the same terms like *high potential* or *best athlete* or even *best fit*.

Actionable suggestion: Question your own assumptions

If you are hiring *and* making a long-term commitment to a candidate, it perhaps makes sense to find out whether others in your unit have similar views regarding what makes a quality candidate. If people disagree, you could decide to alternate the criteria you use, and/or begin the longer process of managing value (or identity) conflicts. Either way, it helps to have some agreed upon (even if temporarily) standards from which to hire. If you want to be really radical, you could even communicate these quality standards to the candidates you are assessing. It would also help the quality assessors if the PhD granting institutions communicated how they view quality, and how they attempted to achieve it in their programs. (Similarly, it would also be tremendously helpful in the tenure letter writing process if the letter writers were allowed some insight into how quality is viewed at the tenure-granting institution).

Lesson 2: Seek the "Voice" of the People You Are Studying

To write a really good ethnography, you have to immerse yourself in the context of those whom you are studying. This is because the purpose of doing an ethnography is to understand the "voice" of the people you are studying: to hear how they understand their world in their own words and actions. It is for this reason that ethnographers refer to these people as

"informants" or those that provide you information versus a "subject" whom you are manipulating or a "respondent" who is just answering your questions. In an ethnographic study, one often tries to understand this "voice" by working alongside—for an extended period of time—the people from whom you are learning.

When assessing the quality of a candidate, such an immersive experience is unlikely, but there are things we already do to get at the "voice" of a candidate. Perhaps the most obvious is conducting early and informal interviews with the candidates and gathering and reviewing artifacts and archives (e.g., publications, papers, cover letters, research and teaching statements). Others may attempt to understand a candidate better by talking with his or her colleagues, reflecting on personal interactions with her or him, or gathering other types of qualitative data. For these data to be useful, however, *they must be used as a window into how the candidate views the world*, not on how the data support the assessor's own worldview. Given the sheer number of candidates, it is unlikely that any given assessor or assessing group will have the time or resources to immerse deeply into the worldview of many (if any) candidates. In fact, it is for these reasons that some rely on the number of top publications and citation counts.

Actionable suggestions

Last summer, at the Academy of Management meetings, I was in a conversation about assessing candidates' quality via the number of their publications and citations and someone shouted out, "Why don't people just read their damn papers?" While a glib remark, it seems essential that people read what the candidates write. It is even better, though unlikely if the candidate is a freshly minted PhD, that there is a solo-authored piece available so as to better get at the "voice" of the candidate. Another, more radical, suggestion for increasing immersion came from an accounting colleague of mine. He made an offhand remark in a conversation about the growth of a "Rookie Camp" in his field—a conference to assess academic accountants seeking employment. According to their website, this Rookie Camp has the following goals:

> The 2014 Accounting PhD Rookie Recruiting & Research Camp is a two-day forum for faculty and recruiters to meet and network with PhD candidates, attend 15-minute research presentations by job-seeking candidates, and interview a number of candidates interested in faculty or private sector position. (http://aaahq.org/meetings/RookieCamp/2014/faq.cfm)

Recruiters can sign up to attend any number of talks, which happen concurrently. Often, a school will send a delegation of faculty to attend multiple talks occurring at the same time. Before the talk, recruiters have access to a "bio book" that contains a resume and other information. The recruiter can, based on the bio book or the presentation, sign up to interview a candidate. Taken together, this allows assessors to more directly access the "voice" of a greater number of candidates who can talk about their dissertations in their own words; while also providing some time to immerse themselves more deeply in conversations with preferred candidates.

The growth of this conference started with one university—along with a planning committee from multiple schools—initiating and implementing it. Over time, other schools joined. Initially, elite schools did not send candidates. However, eventually they came to send candidates and even to participate as recruiters. As the conference grew, the American Accounting Association also came to endorse the conference. Since its founding in 2008, this conference has grown into a mainstay for recruiting accounting PhDs. What if we did something similar under the auspices of the Academy of Management, or even a brave consortium of schools? I mentioned this to a recently hired assistant professor and his reaction was, "sounds great for recruiters, but a nightmare for candidates." However, I would argue that this forum gives a broader number of candidates a shot at being heard.

Lesson 3: Understanding and Learning From Others Takes Time

Learning from others and understanding the "voice" of a candidate takes a lot of time. While I think the "Rookie Camp" allows for a relatively efficient use of time, it still involves more than a 2-day commitment. For some academics, adding this request onto teaching, administrative duties, mentoring, writing, and possibly even non-work commitments may seem overwhelming. Given the investment one makes in hiring a candidate (at many universities, this may well be more than a million dollar commitment before contract renewal, and you may also be hiring a colleague for life), one could argue that the time is well-spent. But it is still time spent.

Actionable items

Investment of time is a given and there are few shortcuts. But at my university, faculty who attend Rookie Camp are often excused (if they want to be) from other duties during recruiting. However, the only way that "sharing the pain" like this is going to produce good results is if the assessors have some idea of how each is assessing quality (see Lesson 1). Otherwise, you could be using the same assessment language, but be talking about very different things.

Concluding Thoughts

I want to close with two thoughts. First, I opened this article by asking personal questions about what you, the reader,

have or have not experienced. I did so to underscore a point: While critiques such as these are often aimed at institutions, let's not forget that institutions are "inhabited." How we assess quality influences people's lives. Thus, changes we make (or do not make) not only reflect who we are as professionals but also how we treat each other. Second, I have never before written a paper that I hope is quickly passé. But if people could re-read the first two sentences, and based on their answers could honestly set this article aside, I would be incredibly gratified.

Acknowledgments

I would like to thank Denny Gioia as well as Teresa Amabile, Michel Anteby, Sue Ashford, Blake Ashforth, Jean Bartunek, Stephanie Creary, Spencer Harrison, Doug Lepisto, and Sugata Roychowdhury for their comments and support with regard to this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. This list is certainly not all-inclusive. At times, assessors will engage in "cohort analyses" to compare a candidate vis a vis his or her peers. Others may argue for a measure of "fit" with a particular group at a particular time. There may be other criteria as well (e.g., progress on dissertation for a candidate's first job), but in the interest of space as well as the reasons provided in the body of this essay, I only address these four.
2. Institutional ranking also comes into play directly when assessing the quality of an applicant. However, institutional rankings are likely based on the quality of MBA programs, not doctoral programs (and one could argue that given limited resources, strength in one area my not necessarily translate to strength in the other).
3. The letter writer is, of course, also assessed. Interestingly, scholars have even called into question what it means to be a "star" (i.e., highly published) in our profession. While reputation may reflect quality work, some suggest it may also be a product of luck, political considerations, or may even be a statistical artifact of a skewed distribution (Baum, 2012; Macdonald & Kam, 2011). But a fuller consideration of this topic would, in itself, be an entire essay.

References

Aguinis, H., Shapiro, D., Antonacopoulou, E., & Cummings, T. (2014). Scholarly impact: A pluralist conceptualization. *Academy of Management, Learning & Education*, *13*, 623-639.

Baum, J. (2012). The skewed few: Does "skew" signal quality among journals, articles, and academics? *Journal of Management Inquiry*, *21*, 349-354.

Callaway, E. (2011). Report finds massive fraud at Dutch universities. *Nature*, *479*, 15.

Davis, G. (2014). Editorial essay: Why do we still have journals? *Administrative Science Quarterly*, *59*, 193-201.

Ferguson, C. (2012, July 12). Can we trust psychological research? *Time.com*. Retrieved from http://ideas.time.com/2012/07/17/can-we-trust-psychological-research/print/

Macdonald, S., & Kam, J. (2011). The skewed few: People and papers of quality in management studies. *Organization*, *18*, 467-475.

Markus, J. (2013, March 20). Caught cheating: Colleges falsify admissions data for higher rankings. *NBCnews.com*. Retrieved from http://investigations.nbcnews.com/_news/2013/03/20/17376664-caught-cheating-colleges-falsify-admissions-data-for-higher-rankings

Nuzzo, R. (2014). Statistical errors. *Nature*, *506*, 150-152.

Tushman, G. (2009). *Wannabe U: Inside the corporate university*. Chicago, IL: University of Chicago Press.

Walsh, J. (2011). 2010 Presidential address: Embracing the sacred in our secular scholarly world. *Academy of Management Review*, *36*, 215-234.

Author Biography

Michael G. Pratt (PhD, University of Michigan) is the O'Connor Family Professor in the Carroll School of Management at Boston College. His research interests are cross-level and include how people connect to their organizations, professions, and work. Theoretically, his research draws heavily from theories of identity and identification, meaning, emotion, intuition, and culture (e.g., artifacts). Methodologically, although he has published work that utilizes lab research and surveys, much of his work is ethnographic or otherwise qualitative in nature. His recent research projects center on meaningful work, collective ambivalence, professional identity work, and the social construction of trust among firefighters.